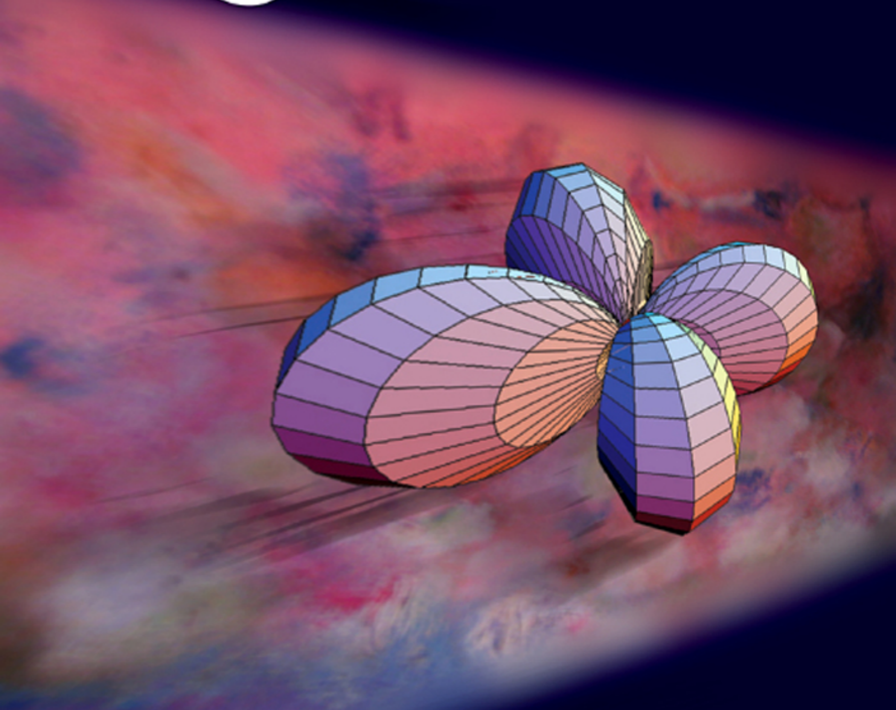
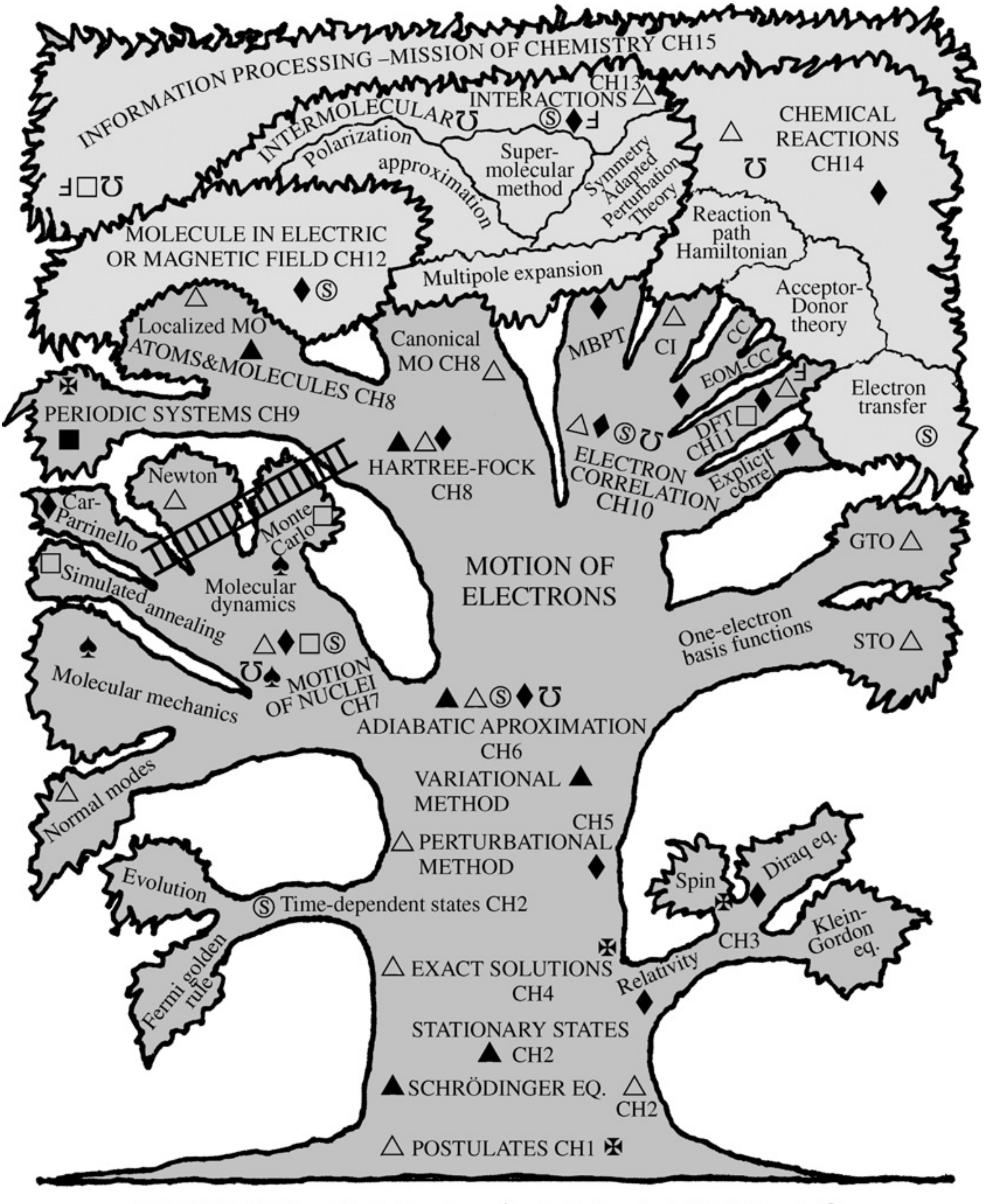
ELSEVIER

IDEAS OF QUANTUM CHEMISTRY



LUCJAN PIELA

IDEAS OF
QUANTUM CHEMISTRY


INFORMATION PROCESSING –MISSION OF CHEMISTRY CH15
Ⅎ□℧
INTERMOLECULAR ℧ INTERACTIONS CH13 △ Ⓢ♦Ⅎ
Polarization approximation
Super-molecular method
Symmetry Adapted Perturbation Theory
CHEMICAL REACTIONS CH14 △ ℧ ♦
Reaction path Hamiltonian
Acceptor-Donor theory
Electron transfer Ⓢ
MOLECULE IN ELECTRIC OR MAGNETIC FIELD CH12 ♦Ⓢ
Multipole expansion
△ Localized MO ▲
ATOMS&MOLECULES CH8
Canonical MO CH8 △
♦ MBPT
△ CI
CC
EOM-CC ♦
DFT □ ♦ △ Ⅎ
CH11
Explicit correl ♦
△♦Ⓢ℧ ELECTRON CORRELATION CH10
✠ PERIODIC SYSTEMS CH9 ■
▲△♦ HARTREE-FOCK CH8
Newton △
♦ Car-Parrinello
Monte Carlo □
□ Simulated annealing
♠ Molecular dynamics
MOTION OF ELECTRONS
GTO △
One-electron basis functions
STO △
♠ Molecular mechanics
△♦□Ⓢ
℧♠ MOTION OF NUCLEI CH7
▲△Ⓢ♦℧ ADIABATIC APROXIMATION CH6
△ Normal modes
VARIATIONAL METHOD ▲
CH5
△ PERTURBATIONAL METHOD ♦
Evolution
Ⓢ Time-dependent states CH2
Fermi golden rule
Spin
Diraq eq.
✠ ♦
CH3
Klein-Gordon eq.
✠
△ EXACT SOLUTIONS CH4
Relativity ♦
STATIONARY STATES ▲ CH2
▲ SCHRÖDINGER EQ. △ CH2
△ POSTULATES CH1 ✠

WHERE ARE WE?

IDEAS OF QUANTUM CHEMISTRY

by

LUCJAN PIELA

Department of Chemistry University of Warsaw
Warsaw, Poland

ELSEVIER

Amsterdam • Boston • Heidelberg • London • New York • Oxford
Paris • San Diego • San Francisco • Singapore • Sydney • Tokyo

ELSEVIER
Radarweg 29, PO Box 211, 1000 AE Amsterdam, The Netherlands
The Boulevard, Langford Lane, Kidlington, Oxford OX5 1GB, UK

First edition 2007

Library of Congress Cataloging-in-Publication Data
A catalog record for this book is available from the Library of Congress

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

ISBN-13: 978-0-444-52227-6
ISBN-10: 0-444-52227-1

Printed and bound in The Netherlands

07 08 09 10 11   10 9 8 7 6 5 4 3 2 1

To all whom I love

CONTENTS

... "to see a world in a grain of sand
and Heaven in a wildflower
hold infinity in the palm of your hand
and eternity in an hour ... "

William Blake *"Auguries of Innocence"*

INTRODUCTION

Our wonderful world

Colours! The most beautiful of buds – an apple bud in my garden changes colour from red to rosy after a few days. Why? It then explodes into a beautiful pale rosy flower. After a few months what was once a flower looks completely different: it has become a big, round and red apple. Look at the apple skin. It is pale green, but moving along its surface the colour changes quite abruptly to an extraordinary vibrant red. The apple looks quite different when lit by full sunlight, or when placed in the shade.

Touch the apple, you will feel it smooth as silk.

How it *smells*! An exotic mixture of subtle scents.

What a *taste*: a fantastic juicy pulp!

Sounds ... the amazing melody of a finch is repeated with remarkable regularity. My friend Jean-Marie André says it is the same here as it is in Belgium. The same? Is there any program that forces finches to make the same sound in Belgium as in Poland? A woodpecker hits a tree with the regularity of a machine gun, my Kampinos forest echoes that sound. Has the woodpecker also been programmed? What kind of program is used by a blackbird couple that forces it to prepare, with enormous effort and ingenuity, a nest necessary for *future* events?

What we do know

Our senses connect us to what we call the Universe. Using them we feel its presence, while at the same time we are a part of it. Sensory operations are the direct result of interactions, both between molecules and between light and matter. *All* of these phenomena deal with chemistry, physics, biology and even psychology. In these complex events it is impossible to discern precisely where the disciplines of chemistry, physics, biology, and psychology begin and end. Any separation of these domains is artificial. The only reason for making such separations is to focus our attention on *some* aspects of one indivisible phenomenon. Touch, taste, smell, sight, hearing, are these our only links and information channels to the Universe? How little we know about it! To feel that, just look up at the sky. A myriad of stars around us points to new worlds, which will remain unknown forever. On the other hand, imagine how incredibly complicated the chemistry of friendship is.

We try to understand what is around us by constructing in our minds pictures representing a "reality", which we call models. Any model relies on our perception of reality (on the appropriate scale of masses and time) emanating from our experience, and on the other hand, on our ability to abstract by creating ideal beings. Many such models will be described in this book.

It is fascinating that man is able to magnify the realm of his senses by using sophisticated tools, e.g., to see quarks sitting in a proton,[1] to discover an amazingly simple equation of motion[2] that describes both cosmic catastrophes, with an intensity beyond our imagination, as well as the flight of a butterfly. A water molecule has exactly the same properties in the Pacific as on Mars, or in another galaxy. The conditions over there may sometimes be quite different from those we have here in our laboratory, but we *assume* that if these conditions could be imposed on the lab, the molecule would behave in exactly the same way. We hold out hope that a set of universal physical laws applies to the entire Universe.

The set of these basic laws is not yet complete or unified. Given the progress and important generalizations of physics in the twentieth century, much is currently understood. For example, forces with seemingly disparate sources have been reduced to only three kinds:

- those attributed to *strong interactions* (acting in nuclear matter),
- those attributed to *electroweak interactions* (the domain of chemistry, biology, as well as β-decay),
- those attributed to *gravitational interaction* (showing up mainly in astrophysics).

Many scientists believe other reductions are possible, perhaps up to a single fundamental interaction, one that explains Everything (quoting Feynman: the frogs as well as the composers). This assertion is based on the conviction, supported by developments in modern physics, that the laws of nature are not only universal, but simple.

Which of the three basic interactions is the most important? This is an ill conceived question. The answer depends on the external conditions imposed (pressure, temperature) and the magnitude of the energy exchanged amongst the interacting objects. A measure of the energy exchanged[3] may be taken to be the percentage of the accompanying mass deficiency according to Einstein's relation $\Delta E = \Delta mc^2$. At a given magnitude of exchanged energies some particles are stable.

---

[1] A proton is 10^{15} times smaller than a human being.

[2] Acceleration is directly *proportional* to force. Higher derivatives of the trajectory with respect to time do not enter this equation, neither does the nature or cause of the force. The equation is also invariant with respect to any possible starting point (position, velocity, and mass). What remarkable simplicity and generality (within limits, see Chapter 3)!

[3] This is also related to the areas of operation of particular branches of science.

Strong interactions produce the huge pressures that accompany the gravitational collapse of a star and lead to the formation of neutron stars, where the mass deficiency approaches 40%. At smaller pressures, where individual nuclei may exist and undergo nuclear reactions (strong interactions[4]), the mass deficiency is of the order of 1%. At much smaller pressures the electroweak forces dominate, nuclei are stable, atomic and molecular structures emerge. Life (as we know it) becomes possible. The energies exchanged are much smaller and correspond to a mass deficiency of the order of only about $10^{-7}\%$. The weakest of the basic forces is gravitation. Paradoxically, this force is the most important on the macro scale (galaxies, stars, planets, etc.). There are two reasons for this. Gravitational interactions share with electric interactions the longest range known (both decay as $1/r$). However, unlike electric interactions[5] those due to gravitation are not shielded. For this reason the Earth and Moon attract each other by a huge gravitational force[6] while their electric interaction is negligible. This is how David conquers Goliath, since at any distance electrons and protons attract each other by electrostatic forces, about 40 orders of magnitude stronger than their gravitational attraction.

Gravitation does not have any measurable influence on the collisions of molecules leading to chemical reactions, since reactions are due to much stronger electric interactions.[7]

A narrow margin

Due to strong interactions, protons overcome mutual electrostatic repulsion and form (together with neutrons) stable nuclei leading to the variety of chemical elements. Therefore, strong interactions are the prerequisite of any chemistry (except hydrogen chemistry). However, chemists deal with already prepared stable nuclei[8] and these strong interactions have a very small range (of about 10^{-13} cm) as compared to interatomic distances (of the order of 10^{-8} cm). *This is why a chemist may treat nuclei as stable point charges that create an electrostatic field.* Test tube conditions allow for the presence of electrons and photons, thus completing the set of particles that one might expect to see (some exceptions are covered in this book). This has to do with the order of magnitude of energies exchanged (under the conditions where we carry out chemical reactions, the energies exchanged exclude practically all nuclear reactions).

---

[4] With a corresponding large energy output; the energy coming from the fusion $\mathrm{D}+\mathrm{D}\rightarrow\mathrm{He}$ taking place on the Sun makes our existence possible.

[5] In electrostatic interactions charges of opposite sign attract each other while charges of the same sign repel each other (Coulomb's law). This results in the fact that large bodies (built of a huge number of charged particles) are nearly electrically *neutral* and interact electrically only very weakly. This dramatically reduces the range of their electrical interactions.

[6] Huge tides and deformations of the whole Earth are witness to that.

[7] It does not mean that gravitation has no influence on reagent concentration. Gravitation controls the convection flow in liquids and gases (and even solids) and therefore a chemical reaction or even crystallization may proceed in a different manner on the Earth's surface, in the stratosphere, in a centrifuge or in space.

[8] At least in the time scale of a chemical experiment. Instability of some nuclei is used in nuclear chemistry and radiation chemistry.

On the vast scale of attainable temperatures[9] chemical structures may exist in the narrow temperature range of 0 K to thousands of K. Above this range one has plasma, which represents a soup made of electrons and nuclei. Nature, in its vibrant living form, requires a temperature range of about 200–320 K, a margin of only 120 K. One does not require a chemist for chemical structures to exist. However, to develop a chemical science one has to have a chemist. This chemist can survive a temperature range of 273 K ± 50 K, i.e. a range of only 100 K. The reader has to admit that a chemist may think of the job only in the narrow range[10] of 290–300 K, only 10 K.

A fascinating mission

Suppose our dream comes true and the grand unification of the three remaining basic forces is accomplished one day. We would then know the first principles of constructing everything. One of the consequences of such a feat would be a catalogue of all the elementary particles. Maybe the catalogue would be finite, perhaps it would be simple.[11] We might have a catalogue of the conserved symmetries (which seem to be more elementary than the particles). Of course, knowing such first principles would have an enormous impact on all the physical sciences. It could, however, create the impression that everything is clear and that physics is complete. Even though structures and processes are governed by first principles, it would still be very difficult to predict their existence by such principles alone. The resulting structures would depend not only on the principles, but also on the initial conditions, complexity, self-organization, etc.[12] *Therefore, if it does happen, the Grand Unification will not change the goals of chemistry.*

Chemistry currently faces the enormous challenge of information processing, quite different to this posed by our computers. This question is discussed in the last chapter of this book.

BOOK GUIDELINES

TREE

Any book has a linear appearance, i.e. the text goes from page to page and the page numbers remind us of that. However, the *logic* of virtually any book is *non*-linear, and in many cases can be visualized by a diagram connecting the chapters that

---

[9] Millions of degrees.

[10] The chemist may enlarge this range by isolation from the specimen.

[11] None of this is certain. Much of elementary particle research relies on large particle accelerators. This process resembles discerning the components of a car by dropping it from increasing heights from a large building. Dropping it from the first floor yields five tires and a jack. Dropping from the second floor reveals an engine and 11 screws of similar appearance. Eventually a problem emerges: after landing from a very high floor new components appear (having nothing to do with the car) and reveal that some of the collision energy has been converted to the new particles!

[12] The fact that Uncle John likes to drink coffee with cream at 5 p.m. possibly follows from first principles, but it would be very difficult to trace that dependence.

(logically) follow from one another. Such a diagram allows for multiple branches emanating from a given chapter, particularly if the branches are placed logically on an equal footing. Such logical connections are illustrated in this book as a TREE diagram (inside front cover). This TREE diagram plays a very important role in our book and is intended to be a study guide. An author leads the reader in a certain direction and the reader expects to know what this direction is, why he needs this direction, what will follow, and what benefits he will gain after such study. If studying were easy and did not require time, a TREE diagram might be of little importance. However, the opposite is usually true. In addition, knowledge represents much more than a registry of facts. Any understanding gained from seeing relationships among those facts and methods plays a key role.[13] The primary function of the TREE diagram is to make these relationships clear.

The use of hypertext in information science is superior to a traditional linear presentation. It relies on a tree structure. However, it has a serious drawback. Sitting on a branch, we have no idea what that branch represents in the whole diagram, whether it is an important branch or a remote tiny one; does it lead further to important parts of the book or it is just a dead end, and so on. At the same time, a glimpse of the TREE shows us that the thick trunk is the most important structure. What do we mean by important? At least two criteria may be used. Important for the majority of *readers*, or important because the material is fundamental for an understanding of the *laws of nature*. I have chosen the first. For example, relativity theory plays a pivotal role as the foundation of physical sciences, but for the vast majority of chemists its practical importance and impact are much smaller. Should relativity be represented therefore as the base of the trunk, or as a minor branch? I have decided to make the second choice *not* to create the impression that this topic is absolutely necessary for the student. Thus, the trunk of the TREE corresponds to the pragmatic way to study this book.

The trunk is the backbone of this book:

- it begins by presenting Postulates, which play a vital role in formulating the foundation of quantum mechanics. Next, it goes through
- the Schrödinger equation for stationary states, so far the most important equation in quantum chemical applications,
- the separation of nuclear and electronic motion,
- it then develops the mean-field theory of electronic structure and
- finally, develops and describes methods that take into account electronic correlation.

The trunk thus corresponds to a traditional course in quantum chemistry for undergraduates. This material represents the necessary basis for further extensions into other parts of the TREE (appropriate for graduate students). In particular, it makes it possible to reach the crown of the TREE, where the reader may find tasty fruit. Examples include the theory of molecule-electric/magnetic field inter-

---

[13] This advice comes from Antiquity: ...“*knowledge is more precious than facts, understanding is more precious than knowledge, wisdom is more precious than understanding*”.

actions, as well as the theory of intermolecular interactions (including chemical reactions), which form the very essence of chemistry. We also see that our TREE has an important branch concerned with nuclear motion, including molecular mechanics and several variants of molecular dynamics. At its base, the trunk has two thin branches: one pertains to relativity mechanics and the other to the time-dependent Schrödinger equation. The motivation for this presentation is different in each case. I do not highlight relativity theory: its role in chemistry is significant,[14] but not crucial. The time-dependent Schrödinger equation is not highlighted, because, for the time being, quantum chemistry accentuates stationary states. I am confident, however, that the 21st century will see significant developments in the methods designed for time-dependent phenomena.

Traversing the TREE

The TREE serves not only as a diagram of logical chapter connections, but also enables the reader to make important decisions:

- the choice of a logical path of study (“itinerary”) leading to topics of interest,
- elimination of chapters that are irrelevant to the goal of study.[15]

Of course, all readers are welcome to find their own itineraries when traversing the TREE. Some readers might wish to take into account the author’s suggestions as to how the book can be shaped..

First of all we can follow two basic paths:

- *minimum minimorum* for those who want to proceed as quickly as possible to get an idea what quantum chemistry is all about[16] following the chapters designated by (▲).
- *minimum* for those who seek basic information about quantum chemistry, e.g., in order to use popular computer packages for the study of molecular electronic structure,[17] they may follow the chapters designated by the symbols ▲ *and* △.

Other paths proposed consist of a *minimum itinerary* (i.e. ▲ *and* △) plus special excursions: which we term **additional**.

- Those who want to use computer packages with molecular mechanics and molecular dynamics in a knowledgeable fashion, may follow the chapters designated by this symbol (♠).
- Those interested in spectroscopy may follow chapters designated by this symbol (Ⓢ).

---

[14] Contemporary inorganic chemistry and metallo-organic chemistry concentrate currently on heavy elements, where relativity effects are important.

[15] It is, therefore, possible to prune some of the branches.

[16] I imagine someone studying material science, biology, biochemistry, or a similar subject. They have heard that quantum chemistry explains chemistry, and want to get the flavour and grasp the most important information. They should read only 47 pages.

[17] I imagine here a student of chemistry, specializing in, say, analytical or organic chemistry (not quantum chemistry). This path involves reading something like 300 pages + the appropriate Appendices (if necessary).

- Those interested in chemical reactions may follow chapters designated by this symbol (℧).
- People interested in large molecules may follow chapters designated by this symbol (□).
- People interested in exact calculations on atoms or small molecules[18] may follow chapters designated by this symbol (♦).
- People interested in solid state physics and chemistry may follow chapters designated by this symbol (■).

For readers interested in particular aspects of this book rather than any systematic study, the following itineraries are proposed.

- Just before an exam read in each chapter these sections. "*Where are we*", "*An example*", "*What is it all about*", "*Why is this important*", "*Summary*", "*Questions*" and "*Answers*".
- For those interested in recent progress in quantum chemistry, we suggest sections "*From the research front*" in each chapter.
- For those interested in the future of quantum chemistry we propose the sections labelled, "**Ad futurum**" in each chapter, and the chapters designated by (Ⅎ).
- For people interested in the "magical" aspects of quantum chemistry we suggest sections with the label (✠).
  - Is the world real? We suggest looking at p. 38 and subsequent material.
  - For those interested in teleportation please look at p. 47 and subsequent material.
  - For those interested in the creation of matter, we suggest p. 134 and subsequent material.
  - For those interested in tunnelling through barriers, please look at p. 153 and subsequent material.

The target audience

I hope that the TREE structure presented above will be useful for those with varying levels of knowledge in quantum chemistry as well as for those whose goals and interests differ from those of traditional quantum chemistry.

This book is a direct result of my lectures at the Department of Chemistry, University of Warsaw, for students specializing in theoretical rather than experimental chemistry. Are such students the target audience of this book? Yes, but not exclusively. At the beginning I assumed that the reader would have completed a basic quantum chemistry course[19] and, therefore, in the first version I omitted the basic material. However, that version became inconsistent, devoid of several

---

[18] Suppose the reader is interested in an accurate theoretical description of small molecules. I imagine such a Ph.D. student working in quantum chemistry. Following their itinerary, they have, in addition to the minimum program (300 pages), an additional 230 pages, which gives about 530 pages plus the appropriate Appendices, in total about 700 pages.

[19] Say at the level of P.W. Atkins, "*Physical Chemistry*", sixth edition, Oxford University Press, Oxford, 1998, chapters 11–14.

fundamental problems. This is why I have decided to explain, mainly very briefly,[20] these problems too. Therefore, a student who chooses the *minimum* path along the TREE diagram (mainly along the TREE trunk) will obtain an introductory course in quantum chemistry. On the other hand, the complete collection of chapters provides the student with a set of advanced topics in quantum chemistry, appropriate for graduate students. For example, a number of chapters such as relativity mechanics, global molecular mechanics, solid state physics and chemistry, electron correlation, density functional theory, intermolecular interactions and theory of chemical reactions, present material that is usually accessible in monographs or review articles.

In writing this book I imagined students sitting in front of me. In discussions with students I often saw their enthusiasm, their eyes showed me a glimpse of curiosity. First of all, this book is an acknowledgement of my young friends, my students, and an expression of the joy of being with them. Working with them formulated and influenced the way I decided to write this book. When reading textbooks one often has the impression that all the outstanding problems in a particular field have been solved, that everything is complete and clear, and that the student is just supposed to learn and absorb the material at hand. In science the opposite is true. All areas can benefit from careful probing and investigation. Your insight, your different perspective or point of view, even on a fundamental question, may open new doors for others.

Fostering this kind of new insight is one of my main goals. I have tried, whenever possible, to present the reasoning behind a particular method and to avoid rote citation of discoveries. I have tried to avoid writing too much about details, because I know how difficult it is for a new student to see the forest through the trees. I wanted to focus on the main ideas of quantum chemistry.

I have tried to stress this integral point of view, and this is why the book sometimes deviates from what is normally considered as quantum chemistry. I sacrificed, not only in full consciousness, but also voluntarily "quantum cleanness" in favour of exposing the inter-relationships of problems. In this respect, any division between physics and chemistry, organic chemistry and quantum chemistry, quantum chemistry for chemists and quantum chemistry for biologists, intermolecular interactions for chemists, for physicists or for biologists is completely artificial, and sometimes even absurd. I tried to cross these borders[21] by supplying examples and comparisons from the various disciplines, as well as from everyday life, by incorporating into intermolecular interactions not only supramolecular chemistry, but also molecular computers, and particularly by writing a "holistic" (last) chapter about the mission of chemistry.

My experience tells me that the new talented student who loves mathematics courts danger. They like complex derivations of formulae so much that it seems that the more complex the formalism, the happier the student. However, all these formulae represent no more than an approximation, and sometimes it would be

---

[20] Except where I wanted to stress some particular topics.

[21] The above described itineraries cross these borders.

better to have a simple formula. The simple formula, even if less accurate, may tell us more and bring more understanding than a very complicated one. Behind complex formulae are usually hidden some very simple concepts, e.g., that two molecules are unhappy when occupying the same space, or that in a tedious iteration process we approach the final ideal wave function in a way similar to a sculptor shaping his masterpiece. All the time, in everyday life, we unconsciously use these variational and perturbational methods – the most important tools in quantum chemistry. This book may be considered by some students as "too easy". However, I prize easy explanations very highly. In later years the student will not remember long derivations, but will know exactly why something *must* happen. Also, when deriving formulae, I try to avoid presenting the final result right away, but instead proceed with the derivation step by step.[22] The reason is psychological. Students have much stronger motivation knowing they *control* everything, even by simply *accepting* every step of a derivation. It gives them a kind of psychological integrity, very important in any study. Some formulae may be judged as right just by inspection. This is especially valuable for students and I always try to stress this.

In the course of study, students should master material that is both simple and complex. Much of this involves familiarity with the set of mathematical tools repeatedly used throughout this book. The Appendices provide ample reference to such a toolbox. These include matrix algebra, determinants, vector spaces, vector orthogonalization, secular equations, matrix diagonalization, point group theory, delta functions, finding conditional extrema (Lagrange multipliers, penalty function methods), Slater–Condon rules, as well as secondary quantization.

The tone of this book should bring to mind a lecture in an interactive mode. To some, this is not the way books are supposed to be written. I apologize to any readers who may not feel comfortable with this approach.

I invite cordially all readers to share with me their comments on my book: **piela@chem.uw.edu.pl**

My goals

- To arouse the reader's interest in the field of quantum chemistry.
- To show the reader the structure of this field, through the use of the TREE diagram. Boxed text is also used to highlight and summarize important concepts in each chapter.
- To provide the reader with fundamental theoretical concepts and tools, and the knowledge of how to use them.
- To highlight the simple philosophy behind these tools.
- To indicate theoretical problems that are unsolved and worthy of further theoretical consideration.
- To indicate the anticipated and most important directions of research in chemistry (including quantum chemistry).

---

[22] Sometimes this is not possible. Some formulae require painstaking effort in their derivation. This was the case, for example, in the coupled cluster method, p. 546.

To begin with

It is suggested that the reader start with the following.

- A study of the TREE diagram.
- Read the table of contents and compare it with the TREE.
- Address the question of what is *your* goal, i.e. why you would like to read such a book?
- Choose a personal path on the TREE, the suggested itineraries may be of some help.[23]
- Become acquainted with the organization of any chapter.

CHAPTER ORGANIZATION

Once an itinerary is chosen the student will cover different chapters. All the chapters have the same structure, and are divided into sections.

- **Where are we**

  In this section readers are made aware of their current position on the TREE diagram. In this way, they know the relationship of the current chapter to other chapters, what chapters they are expected to have covered already, and the remaining chapters for which the current chapter provides a preparation. The current position shows *whether they should invest time and effort in studying the current chapter*. Without the TREE diagram it may appear, after tedious study of the current chapter, that this chapter was of little value and benefit to the reader.
- **An example**

  Here the reader is confronted with a practical problem that the current chapter addresses. Only after posing a clear-cut problem without an evident solution, will the student see the purpose of the chapter and how the material presented sheds light on the stated problem.
- **What is it all about**

  In this section the essence of the chapter is presented and a detailed exposition follows. This may be an occasion for the students to review the relationship of the current chapter to their chosen path. In surveying the subject matter of a given chapter, it is also appropriate to review student expectations. Those who have chosen a special path will find only some of the material pertinent to their needs. Such recommended paths are also provided within each chapter.

---

[23] This choice may still be tentative and may become clear in the course of reading this book. The subject index may serve as a significant help. For example a reader interested in drug design, that is based in particular on enzymatic receptors, should cover the chapters with ▲ (those considered most important) and then those with △ (at the very least, intermolecular interactions). They will gain the requisite familiarity with the energy which is minimized in computer programs. The reader should then proceed to those branches of the TREE diagram labelled with □. Initially they may be interested in force fields (where the above mentioned energy is approximated), and then in molecular mechanics and molecular dynamics. Our students may begin this course with only the □ labels. However, such a course would leave them without any link to quantum mechanics.

- **Why is this important**
  There is simply not enough time for a student to cover and become familiar with all extent textbooks on quantum mechanics. Therefore, one has to choose a set of important topics, those that represent a key to an understanding of the broad domains of knowledge. To this end, it often pays to master a complex mathematical apparatus. Such mastery often leads to a generalization or simplification of the internal structure of a theory. Not all chapters are of equal importance. At this point, the reader has the opportunity to judge whether the author's arguments about the importance of a current chapter are convincing.
- **What is needed**
  It is extremely disappointing if, after investing time and effort, the reader is stuck in the middle of a chapter, simply for lack of a particular theoretical tool. This section covers the prerequisites necessary for the successful completion of the current chapter. Material required for understanding the text is provided in the Appendices at the end of this book. The reader is asked not to take this section too literally, since a tool may be needed only for a minor part of the material covered, and is of secondary importance. This author, however, does presuppose that the student begins this course with a basic knowledge of mathematics, physics and chemistry.
- **Classical works**
  Every field of science has a founding parent, who identified the seminal problems, introduced basic ideas and selected the necessary tools. Wherever appropriate, we mention these classical investigators, their motivation and their most important contributions. In many cases a short biographical note is also given.
- **The Chapter's Body**
  The main body of each chapter is presented in this section. An attempt is made to divide the contents logically into sub-sections, and to have these sections as small as possible in order to make them easy to swallow. A small section makes for easier understanding.
- **Summary**
  The main body of a chapter is still a big thing to digest and a student may be lost in seeing the logical structure of each chapter.[24] A short summary gives the motivation for presenting the material at hand, and why one should expend the effort, what the main benefits are and why the author has attached importance to this subject. This is a useful point for reflection and consideration. What have we learnt, where are we heading, and where can this knowledge be used and applied?
- **Main concepts, new terms**
  New terms, definitions, concepts, relationships are introduced. In the current chapter they become familiar tools. The reader will appreciate this section (as well as sections *Why is this important* and *Summaries*) just before an examination.

---

[24] This is most dangerous. A student at *any* stage of study has to be able to answer easily what the purpose of each stage is.

- **From the research front**

  It is often ill advised to present state of the art results to students. For example, what value is it to present the best wave function consisting of thousands of terms for the helium atom? The logistics of such a presentation are difficult to contemplate. There is significant didactic value in presenting a wavefunction with one or only a few terms where significant concepts are communicated. On the other hand the student should be made aware of recent progress in generating new results and how well they agree with experimental observations.
- **Ad futurum...**

  The reader deserves to have a learned assessment of the subject matter covered in a given chapter. For example, is this field stale or new? What is the prognosis for future developments in this area? These are often perplexing questions and the reader deserves an honest answer.
- **Additional literature**

  The present text offers only a general panorama of quantum chemistry. In most cases there exists an extensive literature, where the reader will find more detailed information. The role of review articles, monographs and textbooks is to provide an up-to-date description of a particular field. References to such works are provided in this section, often combined with the author's comments on their appropriateness for students.
- **Questions**

  In this section the reader will find ten questions related to the current chapter. Each question is supplied with four possible answers. The student is asked to choose the correct answer. Sometimes the answer will come easily. In other cases, the student will have to decide between two or more similar possibilities that may differ only in some subtle way. In other cases the choice will come down to the truth or an absurdity (I beg your pardon for this). Life is filled with situations where such choices have to be made.
- **Answers**

  Here answers to the above questions are provided.

WEB ANNEX http://www.chem.uw.edu.pl/ideas

The role of the Annex is to expand the readers' knowledge *after* they read a given chapter. At the heart of the web Annex are useful links to other people's websites. The Annex will be updated every several months. The Annex adds at least four new dimensions to my book: colour, motion, an interactive mode of learning and connection to the web (with a plethora of possibilities to go further and further). The *living erratum* in the Annex (with the names of those readers who found the errors) will help to keep improving the book after it was printed.

ACKNOWLEDGEMENTS

The list of people given below is ample evidence that the present book is not just the effort of a single individual, although I alone am responsible for any remaining errors or problems. Special thanks are reserved for Professor Andrzej Sadlej (University of Toruń, Poland). I appreciate very, very much his extraordinary work. I would like to acknowledge the special effort provided by Miss Edyta Małolepsza, who devoted all her strength and talents (always smiling) to keep the whole longtime endeavour running. I acknowledge also the friendly help of Professor Andrzej Holas from the Polish Academy of Sciences, Professors Bogumił Jeziorski and Wojciech Grochala from the University of Warsaw and Professor Stanisław Kucharski from the Silesian University, who commented on Chapters 1, 8, 10 and 11, as well as of Eva Jaroszkiewicz and my other British friends for their linguistic expertise.

My warmest thoughts are always associated with my friends, with whom discussions were unbounded, and contained what we all appreciated most, fantasy and surrealism. I think here of Professor Jean-Marie André (Facultés Universitaires de Namur, Belgium) and of Professor Andrzej J. Sadlej, Professor Leszek Stolarczyk and Professor Wojciech Grochala (from the University of Warsaw). Thank you all for the intellectual glimmers in our discussions.

Without my dearest wife Basia this book would not be possible. I thank her for her love and patience.

Izabelin,
in Kampinos Forest (central Poland),
hot August 2006

SOURCES OF PHOTOGRAPHS AND FIGURES

★ Courtesy of The Nobel Foundation, © The Nobel Foundation (pp. 5, 8, 9, 11, 12, 26, 29, 67, 70, 94, 97, 111, 113, 132, 140, 220, 252, 268, 285, 294, 395, 425, 428, 434, 511, 523, 565, 579, 617, 619, 655, 663, 663, 683, 753, 765, 765, 765, 806, 830, 832, 850). ★ Wikipedia public domain (1, 35, 57, 97, 229, 279, 310, 380, 525, 745, 880, 898) ★ p. 391 – courtesy of Nicolaus Copernicus University, Toruń, Poland ★ Fig. 1.14 – courtesy of Professor Akira Tonomura, Japan ★ p. 112 – with permission from National Physical Laboratory, courtesy of Dr. R.C. McGuiness, UK ★ p. 220 – courtesy of Professor J.D. Roberts, California Institute of Technology, USA ★ p. 287 – reused from International Journal of Quantum Chemistry (1989), Symp. 23, © 1989 John Wiley&Sons, Inc., reprinted with permission of John Wiley&Sons, Inc. ★ p. 392 – courtesy of Mr Gerhard Hund, Germany ★ p. 575 – courtesy of Professor Richard Bader, Canada ★ p. 618 – reproduced from a painting by Ms Tatiana Livschitz, courtesy of Professor W.H. Eugen Schwartz, Germany ★ p. 863 – courtesy of Bradley University, USA ★ p. 131 – courtesy of Alburtus Yale News Bureau (through Wikipedia) ★ p. 219 – courtesy Livermore National Laboratory ★ photographs by the author (377, 456, 509, 509, 311, 805).

Sources of Figures: besides those acknowledged *in situ*: ★ courtesy of Wydawnictwo Naukowe PWN, Poland from “Idee chemii kwantowej”, © 1993 PWN (Figs. 1.1–1.7, 1.9, 1.10, 1.15, 2.1–2.4, 2.6, 3.1, 3.2, 3.3, 3.4, 4.1–4.6, 4.8–4.16, 4.23, 5.1–5.3, 6.1, 6.3, 6.6–6.10, 6.13, 6.14, 6.16, 7.1–7.11, 7.13, 7.14, 8.3, 8.4, 8.6–8.12, 8.18–8.27, 9.1–9.6, 9.21, 9.22, 10.2–10.12, 12.1–12.9, 12.13, 12.14, 13.1–13.6, 13.8, 13.10–13.13, 13.18, 14.1–14.8, 14.14–14.22, 14.26, 14.27b, 15.1–15.3, 15.5, 15.7–15.10, B.1, B.2, C.1–C.7, G.1, J.1, J.2, O.1, R.1, R.2, X.1–X.4, Y.1, Y.2) ★ Figs. 14.14, 14.15, 14.17 and Tables 14.1–14.6 reused and adapted with permission from the American Chemical Society from S. Shaik, Journal of the American Chemical Society, 103 (1981) 3692. © 1981 American Chemical Society. Courtesy of Professor Sason Shaik. Figures 4.3, 4.4, 4.6–4.13, 4.15, 4.17–4.21, 6.12, 7.8, 8.5, 8.8, 8.9, 8.26, 10.1–10.4, 11.2, 12.5, 12.6, 14.1, 14.2, 14.22, 14.25b, O.1, R.2, T.1, Y.1 (also a part of the cover figure) have been made using the software of *Mathematica* by Stephen Wolfram.

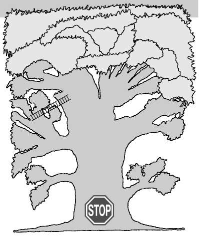
Quotations: from John C. Slater’s book “Solid State and Molecular Theory”, London, Wiley, 1975 (pp. 180, 329, 343, 986); from Werner Heisenberg’s “Der Teil und das Ganze” (p. 11); from Richard Feynman’s Nobel lecture, © The Nobel Foundation (p. 105); from Simon Schnoll’s “Gieroi i zladiei rossiyskoi nauki” (p. 524); from Richard Zare in “Chemical and Engineering News” 4 (1990) 32 (p. 766); Erwin Schrödinger’s curriculum vitae (unpublished) – courtesy of the University of Wrocław, The Rector Leszek Pacholski, Poland (p. 70).

Despite reasonable attempts made, we were unable to contact the owners of the copyright of the following images: ★ p. 850 – from S. Schnoll’s book “Geroi i zladiei rossiyskoi nauki”, Kronpress, Moscow, 1997 ★ p. 744 – reproduced from B.W. Nekrasov, “Kurs obshchei khimii”, Goskhimizdat, Moscow, 1960 ★ Figs. 13.21 and 13.22 reproduced from “Biology Today”, CRM Books, Del Mar, USA, ©1972 Communications Research Machines ★ *and the copyright owners of the items found in the websites* ★ http://www-gap.dcs.st-and.ac.uk/~history (pp. 4, 8, 9, 26, 42, 43, 72, 94, 105, 110, 219, 278, 318, 329, 446, 482, 512, 575, 580, 683, 683, 849, 851, 858, 866, 866, 876, 878, 879, 903, 997) ★ http://nuclphys.sinp.msu.ru (p. 252) ★ http://www.phy.duke.edu, photo Lotte Meitner-Graf (p. 521) ★ http://www.volny.cz (p. 7) ★ http://metwww.epfl.ch (p. 438) ★ http://osulibrary.orst.edu (p. 331) ★ http://www.mathsoc.spb.ru (p. 329) ★ http://www.stetson.edu (p. 311) ★ http://www.aeiou.at (p. 93) ★ http://www.quantum-chemistry-history.com (pp. 361, 364).

If you are the copyright owner to any of the images we have used without your explicit permission (because we were unable to identify you during our search), please contact Prof. Lucjan Piela, Chemistry Department, Warsaw University, 02093 Warsaw, Poland, e-mail: piela@chem.uw.edu.pl, phone (48)-22-8220211 ext. 525. We will place the appropriate information on our website http://www.chem.uw.edu.pl/ideas which represents an appendix to the present book, as well as correct the next editions of the book.

Chapter 1

THE MAGIC OF QUANTUM MECHANICS



WHERE ARE WE?

Where are we?

We are at the beginning of all the paths, at the base of the TREE.

An example

Since 1911 we have known that atoms and molecules are built of two kinds of particles: electrons and nuclei. Experiments show the particles may be treated as point-like objects of certain mass and electric charge. The electronic charge is equal to $-e$, while the nuclear charge amounts to Ze, where $e = 1.6 \cdot 10^{-19}$ C and Z is a natural number. Electrons and nuclei interact according to the Coulomb law, and classical mechanics and electrodynamics predict that *any atom or molecule is bound to collapse in a matter of a femtosecond* emitting an infinite amount of energy. Hence, according to the classical laws, the complex matter we see around us (also our bodies) should simply not exist at all.

However, atoms and molecules do exist, and their existence may be described in detail by quantum mechanics using what is known as the wave function. The axioms of quantum mechanics provide the rules for the derivation of this function and for the calculation of all the observable properties of atoms and molecules.

Charles Augustin de Coulomb (1736–1806), French military engineer, one of the founders of quantitative physics. In 1777 he constructed a torsion balance for measuring very weak forces, with which he was able to demonstrate the inverse square law for electric and magnetic forces. He also studied charge distribution on the surfaces of dielectrics.

What is it all about

Any branch of science has a list of axioms, on which the entire construction is built.[1] For quantum mechanics, six such axioms (postulates) have been established. The postulates have evolved in the process of reconciling theory and experiment, and may sometimes be viewed as non-intuitive. These axioms stand behind any tool of quantum mechanics used in practical applications. They also lead to some striking conclusions concerning the reality of our world, for example, the possibilities of bilocation, teleportation, and so on. These unexpected conclusions have recently been experimentally confirmed.

Why is this important?

The axioms given in this chapter represent the *foundation* of quantum mechanics, and justify all that follows in this book. In addition, our ideas of what the world is really like acquire a new and unexpected dimension.

What is needed?

- Complex numbers (necessary).
- Operator algebra and vector spaces, Appendix B, p. 895 (necessary).
- Angular momentum, Appendix F, p. 955 (recommended).
- Some background in experimental physics: black body radiation, photoelectric effect (recommended).

Classical works

The beginning of quantum theory was the discovery, by Max Planck, of the electromagnetic *energy quanta emitted* by a black body. The work was published under the title: "*Über das Gesetz der Energieverteilung im Normalspektrum*"[2] *in Annalen der Physik*, 4 (1901) 553. ★ Four years later Albert Einstein published a paper "*Über die Erzeugung und Verwandlung des Lichtes betreffenden heuristischen Gesichtspunkt*" in *Annalen der Physik*, 17 (1905) 132, in which he explained the photoelectric effect by assuming that the energy is *absorbed* by a metal as quanta of energy. ★ In 1911 Ernest Rutherford discovered that atoms are composed of a massive nucleus and electrons: "*The Scattering of the α and β Rays and the Structure of the Atom*", in *Proceedings of the Manchester Literary and Philosophical Society*, IV,

---

[1] And which are not expected to be proved.

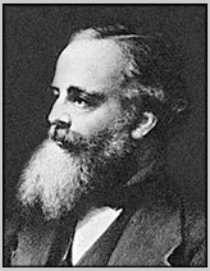
[2] *Or "On the Energy Distribution Law in the Normal Spectrum"* with a note saying that the material had already been presented (in another form) at the meetings of the German Physical Society on Oct. 19 and Dec. 14, 1900.
On p. 556 one can find the following historical sentence on the total energy denoted as U_N: "*Hierzu ist es notwendig, U_N nicht als eine stetige, unbeschränkt teilbare, sondern als eine diskrete, aus einer ganzen Zahl von endlichen gleichen Teilen zusammengesetzte Grösse aufzufassen*", which translates as: "*Therefore, it is necessary to assume that U_N does not represent any continuous quantity that can be divided without any restriction. Instead, one has to understand that it as a discrete quantity composed of a finite number of equal parts.*

55 (1911) 18. ★ Two years later Niels Bohr introduced a planetary model of the hydrogen atom in "*On the Constitution of Atoms and Molecules*" in *Philosophical Magazine*, Series 6, vol. 26 (1913). ★ Louis de Broglie generalized the corpuscular and wave character of any particle in his PhD thesis "*Recherches sur la théorie des quanta*", Sorbonne, 1924. ★ The first mathematical formulation of quantum mechanics was developed by Werner Heisenberg in "*Über quantentheoretischen Umdeutung kinematischer und mechanischer Beziehungen*", *Zeitschrift für Physik*, 33 (1925) 879. ★ Max Born and Pascual Jordan recognized matrix algebra in the formulation [in "*Zur Quantenmechanik*", *Zeitschrift für Physik*, 34 (1925) 858] and then all three [the famous "*Drei-Männer Arbeit*" entitled "*Zur Quantenmechanik. II.*" and published in *Zeitschrift für Physik*, 35 (1925) 557] expounded a coherent mathematical basis for quantum mechanics. ★ Wolfgang Pauli introduced his "two-valuedness" for the non-classical electron coordinate in "*Über den Einfluss der Geschwindigkeitsabhängigkeit der Elektronenmasse auf den Zeemaneffekt*", published in *Zeitschrift für Physik*, 31 (1925) 373, the next year George Uhlenbeck and Samuel Goudsmit described their concept of particle spin in "*Spinning Electrons and the Structure of Spectra*", *Nature*, 117 (1926) 264. ★ Wolfgang Pauli published his famous exclusion principle in "*Über den Zusammenhang des Abschlusses der Elektronengruppen im Atom mit der Komplexstruktur der Spektren*" which appeared in *Zeitschrift für Physik B*, 31 (1925) 765. ★ The series of papers by Erwin Schrödinger "*Quantisierung als Eigenwertproblem*" in *Annalen der Physik*, 79 (1926) 361 (other references in Chapter 2) was a major advance. He proposed a different mathematical formulation (from Heisenberg's) and introduced the notion of the wave function. ★ In the same year Max Born, in "*Quantenmechanik der Stossvorgänge*" which appeared in *Zeitschrift für Physik*, 37 (1926) 863 gave an interpretation of the wave function. ★ The uncertainty principle was discovered by Werner Heisenberg and described in "*Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik*", *Zeitschrift für Physik*, 43 (1927) 172. ★ Paul Adrien Maurice Dirac reported an attempt to reconcile quantum and relativity theories in a series of papers published from 1926–1928 (references in Chapter 3). ★ Albert Einstein, Boris Podolsky and Natan Rosen proposed a test (then a Gedanken or thinking-experiment, now a real one) of quantum mechanics "*Can quantum-mechanical description of physical reality be considered complete*?" published in *Physical Review*, 47 (1935) 777. ★ Richard Feynman, Julian Schwinger and Shinichiro Tomonaga independently developed quantum electrodynamics in 1948. ★ John Bell, in "*On the Einstein–Podolsky–Rosen Paradox*", *Physics*, 1 (1964) 195 reported inequalities which were able to verify the very foundations of quantum mechanics. ★ Alain Aspect, Jean Dalibard and Géard Roger in "*Experimental Test of Bell's Inequalities Using Time-Varying Analyzers*", *Physical Review Letters*, 49 (1982) 1804 reported measurements which violated the Bell inequality and proved the non-locality or/and (in a sense) non-reality of our world. ★ Akira Tonomura, Junji Endo, Tsuyoshi Matsuda and Takeshi Kawasaki in "*Demonstration of Single-Electron Buildup of an Interference Pattern*", *American Journal of Physics*, 57 (1989) 117 reported direct electron interference in a two-slit experiment. ★ Charles H. Bennett, Gilles Brassard, Claude Crépeau, Richard Jozsa, Asher Peres and William K. Wootters, in "*Teleporting an unknown quantum state via dual classical and Einstein–Podolsky–Rosen channels*" in *Physical Review Letters*, 70 (1993) 1895 designed a teleportation experiment, which has subsequently been successfully accomplished by Dik Bouwmeester, Jian-Wei Pan, Klaus Mattle, Manfred Eibl, Harald Weinfurter and Anton Zeilinger, "*Experimental Quantum Teleportation*" in *Nature*, 390 (1997) 575.

1.1 HISTORY OF A REVOLUTION

The end of the nineteenth century saw itself as a proud period for physics, which seemed to finally achieve a state of coherence and clarity. Physics at that time believed the world consisted of two kingdoms: a kingdom of particles and a kingdom of electromagnetic waves. Motion of particles had been described by Isaac Newton's equation, with its striking simplicity, universality and beauty. Similarly, electromagnetic waves had been accurately described by James Clerk Maxwell's simple and beautiful equations.

James Clerk Maxwell (1831–1879), British physicist, professor at the University of Aberdeen, Kings College, London, and Cavendish Professor in Cambridge. His main contributions are his famous equations for electromagnetism (1864), and the earlier discovery of velocity distribution in gases (1860).



Young Planck was advised to abandon the idea of studying physics, because everything had already been discovered. *This beautiful idyll was only slightly incomplete, because of a few annoying details: the strange black body radiation, the photoelectric effect and the mysterious atomic spectra*. Just some rather exotic problems to be fixed in the near future…

As it turned out, they opened a New World. The history of quantum theory, one of most revolutionary and successful theories ever designed by man, will briefly be given below. Many of these facts have their continuation in the present textbook.

Black body radiation

1900 – Max Planck

Max Planck wanted to understand black body radiation. The black body may be modelled by a box, with a small hole, Fig. 1.1. We heat the box up, wait for the system to reach a stationary state (at a fixed temperature) and see what kind of electromagnetic radiation (intensity as a function of frequency) comes out of the hole. In 1900 Rayleigh and Jeans[3] tried to apply classical mechanics to this problem, and calculated correctly that the black body would emit electromagnetic radiation having a distribution of frequencies. However, the larger the frequency the larger its intensity, leading to what is known as ultraviolet catastrophe, an absurd conclusion. Experiment contradicted theory (Fig. 1.1).

UV catastrophe

At a given temperature T the intensity distribution (at a given frequency ν, Fig. 1.1.b) has a single maximum. As the temperature increases, the maximum should shift towards higher frequencies (a piece of iron appears red at 500 °C, but bluish at 1000 °C). Just like Rayleigh and Jeans, Max Planck was unable to derive

---

[3] James Hopwood Jeans (1877–1946), British physicist, professor at the University of Cambridge and at the Institute for Advanced Study in Princeton. Jeans also made important discoveries in astrophysics (e.g., the theory of double stars).

Max Karl Ernst Ludwig Planck (1858–1947), German physicist, professor at the universities in Munich, Kiel and Berlin, first director of the Institute of Theoretical Physics in Berlin. Planck was born in Kiel, where his father was a university professor of law. Max Planck was a universally talented school pupil, then an outstanding physics student at the University of Berlin, where he was supervised by Gustaw Kirchhoff and Hermann Helmholtz. Music was his passion throughout his life, and he used to play piano duets with Einstein (who played the violin). This hard-working, middle-aged, old-fashioned, professor of thermodynamics made a major breakthrough as if in an act of scientific desperation. In 1918 Planck received the Nobel Prize "*for services rendered to the advancement of Physics by his discovery of energy quanta*". Einstein recalls jokingly Planck's reported lack of full confidence in general relativity theory: "*Planck was one of the most outstanding people I have ever known, (...) In reality, however, he did not understand physics. During the solar eclipse in 1919 he stayed awake all night, to see whether light bending in the gravitational field will be confirmed. If he understood the very essence of the general relativity theory, he would quietly go to bed, as I did*". (Cited by Ernst Straus in "*Einstein: A Centenary Volume*", p. 31).

John William Strutt, Lord Rayleigh (1842–1919), British physicist, Cavendish Professor at Cambridge, contributed greatly to physics (wave propagation, light scattering theory – Rayleigh scattering). In 1904 Rayleigh received the Nobel Prize "*for his investigations of the densities of the most important gases and for his discovery of argon in connection with these studies*".

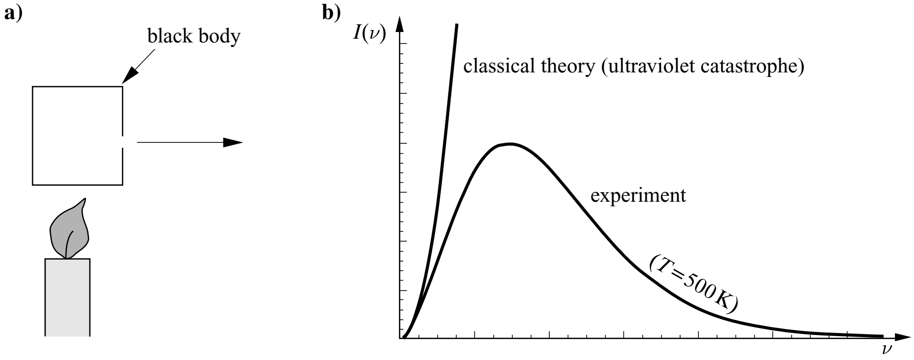


Fig. 1.1. Black body radiation. (a) As one heats a box to temperature T, the hole emits electromagnetic radiation with a wide range of frequencies. The distribution of intensity $I(\nu)$ as a function of frequency ν is given in Fig. (b). There is a serious discrepancy between the results of classical theory and the experiment, especially for large frequencies. Only after assuming the existence of energy quanta can theory and experiment be reconciled.

this simple qualitative picture from classical theory – something had to be done. On 14 December 1900, the generally accepted date for the birth of quantum theory, Planck presented his theoretical results for the black body treated as an ensemble of harmonic oscillators. With considerable reluctance he postulated[4] that matter cannot emit radiation otherwise than by equal portions ("*quanta*") of energy $h\nu$, proportional to the frequency ν of vibrations of a single oscillator of the black body. The famous Planck constant h followed soon after ($h = 6.62607 \cdot 10^{-34}$ J s; but in this book, we will use a more convenient constant[5] $\hbar = \frac{h}{2\pi}$). It is exactly this hypothesis about energy quanta that led to the agreement of theory with experiment and the elimination of the ultraviolet catastrophe.

quanta

Planck constant

Photoelectric effect

1905 – Albert Einstein

photon

The second worrying problem, apart from the black body, was the photoelectric effect.[6] Light knocks electrons[7] out of metals, but only when its frequency exceeds a certain threshold. Classical physics was helpless. In classical theory, light energy should be stored in the metal in a continuous way and *independent of the frequency used, after a sufficient period of time, the electrons should be ejected from the metal.* Nothing like that was observed. Einstein introduced the idea of electromagnetic radiation quanta as *particles*, later baptised *photons* by Gilbert Lewis. Note that Planck's idea of a quantum concerned energy transfer *from the black body to the electromagnetic field*, while Einstein introduced it *for the opposite direction* with the energy corresponding to Planck's quantum. Planck considered the quantum as a portion of energy, while for Einstein the quantum meant a particle.[8] Everything became clear: energy goes to electrons by *quanta* and this is why only quanta ex-

---

[4] He felt uncomfortable with this idea for many years.

[5] Known as "h bar".

[6] Experimental work on the effect had been done by Philipp Eduard Anton Lenard (1862–1947), German physicist, professor at Breslau (now Wrocław), Köln and Heidelberg. Lenard discovered that the number of photoelectrons is proportional to the intensity of light, and that their kinetic energy *does not depend at all on the intensity*, depending instead on the *frequency* of light. Lenard received the Nobel Prize in 1905 "*for his work on cathode rays*". A faithful follower of Adolf Hitler, and devoted to the barbarous Nazi ideas, Lenard terrorized German science. He demonstrates that scientific achievement and decency are two separate human characteristics.

[7] The electron was already known, having been predicted as early as 1868 by the Irish physicist George Johnstone Stoney (1826–1911), and finally discovered in 1897 by the British physicist Joseph John Thomson (1856–1940). Thomson also discovered a strange pattern: the number of electrons in light elements was equal to about one half of their atomic mass. Free electrons were obtained much later (1906). The very existence of atoms was still a hypothesis. The atomic nucleus was to be discovered only in 1911. Physicists were also anxious about the spectra of even the simplest substances such as hydrogen. Johann Jacob Balmer, a teacher from Basel, was able to design an astonishingly simple formula which fitted perfectly some of the observed lines in the hydrogen spectrum ("*Balmer series*"). All that seemed mysterious and intriguing.

[8] It is true that Einstein wrote about "*point-like quanta*" four years later, in a careful approach identifying the quantum with the particle. Modern equipment enables us to count photons, the individual particles of light. The human eye is also capable of detecting 6–8 photons striking a neuron.

Gilbert Newton Lewis (1875–1946), the greatest American chemist, who advanced American chemistry internationally through his research and teaching. In a 1926 article in *Nature* Lewis introduced the name of the "photon". He also developed an early theory of chemical bonding ("Lewis structures") based on counting the valence electrons and forming "octets" from them. The idea that atoms in molecules tend to form octets in order to complete their electron shells turned out to be surprisingly useful in predicting bond patterns in molecules. A drawback of this concept is that it was not related to the ideas of theoretical physics. It is an example of an extremely clever concept rather than of a coherent theory. Lewis also introduced a new definition of acids and bases, which is still in use.

ceeding some threshold (the binding energy of an electron in the metal) are able to eject electrons from a metal.

1911 – Ernest Rutherford

Rutherford proved experimentally that an atom has massive nucleus, but it is however very small when compared to the size of the atom. The positive charge is concentrated in the nucleus, which is about 10^{-13} cm in size. The density of the nuclear matter boggles the imagination: 1 cm^3 has a mass of about 300 million tonnes. This is how researchers found out that an atom is composed of a massive nucleus and electrons. atomic nucleus

The model of the hydrogen atom

1913 – Niels Bohr

Atomic spectra were the third great mystery of early 20th century physics. Even interpretation of the spectrum of the hydrogen atom represented a challenge. At the age of 28 Bohr proposed (in 1913) a simple planetary model of this atom, in which the electron, contrary to classical mechanics, did not fall onto the nucleus. Instead, it changed its orbit with accompanying absorption or emission of energy quanta. Bohr assumed that angular orbital momentum is quantized and that the centrifugal force is compensated by the Coulomb attraction between the electron and the nucleus. He was able to reproduce part of the spectrum of the hydrogen

---

In 1905, the accuracy of experimental data was too poor to confirm Einstein's theory as the only one which could account for the experimental results. Besides, the wave nature of light was supported by thousands of crystal clear experiments. Einstein's argument was so breathtaking (...particles???), that Robert Millikan decided to falsify experimentally Einstein's hypothesis. However, after ten years of investigations, Millikan acknowledged that he was forced to support undoubtedly Einstein's explanation "however absurd it may look" (*Rev. Modern Phys*. 21 (1949) 343). This conversion of a sceptic inclined the Nobel Committee to grant Einstein the Nobel Prize in 1923 "*for his work on the elementary charge of electricity and on the photo-electric effect*".

Niels Hendrik Bohr (1885–1962), Danish physicist, a professor at Copenhagen University, played a key role in the creation and interpretation of quantum mechanics (see end of this chapter). Bohr was born in Copenhagen, the son of a professor of physiology. He graduated from Copenhagen university and in 1911 obtained his doctorate there. Then he went to Cambridge to work under the supervision of J.J. Thomson, the discoverer of the electron. The collaboration did not work out, and in 1912 Bohr began to cooperate with Ernest Rutherford at the University of Manchester. In Manchester Niels Bohr made a breakthrough by introducing a planetary model of hydrogen atom. He *postulated* that the angular orbital momentum must be quantized. Using this Bohr reproduced the experimental spectrum of hydrogen atom with high accuracy. In 1922 Bohr received the Nobel Prize "*for his investigation of the structure of atoms*". In the same year he became the father of Aage Niels Bohr – a future winner of the Nobel Prize (1975, for studies of the structure of nuclei). In October 1943, Bohr and his family fled from Denmark to Sweden, and then to Great Britain and the USA, where he worked on the Manhattan Project. After the war the Bohr family returned to Denmark.

atom very accurately. Bohr then began work on the helium atom, which turned out to be a disaster, but he was successful again with the helium cation[9] He^{+}.

Niels Bohr played an inspiring role in the development and popularization of quantum mechanics. His Copenhagen Institute for Theoretical Physics, founded in 1921, was the leading world centre in the twenties and thirties, where many young theoreticians from all over the world worked on quantum mechanical problems.[10] Bohr, with Werner Heisenberg, Max Born and John von Neumann, contributed greatly to the elaboration of the philosophical foundations of quantum mechanics. According to this, quantum mechanics represents a coherent and complete model of reality ("the world"), and the discrepancies with the classical mechanics have a profound and fundamental character,[11] and both theories coincide in the limit $h \to 0$ (where h is the Planck constant), and thus the predictions of quantum

---

[9] Bohr did not want to publish without good results for all other atoms, something he would never achieve. Rutherford argued: "*Bohr, you explained hydrogen, you explained helium, people will believe you for other atoms*".

[10] John Archibald Wheeler recalls that, when he first came to the Institute, he met a man working in the garden and asked him where he could find Professor Bohr. The gardener answered: "*That's me*".

[11] The centre of the controversy was that quantum mechanics is indeterministic, while classical mechanics is deterministic, although this indeterminism is not all it seems. As will be shown later in this chapter, quantum mechanics is a *fully deterministic theory in the Hilbert space* (the space of all possible wave functions of the system), its indeterminism pertains to the physical space in which we live.

mechanics reduce to those of classical mechanics (known as Bohr's correspondence principle).

"Old quantum theory"

1916 – Arnold Sommerfeld

In 1916 Arnold Sommerfeld generalized the Bohr quantization rule beyond the problem of the one-electron atom. Known as "old quantum theory", it did not represent any coherent theory of general applicability. As a matter of fact, this quantization was achieved by assuming that for every periodic variable (like an angle), an integral is equal to an integer times the Planck constant.[12] Sommerfeld also tried to apply the Bohr model to atoms with a single valence electron (he had to modify the Bohr formula by introducing the quantum defect, i.e. a small change in the principal quantum number, see p. 179).

Arnold Sommerfeld (1868–1951), German physicist, professor at the Mining Academy in Clausthal, then at the Technical University of Aachen, in the key period 1906–1938, was professor at Munich University. Sommerfeld considered not only circular (Bohr-like) orbits, but also elliptical ones, and introduced the angular quantum number. He also investigated X rays and the theory of metals. The scientific father of many Nobel Prize winners he did not get this distinction himself.

Waves of matter

1923 – Louis de Broglie

In his doctoral dissertation, stuffed with mathematics, Louis de Broglie introduced the concept of "waves of matter". He postulated that not only photons, but *also any other particle*, has, besides its corpuscular characteristics, some wave properties (those corresponding to light had been known for a long, long time). According to de Broglie, the wave length corresponds to momentum p,

dualism

Louis Victor Pierre Raymond de Broglie (1892–1987) was studying history at the Sorbonne, carefully preparing himself for a diplomatic career. His older brother Maurice, a radiographer, aroused his interest in physics. The first World War (Louis did his military service in a radio communications unit) and the study of history delayed his start in physics. He was 32 when he presented his doctoral dissertation, which embarrassed his supervisor, Paul Langevin. The thesis, on the wave nature of all particles, was so revolutionary, that only a positive opinion from Einstein, who was asked by Langevin to take a look of the dissertation, convinced the doctoral committee. Only five years later (in 1929), Louis de Broglie received the Nobel Prize "*for his discovery of the wave nature of electrons*".

---

[12] Similar periodic integrals were used earlier by Bohr.

$$p = \frac{h}{\lambda}$$

where h is again the Planck constant! What kind of momentum can this be, in view of the fact that momentum depends on the laboratory coordinate system chosen? Well, it is the momentum measured in the same laboratory coordinate system as that used to measure the corresponding wave length.

Electron–photon scattering

1923 – Arthur Compton[13]

It turned out that an electron–photon collision obeys the same laws of dynamics as those describing collision of two particles: the energy conservation law and the momentum conservation law. This result confirmed the wave–corpuscular picture emerging from experiments.

Discovery of spin

1925 – George E. Uhlenbeck and Samuel A. Goudsmit

Two Dutch students explained an experiment (Stern–Gerlach) in which a beam of silver atoms passing through a magnetic field split into two beams. In a short paper, they suggested that the silver atoms have (besides their orbital angular momentum) an additional internal angular momentum (spin), *similar* to a macroscopic body, which besides its centre-of-mass motion, also has a rotational (spinning) motion.[14] Moreover, the students demonstrated that the atomic spin follows from the spin of the electrons: among the 47 electrons of the silver atom, 46 have their spin compensated (23 "down" and 23 "up"), while the last "unpaired" electron gives the net spin of the atom.

Pauli Exclusion Principle

1925 – Wolfgang Pauli[15]

Pauli postulated that *in any system two electrons cannot be in the same state* (including their spins). This "Pauli exclusion principle" was deduced from spectroscopic data (some states were not allowed).

---

[13] Arthur Holly Compton (1892–1962), American physicist, professor at the universities of Saint Louis and Chicago. He obtained the Nobel Prize in 1927 "*for the discovery of the effect named after him*", i.e. for investigations of electron–photon scattering.

[14] Caution: *identifying* the spin with the rotation of a rigid body leads to physical inconsistencies.

[15] Pauli also introduced the idea of spin when interpreting spectra of atoms with a single valence electron. He was inspired by Sommerfeld, who interpreted the spectra by introducing the quantum number $j = l \pm \frac{1}{2}$, where the quantum number l quantized the orbital angular momentum of the electron. Pauli described spin as *a bivalent non-classical characteristic of the electron* [W. Pauli, *Zeit. Phys. B* 3 (1925) 765].

Matrix quantum mechanics

1925 – Werner Heisenberg

A paper by 24 year old Werner Heisenberg turned out to be a breakthrough in quantum theory.[16] Max Born recognized matrix algebra in Heisenberg's formulation (who, himself, had not yet realised it) and in the same year a more solid formulation of the new mechanics ("matrix mechanics") was proposed by Werner Heisenberg, Max Born and Pascual Jordan.[17]

Schrödinger equation

1926 – Erwin Schrödinger

In November 1925, Erwin Schrödinger delivered a lecture at the Technical University in Zurich (ETH), in which he presented the results of de Broglie. Professor Peter Debye stood up and asked the speaker:

Peter Joseph Wilhelm Debye, more exactly, Peter Josephus Wilhelmus Debye (1884–1966), Dutch physicist and chemist, professor in the Technical University (ETH) of Zurich (1911, 1920–1937) as well as at Göttingen, Leipzig and Berlin, won the Nobel Prize in chemistry in 1936 "*for his contribution to our knowledge of molecular structure through his investigations on dipole moments and on the diffraction of X-rays and electrons in gases*". Debye emigrated to the USA in 1940, where he obtained a professorship at Cornell University in Ithaca, NY (and remained in this beautiful town to the end of his life). His memory is still alive there. Professor Scheraga remembers him as an able chair in seminar discussions, in the tradition of the Zurich seminar of 1925.

---

[16] On June 7, 1925, Heisenberg was so tired after a bad attack of hay fever that he decided to go and relax on the North Sea island of Helgoland. Here, he divided his time between climbing the mountains, learning Goethe's poems by heart and (despite his intention to rest) hard work on the spectrum of the hydrogen atom with which he was obsessed. It was at night on 7 or 8 June that he saw something – the beginning of the new mechanics. In later years he wrote in his book "*Der Teil and das Ganze*": "*It was about three o'clock in the morning when the final result of the calculation lay before me. At first I was deeply shaken. I was so excited that I could not think of sleep. So I left the house and awaited the sunrise on the top of a rock*." The first man with whom Heisenberg shared his excitement a few days later was his schoolmate Wolfgang Pauli, and, after another few days, also with Max Born.

[17] Jordan, despite his talents and achievements, felt himself underestimated and even humiliated in his native Germany. For example, he had to accept a position at Rostock University, which the German scientific elite used to call the "Outer-Mongolia of Germany". The best positions seemed to be reserved. When Hitler came to power, Jordan became a fervent follower…

Max Born (1882–1970), German physicist, professor at the universities of Göttingen, Berlin, Cambridge and Edinburgh, born in Breslau (now Wrocław) to the family of a professor of anatomy in Breslau. Born studied first in Breslau, then at Heidelberg and Zurich. He received his PhD in physics and astronomy in 1907 at Göttingen, where he began his swift academic career. Born obtained a chair at the University of Berlin in 1914, and returned to Göttingen in 1921, where he founded an outstanding school of theoretical physics, which competed with the famous institute of Niels Bohr in Copenhagen. Born supervised Werner Heisenberg, Pascual Jordan and Wolfgang Pauli. It was Born who recognized, in 1925, that Heisenberg's quantum mechanics could be formulated in terms of matrix algebra. Together with Heisenberg and Jordan, he created the first consistent quantum theory (the famous "*drei-Männer Arbeit*"). After Schrödinger's formulation of quantum mechanics, Born proposed the probabilistic interpretation of the wave function. Despite such seminal achievements, the Nobel Prizes in the thirties were received by his colleagues. Finally, when in 1954 Born obtained the Nobel Prize "*for his fundamental research in quantum mechanics, especially for his statistical interpretation of the wave-function*", there was a great relief among his famous friends.

"*You are telling us about waves, but where is the wave equation in your talk?*" Indeed, there wasn't any! Schrödinger began to work on this and the next year formulated what is now called wave mechanics based on the wave equation. Both formulations, Heisenberg's and Schrödinger's[18] turned out to be equivalent and are now known as the foundation for (non-relativistic) quantum mechanics.

Statistical interpretation of wave function

1926 – Max Born

Max Born proposed interpreting the square of the complex modulus of Schrödinger's wave function as the probability density for finding the particle.

Uncertainty principle

1927 – Werner Heisenberg

Heisenberg concluded that it is not possible to measure simultaneously the position (x) and momentum (p_x) of a particle with any desired accuracy. The more exactly we measure the position (small Δx), the larger the error we make in measuring the momentum (large Δp_x) and *vice versa*.

---

[18] And the formulation proposed by Paul A.M. Dirac.

Electron diffraction

1927 – Clinton Davisson, Lester H. Germer, George Thomson[19]

Davisson and Germer, and Thomson, demonstrated in ingenious experiments that indeed electrons do exhibit wave properties (using crystals as diffraction gratings).

The birth of quantum chemistry

1927 – Walter Heitler, Fritz Wolfgang London

Walter Heitler and Fritz Wolfgang London convincingly explained why two neutral atoms (like hydrogen) attract each other with a force so strong as to be comparable with the Coulomb forces between ions. Applying the Pauli exclusion principle when solving the Schrödinger equation is of key importance. Their paper was received on June 30, 1927, by *Zeitschrift für Physik*, and this may be counted as the birthday of quantum *chemistry*.[20]

Dirac equation for the electron and positron

1928 – Paul Dirac

Paul Dirac made a magnificent contribution to quantum theory. His main achievements are the foundations of quantum electrodynamics and construction of the relativistic wave equation (1926–1928) which now bears his name. The equation not only described the electron, but also its anti-matter counterpart – the positron (predicting anti-matter). Spin was also inherently present in the equation.

Quantum field theory

1929 – Werner Heisenberg and Wolfgang Pauli

These classmates developed a theory of matter, and the main features still survive there. In this theory, the elementary particles (the electron, photon, and so on) were viewed as excited states of the corresponding fields (the electron field, electromagnetic field and so on).

---

[19] Clinton Joseph Davisson (1881–1958), American physicist at Bell Telephone Laboratories. He discovered the diffraction of electrons with L.H. Germer, and they received the Nobel Prize in 1937 "*for their experimental discovery of the diffraction of electrons by crystals*". The prize was shared with G.P. Thomson, who used a different diffraction method. George Paget Thomson (1892–1975), son of the discoverer of the electron, Joseph John Thomson, and professor at universities in Aberdeen, London and Cambridge.

[20] The term "quantum chemistry" was first used by Arthur Haas in his lectures to the Physicochemical Society of Vienna in 1929 (A. Haas, "*Die Grundlagen der Quantenchemie. Eine Einleitung in vier Vortragen*", Akademische Verlagsgesellschaft, Leipzig, 1929).

Discovery of anti-matter (the positron)

1932 – Carl Anderson[21]

One of Dirac's important results was the observation that his relativistic wave equation is satisfied, not only by the electron but also by a mysterious unknown particle, the positive electron (positron). This anti-matter hypothesis was confirmed by Carl Anderson, who found the positron experimentally – a victorious day for quantum theory.

anti-matter

Quantum electrodynamics

1948 – Richard Feynman, Julian Schwinger, Shinichiro Tomonaga[22]

The Dirac equation did not take all the physical effects into account. For example, the strong electric field of the nucleus polarizes a vacuum so much, that electron–positron pairs emerge from the vacuum and screen the electron–nucleus interaction. The quantum electrodynamics (QED) developed independently by Feynman, Schwinger and Tomonaga accounts for this, and for similar effects, and brings theory and experiment to an agreement of unprecedented accuracy.

Bell inequalities

1964 – John Bell

The mathematician John Bell proved that, if particles have certain properties *before measurement* (so that they were small but *classical objects*), then the measurement results would have to satisfy some inequalities which contradict the predictions of quantum mechanics (further details at the end of this chapter).

Is the world non-local?

1982 – Alain Aspect

Experiments with photons showed that the Bell inequalities are *not* satisfied. This means that either there is *instantaneous communication* even between extremely distant particles ("entangled states"), or that the particles *do not have some definite properties* before the measurement is performed (more details at the end of this chapter).

Teleportation of the photon state

1997 – Anton Zeilinger

A research group at the University of Innsbruck used entangled quantum states (see p. 39) to perform teleportation of a photon state[23] that is, to prepare at a

---

[21] More details in Chapter 3.

[22] All received the Nobel Prize in 1965 *"for their fundamental work in quantum electrodynamics, with fundamental implications for the physics of elementary particles"*.

[23] D. Bouwmeester, J. Pan, K. Mattle, M. Eibl, H. Weinfurter, A. Zeilinger, *Nature* 390 (1997) 575.

distance any state of a photon with simultaneous disappearance of this state from the teleportation site (details at the end of this chapter).

1.2 POSTULATES

All science is based on a number of axioms (postulates). Quantum mechanics is based on a system of axioms that have been formulated to be as simple as possible and yet reproduce experimental results. Axioms are not supposed to be proved, their justification is efficiency. Quantum mechanics, the foundations of which date from 1925–26, still represents the basic theory of phenomena within atoms and molecules. This is the domain of chemistry, biochemistry, and atomic and nuclear physics. Further progress (quantum electrodynamics, quantum field theory, elementary particle theory) permitted deeper insights into the structure of the atomic nucleus, but did not produce any fundamental revision of our understanding of atoms and molecules. Matter as described at a non-relativistic[24] quantum mechanics represents a system of electrons and nuclei, treated as point-like particles with a definite mass and electric charge, moving in three-dimensional space and interacting by *electrostatic* forces.[25] This model of matter is at the core of quantum chemistry, Fig. 1.2.

The assumptions on which quantum mechanics is based are given by the following postulates I–VI. For simplicity, we will restrict ourselves to a single particle

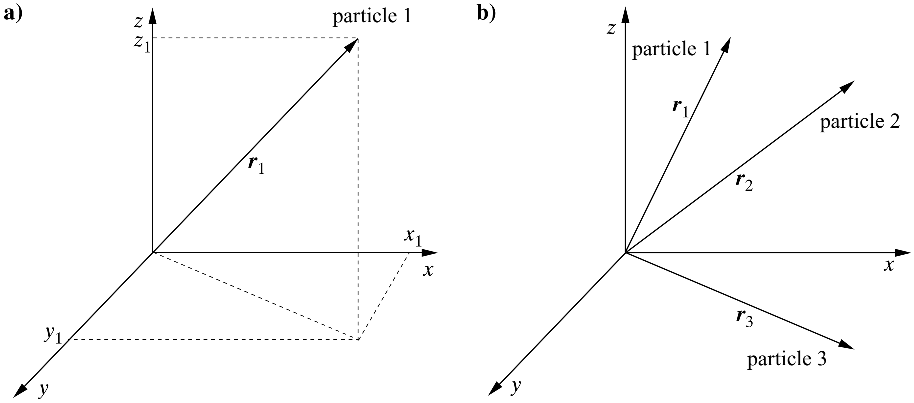


Fig. 1.2. An atom (molecule) in non-relativistic quantum mechanics. A Cartesian ("laboratory") coordinate system is introduced into three-dimensional space (a). We assume (b) that all the particles (electrons and nuclei) are point-like (figure shows their instantaneous positions) and interact only by electrostatic (Coulomb) forces.

---

[24] Assuming that the speed of light is infinite.

[25] Yes, we take only electrostatics, that is, Coulomb interactions. It is true that a moving charged particle creates a magnetic field, which influences its own and other particles' motion. This however (the Lorentz force) is taken into account in the *relativistic* approach to quantum mechanics.

moving along a single coordinate axis x (the mathematical foundations of quantum mechanics are given in Appendix B on p. 895).

Postulate I (on the quantum mechanical state)

wave function

The state of the system is described by the *wave function* $\Psi = \Psi(x, t)$, which depends on the coordinate of particle x at time t. Wave functions in general are complex functions of real variables. The symbol $\Psi^*(x, t)$ denotes the complex conjugate of $\Psi(x, t)$. The quantity

$$p(x, t) = \Psi^*(x, t)\Psi(x, t)\,\mathrm{d}x = \left|\Psi(x, t)\right|^2 \mathrm{d}x \tag{1.1}$$

gives the *probability that at time t the x coordinate of the particle lies in the small interval* $[x, x + \mathrm{d}x]$ (Fig. 1.3.a). The probability of the particle being in the interval (a, b) on the x axis is given by (Fig. 1.3.b): $\int_a^b |\Psi(x, t)|^2\,\mathrm{d}x$.

statistical interpretation

probability density

The probabilistic interpretation of the wave function was proposed by Max Born.[26] By analogy with the formula: mass = density × volume, the quantity $\Psi^*(x, t)\Psi(x, t)$ is called the *probability density* that a particle at time t has position x.

normalization

In order to treat the quantity $p(x, t)$ as a probability, at any instant t the wave function must satisfy the *normalization condition*:

$$\int_{-\infty}^{\infty} \Psi^*(x, t)\Psi(x, t)\,\mathrm{d}x = 1. \tag{1.2}$$

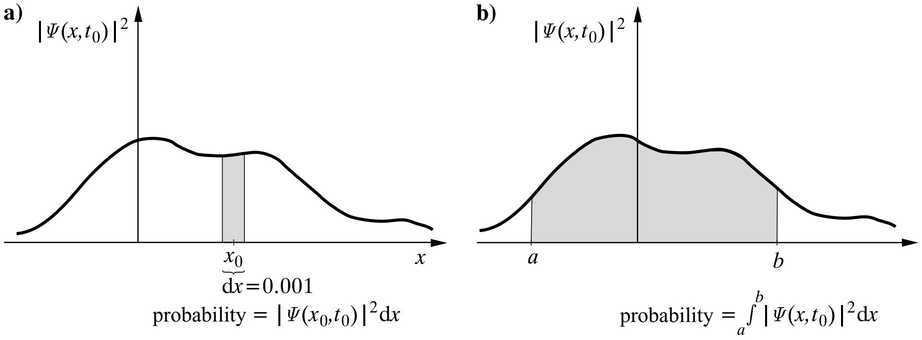


Fig. 1.3. A particle moves along the x axis and is in the state described by the wave function $\Psi(x, t)$. Fig. (a) shows how the probability of finding particle in an *infinitesimally small* section of the length dx at x_0 (at time $t = t_0$) is calculated. Fig. (b) shows how to calculate the probability of finding the particle at $t = t_0$ in a *section* (a, b).

---

[26] M. Born, *Zeitschrift für Physik* 37 (1926) 863.

All this may be generalized for more complex situations. For example, in three-dimensional space, the wave function of a single particle depends on position $\boldsymbol{r} = (x, y, z)$ and time: $\Psi(\boldsymbol{r}, t)$, and the *normalization condition* takes the form

$$\int_{-\infty}^{\infty} \mathrm{d}x \int_{-\infty}^{\infty} \mathrm{d}y \int_{-\infty}^{\infty} \mathrm{d}z\, \Psi^*(x, y, z, t)\Psi(x, y, z, t) \equiv \int \Psi^*(\boldsymbol{r}, t)\Psi(\boldsymbol{r}, t)\, \mathrm{d}V$$
$$\equiv \int \Psi^*(\boldsymbol{r}, t)\Psi(\boldsymbol{r}, t)\, \mathrm{d}^3\boldsymbol{r} = 1. \tag{1.3}$$

When integrating over whole space, for simplicity, the last two integrals are given without the integration limits, but they are there *implicitly, and this convention will be used by us throughout the book unless stated otherwise*.

For n particles (Fig. 1.4), shown by vectors $\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_n$ in three-dimensional space, the interpretation of the wave function is as follows. The probability P, that at a given time $t = t_0$, particle 1 is in the domain V_1, particle 2 is in the domain V_2 etc., is calculated as

$$P = \int_{V_1} \mathrm{d}V_1 \int_{V_2} \mathrm{d}V_2 \ldots \int_{V_n} \mathrm{d}V_n\, \Psi^*(\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_n, t_0)\Psi(\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_n, t_0)$$
$$\equiv \int_{V_1} \mathrm{d}^3\boldsymbol{r}_1 \int_{V_2} \mathrm{d}^3\boldsymbol{r}_2 \ldots \int_{V_n} \mathrm{d}^3\boldsymbol{r}_n\, \Psi^*(\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_n, t_0)\Psi(\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_n, t_0).$$

Often in this book we will perform what is called the normalization of a function, which is required if a probability is to be calculated. Suppose we have a unnormal-

normalization

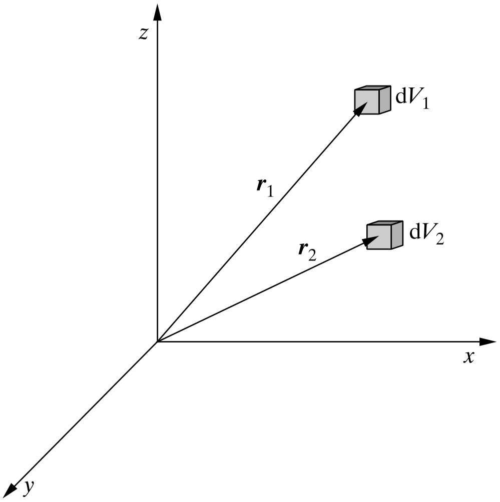


Fig. 1.4. Interpretation of a many-particle wave function, an example for two particles. The number $|\psi(\boldsymbol{r}_1, \boldsymbol{r}_2, t_0)|^2\, \mathrm{d}V_1\, \mathrm{d}V_2$ represents the probability that at $t = t_0$ particle 1 is in its box of volume dV_1 shown by vector $\boldsymbol{r}_1$ and particle 2 in its box of volume dV_2 indicated by vector $\boldsymbol{r}_2$.

ized function[27] ψ, that is

$$\int_{-\infty}^{\infty} \psi(x,t)^* \, \psi(x,t) \, \mathrm{d}x = A, \tag{1.4}$$

with $0 < A \neq 1$. To compute the probability ψ must be normalized, i.e. multiplied by a *normalization constant* N, such that the new function $\Psi = N\psi$ satisfies the normalization condition:

$$1 = \int_{-\infty}^{\infty} \Psi^*(x,t)\Psi(x,t)\,\mathrm{d}x = N^*N \int_{-\infty}^{\infty} \psi^*(x,t)\psi(x,t)\,\mathrm{d}x = A|N|^2.$$

Hence, $|N| = \frac{1}{\sqrt{A}}$. How is N calculated? One person may choose it as equal to $N = \frac{1}{\sqrt{A}}$, another: $N = -\frac{1}{\sqrt{A}}$, a third: $N = e^{1989i}\frac{1}{\sqrt{A}}$, and so on. There are, therefore, an infinite number of legitimate choices of the *phase* ϕ of the wave function $\Psi(x,t) = e^{i\phi}\frac{1}{\sqrt{A}}\psi$. Yet, when $\Psi^*(x,t)\Psi(x,t)$, is calculated, everyone will obtain *the same* result, $\frac{1}{A}\psi^*\psi$, because the phase disappears. In most applications, this is what will happen and therefore the computed physical properties will not depend on the choice of phase. There are cases, however, where the phase will be of importance.

phase

Postulate II (on operator representation of mechanical quantities)

The mechanical quantities that describe the particle (energy, the components of vectors of position, momentum, angular momentum, etc.) are represented by linear operators acting in Hilbert space (see Appendix B). There are two important examples of the operators: the operator of the particle's position $\hat{x} = x$ (i.e. multiplication by x, or $\hat{x} = x\cdot$, Fig. 1.5), as well as the operator of the (x-component) momentum $\hat{p}_x = -i\hbar\frac{\mathrm{d}}{\mathrm{d}x}$, where i stands for the imaginary unit.

Note that the mathematical form of the operators is always defined with respect to a *Cartesian* coordinate system.[28] From the given operators (Fig. 1.5) the operators of some other quantities may be constructed. The potential energy operator $\hat{V} = V(x)$, where $V(x)$ [the multiplication operator by the function $\hat{V}f = V(x)f$] represents a function of x called a potential. The kinetic energy operator of a single particle (in one dimension) $\hat{T} = \frac{\hat{p}_x{}^2}{2m} = -\frac{\hbar^2}{2m}\frac{\mathrm{d}^2}{\mathrm{d}x^2}$, and in three dimensions:

$$\hat{T} = \frac{\hat{p}^2}{2m} = \frac{\hat{p}_x{}^2 + \hat{p}_y{}^2 + \hat{p}_z{}^2}{2m} = -\frac{\hbar^2}{2m}\Delta, \tag{1.5}$$

---

[27] Eq. (1.2) not satisfied.

[28] Although they may then be transformed to other coordinates systems.

Mechanical quantity	Classical formula	Operator acting on f
coordinate	x	$\hat{x}f \overset{\text{def}}{=} xf$
momentum component	p_x	$\hat{p}_x f \overset{\text{def}}{=} -i\hbar\frac{\partial f}{\partial x}$
kinetic energy	$T = \frac{mv^2}{2} = \frac{p^2}{2m}$	$\hat{T}f = -\frac{\hbar^2}{2m}\Delta f$

Fig. 1.5. Mechanical quantities and the corresponding operators.

where the Laplacian Δ is

$$\Delta \equiv \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \tag{1.6}$$

and m denotes the particle's mass. The total energy operator, or *Hamiltonian* is the most frequently used:

Hamiltonian

$$\hat{H} = \hat{T} + \hat{V}. \tag{1.7}$$

An important feature of operators is that they may not *commute*,[29] i.e. for two particular operators $\hat{A}$ and $\hat{B}$ one *may* have $\hat{A}\hat{B} - \hat{B}\hat{A} \neq 0$. This property has important physical consequences (see below, postulate IV and the Heisenberg uncertainty principle). Because of the possible non-commutation of the operators, transformation of the classical formula (in which the commutation or non-commutation did not matter) may be non-unique. In such a case, from all the possibilities one has to choose an operator which is Hermitian. The operator $\hat{A}$ is Hermitian if, for any functions ψ and ϕ from its domain, one has

commutation

$$\int_{-\infty}^{\infty} \psi^*(x)\hat{A}\phi(x)\,\mathrm{d}x = \int_{-\infty}^{\infty} [\hat{A}\psi(x)]^*\phi(x)\,\mathrm{d}x. \tag{1.8}$$

Using what is known as *Dirac notation*, Fig. 1.6, the above equality may be written in a concise form:

$$\langle\psi|\hat{A}\phi\rangle = \langle\hat{A}\psi|\phi\rangle. \tag{1.9}$$

In Dirac notation[30] (Fig. 1.6) the key role is played by vectors *bra*: $\langle\ |$ and *ket*: $|\ \rangle$ denoting respectively $\psi^* \equiv \langle\psi|$ and $\phi \equiv |\phi\rangle$. Writing the bra and ket as $\langle\psi||\phi\rangle$

bra and ket

---

[29] Commutation means $\hat{A}\hat{B} = \hat{B}\hat{A}$.

[30] Its deeper meaning is discussed in many textbooks of quantum mechanics, e.g., A. Messiah, "*Quantum Mechanics*", vol. I, Amsterdam (1961), p. 245. Here we treat it as a convenient tool.

$\int \psi^* \phi \, d\tau \equiv \langle \psi \vert \phi \rangle$	Scalar product of two functions
$\int \psi^* \hat{A} \phi \, d\tau \equiv \langle \psi \vert \hat{A} \phi \rangle$ or $\langle \psi \vert \hat{A} \vert \phi \rangle$	Scalar product of ψ and $\hat{A}\phi$ or a matrix element of the operator $\hat{A}$
$\hat{Q} = \vert \psi \rangle \langle \psi \vert$	Projection operator on the direction of the vector ψ
$1 = \sum_k \vert \psi_k \rangle \langle \psi_k \vert$	Spectral resolution of identity. Its sense is best seen when acting on χ: $\chi = \sum_k \vert \psi_k \rangle \langle \psi_k \vert \chi \rangle = \sum_k \vert \psi_k \rangle c_k$.

Fig. 1.6. Dirac notation.

denotes $\langle \psi | \phi \rangle$, or the scalar product of ψ and ϕ in a unitary space (Appendix B), while writing it as $|\psi\rangle\langle\phi|$ means the operator $\hat{Q} = |\psi\rangle\langle\phi|$, because of its action on function $\xi = |\xi\rangle$ shown as: $\hat{Q}\xi = |\psi\rangle\langle\phi|\xi = |\psi\rangle\langle\phi|\xi\rangle = c\psi$, where $c = \langle\phi|\xi\rangle$.

The last formula in Fig. 1.6 (with $\{\psi_k\}$ standing for the complete set of functions) represents what is known as "spectral resolution of identity", best demonstrated when acting on an arbitrary function χ:

$$\chi = \sum_k |\psi_k\rangle\langle\psi_k|\chi\rangle = \sum_k |\psi_k\rangle c_k.$$

spectral resulution of identity

We have obtained the decomposition of the function (i.e. a vector of the Hilbert space) χ on its components $|\psi_k\rangle c_k$ along the basis vectors $|\psi_k\rangle$ of the Hilbert space. The coefficient $c_k = \langle\psi_k|\chi\rangle$ is the corresponding scalar product, the basis vectors ψ_k are normalized. This formula says something trivial: any vector can be retrieved when adding *all* its components together.

Postulate III (on time evolution of the state)

time evolution

TIME-DEPENDENT SCHRÖDINGER EQUATION
The time-evolution of the wave function Ψ is given by the equation

$$i\hbar \frac{\partial \Psi(x,t)}{\partial t} = \hat{H}\Psi(x,t), \tag{1.10}$$

where $\hat{H}$ is the system Hamiltonian, see eq. (1.7).

$\hat{H}$ may be time-dependent (energy changes in time, interacting system) or time-independent (energy conserved, isolated system). Eq. (1.10) is called the time-dependent Schrödinger equation (Fig. 1.7).

$$\boxed{\begin{array}{c}\Psi(x,t_0)\\ \downarrow\\ \hat{H}\Psi(x,t_0)\\ \downarrow\\ i\hbar\left(\dfrac{\partial\Psi}{\partial t}\right)_{t=t_0}\\ \downarrow\\ \Psi(x,t_0+\mathrm{d}t)=\Psi(x,t_0)-\dfrac{i}{\hbar}\hat{H}\Psi\,\mathrm{d}t\end{array}}$$

Fig. 1.7. Time evolution of a wave function. Knowing $\Psi(x,t)$ at a certain time $t=t_0$ makes it possible to compute $\hat{H}\Psi(x,t_0)$, and from this (using eq. (1.10)) one can calculate the time derivative $\frac{\partial\Psi(x,t_0)}{\partial t}=-\frac{i\hat{H}\Psi(x,t_0)}{\hbar}$. Knowledge of the wave function at time $t=t_0$, and of its time derivative, is sufficient to calculate the function a little later ($t=t_0+\mathrm{d}t$): $\Psi(x,t_0+\mathrm{d}t)\cong\Psi(x,t_0)+\frac{\partial\Psi}{\partial t}\,\mathrm{d}t$.

When $\hat{H}$ *is time-independent*, the general solution to (1.10) can be written as

$$\Psi(x,t)=\sum_{n=1}^{\infty}c_n\Psi_n(x,t), \tag{1.11}$$

where $\Psi_n(x,t)$ represent special solutions to (1.10), that have the form

$$\Psi_n(x,t)=\psi_n(x)\,e^{-i\frac{E_n}{\hbar}t}, \tag{1.12}$$

and c_n stand for some constants. Substituting the special solution into (1.10) leads to[31] what is known as the *time-independent Schrödinger equation*:

time-independent Schrödinger equation

SCHRÖDINGER EQUATION FOR STATIONARY STATES

$$\hat{H}\psi_n=E_n\psi_n,\quad n=1,2,\dots,M. \tag{1.13}$$

The equation represents an example of an *eigenvalue equation of the operator*; the functions ψ_n are called the eigenfunctions, and E_n are the eigenvalues of the operator $\hat{H}$ (M may be finite or infinite). It can be shown that E_n are real (see Appendix B, p. 895). The eigenvalues are the permitted energies of the system,

---

[31] $i\hbar\frac{\partial\Psi_n(x,t)}{\partial t}=i\hbar\frac{\partial\psi_n(x)\,e^{-i\frac{E_n}{\hbar}t}}{\partial t}=i\hbar\psi_n(x)\,\frac{\partial e^{-i\frac{E_n}{\hbar}t}}{\partial t}=i\hbar\psi_n(x)\,(-i\frac{E_n}{\hbar})e^{-i\frac{E_n}{t}}=E_n\psi_n e^{-i\frac{E_n}{\hbar}t}$. However, $\hat{H}\Psi_n(x,t)=\hat{H}\psi_n(x)\,e^{-i\frac{E_n}{\hbar}t}=e^{-i\frac{E_n}{\hbar}t}\hat{H}\psi_n(x)$, because the Hamiltonian does not depend on t. Hence, after dividing both sides of the equation by $e^{-i\frac{E_n}{\hbar}t}$ one obtains the time independent Schrödinger equation.

and the corresponding eigenfunctions Ψ_n are defined in eqs. (1.12) and (1.13). These states have a special character, because the probability given by (1.1) does not change in time (Fig. 1.8):

$$p_n(x,t) = \Psi_n^*(x,t)\Psi_n(x,t)\,\mathrm{d}x = \psi_n^*(x)\psi_n(x)\,\mathrm{d}x = p_n(x). \tag{1.14}$$

stationary states

Therefore, in determining these states, known as *stationary states*, one can apply the time–independent formalism based on the Schrödinger equation (1.13).

Postulate IV (on interpretation of experimental measurements)

This postulate pertains to ideal measurements, i.e. such that no error is introduced through imperfections in the measurement apparatus. We assume the measurement of the physical quantity A, represented by its time-independent operator $\hat{A}$ and, for the sake of simplicity, that the system is composed of a single particle (with one variable only).

- The result of a single measurement of a mechanical quantity A can *only* be an eigenvalue a_k of the operator $\hat{A}$.

The eigenvalue equation for operator $\hat{A}$ reads

$$\hat{A}\phi_k = a_k\phi_k, \quad k = 1, 2, \ldots, M. \tag{1.15}$$

quantization

The eigenfunctions ϕ_k are orthogonal[32] (cf. Appendix on p. 895). When the eigenvalues do not form a continuum, they are *quantized*, and then the corresponding eigenfunctions ϕ_k, $k = 1, 2, \ldots, M$, satisfy the orthonormality relations:[33]

$$\int_{-\infty}^{\infty} \phi_k^*(x)\phi_l(x)\,\mathrm{d}x \equiv \langle\phi_k|\phi_l\rangle \equiv \langle k|l\rangle = \delta_{kl} \equiv \begin{cases} 1, & \text{when } k = l, \\ 0, & \text{when } k \neq l, \end{cases} \tag{1.16}$$

where we have given several equivalent notations of the scalar product, which will be used in the present book, δ_{kl} means the Kronecker delta.

- Since eigenfunctions $\{\phi_k\}$ form the complete set, then the wave function of the system may be expanded as (M is quite often equal to ∞)

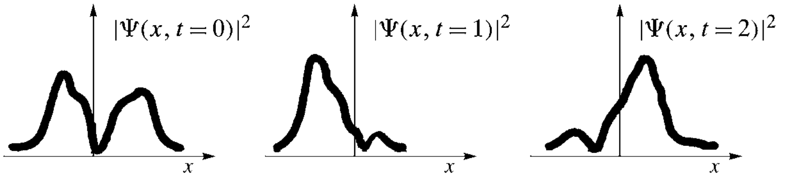
$$\psi = \sum_{k=1}^{M} c_k\phi_k, \tag{1.17}$$

---

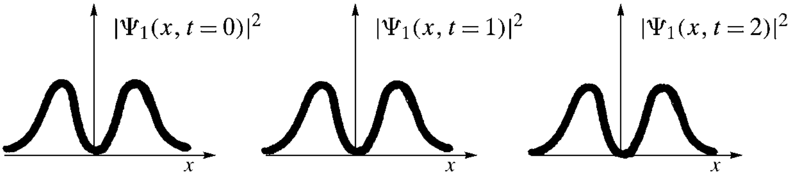
[32] If two eigenfunctions correspond to the same eigenvalue, they are not necessarily orthogonal, but they *can* still be orthogonalized (if they are linearly independent, see Appendix J, p. 977). Such orthogonal functions still remain the eigenfunctions of $\hat{A}$. Therefore, *one can always construct the orthonormal set of the eigenfunctions of a Hermitian operator*.

[33] If ϕ_k belong to continuum they cannot be normalized, but still can be made mutually orthogonal.

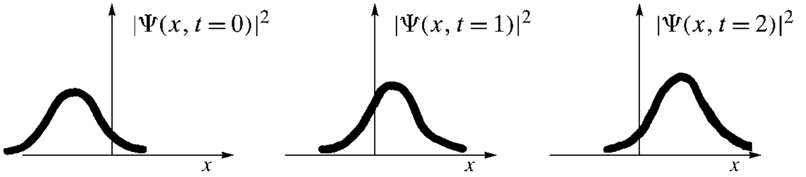
a) non-stationary state



b) stationary state



c) non-stationary state



d) stationary state

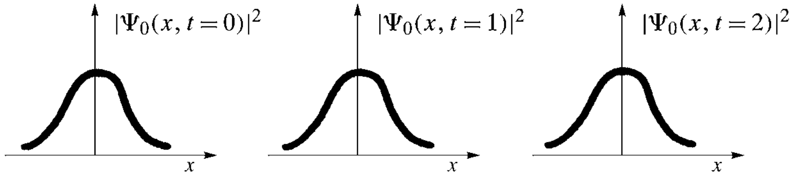


Fig. 1.8. Evolution of a starting wave function Ψ for a system shown as three snapshots ($t=0,1,2$) of $|\Psi(x,t)|^2$. In cases (a) and (c) it is seen that $|\Psi(x,t)|^2$ changes considerably when the time goes on: in (a) the function changes its overall shape, in (c) the function preserves its shape but travels along x axis. Both cases are therefore *non-stationary*. Cases (b) and (d) have a remarkable property that $|\Psi(x,t)|^2$ does not change at all in time. Hence, they represent examples of the *stationary* states. The reason why $|\Psi(x,t)|^2$ changes in cases (a) and (c) is that $\Psi(x,t)$ does not represent a pure stationary state [as in (b) and (d)], but instead is a linear combination of some stationary states.

where the c_k are in general, complex coefficients. From the normalization condition for ψ we have[34]

$$\sum_{k=1}^{M} c_k^* c_k = 1. \tag{1.18}$$

---

[34] $\langle\psi|\psi\rangle = 1 = \sum_{k=1}^{M}\sum_{l=1}^{M} c_k^* c_l \langle\phi_k|\phi_l\rangle = \sum_{k,l=1} c_k^* c_l \delta_{kl} = \sum_{k=1}^{M} c_k^* c_k$.

According to the axiom, the probability that the result of the measurement is a_k, is equal to $c_k^* c_k$.

collapse

If the wave function that describes the state of the system has the form given by (1.17) and does not reduce to a single term $\psi = \phi_k$, then the result of the measurement of the quantity A *cannot* be foreseen. We will measure *some* eigenvalue of the operator $\hat{A}$, but cannot predict which one. After the measurement is completed the wave function of the system represents the eigenstate that corresponds to the measured eigenvalue (known as the *collapse of the wave function*). According to the axiom, the only thing one may say about the measurements is that the *mean value* $\overline{a}$ of the quantity A (from many measurements) is to be compared with the following theoretical result[35] (Fig. 1.9)

$$\overline{a} = \sum_{k=1}^{M} c_k^* c_k a_k = \langle \psi | \hat{A} \psi \rangle, \tag{1.19}$$

where the normalization of ψ has been assumed.

If we have a special case, $\psi = \phi_k$ (all coefficients $c_l = 0$, except $c_k = 1$), the measured quantity is exactly equal a_k. From this it follows that if the wave function is an eigenfunction of operators of several quantities (this happens when the operators commute, Appendix B), then all these quantities when measured, produce with certainty, the eigenvalues corresponding to the eigenfunction.

mean value

The coefficients c can be calculated from (1.17). After multiplying by ϕ_l^* and integration, one has $c_l = \langle \phi_l | \psi \rangle$, i.e. c_l is identical to the overlap integral of the function ψ describing the state of the system and the function ϕ_l that corresponds to the eigenvalue a_l of the operator $\hat{A}$. *In other words, the more the eigenfunction corresponding to a_l resembles the wave function ψ*, the more frequently a_l will be measured.

---

[35]

$$\langle \psi | \hat{A} \psi \rangle = \left\langle \sum_{l=1}^{M} c_l \phi_l \middle| \hat{A} \sum_{k=1}^{M} c_k \phi_k \right\rangle = \sum_{k=1}^{M} \sum_{l=1}^{M} c_k^* c_l \langle \phi_l | \hat{A} \phi_k \rangle$$

$$= \sum_{k=1}^{M} \sum_{l=1}^{M} c_k^* c_l a_k \langle \phi_l | \phi_k \rangle = \sum_{k=1}^{M} \sum_{l=1}^{M} c_k^* c_l a_k \delta_{kl} = \sum_{k=1}^{M} c_k^* c_k a_k .$$

In case of degeneracy ($a_k = a_l = \cdots$) the probability is $c_k^* c_k + c_l^* c_l + \cdots$. This is how one computes the mean value of anything. Just take all possible distinct results of measurements, multiply each by its probability and sum up all resulting numbers.

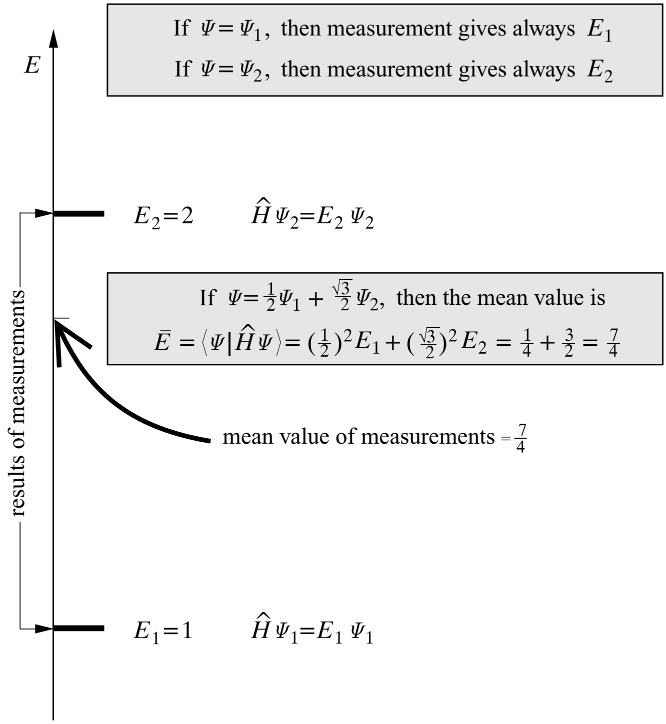


Fig. 1.9. The results of measurements of a quantity *A are the eigenvalues* of the operator $\hat{A}$: E_1 and E_2.

Postulate V (spin angular momentum)

spin

Spin of elementary particles. As will be shown in Chapter 3 (about relativistic effects) spin angular momentum will appear in a natural way. However, in nonrelativistic theory the existence of spin is *postulated*.[36]

An elementary particle has, besides its orbital angular momentum $\boldsymbol{r}\times\boldsymbol{p}$, an internal angular momentum (analogous to that associated with the rotation of a body about its own axis) called spin $\boldsymbol{S}=(S_x,S_y,S_z)$. Two quantities are measurable: the square of the spin length: $|\boldsymbol{S}|^2=S_x^2+S_y^2+S_z^2$ and one of its components, by convention, S_z. These quantities only take some particular values: $|\boldsymbol{S}|^2=s(s+1)\hbar^2$, $S_z=m_s\hbar$, where the spin magnetic quantum number $m_s=-s,-s+1,\ldots,s$.

---

[36] This has been forced by experimental facts, e.g., energy level splitting in a magnetic field suggested two possible electron states connected to internal angular momentum.

Enrico Fermi (1901–1954), Italian physicist, professor at universities in Florence, Rome, New York, and in 1941–1946 at the University of Chicago. Fermi introduced the notion of statistics for the particles with a half-integer spin number (called fermions) during the Florence period. Dirac made the same discovery independently, hence this property is called the Fermi–Dirac statistics. Young Fermi was notorious for being able to derive a formula from any domain of physics faster than someone sent to find it in textbooks. His main topic was nuclear physics. He played an important role in the A bomb construction in Los Alamos, and in 1942 he built the world's first nuclear reactor on a tennis court at the University of Chicago.

Fermi was awarded the Nobel Prize in 1938 "*for his demonstration of the existence of new radioactive elements and for results obtained with them, especially with regard to artificial radioactive elements*".

The spin quantum number s, *characteristic of the type of particle*[37] (often called simply its spin), can be written as: $s = \frac{n}{2}$, where n may be zero or a natural number ("an integer or half-integer" number). The particles with a half-integer s (e.g., $s = \frac{1}{2}$ for electron, proton, neutron, neutrino) are called fermions, the particles with an integer s (e.g., $s = 1$ for deuteron, photon;[38] $s = 0$ for meson π and meson K) are called bosons.

fermions

bosons

The magnetic[39] spin quantum number m_s quantizes the z component of the spin angular momentum.

Satyendra Nath Bose (1894–1974), Indian physicist, professor at Dakka and Calcutta, first recognized that particles with integer spin number have different statistical properties. Einstein contributed to a more detailed description of this statistics.

Thus, a particle with spin quantum number s has an additional (spin) degree of freedom, or an additional coordinate – *spin coordinate* σ. The spin coordinate differs widely from a spatial coordinate, because it takes only $2s + 1$ *discrete values* (Fig. 1.10) associated to $-s, -s + 1, \ldots, 0, \ldots, +s$.

Most often one will have to deal with electrons. For electrons, the spin coordinate σ takes two values, often called "up" and "down". We will (arbitrarily) choose $\sigma = -\frac{1}{2}$ and $\sigma = +\frac{1}{2}$, Fig. 1.11.a,b.

---

[37] Note, the length of the spin vector for an elementary particle is given by Nature once and for all. Thus, if there is any relation between the spin and the rotation of the particle about its own axis, it has to be a special relation. One cannot change the angular momentum of such a rotation.

[38] The photon represents a particle of zero mass. One can show that, instead of three possible m_s one has only two: $m_s = 1, -1$. We call these two possibilities "polarizations" ("parallel" and "perpendicular").

[39] The name is related to energy level splitting in a magnetic field, from which the number is deduced. A non-zero s value is associated to the magnetic dipole, which in magnetic field acquires $2s + 1$ energetically non-equivalent positions.

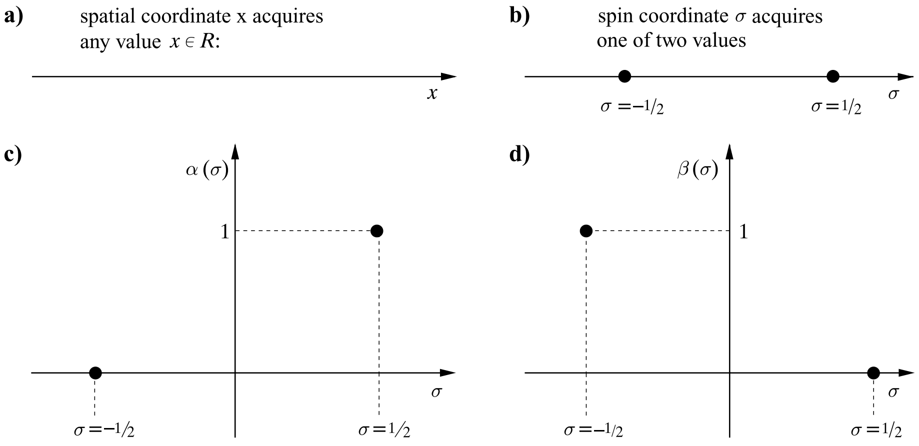


Fig. 1.10. Main differences between the spatial coordinate (x) and spin coordinate (σ) of an electron. (a) the spatial coordinate is *continuous*: it may take any value being a real number (b) the spin coordinate σ has a granular character (*discrete* values): for $s = \frac{1}{2}$ it can take only one of two values. One of the values is represented by $\sigma = -\frac{1}{2}$, the other to $\sigma = \frac{1}{2}$. Figs. (c,d) show, respectively, two widely used basis functions in the spin space: $\alpha(\sigma)$ and $\beta(\sigma)$.

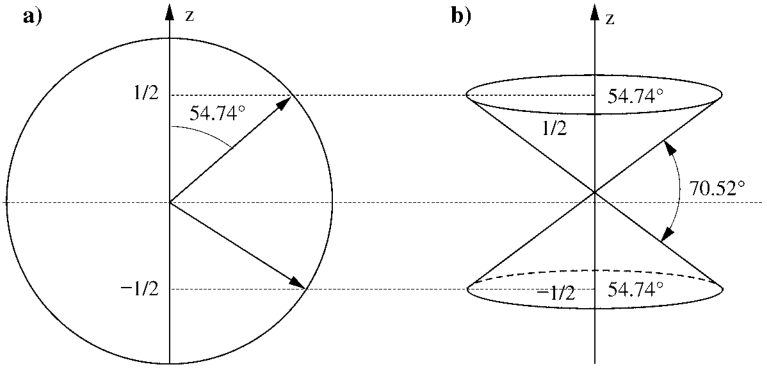


Fig. 1.11. Diagram of the spin angular momentum vector for a particle with spin quantum number $s = \frac{1}{2}$. The only measurable quantities are the spin length $\sqrt{s(s+1)}\hbar = \frac{\sqrt{3}}{2}\hbar$ and the projection of the spin on the quantization axis (chosen as coincident with the vertical axis z), which takes only the values $-s, -s+1, \ldots, +s$ in units $\hbar$, i.e. $S_z = -\frac{1}{2}\hbar, \frac{1}{2}\hbar$ (a). Possible positions of the spin angular momentum with respect to the quantization axis z (b) since the x and y components of the spin remain indefinite, one may visualize the same by locating the spin vector (of constant length $\sqrt{s(s+1)}\hbar$) *anywhere* on a cone surface that assures a given z component. Thus, one has $2s + 1 = 2$ such cones.

According to the postulate (p. 25), the square of the spin length is always the same and equal to $s(s+1)\hbar^2 = \frac{3}{4}\hbar^2$. The maximum projection of a vector on a chosen axis is equal to $\frac{1}{2}\hbar$, while the length of the vector is larger, equal to $\sqrt{s(s+1)}\hbar = \frac{\sqrt{3}}{2}\hbar$. We conclude that the vector of the spin angular momentum makes an angle θ with the axis, with $\cos\theta = \frac{1}{2}/\frac{\sqrt{3}}{2} = \frac{1}{\sqrt{3}}$. From this one obtains[40] $\theta = \arccos\frac{1}{\sqrt{3}} \approx 54.74°$. Fig. 1.11.b shows that the spin angular momentum has indefinite x and y components, while always preserving its length and projection on the z axis.

Spin basis functions for $s = \frac{1}{2}$. One may define (see Fig. 1.10.c,d) the complete set of orthonormal basis functions of the spin space of an electron:

$$\alpha(\sigma) = \begin{cases} 1 & \text{for } \sigma = \frac{1}{2} \\ 0 & \text{for } \sigma = -\frac{1}{2} \end{cases} \quad \text{and} \quad \beta(\sigma) = \begin{cases} 0 & \text{for } \sigma = \frac{1}{2} \\ 1 & \text{for } \sigma = -\frac{1}{2} \end{cases}$$

or, in a slightly different notation, as orthogonal unit vectors:[41]

$$|\alpha\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}; \qquad |\beta\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

α and β spin functions

Orthogonality follows from

$$\langle\alpha|\beta\rangle \equiv \sum_\sigma \alpha(\sigma)^*\beta(\sigma) = 0\cdot 1 + 1\cdot 0 = 0.$$

Similarly, normalization means that

$$\langle\alpha|\alpha\rangle \equiv \sum_\sigma \alpha(\sigma)^*\alpha(\sigma) = \alpha\left(-\frac{1}{2}\right)^*\alpha\left(-\frac{1}{2}\right) + \alpha\left(\frac{1}{2}\right)^*\alpha\left(\frac{1}{2}\right) = 0\cdot 0 + 1\cdot 1 = 1$$

etc.

Pauli matrices

We shall now construct operators of the spin angular momentum.

The following definition of spin operators is consistent with the postulate about spin.

$$\hat{S}_x = \frac{1}{2}\hbar\sigma_x$$

$$\hat{S}_y = \frac{1}{2}\hbar\sigma_y$$

$$\hat{S}_z = \frac{1}{2}\hbar\sigma_z\ ,$$

where the Pauli matrices of rank 2 are defined as:

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \qquad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

---

[40] In the general case, the spin of a particle may take the following angles with the quantization axis: $\arccos\frac{m_s}{\sqrt{s(s+1)}}$ for $m_s = -s, -s+1, \ldots, +s,$.

[41] In the same spirit as wave functions represent vectors: vector components are values of the function for various values of the variable.

Indeed, after applying $\hat{S}_z$ to the spin basis functions one obtains:

$$\hat{S}_z|\alpha\rangle \equiv \hat{S}_z \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \frac{1}{2}\hbar \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \frac{1}{2}\hbar \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \frac{1}{2}\hbar|\alpha\rangle,$$

$$\hat{S}_z|\beta\rangle \equiv \hat{S}_z \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \frac{1}{2}\hbar \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \frac{1}{2}\hbar \begin{pmatrix} 0 \\ -1 \end{pmatrix} = -\frac{1}{2}\hbar|\alpha\rangle.$$

Therefore, functions α and β represent the eigenfunctions of the $\hat{S}_z$ operator with corresponding eigenvalues $\frac{1}{2}\hbar$ and $-\frac{1}{2}\hbar$. How to construct the operator $\hat{\boldsymbol{S}}^2$? From Pythagoras' theorem, after applying Pauli matrices one obtains:

$$\begin{aligned}
\hat{S}^2|\alpha\rangle &= \hat{S}^2 \begin{pmatrix} 1 \\ 0 \end{pmatrix} = (\hat{S}_x^2 + \hat{S}_y^2 + \hat{S}_z^2) \begin{pmatrix} 1 \\ 0 \end{pmatrix} \\
&= \frac{1}{4}\hbar^2 \left\{ \begin{matrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} + \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \\ + \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \end{matrix} \right\} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \\
&= \frac{1}{4}\hbar^2 \begin{pmatrix} 1+1+1 & 0+0+0 \\ 0+0+0 & 1+1+1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \frac{3}{4}\hbar^2 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \\
&= \frac{3}{4}\hbar^2 \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \left[\frac{1}{2}\left(\frac{1}{2}+1\right)\hbar^2 \right] |\alpha\rangle.
\end{aligned}$$

The function $|\beta\rangle$ gives an identical eigenvalue.

Therefore, both basis functions α and β represent the eigenfunctions of $\hat{S}^2$ and correspond to the same eigenvalue. Thus, the definition of spin operators through Pauli matrices gives results identical to those postulated for S^2 and S_z, and the two formulations are equivalent. From Pauli matrices, it follows that the functions α and β are not eigenfunctions of $\hat{S}_x$ and $\hat{S}_y$ and that the following relations are satisfied[42]

$$[\hat{S}^2, \hat{S}_z] = 0, \quad [\hat{S}_x, \hat{S}_y] = i\hbar\hat{S}_z,$$

$$[\hat{S}_y, \hat{S}_z] = i\hbar\hat{S}_x, \quad [\hat{S}_z, \hat{S}_x] = i\hbar\hat{S}_y,$$

Wolfgang Pauli (1900–1958), German physicist, professor in Hamburg, at Technical University of Zurich, Institute for Advanced Studies in Princeton (USA), son of a physical chemistry professor in Vienna and a classmate of Werner Heisenberg. At the age of 20 he wrote a famous 200-page article on relativity theory for *Mathematical Encyclopedia*, afterwards edited as a book. A year later Pauli defended his doctoral dissertation under the supervision of Sommerfeld in Munich.

The renowned Pauli exclusion principle was proposed in 1924. Wolfgang Pauli received the Nobel Prize in 1945 "*for the discovery of the Exclusion Principle, also called the Pauli Principle*".

---

[42] These formulae are easy to memorize, since the sequence of the indices is always "rotational", i.e. $x, y, z, x, y, z, \ldots$

which is in agreement with the general properties of angular momenta[43] (Appendix on p. 955).

Spin of non-elementary particles. The postulate on spin pertains to an elementary particle. What about a system composed of such particles? Do they have spin? Spin represents angular momentum (a vector) and therefore the angular momentum vectors of elementary particles have to be added. A system composed of a number of elementary particles (each with its spin $\boldsymbol{s}_i$) has as a measurable quantity (an observable quantity), the square

$$|\boldsymbol{S}|^2 = S(S+1)\hbar^2$$

of the total spin vector,

$$\boldsymbol{S} = \boldsymbol{s}_1 + \boldsymbol{s}_2 + \cdots + \boldsymbol{s}_N,$$

and one of the components of $\boldsymbol{S}$ (denoted by $S_z = \sum_{i=1}^{N} s_{iz} = \hbar \sum_{i=1}^{N} m_{si}$):

$$S_z = M_S\hbar, \quad \text{for } M_S = -S, -S+1, \ldots, S,$$

where the number S stands (as in the case of a single particle) for an integer or half-integer non-negative number. Particular values of S (often called simply the spin) and of the spin magnetic number M_S depend on the directions of vectors $\boldsymbol{s}_i$. It follows that no excitation of a non-elementary boson (that causes another summing of the individual spin vectors) can change the particle to a fermion and *vice versa*. Systems with an even number of fermions are always bosons, while these with an odd number of fermions are always fermions.

Nuclei. The ground states of the important *nuclei* ^{12}C and ^{16}O correspond to $S = 0$, while those of ^{13}C, ^{15}N, ^{19}F have $S = \frac{1}{2}$.

Atoms and molecules. Does an atom as a whole represent a fermion or a boson? This depends on which atom and which molecule. Consider the hydrogen atom, composed of two fermions (proton and electron, both with spin number $\frac{1}{2}$). This is sufficient to deduce that here we are dealing with a boson. For similar reasons,

---

[43] Also, note that the mean values of S_x and S_y are both equal to zero in the α and β state, e.g., for the α state one has

$$\langle\alpha|\hat{S}_x\alpha\rangle == \left\langle\alpha\middle|\frac{1}{2}\hbar\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}\begin{pmatrix}1\\ 0\end{pmatrix}\right\rangle = \frac{1}{2}\hbar\langle\alpha|\beta\rangle = 0.$$

This means that in an external vector field (of direction z), when the space is no longer isotropic, only the projection of the total angular momentum on the field direction is conserved. A way to satisfy this is to recall the behaviour of a top in a vector field. The top rotates about its own axis, but the axis precesses about the field axis. This means that the total electron spin momentum moves on the cone surface making an angle of 54.74° with the external field axis in α state and an angle 180° − 54.74° in the β state. Whatever the motion, it must satisfy $\langle\alpha|\hat{S}_x\alpha\rangle = \langle\alpha|\hat{S}_y\alpha\rangle = 0$ and $\langle\beta|\hat{S}_x\beta\rangle = \langle\beta|\hat{S}_y\beta\rangle = 0$. No more information is available, but one may imagine the motion as a precession just like that of the top.

the sodium atom with 23 nucleons (each of spin $\frac{1}{2}$) in the nucleus and 11 electrons moving around it, also represents a boson.

When one adds together two electron spin vectors $\boldsymbol{s}_1 + \boldsymbol{s}_2$, then the maximum z component of the spin angular momentum will be (in $\hbar$ units): $|M_S| = |m_{s1} + m_{s2}| = \frac{1}{2} + \frac{1}{2} = 1$. This corresponds to the vectors $\boldsymbol{s}_1, \boldsymbol{s}_2$, called "parallel" to each other, while the minimum $|M_S| = |m_{s1} + m_{s2}| = \frac{1}{2} - \frac{1}{2} = 0$ means an "antiparallel" configuration of $\boldsymbol{s}_1$and $\boldsymbol{s}_2$ (Fig. 1.12).

The first situation indicates that for the state with parallel spins $S = 1$, for this S the possible $M_S = 1, 0, -1$. This means there are three states: $(S, M_S) = (1,1), (1,0), (1,-1)$. If no direction in space is privileged, then all the three states correspond to the same energy (triple degeneracy). This is why such a set of three states is called a *triplet state.* The second situation witnesses the existence of a state with $S = 0$, which obviously corresponds to $M_S = 0$. This state is called a *singlet state*.

triplet

singlet

Let us calculate the angle ω between the individual electronic spins:

$$\begin{aligned}
|\boldsymbol{S}|^2 &= (\boldsymbol{s}_1 + \boldsymbol{s}_2)^2 = s_1^2 + s_2^2 + 2\boldsymbol{s}_1 \cdot \boldsymbol{s}_2 \\
&= s_1^2 + s_2^2 + 2s_1 \cdot s_2 \cos\omega \\
&= \frac{1}{2}\left(\frac{1}{2} + 1\right)\hbar^2 \cdot 2 + 2\sqrt{\frac{1}{2}\left(\frac{1}{2}+1\right)}\sqrt{\frac{1}{2}\left(\frac{1}{2}+1\right)}\hbar^2 \cos\omega \\
&= \left(\frac{3}{2} + \frac{3}{2}\cos\omega\right)\hbar^2 = \frac{3}{2}(1+\cos\omega)\hbar^2.
\end{aligned}$$

SINGLET AND TRIPLET STATES:
For the singlet state $|\boldsymbol{S}|^2 = S(S+1)\hbar^2 = 0$, hence $1 + \cos\omega = 0$ and $\omega = 180°$. *This means the two electronic spins in the singlet state are antiparallel.*
For the triplet state $|\boldsymbol{S}|^2 = S(S+1)\hbar^2 = 2\hbar^2$, and hence $\frac{3}{2}(1+\cos\omega)\hbar^2 = 2\hbar^2$, i.e. $\cos\omega = \frac{1}{3}$, or $\omega = 70.52°$, see Fig. 1.12. *Despite forming the angle* $\omega = 70.52°$ *the two spins in the triplet state are said to be "parallel".*

The two electrons which we have considered may, for example, be part of a hydrogen molecule. Therefore, when considering electronic states, we may have to deal with singlets or triplets. However, in the same hydrogen molecule we have two protons, whose spins may also be "parallel" (orthohydrogen) or antiparallel (parahydrogen) . In parahydrogen the nuclear spin is $S = 0$, while in orthohydrogen $S = 1$. In consequence, there is only one state for parahydrogen ($M_S = 0$), and three states for orthohydrogen ($M_S = 1, 0, -1$).[44]

parahydrogen
and
orthohydrogen

---

[44] Since all the states have very similar energies (and therefore at high temperatures the Boltzmann factors are practically the same), there are three times as many molecules of orthohydrogen as of parahydrogen. Both states (ortho and para) differ slightly in their physicochemical characteristics.

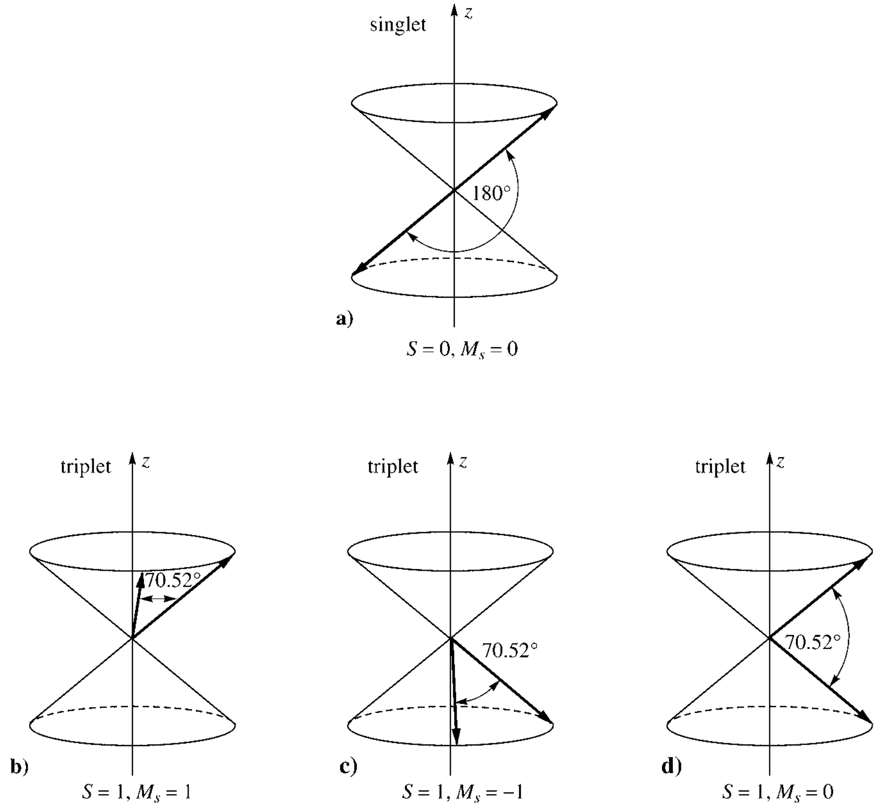


Fig. 1.12. Spin angular momentum for a system with two electrons (in general, particles with $s = \frac{1}{2}$). The quantization axis is arbitrarily chosen as the vertical axis z. Then, the spin vectors of individual electrons (see Fig. 1.11.b) may be thought to reside somewhere on the upper cone that corresponds to $m_{s1} = \frac{1}{2}$, or on the lower cone corresponding to $m_{s1} = -\frac{1}{2}$. For two electrons there are two spin eigenstates of $\hat{S}^2$. One has total spin quantum number $S = 0$ (singlet state); the other is triply degenerate (triplet state), and the three components of the state have $S = 1$ and $S_z = 1, 0, -1$ in $\hbar$ units. In the singlet state (a) the vectors s_1 and s_2 remain on the cones of different orientation, and have the opposite ("*antiparallel*") orientations, so that $s_1 + s_2 = \mathbf{0}$. Although their exact positions on the cones are undetermined (and moreover the cones themselves follow from the arbitrary choice of the quantization axis in space), they are always pointing in opposite directions. The three triplet components (b,c,d) differ by the direction of the total spin angular momentum (of constant length $\sqrt{S(S+1)}\hbar = \sqrt{2}\hbar$). The three directions correspond to three projections $M_S\hbar$ of spin momentum: $\hbar, -\hbar, 0$ for Figs. b, c, d, respectively. In each of the three cases the angle between the two spins equals $\omega = 70.52°$ (although in textbooks – including this one – they are said to be "*parallel*". In fact they are not, see the text).

Postulate VI (on the permutational symmetry)

Unlike classical mechanics, quantum mechanics is radical: it requires that two particles of the same kind (two electrons, two protons, etc.) should play the same role in the system, and therefore in its description enshrined in the wave function.[45] Quantum mechanics *guarantees* that the roles played in the Hamiltonian by two identical particles are identical. Within this philosophy, exchange of the labels of two identical particles (i.e. the exchange of their coordinates $x_1, y_1, z_1, \sigma_1 \leftrightarrow x_2, y_2, z_2, \sigma_2$. In short, $1 \leftrightarrow 2$) leads, at most, to a change of the phase ϕ of the wave function: $\psi(2,1) \to e^{i\phi}\psi(1,2)$, because in such a case $|\psi(2,1)| = |\psi(1,2)|$ (and this guarantees equal probabilities of both situations). However, when we exchange the two labels once more, we have to return to the initial situation: $\psi(1,2) = e^{i\phi}\psi(2,1) = e^{i\phi}e^{i\phi}\psi(1,2) = (e^{i\phi})^2\psi(1,2)$. Hence, $(e^{i\phi})^2 = 1$, i.e. $e^{i\phi} = \pm 1$. *Postulate VI says that* $e^{i\phi} = +1$ *refers to bosons, while* $e^{i\phi} = -1$ *refers to fermions*.[46]

bosons – symmetric function

fermions – antisymmetric function

The wave function ψ which describes identical *bosons* (i.e. spin integer particles) $1, 2, 3, \ldots, N$ has to be *symmetric* with respect to the exchange of coordinates x_i, y_i, z_i, σ_i and x_j, y_j, z_j, σ_j, i.e. if $x_i \leftrightarrow x_j$, $y_i \leftrightarrow y_j$, $z_i \leftrightarrow z_j$, $\sigma_i \leftrightarrow \sigma_j$, then $\psi(1, 2, \ldots, i, \ldots, j, \ldots, N) = \psi(1, 2, \ldots, j, \ldots, i, \ldots, N)$. If particles i and j denote identical *fermions*, the wave function must be *antisymmetric*, i.e. $\psi(1, 2, \ldots, i \ldots, j, \ldots, N) = -\psi(1, 2, \ldots, j, \ldots, i, \ldots, N)$.

Let us see the probability density that two identical fermions occupy the same position in space and, additionally, that they have the same spin coordinate $(x_1, y_1, z_1, \sigma_1) = (x_2, y_2, z_2, \sigma_2)$. We have: $\psi(1, 1, 3, 4, \ldots, N) = -\psi(1, 1, 3, 4, \ldots, N)$, hence $\psi(1, 1, 3, 4, \ldots, N) = 0$ and, of course, $|\psi(1, 1, 3, 4, \ldots, N)|^2 = 0$. Conclusion: two electrons of the same spin coordinate (we will sometimes say: "of the same spin") avoid each other. This is called the exchange or Fermi hole around each electron.[47] The reason for the hole is the antisymmetry of the electronic wave function, or in other words, the Pauli exclusion principle.[48]

Pauli exclusion principle

Thus, the probability density of finding two identical fermions *in the same position* and with *the same spin coordinate* is equal to zero. There is no such restriction for two identical bosons or two identical fermions with *different* spin coordinates. They can be at the same point in space.

---

[45] Everyday experience in classical world tells us the opposite, e.g., a car accident involving a Mercedes does not cause all copies of that particular model to have identical crash records.

[46] The postulate requires more than just making identical particles indistinguishable. It requires that *all* pairs of the identical particles follow the same rule.

[47] Besides that any two electrons avoid each other because of the same charge (Coulombic hole). Both holes (Fermi and Coulomb) have to be reflected in a good wave function. We will come back to this problem in Chapter 10.

[48] The Pauli exclusion principle is sometimes formulated in another way: two electrons cannot be in the same state (including spin). The connection of this strange phrasing (what does electron state mean?) with the above will become clear in Chapter 8.

Bose condensation

This is related to what is known as Bose condensation.[49]

* * *

Among the above postulates, the strongest controversy has always been associated with Postulate IV, which says that, except of some special cases, one cannot predict the result of a particular single measurement, but only its probability. More advanced considerations devoted to Postulate IV lead to the conclusion that there is no way (neither experimental protocol nor theoretical reasoning), to predict when and in which direction an excited atom will emit a photon. This means that quantum mechanics is not a deterministic theory.

The indeterminism appears however only in the physical space, while in the space of all states (Hilbert space) everything is perfectly deterministic. The wave function evolves in a deterministic way according to the time-dependent Schrödinger equation (1.10).

The puzzling way in which indeterminism operates will be shown below.

1.3 THE HEISENBERG UNCERTAINTY PRINCIPLE

Consider two mechanical quantities A and B, for which the corresponding Hermitian operators (constructed according to Postulate II), $\hat{A}$ and $\hat{B}$, give the commutator $[\hat{A}, \hat{B}] = \hat{A}\hat{B} - \hat{B}\hat{A} = i\hat{C}$, where $\hat{C}$ is a Hermitian operator.[50] This is what happens for example for $A = x$ and $B = p_x$. Indeed, for any differentiable function ϕ one has: $[\hat{x}, \hat{p_x}]\phi = -xi\hbar\phi' + i\hbar(x\phi)' = i\hbar\phi$, and therefore the operator $\hat{C}$ in this case means simply multiplication by $\hbar$.

From axioms of quantum mechanics one can prove that a product of errors (in the sense of standard deviation) of measurements of two mechanical quantities is greater than or equal to $\frac{1}{2}\langle[\hat{A}, \hat{B}]\rangle$, where $\langle[\hat{A}, \hat{B}]\rangle$ is the mean value of the commutator $[\hat{A}, \hat{B}]$.

This is known as the Heisenberg uncertainty principle.

---

[49] Carried out by Eric A. Cornell, Carl E. Wieman and Wolfgang Ketterle (Nobel Prize 2001 "*for discovering a new state of matter*"). In the Bose condensate the bosons (alkali metal atoms) are in the same place in a peculiar sense. The total wave function for the bosons was, to a first approximation, a product of *identical* nodeless wave functions for the particular bosons (this assures proper symmetry). Each of the wave functions extends considerably in space (the Bose condensate is as large as a fraction of a millimetre), but all have been centred *in the same point in space.*

[50] This is guaranteed. Indeed, $\hat{C} = -i[\hat{A}, \hat{B}]$ and then the Hermitian character of $\hat{C}$ is shown by the following chain of transformations $\langle f|\hat{C}g\rangle = -i\langle f|[\hat{A}, \hat{B}]g\rangle = -i\langle f|(\hat{A}\hat{B} - \hat{B}\hat{A})g\rangle = -i\langle(\hat{B}\hat{A} - \hat{A}\hat{B})f|g\rangle = \langle -i(\hat{A}\hat{B} - \hat{B}\hat{A})f|g\rangle = \langle\hat{C}f|g\rangle$.

Werner Karl Heisenberg (1901–1976) was born in Würzburg (Germany), attended high school in Munich, then (with his friend Wolfgang Pauli) studied physics at the Munich University under Sommerfeld's supervision. In 1923 he defended his doctoral thesis on turbulence in liquids. Reportedly, during the doctoral examination he had problems writing down the chemical reaction in lead batteries. He joined the laboratory of Max Born at Göttingen (following his friend Wolfgang) and in 1924 the Institute of Theoretical Physics in Copenhagen working under the supervision of Niels Bohr. A lecture delivered by Niels Bohr decided the future direction of his work. Heisenberg later wrote: "*I was taught optimism by Sommerfeld, mathematics in Göttingen, physics by Bohr*". In 1925 (only a year after being convinced by Bohr) Heisenberg developed a formalism, which became the first successful quantum theory. Then, in 1926 Heisenberg, Born and Jordan elaborated the formalism, which resulted in a coherent theory ("matrix mechanics"). In 1927 Heisenberg obtained a chair at Leipzig University, which he held until 1941 (when he became director of the Kaiser Wilhelm Physics Institute in Berlin). Heisenberg received the Nobel Prize in 1932 "*for the creation of quantum mechanics, the application of which has, inter alia, led to the discovery of the allotropic forms of hydrogen*".

In 1937 Werner Heisenberg was at the height of his powers. He was nominated professor and got married. However, just after returning from his honeymoon, the rector of the university called him, saying that there was a problem. In the SS weekly an article by Prof. Johannes Stark (a Nobel Prize winner and faithful Nazi) was about to appear claiming that Professor Heisenberg is not such a good patriot as he pretends, because he socialized in the past with Jewish physicists... Soon Professor Heisenberg was invited to SS headquarters at Prinz Albert Strasse in Berlin. The interrogation took place in the basement. On the raw concrete wall there was the interesting slogan "*Breath deeply and quietly*". One of the questioners was a Ph.D. student from Leipzig, who had once been examined by Heisenberg. The terrified Heisenberg told his mother about the problem. She recalled that in her youth she had made the acquaintance of Heinrich Himmler's mother. Frau Heisenberg paid a visit to Frau Himmler and asked her to pass a letter from her son to Himmler. At the beginning Himmler's mother tried to separate her maternal feelings for her beloved son from politics. She was finally convinced after Frau Heisenberg said "*we mothers should care about our boys*". After a certain time, Heisenberg received a letter from Himmler saying that his letter "coming through unusual channels" has been examined especially carefully. He promised to stop the attack. In the *post scriptum* there was a precisely tailored phrase: "*I think it best for your future, if for the benefit of your students, you would carefully separate scientific achievements from the personal and political beliefs of those who carried them out. Yours faithfully, Heinrich Himmler*" (after D. Bodanis, "$E = mc^2$", Fakty, Warsaw, 2001, p. 130).

Werner Heisenberg did not carry out any formal proof, instead he analyzed a Gedankenexperiment (an imaginary ideal experiment) with an electron interacting with an electromagnetic wave ("Heisenberg's microscope").

The formal proof goes as follows.

Recall the definition of the variance, or the square of the standard deviation $(\Delta A)^2$, of measurements of the quantity A:

$$(\Delta A)^2 = \langle \hat{A}^2 \rangle - \langle \hat{A} \rangle^2, \tag{1.20}$$

where $\langle X\rangle$ means the mean value of many measurements of the quantity X. The *standard deviation* ΔA *represents the width of the distribution of* A, i.e. *measures the error made.* Eq. (1.20) is equivalent to

$$(\Delta A)^2 = \big\langle (\hat{A} - \langle \hat{A}\rangle)^2 \big\rangle, \tag{1.21}$$

because $\langle(\hat{A} - \langle\hat{A}\rangle)^2\rangle = \langle\hat{A}^2 - 2\hat{A}\langle\hat{A}\rangle + \langle\hat{A}\rangle^2\rangle = \langle\hat{A}^2\rangle - 2\langle\hat{A}\rangle^2 + \langle\hat{A}\rangle^2 = \langle\hat{A}^2\rangle - \langle\hat{A}\rangle^2$. Consider the product of the standard deviations for the operators $\hat{A}$ and $\hat{B}$, taking into account that $\langle\hat{u}\rangle$ denotes (Postulate IV) the integral $\langle\Psi|\hat{u}|\Psi\rangle$ according to (1.19). One obtains (denoting $\hat{\mathcal{A}} = \hat{A} - \langle\hat{A}\rangle$ and $\hat{\mathcal{B}} = \hat{B} - \langle\hat{B}\rangle$; of course, $[\hat{\mathcal{A}}, \hat{\mathcal{B}}] = [\hat{A}, \hat{B}]$):

$$(\Delta A)^2 \cdot (\Delta B)^2 = \langle\Psi|\hat{\mathcal{A}}^2\Psi\rangle\langle\Psi|\hat{\mathcal{B}}^2\Psi\rangle = \langle\hat{\mathcal{A}}\Psi|\hat{\mathcal{A}}\Psi\rangle\langle\hat{\mathcal{B}}\Psi|\hat{\mathcal{B}}\Psi\rangle,$$

where the Hermitian character of the operators $\hat{\mathcal{A}}$ and $\hat{\mathcal{B}}$ is used. Now, let us use the Schwarz inequality (Appendix B) $\langle f_1|f_1\rangle\langle f_2|f_2\rangle \geqslant |\langle f_1|f_2\rangle|^2$:

$$(\Delta A)^2 \cdot (\Delta B)^2 = \langle\hat{\mathcal{A}}\Psi|\hat{\mathcal{A}}\Psi\rangle\langle\hat{\mathcal{B}}\Psi|\hat{\mathcal{B}}\Psi\rangle \geqslant |\langle\hat{\mathcal{A}}\Psi|\hat{\mathcal{B}}\Psi\rangle|^2.$$

Next,

$$\begin{aligned}\langle\hat{\mathcal{A}}\Psi|\hat{\mathcal{B}}\Psi\rangle &= \langle\Psi|\hat{\mathcal{A}}\hat{\mathcal{B}}\Psi\rangle = \langle\Psi|\{[\hat{\mathcal{A}}, \hat{\mathcal{B}}] + \hat{\mathcal{B}}\hat{\mathcal{A}}\}\Psi\rangle = i\langle\Psi|\hat{C}\Psi\rangle + \langle\Psi|\hat{\mathcal{B}}\hat{\mathcal{A}}\Psi\rangle\\ &= i\langle\Psi|\hat{C}\Psi\rangle + \langle\hat{\mathcal{B}}\Psi|\hat{\mathcal{A}}\Psi\rangle = i\langle\Psi|\hat{C}\Psi\rangle + \langle\hat{\mathcal{A}}\Psi|\hat{\mathcal{B}}\Psi\rangle^*.\end{aligned}$$

Hence,

$$i\langle\Psi|\hat{C}\Psi\rangle = 2i\,\mathrm{Im}\big\{\langle\hat{\mathcal{A}}\Psi|\hat{\mathcal{B}}\Psi\rangle\big\}$$

This means that $\mathrm{Im}\{\langle\hat{\mathcal{A}}\Psi|\hat{\mathcal{B}}\Psi\rangle\} = \frac{\langle\Psi|\hat{C}\Psi\rangle}{2}$, which gives $|\langle\hat{\mathcal{A}}\Psi|\hat{\mathcal{B}}\Psi\rangle| \geqslant \frac{|\langle\Psi|\hat{C}\Psi\rangle|}{2}$. Hence,

$$(\Delta A)^2 \cdot (\Delta B)^2 \geqslant \big|\langle\hat{\mathcal{A}}\Psi|\hat{\mathcal{B}}\Psi\rangle\big|^2 \geqslant \frac{|\langle\Psi|\hat{C}\Psi\rangle|^2}{4} \tag{1.22}$$

or, taking into account that $|\langle\Psi|\hat{C}\Psi\rangle| = |\langle\Psi|[\hat{A}, \hat{B}]\Psi\rangle|$ we have

$$\Delta A \cdot \Delta B \geqslant \frac{1}{2}|\langle\Psi|[\hat{A}, \hat{B}]\Psi\rangle|. \tag{1.23}$$

There are two important special cases:

(a) $\hat{C} = 0$, i.e. the operators $\hat{A}$ and $\hat{B}$ commute. We have $\Delta A \cdot \Delta B \geqslant 0$, i.e. *the errors can be arbitrarily small*. Both quantities therefore *can be* measured simultaneously without error.

(b) $\hat{C} = \hbar$, as in the case of $\hat{x}$ and $\hat{p}_x$. Then, $(\Delta A) \cdot (\Delta B) \geqslant \frac{\hbar}{2}$.

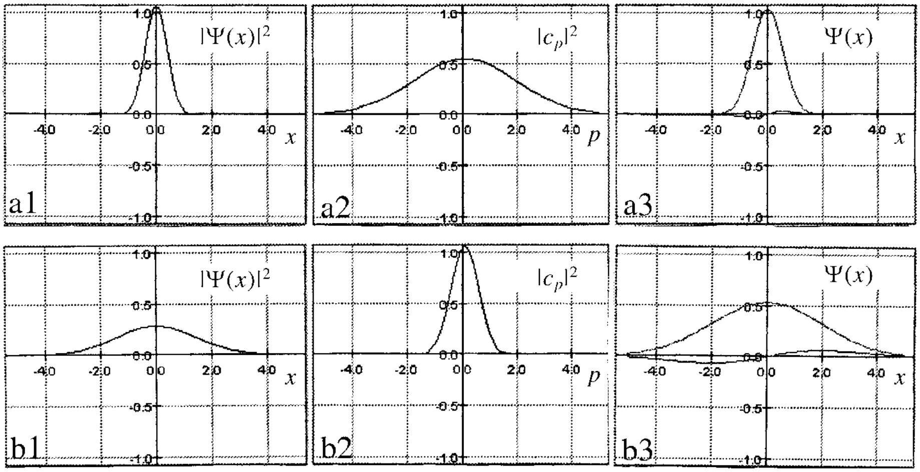


Fig. 1.13. Illustration of the Heisenberg uncertainty principle. (a1) $|\Psi(x)|^2$ as function of coordinate x. Wave function $\Psi(x)$ can be expanded in the infinite series $\Psi(x) = \sum_p c_p \exp(ipx)$, where p denotes the momentum. Note that each individual function $\exp(ipx)$ is an eigenfunction of momentum, and therefore if $\Psi(x) = \exp(ipx)$, a measurement of momentum gives exactly p. If however $\Psi(x) = \sum_p c_p \exp(ipx)$, then such a measurement yields a given p with the probability $|c_p|^2$. Fig. (a2) shows $|c_p|^2$ as function of p. As one can see a broad range of p (large uncertainty of momentum) assures a sharp $|\Psi(x)|^2$ distribution (small uncertainty of position). Simply the waves $\exp(ipx)$ to obtain a sharp peak of $\Psi(x)$ should exhibit a perfect constructive interference in a small region and a destructive interference elsewhere. This requires a lot of different p's, i.e. a broad momentum distribution. Fig. (a3) shows $\Psi(x)$ itself, i.e. its real (large) and imaginary (small) part. The imaginary part is non-zero because of small deviation from symmetry. Figs. (b1–b3) show the same, but this time a narrow p distribution gives a broad x distribution.

In particular, for $\hat{A} = \hat{x}$ and $\hat{B} = \hat{p}_x$, if quantum mechanics is valid, *one cannot measure the exact position and the exact momentum of a particle*. When the precision with which x is measured increases, the particle's momentum has so wide a distribution that the error in determining p_x is huge, Fig. 1.13.[51]

1.4 THE COPENHAGEN INTERPRETATION

In the 1920s and 1930s, Copenhagen for quantum mechanics was like Rome for catholics, and Bohr played the role of the president of the Quantum Faith Congregation.[52] The picture of the world that emerged from quantum mechanics was "diffuse" compared to classical mechanics. In classical mechanics one could mea-

---

[51] There is an apocryphal story about a police patrol stopping Professor Heisenberg for speeding. The policeman asks: "*Do you know how fast you were going when I stopped you?*" Heisenberg answered: "*I have no idea, but can tell you precisely where you stopped me*".

[52] Schrödinger did not like the Copenhagen interpretation. Once Bohr and Heisenberg invited him for a Baltic Sea cruise and indoctrinated him so strongly, that Schrödinger became ill and stopped participating in their discussions.

sure a particle's position and momentum with a desired accuracy,[53] whereas the Heisenberg uncertainty principle states that this is simply *impossible*.

Bohr presented a philosophical interpretation of the world, which at its foundation had in a sense a *non-reality of the world*.

According to Bohr, before a measurement on a particle is made, *nothing* can be said about the value of a given mechanical quantity, unless the wave function represents an eigenfunction of the operator of this mechanical quantity. Moreover, except in this case, the particle *does not have* any fixed value of mechanical quantity at all.

collapse

decoherence

A measurement gives a value of the mechanical property (A). Then, according to Bohr, after the measurement is completed, the state of the system changes (the so called *wave function collapse* or, more generally, *decoherence*) to the state described by *an eigenfunction of the corresponding operator* $\hat{A}$, and as the *measured value one obtains the eigenvalue corresponding to the wave function*. According to Bohr, there is no way to foresee which eigenvalue one will get as the result of the measurement. However, one can calculate the *probability* of getting a particular eigenvalue. This probability may be computed as the square of the overlap integral (cf. p. 24) of the initial wave function and the eigenfunction of $\hat{A}$.

1.5 HOW TO DISPROVE THE HEISENBERG PRINCIPLE? THE EINSTEIN–PODOLSKY–ROSEN RECIPE

EPR "experiment"

The Heisenberg uncertainty principle came as a shock. Many scientists felt a strong imperative to prove that the principle is false. One of them was Albert Einstein, who used to play with ideas by performing some (as he used to say) imaginary ideal experiments (in German *Gedankenexperiment*) in order to demonstrate internal contradictions in theories. Einstein believed in the reality of our world. With his colleagues Podolsky and Rosen ("EPR team") he designed a special Gedankenexperiment.[54] It represented an attempt to disprove the Heisenberg uncertainty principle and to show that one *can* measure the position and momentum of a particle *without any error*. To achieve this, the gentlemen invoked a second particle.

The key statement of the whole reasoning, given in the EPR paper, was the following: "*If, without in any way disturbing a system, we can predict with certainty (i.e. with probability equal to unity) the value of a physical quantity, then there exists an element of physical reality corresponding to this physical quantity*". EPR considered

---

[53] This is an exaggeration. Classical mechanics also has its own problems with uncertainty. For example, obtaining the same results for a game of dice would require a perfect reproduction of the initial conditions, which is never feasible.

[54] A. Einstein, B. Podolsky, N. Rosen, *Phys. Rev.* 47 (1935) 777.

a coordinate system fixed in space and two particles: 1 with coordinate x_1 and momentum p_{x1} and 2 with coordinate x_2 and momentum p_{x2}, the total system being in a state with a well defined total momentum: $P = p_{x1} + p_{x2}$ and well defined relative position $x = x_1 - x_2$. The meaning of the words "well defined" is that, according to quantum mechanics, there is a possibility of the *exact measurement* of the two quantities (x and P), because the two operators $\hat{x}$ and $\hat{P}$ do commute.[55] At this point, Einstein and his colleagues and the great interpreters of quantum theory, agreed.

We now come to the crux of the real controversy.

The particles interact, then separate and fly far away (at any time we are able to measure exactly both x and P). When they are extremely far from each other (e.g., one close to us, the other one millions of light years away), we begin to suspect that each of the particles may be treated as free. Then, we decide to measure p_{x1}. However, after we do it, we know *with absolute certainty the momentum of the second particle* $p_{x2} = P - p_{x1}$, and this knowledge has been acquired *without any perturbation of particle* 2. According to the above cited statement, one has to admit that p_{x2} represents an element of physical reality. So far so good. However, we *might have decided* with respect to particle 1 to measure its *coordinate* x_1. If this happened, then we would know *with absolute certainty the position of the second particle*, $x_2 = x - x_1$, without perturbing particle 2 at all. Therefore, x_2, as p_{x2}, is an element of physical reality. The Heisenberg uncertainty principle says that it is *impossible* for x_2 and p_{x2} to be exactly measurable quantities. Conclusion: the Heisenberg uncertainty principle is wrong, and quantum mechanics is at least incomplete!

A way to defend the Heisenberg principle was to treat the two particles as an indivisible total system and reject the supposition that the particles are independent, even if they are millions light years apart. This is how Niels Bohr defended himself against Einstein (and his two colleagues). He said that the state of the total system in fact never fell apart into particles 1 and 2, and still is in what is known as *entangled quantum state*[56] of the system of particles 1 and 2 and

entangled states

> any measurement influences the state of the system as a whole, *independently of the distance of particles* 1 and 2.

This reduces to the statement that measurement manipulations on particle 1 influence the results of measurements on particle 2. This correlation between measurements on particles 1 and 2 has to take place *immediately, regardless of the space*

---

[55] Indeed, $\hat{x}\hat{P} - \hat{P}\hat{x} = (\hat{x}_1 - \hat{x}_2)(\hat{p}_{x1} + \hat{p}_{x2}) - (\hat{p}_{x1} + \hat{p}_{x2})(\hat{x}_1 - \hat{x}_2) = [\hat{x}_1, \hat{p}_{x1}] - [\hat{x}_2, \hat{p}_{x2}] + [\hat{x}_1, \hat{p}_{x2}] - [\hat{x}_2, \hat{p}_{x1}] = +i\hbar - i\hbar + 0 - 0 = 0$.

[56] To honour Einstein, Podolsky and Rosen the entanglement of states is sometimes called the EPR effect.

that separates them. This is a shocking and non-intuitive feature of quantum mechanics. This is why it is often said, also by specialists, that quantum mechanics cannot be understood. One can apply it successfully and obtain an excellent agreement with experiment, but there is something strange in its foundations. This represents a challenge: an excellent theory, but based on some unclear foundations.

In the following, some precise experiments will be described, in which it is shown that quantum mechanics is right, however absurd it looks.

1.6 IS THE WORLD REAL?

BILOCATION

Assume that the world (stars, Earth, Moon, you and me, table, proton, etc.) exists objectively. This one may suspect from everyday observations. For example, the Moon is seen by many people, who describe it in a similar way.[57] Instead of the Moon, let us begin with something simpler: how about electrons, protons or other elementary particles? This is an important question because the world as we know it – including the Moon – is mainly composed of protons.[58] Here one encounters a mysterious problem. I will try to describe it by reporting results of several experiments.

Following Richard Feynman,[59] imagine two slits in a wall. Every second (the time interval has to be large enough to be sure that we deal with properties of a *single* particle) we send an electron towards the slits. There is a screen behind the two slits, and when an electron hits the screen, there is a flash (fluorescence) at the point of collision. Nothing special happens. Some electrons will not reach the screen at all, but traces of others form a pattern, which seems quite chaotic. The experiment looks monotonous and boring. Just a flash here, and another there. One cannot predict where a particular electron will hit the screen. But suddenly we begin to suspect that there is some regularity in the traces, Fig. 1.14.

---

[57] This *may* indicate that the Moon exists independently of our observations and overcome importunate suspicions that the Moon ceases to exist, when we do not look at it. Besides, there are people who claim to have seen the Moon from very close and even touched it (admittedly through a glove) and this slightly strengthens our belief in the Moon's existence. First of all, one has to be cautious. For example, some chemical substances, hypnosis or an ingenious set of mirrors may cause some people to be convinced about the reality of some phenomena, while others do not see them. Yet, would it help if even everybody saw? We should not verify serious things by voting. The example of the Moon also intrigued others, cf. D. Mermin, "*Is the Moon there, when nobody looks?*", *Phys. Today* 38 (1985) 38.

[58] In the darkest communist times a colleague of mine came to my office. Conspiratorially, very excited, he whispered: "*The proton decays!!!*" He just read in a government newspaper that the lifetime of proton turned out to be finite. When asked about the lifetime, he gave an astronomical number, something like 10^{30} years or so. I said: "*Why do you look so excited then and why all this conspiracy?*" He answered: "*The Soviet Union is built of protons, and therefore is bound to decay as well!*"

[59] After Richard Feynman, "The Character of Physical Law", MIT Press, 1967.

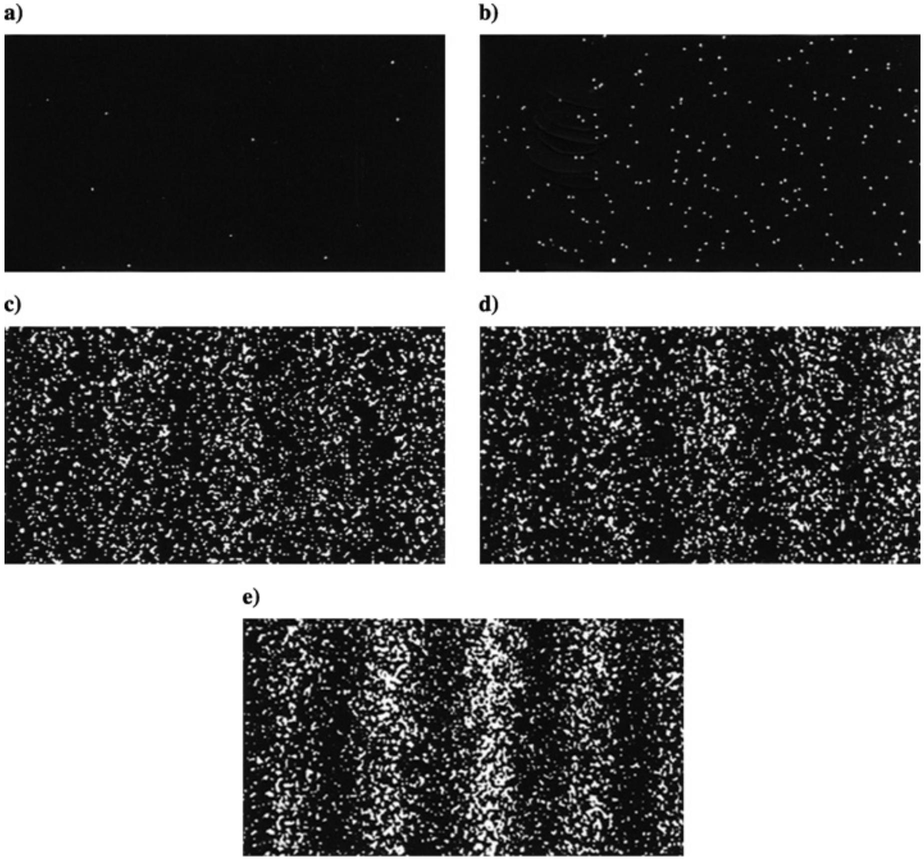


Fig. 1.14. Two-slit electron interference pattern registered by Akira Tonomura. (a) 10 electrons (b) 100 electrons (c) 3000 electrons – one begins to suspect something (d) 20000 electrons – no doubt, we will have a surprize (e) 70000 electrons – here it is! Conclusion: there is only one possibility – each electron went through the two slits. Courtesy of Professor Akira Tonomura.

A strange pattern appears on the screen: a number of high concentrations of traces is separated by regions of low concentration. This resembles the interference of waves, e.g., a stone thrown into water causes interference behind two slits: an alternation of high and low amplitude of water level. Well, but what has an electron in common with a wave on the water surface? The interference on water was possible, because there were two sources of waves (the Huygens principle) – two slits.

The common sense tells us that nothing like this could happen with the electron, because, firstly, the electron could not pass through *both* slits, and, secondly, unlike the waves, the electron has hit a tiny spot on the screen (transferring its energy). Let us repeat the experiment with a *single* slit. The electrons

Christiaan Huygens (1629–1695), Dutch mathematician, physicist and astronomer. Huygens was the first to construct a useful pendulum clock.

go through the slit and make flashes on the screen here and there, but there is only a single major concentration region (just facing the slit) fading away from the centre (with some minor minima).

This result should make you feel faint. Why? You would like the Moon, a proton or an electron to be solid objects, wouldn't you? All investigations made so far indicate that the electron is a point-like elementary particle. If, in the experiments we have considered, the electrons were to be divided into two classes: those that went through slit 1 and those that passed slit 2, then the electron patterns would be different. The pattern with the two slits *had to be* the sum of the patterns corresponding to only one open slit (facing slit 1 and slit 2). *We do not have that picture*.

The *only* explanation for this interference of the electron with itself is that with the two slits open it went through *both*.

Clearly, the two parts of the electron united somehow and caused the flash at a single point on the screen. The quantum world is really puzzling. Despite the fact that the wave function is delocalized, the measurement gives its single point position (decoherence). How could an electron pass simultaneously through two slits? We do not understand this, but this is what happens.

Maybe it is possible to pinpoint the electron passing through two slits? Indeed, one may think of the Compton effect: a photon collides with an electron, changes its direction and this can be detected ("a flash on the electron"). When one prepares two such ambushes at the two open slits, it turns out that the flash is *always on a single slit, not on both*. This cannot be true! If it were true, then the pattern would be of a NON-interference character (and had to be the sum of the two one-slit patterns), but we have the interference. No. There is no interference. *Now*, the pattern does not show the interference. The interference was when the electrons were not observed. When we observe them, there is no interference.[60] Somehow we perturb the electron's momentum (the Heisenberg principle) and the interference disappears.

We have to accept that the electron passes through two slits. This is a blow to those who believe in the reality of the world. Maybe it only pertains to the electron, maybe the Moon is something completely different? A weak hope. The same thing

---

[60] Even if an electron has been pinpointed just *after* passing the slit region, i.e. already on the screen side (leaving the slit system *behind*). One might think it is too late, it has already passed the interference region. This has serious consequences, known as the problem of "*delayed choice*" (cf. the experiments with photons at the end of this chapter).

happens to proton. Sodium atoms were also found to interfere.[61] A sodium atom, of diameter of a few Å, looks like an ocean liner, when compared to a child's toy boat of a tiny electron (42 000 times less massive). And this ocean liner passed through two slits separated by thousands of Å. A similar interference was observed for the fullerene,[62] a giant C_{60} molecule (in 2001 also for C_{70}), about million times more massive than the electron. It is worth noting that after such adventure the fullerene molecule remained intact: somehow all its atoms, with the details of their chemical bonds, preserved their nature. There is something intriguing in this.

1.7 THE BELL INEQUALITY WILL DECIDE

John Bell proved a theorem in 1964 that pertains to the results of measurements carried out on particles and some of the inequalities they have to fulfil. The theorem pertains to the basic logic of the measurements and is valid independently of the kind of particles and of the nature of their interaction. The theorem soon became very famous, because it turned out to be a useful tool allowing us to verify some fundamental features of our knowledge about the world.

Imagine a launching gun[63] (Fig. 1.15), which ejects a series of pairs of identical rectangular bars flying along a straight line (no gravitation) in opposite directions (opposite velocities). The axes of the bars are always parallel to each other and always perpendicular to the straight line. The launching machine is constructed in such a way that it can rotate about the straight line, and that any two launching series are absolutely identical. At a certain distance from the launching machine there are two rectangular slits A and B (the same on both sides). If the bar's longer axis coincides with the longer dimension of the slit then the bar will go through for sure and will be registered as "1", i.e. "it has arrived" by the detector. If the bar's longer axis coincides with the shorter axis of the slit, then the bar will not go through for sure, and will be detected as "0". For other angles between the bar and slit axes the bar will sometimes go through (when it fits the slit), sometimes not (when it does not fit the slit).[64]

John Stuart Bell (1928–1990), Irish mathematician at Centre Européen de la Recherche Nucleaire (CERN) in Geneva. In the 1960s Bell reconsidered an old controversy of locality versus non-locality, hidden variables, etc., a subject apparently exhausted after exchange of ideas between Einstein and Bohr.

---

[61] To observe such phenomena the slit distance has to be of the order of the de Broglie wave length, $\lambda = h/p$, where h is the Planck constant, and p is the momentum. Cohen-Tannoudji lowered the temperature to such an extent that the momentum was close to 0, and λ could be of the order of thousands of Å.

[62] M. Arndt, O. Nairz, J. Voss-Andreae, C. Keller, G. van der Zouw, A. Zeilinger, *Nature* 401 (1999) 680.

[63] See, e.g., W. Kołos, Proceedings of the IV Castel Gandolfo Symposium, 1986.

[64] Simple reasoning shows that for a bar of length L, two possibilities: "to go through" and "not to go through" are equally probable (for a bar of zero width) if the slit width is equal to $\frac{L}{\sqrt{2}}$.

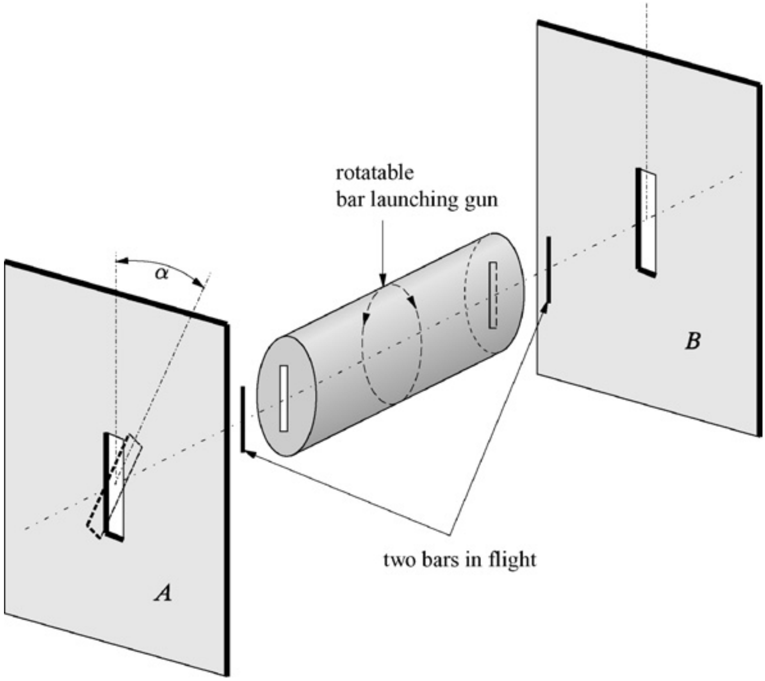


Fig. 1.15. Bell inequalities. A bar launching gun adopts some positions when rotating about the axis. Each time the full magazine of bars is loaded. The slits also may be rotated about the axis. The bars arrive at slits A and B. Some will go through and be detected.

Having prepared the launching gun (our magazine contains 16 pairs of bars) we begin our experiments. Four experiments will be performed. Each experiment will need the full magazine of bars. In the first experiment the two slits will be parallel. This means that the fate of both bars in any pair will be exactly the same: if they go through, they will both do it, if they are stopped by the slits, they will both be stopped. Our detectors have registered (we group the 16 pairs in clusters of 4 to make the sequence more transparent):

Experiment I (angle 0)				
Detector A:	1001	0111	0010	1001
Detector B:	1001	0111	0010	1001

Now, we repeat Experiment I, but this time slit A will be rotated by a small angle α (Experiment II). At the slit B nothing has changed, and therefore we must obtain there exactly the same sequence of zeros and ones as in Experiment I. At slit A, however, the results may be different. Since the rotation angle is small, the difference list will be short. We might get the following result

Experiment II (angle α)				
Detector A:	10**1**1	0111	0010	**0**001
Detector B:	1001	0111	0010	1001

There are two differences (highlighted in bold) between the lists for the two detectors.

Now for Experiment III. This time slit A comes back to its initial position, but slit B is rotated by $-\alpha$. Because of our perfect gun, we must obtain at detector A the same result as in Experiment I. However, at B we find some difference with respect to Experiments I and II:

Experiment III (angle $-\alpha$)				
Detector A:	1001	0111	0010	1001
Detector B:	1001	0**0**11	0**1**10	1001

There are two differences (bold) between the two detectors.

We now carry out Experiment IV. We rotate slit A by angle α, and slit B by angle $-\alpha$. Therefore, at Detector A we obtain the same results as in Experiment II, while at Detector B – the same as in Experiment III. Therefore, we detect:

Experiment IV (angle 2α)				
Detector A:	10**1**1	0111	0010	**0**001
Detector B:	1001	0**0**11	0**1**10	1001

Now there are four differences between Detector A and Detector B. In Experiment IV *the number of differences could not be larger (Bell inequality)*. In our case it could be four or fewer. When would it be fewer? When accidentally the bold figures (i.e. the differences of Experiments II and III with respect to those of Experiment I) coincide. In this case this would be counted as a difference in Experiments II and III, while in Experiment IV it would not be counted as a difference.

Thus, we have demonstrated

BELL INEQUALITY:

$$N(2\alpha) \leqslant 2N(\alpha), \tag{1.24}$$

where N stands for the number of measurement differences. The Bell inequality was derived under assumption that whatever happens at slit A it does not influence that which happens at slit B (this is how we constructed the counting tables) and that the two flying bars have, maybe unknown for the observer, only a real (definite) direction in space (the same for both bars).

It would be interesting to perform a real experiment similar to Bell's to confirm the Bell inequality. This opens the way for deciding in a physical experiment whether:

- elementary particles are classical (though extremely small) objects that have some well defined attributes irrespective of whether we observe them or not (Einstein's view)
- elementary particles do not have such attributes and only measurements themselves make them have measured values (Bohr's view).

1.8 INTRIGUING RESULTS OF EXPERIMENTS WITH PHOTONS

Aspect et al., French scientists from the Institute of Theoretical and Applied Optics in Orsay published the results of their experiments with photons.[65] The excited calcium atom emitted pairs of photons (analogues of our bars), which moved in opposite directions and had the same polarization. After flying about 6 m they both met the polarizers – analogues of slits A and B in the Bell procedure. A polarizer allows a photon with polarization state $|0\rangle$, or "parallel" (to the polarizer axis), always pass through, and always rejects any photon in the polarization state $|1\rangle$, or "perpendicular" (indeed perpendicular to the above "parallel" setting). When the polarizer is rotated about the optical axis by an angle, it will pass through a percentage of the photons in state $|0\rangle$ and a percentage of the photons in state $|1\rangle$. When both polarizers are in the "parallel" setting, there is perfect correlation between the two photons of each pair, i.e. exactly as in Bell's Experiment I. In the photon experiment, this correlation was checked for 50 million photons every second for about 12 000 seconds.

Bell's experiments II–IV have been carried out. Common sense indicates that, even if the two photons in a pair have random polarizations (perfectly correlated though always the same – like the bars), they still have *some* polarizations, i.e. maybe unknown but definite (as in the case of the bars, i.e. what E, P and R believed happens). *Hence, the results of the photon experiments would have to fulfil the Bell inequality.* However, the photon experiments have shown that the Bell inequality is violated, but still the results are in accordance with the prediction of quantum mechanics.

There are therefore only two possibilities (compare the frame at the end of the previous section):

(a) either the measurement on a photon carried out at polarizer A (B) results in some *instantaneous interaction* with the photon at polarizer B(A), or/and
(b) the polarization of any of these photons is completely *indefinite* (even if the polarizations of the two photons are fully correlated, i.e. the same) and only the *measurement on one of the photons at A (B) determines its polarization*, which

---

[65] A. Aspect, J. Dalibard, G. Roger, *Phys. Rev. Lett*. 49 (1982) 1804.

results in the automatic determination of the polarization of the second photon at B(A), even if they are separated by millions of light years.

Both possibilities are sensational. The first assumes a strange form of communication between the photons or the polarizers. This communication must be propagated with a velocity *exceeding* the speed of light, because an experiment was performed in which the polarizers were switched (this took something like 10 nanoseconds) *after* the photons started (their flight took about 40 nanoseconds). Despite this, communication between the photons did exist.[66] The possibility b) as a matter of fact represents Bohr's interpretation of quantum mechanics: elementary particles do not have definite attributes (e.g., polarization).

> As a result there is dilemma: either the world is "non-real" (in the sense that the properties of particles are not determined before measurement) or/and there is instantaneous (i.e. faster than light) communication between particles which operates independently of how far apart they are ("non-locality").

This dilemma may make everybody's metaphysical shiver!

1.9 TELEPORTATION

The idea of teleportation comes from science fiction and means:

- *acquisition of full information* about an object located at A,
- *its transmission* to B,
- *creation* (materialization) of an identical object at B
- and *at the same time, the disappearance* of the object at A.

At first sight *it seems* that this contradicts quantum mechanics. The Heisenberg uncertainty principle says that it is not possible to prepare a perfect copy of the object, because, in case of mechanical quantities with non-commuting operators (like positions and momenta), there is *no way to have them measured exactly*, in order to rebuild the system elsewhere with the same values of the quantities.

The trick is, however, that the quantum teleportation we are going to describe, *will not violate* the Heisenberg principle, because the mechanical quantities needed *will not be measured and the copy made based on their values*.

The teleportation protocol was proposed by Bennett and coworkers,[67] and applied by the Anton Zeilinger group.[68] The latter used the entangled states (EPR effect) of two photons described above.[69] teleportation

---

[66] This again is the problem of delayed choice. It seems that when starting the photons have a knowledge of the *future* setting of the aparatus (the two polarizers)!

[67] C.H. Benneth, G. Brassard, C. Crépeau, R. Josza, A. Peres, W.K. Wootters, *Phys. Rev. Letters* 70 (1993) 1895.

[68] D. Bouwmeester, J. Pan, K. Mattle, M. Eibl, H. Weinfurter, A. Zeilinger, *Nature* 390 (1997) 575.

[69] A UV laser beam hits a barium borate crystal (known for its birefringence). Photons with parallel polarization move along the surface of a cone (with the origin at the beam-surface collision point),

Assume that photon A (number 1) from the entangled state belongs to Alice, and photon B (number 2) to Bob. Alice and Bob introduce a common fixed coordinate system. Both photons have *identical* polarizations in this coordinate system, *but neither Alice nor Bob know which*. Alice may measure the polarization of her photon and send this information to Bob, who may prepare his photon in that state. This, however, does not amount to teleportation, because the original state could be a linear combination of the $|0\rangle$ (parallel) and $|1\rangle$ (perpendicular) states, and in such a case Alice's measurement would "falsify" the state due to wave function collapse (it would give either $|0\rangle$ or $|1\rangle$), cf. p. 23.

Since Alice and Bob have two entangled photons of the same polarization, then let us assume that the state of the two photons is the following superposition:[70] $|00\rangle + |11\rangle$, where the first position in every ket pertains to Alice's photon, the second to Bob's.

qubit

Now, Alice wants to carry out teleportation of her additional photon (number 3) in an unknown quantum state $\phi = a|0\rangle + b|1\rangle$ (known as *qubit*), where a and b stand for *unknown* coefficients[71] satisfying the normalization condition $a^2 + b^2 = 1$. Therefore, the state of three photons (Alice's: the first and the third position in the three-photon ket, Bob's: the second position) will be $[|00\rangle + |11\rangle][a|0\rangle + b|1\rangle] = a|000\rangle + b|001\rangle + a|110\rangle + b|111\rangle$.

Alice prepares herself for teleportation of the qubit ϕ corresponding to her second photon. She first prepares a device called the *XOR gate*.[72]

XOR gate

What is the XOR gate? The device manipulates two photons, one is treated as the steering photon, the second as the steered photon. The device operates thus: if the steering photon is in state $|0\rangle$, then no change is introduced for the state of the steered photon. If, however, the steering photon is in the state $|1\rangle$, the steered photon will be switched over, i.e. it will be changed from $|0\rangle$ to $|1\rangle$ or from $|1\rangle$ to $|0\rangle$. Alice chooses the photon in the state ϕ as her steering photon, and photon 1 as her steered photon.

After the XOR gate is applied, the state of the three photons will be as follows: $a|000\rangle + b|101\rangle + a|110\rangle + b|011\rangle$.

Hadamard gate

Alice continues her preparation by using another device called the *Hadamard gate* that operates on a single photon and does the following

$$|0\rangle \rightarrow \frac{1}{\sqrt{2}}\big(|0\rangle + |1\rangle\big),$$

$$|1\rangle \rightarrow \frac{1}{\sqrt{2}}\big(|0\rangle - |1\rangle\big).$$

---

the photons with perpendicular polarization move on another cone, the two cones intersecting. From time to time a single UV photon splits into two equal energy photons of different polarizations. Two such photons when running along the intersection lines of the two cones, *and therefore not having a definite polarization* (i.e. being in a superposition state composed of both polarizations) represent the two entangled photons.

[70] The teleportation result does not depend on the state.

[71] Neither Alice nor Bob will know these coefficients up to the end of the teleportation procedure, but still Alice will be able to send her qubit to Bob!

[72] Abbreviation of "*eXclusive OR*".

Alice applies this operation to her photon 3, and after that the three-photon state is changed to the following

$$\begin{aligned}
&\frac{1}{\sqrt{2}}[a|000\rangle + a|001\rangle + b|100\rangle - b|101\rangle + a|110\rangle + a|111\rangle + b|010\rangle - b|011\rangle]\\
&\quad= \frac{1}{\sqrt{2}}\big[\big|0\,(a|0\rangle + b|1\rangle)\,0\big\rangle + \big|0\,(a|0\rangle - b|1\rangle)\,1\big\rangle - \big|1\,(a|1\rangle + b|0\rangle)\,0\big\rangle\\
&\qquad+ \big|1\,(a|1\rangle - b|0\rangle)\,1\big\rangle\big]. \qquad (1.25)
\end{aligned}$$

There is a superposition of four three-photon states in the last row. Each state shows the state of Bob's photon (number 2 in the ket), at any given state of Alice's two photons. Finally, Alice carries out the measurement of the polarization states of her photons (1 and 3). This inevitably causes her to get (for each of the photons) either $|0\rangle$ or $|1\rangle$ (collapse). This causes her to know the state of Bob's photon from the three-photon superposition (1.25):

- Alice's photons 00, i.e. Bob has his photon in state $(a|0\rangle + b|1\rangle) = \phi$,
- Alice's photons 01, i.e. Bob has his photon in state $(a|0\rangle - b|1\rangle)$,
- Alice's photons 10, i.e. Bob has his photon in state $(a|1\rangle + b|0\rangle)$,
- Alice's photons 11, i.e. Bob has his photon in state $(a|1\rangle - b|0\rangle)$.

Then Alice calls Bob and tells him the result of her measurements of the polarization of her two photons. Bob has derived (1.25) as we did.

Bob knows therefore, that if Alice tells him 00 this means that the teleportation is over: he already has his photon in state ϕ! If Alice sends him one of the remaining possibilities, he would know exactly what to do with his photon to prepare it in state ϕ and he does this with his equipment. The *teleportation is over: Bob has the teleported state ϕ, Alice has lost her photon state ϕ when performing her measurement (wave function collapse).*

Note that to carry out the successful teleportation of a photon state Alice had to communicate something to Bob.

1.10 QUANTUM COMPUTING

Richard Feynman pointed out that contemporary computers are based on the "all" or "nothing" philosophy (two bits: $|0\rangle$ *or* $|1\rangle$), while in quantum mechanics one may also use a linear combination (superposition) of these two states with arbitrary coefficients a and b: $a|0\rangle + b|1\rangle$, a qubit. Would a quantum computer based on such superpositions be better than traditional one? The hope associated with quantum computers relies on a multitude of quantum states (those obtained using variable coefficients $a, b, c, \ldots$) and possibility of working with many of them using a single processor. It was (theoretically) proved in 1994 that quantum computers could factorize natural numbers much faster than traditional computers. This sparked intensive research on the concept of quantum computation, which uses the idea of entangled states. According to many researchers, any entangled state (a superposition) is extremely sensitive to the slightest interaction with the environment, and as a result decoherence takes place very easily, which is devastating for quantum computing.[73] First attempts at constructing quantum computers were based on protecting the entangled states, but, after a few simple operations, decoherence took place.

In 1997 Neil Gershenfeld and Isaac Chuang realized that any routine nuclear magnetic resonance (NMR) measurement represents nothing but a simple quantum computation. The breakthrough was recognizing that a qubit may be also represented by the huge number of molecules in a liquid.[74] The nuclear spin angular momentum (say, corresponding to $s = \frac{1}{2}$) is associated with a magnetic dipole moment and those magnetic dipole moments interact with an external magnetic field and with themselves (Chapter 12). An isolated magnetic dipole moment has two states in a magnetic field: a lower energy state corresponding to the antiparallel configuration (state $|0\rangle$) and of higher energy state related to the parallel configuration (state $|1\rangle$). By exposing a sample to a carefully tailored nanosecond radiowave impulse one obtains a rotation of the nuclear magnetic dipoles, which corresponds to their state being a superposition $a|0\rangle + b|1\rangle$.

Here is a prototype of the XOR gate. Take chloroform[75] [$^{13}CHCl_3$]. Due to the interaction of the magnetic dipoles of the proton and of the carbon nucleus (both either in parallel or antiparallel configurations with respect to the external magnetic field) a radiowave impulse of a certain frequency causes the carbon nuclear spin magnetic dipole to rotate by 180° *provided* the proton spin dipole moment is parallel to that of the carbon. Similarly, one may conceive other logical gates. The spins changes their orientations according to a sequence of impulses, which play the role of a computer program. There are many technical problems to overcome in "liquid quantum computers": the magnetic interaction of distant nuclei is very weak, decoherence remains a worry and for the time being, limits the number of operations to several hundred. However, this is only the beginning of a new computer technology. It is most important that chemists know the future computers well – they are simply molecules.

---

[73] It pertains to an entangled state of (already) distant particles. When the particles interact strongly the state is more stable. The wave function for H_2 also represents an entangled state of two electrons, yet the decoherence does not take place even at short internuclear distances. As we will see, entangled states can also be obtained in liquids.

[74] Interaction of the molecules with the environment does not necessarily result in decoherence.

[75] The NMR operations on spins pertain in practise to a tiny fraction of the nuclei of the sample (of the order of 1 : 1000000).

Summary

Classical mechanics was unable to explain certain phenomena: black body radiation, the photoelectric effect, the stability of atoms and molecules as well as their spectra. Quantum mechanics, created mainly by Werner Heisenberg and Erwin Schrödinger, explained these effects. The new mechanics was based on six postulates:

- Postulate I says that all information about the system follows from the wave function ψ. The quantity $|\psi|^2$ represents the probability density of finding particular values of the coordinates of the particles, the system is composed of.
- Postulate II allows *operators* to be ascribed to mechanical quantities (e.g., energy). One obtains the operators by writing down the classical expression for the corresponding quantity, and replacing momenta (e.g., p_x) by momenta operators (here, $\hat{p}_x = -i\hbar\frac{\partial}{\partial x}$).
- Postulate III gives the time evolution equation for the wave function ψ (time-dependent Schrödinger equation $\hat{H}\psi = i\hbar\frac{\partial\psi}{\partial t}$), using the energy operator (*Hamiltonian* $\hat{H}$).
- Postulate IV pertains to ideal measurements. When making a measurement of a quantity A, one can obtain only *an eigenvalue of the corresponding operator* $\hat{A}$. If the wave function ψ represents an eigenfunction of $\hat{A}$, i.e. $(\hat{A}\psi = a\psi)$, then one obtains always as a result of the measurement the eigenvalue corresponding to ψ (i.e., a). If, however, the system is described by a wave function, which *does not represent any eigenfunction of* $\hat{A}$, then one obtains also an eigenvalue of $\hat{A}$, but there is no way to predict which eigenvalue. The only thing one can predict is the mean value of many measurements, which may be computed as $\langle\psi|\hat{A}\psi\rangle$ (for the normalized function ψ).
- Postulate V says that an elementary particle has an internal angular momentum (spin). One can measure only two quantities: the square of the spin length $s(s+1)\hbar^2$ and one of its components $m_s\hbar$, where $m_s = -s, -s+1, \ldots, +s$, with spin quantum number $s \geqslant 0$ characteristic for the type of particle (integer for bosons, half-integer for fermions). The spin magnetic quantum number m_s takes $2s+1$ values.
- Postulate VI has to do with symmetry of the wave function with respect to the different labelling of identical particles. If one exchanges the labels of two identical particles (we sometimes call it the exchange of all the coordinates of the two particles), then for two identical fermions the wave function has to change its sign (antisymmetric), while for two identical bosons the function does not change (symmetry). As a consequence, two identical fermions with the same spin coordinate cannot occupy the same point in space.

Quantum mechanics is one of the most peculiar theories. It gives numerical results that agree extremely well with experiments, but on the other hand the relation of these results to our everyday experience sometimes seems shocking. For example, it turned out that a particle or even a molecule may somehow exist in two locations (they pass through two slits simultaneously), but when one checks that out they are always in one place. It also turned out that

- either a particle has no definite properties ("*the world is unreal*"), and the measurement fixes them somehow
- or/and, there is instantaneous communication between particles however distant they are from each other ("*non-locality of interactions*").

It turned out that in the Bohr–Einstein controversy Bohr was right. The Einstein–Podolsky–Rosen paradox resulted (in agreement with Bohr's view) in the concept of entangled states. These states have been used experimentally to teleport a photon state without violating the Heisenberg uncertainty principle. Also the entangled states stand behind the idea of quantum computing: with a superposition of two states (qubit) $a|0\rangle + b|1\rangle$ instead of $|0\rangle$ and $|1\rangle$ as information states.

Main concepts, new terms

wave function (p. 16)
operator of a quantity (p. 18)
Dirac notation (p. 19)
time evolution equation (p. 20)
eigenfunction (p. 21)
eigenvalue problem (p. 21)
stationary state (p. 22)
measurement (p. 22)
mean value of an operator (p. 24)
spin angular momentum (p. 25)
spin coordinate (p. 26)
Pauli matrices (p. 28)
symmetry of wave function (p. 33)
antisymmetric function (p. 33)
symmetric function (p. 33)
Heisenberg uncertainty principle (p. 36)
Gedankenexperiment (p. 38)
EPR effect (p. 38)
entangled states (p. 39)
delayed choice (p. 42)
interference of particles (p. 42)
bilocation (p. 42)
Bell inequality (p. 43)
experiment of Aspect (p. 46)
teleportation (p. 47)
logical gate (p. 47)
qubit (p. 48)
XOR and Hadamard gates (p. 48)

From the research front

Until recently, the puzzling foundations of quantum mechanics could not be verified directly by experiment. As a result of enormous technological advances in quantum electronics and quantum optics it became possible to carry out experiments on single atoms, molecules, photons, etc. It was possible to carry out teleportation of a photon state across the Danube River. Even molecules such as fullerene were subjected to successful interference experiments. Quantum computer science is just beginning to prove that its principles are correct.

Ad futurum

Quantum mechanics has been proved in the past to give excellent results, but its foundations are still unclear.[76] There is no successful theory of decoherence, that would explain why and how a delocalized state becomes localized after the measurement. It is possible to make fullerene interfere, and it may be that in the near future we will be able to do this with a virus.[77] It is interesting that fullerene passes instantaneously through two slits with its whole complex electronic structure as well as nuclear framework, although the de Broglie wave length is quite different for the electrons and for the nuclei. Visibly the "overweighted" electrons interfere differently from free ones. After the fullerene passes the slits, one sees it in a single spot on the screen (decoherence). It seems that there are cases when even strong interaction does not make decoherence necessary. Sławomir Szymański presented his theoretical and experimental results[78] and showed that the functional group –CD_3 exhibits a delocalized state (which corresponds to its rotation instantaneously in both directions, a coherence) and, which makes the thing more peculiar, interaction with the environment *not only does not destroy the coherence, but makes it more robust*. This type of phenomenon might fuel investigations towards future quantum computer architectures.

---

[76] A pragmatic viewpoint is shared by the vast majority: "*do not wiseacre, just compute!*"

[77] As announced by Anton Zeilinger.

[78] S. Szymański, *J. Chem. Phys*. 111 (1999) 288.

Additional literature

"The Ghost in the atom: a discussion of the mysteries of quantum physics", P.C.W. Davies and J.R. Brown, eds, Cambridge University Press, 1986.

Two BBC journalists interviewed eight outstanding physicists: Alain Aspect (photon experiments), John Bell (Bell inequalities), John Wheeler (Feynman's PhD supervisor), Rudolf Peierls ("Peierls metal-semiconductor transition"), John Taylor ("black holes"), David Bohm ("hidden parameters") and Basil Hiley ("mathematical foundations of quantum physics"). It is most striking that all these physicists give *very* different theoretical interpretations of quantum mechanics (summarized in Chapter I).

R. Feynman, "QED – the Strange Theory of Light and Matter", Princeton University Press, Princeton (1985).

Excellent popular presentation of quantum electrodynamics written by one of the outstanding physicists of the 20th century.

A. Zeilinger, "Quantum teleportation", *Scientific American* 282 (2000) 50.

The leader in teleportation describes this new domain.

N. Gershenfeld, I.L. Chuang, "Quantum computing with molecules", *Scientific American* 278 (1998) 66.

First-hand information about NMR computing.

Ch.H. Bennett, "Quantum Information and Computation", *Physics Today* 48 (1995) 24.

Another first-hand description.

Questions

1. The state of the system is described by the wave function ψ. If $|\psi|^2$ is computed by inserting some particular values of the coordinates, then one obtains:
   a) the probability of finding the system with these coordinates; b) a complex number; c) 1; d) the probability density of finding the system with these coordinates.
2. The kinetic energy operator (one dimension) is equal to:
   a) $\frac{mv^2}{2}$; b) $-i\hbar\frac{\partial}{\partial x}$; c) $-\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2}$; d) $\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2}$.
3. The length of the electron spin vector is equal to:
   a) $\sqrt{\frac{3}{4}}\hbar$; b) $\frac{1}{2}\hbar$; c) $\pm\frac{1}{2}\hbar$; d) $\hbar$.
4. The probability density of finding two identical fermions in a single point and with the same spin coordinate is:
   a) > 0; b) 0; c) 1; d) 1/2.
5. The measurement error ΔA of quantity A, which the Heisenberg uncertainty principle speaks about is equal to:
   a) $\Delta A = \sqrt{\langle\psi|\hat{A}\psi\rangle}$; b) $\Delta A = \sqrt{\langle\psi|\hat{A}^2\psi\rangle - \langle\psi|\hat{A}\psi\rangle^2}$;
   c) $\Delta A = \langle\psi|(\hat{A}^2 - \hat{A})\psi\rangle$; d) $\Delta A = \langle\psi|\hat{A}\psi\rangle$.
6. The Heisenberg uncertainty principle $\Delta A \cdot \Delta B \geqslant \frac{\hbar}{2}$ pertains to:
   a) any two mechanical quantities A and B; b) such mechanical quantities A and B, for which $\hat{A}\hat{B} = \hat{B}\hat{A}$; c) such mechanical quantities A and B, for which the operators do not commute; d) only to a coordinate and the corresponding momentum.

7. The product of the measurement errors for a coordinate and the corresponding momentum is:
   a) $\geqslant \frac{\hbar}{2}$; b) $> \frac{\hbar}{2}$; c) $= \frac{\hbar}{2}$; d) $> \hbar$.

8. The Einstein–Podolsky–Rosen experiment aimed at falsifying the Heisenberg uncertainty principle:
   a) measuring the coordinate of the first particle and the momentum of the second particle; b) measuring exactly the coordinates of two particles; c) measuring exactly the momenta of two particles; d) by exact measuring whatever one chooses: either the coordinate or the momentum of a particle (one of two particles).

9. Entangled states mean:
   a) the real and imaginary parts of the wave function; b) a single state of two separated particles causing dependence of the results of measurements carried out on both particles; c) a product of the wave function for the first and for the second particle; d) wave functions with a very large number of nodes.

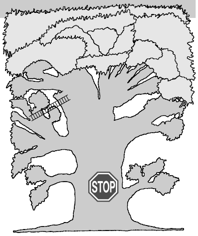
10. The experiment of Aspect has shown that:
    a) the world is local; b) the photon polarizations are definite before measurement; c) the world is non-local or/and the photon polarizations are indefinite before measurement; d) the Bell inequality is satisfied for photons.

Answers

1d, 2c, 3a, 4b, 5b, 6c, 7a, 8d, 9b, 10c

Chapter 2

THE SCHRÖDINGER EQUATION



WHERE ARE WE?

Where are we?

The postulates constitute the foundation of quantum mechanics (the base of the TREE trunk). One of their consequences is the Schrödinger equation for stationary states. Thus we begin our itinerary on the TREE. The second part of this chapter is devoted to the time-dependent Schrödinger equation, which, from the pragmatic point of view, is outside the main theme of this book (this is why it is a side branch on the left side of the TREE).

An example

A friend asked us to predict what the UV spectrum of antracene[1] looks like. One can predict any UV spectrum if one knows the electronic stationary states of the molecule. The *only* way to obtain such states and their energies is to solve the time-independent Schrödinger equation. Thus, one has to solve the equation for the Hamiltonian for antracene, then find the ground (the lowest) and the excited stationary states. The energy differences of these states will tell us where (in the energy scale), to expect light absorption, and finally then the wave functions will enable us to compute the intensity of this absorption.

What is it all about

---

[1] Three condensed benzene rings.

The *time-independent* Schrödinger equation is the one place where stationary states can be produced as solutions of the equation. The *time-dependent* Schrödinger equation plays a role as the equation of motion, describing the evolution of a given wave function as time passes. As always for an equation of motion, one has to provide an initial state (starting point), i.e. the wave function for $t = 0$. Both the stationary states, and the evolution of the non-stationary states, depend on the energy operator (Hamiltonian). If one finds some symmetry of the Hamiltonian, this will influence the symmetry of the wave functions. At the end of this chapter we will be interested in the evolution of a wave function after applying a perturbation.

Why is this important?

The wave function is a central notion in quantum mechanics, and is obtained as a solution of the Schrödinger equation. Hence this chapter is necessary for understanding quantum chemistry.

What is needed?

- Postulates of quantum mechanics, Chapter 1 (necessary).
- Matrix algebra, Appendix A, p. 889 (advised).
- Centre-of-mass separation, Appendix I, p. 971 (necessary).
- Translation vs momentum and rotation vs angular momentum, Appendix F, p. 955 (necessary).
- Dirac notation, p. 19 (necessary).
- Two-state model, Appendix D, p. 948 (necessary).
- Dirac delta, Appendix E, p. 951 (necessary).

Classical works

A paper by the mathematician Emmy Noether "*Invariante Variationsprobleme*" published in *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen*, 1918, pp. 235–257 was the first to follow the conservation laws of certain physical quantities with the symmetry of theoretical descriptions of the system. ★ Four papers by Erwin Schrödinger, which turned out to cause an "earth-quake" in science: *Annalen der Physik*, 79 (1926) 361, *ibid*. 79 (1926) 489, *ibid.* 80 (1926) 437, *ibid.* 81 (1926) 109, all under the title "*Quantisierung als Eigenwertproblem*" presented quantum mechanics as an eigenvalue problem (known from the developed differential equation theory), instead of an abstract Heisenberg algebra. Schrödinger proved the equivalence of both theories, gave the solution for the hydrogen atom, and introduced the variational principle. ★ The time-dependent perturbation theory described in this chapter was developed by Paul Adrien Maurice Dirac in 1926. Twenty years later, Enrico Fermi, lecturing at the University of Chicago coined the term "The Golden Rule" for these results. From then on, they are known as the Fermi Golden Rule.

2.1 SYMMETRY OF THE HAMILTONIAN AND ITS CONSEQUENCES

2.1.1 THE NON-RELATIVISTIC HAMILTONIAN AND CONSERVATION LAWS

From classical mechanics it follows that for an isolated system (and assuming the forces to be central and obeying the action-reaction principle), its *energy*, *momentum* and *angular momentum* are conserved.

Imagine a well isolated space ship observed in an inertial coordinate system. Its *energy is preserved*, its centre of mass moves along a straight line with constant velocity (*the total, or centre-of-mass, momentum vector is preserved*), it rotates about an axis with an angular velocity (*total angular momentum preserved*[2]). The same is true for a molecule or atom, but the conservation laws have to be formulated in the language of quantum mechanics.

Where did the conservation laws come from? Emmy Noether proved that they are related to the symmetry operations, with respect to which the equation of motion is invariant.[3]

Emmy Noether (1882–1935), German mathematician, informally professor, formally only the assistant of David Hilbert at the University of Göttingen (in the first quarter of the twentieth century women were not allowed to be professors in Germany). Her outstanding achievements in mathematics meant nothing to the Nazis, because Noether was Jewish (people should reminded of such problems) and in 1933 Noether has been forced to emigrate to the USA (Institute for Advanced Study in Princeton).

---

[2] I.e. its *length and direction*. Think of a skater performing a spin: extending the arms sideways slows down her rotation, while stretching them along the axis of rotation results in faster rotation. *But all the time the total angular momentum vector is the same*. If the space ship captain wanted to stop the rotation of the ship which is making the crew sick, he could either throw something (e.g., gas from a steering jet) away from the ship, or spin a well oriented body, fast, inside the ship. But even the captain is unable to change the *total* angular momentum.

[3] In case of a one-parameter family of operations $\hat{S}_\alpha \hat{S}_\beta = \hat{S}_{\alpha+\beta}$, e.g., translation ($\alpha$, β stand for the translation vectors), rotation (α, β are rotational angles), etc. Some other operations may not form such

Thus, it turned out that invariance of the equation of motion with respect to an arbitrary:

- *translation in time* (time homogeneity) results in the *energy conservation principle*
- *translation in space* (space homogeneity) gives the *total momentum conservation principle*
- *rotation in space* (space isotropy) implies the *total angular momentum conservation principle*.

These may be regarded as the foundations of science. The homogeneity of time allows one to expect that repeating experiments give the same results. The homogeneity of space makes it possible to compare the results of the same experiments carried out in two different laboratories. Finally, the isotropy of space allows one to reject any suspicion that a different orientation of our laboratory bench with respect to distant stars changes the result.

Now, let us try to incorporate this into quantum mechanics.

All symmetry operations (e.g. translation, rotation, reflection in a plane) are isometric, i.e. $\hat{U}^\dagger = \hat{U}^{-1}$ and $\hat{U}$ does not change distances between points of the transformed object (Figs. 2.1 and 2.2).

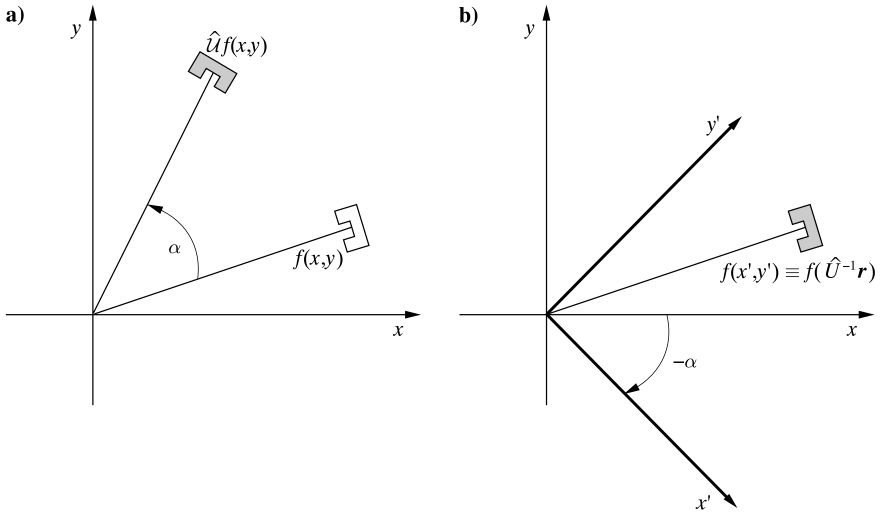


Fig. 2.1. (a) An *object* is rotated by angle α. (b) The *coordinate system* is rotated by angle $-\alpha$. The new position of the object in the old coordinate system (a) is the same as the initial position of the object in the new coordinate system (b).

---

families and then the Noether theorem is no longer valid. This was an important discovery. Symmetry of a theory is much more fundamental than the symmetry of an object. *The symmetry of a theory means that phenomena are described by the same equations no matter what laboratory coordinate system is chosen.*

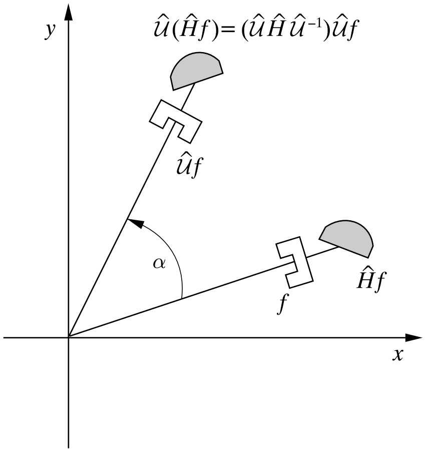


Fig. 2.2. The f and $\hat{H}f$ represent, in general, different functions. Rotation (by α) of function $\hat{H}f$ gives function $\hat{\mathcal{U}}(\hat{H}f)$ and, in consequence, is bound to denote the *rotation of f (i.e. $\hat{U}f$) and the transformation $\hat{U}\hat{H}\hat{U}^{-1}$ of the operator $\hat{H}$*. Indeed, only then does $\hat{\mathcal{U}}\hat{H}\hat{\mathcal{U}}^{-1}$ acting on the *rotated* function, i.e. $\hat{\mathcal{U}}f$ give $\hat{\mathcal{U}}\hat{H}\hat{\mathcal{U}}^{-1}(\hat{\mathcal{U}}f) = \hat{\mathcal{U}}(\hat{H}f)$, i.e. the rotation of the result. Because of $\hat{\mathcal{U}}(\hat{H}f) = (\hat{\mathcal{U}}\hat{H})(\hat{\mathcal{U}}f)$, when verifying the invariance of $\hat{H}$ with respect to transformation $\hat{\mathcal{U}}$, it is sufficient to check whether $\hat{\mathcal{U}}\hat{H}$ has the same formula as $\hat{H}$, but expressed in the new coordinates. Only this $\hat{\mathcal{U}}\hat{H}$ will fit to f expressed in the new coordinates, i.e. to $\hat{\mathcal{U}}f$. This is how we will proceed shortly.

The operator $\hat{U}$ acting in 3D Cartesian space corresponds to the operator $\hat{\mathcal{U}}$ acting in the Hilbert space, cf. eq. (C.2), p. 905. Thus the function $f(\boldsymbol{r})$ transforms to $f' = \hat{\mathcal{U}}f = f(\hat{U}^{-1}\boldsymbol{r})$, while the operator $\hat{A}$ transforms to $\hat{A}' = \hat{\mathcal{U}}\hat{A}\hat{\mathcal{U}}^{-1}$ (Fig. 2.2). The formula for $\hat{A}'$ *differs* in general from $\hat{A}$, but when *it does not*, i.e. $\hat{A}' = \hat{A}$, then $\hat{\mathcal{U}}$ commutes with $\hat{A}$.

Indeed, then $\hat{A} = \hat{\mathcal{U}}\hat{A}\hat{\mathcal{U}}^{-1}$, i.e. one has the commutation relation $\hat{A}\hat{\mathcal{U}} = \hat{\mathcal{U}}\hat{A}$, which means that $\hat{\mathcal{U}}$ and $\hat{A}$ share their eigenfunctions (Appendix B, p. 895).

Let us take the Hamiltonian $\hat{H}$ as the operator $\hat{A}$. Before writing it down let us introduce *atomic units*. Their justification comes from something similar to laziness. The quantities one calculates in quantum mechanics are stuffed up by some constants: $\hbar = \frac{h}{2\pi}$, where h is the Planck constant, electron charge $-e$, its (rest) mass m_0, etc. These constants appear in clumsy formulae with various powers, in the nominator and denominator (see Table of units, p. 1062). We always know, however, that the quantity we calculate is energy, length, time or something similar and we know how the unit energy, the unit length, etc. is expressed by $\hbar$, e, m_0. atomic units

ATOMIC UNITS
If one inserts: $\hbar = 1$, $e = 1$, $m_0 = 1$, this gives a dramatic simplification of the formulae. One has to remember though, that these units have been introduced and, whenever needed, one can evaluate the result in other units (see Table of conversion coefficients, p. 1063).

The Hamiltonian for a system of M nuclei (with charges Z_I and masses m_I, non-relativistic Hamiltonian

$I = 1, \ldots, M$) and N electrons, in the non-relativistic approximation and assuming point-like particles without any internal structure,[4] takes [in atomic units (a.u.)] the following form (see p. 18)

$$\hat{H} = \hat{T}_n + \hat{T}_e + \hat{V}, \tag{2.1}$$

where the kinetic energy operators for the nuclei and electrons (in a.u.) read as:

$$\hat{T}_n = -\frac{1}{2}\sum_{I=1}^{M} \frac{1}{m_I}\Delta_I, \tag{2.2}$$

$$\hat{T}_e = -\frac{1}{2}\sum_{i=1}^{N} \Delta_i, \tag{2.3}$$

where the Laplacians are

$$\Delta_I = \frac{\partial^2}{\partial X_I^2} + \frac{\partial^2}{\partial Y_I^2} + \frac{\partial^2}{\partial Z_I^2},$$

$$\Delta_i = \frac{\partial^2}{\partial x_i^2} + \frac{\partial^2}{\partial y_i^2} + \frac{\partial^2}{\partial z_i^2},$$

---

[4] No internal structure of the electron has yet been discovered. The electron is treated as a point-like particle. Contrary to this nuclei have a rich internal structure and non-zero dimensions. A clear multi-level-like structure appears (which has to a large extent forced a similar structure on the corresponding scientific methodologies):

- Level I. A nucleon (neutron, proton) consists of three (the valence) quarks, clearly seen on the scattering image obtained for the proton. Nobody has yet observed a free quark.
- Level II. The strong forces acting among nucleons have a range of about 1–2 fm (1 fm $= 10^{-15}$ m). Above 0.4–0.5 fm they are attractive, at shorter distances they correspond to repulsion. One need not consider their quark structure when computing the forces among nucleons, but they may be treated as particles *without internal structure*. The attractive forces between nucleons practically do not depend on the nucleon's charge and are so strong that they may overcome the Coulomb repulsion of protons. Thus the nuclei composed of many nucleons (various chemical elements) may be formed, which exhibit a shell structure (analogous to electronic structure, cf. Chapter 8) related to the packing of the nucleons. The motion of the nucleons is strongly correlated. A nucleus may have various energy states (ground and excited), may be distorted, may undergo splitting, etc. About 2000 nuclei are known, of which only 270 are stable. The smallest nucleus is the proton, the largest known so far is ^{209}Bi (209 nucleons). The largest observed number of protons in a nucleus is 118. Even the largest nuclei have diameters about 100000 times smaller than the electronic shells of the atom. Even for an atom with atomic number 118, the first Bohr radius is equal to $\frac{1}{118}$ a.u. or $5 \cdot 10^{-13}$ m, still about 100 times larger than the nucleus.
- Level III. Chemists *can* neglect the internal structure of nuclei. A nucleus *can* be treated as a structureless point-like particle and using the theory described in this book, one is able to predict extremely precisely virtually all the chemical properties of atoms and molecules. Some interesting exceptions will be given in 6.11.2.

and x, y, z stand for the Cartesian coordinates of the nuclei and electrons indicated by vectors $\boldsymbol{R}_I = (X_I, Y_I, Z_I)$ and $\boldsymbol{r}_i = (x_i, y_i, z_i)$, respectively.

The operator $\hat{V}$ corresponds to the electrostatic interaction of all the particles (nucleus–nucleus, nucleus–electron, electron–electron):

$$\hat{V} = \sum_{I=1}^{M}\sum_{J>I}^{M} \frac{Z_I Z_J}{|\boldsymbol{R}_I - \boldsymbol{R}_J|} - \sum_{I=1}^{M}\sum_{i=1}^{N} \frac{Z_I}{|\boldsymbol{r}_i - \boldsymbol{R}_I|} + \sum_{i=1}^{N}\sum_{j>i}^{N} \frac{1}{|\boldsymbol{r}_i - \boldsymbol{r}_j|}, \tag{2.4}$$

or, in a simplified form

$$\hat{V} = \sum_{I=1}^{M}\sum_{J>I}^{M} \frac{Z_I Z_J}{R_{IJ}} - \sum_{I=1}^{M}\sum_{i=1}^{N} \frac{Z_I}{r_{iI}} + \sum_{i=1}^{N}\sum_{j>i}^{N} \frac{1}{r_{ij}}. \tag{2.5}$$

If the Hamiltonian turned out to be invariant with respect to a symmetry operation $\hat{U}$ (translation, rotation, etc.), this would imply the commutation of $\hat{\mathcal{U}}$ and $\hat{H}$. We will check this in more detail below.

Note that the distances R_{IJ}, r_{iI} and r_{ij} in the Coulombic potential energy in eq. (2.5) witness the assumption of instantaneous interactions in non-relativistic theory (infinite speed of travelling the interaction through space).

2.1.2 INVARIANCE WITH RESPECT TO TRANSLATION

Translation by vector $\boldsymbol{T}$ of function $f(\boldsymbol{r})$ in space means the function $\hat{\mathcal{U}} f(\boldsymbol{r}) = f(\hat{U}^{-1}\boldsymbol{r}) = f(\boldsymbol{r} - \boldsymbol{T})$, i.e. an opposite (by vector $-\boldsymbol{T}$) translation of the coordinate system (Fig. 2.3).

Transformation $\boldsymbol{r}' = \boldsymbol{r} + \boldsymbol{T}$ does not change the Hamiltonian. This is evident for the potential energy $\hat{V}$, because the translations $\boldsymbol{T}$ cancel, leaving the interparticle distances unchanged. For the kinetic energy one obtains

$$\frac{\partial}{\partial x'} = \sum_{\sigma = x,y,z} \frac{\partial \sigma}{\partial x'} \frac{\partial}{\partial \sigma} = \frac{\partial x}{\partial x'} \frac{\partial}{\partial x} = \frac{\partial}{\partial x},$$

and all the kinetic energy operators (eqs. (2.2) and (2.3)) are composed of the operators having this form.

> The Hamiltonian is therefore invariant with respect to any translation of the coordinate system.

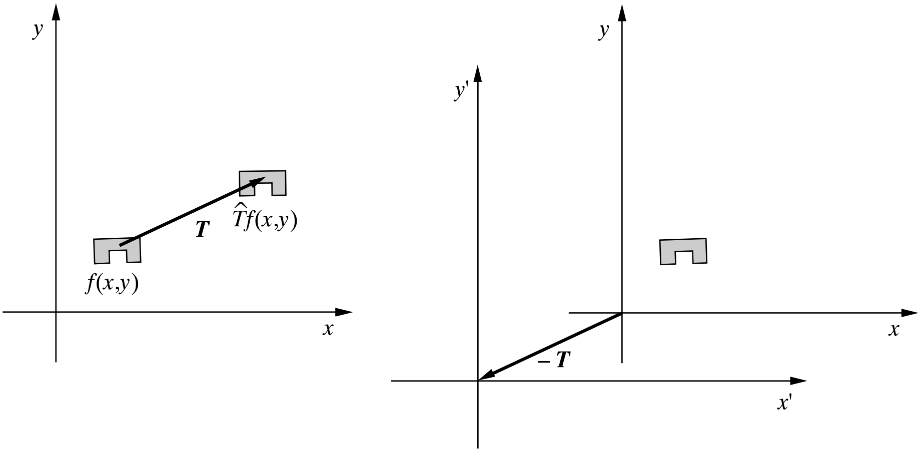


Fig. 2.3. A *function* f shifted by vector $\boldsymbol{T}$ (symmetry operation $\hat{T}$), i.e. $\hat{T}f(x,y)$ in the coordinate system (x,y) is the same as function $f(x',y')$, in the *coordinate system* (x',y') shifted by $-\boldsymbol{T}$.

space homogeneity

There are two main consequences of translational symmetry:

- No matter, whether the coordinate system used is fixed in Trafalgar Square, or in the centre of mass of the system, one has to solve the same mathematical problem.
- The solution to the Schrödinger equation corresponding to the space fixed coordinate system (SFS) located in Trafalgar Square is Ψ_{pN}, whereas Ψ_{0N} is calculated in the body-fixed coordinate system (see Appendix I) located in the centre of mass at $\boldsymbol{R}_{CM}$ with the (total) momentum $\boldsymbol{p}_{CM}$. These two solutions are related by $\Psi_{pN} = \Psi_{0N}\ \exp(i\boldsymbol{p}_{CM} \cdot \boldsymbol{R}_{CM})$. The number $N = 0, 1, 2, \ldots$ counts the energy states after the centre-of-mass motion is separated.

This means that the energy spectrum represents a continuum, because the centre of mass may have any (non-negative) kinetic energy $p_{CM}^2/(2m)$. If, however, one assumes that $p_{CM} = \text{const}$, then the energy spectrum is *discrete* for low-energy eigenvalues (see eq. (1.13)).

This spectrum corresponds to the bound states, i.e. those states which do not correspond to any kind of dissociation (including ionization). Higher energy states lead to dissociation of the molecule, and the fragments may have any kinetic energy. Therefore, above the discrete spectrum one has a continuum of states. The states Ψ_{0N} will be called *spectroscopic states.* The bound states Ψ_{0N} are square integrable, as opposed to Ψ_{pN}, which are not because of function $\exp(i\boldsymbol{p}\boldsymbol{R}_{CM})$, which describes the free motion of the centre of mass.

spectroscopic states

2.1.3 INVARIANCE WITH RESPECT TO ROTATION

isotropy of space

The Hamiltonian is also invariant with respect to any rotation in space $\hat{U}$ of the coordinate system about a fixed axis. The rotation is carried out by applying an orthogonal matrix transformation $\boldsymbol{U}$ of vector $\boldsymbol{r} = (x, y, z)^{\mathrm{T}}$ that describes any particle of coordinates x, y, z. Therefore all the particles undergo the same rotation and the new coordinates are $\boldsymbol{r}' = \hat{U}\boldsymbol{r} = \boldsymbol{U}\boldsymbol{r}$. Again there is no problem with the potential energy, because a rotation does not change the interparticle distances. What about the Laplacians in the kinetic energy operators? Let us see.

$$\begin{aligned}
\Delta &= \sum_{k=1}^{3} \frac{\partial^2}{\partial x_k^2} = \sum_{k=1}^{3} \frac{\partial}{\partial x_k}\frac{\partial}{\partial x_k} = \sum_{k=1}^{3}\left(\sum_{i=1}^{3}\frac{\partial}{\partial x_i'}\frac{\partial x_i'}{\partial x_k}\right)\left(\sum_{i=1}^{3}\frac{\partial}{\partial x_i'}\frac{\partial x_i'}{\partial x_k}\right)\\
&= \sum_{i=1}^{3}\sum_{j=1}^{3}\sum_{k=1}^{3}\left(\frac{\partial}{\partial x_i'}\frac{\partial x_i'}{\partial x_k}\right)\left(\frac{\partial}{\partial x_j'}\frac{\partial x_j'}{\partial x_k}\right)\\
&= \sum_{i=1}^{3}\sum_{j=1}^{3}\sum_{k=1}^{3}\left(\frac{\partial}{\partial x_i'}U_{ik}\right)\left(\frac{\partial}{\partial x_j'}U_{jk}\right) = \sum_{i=1}^{3}\sum_{j=1}^{3}\sum_{k=1}^{3}\left(\frac{\partial}{\partial x_i'}U_{ik}\right)\left(\frac{\partial}{\partial x_j'}U_{kj}^{\dagger}\right)\\
&= \sum_{i=1}^{3}\sum_{j=1}^{3}\left(\frac{\partial}{\partial x_i'}\right)\left(\frac{\partial}{\partial x_j'}\right)\sum_{k=1}^{3}U_{ik}U_{kj}^{\dagger}\\
&= \sum_{i=1}^{3}\sum_{j=1}^{3}\left(\frac{\partial}{\partial x_i'}\right)\left(\frac{\partial}{\partial x_j'}\right)\delta_{ij} = \sum_{k=1}^{3}\frac{\partial^2}{\partial (x_k')^2}.
\end{aligned}$$

Thus, one has invariance of the Hamiltonian with respect to any rotation about the origin of the coordinate system. This means (see p. 955) that the Hamiltonian and the operator of the square of the total angular momentum $\hat{J}^2$ (as well as of one of its components, denoted by $\hat{J}_z$) commute. One is able, therefore, to measure simultaneously the energy, the square of total angular momentum as well as one of the components of total angular momentum, and (as it will be shown in (4.6)) one has

$$\hat{J}^2\Psi_{0N}(\boldsymbol{r}, \boldsymbol{R}) = J(J+1)\hbar^2\Psi_{0N}(\boldsymbol{r}, \boldsymbol{R}), \tag{2.6}$$

$$\hat{J}_z\Psi_{0N}(\boldsymbol{r}, \boldsymbol{R}) = M_J\hbar\Psi_{0N}(\boldsymbol{r}, \boldsymbol{R}), \tag{2.7}$$

where $J = 0, 1, 2\ldots$ and $M_J = -J, -J+1, \ldots, +J$.

Any rotation matrix may be shown as a product of "elementary" rotations, each about axes x, y or z. For example, rotation about the y axis by angle θ corresponds to the matrix

$$\begin{pmatrix} \cos\theta & 0 & -\sin\theta \\ 0 & 1 & 0 \\ \sin\theta & 0 & \cos\theta \end{pmatrix}.$$

The pattern of such matrices is simple: one has to put in some places sines, cosines, zeros and ones with the proper signs.[5] This matrix is orthogonal,[6] i.e. $\boldsymbol{U}^{\mathrm{T}} = \boldsymbol{U}^{-1}$, which you may easily check. The product of two orthogonal matrices represents an orthogonal matrix, therefore any rotation corresponds to an orthogonal matrix.

2.1.4 INVARIANCE WITH RESPECT TO PERMUTATION OF IDENTICAL PARTICLES (FERMIONS AND BOSONS)

The Hamiltonian has also permutational symmetry. This means that if someone exchanged labels numbering the identical particles, independently of how it was done, they would always obtain the identical mathematical expression for the Hamiltonian. This implies that any wave function has to be symmetric (for bosons) or antisymmetric (fermions) with respect to the exchange of labels between two identical particles (cf. p. 33).

2.1.5 INVARIANCE OF THE TOTAL CHARGE

The total electric charge of a system does not change, whatever happens. In addition to the energy, momentum and angular momentum, strict conservation laws are obeyed exclusively for the total electric charge and the baryon and lepton numbers (a given particle contributes $+1$, the corresponding the antiparticle -1).[7] The charge conservation law follows from the gauge symmetry. This symmetry means the invariance of the theory with respect to partition of the total system into subsystems. Total electric charge conservation follows from the fact that the description of the system has to be invariant with respect to the mixing of the particle and antiparticle states, which is analogous to rotation.

---

[5] Clockwise and anticlockwise rotations and two possible signs at sines cause a problem with memorizing the right combination. In order to choose the correct one, one may use the following trick. First, we decide that what moves is an *object* (e.g., a function, not the coordinate system). Then, you take my book from your pocket. With Fig. 2.1.a one sees that the rotation of the point with coordinates $(1,0)$ by angle $\theta = 90°$ should give the point $(0,1)$, and this is assured only by the rotation matrix: $\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$.

[6] And therefore also unitary (cf. Appendix A, p. 889).

[7] For example, in the Hamiltonian (2.1) it is assumed that whatever might happen to our system, the numbers of the nucleons and electrons will remain constant.

2.1.6 FUNDAMENTAL AND LESS FUNDAMENTAL INVARIANCES

The conservation laws described are of a fundamental character, because they are related to the homogeneity of space and time, the isotropy of space and the non-distinguishability of identical particles.

Besides these strict conservation laws, there are also some approximate laws. Two of these: parity and charge conjugation, will be discussed below. They are rooted in these strict laws, but are valid only in some conditions. For example, in most experiments, not only the baryon number, but also the number of nuclei of each kind are conserved. Despite the importance of this law in chemical reaction equations, this does not represent any strict conservation law as shown by radioactive transmutations of elements.

Some other approximate conservation laws will soon be discussed.

2.1.7 INVARIANCE WITH RESPECT TO INVERSION – PARITY

There are orthogonal transformations which are not equivalent to any rotation, e.g., the matrix of inversion

$$\begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix},$$

which corresponds to changing $\boldsymbol{r}$ to $-\boldsymbol{r}$ for all the particles and does not represent any rotation. If one performs such a symmetry operation, the Hamiltonian remains invariant. This is evident, both for $\hat{V}$ (the interparticle distances do not change), and for the Laplacian (single differentiation changes sign, double does not). Two consecutive inversions mean an identity operation. Hence,

$$\Psi_{0N}(-\boldsymbol{r}, -\boldsymbol{R}) = \Pi\Psi_{0N}(\boldsymbol{r}, \boldsymbol{R}), \quad \text{where } \Pi \in \{1, -1\}.$$

Therefore,

> the wave function of a stationary state represents an eigenfunction of the inversion operator, and the eigenvalue can be either $\Pi = 1$ or $\Pi = -1$ (this property is called parity, or P).

Now the reader will be taken by surprise. From what we have said, it follows that *no molecule has a non-zero dipole moment*. Indeed, the dipole moment is calculated as the mean value of the dipole moment operator, i.e. $\boldsymbol{\mu} = \langle\Psi_{0N}|\hat{\boldsymbol{\mu}}\Psi_{0N}\rangle = \langle\Psi_{0N}|(\sum_i q_i\boldsymbol{r}_i)\Psi_{0N}\rangle$. This integral will be calculated very easily: the integrand is antisymmetric with respect to inversion[8] and therefore $\boldsymbol{\mu} = \boldsymbol{0}$.

---

[8] Ψ_{0N} may be symmetric or antisymmetric, but $|\Psi_{0N}|^2$ is bound to be symmetric. Therefore, since $\sum_i q_i\boldsymbol{r}_i$ is antisymmetric, then indeed, the integrand is antisymmetric (the integration limits are symmetric).

Is therefore the very meaning of the dipole moment, a quantity often used in chemistry and physics, a fairy tale? If HCl has no dipole moment, then it is more understandable that H_2 does not have either. All this seems absurd. What about this dipole moment?

Let us stress, that our conclusion pertains to the *total* wave function, which has to reflect the space isotropy leading to the zero dipole moment, because all orientations in space are equally probable. If one applied the transformation $\boldsymbol{r} \to -\boldsymbol{r}$ only to *some* particles in the molecule (e.g., electrons), and not to others (e.g., the nuclei), the wave function will show no parity (it would be neither symmetric nor antisymmetric). We will introduce the Hamiltonian in Chapter 6, which corresponds to immobilizing the nuclei (clamped nuclei approximation) in certain positions in space, and in such a case the wave function depends on the electronic coordinates only. This wave function may be neither symmetric nor antisymmetric with respect to the *partial* inversion transformation $\boldsymbol{r} \to -\boldsymbol{r}$ (for the electrons only). To give an example, let us imagine the HF molecule in a coordinate system, its origin in the middle between the H and F nuclei. Consider a particular configuration of the 10 electrons of the molecule; all close to the fluorine nucleus in some well defined points. One may compute the value of the wave function for this configuration of electrons. Its square gives us the probability density of finding this particular configuration of electrons. Now, imagine the (*partial*) inversion $\boldsymbol{r} \to -\boldsymbol{r}$ applied to all the electrons. Now they will all be close to the proton. If one computes the probability density for the new situation, one would obtain a different value (much, much smaller, because the electrons prefer the fluorine, not the hydrogen). No symmetry or antisymmetry. No wonder therefore that if one computed $\boldsymbol{\mu} = \langle \Psi_{0N} | \hat{\boldsymbol{\mu}} \Psi_{0N} \rangle$ with such a function (integration is over the electronic coordinates only), the result would differ from zero. This is why chemists believe the HF molecule has a nonzero dipole moment.[9] On the other hand, if the molecule taken as the example were B_2 (also ten electrons), then the two values have had to be equal, because they describe the same physical situation. This corresponds, therefore, to a wave function with definite parity (symmetric or antisymmetric), and therefore, in this case $\boldsymbol{\mu} = \boldsymbol{0}$. This is why chemists believe such molecules as H_2, B_2, O_2 have no dipole moment.

Product of inversion and rotation

The Hamiltonian is also invariant with respect to some other symmetry operations like changing the sign of the x coordinates of all particles, or similar operations which are products of inversion and rotation. If one changed the sign of all the x coordinates, it would correspond to a mirror reflection. Since rotational symmetry stems from space isotropy (which we will treat as "trivial"), the mirror reflection may be identified with parity P.

mirror reflection

P symmetry

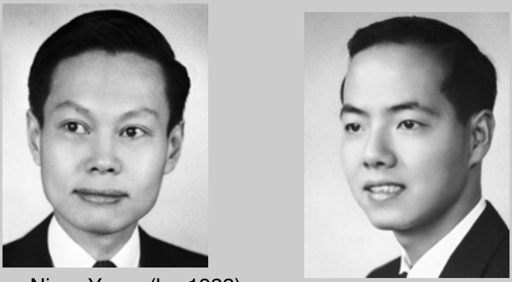
---

[9] What therefore do they measure? The answer will be given in Chapter 12.

Enantiomers

A consequence of inversion symmetry is that the wave functions have to be eigenfunctions of the inversion operator with eigenvalues $\Pi = 1$, i.e. the wave function is symmetric, or $\Pi = -1$, i.e. the wave function is antisymmetric. Any asymmetric wave function corresponding to a stationary state is therefore excluded ("illegal"). However, two optical isomers (enantiomers), corresponding to an object and its mirror image, do exist (Fig. 2.4).[10]

We ask in a pharmacy for D-glucose, strangely enough the pharmacist is fully cooperative and does not make trouble. We pay a small sum and he gives us something which should not exist[11] – a substance with a single enantiomer. We should obtain a substance composed of molecules in their stationary states, which have therefore to have definite parity, either as a sum of the wave functions for the two enantiomers D and L ($\Pi = 1$, cf. Appendix D on p. 948, Example I): $\psi_+ = \psi_D + \psi_L$ or as the difference ($\Pi = -1$): $\psi_- = \psi_D - \psi_L$. The energies corresponding to ψ_+ and ψ_- differ, but the difference is extremely small (quasidegeneracy). The brave shopkeeper has given us something with the wave function $\psi = N(\psi_+ + \psi_-) = \psi_D$ (as result of de-



Chen Ning Yang (b. 1922) and Tsung Dao Lee (b. 1926) American physicists, professors at the Advanced Study Institute in Princeton predicted in 1956 parity breaking in the weak interactions, which a few months later has been confirmed experimentally by Madam Wu. In 1957 Yang and Lee received the Nobel Prize *"for their penetrating investigation of parity laws, which led to important discoveries regarding elementary particles".*

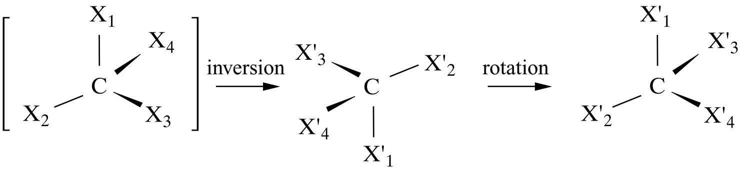


Fig. 2.4. If one superposed the X_1–C–X_2 and X_1'–C′–X_2' fragments of both molecules, the other two substituents could not match: X_4' in place of X_3 and X_3' in place of X_4. The two molecules represent two enantiomeric isomers. A wave function that describes one of the enantiomers does not have a definite parity and is therefore "illegal".

---

[10] The property that distinguishes them is known as chirality (your hands are an example of chiral objects). The chiral molecules (enantiomers) exhibit optical activity, i.e. polarized light passing through a solution of one of the enantiomers undergoes a rotation of the polarization plane *always in the same direction* (which may be easily seen by reversing the direction of the light beam). Two enantiomeric molecules have the same properties, provided one is checking this by using non-chiral objects. If the probe were chiral, one of the enantiomers would interact with it differently (for purely sterical reasons). Enantiomers (e.g., molecular associates) may be formed from chiral or non-chiral subunits.

[11] More exactly, should be unstable.

coherence), which therefore describes a non-stationary state.[12] As we will see in a moment (p. 82), the approximate lifetime of the state is proportional to the inverse of the integral $\langle\psi_D|\hat{H}\psi_L\rangle$. If one calculated this integral, one would obtain an extremely small number.[13] It would turn out that the pharmacy could safely keep the stock of glucose for millions of years. Maybe the reason for decoherence is interaction with the rest of the Universe, maybe even interaction with a vacuum. The very existence of enantiomers, or even the prevailence of one of them on Earth, does not mean breaking parity symmetry. This would happen if one of the enantiomers corresponded to a lower energy than the other.[14]

2.1.8 INVARIANCE WITH RESPECT TO CHARGE CONJUGATION

> If one changed the signs of the charges of all particles, the Hamiltonian would not change.

C symmetry

This therefore corresponds to exchanging particles and antiparticles.[15] Such a symmetry operation is called the charge conjugation and denoted as C symmetry. This symmetry will not be marked in the wave function symbol (because, as a rule, we have to do with matter, not antimatter), but we will remember. Sometimes it may turn out unexpectedly to be useful (see Chapter 13, p. 702). After Wu's experiment, physicists tried to save the hypothesis that what is conserved is the CP symmetry, i.e. the product of charge conjugation and inversion. However, analysis of experiments with the meson K decay has shown that even this symmetry is approximate (although the deviation is extremely small).

2.1.9 INVARIANCE WITH RESPECT TO THE SYMMETRY OF THE NUCLEAR FRAMEWORK

In many applications the positions of the nuclei are fixed (clamped nuclei approximation, Chapter 6), often in a high-symmetry configuration (cf. Appendix C, p. 903). For example, the benzene molecule in its ground state (after minimizing the energy with respect to the positions of the nuclei) has the symmetry of a regular hexagon. In such cases the electronic Hamiltonian additionally exhibits invariance with respect to some symmetry operations and therefore the wave functions are

---

[12] Only ψ_+ and ψ_- are stationary states.

[13] This is seen even after attempting to overlap two molecular models physically, Fig. 2.4. The overlap of the wave functions will be small for the same reasons (the wave functions decay exponentially with distance).

[14] This is what happens in reality, although the energy difference is extremely small. Experiments with β-decay have shown that Nature breaks parity in weak interactions. Parity conservation law therefore has an approximate character.

[15] Somebody thought he had carried out computations for benzene, but he also computed antibenzene. The wave function for benzene and antibenzene are the same.

the eigenstates of these symmetry operations. Therefore, any wave function may have an additional label; the symbol of the irreducible representation[16] it belongs to.

2.1.10 CONSERVATION OF TOTAL SPIN

In an isolated system the *total* angular momentum $\boldsymbol{J}$ is conserved. However, $\boldsymbol{J} = \boldsymbol{L} + \boldsymbol{S}$, where $\boldsymbol{L}$ and $\boldsymbol{S}$ stand for the orbital and spin angular momenta (sum over all particles), respectively. The spin angular momentum $\boldsymbol{S}$, a sum of spins of all particles, is not conserved.

However, the (non-relativistic) Hamiltonian does not contain any spin variables. This means that it commutes with the operator of the square of the total spin as well as with the operator of one of the spin components (by convention the z component). Therefore, in the non-relativistic approximation one can simultaneously measure the energy E, the square of the spin S^2 and one of its components S_z.

2.1.11 INDICES OF SPECTROSCOPIC STATES

In summary, assumptions about the homogeneity of space and time, isotropy of space and parity conservation lead to the following quantum numbers (indices) for the spectroscopic states:

- N quantizes energy,
- J quantizes the length of total angular momentum,
- M quantizes the z component of total angular momentum,
- Π determines parity:

$$\Psi_{N,J,M,\Pi}(\boldsymbol{r}, \boldsymbol{R}).$$

Besides these indices following from the fundamental laws (in the case of parity it is a little too exaggerated), there may be also some indices related to less fundamental conservation laws:

- the irreducible representation index of the symmetry group of the clamped nuclei Hamiltonian (Appendix C)
- the values of S^2 (traditionally one gives the *multiplicity* $2S + 1$) and S_z.

---

[16] Of the symmetry group composed of the symmetry operations mentioned above.

2.2 SCHRÖDINGER EQUATION FOR STATIONARY STATES

It may be instructive to see how Erwin Schrödinger invented his famous equation (1.13) for *stationary states* ψ of energy E ($\hat{H}$ denotes the Hamiltonian of the system)

$$\hat{H}\psi = E\psi. \tag{2.8}$$

Schrödinger surprised the contemporary quantum elite (associated mainly with

Erwin Rudolf Josef Alexander Schrödinger (1887–1961), Austrian physicist, professor at the universities of Jena, Stuttgart, Graz, Breslau, Zurich, Berlin and Vienna. In later years Schrödinger recalled the Zurich period most warmly, in particular, discussions with the mathematician Hermann Weyl and physicist Peter Debye. In 1927 Schrödinger succeeded Max Planck at the University of Berlin, and in 1933 received the Nobel Prize "*for the discovery of new productive forms of atomic theory*". Hating the Nazi regime, he left Germany in 1933 and moved to the University of Oxford. However, homesick for his native Austria he went back in 1936 and took a professorship at the University of Graz. Meanwhile Hitler carried out his Anschluss with Austria in 1938, and Schrödinger even though not a Jew, could have been an easy target as one who fled Germany because of the Nazis. He emigrated to the USA (Princeton), and then to Ireland (Institute for Advanced Studies in Dublin), worked there till 1956, then returned to Austria and remained there, at the Vienna University, until his death.

In his scientific work as well as in his personal life Schrödinger did not strive for big goals, he worked by himself. Maybe what characterizes him best is that he was always ready to leave having belongings packed in his rucksack. Among the goals of this textbook listed in the Introduction there is not demoralization of youth. This is why I will stop here, limit myself to the carefully selected information given above and refrain from describing the circumstances, in which quantum mechanics was born. For those students who read the material recommended in the Additional Literature, I provide some useful references: W. Moore, "*Schrödinger: Life and Thought*", Cambridge University Press, 1989, and the comments on the book given by P.W. Atkins, *Nature*, 341 (1989), also http://www-history.mcs.st-andrews.ac.uk/history/Mathematicians/Schrodinger.html.

Schrödinger's curriculum vitae found in Breslau (now Wrocław):

"*Erwin Schrödinger, born on Aug., 12, 1887 in Vienna, the son of the merchant Rudolf Schrödinger and his wife née Lauer. The family of my father comes from the Upper Palatinate and Wirtemberg region, and the family of my mother from German Hungary and (from the maternal side) from England. I attended a so called "academic" high school (once part of the university) in my native town. Then from 1906–1910 I studied physics at Vienna University, where I graduated in 1910 as a doctor of physics. I owe my main inspiration to my respected teacher Fritz Hasenöhrl, who by an unlucky fate was torn from his diligent students – he fell gloriously as an attack commander on the battlefield of Vielgereuth. As well as Hasenöhrl, I owe my mathematical education to Professors Franz Mertens and Wilhelm Wirtinger, and a certain knowledge of experimental physics to my principal of many years (1911–1920) Professor Franz Exner and my intimate friend R.M.F. Rohrmuth. A lack of experimental and some mathematical skills oriented me basically towards theory.*

Presumably the spirit of Ludwig Boltzmann (deceased in 1906), operating especially intensively in Vienna, directed me first towards the probability theory in physics. Then, (...) a closer contact with the experimental works of Exner and Rohrmuth oriented me to the physiological theory of colours, in which I tried to confirm and develop the achievements of Helmholtz. In 1911–1920 I was a laboratory assistant under Franz Exner in Vienna, of course, with $4\frac{1}{2}$ years long pause caused by war. I have obtained my habilitation in 1914 at the *University of Vienna, while in 1920 I accepted an offer from Max Wien and became his assistant professor at the new theoretical physics department in Jena. This lasted, unfortunately, only one semester, because I could not refuse a professorship at the Technical University in Stuttgart. I was there also only one semester, because April 1921 I came to the University of Hessen in succession to Klemens Schrafer. I am almost ashamed to confess, that at the moment I sign the present curriculum vitae I am no longer a professor at the University of Breslau, because on Oct. 15. I received my nomination to the University of Zurich. My instability may be recognized exclusively as a sign of my ingratitude!*

Breslau, Oct., 5, 1921. *Dr Erwin Schrödinger*

(found in the archives of the University of Wrocław (Breslau) by Professor Zdzisław Latajka and Professor Andrzej Sokalski, translated by Professor Andrzej Kaim and the Author. Since the manuscript (see web annex, Supplements) was hardly legible due to Schrödinger's difficult handwriting, some names may have been misspelled.)

Copenhagen and Göttingen) by a clear formulation of quantum mechanics as wave mechanics. January 27, 1926, when Schrödinger submitted a paper entitled "*Quantisierung als Eigenwertproblem*"[17] to *Annalen der Physik*, may be regarded as the birthday of wave mechanics.

Most probably Schrödinger's reasoning was as follows. De Broglie discovered that what people called a particle also had a wave nature (Chapter 1). That is really puzzling. If a wave is involved, then according to Debye's suggestion at the November seminar in Zurich, it might be possible to write the standing wave equation with $\psi(x)$ as its amplitude at position x:

$$v^2\frac{\mathrm{d}^2\psi}{\mathrm{d}x^2}+\omega^2\psi=0, \tag{2.9}$$

where v stands for the (phase) velocity of the wave, and ω represents its angular frequency ($\omega=2\pi\nu$, where ν is the usual frequency) which is related to the wave length λ by the well known formula:[18]

$$\omega/v=\frac{2\pi}{\lambda}. \tag{2.10}$$

Besides, Schrödinger knew from the de Broglie's thesis, himself having lectured in Zurich about this, that the wavelength, λ, is related to a particle's momentum p through $\lambda=h/p$, where $h=2\pi\hbar$ is the Planck constant. This equation is the most

---

[17] Quantization as an eigenproblem. Well, once upon a time quantum mechanics was discussed in German. Some traces of that period remain in the nomenclature. One is the "*eigen*value problem or *eigen*problem" which is a German–English hybrid.

[18] In other words $\nu=\frac{v}{\lambda}$ or $\lambda=vT$ (i.e. wave length is equal to the velocity times the period). Eq. (2.9) represents an oscillating function $\psi(x)$. Indeed, it means that $\frac{\mathrm{d}^2\psi}{\mathrm{d}x^2}$ and ψ differ by sign, i.e. if ψ is above the x axis, then it curves down, while if it is below the x axis, then it curves up.

famous achievement of de Broglie, and relates the corpuscular (p) character and the wave (λ) character of any particle.

On the other hand the momentum p is related to the total energy (E) and the potential energy (V) of the particle through: $p^2 = 2m(E - V)$, which follows, from the expression for the kinetic energy $T = \frac{mv^2}{2} = p^2/(2m)$ and $E = T + V$. Therefore, eq. (2.9) can be rewritten as:

$$\frac{\mathrm{d}^2\psi}{\mathrm{d}x^2} + \frac{1}{\hbar^2}\bigl[2m(E - V)\bigr]\psi = 0, \tag{2.11}$$

The most important step towards the great discovery was the transfer of the term involving E to the left hand side. Let us see what Schrödinger obtained:

$$\left[-\frac{\hbar^2}{2m}\frac{\mathrm{d}^2}{\mathrm{d}x^2} + V\right]\psi = E\psi. \tag{2.12}$$

This was certainly a good moment for a discovery. Schrödinger obtained a kind of *eigenvalue equation* (1.13), recalling his experience with eigenvalue equations in the theory of liquids.[19] What is striking in eq. (2.12) is the odd fact that an operator $-\frac{\hbar^2}{2m}\frac{\mathrm{d}^2}{\mathrm{d}x^2}$ amazingly plays the role of the kinetic energy. Indeed, keeping calm we see the following: *something* plus potential energy, all that multiplied by ψ, equals total energy times ψ. Therefore, clearly this *something* must be the kinetic energy! But, wait a minute, the kinetic energy is equal to $\frac{p^2}{2m}$. From this it follows that, in the equation obtained instead of p there is a certain *operator* $i\hbar\frac{\mathrm{d}}{\mathrm{d}x}$ or $-i\hbar\frac{\mathrm{d}}{\mathrm{d}x}$, because only then does the squaring give the right answer.

Would the key to the puzzle be simply taking the classical expression for total energy and inserting the above operators instead the momenta? What was the excited Schrödinger supposed to do? The best choice is always to begin with the simplest toys, such as the free particle, the particle in a box, the harmonic oscillator, the rigid rotator or hydrogen atom. Nothing is known about whether Schrödinger himself had a sufficiently deep knowledge of mathematics to be able to solve the (sometimes non-trivial) equations related to these problems, or whether he had advice from a friend versed in mathematics, such as Hermann Weyl.

Hermann Weyl (1885–1955), German mathematician, professor at ETH Zurich, then the University of Göttingen and the Institute for Advanced Studies at Princeton (USA), expert in the theory of orthogonal series, group theory and differential equations. Weyl adored Schrödinger's wife, was a friend of the family, and provided an ideal partner for Schrödinger in conversations about the eigenproblem.

It turned out that instead of p, $-i\hbar\frac{\mathrm{d}}{\mathrm{d}x}$ had to be inserted, and not $i\hbar\frac{\mathrm{d}}{\mathrm{d}x}$ (Postulate II, Chapter 1).

---

[19] Very interesting coincidence: Heisenberg was also involved in fluid dynamics. At the beginning, Schrödinger did not use operators. They appeared after he established closer contacts with the University of Göttingen.

2.2.1 WAVE FUNCTIONS OF CLASS Q

The postulates of quantum mechanics, especially the probabilistic interpretation of the wave function given by Max Born, limits the class of functions allowed (to "class Q", or "quantum").

Any wave function

- cannot be zero everywhere (Fig. 2.5.a), because the system *is somewhere* in space;
- has to be continuous, (Fig. 2.5.b). This also means it cannot take infinite values at any point in space[20] (Fig. 2.5.c,d);
- has to have a continuous first derivative as well (everywhere in space except isolated points (Fig. 2.5.e,f), where the potential energy tends to $-\infty$), because the Schrödinger equation is a second order differential equation and the second derivative must be defined;
- has to have a uniquely defined value in space,[21] Fig. 2.5.g,h;
- for bound states has to tend to zero at infinite values of any of the coordinates (Fig. 2.5.i,j), because such a system is compact and does not disintegrate in space. In consequence (from the probabilistic interpretation), the wave function is square integrable, i.e. $\langle\Psi|\Psi\rangle < \infty$.

2.2.2 BOUNDARY CONDITIONS

The Schrödinger equation is a differential equation. In order to obtain a special solution to such equations, one has to insert the particular boundary conditions to be fulfilled. Such conditions follow from the physics of the problem, i.e. with which kind of experiment are we going to compare the theoretical results? For example:

bound states

- for the *bound states* (i.e. square integrable states) we put the condition that the wave function has to vanish at infinity, i.e. if any of the coordinates tends to infinity: $\psi(x = \infty) = \psi(x = -\infty) = 0$;
- for cyclic systems of circumference L, the natural conditions will be: $\psi(x) = \psi(x + L)$ and $\psi'(x) = \psi'(x + L)$, because they ensure a smooth matching of the wave function for $x < 0$ and of the wave function for $x > 0$ at $x = 0$;
- for scattering states (not discussed here) the boundary conditions are more complex.[22]

There is a countable number of bound states. Each state corresponds to eigenvalue E.

---

[20] If this happened in any non-zero volume of space (Fig. 2.5.d) the probability would tend to infinity (which is prohibited). However, the requirement is stronger than that: a wave function cannot take an infinite value even at a single point, Fig. 2.5.c. Sometimes such functions appear among the solutions of the Schrödinger equation, and have to be rejected. The formal argument is that, if not excluded from the domain of the Hamiltonian, the latter would be non-Hermitian when such a function were involved in $\langle f|\hat{H}g\rangle = \langle\hat{H}f|g\rangle$. A non-Hermitian Hamiltonian might lead to complex energy eigenvalues, which is prohibited.

[21] At any point in space the function has to have a single value. This plays a role only if we have an angular variable, say ϕ. Then, ϕ and $\phi + 2\pi$ have to give the same value of the wave function. We will encounter this problem in the solution for the rigid rotator.

[22] J.R. Taylor, "*Scattering Theory*", Wiley, New York, 1972 is an excellent reference.

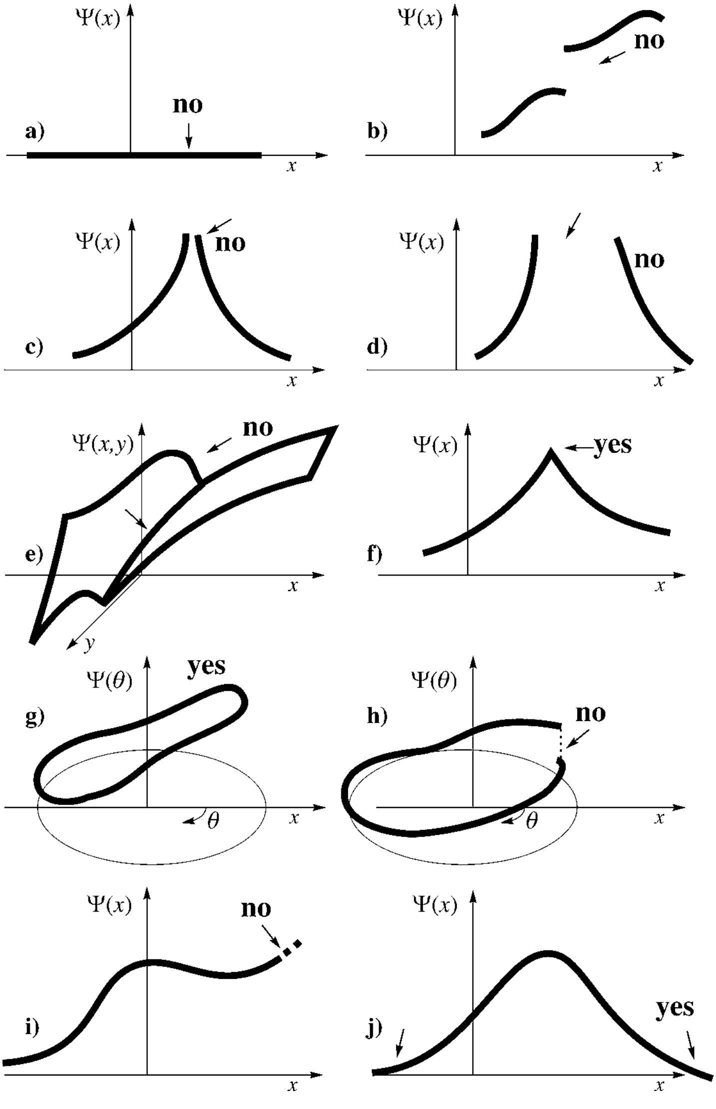
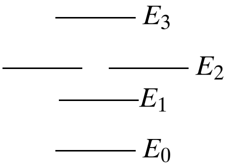


Fig. 2.5. Functions of class Q (i.e. wave functions allowed in quantum mechanics) – examples and counterexamples. A wave function (a) must not be zero everywhere in space (b) has to be continuous (c) cannot tend to infinity even at a single point (d) cannot tend to infinity (e) its first derivative cannot be discontinuous for infinite number of points (f) its first derivative may be discontinuous for a finite number of points (g) has to be defined uniquely in space (for angular variable θ) (h) cannot correspond to multiple values at a point in space (for angular variable θ) (i) for bound states: must not be non-zero in infinity (j) for bound states: has to vanish in infinity.

degeneracy

An energy level may be *degenerate*, that is, more than one wave function may correspond to it, all the wave functions being linearly independent (their number is the *degree of degeneracy*). The eigenvalue spectrum is usually represented by putting a single horizontal section (in the energy scale) for each wave function:



2.2.3 AN ANALOGY

Let us imagine all the stable positions of a chair on the floor (Fig. 2.6).

Consider a simple chair, very uncomfortable for sitting, but very convenient for a mathematical exercise. Each of the four legs represents a rod of length a, the „seat” is simply a square built of the rods, the back consists of three such rods making a C shape. The potential energy of the chair (in position i) in a gravitational field equals mgh_i, where m stands for the mass of the chair, g gravitational acceleration, and h_i denotes the height of the centre of mass with respect to the floor. We obtain the following energies, E_i, of the stationary states (in units of mga):

- the chair is lying on the support: $E_0 = \frac{4}{11}$;
- the chair is lying inclined: the support and the seat touch the floor $E_1 = \frac{7\sqrt{2}}{22} = 0.45$;
- the chair is lying on the side: $E_2 = \frac{1}{2}$.

Note, however, that we have *two* sides. The energy is the same for the chair lying on the first and second side (because the chair is symmetric), but these are *two* states of the chair, not one. The degree of degeneracy equals *two*, and therefore on the energy diagram we have two horizontal sections. Note how naturally the problem of degeneracy has appeared. The degeneracy of the energy eigenstates of molecules results from symmetry, exactly as in the case of the chair. In some cases,

accidental degeneracy

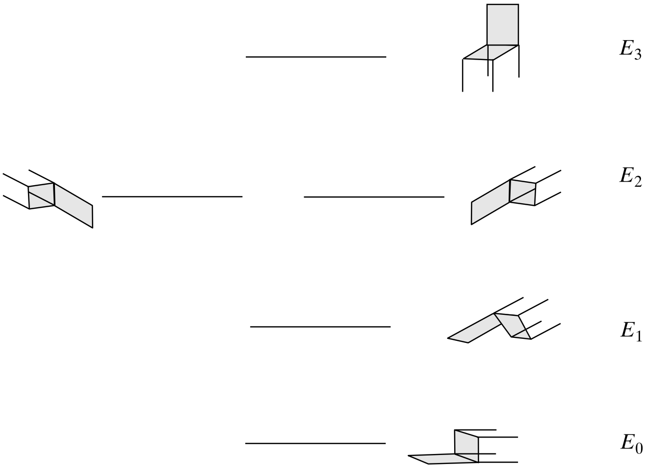


Fig. 2.6. The stable positions of a chair on the floor (arbitrary energy scale). In everyday life we most often use the third excited state.

one may obtain an *accidental degeneracy*, which does not follow from the symmetry of an object like a chair, but from the properties of the potential field, and is called *dynamical symmetry*.[23]

dynamical symmetry

- the chair is in the normal position: $E_3 = 1$.

There are no more stable states of the chair[24] and there are only four energy levels, Fig. 2.6. The stable states of the chair are analogues to the stationary quantum states of Fig. 1.8.b,d, on p. 23, while unstable states of the chair on the floor are analogues of the non-stationary states of Fig. 1.8.a,c.

2.2.4 MATHEMATICAL AND PHYSICAL SOLUTIONS

> It is worth noting that not all solutions of the Schrödinger equation are physically acceptable.

For example, for bound states, all solutions other than those of class Q (see p. 895) must be rejected. In addition, these solutions ψ, which do not exhibit the proper symmetry, even if $|\psi|^2$ does, have also to be rejected. They are called mathematical (non-physical) solutions to the Schrödinger equation. Sometimes such mathematical solutions correspond to a *lower* energy than any physically acceptable energy (known as *underground states*). In particular, such illegal, non-acceptable functions are *asymmetric* with respect to the label exchange for electrons (e.g., symmetric for some pairs and antisymmetric for others). Also, a fully symmetric function would also be such a non-physical (purely mathematical) solution.

underground states

2.3 THE TIME-DEPENDENT SCHRÖDINGER EQUATION

What would happen if one prepared the system in a given state ψ, which does not represent a stationary state? For example, one may deform a molecule by using an electric field and then switch the field off.[25] The molecule will suddenly turn out to be in state ψ, that is not its stationary state. Then, according to quantum mechanics, the state of the molecule will start to change according to the time evolution equation (time-dependent Schrödinger equation)

$$\hat{H}\psi = i\hbar\frac{\partial\psi}{\partial t}. \tag{2.13}$$

---

[23] Cf. the original works C. Runge, “*Vektoranalysis*”, vol. I, p. 70, ed. S. Hirzel, Leipzig, 1919, W. Lenz, *Zeit. Physik* 24 (1924) 197 as well as L.I. Schiff, “*Quantum Mechanics*”, McGraw Hill (1968).

[24] Of course, there are plenty of unstable positions of the chair with respect to the floor. The stationary states of the chair have more in common with chemistry than we might think. A chair-like molecule (organic chemists have already synthesized much more complex molecules) interacting with a crystal surface would very probably have similar stationary (vibrational) states.

[25] We neglect the influence of the magnetic field that accompanies any change of electric field.

The equation plays a role analogous to Newton's equation of motion in classical mechanics. The position and momentum of a particle change according to Newton's equation. In the time-dependent Schrödinger equation the evolution proceeds in a completely different space – in the space of states or Hilbert space (cf. Appendix B, p. 895).

> Therefore, in quantum mechanics one has absolute determinism, but in the state space. Indeterminism begins only in our space, when one asks about the coordinates of a particle.

2.3.1 EVOLUTION IN TIME

As is seen from eq. (2.13), knowledge of the Hamiltonian and of the wave function at a given time (left-hand side), represents sufficient information to determine the time derivative of the wave function (right-hand side). This means that we may compute the wave function after an infinitesimal time dt:

$$\psi + \frac{\partial\psi}{\partial t}\,\mathrm{d}t = \psi - \frac{i}{\hbar}\hat{H}\psi\,\mathrm{d}t = \left[1 + \left(-i\frac{t}{N\hbar}\right)\hat{H}\right]\psi,$$

where we have set $\mathrm{d}t = t/N$ with N (natural number) very large. Thus, the new wave function results from action of the operator $[1 + (-i\frac{t}{N\hbar})\hat{H}]$ on the old wave function. Now, we may pretend that we did not change any function and apply the operator again and again. We assume that $\hat{H}$ is time-independent. The total operation is nothing but the action of the operator:

$$\lim_{N\to\infty}\left[1 + \left(-i\frac{t}{N\hbar}\right)\hat{H}\right]^N.$$

Please recall that $e^x = \lim_{N\to\infty}[1 + \frac{x}{N}]^N$.

> Hence, the time evolution corresponds to action on the initial ψ of the operator $\exp(-\frac{it}{\hbar}\hat{H})$:
>
> $$\psi' = \exp\left(-\frac{it}{\hbar}\hat{H}\right)\psi. \tag{2.14}$$

Our result satisfies the time-dependent Schrödinger equation,[26] if $\hat{H}$ does not depend on time (as we assumed when constructing ψ').

Inserting the spectral resolution of the identity[27] (cf. Postulate II in Chapter 1)

---

[26] One may verify inserting ψ' into the Schrödinger equation. Differentiating ψ' with respect to t, the left-hand-side is obtained.

[27] The use of the spectral resolution of the identity in this form is not fully justified. A sudden cut in the electric field may leave the molecule with a non-zero translational energy. However, in the above spectral resolution one has the stationary states computed in the centre-of-mass coordinate system, and therefore translation is not taken into account.

one obtains[28]

$$\psi' = \exp\left(-i\frac{t}{\hbar}\hat{H}\right)1\psi = \exp\left(-i\frac{t}{\hbar}\hat{H}\right)\sum_n |\psi_n\rangle\langle\psi_n|\psi\rangle$$
$$= \sum_n \langle\psi_n|\psi\rangle \exp\left(-i\frac{t}{\hbar}\hat{H}\right)|\psi_n\rangle = \sum_n \langle\psi_n|\psi\rangle \exp\left(-i\frac{t}{\hbar}E_n\right)|\psi_n\rangle.$$

This is how the state ψ will evolve. It will be similar to one or another stationary state ψ_n, more often to those ψ_n which overlap significantly with the starting function (ψ) and/or correspond to low energy (low frequency). If the overlap $\langle\psi_n|\psi\rangle$ of the starting function ψ with a stationary state ψ_n is zero, then during the evolution no admixture of the ψ_n state will be seen, i.e. *only those stationary states that constitute the starting wave function ψ contribute to the evolution*.

2.3.2 NORMALIZATION IS PRESERVED

Note that the imaginary unit i is important in the formula for ψ'. If "i" were absent in the formula, then ψ' would be unnormalized (even if ψ is). Indeed,

$$\langle\psi'|\psi'\rangle = \sum_n\sum_m \langle\psi|\psi_m\rangle\langle\psi_n|\psi\rangle \exp\left(-i\frac{t}{\hbar}E_n\right)\exp\left(+i\frac{t}{\hbar}E_m\right)\langle\psi_m|\psi_n\rangle$$
$$= \sum_n\sum_m \langle\psi|\psi_m\rangle\langle\psi_n|\psi\rangle \exp\left(-i\frac{t}{\hbar}(E_n - E_m)\right)\delta_{mn} = \langle\psi|\psi\rangle = 1.$$

Therefore, the evolution preserves the normalization of the wave function.

2.3.3 THE MEAN VALUE OF THE HAMILTONIAN IS PRESERVED

The mean value of the Hamiltonian is a time-independent quantity. Indeed,

$$\left\langle\psi'\middle|\hat{H}\psi'\right\rangle = \left\langle \exp\left(-i\frac{t}{\hbar}\hat{H}\right)\psi \middle| \hat{H}\exp\left(-i\frac{t}{\hbar}\hat{H}\right)\psi\right\rangle$$
$$= \left\langle\psi\middle|\exp\left(i\frac{t}{\hbar}\hat{H}\right)\hat{H}\exp\left(-i\frac{t}{\hbar}\hat{H}\right)\psi\right\rangle$$
$$= \left\langle\psi\middle|\hat{H}\exp\left(i\frac{t}{\hbar}\hat{H}\right)\exp\left(-i\frac{t}{\hbar}\hat{H}\right)\psi\right\rangle = \langle\psi|\hat{H}\psi\rangle,$$

because $\exp(-i\frac{t}{\hbar}\hat{H})^\dagger = \exp(i\frac{t}{\hbar}\hat{H})$ (Appendix B, p. 895) and, of course, $\exp(i\frac{t}{\hbar}\hat{H})$ commutes with $\hat{H}$. The evolution of a non-degenerate stationary state has a trivial time dependence through the factor $\exp(-i\frac{t}{\hbar}E_n)$. If a stationary state is degenerate then it may evolve in a non-trivial way, but at a given time the wave function is always a linear combination of the wave functions corresponding to this energy

---

[28] We used here the property of an analytical function f, that for any eigenfunction ψ_n of the operator $\hat{H}$ one has $f(\hat{H})\psi_n = f(E_n)\psi_n$. This follows from the Taylor expansion of $f(\hat{H})$ acting on eigenfunction ψ_n.

level only. However, starting from a non-stationary state (even if the mean energy is equal to the energy of a stationary state), one never reaches a pure stationary state during evolution.

> Until a coupling of the system with an electromagnetic field is established, the excited states have an infinite lifetime. However, in reality the excited states have a *finite* lifetime, emit photons, and as a result the energy of the system is lowered (although *together with the photons* the energy remains constant). Quantitative description of spontaneous photon emission has been given by Einstein.

2.3.4 LINEARITY

The most mysterious feature of the Schrödinger equation (2.13) is its linear character. The world is non-linear, because effect is never strictly proportional to cause. However, if $\psi_1(x,t)$ and $\psi_2(x,t)$ satisfy the time dependent Schrödinger equation, then their arbitrary linear combination also represents a solution.[29]

2.4 EVOLUTION AFTER SWITCHING A PERTURBATION

Let us suppose that we have a system with the Hamiltonian $\hat{H}^{(0)}$ and its stationary states $\psi_k^{(0)}$:

$$\hat{H}^{(0)}\psi_k^{(0)} = E_k^{(0)}\psi_k^{(0)}, \tag{2.15}$$

that form the orthonormal complete set[30]

$$\psi_k^{(0)}(x,t) = \phi_k^{(0)}(x)\exp\left(-i\frac{E_k^{(0)}}{\hbar}t\right), \tag{2.16}$$

where x represents the coordinates, and t denotes time.

> Let us assume, that at time $t=0$ the system is in the stationary state $\psi_m^{(0)}$. At $t=0$ a drama begins: one switches on the perturbation $V(x,t)$ that in general depends on all the coordinates (x) and time (t), and after time τ the perturbation is switched off. Now we ask question about the probability of finding the system in the stationary state $\psi_k^{(0)}$.

---

[29] Indeed, $\hat{H}(c_1\psi_1 + c_2\psi_2) = c_1\hat{H}\psi_1 + c_2\hat{H}\psi_2 = c_1 i\hbar\frac{\partial\psi_1}{\partial t} + c_2 i\hbar\frac{\partial\psi_2}{\partial t} = i\hbar\frac{\partial(c_1\psi_1 + c_2\psi_2)}{\partial t}$.

[30] This *can* always be assured (by suitable orthogonalization and normalization) and follows from the Hermitian character of the operator $\hat{H}^{(0)}$.

After the perturbation is switched on, the wave function $\psi_m^{(0)}$ is no longer stationary and begins to evolve in time according to the time-dependent Schrödinger equation $(\hat{H}^{(0)} + \hat{V})\psi = i\hbar\frac{\partial\psi}{\partial t}$. This is a differential equation with partial derivatives with the boundary condition $\psi(x, t = 0) = \phi_m^{(0)}(x)$. The functions $\{\psi_n^{(0)}\}$ form a complete set and therefore the wave function $\psi(x, t)$ that fulfils the Schrödinger equation at any time can be represented as a linear combination with time-dependent coefficients c:

$$\psi(x, t) = \sum_{n=0}^{\infty} c_n(t)\psi_n^{(0)}. \tag{2.17}$$

Inserting this to the left-hand side of the time-dependent Schrödinger equation one obtains:

$$\big(\hat{H}^{(0)} + \hat{V}\big)\psi = \sum_n c_n\big(\hat{H}^{(0)} + \hat{V}\big)\psi_n^{(0)} = \sum_n c_n\big(E_n^{(0)} + V\big)\psi_n^{(0)},$$

whereas its right-hand side gives:

$$i\hbar\frac{\partial\psi}{\partial t} = i\hbar\sum_n\left[\psi_n^{(0)}\frac{\partial c_n}{\partial t} + c_n\frac{\partial\psi_n^{(0)}}{\partial t}\right] = \sum_n\left[i\hbar\psi_n^{(0)}\frac{\partial c_n}{\partial t} + c_n E_n^{(0)}\psi_n^{(0)}\right].$$

Both sides give:

$$\sum_n c_n\hat{V}\psi_n^{(0)} = \sum_n\left(i\hbar\frac{\partial c_n}{\partial t}\right)\psi_n^{(0)}.$$

Multiplying the left-hand side by $\psi_k^{(0)*}$ and integrating result in:

$$\boxed{\sum_n^{\infty} c_n V_{kn} = i\hbar\frac{\partial c_k}{\partial t},} \tag{2.18}$$

for $k = 0, 1, 2, \ldots,$ where

$$V_{kn} = \langle\psi_k^{(0)}|\hat{V}\psi_n^{(0)}\rangle. \tag{2.19}$$

The formulae obtained are equivalent to the Schrödinger equation. These are differential equations, which we would generally like, provided the summation is not infinite, but in fact it is.[31] In practice, however, one has to keep the summation finite.[32] If the assumed number of terms in the summation is not too large, then problem is soluble using modern computer techniques.

---

[31] Only then the equivalence to the Schrödinger equation is ensured.

[32] This is typical for expansions into the complete set of functions (the so called *algebraic approximation*).

2.4.1 THE TWO-STATE MODEL

Let us take the *two-state model* (cf. Appendix D, p. 948) with two orthonormal eigenfunctions $|\psi_1^{(0)}\rangle = |1\rangle$ and $|\psi_2^{(0)}\rangle = |2\rangle$ of the Hamiltonian $\hat{H}^{(0)}$

$$\hat{H}^{(0)} = E_1^{(0)}|1\rangle\langle 1| + E_2^{(0)}|2\rangle\langle 2|$$

with the perturbation ($v_{12} = v_{21}^* = v$):

$$V = v_{12}|1\rangle\langle 2| + v_{21}|2\rangle\langle 1|.$$

This model has an exact solution (even for a large perturbation V). One may introduce various time-dependences of V, including various regimes for switching on the perturbation.

The differential equations (2.18) for the coefficients $c_1(t)$ and $c_2(t)$ are (in a.u., $\omega_{21} = E_2^{(0)} - E_1^{(0)}$)

$$c_1 v \exp(i\omega_{21}t) = i\frac{\partial c_2}{\partial t},$$

$$c_2 v \exp(-i\omega_{21}t) = i\frac{\partial c_1}{\partial t}.$$

Let us assume as the initial wave function $|2\rangle$, i.e. $c_1(0) = 0$, $c_2(0) = 1$. In such a case one obtains

$$c_1(t) = -\frac{i}{a}\exp\left(-i\frac{\omega_{21}}{2}t\right)\sin(avt),$$

$$c_2(t) = \frac{1}{a}\exp(i\omega_{21}t)\cos\left[-avt + \operatorname{arcSec}\left(\frac{1}{a}\right)\right],$$

where $a = \sqrt{1 + (\frac{\omega_{21}}{2v})^2}$, and Sec denotes $\frac{1}{\cos}$.

Two states – degeneracy

One of the most important cases corresponds to the degeneracy $\omega_{21} = E_2^{(0)} - E_1^{(0)} = 0$. One obtains $a = 1$ and

$$c_1(t) = -i\sin(vt),$$

$$c_2(t) = \cos(vt).$$

A very interesting result. The functions $|1\rangle$ and $|2\rangle$ may be identified with the ψ_D and ψ_L functions for the D and L enantiomers (cf. p. 68) or, with the wave functions 1s, centred on the two nuclei in the H_2^+ molecule. As one can see from the last two equations, the two wave functions oscillate, transforming one to the oscillations other with an oscillation period $T = \frac{2\pi}{v}$. If v were very small (as in the case of D- and L-glucose), then the oscillation period would be very large. This happens to D- and L-enantiomers of glucose, where changing the nuclear configuration from one to the other enantiomer means breaking a chemical bond (a high and wide energy barrier to overcome). This is why the drugstore keeper can safely stock a single enantiomer for a very long time.[33] This however may not be true for other enantiomers. For example, imagine a pair of enantiomers that represent some intermolecular complexes and a small change of the nuclear framework may cause one of them to transform into the other. In such a case, the oscillation period may be much, much smaller than the lifetime of the Universe, e.g., it may be comparable to the time of an experiment. In such a case one could observe the oscillation between the two enantiomers. This is what happens in reality. One observes a spontaneous racemization, which is of dynamical character, i.e. a single molecule oscillates between D and L forms.

2.4.2 FIRST-ORDER PERTURBATION THEORY

If one is to apply first-order perturbation theory, two things have to be assured: the perturbation V has to be small and the time of interest has to be small (switching the perturbation in corresponds to $t = 0$). This is what we are going to assume from now on. At $t = 0$ one starts from the m-th state and therefore $c_m = 1$, while other coefficients $c_n = 0$. Let us assume that to first approximation this will be true even *after* switching the perturbation on, and we will be interested in the tendencies in the time-evolution of c_n for $n \neq m$. These assumptions (based on first-order perturbation theory) lead to a considerable simplification[34] of eqs. (2.18):

$$V_{km} = i\hbar \frac{\partial c_k}{\partial t} \quad \text{for } k = 1, 2, \ldots, N.$$

In this, and the further equations of this chapter, the coefficients c_k will depend implicitly on the initial state m.

The quantity V_{km} depends on time for two or even three reasons: firstly and secondly, the stationary states $\psi_m^{(0)}$ and $\psi_k^{(0)}$ of eq. (2.16) are time-dependent, and thirdly, in addition the perturbation V may also depend on time. Let us highlight the time-dependence of the wave functions by introducing the frequency

$$\omega_{km} = \frac{E_k^{(0)} - E_m^{(0)}}{\hbar}$$

and the definition

$$v_{km} \equiv \langle \phi_k^{(0)} | \hat{V} \phi_m^{(0)} \rangle.$$

---

[33] Let us hope no longer than the sell-by date.

[34] For the sake of simplicity we will not introduce a new notation for the coefficients c_n corresponding to a first-order procedure. If the above simplified equation were introduced to the left-hand side of eq. (2.18), then its solution would give c accurate up to the second order, etc.

One obtains

$$-\frac{i}{\hbar}v_{km}\, e^{i\omega_{km}t} = \frac{\partial c_k}{\partial t}.$$

Subsequent integration with the boundary condition $c_k(\tau = 0) = 0$ for $k \neq m$ gives:

$$c_k(\tau) = -\frac{i}{\hbar}\int_0^{\tau} \mathrm{d}t\, v_{km}(t)e^{i\omega_{km}t}. \tag{2.20}$$

The square of $c_k(\tau)$ represents (to the accuracy of first-order perturbation theory), the probability that at time τ the system will be found in state $\psi_k^{(0)}$. Let us calculate this probability for a few important cases of the perturbation $\hat{V}$.

2.4.3 TIME-INDEPENDENT PERTURBATION AND THE FERMI GOLDEN RULE

From formula (2.20) we have

$$c_k(\tau) = -\frac{i}{\hbar}v_{km}\int_0^{\tau} \mathrm{d}t\, e^{i\omega_{km}t} = -\frac{i}{\hbar}v_{km}\frac{e^{i\omega_{km}\tau}-1}{i\omega_{km}} = -v_{km}\frac{e^{i\omega_{km}\tau}-1}{\hbar\omega_{km}}. \tag{2.21}$$

Now let us calculate the probability density $P_m^k = |c_k|^2$, that at time τ the system will be in state k (the initial state is m):

$$\begin{aligned} P_m^k(\tau) &= |v_{km}|^2\frac{(-1+\cos\omega_{km}\tau)^2 + \sin^2\omega_{km}\tau}{(\hbar\omega_{km})^2} = |v_{km}|^2\frac{(2-2\cos\omega_{km}\tau)}{(\hbar\omega_{km})^2} \\ &= |v_{km}|^2\frac{(4\sin^2\frac{\omega_{km}\tau}{2})}{(\hbar\omega_{km})^2} = |v_{km}|^2\frac{1}{\hbar^2}\frac{(\sin^2\frac{\omega_{km}\tau}{2})}{(\frac{\omega_{km}}{2})^2}. \end{aligned}$$

In order to undergo the transition from state m to state k one has to have a large v_{km}, i.e. a large coupling of the two states through perturbation $\hat{V}$. Note that probability P_m^k strongly depends on the time τ chosen; the probability oscillates as the square of the sine when τ increases, for some τ it is large, for others it is just zero. From Example 4 in Appendix E, p. 951, one can see that for large values of τ one may write the following approximation[35] to P_m^k:

$$P_m^k(\tau) \cong |v_{km}|^2\pi\frac{\tau}{\hbar^2}\delta\left(\frac{\omega_{km}}{2}\right) = \frac{2\pi\tau}{\hbar^2}|v_{km}|^2\delta(\omega_{km}) = \frac{2\pi\tau}{\hbar}|v_{km}|^2\delta\big(E_k^{(0)} - E_m^{(0)}\big),$$

where we have used twice the Dirac delta function property that $\delta(ax) = \frac{\delta(x)}{|a|}$.

---

[35] Large when compared to $2\pi/\omega_{km}$, but not too large in order to keep the first-order perturbation theory valid.

As one can see, P_m^k is proportional to time τ, which makes sense only because τ has to be relatively small (first-order perturbation theory has to be valid). Note that the Dirac delta function forces the energies of both states (the initial and the final) to be equal, because of the time-independence of V.

> A time-independent perturbation is unable to change the state of the system when it corresponds to a change of its energy.

A very similar formula is systematically derived in several important cases. Probably this is why the probability per unit time is called, poetically, the Fermi golden rule:[36]

> FERMI GOLDEN RULE
>
> $$w_m^k \equiv \frac{P_m^k(\tau)}{\tau} = |v_{km}|^2 \frac{2\pi}{\hbar} \delta\big(E_k^{(0)} - E_m^{(0)}\big). \tag{2.22}$$

2.4.4 THE MOST IMPORTANT CASE: PERIODIC PERTURBATION

Let us assume a time-dependent periodic perturbation

$$\hat{V}(x,t) = \hat{v}(x)e^{\pm i\omega t}.$$

Such a perturbation corresponds, e.g., to an oscillating electric field[37] of angular frequency ω.

Let us take a look at successive equations, which we obtained at the time-independent $\hat{V}$. The only change will be, that V_{km} will have the form

$$V_{km} \equiv \langle \psi_k^{(0)} | \hat{V} \psi_m^{(0)} \rangle = v_{km} e^{i(\omega_{km} \pm \omega)t}$$

instead of

$$V_{km} \equiv \langle \psi_k^{(0)} | \hat{V} \psi_m^{(0)} \rangle = v_{km} e^{i\omega_{km} t}.$$

The whole derivation will be therefore identical, except that the constant ω_{km} will be replaced by $\omega_{km} \pm \omega$. Hence, we have a new form of the Fermi golden rule for the probability per unit time of transition from the m-th to the k-th state:

---

[36] E. Fermi, *Nuclear Physics*, University of Chicago Press, Chicago, 1950, p. 142.

[37] In the homogeneous field approximation, the field interacts with the dipole moment of the molecule (cf. Chapter 12) $V(x,t) = V(x)e^{\pm i\omega t} = -\hat{\boldsymbol{\mu}} \cdot \boldsymbol{\mathcal{E}} e^{\pm i\omega t}$, where $\boldsymbol{\mathcal{E}}$ denotes the electric field intensity of the light wave and $\hat{\boldsymbol{\mu}}$ is the dipole moment operator.

FERMI GOLDEN RULE

$$w_m^k \equiv \frac{P_m^k(\tau)}{\tau} = |v_{km}|^2 \frac{2\pi}{\hbar} \delta\big(E_k^{(0)} - E_m^{(0)} \pm \hbar\omega\big). \tag{2.23}$$

Note, that V with $\exp(+i\omega t)$ gives the equality $E_k^{(0)} + \hbar\omega = E_m^{(0)}$, which means that $E_k^{(0)} \leqslant E_m^{(0)}$ and therefore one has emission from the m-th to the k-th states. On the other hand, V with $\exp(-i\omega t)$ forces the equation $E_k^{(0)} - \hbar\omega = E_m^{(0)}$, which corresponds to absorption from the m-th to the k-th state.

Therefore, a periodic perturbation is able to make a transition between states of different energy.

Summary

The Hamiltonian of any isolated system is invariant with respect to the following transformations (operations):

- any translation in time (homogeneity of time)
- any translation of the coordinate system (space homogeneity)
- any rotation of the coordinate system (space isotropy)
- inversion ($\boldsymbol{r} \to -\boldsymbol{r}$)
- reversing all charges (charge conjugation)
- exchanging labels of identical particles.

This means that the wave function corresponding to a stationary state (the eigenfunction of the Hamiltonian) also has to be an eigenfunction of the:

- total momentum operator (due to the translational symmetry)
- total angular momentum operator and one of its components (due to the rotational symmetry)
- inversion operator
- any permutation (of identical particles) operator (due to the non-distinguishability of identical particles)
- $\hat{S}^2$ and $\hat{S}_z$ operators (for the non-relativistic Hamiltonian (2.1) due to the absence of spin variables in it).

Such a wave function corresponds to the energy belonging to the energy continuum.[38] Only after separation of the centre-of-mass motion does one obtain the spectroscopic states (belonging to a discrete spectrum) $\Psi_{N,J,M,\Pi}(\boldsymbol{r},\boldsymbol{R})$, where $N = 0, 1, 2, \ldots$ denotes the number of the electronic state, $J = 0, 1, 2, \ldots$ quantizes the total angular momentum,

---

[38] Because the molecule as a whole (i.e. its centre of mass) may have an arbitrary kinetic energy. Sometimes it is rewarding to introduce the notion of the *quasicontinuum* of states, which arises if the system is enclosed in a large box instead of considering it in infinite space. This simplifies the underlying mathematics.

M, $-J \leqslant M \leqslant J$, quantizes its component along the z axis, and $\Pi = \pm 1$ represents the parity with respect to the inversion. As to the invariance with respect to permutations of identical particles: an acceptable wave function has to be antisymmetric with respect to the exchange of identical fermions, whereas it has to be symmetric when exchanging bosons.

The time-independent Schrödinger equation $\hat{H}\psi = E\psi$ has been "derived" from the wave equation and the de Broglie equation. Solving this equation results in the stationary states and their energies. This is the basic equation of quantum chemistry. The prevailing weight of research in this domain is concentrated on solving this equation for various systems.

The time-dependent Schrödinger equation $\hat{H}\psi = i\hbar\frac{\partial\psi}{\partial t}$ represents the time evolution of an arbitrary initial wave function. The assumption that translation in time is a unitary operator leads to preserving the normalization of the wave function and of the mean value of the Hamiltonian. If the Hamiltonian is time-independent, then one obtains the formal solution to the Schrödinger equation by applying the operator $\exp(-\frac{it}{\hbar}\hat{H})$ to the initial wave function. The time evolution of the stationary state $\psi_m^{(0)}$ is most interesting in the case of suddenly switching the perturbation $\hat{V}$. The state is no longer stationary and the wave function begins to change as time passes. Two cases have been considered:

- time-independent perturbation
- periodic perturbation.

Only in the case of a time-dependent perturbation may the system change the energy state.

Main concepts, new terms

symmetry of the Hamiltonian (p. 57)
invariance of theory (p. 58)
translational symmetry (p. 61)
spectroscopic state (p. 62)
rotational symmetry (p. 63)
baryon number (p. 64)
lepton number (p. 64)
gauge symmetry (p. 64)
inversion (p. 65)
symmetry P (p. 65)
enantiomers (p. 67)
symmetry C (p. 68)
charge conjugation (p. 68)
stationary state (p. 70)
Schrödinger equation (p. 70)
bound state (p. 73)
wave function "matching" (p. 73)
mathematical solution (p. 76)
physical solutions (p. 76)
wave function evolution (p. 76)
time-evolution operator (p. 77)
algebraic approximation (p. 80)
two-state model (p. 81)
first-order perturbation theory (p. 82)
time-independent perturbation (p. 83)
Fermi golden rule (p. 84)
periodic perturbation (p. 84)

From the research front

The overwhelming majority of research in the domain of quantum chemistry is based on the solution of the time-independent Schrödinger equation. Without computers it was possible to solve (in an approximate way) the equation for H_2^+ by conducting a hall full of secretaries with primitive calculators for many hours (what a determination!). Thanks to computers, such problems became easy as early as the 1960s. Despite enormous progress in computer science,[39] by the end of the 1980s the molecules studied were rather small when compared

---

[39] A boss of the computer industry reportedly declared that if progress similar to that in his branch would occur in the car industry, a Mercedes would cost 1$ and would take one gallon of gas to go

to the expectations of experimentalists. They could be treated only as models, because they were usually devoid of the substituents theoreticians considered irrelevant in the behaviour of the molecule. The last years of the twentieth century were marked by the unprecedented delivery by theoreticians of powerful high-tech efficient tools of quantum chemistry available to everybody: chemists, physicists, etc. Also to those who knew little of the underlying theory. The software computes millions of integrals, uses sophisticated mathematics, literally the whole arsenal of quantum chemistry, but users need not know about it. It is sufficient to make … a mouse click on a quantum chemistry method icon.[40] Despite such progress, the time-dependent Schrödinger equation is rarely solved. For the time being, researchers are interested mainly in stationary states. The quality of results depends on the size of the molecules investigated. Very accurate computations (accuracy ~ 0.01 kcal/mol) are feasible for the smallest molecules containing a few electrons, less accurate ones use first principles (*ab initio methods*) and are feasible for hundreds of atoms (accuracy to a few kcals/mol). Semiempirical quantum calculations[41] of even poorer accuracy are applicable to thousands of atoms.

oscillations

Ad futurum...

The numerical results routinely obtained so far indicate that, for the vast majority of chemical problems (yet not all, cf. Chapter 3) there is no need to search for a better tool than the Schrödinger equation. Future progress will be based on more and more accurate solutions for larger and larger molecules. The appetite is unlimited here, but the numerical difficulties increase much faster than the size of the system. However, progress in computer science has systematically opened new possibilities, always more exciting than previous ones. Some simplified alternatives to the Schrödinger equation (e.g., such as described in Chapter 11) will also become more important.

Undoubtedly, methods based on the time-dependent Schrödinger equation will also be developed. A typical possible target might be to plan a sequence of laser pulses,[42] such that the system undergoes a change of state from ψ_1 to ψ_2 ("*state-to-state reaction*"). The way we carry out chemical reactions, usually by rather primitive heating, may change to a precise transformation of the system from state to state.

state-to-state

It seems that the essence of science comes down to the fundamental question "*why?*" and a clear answer to this question follows from a deep understanding of Nature's machinery. We *cannot* tell a student: "well, this is what the computer says", because it implies that the computer understands, but what about you and me? Hence, interpretation of the results will be of crucial importance (a sort of Bader analysis, cf. Chapter 11). Progress here seems to be rather modest for the time being.

---

around the Earth. Somebody from the car sector answered that if cars were produced like computers, they would break down twice a day.

[40] I hope all students understand that a quantum chemist has to be equipped with something more than a strong forefinger for clicking.

[41] In such calculations many integrals are approximated by simple formulae (sometimes involving experimental data), the main goal of which is efficiency.

[42] I.e. a sinusoidal impulse for each of the sequences: switching on time, duration, intensity, phase. For contemporary laser technique it is an easy task. Now chemists should consider transforming reagents to products. The beginnings of such an approach are already present in the literature (e.g., J. Manz, G.K. Paramonov, M. Polášek, C. Schütte, *Isr. J. Chem.* 34 (1994) 115).

Additional literature

R. Feynman, The Character of Physical Law, Cox and Wyman, Ltd, London, (1965).
The best recommendation is that the Feynman's books need no recommendation.

J. Ciosłowski, in "Pauling's Legacy: Modern Modelling of the Chemical Bond", Elsevier, *Theor. Comput. Chem.*, 6 (1999) 1, eds. Z.B. Maksić, W.J. Orville-Thomas.
A concise presentation of the symmetry requirements.

Questions

1. If the Hamiltonian $\hat{H}$ is invariant with respect to a unitary operation $\hat{\mathcal{U}}$, then:
   a) the stationary states are the eigenfunctions of $\hat{\mathcal{U}}$ and $\hat{H}$; b) $\hat{\mathcal{U}}\hat{H} + \hat{H}\hat{\mathcal{U}} = 0$; c) $\hat{\mathcal{U}}\hat{H}\hat{\mathcal{U}} = \hat{H}$; d) $\hat{\mathcal{U}} = 1$.
2. The symmetry broken in the wave function for an optical isomer is:
   a) rotation; b) inversion; c) charge conjugation; d) permutational symmetry.
3. The spectroscopic states represent:
   a) such states of a molecule, that the optical transitions between them are especially intensive; b) the ground and the first excited states; c) the states that arise after switching on the electric field; d) the wave functions of the Hamiltonian that are calculated in the centre-of-mass coordinate system.
4. Degeneracy of the energy levels comes from:
   a) symmetry of the problem; b) requirement that the wave functions are of class Q; c) energy conservation rule; d) angular momentum conservation rule.
5. The time evolution operator in the Schrödinger equation (time-independent Hamiltonian $\hat{H}$) is equal to:
   a) $\exp(-\frac{i\hbar}{t})$; b) $\exp(-\frac{i\hat{H}}{t})$; c) $\exp(-\frac{t}{\hbar}\hat{H})$; d) $\exp(-\frac{i\hat{H}}{\hbar}t)$.
6. If one prepares the system in state ψ_D (one of the two non-stationary states ψ_D and ψ_L of equal energy), the system will:
   a) go to state ψ_L and remain in it; b) remain in ψ_D; c) oscillate between ψ_D and ψ_L; d) evolve to state $\psi_D \pm \psi_L$.
7. During time evolution of the wave function $\psi(t)$ of the system (the initial function is normalized, $\bar{E}$ stands for the mean value of the time-independent Hamiltonian)
   a) $|\psi(t)| = 1$, $\bar{E} = \text{const}$; b) $\int |\psi(t)|^2 = 1$, and $\bar{E}$ decreases; c) $\int |\psi(t)|^2 \, d\tau = 1$, and $\bar{E} = \text{const}$; d) $\bar{E}$ lowers, while the wave function stays normalized.
8. The Dirac delta $\delta(x)$ has the following properties:
   a) $\int_1^\infty f(x)\delta(x)\,dx = f(0)$ and $\delta(ax) = \delta(x)/|a|$; b) $\int_{-\infty}^\infty f(x)\delta(x)\,dx = f(0)$ and $\delta(ax) = \frac{\delta(x)}{|a|}$; c) $\int_{-\infty}^{-1} f(x)\delta(x)\,dx = f(0)$ and $\delta(ax) = \delta(x)|a|$;
   d) $\int_{-\infty}^\infty f(x)\delta(x)\,dx = f(0)$ and $\delta(ax) = \delta(x)/a$.
9. The state of the system corresponds to the energy $E_m^{(0)} < E_k^{(0)}$. During a short time period τ one applies a homogeneous time-independent electric field of intensity $\mathcal{E}$. The probability that after time τ the system will have energy $E_k^{(0)}$ is equal to:
   a) $1/2$; b) 0; c) 1; d) $\frac{1}{\mathcal{E}^2}$.

10. The state of the system corresponds to energy $E_m^{(0)} < E_k^{(0)}$. During a short time period τ one applies a periodic perturbation of amplitude $V(x)$, corresponding to the matrix elements denoted by v_{km}. The probability that after time τ the system will have energy $E_k^{(0)}$ is proportional to:

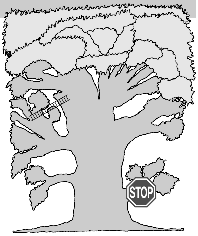
    a) $|v_{km}|$ and τ; b) $|v_{km}|^{-2}$ and $\tau^{\frac{1}{2}}$; c) $|v_{km}|^2$ and τ^{-1}; d) $|v_{km}|^2$ and τ.

Answers

1a, 2b, 3d, 4a, 5d, 6c, 7c, 8b, 9b, 10d

Chapter 3

BEYOND THE SCHRÖDINGER EQUATION



WHERE ARE WE?

Where are we?

The problems considered in the present chapter are shown as a small side-branch at the base of the TREE.[1]

An example

Copper, silver and gold – many people want to know everything about them (especially about the latter). The yellow shine of this metal has hypnotized humanity for centuries. Few people know that the colour of gold, as calculated assuming the infinite velocity of light (as it is in the Schrödinger equation), would be silver.[2] The Schrödinger equation fails. Here is an example of three diatomics: Cu_2, Ag_2, Au_2 ($Z_{Cu} = 29$, $Z_{Ag} = 47$, $Z_{Au} = 79$).

BOND LENGTH (Å)	Cu	Ag	Au
Non-relativistic calculations	2.26	2.67	**2.90**
Relativistic calculations	2.24	2.52	2.44
Experimental results	2.22	2.48	**2.47**

It is seen that the error of calculations within relativistic theories is of the order of 1%–2%, while the non-relativistic calculations lead to errors of the order of 2%, 8%, 20%, respectively. This is a *huge* discrepancy for such a quantity as bond length.

What is it all about

A glimpse of classical relativity theory (♦) **p. 93**

- The vanishing of apparent forces
- The Galilean transformation
- The Michelson–Morley experiment
- The Galilean transformation crashes

---

[1] This chapter owes much to the presentation given by L. Pisani, J.-M. André, M.-C. André, E. Clementi, *J. Chem. Educ*. 70 (1993) 894–901, as well as to the work of my friends J.-M. André, D.H. Mosley, M.-C. André, B. Champagne, E. Clementi, J.G. Fripiat, L. Leherte, L. Pisani, D. Vercauteren, M. Vracko, *Exploring Aspects of Computational Chemistry: Vol. I, Concepts*, Presses Universitaires de Namur, pp. 150–166 (1997), *Vol. II, Exercises*, Presses Universitaires de Namur, pp. 249–272 (1997), and J.-M. André, M.-C. André, "*Une introduction à la théorie de la relativité classique et quantique à l'usage des chimistes"*, Namur, 1999.

[2] P. Pyykkö, *Chem. Rev*. 88 (1988) 563; also P. Pyykkö, *ibid*. 97 (1997) 597.

The greater the velocity of an object the greater the errors in Newton dynamics. Electrons have a greater velocity when close to nuclei of large electric charges.[3] This is why relativistic corrections may turn out to be important for heavy elements.

relativistic mass effect

The Schrödinger equation is incompatible with special relativity theory. This has to be corrected somehow. This is far from being solved, but progress so far shows the Schrödinger equation, the spin of a particle, etc. in a new light.

Why is this important?

The subject of the present chapter addresses the very foundations of physics, and in principle has to be treated on an equal footing with the postulates of quantum mechanics. The Schrödinger equation of Chapter 2 does not fulfil (as will be shown in the present chapter) the requirements of relativity theory, and therefore is in principle "illegal". In the present chapter, Dirac's attempt to generalize the Schrödinger equation to adapt it to relativity theory will be described. If one assumes that particle velocities are small compared to that of light, then from this more general theory one obtains the Schrödinger equation. Also the

---

[3] This is easy to estimate. From Appendix H on p. 969 it follows that the mean value of the kinetic energy of an electron described by the $1s$ orbital in an atom of atomic number Z is equal to $\bar{T} = \frac{1}{2}Z^2$ (in a.u.). On the other hand, for a rough estimation of the electron velocity v, one may write $\bar{T} = \frac{mv^2}{2}$. This results in the expression $v = Z$ valid in a.u., while the velocity of light $c = 137.036$ a.u. The largest Z known hardly exceeds a hundred. It is seen, therefore, that if an atom with $Z > 137$ existed, then 1s electrons would attain velocities exceeding the velocity of light. Even if this calculation is nothing but a rule of thumb, there is no doubt that when Z increases a certain critical Z value is approached (the so called *relativistic mass effect*).

notion of the spin, which was introduced as a postulate in Chapter 1, follows as a natural consequence of the relativistic theory. One may draw the conclusion that the present chapter addresses "the foundations of foundations" and therefore should occupy a prominent position in the TREE, instead of representing a small side branch (as it does now). However, the relativistic effects, even if visible in chemistry, do not play an important role in the case of the light elements (covering almost the whole of organic chemistry as well as almost the whole of biology). This is why I have chosen a rather pragmatic ("non-fundamental") way of presentation. This chapter is mainly for those readers who are interested in:

- "the foundations of foundations"
- very accurate calculations for small atoms and molecules
- calculations for systems containing heavy elements

What is needed?

- The postulates of quantum mechanics (Chapter 1, necessary).
- Operator algebra (Appendix A, p. 889, necessary).
- Vector and scalar potentials (Appendix G, p. 962, necessary).

Classical works

In 1881 the American physicist Albert Michelson and in 1887 with Edward Morley carried out some experiments showing that the speed of light is the same in the directions perpendicular and parallel to the Earth's orbit, i.e. the Earth's orbital velocity did not change the speed of light with respect to the Earth. The results were published in the *American Journal of Science*, 22 (1881) 120 under the title "*The Relative Motion of the Earth and the Luminiferous Aether*", and *ibid.*, 34 (1887) 333 (with similar title). ★ In 1889 the Irish physicist George Francis FitzGerald made the conjecture that, if all moving objects were foreshortened in the direction of their motion, this would account for the strange results of the Michelson–Morley experiment. This was published in *Science*, 13 (1889) 390 with the title "*The Ether and the Earth's Atmosphere*". ★ The revolutionary special relativity theory (that explained this in detail) was developed by Albert Einstein in an article entitled "*Zur Elektrodynamik bewegter Körper*" published in *Annalen der Physik (Leipzig)*, 17 (1905) 891. ★ The article is based largely on the ideas of the Dutchman Hendrik Antoon Lorentz, who independently of FitzGerald[4] proposed the Lorentz transformation (of space and time) in 1904. The transformation accounted for the contraction of moving objects, as predicted by FitzGerald. The paper "*Electromagnetic Phenomena in a System Moving with any Velocity less than that of Light*" was published in *Proceedings of the Academy of Sciences of Amsterdam*, 6 (1904) 809. ★ The German mathematician Hermann Minkowski realized that the work of Lorentz and Einstein could best be understood using a non-Euclidean space of the space and time variables. His first paper on this subject was *"Die Grundgleichungen für die elektromagnetischen Vorgänge in bewegten Körper*" published in *Nachrichten der königlichen Gesellschaft der Wissenschaften zu Göttingen* (1908). ★ The Soviet physicist Vladimir A. Fock derived the first relativistic wave equation for a particle (published in *Zeitschrift für Physik*, 39 (1926) 226),

---

[4] It was pointed out to Lorentz in 1894 that FitzGerald had published something similar. He wrote to FitzGerald, but the latter replied that indeed he has sent a half-page article to *Science*, but he did not know "*whether they ever published it*". Afterwards Lorentz took every opportunity to stress that FitzGerald was first to present the idea.

then the German Walter Gordon did the same and also published in *Zeitschrift für Physik*, 40 (1926) 117. Finally, a similar theory was proposed independently by the Swede Oskar Klein in *Zeitschrift für Physik*, 41 (1927) 407. The Austrian Erwin Schrödinger also derived the same equation, and this is why it is sometimes called "*the equation with many fathers*". ★ A more advanced quantum mechanical theory (for a single particle) adapted to the principles of relativity was given by the British Paul Adrien Maurice Dirac in several articles in *Proceedings of the Royal Society (London)* entitled "*The Fundamental Equations of Quantum Mechanics*", A109 (1926) 642, "*Quantum Mechanics and a Preliminary Investigation of the Hydrogen Atom*", *ibid*., A110 (1926) 561, "*The Quantum Theory of Radiation*", *ibid*., A114 (1927) 243, "*The Quantum Theory of the Electron*", *ibid*., A117 (1928) 610, and "*The Quantum Theory of the Electron. Part II*" *ibid*., A118 (1928) 351. ★ An extension of relativistic quantum theory to many-electron problems (still approximate) was published by the American Gregory Breit in *Physical Review* with the title "*The Effect of Retardation on the Interaction of Two Electrons*", 34 (1929) 553, and then in two other papers entitled "*Fine Structure of He as a Test of the Spin Interaction of Two Electrons*", *ibid.*, 36 (1930) 383, and "*Dirac's Equation and the Spin–Spin Interactions of Two Electrons*", *ibid.*, 39 (1932) 616. ★ In 1948 the Americans Richard Feynman and Julian Schwinger as well as the Japanese Shinichiro Tomonaga independently invented the quantum electrodynamics (QED), which successfully combined quantum theory with the special theory of relativity and produced extremely accurate results.

3.1 A GLIMPSE OF CLASSICAL RELATIVITY THEORY

3.1.1 THE VANISHING OF APPARENT FORCES

The three principles of Newtonian[5] dynamics were taught to us in school. The first principle, that a free body (with no acting force) moves uniformly along a straight line, seems to be particularly simple. It was not so simple for Ernest Mach though.

Mach wondered how one recognizes that no force is acting on a body. The contemporary meaning of the first principle of Newton dynamics is the following. First, we introduce a Cartesian coordinate system x, y, z to the Universe, then remove all objects except one from the Universe, to avoid any interactions. Then, we measure equal time intervals using a spring clock and put the corresponding positions of the body in the coordinate system (we are thus there with our clock and our ruler...). The first principle says that the positions of the body are along a straight line and equidistant. What a crazy procedure! The doubts and dilemmas of Mach were implanted in the mind of Albert Einstein.

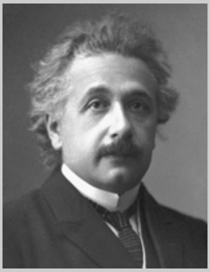
Ernest Mach (1838–1916), Austrian physicist and philosopher, professor at the Universities of Graz, Prague, and Vienna, godfather of Wolfgang Pauli. Mach investigated supersonic flows. In recognition of his achievements the velocity of sound in air (1224 km/hour) is called Mach 1.

---

[5] For Newton's biography see Chapter 7.

Albert Einstein (1879–1955) born in Ulm (Germany) studied at the ETH, Zurich. Considered by many as genius of all times. As a teenager and student, Einstein rejected many social conventions. This is why he was forced to begin his scientific career at a secondary position in the Federal Patent Office. Being afraid of his supervisor, he used to read books hidden in a drawer (he called it the "Department of Physics").

The year of his 26th birthday was particularly fruitful ("miraculous year" 1905). He published three fundamental papers: about relativity theory, about Brownian motion and about the photoelectric effect. For the last, Einstein received the Nobel Prize in 1921. After these publications he was appointed professor at the University of Zurich and then at the University of Prague. From 1914 Einstein headed the Physics Institute in Berlin, which was founded especially for him. He emigrated to the USA in 1933, because of menacing persecution, because of his Jewish origin. Einstein worked at the Institute for Advanced Study in Princeton in the USA. He died there in 1955. According to his will, his ashes were dispersed over America from the air.



This Bern Patent Office employee also knew about the dramatic dilemmas of Lorentz, which we will talk about in a moment. Einstein recalls that there was a clock at a tram stop in Bern. Whenever his tram moved away from the stop, the modest patent office clerk asked himself what would the clock show, if the tram had the velocity of light. While other passengers probably read their newspapers, Einstein had questions which led humanity on new pathways.

Let us imagine two coordinate systems (each in 1D): O "at rest" (we assume it inertial[6]) while the coordinate system O' moves with respect to the first in a certain way (possibly very complicated). The position of the moving point may be measured in O giving the number x as the result, while in O' on gets the result x'. These numbers are related one to another (f is a function of time t):

$$x' = x + f(t). \tag{3.1}$$

If a scientist working in a lab associated with the coordinate system O would like to calculate the force acting on the above mentioned point body, he would get a result proportional to the acceleration, i.e. to $\frac{d^2x}{dt^2}$. If the same were done by

---

[6] That is, in which the Newton equation is satisfied. A coordinate system associated with accelerating train is not inertial, because there is a non-zero force acting on everybody in the train, while the acceleration with respect to the train coordinate system is zero.

another scientist working in a lab in O', then he would obtain *another force*, this time proportional to the acceleration computed as $\frac{d^2x'}{dt^2} = \frac{d^2x}{dt^2} + \frac{d^2f}{dt^2}$. The second term in this force is the *apparent* force. Such apparent forces (from the point of view of an observer on the ground) are encountered in lifts, on a carousel, etc.

Let us note an important consequence: if one postulates the same forces (and therefore the same dynamics) in two coordinate systems, $f(t)$ has to be a *linear* function (because its second derivative is equal to zero). This means that a family of all coordinate systems that moved uniformly with respect to one another would be characterized by the same description of phenomena because the forces computed would be the same (*inertial systems*).

Physics textbooks written in the two laboratories associated to O and O' would describe all the phenomena in the same way.

inertial systems

The linearity condition gives $x' = x + vt$, if at $t = 0$ O and O' coincide. Let us take a fresh look at this equation: x' represents a linear combination of x and t, which means that *time and the linear coordinate mix together*. One has two coordinates: one in the O coordinate system and the other in the O' coordinate system. Wait a minute! Since the time and the coordinate are on an equal footing (they mix together), maybe one may also have the time (t) appropriate for (i.e. running in) the O and the time (t') running in the O' coordinate system?

Now, a crucial step in the reasoning. Let us write *in a most general way* a linear transformation of coordinates and time that ensures the two systems equivalent (no apparent forces):

$$x' = Ax + Bt,$$
$$t' = Cx + Dt.$$

First of all the corresponding transformation matrix *has to be* invertible (i.e. non-singular), because inversion simply means exchanging the roles of the two coordinate systems and of the observers flying with them. Thus, one has:

$$x = \bar{A}x' + \bar{B}t',$$
$$t = \bar{C}x' + \bar{D}t'.$$

Next, A has to be equal to $\bar{A}$, because the measurements of length in O and O', i.e. x and x', cannot depend on whether one looks at the O coordinate system from O', or at O' from O. If the opposite were true, then one of the coordinate systems would be privileged (treated in a special way). This, however, is impossible, because the two coordinate systems differ *only* in that O' flies from O with velocity

v, while O flies from O' with velocity $-v$, but the space is isotropic. The same *has to happen* with the time measurements: on board O, i.e. t, and on board O', i.e. t', therefore $D = \bar{D}$. Since (from the inverse transformation matrix) $\bar{A} = \frac{D}{AD-BC}$ and $\bar{D} = \frac{A}{AD-BC}$, therefore we have

$$\frac{D}{AD-BC} = A, \qquad \frac{A}{AD-BC} = D.$$

From this $\frac{D}{A} = \frac{A}{D}$ follows, or:

$$A^2 = D^2. \tag{3.2}$$

From the two solutions: $A = D$ and $A = -D$, one has to choose only $A = D$, because the second solution would mean that the times t and t' have opposite signs, i.e. when time run forwards in O it would run backwards in O'. Thus, we have

$$\boxed{A = D. \qquad (3.3)}$$

3.1.2 THE GALILEAN TRANSFORMATION

The equality condition $A = D$ is satisfied by the Galilean transformation, in which the two coefficients are equal to 1:

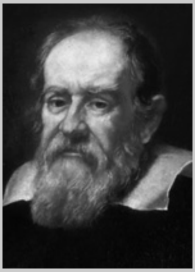
$$x' = x - vt,$$
$$t' = t,$$

where position x and time t, say, of a passenger in a train, is measured in a platform-fixed coordinate system, while x' and t' are measured in a train-fixed coordinate system. There are no apparent forces in the two coordinate systems related by the Galilean transformation. Also, the Newtonian equation is consistent with our intuition, saying that time flows at the same pace in any coordinate system.

3.1.3 THE MICHELSON–MORLEY EXPERIMENT

Hendrik Lorentz indicated that the Galilean transformation represents only *one possibility* of making the apparent forces vanish, i.e. assuring that $A = D$. Both constants need not be equal to 1. As it happens that such a generalization is forced by an intriguing experiment performed in 1887.

Michelson and Morley were interested in whether the speed of light differs, when measured in two laboratories moving with respect to one another. According to the Galilean transformation, the two velocities of light should be different, in the same way as the speed of train passengers (measured with respect to the platform)



Galileo Galilei (1564–1642), Italian scientist, professor of mathematics at the University of Pisa. Only those who have visited Pisa are able to appreciate the inspiration (for studying the free fall of bodies of different materials) from the *incredibly* leaning tower. Galileo's *opus magnum* (right-hand side) has been published by Elsevier in 1638. Portrait by Justus Sustermans (XVII century).

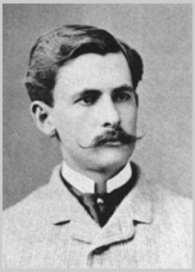
DISCORSI
E
DIMOSTRAZIONI
MATEMATICHE,
intorno à due nuoue ſcienze
Attenenti alla
MECANICA & I MOVIMENTI LOCALI,
del Signor
GALILEO GALILEI LINCEO,
Filoſofo e Matematico primario del Sereniſſimo
Grand Duca di Toſcana.
Con vna Appendice del centro di grauità d'alcuni Solidi.

IN LEIDA,
Appreſſo gli Elſevirii. M. D. C. XXXVIII.

Hendrik Lorentz (1853–1928), Dutch scientist, professor at Leiden. Lorentz was very close to formulating the special theory of relativity.

Albert Michelson (1852–1931), American physicist, professor in Cleveland and Chicago, USA. He specialized in the precise measurements of the speed of light.

His older colleague Edward Williams Morley was American physicist and chemist, professor of chemistry at Western Reserve University in Cleveland, USA.



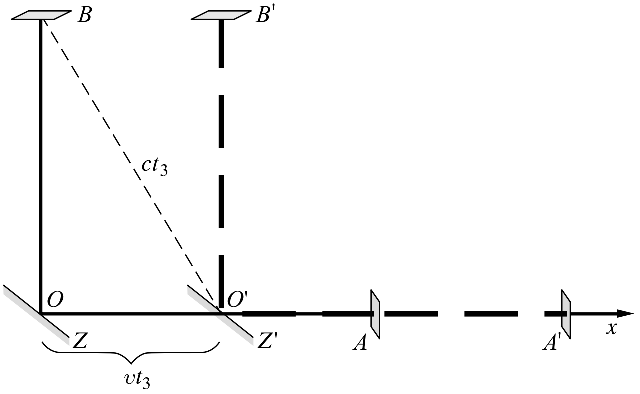


Fig. 3.1. The Michelson–Morley experimental framework. We have two identical V-shaped right-angle objects, each associated with a Cartesian coordinate system (with origins O and O'). The first is at rest, while the second moves with velocity v with respect to the first (along coordinate x). We are going to measure the velocity of light in two laboratories rigidly bound to the two coordinate systems. The mirrors are at the ends of the objects: A, B in O and A', B' in O', while at the origins two semi-transparent mirrors Z and Z' are installed. Time $2t_3 \equiv t_\downarrow$ is the time for light to go down and up the vertical arm.

differs depending on whether they walk in the same or the opposite direction with respect to the train motion. Michelson and Morley replaced the train by Earth, which moves along its orbit around the Sun with a speed of about 40 km/s. Fig. 3.1 shows the Michelson–Morley experimental framework schematically. Let us imagine two identical right-angle V-shaped objects with all the arm lengths equal to L.

Each of the objects has a semi-transparent mirror at its vertex,[7] and ordinary mirrors at the ends. We will be interested in how much time it takes the light to travel along the arms of our objects (back and forth). One of the two arms of any object is oriented along the x axis, while the other one must be orthogonal to it. The mirror system enables us to overlap the light beam from the horizontal arm (x axis) with the light beam from the perpendicular arm. If there were any difference in phase between them we would immediately see the interference pattern.[8] The second object moves along x with velocity v (and is associated with coordinate system O') with respect to the first ("at rest", associated with coordinate system O).

3.1.4 THE GALILEAN TRANSFORMATION CRASHES

In the following we will *suppose that the Galilean transformation is true.* In coordinate system O the time required for light to travel (round-trip) the arm along the

---

[7] Such a mirror is made by covering glass with a silver coating.

[8] From my own experience I know that interference measurement is very sensitive. A laser installation was fixed to a steel table 10 cm thick concreted into the foundations of the Chemistry Department building, and the interference pattern was seen on the wall. My son Peter (then five-years-old) just touched the table with his finger. Everybody could see immediately a large change in the pattern, because the table bent.

x axis ($T_{\rightarrow}$) and that required to go perpendicularly to axis ($T_{\downarrow}$) are the same:

$$T_{\rightarrow} = \frac{2L}{c}, \qquad T_{\downarrow} = \frac{2L}{c}.$$

Thus, in the O coordinate system, there will be no phase difference between the two beams (one coming from the parallel, the other from the perpendicular arm) and therefore no interference will be observed. Let us consider now a similar measurement in O'. In the arm co-linear with x, when light goes in the direction of $\boldsymbol{v}$, it has to take more time (t_1) to get to the end of the arm:

$$ct_1 = L + vt_1, \tag{3.4}$$

than the time required to come back (t_2) along the arm:

$$ct_2 = L - vt_2. \tag{3.5}$$

Thus, the total round-trip time $t_{\rightarrow}$ is[9]

$$t_{\rightarrow} = t_1 + t_2 = \frac{L}{c-v} + \frac{L}{c+v} = \frac{L(c+v)+L(c-v)}{(c-v)(c+v)} = \frac{2Lc}{c^2-v^2} = \frac{\frac{2L}{c}}{1-\frac{v^2}{c^2}}. \tag{3.6}$$

What about the perpendicular arm in the coordinate system O'? In this case the time for light to go down (t_3) and up will be the same (let us denote total flight time by $t_{\downarrow} = 2t_3$, Fig. 3.1). Light going down goes along the hypotenuse of the rectangular triangle with sides: L and $\frac{vt_{\downarrow}}{2}$ (because it goes down, but not only, since after $\frac{t_{\downarrow}}{2}$ it is found at $x = \frac{vt_{\downarrow}}{2}$). We will find, therefore, the time $t_{\downarrow}$ from Pythagoras' theorem:

$$\left(c\frac{t_{\downarrow}}{2}\right)^2 = L^2 + \left(v\frac{t_{\downarrow}}{2}\right)^2, \tag{3.7}$$

or

$$t_{\downarrow} = \sqrt{\frac{4L^2}{c^2-v^2}} = \frac{2L}{\sqrt{c^2-v^2}} = \frac{\frac{2L}{c}}{\sqrt{1-\frac{v^2}{c^2}}}. \tag{3.8}$$

The times $t_{\downarrow}$ and $t_{\rightarrow}$ do not equal each other for the moving system and there will be the interference, we were talking about a little earlier.

However, there is absolutely no interference!
Lorentz was forced to put the Galilean transformation into doubt
(apparently the foundation of the whole science).

---

[9] Those who have some experience with relativity theory, will certainly recognize the characteristic term $1 - \frac{v^2}{c^2}$.

3.1.5 THE LORENTZ TRANSFORMATION

The interference predicted by the Galilean transformation is impossible, because physical phenomena would experience the two systems in a different way, while they differ only by their relative motions ($\boldsymbol{v}$ has to be replaced by $-\boldsymbol{v}$).

> To have everything back in order, Lorentz assumed that, when a body moves, its length (measured by using the unit length at rest in the coordinate system O) along the direction of the motion, *contracts* according to equation
>
> $$l = L\sqrt{1-\frac{v^2}{c^2}}. \tag{3.9}$$

length contraction

If we insert such a length l, instead of L, in the expression for $t_{\rightarrow}$, then we obtain

$$t_{\rightarrow} = \frac{\frac{2l}{c}}{1-\frac{v^2}{c^2}} = \frac{\frac{2L\sqrt{1-\frac{v^2}{c^2}}}{c}}{1-\frac{v^2}{c^2}} = \frac{\frac{2L}{c}}{\sqrt{1-\frac{v^2}{c^2}}} \tag{3.10}$$

and everything is perfect again: $t_{\downarrow} = t_{\rightarrow}$. No interference. This means that x' (i.e. the position of a point belonging to a rigid body as measured in O') and x (the position of the same point measured in O) have to be related by the following formula. The coordinate x measured by an observer in his O is composed of the intersystem distance OO', i.e. vt plus the distance O' – point, but measured using the length unit of the observer in O, i.e. the unit that resides in O (thus, *non-contracted* by the motion). Because of the contraction 1: $\sqrt{1-\frac{v^2}{c^2}}$ of the rigid body the latter result will be *smaller* than x' (recall, please, that x' is what the observer measuring the position in his O' obtains), hence:

$$x = x'\sqrt{1-\frac{v^2}{c^2}} + vt, \tag{3.11}$$

or:

$$x' = \frac{x}{\sqrt{1-\frac{v^2}{c^2}}} - \frac{vt}{\sqrt{1-\frac{v^2}{c^2}}}, \tag{3.12}$$

which means that in the linear transformation

$$A = \frac{1}{\sqrt{1-\frac{v^2}{c^2}}}, \tag{3.13}$$

$$B = -\frac{v}{\sqrt{1-\frac{v^2}{c^2}}}. \tag{3.14}$$

As we have already shown, in linear transformation $(x', t') \to (x, t)$ the diagonal coefficients have to be equal $(A = D)$, therefore

$$t' = Cx + Dt, \tag{3.15}$$

$$D = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}}. \tag{3.16}$$

To complete determination of the linear transformation we have to calculate the constant C. Albert Einstein assumed, that if Professors Oconnor and O'connor began (in their own coordinate systems O and O') measurements on the velocity of light, then despite the different distances gone (x and x') and different flight times[10] (t and t'), both scientists would get *the same velocity of light* (denoted by c).

In other words $x = ct$ and $x' = ct'$.

Using this assumption and eqs. (3.12) and (3.16) we obtain:

$$ct' = Dct - vDt, \tag{3.17}$$

while multiplying equation (3.15) for t' by c we get:

$$ct' = cCx + Dct. \tag{3.18}$$

Subtracting both equations we have

$$0 = -vDt - cCx \tag{3.19}$$

or

$$C = -\frac{vtD}{cx} = -\frac{vtD}{cct} = -\frac{vD}{c^2}. \tag{3.20}$$

Thus we obtain the full Lorentz transformation, which assures that no of the systems is privileged, and the same speed of light in *both* systems:

$$x' = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}} x - \frac{v}{\sqrt{1 - \frac{v^2}{c^2}}} t,$$

$$t' = -\frac{v}{c^2} \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}} x + \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}} t.$$

---

[10] At the moment of separation $t = t' = 0$.

Let us check first of all, whether if $v = 0$, then everything is OK. Yes it is. Indeed, the denominator equals 1 and we have $t' = t$ and $x' = x$. Let us see what would happen if the velocity of light were equal to infinity. Then, the Lorentz transformation becomes identical with the Galilean. In general, after expanding t' and x' in a power series of v^2/c^2 we obtain

$$x' = -vt + x + \frac{1}{2}(-vt + x)\frac{v^2}{c^2} + \cdots,$$

$$t' = t + \left(-\frac{x}{v} + \frac{t}{2}\right)\frac{v^2}{c^2} + \cdots$$

This means that only at very high velocity v, may we expect differences between both transformations.

Contraction is relative

Of course, Professor O'connor in his laboratory O' would not believe in Professor Oconnor (sitting in his O lab) saying that he (O'connor) has a contraction of the rigid body. And indeed, if Professor O'connor measured the rigid body using his standard length unit (he would not know his unit is contracted), then the length measured would be exactly the same as that measured just before separation of the two systems, when both systems were at rest. In a kind of retaliation, Professor O'connor could say (smiling) that it is certainly not him who has the contraction, but his colleague Oconnor. He would be right, because for him, his system is at rest and his colleague Oconnor flies away from him with velocity $-v$. Indeed, our formula (3.11) makes that very clear: expressing in (3.11) t by t' from the Lorentz transformation leads to the point of view of Professor O'connor

$$x' = x\sqrt{1 - \frac{v^2}{c^2}} - vt', \tag{3.21}$$

and one can indeed see an evident contraction of the rigid body of Professor Oconnor. This way, neither of these two coordinate systems is privileged. That is very, very good.

3.1.6 NEW LAW OF ADDING VELOCITIES

Our intuition was worked out for small velocities, much smaller than the velocity of light. The Lorentz transformation teaches us something, which is against intuition. What does it mean that the velocity of light is constant? Suppose we are flying with the velocity of light and send the light in the direction of our motion. Our intuition tells us: the light will have the velocity equal to $2c$. Our intuition has to be wrong. How it will happen?

Let us see. We would like to have the velocity in the coordinate system O, but first let us find the velocity in the coordinate system O', i.e. $\frac{\mathrm{d}x'}{\mathrm{d}t'}$. From the Lorentz transformation one obtains step by step:

$$\frac{\mathrm{d}x'}{\mathrm{d}t'} = \frac{\frac{1}{\sqrt{1-\frac{v^2}{c^2}}}\,\mathrm{d}x - \frac{v}{\sqrt{1-\frac{v^2}{c^2}}}\,\mathrm{d}t}{-\frac{v}{c^2}\frac{1}{\sqrt{1-\frac{v^2}{c^2}}}\,\mathrm{d}x + \frac{1}{\sqrt{1-\frac{v^2}{c^2}}}\,\mathrm{d}t} = \frac{\frac{\mathrm{d}x}{\mathrm{d}t} - v}{1 - \frac{v}{c^2}\frac{\mathrm{d}x}{\mathrm{d}t}}. \tag{3.22}$$

By extracting $\frac{\mathrm{d}x}{\mathrm{d}t}$ or using the symmetry relation (when $O' \to O$, then $v \to -v$) we obtain:

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \frac{\frac{\mathrm{d}x'}{\mathrm{d}t'} + v}{1 + \frac{v}{c^2}\frac{\mathrm{d}x'}{\mathrm{d}t'}} \tag{3.23}$$

or

VELOCITY ADDITION LAW

$$V = \frac{v' + v}{1 + \frac{vv'}{c^2}}. \tag{3.24}$$

In this way we have obtained a new rule of adding the velocities of the train and its passenger. Everybody naively thought that if the train velocity is v and, the passenger velocity with respect to the train corridor is v', then the velocity of the passenger with respect to the platform is $V = v + v'$. It turned out that this is not true. On the other hand when both velocities are small with respect to c, then indeed one restores the old rule

$$V = v' + v. \tag{3.25}$$

Now, let us try to fool Mother Nature. Suppose our train is running with the velocity of light, i.e. $v = c$, and we take out a torch and shine the light forward, i.e. $\frac{\mathrm{d}x'}{\mathrm{d}t'} = v' = c$. What will happen? What will be the velocity V of the light with respect to the platform? $2c$? From (3.24) we have $V = \frac{2c}{2} = c$. This is precisely what is called the universality of the speed of light. Now, let us make a bargain with Nature. We are hurtling in the train with the speed of light $v = c$ and walking along the corridor with velocity $v' = 5$ km/h. What will our velocity be with respect to the platform? Let us calculate again:

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \frac{5 + c}{1 + \frac{c}{c^2}5} = \frac{5 + c}{1 + \frac{5}{c}} = c\frac{5 + c}{5 + c} = c. \tag{3.26}$$

Once more we have been unable to exceed the speed of light c. One last attempt. Let us take the train velocity as $v = 0.95c$, and fire along the corridor a powerful missile with speed $v' = 0.10c$. Will the missile exceed the speed of light or not? We have

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \frac{0.10c + 0.95c}{1 + \frac{0.95c}{c^2}0.10c} = \frac{1.05c}{1 + 0.095} = \frac{1.05}{1.095}c = 0.9589c. \tag{3.27}$$

c is not exceeded. Wonderful formula.

3.1.7 THE MINKOWSKI SPACE-TIME CONTINUUM

The Lorentz transformation may also be written as:

$$\begin{bmatrix} x' \\ ct' \end{bmatrix} = \frac{1}{\sqrt{1-\frac{v^2}{c^2}}} \begin{bmatrix} 1 & -\frac{v}{c} \\ -\frac{v}{c} & 1 \end{bmatrix} \begin{bmatrix} x \\ ct \end{bmatrix}.$$

What would happen if the roles of the two systems were interchanged? To this end let us express x, t by x', t'. By inversion of the transformation matrix we obtain[11]

$$\begin{bmatrix} x \\ ct \end{bmatrix} = \frac{1}{\sqrt{1-\frac{v^2}{c^2}}} \begin{bmatrix} 1 & \frac{v}{c} \\ \frac{v}{c} & 1 \end{bmatrix} \begin{bmatrix} x' \\ ct' \end{bmatrix}. \tag{3.28}$$

relativity principle

We have perfect symmetry, because it is clear that the sign of the velocity has to change. Therefore:

none of the systems is privileged (relativity principle).

Now let us come back to Einstein's morning tram meditation[12] about what he would see on the tramstop clock if the tram had the velocity of light. Now we have the tools to solve the problem. It concerns the two events – two ticks of the clock observed in the coordinate system associated with the tramstop, i.e. $x_1 = x_2 \equiv x$, but happening at two different times t_1 and t_2 (differing by, say, one second, i.e. $t_2 - t_1 = 1$, this is associated with the corresponding movement of the clock hand).

---

[11] You may check this by multiplying the matrices of both transformations – we obtain the unit matrix.

[12] Even today Bern looks quite provincial. In the centre Albert Einstein lived at Kramgasse 49. A small house, squeezed by others, next to a small café, with Einstein's achievements on the walls. Einstein's small apartment is on the second floor showing a room facing the backyard, in the middle a child's room (Einstein lived there with his wife Mileva Marić and their son Hans Albert; the personal life of Einstein is complicated), and a large living room facing the street. A museum employee with oriental features says the apartment looks as it did in the "miraculous year 1905", everything is the same (except the wall-paper, she adds), and then: "*maybe this is the most important place for the history of science*".

What will Einstein see when his tram leaves the stop with velocity v with respect to the stop, or in other words when the tramstop moves with respect to him with velocity $-v$? He will see the same two events, but in his coordinate system they will happen at

$$t_1' = \frac{t_1}{\sqrt{1-\frac{v^2}{c^2}}} - \frac{\frac{v}{c^2}x}{\sqrt{1-\frac{v^2}{c^2}}} \quad \text{and} \quad t_2' = \frac{t_2}{\sqrt{1-\frac{v^2}{c^2}}} - \frac{\frac{v}{c^2}x}{\sqrt{1-\frac{v^2}{c^2}}},$$

i.e. *according to the tram passenger* the two ticks at the tramstop will be separated by the time interval

$$t_2' - t_1' = \frac{t_2 - t_1}{\sqrt{1-\frac{v^2}{c^2}}} = \frac{1}{\sqrt{1-\frac{v^2}{c^2}}}.$$

Thus, when the tram ran through the streets of Bern with velocity $v = c$, the hands on the tramstop clock when seen from the tram would not move at all, and this second would be equivalent to eternity.

This is known as *time dilation*. Of course, for the passengers waiting at the tramstop (for the next tram) and watching the clock, its two ticks would be separated by exactly one second. If Einstein took his watch out of his waistcoat pocket and showed it to them through the window *they* would be amazed. The seconds will pass at the tramstop, while Einstein's watch would seem to be stopped. The effect we are describing has been double checked experimentally many times. For example, the meson lives such a short time (in the coordinate system associated with it), that when created by cosmic rays in the stratosphere, it would have no chance of reaching a surface laboratory before decaying. Nevertheless, as seen from the laboratory coordinate system, the meson's clock ticks very slowly and mesons are observable. time dilation

Hermann Minkowski introduced the seminal concept of the four-dimensional space-time continuum (x, y, z, ct).[13] In our one-dimensional space, the elements of the Minkowski space-time continuum are events, i.e. vectors (x, ct), something happens at space coordinate x at time t, when the event is observed from coordinate system O. When the same event is observed in two coordinate sys-

Hermann Minkowski (1864–1909), German mathematician and physicist, professor in Bonn, Königsberg, Technische Hochschule Zurich, and from 1902 professor at the University of Göttingen.

---

[13] Let me report a telephone conversation between the PhD student Richard Feynman and his supervisor Prof. Archibald Wheeler from Princeton Advanced Study Institute (according to Feynman's Nobel

tems, then the corresponding x, t and x', t' satisfy the Lorentz transformation. It turns out that in both coordinate systems *the distance of the event from the origin of the coordinate system is preserved*. The square of the distance is calculated in a strange way as:

$$(ct)^2 - x^2 \tag{3.29}$$

for the event (x, ct). Indeed, let us check carefully:

$$\begin{aligned}(ct')^2 - (x')^2 &= \frac{1}{1-\frac{v^2}{c^2}}\left(-\frac{v}{c}x + ct\right)^2 - \frac{1}{1-\frac{v^2}{c^2}}\left(x - \frac{v}{c}ct\right)^2 \\ &= \frac{1}{1-\frac{v^2}{c^2}}\left[\frac{v^2}{c^2}x^2 + c^2t^2 - 2vxt - x^2 - \frac{v^2}{c^2}c^2t^2 + 2vxt\right] \\ &= \frac{1}{1-\frac{v^2}{c^2}}\left[\frac{v^2}{c^2}x^2 + c^2t^2 - x^2 - \frac{v^2}{c^2}c^2t^2\right] = (ct)^2 - (x)^2. \end{aligned} \tag{3.30}$$

There it is! This equation enabled Hermann Minkowski to interpret the Lorentz transformation (3.28) as a *rotation* of the event (x, ct) in the Minkowski space about the origin of the coordinate system (since any rotation preserves the distance from the rotation axis).

3.1.8 HOW DO WE GET $E = mc^2$?

The Schrödinger equation is invariant with respect to the Galilean transformation. Indeed, the Hamiltonian contains the potential energy which depends on interparticle distances, i.e. on the differences of the coordinates, whereas the kinetic energy operator contains the second derivative operators which are invariant with respect to the Galilean transformation. Also, since $t = t'$, the time derivative in the time-dependent Schrödinger equation does not change.

Unfortunately, both Schrödinger equations (time-independent and time-dependent) are not invariant with respect to the Lorentz transformation and therefore are illegal.

As a result, one cannot expect the Schrödinger equation to describe accurately objects that move with velocities comparable to the speed of light.

Lecture, 1965): "Wheeler: "*Feynman, I know why all electrons have the same charge and the same mass!*" F: "*Why*?" W: "*Because they are all the same electron!*" Then, Wheeler explained: "*suppose that the world lines which we were ordinarily considering before in time and space – instead of only going up in time were a tremendous knot, and then, when we cut through the knot by the plane corresponding to a fixed time, we would see many, many world lines and that would represent many electrons (...)*".

Let us consider a particle moving in the potential V. The Schrödinger equation has been "derived" (see p. 70) from the total energy expression:

$$E = \frac{p^2}{2m} + V, \tag{3.31}$$

where $\boldsymbol{p}$ is the momentum vector and m is the mass.

Einstein was convinced that nothing could be faster than light.[14] Therefore, what would happen if a particle were subject to a constant force? It would eventually attain the velocity of light and what would happen afterwards? There was a problem, and Einstein assumed that in the laboratory coordinate system in which the particle is speeded up, the particle will increase its … mass. In the coordinate system fixed on the particle no mass increase will be observed, but in the laboratory system it will. We have to admire Einstein's courage. For millions of people the mass of a body represented an invariant characteristic of the body. How was the mass supposed to increase? Well, the increase law – Einstein reasoned – should be such that the particle was able to absorb *any* amount of the kinetic energy. This means that when $v \to c$, then we have to have $m(v) \to \infty$. One of the possible formulae for $m(v)$ may contain a factor typical of relativity theory [cf. eq. (3.16)]:

$$m(v) = \frac{m_0}{\sqrt{1 - \frac{v^2}{c^2}}}, \tag{3.32}$$

where m_0 is the so called rest mass of the particle (i.e. its mass measured in the coordinate system residing on the particle).[15] It is seen that if v/c were zero (as it is in the non-relativistic world), then m would be equal to m_0, i.e. to a constant as it is in non-relativistic physics.[16]

For the time being, the legitimacy of eq. (3.32) is questionable as being just one of the possible *ad hoc* suppositions. However, Einstein has shown that this

---

[14] Maybe this is true, but nothing in the special theory of relativity compels us to think that c is the maximum speed possible.

[15] Because of the speed-dependent mass in relativity theory it is impossible to separate the centre-of-mass motion.

[16] And therefore no corrections to the Schrödinger equation are needed. At the beginning of this chapter we arrived at the conclusion that electron velocity in an atom is close to its atomic number Z (in a.u.). Hence, for the hydrogen atom ($Z_H = 1$) one may estimate $v/c \simeq 0.7\%$, i.e. v of the electron in the 1s state represents a velocity of the order of 2100 km/s, which is probably very impressive for a car driver, but not for an electron. However, for gold ($Z_{\mathrm{Au}} = 79$) we obtain $v/c \simeq 51\%$. This means that in the atom of gold the electron mass is larger by about 15% with respect to its rest mass and therefore the relativistic effects are non-negligible. For such important elements as C, O, N (biology) the relativistic corrections may be safely neglected. A young graduate student, Grzegorz Łach, posed an interesting purely academic question (such questions and the freedom to discuss them represent the cornerstone and the beauty of university life): will the human body survive the switching off of relativistic effects? Most of the biomolecules would function practically without significant changes, but the heavy metal atoms in enzyme active sites might react differently in the chemical reactions in which they are involved. Would they indeed? Would the new direction be destructive for the body? Nobody knows. On the other hand, we have forgotten about the spin concept that follows in a consequent way only from relativistic quantum theory (see below). Without spin no world similar to ours is conceivable.

particular formula fits the existing equation of motion. First, after expanding the mass into the Taylor series one obtains something interesting

$$m(v) = m_0\left\{1 + \frac{1}{2}\frac{v^2}{c^2} + \frac{3}{8}\frac{v^4}{c^4} + \cdots\right\}, \tag{3.33}$$

kinetic energy

especially after multiplying the result by c^2:

$$mc^2 - m_0c^2 = \frac{m_0v^2}{2} + \text{smaller terms.} \tag{3.34}$$

It looks as if indeed the kinetic energy was stored directly in the mass m. Einstein deduced that it may be that the total kinetic energy of the body is equal to:

$$\boxed{E = mc^2.}$$

He convinced himself about this after calculating its time derivative. After assuming that eq. (3.32) is correct, one obtains:

$$\begin{aligned}
\frac{\mathrm{d}E}{\mathrm{d}t} &= c^2\frac{\mathrm{d}m}{\mathrm{d}t} = c^2\frac{\mathrm{d}}{\mathrm{d}t}\frac{m_0}{\sqrt{1-\frac{v^2}{c^2}}} = m_0c^2\frac{\mathrm{d}}{\mathrm{d}t}\frac{1}{\sqrt{1-\frac{v^2}{c^2}}}\\
&= m_0c^2\left(-\frac{1}{2}\right)\left(1-\frac{v^2}{c^2}\right)^{-\frac{3}{2}}\frac{-2v}{c^2}\frac{\mathrm{d}v}{\mathrm{d}t} = m_0\left(1-\frac{v^2}{c^2}\right)^{-\frac{3}{2}} v\frac{\mathrm{d}v}{\mathrm{d}t}\\
&= \frac{m_0}{\sqrt{(1-\frac{v^2}{c^2})}}\frac{1}{1-\frac{v^2}{c^2}}v\frac{\mathrm{d}v}{\mathrm{d}t} = \frac{m_0}{\sqrt{(1-\frac{v^2}{c^2})}}\left(1+\frac{\frac{v^2}{c^2}}{1-\frac{v^2}{c^2}}\right)v\frac{\mathrm{d}v}{\mathrm{d}t}\\
&= \frac{m_0}{\sqrt{(1-\frac{v^2}{c^2})}}v\frac{\mathrm{d}v}{\mathrm{d}t} + \frac{v^2}{c^2}m_0\left(1-\frac{v^2}{c^2}\right)^{-\frac{3}{2}}v\frac{\mathrm{d}v}{\mathrm{d}t} = mv\frac{\mathrm{d}v}{\mathrm{d}t} + v^2\frac{\mathrm{d}m}{\mathrm{d}t} = v\frac{\mathrm{d}(mv)}{\mathrm{d}t}.
\end{aligned}$$

Precisely the same equation is satisfied in non-relativistic mechanics, if E denotes the *kinetic energy*:

$$\frac{\mathrm{d}E}{\mathrm{d}t} = \frac{\mathrm{d}(\frac{mv^2}{2})}{\mathrm{d}t} = mv\frac{\mathrm{d}v}{\mathrm{d}t} = v\frac{\mathrm{d}(mv)}{\mathrm{d}t}. \tag{3.35}$$

Therefore in relativity theory

$$E_{\mathrm{kin}} = mc^2. \tag{3.36}$$

This formula has been verified in laboratories many times. For example, it is possible nowadays to speed electrons in cyclotrons up to a velocity that differs from c by $\frac{1}{8000000}c$. That corresponds to $1-\frac{v^2}{c^2}=\frac{1}{4000000}$, and the electron's mass m becomes 2000 times larger than its m_0. This means that the electron is pumped up with energy to such an extent that its mass is similar to that of the proton. The energy stored in mass is huge. If, from the mass of a 20000 TNT atomic bomb, one subtracted the mass of its ashes after explosion,[17] then one would obtain only about 1g! The energy freed from this 1g gives an effect similar to the apocalypse.

3.2 RECONCILING RELATIVITY AND QUANTUM MECHANICS

"The equation with many fathers" (Klein–Gordon, also Fock and Schrödinger...)

We would like to express the kinetic energy E_{kin} through the particle's momentum p, because we would then know how to obtain the corresponding quantum mechanical operators (Chapter 1, p. 18). To this end let us consider the expression

$$\begin{aligned}E_{\text{kin}}^2-\left(m_0c^2\right)^2&=m^2c^4-m_0^2c^4=m_0^2c^4\left(\frac{1}{1-v^2/c^2}-1\right)\\&=\frac{m_0^2c^4}{1-v^2/c^2}\frac{v^2}{c^2}=m^2v^2c^2=p^2c^2.\end{aligned}\tag{3.37}$$

Therefore,

$$E_{\text{kin}}=c\sqrt{p^2+m_0^2c^2}\tag{3.38}$$

and the total energy E in the external potential V

$$E=c\sqrt{p^2+m_0^2c^2}+V.\tag{3.39}$$

What if the particle is subject to an electromagnetic field, given by the electric field $\boldsymbol{\mathcal{E}}$ and the magnetic field $\boldsymbol{H}$ (or, the magnetic induction $\boldsymbol{B}$) in every point of the space? Instead of $\boldsymbol{\mathcal{E}}$ and $\boldsymbol{H}$ (or $\boldsymbol{B}$) we may introduce two other quantities: the vector field $\boldsymbol{A}$ and the scalar field ϕ (see Appendix G). As we can show in classical electrodynamics,[18] the kinetic energy of the particle subject to an electromagnetic field is very similar to the same expression without the field (eq. (3.38)), namely, for a particle of charge q, the momentum $\boldsymbol{p}$ is to be replaced by $\boldsymbol{p}-\frac{q}{c}\boldsymbol{A}$ and the potential V by $q\phi$. Therefore, we obtain the following expression for the total

---

[17] R. Feynman, R.B. Leighton, M. Sands, "*Feynman Lectures on Physics*", Addison-Wesley Publishing Company, 1964.

[18] For example, H.F. Hameka, *Advanced Quantum Chemistry*, Addison-Wesley, Reading, MA, 1965, p. 40.

energy of the particle in an electromagnetic field

$$E = c\sqrt{\left(\boldsymbol{p} - \frac{q}{c}\boldsymbol{A}\right)^2 + m_0^2c^2} + q\phi, \tag{3.40}$$

where $\boldsymbol{A}$ and ϕ represent functions of the particle's position.

If we wanted to use the last expression to construct the Hamiltonian, then we would find serious difficulty, namely, the momentum operator $\hat{\boldsymbol{p}} = -i\hbar\nabla$ (replacing $\boldsymbol{p}$ according to the appropriate postulate, Chapter 1) is under the square root sign, thus leading to non-linear operators. Brave scientists noted, however, that if you squared both sides, the danger would disappear. We would obtain

$$(E - q\phi)^2 = c^2\left[\left(\boldsymbol{p} - \frac{q}{c}\boldsymbol{A}\right)^2 + m_0^2c^2\right]. \tag{3.41}$$

All this has been, and still is, a sort of groping and guessing from some traces or indications.

> The equations corresponding to physical quantities will be transformed to the corresponding operator equations, and it will be assumed that both sides of them will act on a wavefunction.

What to insert as the operator $\hat{H}$ of the energy E? This was done by Schrödinger (even before Fock, Klein and Gordon). Schrödinger inserted what he had on the right-hand side of his time-dependent equation

$$\hat{H}\Psi = i\hbar\frac{\partial}{\partial t}\Psi, \quad \text{i.e.} \quad \hat{H} = i\hbar\frac{\partial}{\partial t}.$$

Oskar Klein (1894–1977) was the youngest son of the chief rabi of Sweden and professor of mathematics and physics at Stockholm Högskola. Walter Gordon (1893–1940) until 1933 was a professor at the University of Hamburg, and afterwards resided in Sweden.

This way

$$\left(i\hbar\frac{\partial}{\partial t} - q\phi\right)^2 = c^2\left[\left(-i\hbar\nabla - \frac{q}{c}\boldsymbol{A}\right)^2 + m_0^2c^2\right], \tag{3.42}$$

or after acting on the wave function we obtain the equation known as Klein–Gordon:

$$\left(i\hbar\frac{\partial}{\partial t} - q\phi\right)^2\Psi = c^2\left[\left(-i\hbar\nabla - \frac{q}{c}\boldsymbol{A}\right)^2 + m_0^2c^2\right]\Psi. \tag{3.43}$$

This equation has at least one advantage over the Schrödinger equation: ct and x, y, z enter the equation on an equal footing, which is required by special relativity. Moreover, the Klein–Gordon equation is invariant with respect to the Lorentz transformation, whereas the Schrödinger equation is not. This is a prerequisite of any relativity-consistent theory and it is remarkable that such a simple derivation made the theory invariant. The invariance does not however mean that the equation is accurate. The Klein–Gordon equation describes a boson particle.

3.3 THE DIRAC EQUATION

3.3.1 THE DIRAC ELECTRONIC SEA

Paul Dirac used the Klein–Gordon equation to derive a Lorentz transformation invariant equation[19] for a single fermion particle. The Dirac equation is solvable only for several very simple cases. One of them is the free particle (Dirac), the other is an electron in the electrostatic field of a nucleus (Darwin, not *that* one).

One may add here a few other systems, e.g., the harmonic oscillator and that's it.

From eq. (3.38), in the case of a free particle $V = 0$, one obtains two sets of energy eigenvalues, one corresponding to the negative energies

$$E = -\sqrt{p^2c^2 + m_0^2c^4} \tag{3.44}$$

and the other corresponding to the positive energies

$$E = +\sqrt{p^2c^2 + m_0^2c^4}. \tag{3.45}$$

Dirac was not worried by the fact that both roots appear after an *ad hoc* decision to square the expression for the energy (eqs. (3.40) and (3.41)). As we can see,

Paul Adrien Maurice Dirac (1902–1984), British physicist theoretician, professor at universities in Cambridge, and then Oxford. Dirac was keen on hiking and climbing. He used to practise before expeditions by climbing trees near Cambridge, in the black outfit in which always gave his lectures.

He spent his last years at the University of Tallahassee (Florida, USA). On being guided through New York City, Dirac remembered old times. The guide remarked that there were visible changes, among others the buses had been painted pink. Dirac quietly agreed, adding that indeed they had, at least from one side...

---

[19] See J.D. Bjorken, S.D. Drell, "*Relativistic Quantum Mechanics*", McGraw-Hill, 1964.

Charles Galton Darwin (1887–1962), British physicist and mathematician, professor at University of Edinburgh, Scotland, grandson of the evolutionist Sir Charles Robert Darwin. Darwin investigated the scattering of α particles on atoms.

since the momentum may change from 0 to ∞ ($p = mv$, and for $v \to c$, we have $m \to \infty$), we therefore have the *negative energy continuum* and symmetrically located *positive energy continuum*, both continua separated by the energy gap $2m_0c^2$ (Fig. 3.2).

Dirac (when 26 years old) made the apparently absurd assumption that what people call a vacuum is in reality a sea of electrons occupying the negative energy continuum ("*Dirac electronic sea*"). The sea was supposed to consist of an infinite number of electrons, which had to imply catastrophic consequences concerning, for example, the mass of the Universe (infinite), but Dirac did not doubt or discourage: "*We see only those electrons, that have the positive energy*" – said Dirac. Then, he asked himself, whether one could somehow see those electrons that occupy the sea and answered yes, it is possible. According to Dirac it is sufficient to excite such an electron by providing the energy

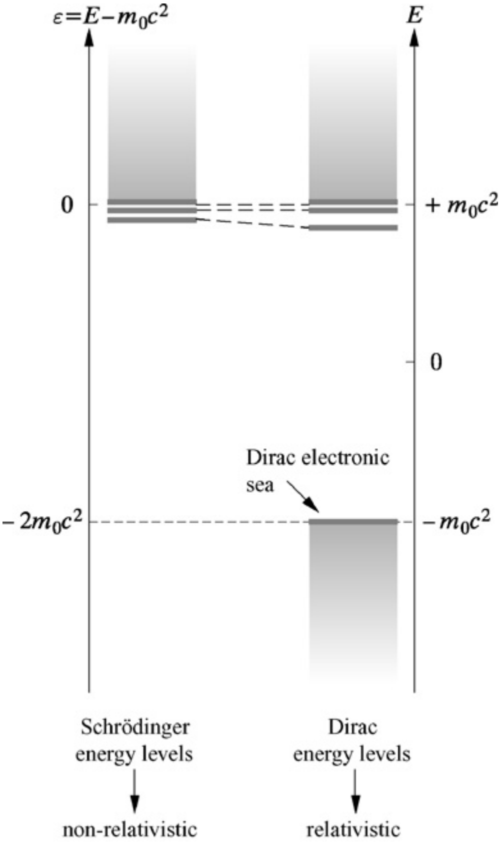


Fig. 3.2. Energy levels for the hydrogen atom according to Schrödinger (left hand side) and Dirac (right hand side). Shadowed areas correspond to the positive and negative energy continua.

of the order of $2m_0c^2$ to cover the energy gap (the energy $2m_0c^2$ is very large, of the order of 1 MeV). Then the sea electron would have positive energy and could be observed as others electrons with positive energy. However, besides the electron one, would leave a hole in the Dirac sea. Dirac has been severely molested about what this strange hole would correspond to in experimental physics. Once, when pushed too strongly, he said desperately that this is a ... proton. Some seemed to be satisfied, but others began to attack him furiously. However, then Dirac has demonstrated that the hole would have the dynamical and electrical properties of an electron, except that its sign would be opposite.[20] This has been nothing but a hypothesis for the existence of antimatter, a state of matter unknown at that time. Please imagine the shock, when three years later Carl Anderson reported the creation of electron–positron pairs from a vacuum after providing energy $2m_0c^2$. This was a day of glory for quantum theory.

positron

antimatter

Carl David Anderson (1905–1991), American physicist, professor at the Pasadena Institute of Technology. In 1932 Anderson discovered the positron, for which he received the Nobel Prize in 1936. He was also a co-discoverer of the muon.

In a moment we will see the determination with which Dirac attacked the Klein–Gordon equation, which we will write down a little differently:

$$\left[\frac{i\hbar\frac{\partial}{\partial t} - q\phi}{c}\right]^2 - \left[\left(-i\hbar\nabla - \frac{q}{c}A\right)^2 + m_0^2c^2\right] = 0. \tag{3.46}$$

Let us first introduce the following abbreviations:

$$\pi_0 = \frac{i\hbar\frac{\partial}{\partial t} - q\phi}{c}, \qquad \pi_\mu = -i\hbar\frac{\partial}{\partial \mu} - \frac{q}{c}A_\mu, \tag{3.47}$$

for $\mu = x, y, z$ or $1, 2, 3$.

---

[20] Paul Dirac, when a pupil in primary school, made his reputation after solving a riddle which goes very well with the person who thought out the positively charged electron (positron).
Three fishermen went overnight fishing and camping at a lake. After heavy fishing, around evening they put the fish in a bucket, and tired, fell asleep in the tent. At midnight one of the fishermen woke up and, tired of the whole escapade decided to take $\frac{1}{3}$ of all the fish, leave the tent quietly and go home. When he counted the fish in the bucket, it turned out that the number of fish was not divisible by 3. However, when he had thrown one fish back to the lake, the number was divisible by 3, he took his $\frac{1}{3}$ and went away. After a while a second fisherman woke up and did the same, and then the third. The question was, how many fish were in the bucket. Dirac's answer was: -2. Indeed, the number is indivisible by 3, but after the first fisherman threw away one fish the number was -3. He took his $\frac{1}{3}$, i.e. -1 fish, wrapped it up using a newspaper and went away leaving -2 fish splashing in the bucket. The same happened to each of the other two fishermen.

Dirac persisted in treating eq. (3.46) as $a^2 - b^2$ and therefore rewriting it in the form $(a+b)(a-b)$, i.e.

$$\left(\pi_0 + \sum_{\mu=x,y,z} \alpha_\mu \pi_\mu + \alpha_0 m_0 c\right)\left(\pi_0 - \sum_{\mu=x,y,z} \alpha_\mu \pi_\mu - \alpha_0 m_0 c\right) = 0. \tag{3.48}$$

He was so self-assured, that he said eq. (3.48) has to be satisfied at any price by finding suitable unknowns α_i (independent of coordinates and time). The α's have to satisfy the following relations (anti-commutation relations)

$$\alpha_\mu^2 = 1, \tag{3.49}$$

$$\alpha_\mu \alpha_\nu + \alpha_\nu \alpha_\mu = 0 \quad \text{for } \mu \neq \nu. \tag{3.50}$$

anti-commutation

Indeed, using the anti-commutation relations one recovers the Klein–Gordon equation:

$$\begin{aligned}
&\left(\pi_0 + \sum_{\mu=x,y,z}^{3} \alpha_\mu \pi_\mu + \alpha_0 m_0 c\right)\left(\pi_0 - \sum_{\mu=x,y,z}^{3} \alpha_\mu \pi_\mu - \alpha_0 m_0 c\right) \\
&= \pi_0^2 - \left[\sum_{\mu=x,y,z}^{3} \alpha_\mu \pi_\mu + \alpha_0 m_0 c\right]^2 = \pi_0^2 - \sum_{\mu,\nu=x,y,z}^{3} \alpha_\mu \alpha_\nu \pi_\mu \pi_\nu \\
&\quad - \sum_{\mu=x,y,z}^{3} (\alpha_\mu \alpha_0 + \alpha_0 \alpha_\mu) \pi_\mu m_0 c - \alpha_0 m_0^2 c^2 \\
&= \pi_0^2 - \sum_{\mu,\nu=x,y,z}^{3} (\alpha_\mu \alpha_\nu + \alpha_\mu \alpha_\nu) \pi_\mu \pi_\nu - m_0^2 c^2 = \pi_0^2 - \sum_{\mu=x,y,z}^{3} \pi_\mu^2 - m_0^2 c^2.
\end{aligned}$$

Note that the α's cannot be just numbers, because no numbers can satisfy the anticommutation relation. They have to be matrices. Since we have four variables x, y, z, t, then we may expect matrices of order 4, but they could be larger. Here is one of the consistent choices of matrices:

$$\alpha_x = \begin{pmatrix} \mathbf{0} & \sigma_x \\ \sigma_x & \mathbf{0} \end{pmatrix}, \qquad \alpha_y = \begin{pmatrix} \mathbf{0} & \sigma_y \\ \sigma_y & \mathbf{0} \end{pmatrix},$$

$$\alpha_z = \begin{pmatrix} \mathbf{0} & \sigma_z \\ \sigma_z & \mathbf{0} \end{pmatrix}, \qquad \alpha_0 \equiv \beta = \begin{pmatrix} \mathbf{1} & \mathbf{0} \\ \mathbf{0} & -\mathbf{1} \end{pmatrix}.$$

Please note the Pauli matrices σ_x, σ_y, σ_z, defined on p. 28, determine electron spin. This is the first sign of what will happen later on: the Dirac equation will automatically describe the spin angular momentum.

3.3.2 THE DIRAC EQUATIONS FOR ELECTRON AND POSITRON

After the factorization described above Dirac obtained two operator equations. The Dirac equations (for the positron and electron) correspond to these operators acting on the wave function Ψ. Thus, we obtain the equation for the negative electron energies (positron)

$$\left(\pi_0 + \sum_{\mu=x,y,z} \alpha_\mu \pi_\mu + \alpha_0 m_0 c\right)\Psi = 0 \tag{3.51}$$

and for the positive electron energies (electron)

$$\left(\pi_0 - \sum_{\mu=x,y,z} \alpha_\mu \pi_\mu - \alpha_0 m_0 c\right)\Psi = 0. \tag{3.52}$$

These two equations are coupled together through the same function Ψ which has to satisfy both of them. This coupling caused a lot of trouble in the past. First, people assumed that the equation with the negative electron energies (positron equation) may be ignored, because the energy gap is so large that the Dirac sea is occupied whatever a chemist does with a molecule. This assumption turned out to cause some really vicious or weird performances of numerical procedures (see later on). The electron equation alone reads as

$$i\hbar\frac{\partial\Psi}{\partial t} = \left(q\phi + c\sum_{\mu=x,y,z} \alpha_\mu \pi_\mu + \alpha_0 m_0 c^2\right)\Psi. \tag{3.53}$$

If one is interested in *stationary states* (cf. p. 21), the wave function has the form $\Psi(x,y,z,t) = \Psi(x,y,z)e^{-i\frac{E}{\hbar}t}$, where we have kept the same symbol for the time independent factor $\Psi(x,y,z)$. After dividing by $e^{-i\frac{E}{\hbar}t}$ we obtain

stationary states

THE DIRAC EQUATION FOR STATIONARY ELECTRONIC STATES

$$\left(E - q\phi - \beta m_0 c^2 - c\boldsymbol{\alpha}\cdot\boldsymbol{\pi}\right)\Psi(x,y,z) = 0, \tag{3.54}$$

where $\beta = \alpha_0$.

The quantity $q\phi = V$ in future applications will denote the Coulomb interaction of the particle under consideration with the external potential.

3.3.3 SPINORS AND BISPINORS

The last equation needs a comment. Because the matrices α have dimension 4, then Ψ has to be a four component vector (known as *bispinor*, its connection to the

spinors and bispinors

spin concept will be shown later on)

$$\Psi = \begin{pmatrix} \psi_1 \\ \psi_2 \\ \phi_1 \\ \phi_2 \end{pmatrix} = \begin{pmatrix} \boldsymbol{\psi} \\ \boldsymbol{\phi} \end{pmatrix},$$

where the first two components (ψ_1 and ψ_2, functions of class Q), for reasons that will become clear in a moment, are called *large components*, are hidden in vector $\boldsymbol{\psi}$, while the two *small components* (ϕ_1 and ϕ_2, functions of class Q)[21] are labelled by vector $\boldsymbol{\phi}$. Vectors $\boldsymbol{\psi}$ and $\boldsymbol{\phi}$ are called the *spinors*.

How to operate the N-component spinor (for $N = 4$ we have called them bispinors)? Let us construct the proper Hilbert space for the N-component spinors. As usual (p. 895), first, we will define the sum of two spinors in the following way:

$$\begin{pmatrix} \Psi_1 \\ \Psi_2 \\ \dots \\ \Psi_N \end{pmatrix} + \begin{pmatrix} \Phi_1 \\ \Phi_2 \\ \dots \\ \Phi_N \end{pmatrix} = \begin{pmatrix} \Psi_1 + \Phi_1 \\ \Psi_2 + \Phi_2 \\ \dots \\ \Psi_N + \Phi_N \end{pmatrix},$$

and then, the product of the spinor by a number γ:

$$\gamma \begin{pmatrix} \Psi_1 \\ \Psi_2 \\ \dots \\ \Psi_N \end{pmatrix} = \begin{pmatrix} \gamma\Psi_1 \\ \gamma\Psi_2 \\ \dots \\ \gamma\Psi_N \end{pmatrix}.$$

Next, we check that the spinors form an Abelian group with respect to the above defined addition (cf. Appendix C, p. 903) and, that the conditions for the vector space are fulfilled (Appendix B). Then, we define the scalar product of two spinors

$$\langle\Phi|\Psi\rangle = \sum_{i=1}^{N} \langle\Phi_i|\Psi_i\rangle,$$

where the scalar products $\langle\Phi_i|\Psi_i\rangle$ are defined as usual in the Hilbert space of class Q functions. Then, using the scalar product $\langle\Phi|\Psi\rangle$ we define the distance between two spinors: $\|\Phi - \Psi\| \equiv \sqrt{\langle\Phi - \Psi|\Phi - \Psi\rangle}$ and afterwards the concept of the Cauchy series (the distances between the consecutive terms tend to zero). The Hilbert space of spinors will contain all the linear combinations of the spinors together with the limits of all the convergent Cauchy series.

---

[21] It will be shown that in the non-relativistic approximation the large components reduce to the wave function known from the Schrödinger equation, and the small components vanish. In eq. (3.54) the constant E as well as the function V individually multiply each component of the bispinor Ψ, while $\boldsymbol{\sigma}\cdot\boldsymbol{\pi} \equiv \alpha_x\pi_x + \alpha_y\pi_y + \alpha_z\pi_z$ denotes the "dot product" of the matrices α_μ, $\mu = x, y, z$, by the operators π_μ (in the absence of the electromagnetic field, it is simply the momentum operator component, see p. 962). The matrix β is multiplied by the constant m_0c^2, then by the bispinor Ψ.

An operator acting on a spinor means a spinor with each component resulting from action on the corresponding component

$$\hat{A}\begin{pmatrix}\Psi_1\\ \Psi_2\\ \dots\\ \Psi_N\end{pmatrix}=\begin{pmatrix}\hat{A}\Psi_1\\ \hat{A}\Psi_2\\ \dots\\ \hat{A}\Psi_N\end{pmatrix}.$$

Sometimes we will use the notation, in which a matrix of operators acts on a spinor. In this case the result corresponds to multiplication of the matrix (of operators) and the vector (spinor)

$$\begin{pmatrix}\hat{A}_{11} & \hat{A}_{12} & \dots & \hat{A}_{1N}\\ \hat{A}_{21} & \hat{A}_{22} & \dots & \hat{A}_{2N}\\ \dots & \dots & \dots & \dots\\ \hat{A}_{N1} & \hat{A}_{N2} & \dots & \hat{A}_{NN}\end{pmatrix}\begin{pmatrix}\Psi_1\\ \Psi_2\\ \dots\\ \Psi_N\end{pmatrix}=\begin{pmatrix}\sum_j \hat{A}_{1j}\Psi_j\\ \sum_j \hat{A}_{2j}\Psi_j\\ \dots\\ \sum_j \hat{A}_{Nj}\Psi_j\end{pmatrix}.$$

3.3.4 WHAT NEXT?

In the following we will show

1. that the first two components of the bispinor are *much larger* than the last two
2. that in the limit $c \to \infty$ the Dirac equation gives the *Schrödinger equation*
3. that the Dirac equation accounts for the spin angular momentum of the electron
4. how to obtain, in a simple way, an approximate solution of the Dirac equation to the electron in the field of a nucleus (*"hydrogen-like atom"*).

3.3.5 LARGE AND SMALL COMPONENTS OF THE BISPINOR

Using matrix multiplication rules, the Dirac equation (3.54) with bispinors can be rewritten in the form of *two equations with spinors* $\boldsymbol{\psi}$ and $\boldsymbol{\phi}$:

$$\left(E - V - m_0c^2\right)\boldsymbol{\psi} - c(\boldsymbol{\sigma}\cdot\boldsymbol{\pi})\boldsymbol{\phi} = \mathbf{0}, \tag{3.55}$$

$$\left(E - V + m_0c^2\right)\boldsymbol{\phi} - c(\boldsymbol{\sigma}\cdot\boldsymbol{\pi})\boldsymbol{\psi} = \mathbf{0}. \tag{3.56}$$

The quantity m_0c^2 represents the energy. Let us use this energy to *shift the energy scale* (we are always free to choose 0 on this scale): $\varepsilon = E - m_0c^2$, in order to make ε comparable in future to the eigenvalue of the Schrödinger equation (p. 70). We obtain

$$(\varepsilon - V)\boldsymbol{\psi} - c(\boldsymbol{\sigma}\cdot\boldsymbol{\pi})\boldsymbol{\phi} = \mathbf{0}, \tag{3.57}$$

$$\left(\varepsilon - V + 2m_0c^2\right)\boldsymbol{\phi} - c(\boldsymbol{\sigma}\cdot\boldsymbol{\pi})\boldsymbol{\psi} = \mathbf{0}. \tag{3.58}$$

This set of equations corresponds to a single matrix equation:

$$\begin{pmatrix} V & c(\boldsymbol{\sigma}\cdot\boldsymbol{\pi}) \\ c(\boldsymbol{\sigma}\cdot\boldsymbol{\pi}) & V-2m_0c^2 \end{pmatrix}\begin{pmatrix} \boldsymbol{\psi} \\ \boldsymbol{\phi} \end{pmatrix}=\begin{pmatrix} \varepsilon & 0 \\ 0 & \varepsilon \end{pmatrix}\begin{pmatrix} \boldsymbol{\psi} \\ \boldsymbol{\phi} \end{pmatrix}. \tag{3.59}$$

3.3.6 HOW TO AVOID DROWNING IN THE DIRAC SEA

When, in the past, the above equation was solved and the energy ε minimized (routine practice in the non-relativistic case) with respect to the variational parameters (see Chapter 5) in the trial spinors $\boldsymbol{\psi}$ and $\boldsymbol{\phi}$, then some serious numerical problems occurred. Either the numerical procedures diverged or the solutions obtained were physically unacceptable. The reason for this was that the existence of the Dirac sea had been totally ignored by neglecting eq. (3.51) for the positron and taking solely eq. (3.52) for electron motion. The variational trial functions felt, however, the presence of the Dirac sea electronic states (there was nothing in the theory that would prevent the electron from attempting to occupy negative energies) and the corresponding variational energies dived down the energy scale towards the abyss of the sea without a bottom.[22] The presence of the Dirac sea makes the Dirac theory, in fact, a theory of an infinite number of particles, whereas formally it was only a theory of a single particle in an external field. This kind of discomfort made people think of the possibility of describing the electron from the Dirac electronic sea by replacing the bispinors by the exact spinor (two components) theory.[23] Such exact separation has been reported by Barysz and Sadlej.[24]

An approximate (and simple) prescription was also invented to avoid the catastrophic drowning described above. Indeed, eq. (3.58) can be transformed without any problem to

$$\boldsymbol{\phi}=\left(1+\frac{(\varepsilon-V)}{2m_0c^2}\right)^{-1}\frac{1}{2m_0c}(\boldsymbol{\sigma}\cdot\boldsymbol{\pi})\boldsymbol{\psi}.$$

Since $2m_0c^2$ represents a huge energy when compared to the kinetic energy $\varepsilon-V$, then the first parenthesis on the right-hand side is to a good approximation

---

[22] How severe the problem might be has been shown by M. Stanke, J. Karwowski, "*Variational Principle in the Dirac Theory: Spurious Solutions, Unexpected Extrema and Other Traps*" in "*New Trends in Quantum Systems in Chemistry and Physics*", vol. I, p. 175–190, eds. J. Maruani et al., Kluwer Academic Publishers. Sometimes an eigenfunction corresponds to a quite different eigenvalue. Nothing of that sort appears in non-relativistic calculations.

[23] Exact within the Dirac model.

[24] M. Barysz, A.J. Sadlej, J.G. Snijders, *Int. J. Quantum Chem.* 65 (1997) 225; M. Barysz, *J. Chem. Phys.* 114 (2001) 9315; M. Barysz, A.J. Sadlej, *J. Mol. Struct. (Theochem)* 573 (2001) 181; M. Barysz, A.J. Sadlej, *J. Chem. Phys.* 116 (2002) 2696. In the latter paper an exact solution to the problem was given. The two-component theory, although more appealing, both from the point of view of physics as well as computationally, implies a change in definition of the operators, e.g., the position operator is replaced by a quite complex expression. This fact, ignored in computations using two-component theories, has been analysed in the articles: V. Kellő, A.J. Sadlej, B.A. Hess, *J. Chem. Phys.* 105 (1996) 1995; M. Barysz, A.J. Sadlej, *Theor. Chem. Acc.* 97 (1997) 260; V. Kellő, A.J. Sadlej, *Int. J. Quantum Chem.* 68 (1998) 159; V. Kellő, A.J. Sadlej, *J. Mol. Struct. (Theochem)* 547 (2001) 35.

equal to 1. This means however that

$$\boldsymbol{\phi} \approx \frac{1}{2m_0c}(\boldsymbol{\sigma}\cdot\boldsymbol{\pi})\boldsymbol{\psi},$$

which is known as "kinetic balancing". It was shown that the "*kinetically balanced*" trial function achieves the miracle[25] of the energy not tending to $-\infty$. The kinetic balancing indicates some fixed relation between $\boldsymbol{\phi}$ and $\boldsymbol{\psi}$.

kinetic balancing

Let us focus now on $\boldsymbol{\sigma}\cdot\boldsymbol{\pi}$. This is a 2×2 matrix and in the absence of an electromagnetic field ($\boldsymbol{\pi}=\boldsymbol{p}$) one has:

$$\begin{aligned}\boldsymbol{\sigma}\cdot\boldsymbol{\pi} &= \sigma_x\hat{p}_x+\sigma_y\hat{p}_y+\sigma_z\hat{p}_z\\ &= \begin{pmatrix}0 & \hat{p}_x\\ \hat{p}_x & 0\end{pmatrix}+\begin{pmatrix}0 & -i\hat{p}_y\\ i\hat{p}_y & 0\end{pmatrix}+\begin{pmatrix}\hat{p}_z & 0\\ 0 & -\hat{p}_z\end{pmatrix}=\begin{pmatrix}\hat{p}_z & \hat{p}_x-i\hat{p}_y\\ \hat{p}_x+i\hat{p}_y & -\hat{p}_z\end{pmatrix}.\end{aligned}$$

It is seen that $\boldsymbol{\sigma}\cdot\boldsymbol{\pi}$ is of the order of momentum mv, and for the small velocities of the order of m_0v.

> Hence, one obtains $\boldsymbol{\phi}\approx\frac{1}{2m_0c}(\boldsymbol{\sigma}\cdot\boldsymbol{\pi})\boldsymbol{\psi}\approx\frac{v}{2c}\boldsymbol{\psi}$, therefore the component $\boldsymbol{\phi}$ is for small v much smaller than the component $\boldsymbol{\psi}$,

which justifies the terms "small" and "large" components.[26]

3.3.7 FROM DIRAC TO SCHRÖDINGER – HOW TO DERIVE THE NON-RELATIVISTIC HAMILTONIAN?

The *approximate* relation ("kinetic balance") between the large and small components of the bispinor (that holds for small v/c) may be used to *eliminate* the small components[27] from (3.57) and (3.58). We obtain

$$(\varepsilon-V)\boldsymbol{\psi}-c(\boldsymbol{\sigma}\cdot\boldsymbol{\pi})\frac{1}{2m_0c}(\boldsymbol{\sigma}\cdot\boldsymbol{\pi})\boldsymbol{\psi}= \tag{3.60}$$

$$(\varepsilon-V)\boldsymbol{\psi}-\frac{1}{2m_0}(\boldsymbol{\sigma}\cdot\boldsymbol{\pi})(\boldsymbol{\sigma}\cdot\boldsymbol{\pi})\boldsymbol{\psi}=\mathbf{0}. \tag{3.61}$$

---

[25] This remedy has not only an *ad hoc* character, but moreover does not work for the heaviest atoms, which are otherwise the most important target of relativistic computations.

[26] These terms refer to the positive part of the energy spectrum. For the negative continuum (Dirac sea) the proportion of the components is reversed.

[27] A more elegant solution was reported by Andrzej W. Rutkowski, *J. Phys. B.* 9 (1986) 3431, *ibid.* 19 (1986) 3431, *ibid.* 19 (1986) 3443. For the one-electron case, this approach was later popularized by Werner Kutzelnigg as Direct Perturbation Theory (DPT).

Let us take a closer look at the meaning of the expression

$$(\boldsymbol{\sigma}\cdot\boldsymbol{\pi})(\boldsymbol{\sigma}\cdot\boldsymbol{\pi}) = \begin{pmatrix} \hat{p}_z & \hat{p}_x - i\hat{p}_y \\ \hat{p}_x + i\hat{p}_y & -\hat{p}_z \end{pmatrix}\begin{pmatrix} \hat{p}_z & \hat{p}_x - i\hat{p}_y \\ \hat{p}_x + i\hat{p}_y & -\hat{p}_z \end{pmatrix}$$

$$= \begin{pmatrix} \hat{p}^2 & 0 \\ 0 & \hat{p}^2 \end{pmatrix} = \hat{p}^2\mathbf{1}.$$

Now please look carefully. Let us insert this into the last equation. We obtain what is sometimes called the Schrödinger equation with spin (because it is satisfied by a two-component spinor)

$$\left(\frac{\hat{p}^2}{2m_0} + V\right)\boldsymbol{\psi} = \varepsilon\boldsymbol{\psi}.$$

Recalling that $\hat{\boldsymbol{p}}$ represents the momentum operator, we observe each of the large components satisfies the familiar Schrödinger equation

$$\left(-\frac{\hbar^2}{2m_0}\Delta + V\right)\psi = \varepsilon\psi.$$

Therefore, the non-relativistic equation has been obtained from the relativistic one, assuming that the velocity of particle v is negligibly small with respect to the speed of light c. The Dirac equation remains valid even for larger particle velocities.

3.3.8 HOW DOES THE SPIN APPEAR?

It will be shown that the Dirac equation for the free electron in an external electromagnetic field is leading to the spin concept. Thus, in relativistic theory, the spin angular momentum appears in a natural way, whereas in the non-relativistic formalism it was the subject of a postulate of quantum mechanics, p. 25.

First let us introduce the following identity:

$$(\boldsymbol{\sigma}\cdot\boldsymbol{a})(\boldsymbol{\sigma}\cdot\boldsymbol{b}) = (\boldsymbol{a}\cdot\boldsymbol{b})\mathbf{1} + i\boldsymbol{\sigma}\cdot(\boldsymbol{a}\times\boldsymbol{b}),$$

where, on the left-hand side, we have a product of two matrices, each formed by a “scalar product” of matrices[28] $\boldsymbol{\sigma}$ and a vector, whereas on the right-hand side we have the scalar product of two vectors multiplied by a unit matrix plus the scalar

---

[28] That is, $\boldsymbol{\sigma}\cdot\boldsymbol{a} = \sigma_x a_x + \sigma_y a_y + \sigma_z a_z$.

product of the matrix $i\boldsymbol{\sigma}$ and the vector $\boldsymbol{a}\times\boldsymbol{b}$. The left-hand side:

$$\begin{aligned}&\left[\begin{pmatrix}0 & a_x\\ a_x & 0\end{pmatrix}+\begin{pmatrix}0 & -ia_y\\ ia_y & 0\end{pmatrix}+\begin{pmatrix}a_z & 0\\ 0 & -a_z\end{pmatrix}\right]\\ &\quad\times\left[\begin{pmatrix}0 & b_x\\ b_x & 0\end{pmatrix}+\begin{pmatrix}0 & -ib_y\\ ib_y & 0\end{pmatrix}+\begin{pmatrix}b_z & 0\\ 0 & -b_z\end{pmatrix}\right]\\ &\quad=\begin{pmatrix}a_z & a_x-ia_y\\ a_x+ia_y & -a_z\end{pmatrix}\begin{pmatrix}b_z & b_x-ib_y\\ b_x+ib_y & -b_z\end{pmatrix}\\ &\quad=\begin{pmatrix}\boldsymbol{a}\cdot\boldsymbol{b}+i(\boldsymbol{a}\times\boldsymbol{b})_z & (\boldsymbol{a}\times\boldsymbol{b})_y+i(\boldsymbol{a}\times\boldsymbol{b})_x\\ -(\boldsymbol{a}\times\boldsymbol{b})_y+i(\boldsymbol{a}\times\boldsymbol{b})_x & \boldsymbol{a}\cdot\boldsymbol{b}-i(\boldsymbol{a}\times\boldsymbol{b})_z\end{pmatrix}\end{aligned}$$

is therefore equal to the right-hand side, which is what we wanted to show.

Now, taking $\boldsymbol{a}=\boldsymbol{b}=\boldsymbol{\pi}$ one obtains the relation

$$(\boldsymbol{\sigma}\cdot\boldsymbol{\pi})(\boldsymbol{\sigma}\cdot\boldsymbol{\pi})=(\boldsymbol{\pi}\cdot\boldsymbol{\pi})\mathbf{1}+i\boldsymbol{\sigma}(\boldsymbol{\pi}\times\boldsymbol{\pi}).$$

If the vector $\boldsymbol{\pi}$ had numbers as its components, the last term would have had to be zero, because the vector product of two parallel vectors would be zero. This, however, need not be true when the vector components are *operators* (as it is in our case). Since $\boldsymbol{\pi}=\boldsymbol{p}-\frac{q}{c}\boldsymbol{A}$, then $(\boldsymbol{\pi}\cdot\boldsymbol{\pi})=\boldsymbol{\pi}^2$ and $(\boldsymbol{\pi}\times\boldsymbol{\pi})=iq\frac{\hbar}{c}\,\mathrm{curl}\,\boldsymbol{A}$. To check this, we will obtain the last equality for the x components of both sides (the proof for the other two components looks the same). Let the operator $(\boldsymbol{\pi}\times\boldsymbol{\pi})$ act on an arbitrary function $f(x,y,z)$. As a result we expect the product of f and the vector $iq\frac{\hbar}{c}\,\mathrm{curl}\,\boldsymbol{A}$. Let us see:

$$\begin{aligned}(\boldsymbol{\pi}\times\boldsymbol{\pi})_x f&=(\hat{p}_y-q/cA_y)(\hat{p}_z-q/cA_z)f-(\hat{p}_z-q/cA_z)(\hat{p}_y-q/cA_y)f\\ &=[\hat{p}_y\hat{p}_z-q/c\hat{p}_yA_z-q/cA_y\hat{p}_z+(q/c)^2A_yA_z\\ &\quad-\hat{p}_z\hat{p}_y+q/c\hat{p}_zA_y+q/cA_z\hat{p}_y-(q/c)^2A_zA_y]f\\ &=-q/c(-i\hbar)\left\{\frac{\partial}{\partial y}(A_zf)-A_z\frac{\partial f}{\partial y}+A_y\frac{\partial f}{\partial z}-\frac{\partial}{\partial z}(A_yf)\right\}\\ &=i\hbar q/c\left\{\frac{\partial A_z}{\partial y}-\frac{\partial A_y}{\partial z}\right\}f=\frac{iq\hbar}{c}(\mathrm{curl}\,A)_xf.\end{aligned}$$

This is what we expected to get. From the Maxwell equations (p. 962), we have $\mathrm{curl}\,\boldsymbol{A}=\boldsymbol{H}$, where $\boldsymbol{H}$ represents the magnetic field intensity. Let us insert this into the Dirac equation (valid for kinetic energy much smaller than $2m_0c^2$, see eq. (3.60))

$$(\varepsilon-V)\boldsymbol{\psi}=\frac{1}{2m_0}(\boldsymbol{\sigma}\cdot\boldsymbol{\pi})(\boldsymbol{\sigma}\cdot\boldsymbol{\pi})\boldsymbol{\psi}$$

$$
\begin{aligned}
&= \frac{1}{2m_0}(\boldsymbol{\pi}\cdot\boldsymbol{\pi})\psi + \frac{i}{2m_0}\boldsymbol{\sigma}\cdot(\boldsymbol{\pi}\times\boldsymbol{\pi})\psi \\
&= \frac{1}{2m_0}(\boldsymbol{\pi}\cdot\boldsymbol{\pi})\psi + \frac{i}{2m_0}\frac{iq\hbar}{c}(\boldsymbol{\sigma}\cdot\boldsymbol{H})\psi \\
&= \left[\frac{\pi^2}{2m_0} - \frac{q\hbar}{2m_0c}\boldsymbol{\sigma}\cdot\boldsymbol{H}\right]\psi = \left[\frac{\pi^2}{2m_0} + \frac{e\hbar}{2m_0c}\boldsymbol{\sigma}\cdot\boldsymbol{H}\right]\psi.
\end{aligned}
$$

In the last parenthesis, beside the kinetic energy operator (first term), there is a strange second term. The term has the appearance of the interaction energy $-\boldsymbol{M}\cdot\boldsymbol{H}$ of a mysterious magnetic dipole moment $\boldsymbol{M}$ with magnetic field $\boldsymbol{H}$ (cf. interaction with magnetic field, p. 659). The operator of this electronic dipole moment $\boldsymbol{M} = -\frac{e\hbar}{2m_0c}\boldsymbol{\sigma} = -\mu_B\boldsymbol{\sigma}$, where μ_B stands for the Bohr magneton equal to $\frac{e\hbar}{2m_0c}$. The spin angular momentum operator of the electron is denoted by (cf. p. 28) $\boldsymbol{s}$. Therefore, one has $\boldsymbol{s} = \frac{1}{2}\hbar\boldsymbol{\sigma}$. Inserting $\boldsymbol{\sigma}$ to the equation for $\boldsymbol{M}$ we obtain

$$\boldsymbol{M} = -2\frac{\mu_B}{\hbar}\boldsymbol{s} = -\frac{e}{m_0c}\boldsymbol{s}. \tag{3.62}$$

It is exactly twice as much as we get for the *orbital* angular momentum and the corresponding *orbital* magnetic dipole (hence the anomalous magnetic spin moment of the electron), see eq. (12.53).

When two values differ by an integer factor (as in our case) this should stimulate our mind, because it may mean something fundamental that might depend on, e.g., the number of dimensions of our space or something similar. However, one of the most precise experiments ever made by humans gave[29] 2.0023193043737 ± 0.0000000000082 instead of 2. Therefore, our excitement must diminish. A more accurate theory (quantum electrodynamics, some of the effects of this will be described later) gave a result that agreed with the experiment within an experimental error of ±0.0000000008. The extreme accuracy achieved witnessed the exceptional status of quantum electrodynamics, because no other theory of mankind has achieved such a level of accuracy.

3.3.9 SIMPLE QUESTIONS

How to interpret a bispinor wave function? Does the Dirac equation describe a single fermion, an electron, a positron, an electron and a Dirac sea of other electrons (infinite number of particles), an effective electron or effective positron (interacting with the Dirac sea)? After eighty years these questions do not have a clear answer.

---

[29] R.S. Van Dyck Jr., P.B. Schwinberg, H.G. Dehmelt, *Phys. Rev. Letters* 59 (1990) 26.

Despite the glorious invariance with respect to the Lorentz transformation and despite spectacular successes, the Dirac equation has some serious drawbacks, including a lack of clear physical interpretation. These drawbacks are removed by a more advanced theory – quantum electrodynamics.

3.4 THE HYDROGEN-LIKE ATOM IN DIRAC THEORY

After this short escapade we are back with Dirac theory. The hydrogen-like atom may be simplified by immobilizing the nucleus and considering a single particle – the electron[30] moving in the *electrostatic field of the nucleus*[31] $-Z/r$. This problem has an exact solution first obtained by Charles Galton Darwin, cf. p. 112. The electron state is described by four quantum numbers n, l, m, m_s, where $n = 1, 2, \ldots$ stands for the principal, $0 \leqslant l \leqslant n-1$ for the angular, $|m| \leqslant l$ for the magnetic and $m_s = \frac{1}{2}, -\frac{1}{2}$ for the spin quantum number. Darwin obtained the following formula for the relativistic energy of the hydrogen-like atom (in a.u.):

$$E_{n,j} = -\frac{1}{2n^2}\left[1 + \frac{1}{nc^2}\left(\frac{1}{j+\frac{1}{2}} - \frac{3}{4n}\right)\right],$$

where $j = l + m_s$, and c is the speed of light (in a.u.). For the ground state (1s, $n = 1$, $l = 0$, $m = 0$, $m_s = \frac{1}{2}$) we have

$$E_{1,\frac{1}{2}} = -\frac{1}{2}\left[1 + \left(\frac{1}{2c}\right)^2\right].$$

Thus, instead of the non-relativistic energy equal to $-\frac{1}{2}$, from the Dirac theory we obtain -0.5000067 a.u., which means a very small correction to the non-relativistic energy. The electron energy levels for the non-relativistic and relativistic cases are shown schematically in Fig. 3.2.

3.4.1 STEP BY STEP: CALCULATION OF THE GROUND STATE OF THE HYDROGEN-LIKE ATOM WITHIN DIRAC THEORY

Matrix form of the Dirac equation

We will use the Dirac equation (3.59). First, the basis set composed of two bispinors will be created: $\boldsymbol{\Psi}_1 = \begin{pmatrix} \psi \\ \mathbf{0} \end{pmatrix}$ and $\boldsymbol{\Psi}_2 = \begin{pmatrix} \mathbf{0} \\ \phi \end{pmatrix}$, and the wave function $\boldsymbol{\Psi}$ will be

---

[30] In the Dirac equation $\boldsymbol{A} = \mathbf{0}$ and $-e\phi = V = -\frac{Ze^2}{r}$ were set.

[31] The centre-of-mass motion can be easily separated from the Schrödinger equation, Appendix I. Nothing like this has been done for the Dirac equation. The atomic mass depends on its velocity with respect to the laboratory coordinate system, the electron and proton mass also depend on their speeds, and there is also a mass deficit as a result of binding between both particles. All this seems to indicate that centre of mass separation is not possible. Nevertheless, for an energy expression accurate to a certain power of c^{-1}, such a separation is, at least in some cases, possible.

sought as a linear combination $\boldsymbol{\Psi} = c_1\boldsymbol{\Psi}_1 + c_2\boldsymbol{\Psi}_2$, which represents an approximation. Within this approximation the Dirac equation looks like this

$$\begin{pmatrix} V-\varepsilon & c(\boldsymbol{\sigma}\cdot\boldsymbol{\pi}) \\ c(\boldsymbol{\sigma}\cdot\boldsymbol{\pi}) & V-2m_0c^2-\varepsilon \end{pmatrix}(c_1\boldsymbol{\Psi}_1 + c_2\boldsymbol{\Psi}_2) = \mathbf{0},$$

which gives

$$c_1\begin{pmatrix} V-\varepsilon & c(\boldsymbol{\sigma}\cdot\boldsymbol{\pi}) \\ c(\boldsymbol{\sigma}\cdot\boldsymbol{\pi}) & V-2m_0c^2-\varepsilon \end{pmatrix}\boldsymbol{\Psi}_1 + c_2\begin{pmatrix} V-\varepsilon & c(\boldsymbol{\sigma}\cdot\boldsymbol{\pi}) \\ c(\boldsymbol{\sigma}\cdot\boldsymbol{\pi}) & V-2m_0c^2-\varepsilon \end{pmatrix}\boldsymbol{\Psi}_2 = \mathbf{0}.$$

By making a scalar product first with $\boldsymbol{\Psi}_1$ and then with $\boldsymbol{\Psi}_2$ we obtain two equations:

$$\begin{aligned} &c_1\left\langle\boldsymbol{\Psi}_1\middle|\begin{pmatrix} V-\varepsilon & c(\boldsymbol{\sigma}\cdot\boldsymbol{\pi}) \\ c(\boldsymbol{\sigma}\cdot\boldsymbol{\pi}) & V-2m_0c^2-\varepsilon \end{pmatrix}\boldsymbol{\Psi}_1\right\rangle \\ &\quad + c_2\left\langle\boldsymbol{\Psi}_1\middle|\begin{pmatrix} V-\varepsilon & c(\boldsymbol{\sigma}\cdot\boldsymbol{\pi}) \\ c(\boldsymbol{\sigma}\cdot\boldsymbol{\pi}) & V-2m_0c^2-\varepsilon \end{pmatrix}\boldsymbol{\Psi}_2\right\rangle = 0 \\ &c_1\left\langle\boldsymbol{\Psi}_2\middle|\begin{pmatrix} V-\varepsilon & c(\boldsymbol{\sigma}\cdot\boldsymbol{\pi}) \\ c(\boldsymbol{\sigma}\cdot\boldsymbol{\pi}) & V-2m_0c^2-\varepsilon \end{pmatrix}\boldsymbol{\Psi}_1\right\rangle \\ &\quad + c_2\left\langle\boldsymbol{\Psi}_2\middle|\begin{pmatrix} V-\varepsilon & c(\boldsymbol{\sigma}\cdot\boldsymbol{\pi}) \\ c(\boldsymbol{\sigma}\cdot\boldsymbol{\pi}) & V-2m_0c^2-\varepsilon \end{pmatrix}\boldsymbol{\Psi}_2\right\rangle = 0. \end{aligned}$$

Taking into account the particular structure of the bispinors $\boldsymbol{\Psi}_1$ and $\boldsymbol{\Psi}_2$, we obtain the same equations expressed in spinors (two component spinors)

$$\begin{aligned} &c_1\langle\boldsymbol{\psi}|(V-\varepsilon)\boldsymbol{\psi}\rangle + c_2\langle\boldsymbol{\psi}|\,c(\boldsymbol{\sigma}\cdot\boldsymbol{\pi})\boldsymbol{\phi}\rangle = 0, \\ &c_1\langle\boldsymbol{\phi}|c(\boldsymbol{\sigma}\cdot\boldsymbol{\pi})\boldsymbol{\psi}\rangle + c_2\langle\boldsymbol{\phi}|\,(V-2m_0c^2-\varepsilon)\boldsymbol{\phi}\rangle = 0. \end{aligned}$$

This is a set of homogeneous linear equations. To obtain a non-trivial solution,[32] the determinant of the coefficients multiplying the unknowns c_1 and c_2 has to be zero (the secular determinant, cf. variational method in Chapter 5)

$$\begin{vmatrix} \langle\boldsymbol{\psi}|(V-\varepsilon)\boldsymbol{\psi}\rangle & \langle\boldsymbol{\psi}|c(\boldsymbol{\sigma}\cdot\boldsymbol{\pi})\boldsymbol{\phi}\rangle \\ \langle\boldsymbol{\phi}|c(\boldsymbol{\sigma}\cdot\boldsymbol{\pi})\boldsymbol{\psi}\rangle & \langle\boldsymbol{\phi}|(V-2m_0c^2-\varepsilon)\boldsymbol{\phi}\rangle \end{vmatrix} = 0.$$

The potential V in the above formula will be taken as $-Z/r$, where r is the electron–nucleus distance.

---

[32] It is easy to give a trivial one, but not acceptable (the wave function cannot equal zero everywhere): $c_1 = c_2 = 0$.

The large component spinor

It is true that we have used an extremely poor basis, however, we will try to compensate for it by allowing a certain flexibility within the large component spinor: $\boldsymbol{\psi} = \begin{pmatrix} 1s \\ 0 \end{pmatrix}$, where the hydrogen-like function $1s = \sqrt{\frac{\zeta^3}{\pi}} \exp(-\zeta r)$. The parameter ζ will be optimized in such a way as to minimize the energy ε of the electron. This idea is similar to the variational method in the non-relativistic theory (Chapter 5 and Appendix H, p. 969), however, it is hardly justified in the relativistic case. Indeed, as proved by numerical experience the variational procedure very often fails. As a remedy we will use kinetic balancing already used to describe the large and small components of the bispinor (p. 119). The spinor of the small components is therefore obtained automatically from the large components (approximation):

$$\begin{aligned}\boldsymbol{\phi} &= \mathcal{N}(\boldsymbol{\sigma} \cdot \boldsymbol{\pi}) \begin{pmatrix} 1s \\ 0 \end{pmatrix} = \mathcal{N} \begin{pmatrix} \hat{p}_z & \hat{p}_x + i\hat{p}_y \\ \hat{p}_x - i\hat{p}_y & \hat{p}_z \end{pmatrix} \begin{pmatrix} 1s \\ 0 \end{pmatrix} \\ &= \mathcal{N} \begin{pmatrix} \hat{p}_z(1s) \\ (\hat{p}_x + i\hat{p}_y)(1s) \end{pmatrix},\end{aligned}$$

where $\mathcal{N}$ is a normalization constant. In the above formula $\hat{\boldsymbol{p}}$ represents the momentum operator. The normalization constant $\mathcal{N}$ will be found from

$$\begin{aligned}\langle \boldsymbol{\phi} | \boldsymbol{\phi} \rangle &= 1 = |\mathcal{N}|^2 \{ \langle \hat{p}_z(1s) \mid \hat{p}_z(1s) \rangle + \langle (\hat{p}_x + i\hat{p}_y)(1s) \mid (\hat{p}_x + i\hat{p}_y)(1s) \rangle \} \\ &= |\mathcal{N}|^2 \cdot \left\{ \begin{array}{c} \langle \hat{p}_z(1s) | \hat{p}_z(1s) \rangle + \langle \hat{p}_x(1s) | \hat{p}_x(1s) \rangle + i \langle \hat{p}_x(1s) | \hat{p}_y(1s) \rangle \\ - i \langle \hat{p}_y(1s) | \hat{p}_x(1s) \rangle + \langle \hat{p}_y(1s) | \hat{p}_y(1s) \rangle \end{array} \right\}.\end{aligned}$$

In the above formula, integrals with the imaginary unit i are equal to zero, because the integrand is an odd function. After using the Hermitian character of the momentum operator we obtain $1 = |\mathcal{N}|^2 \langle 1s | \hat{p}^2 1s \rangle = \zeta^2$. The last equality follows from Appendix H, p. 969. Thus, one may choose $\mathcal{N} = 1/\zeta$.

Calculating integrals in the Dirac matrix equation

We will calculate one by one all the integrals that appear in the Dirac matrix equation. The integral $\langle \boldsymbol{\psi} | -\frac{Z}{r} \boldsymbol{\psi} \rangle = -Z\zeta$, because the scalar product leads to the nuclear attraction integral with a hydrogen-like atomic orbital, and this gives the result above (Appendix H, p. 969). The next integral can be computed as follows

$$\begin{aligned}\left\langle \boldsymbol{\phi} \middle| \frac{1}{r} \boldsymbol{\phi} \right\rangle &= |\mathcal{N}|^2 \left\langle \begin{matrix} \hat{p}_z(1s) \\ (\hat{p}_x + i\hat{p}_y)(1s) \end{matrix} \middle| \frac{1}{r} \begin{pmatrix} \hat{p}_z(1s) \\ (\hat{p}_x + i\hat{p}_y)(1s) \end{pmatrix} \right\rangle \\ &= |\mathcal{N}|^2 \left\langle \hat{p}_z(1s) \middle| \frac{1}{r} \hat{p}_z(1s) \right\rangle + \left\langle (\hat{p}_x + i\hat{p}_y)(1s) \middle| \frac{1}{r} (\hat{p}_x + i\hat{p}y)(1s) \right\rangle\end{aligned}$$

$$
\begin{aligned}
&= |\mathcal{N}|^2\left\langle (1s)\middle|\hat{p}_z\frac{1}{r}\hat{p}_z(1s)\right\rangle + \left\langle (1s)\middle|(\hat{p}_x - i\hat{p}_y)\frac{1}{r}(\hat{p}_x + i\hat{p}_y)(1s)\right\rangle \\
&= |\mathcal{N}|^2\left\langle (1s)\middle|\left(\hat{p}_z\frac{1}{r}\right)\hat{p}_z(1s)\right\rangle + \left\langle (1s)\middle|\left[(\hat{p}_x - i\hat{p}_y)\frac{1}{r}\right](\hat{p}_x + i\hat{p}_y)(1s)\right\rangle \\
&\quad + \left\langle (1s)\middle|\frac{1}{r}\hat{p}_z\hat{p}_z(1s)\right\rangle + \left\langle (1s)\middle|\frac{1}{r}\left(\hat{p}_x - i\hat{p}_y\right)\left(\hat{p}_x + i\hat{p}_y\right)(1s)\right\rangle. \qquad (3.63)
\end{aligned}
$$

In the second row, the scalar product of spinors is used, in the third row, the Hermitian character of the operator $\hat{p}$. Further,

$$
\begin{aligned}
\left\langle \boldsymbol{\phi}\middle|\frac{1}{r}\boldsymbol{\phi}\right\rangle &= |\mathcal{N}|^2\left[\left\langle (1s)\middle|\left(\hat{p}_z\frac{1}{r}\right)\hat{p}_z(1s)\right\rangle + \left\langle (1s)\middle|\frac{1}{r}\left(\hat{p}_x^2 + \hat{p}_y^2 + \hat{p}_z^2\right)(1s)\right\rangle\right. \\
&\quad \left. + \left\langle (1s)\middle|\left[\left(\hat{p}_x - i\hat{p}_y\right)\frac{1}{r}\right]\left(\hat{p}_x + i\hat{p}_y\right)(1s)\right\rangle\right] \\
&= |\mathcal{N}|^2\left[\left\langle (1s)\middle|\left(\hat{p}_z\frac{1}{r}\right)\hat{p}_z(1s)\right\rangle - \left\langle (1s)\middle|\frac{1}{r}\Delta(1s)\right\rangle + \left\langle (1s)\middle|\left(\hat{p}_x\frac{1}{r}\right)\hat{p}_x(1s)\right\rangle\right. \\
&\quad + \left\langle (1s)\middle|\left(\hat{p}_y\frac{1}{r}\right)\hat{p}_y(1s)\right\rangle - i\left\langle (1s)\middle|\left(\hat{p}_y\frac{1}{r}\right)\hat{p}_x(1s)\right\rangle \\
&\quad \left. + i\left\langle (1s)\middle|\left(\hat{p}_x\frac{1}{r}\right)\hat{p}_y(1s)\right\rangle\right]. \qquad (3.64)
\end{aligned}
$$

We used the atomic units and therefore $\hat{p}^2 = -\Delta$, and the momentum operator is equal to $-i\nabla$. The two integrals at the end cancel each other, because each of the integrals does not change when the variables are interchanged: $x \leftrightarrow y$.

Finally, we obtain the following formula

$$
\begin{aligned}
\left\langle \boldsymbol{\phi}\middle|\frac{1}{r}\boldsymbol{\phi}\right\rangle &= -|\mathcal{N}|^2\left\{\left\langle 1s\middle|\frac{1}{r}\Delta(1s)\right\rangle + \left\langle 1s\middle|\left(\nabla\frac{1}{r}\right)\nabla(1s)\right\rangle\right\} \\
&= -\zeta^{-2}\left\{\left(-3\zeta^3 + 2\zeta^3\right)\right\} = \zeta,
\end{aligned}
$$

where the equality follows from a direct calculation of the two integrals.[33]

The next matrix element to calculate is equal to $\langle \boldsymbol{\phi}|c(\boldsymbol{\sigma}\cdot\boldsymbol{\pi})\boldsymbol{\psi}\rangle$. We proceed as follows (please recall kinetic balancing and we also use Appendix H, p. 969):

$$
\langle \boldsymbol{\phi}|c(\boldsymbol{\sigma}\cdot\boldsymbol{\pi})\boldsymbol{\psi}\rangle = \mathcal{N}c\left\langle (\boldsymbol{\sigma}\cdot\boldsymbol{\pi})\begin{pmatrix}1s\\0\end{pmatrix}\middle|(\boldsymbol{\sigma}\cdot\boldsymbol{\pi})\begin{pmatrix}1s\\0\end{pmatrix}\right\rangle
$$

---

[33] In the first integral we have the same situation as a while before. In the second integral we write the nabla operator in Cartesian coordinates, obtain a scalar product of two gradients, then we get three integrals equal to one another (they contain x, y, z), and it is sufficient to calculate one of them by spherical coordinates by formula (H.2) in Appendix H, p. 969.

$$
\begin{aligned}
&= \mathcal{N}c\left\langle \begin{pmatrix} \hat{p}_z(1s) \\ (\hat{p}_x + i\hat{p}_y)(1s) \end{pmatrix} \middle| \begin{pmatrix} \hat{p}_z(1s) \\ (\hat{p}_x + i\hat{p}_y)(1s) \end{pmatrix} \right\rangle \\
&= \mathcal{N}c\big[\langle \hat{p}_z(1s) \mid \hat{p}_z(1s)\rangle + \big\langle(\hat{p}_x + i\hat{p}_y)(1s) \mid (\hat{p}_x + i\hat{p}_y)(1s)\big\rangle\big] \\
&= \mathcal{N}c\langle 1s|\hat{p}^2(1s)\rangle = \frac{1}{\zeta}c\zeta^2 = c\zeta.
\end{aligned}
$$

The last matrix element reads as

$$
\begin{aligned}
\langle \boldsymbol{\psi}|c(\boldsymbol{\sigma}\cdot\boldsymbol{\pi})\boldsymbol{\phi}\rangle &= \mathcal{N}c\left\langle \begin{pmatrix} 1s \\ 0 \end{pmatrix} \middle| (\boldsymbol{\sigma}\cdot\boldsymbol{\pi})^2 \begin{pmatrix} 1s \\ 0 \end{pmatrix} \right\rangle \\
&= \mathcal{N}c\left\langle \begin{pmatrix} 1s \\ 0 \end{pmatrix} \middle| \begin{pmatrix} \hat{p}^2 & 0 \\ 0 & \hat{p}^2 \end{pmatrix} \begin{pmatrix} 1s \\ 0 \end{pmatrix} \right\rangle = \mathcal{N}c\langle 1s \mid \hat{p}^2 1s\rangle = c\frac{1}{\zeta}\zeta^2 = c\zeta.
\end{aligned}
$$

Dirac's secular determinant

We have all the integrals needed and may now write the secular determinant corresponding to the matrix form of the Dirac equation:

$$
\begin{vmatrix} \langle \boldsymbol{\psi}|V\boldsymbol{\psi}\rangle - \varepsilon & \langle \boldsymbol{\psi}|c(\boldsymbol{\sigma}\cdot\boldsymbol{\pi})\boldsymbol{\phi}\rangle \\ \langle \boldsymbol{\phi}|c(\boldsymbol{\sigma}\cdot\boldsymbol{\pi})\boldsymbol{\psi}\rangle & \langle \boldsymbol{\phi}|(V - 2c^2))\boldsymbol{\phi}\rangle - \varepsilon \end{vmatrix} = 0,
$$

and after inserting the calculated integrals

$$
\begin{vmatrix} -Z\zeta - \varepsilon & c\zeta \\ c\zeta & -Z\zeta - 2c^2 - \varepsilon \end{vmatrix} = 0.
$$

Expanding the determinant gives the equation for the energy ε

$$
\varepsilon^2 + \varepsilon\big(2Z\zeta + 2c^2\big) + \big[Z\zeta\big(Z\zeta + 2c^2\big) - c^2\zeta^2\big] = 0.
$$

Hence, we get two solutions

$$
\varepsilon_{\pm} = -\big(c^2 + Z\zeta\big) \pm \sqrt{c^4 + \zeta^2 c^2}.
$$

Note that the square root is of the order of c^2 (in a.u.), and with the (unit) mass of the electron m_0, it is of the order of m_0c^2. Therefore, the minus sign before the square root corresponds to a solution with energy of the order of $-2m_0c^2$, while the plus sign corresponds to energy of the order of zero. Let us recall that we have shifted the energy scale in the Dirac equation and the last solution ε_+ (hereafter denoted by ε) is to be compared to the energy of the non-relativistic hydrogen-like atom

$$
\varepsilon = -\big(c^2 + Z\zeta\big) + \sqrt{c^4 + \zeta^2 c^2} = -\big(c^2 + Z\zeta\big) + c^2\sqrt{1 + \frac{\zeta^2}{c^2}}
$$

$$= -(c^2 + Z\zeta) + c^2\left(1 + \frac{\zeta^2}{2c^2} - \frac{\zeta^4}{8c^4} + \ldots\right)$$

$$= -Z\zeta + \frac{\zeta^2}{2} + \left(-\frac{\zeta^4}{8c^2} + \ldots\right). \tag{3.65}$$

Non-relativistic solution

If $c \to \infty$, i.e. we approach the non-relativistic limit, then $\varepsilon = -Z\zeta + \frac{\zeta^2}{2}$. Minimization of this energy with respect to ζ gives its optimum value $\zeta_{\text{opt}}^{\text{nonrel}} = Z$. In this way one recovers the result known from non-relativistic quantum mechanics (Appendix H) obtained in the variational approach to the hydrogen atom with the 1s orbital as a trial function.

3.4.2 RELATIVISTIC CONTRACTION OF ORBITALS

Minimizing the relativistic energy equation (3.65) leads to an equation for optimum $\zeta \equiv \zeta_{\text{opt}}^{\text{rel}}$:

$$\frac{d\varepsilon}{d\zeta} = 0 = -Z + \frac{1}{2}\left(c^4 + \zeta^2 c^2\right)^{-\frac{1}{2}} 2\zeta c^2 = -Z + \left(c^4 + \zeta^2 c^2\right)^{-\frac{1}{2}} \zeta c^2,$$

giving

$$\zeta_{\text{opt}}^{\text{rel}} = \frac{Z}{\sqrt{1 - \frac{Z^2}{c^2}}}.$$

The result differs remarkably from the non-relativistic value $\zeta_{\text{opt}}^{\text{nonrel}} = Z$, but approaches the non-relativistic value when $c \to \infty$. Note than the difference between the two values increases with atomic number Z, and that the relativistic exponent is always *larger* that its non-relativistic counter-part. This means that the relativistic orbital *decays faster* with the electron–nucleus distance and therefore

> the relativistic orbital 1s is *smaller* (contraction) than the corresponding non-relativistic one.

Let us see how it is for the hydrogen atom. In that case $\zeta_{\text{opt}}^{\text{rel}} = 1.0000266$ as compared to $\zeta_{\text{opt}}^{\text{nonrel}} = Z_H = 1$. And what about 1$s$ orbital of gold? For gold $\zeta_{\text{opt}}^{\text{rel}} = 96.68$, while $\zeta_{\text{opt}}^{\text{nonrel}} = Z_{Au} = 79$! Since for a heavy atom, the effective exponent of the atomic orbitals decreases when moving from the low-energy compact 1s orbital to higher-energy outer orbitals, this means that the most important relativistic orbital contraction occurs for the inner shells. The chemical properties of an atom depend on what happens to its outer shells (valence shell). Therefore, we may conclude that the relativistic corrections are expected to play a secondary role in chemistry.[34]

If we insert $\zeta_{\text{opt}}^{\text{rel}}$ in eq. (3.65) we obtain the minimum value of ε

$$\varepsilon_{\min} = -(c^2 + Z\zeta) + \sqrt{c^4 + \zeta^2}. \tag{3.66}$$

Since Z^2/c^2 is small with respect to 1, we may expand the square root in the Taylor series, $\sqrt{1-x} = 1 - \frac{1}{2}x - \frac{1}{8}x^2 - \cdots$. We obtain

$$\begin{aligned}\varepsilon_{\min} &= -c^2 + c^2\left\{1 - \left(\frac{1}{2}\right)\left(\frac{Z^2}{c^2}\right) - \frac{1}{8}\left(\frac{Z^2}{c^2}\right)^2 - \cdots\right\}\\ &= -\frac{Z^2}{2}\left(1 + \left(\frac{Z}{2c}\right)^2 + \cdots\right).\end{aligned} \tag{3.67}$$

In the case of the hydrogen atom ($Z = 1$) we have

$$\varepsilon_{\min} = -\frac{1}{2}\left(1 + \left(\frac{1}{2c}\right)^2 + \cdots\right), \tag{3.68}$$

where the first two terms shown give Darwin's exact result[35] (discussed earlier). Inserting $c = 137.036$ a.u. we obtain the hydrogen atom ground-state energy $\varepsilon = -0.5000067$ a.u., which agrees with Darwin's result.

3.5 LARGER SYSTEMS

The Dirac equation represents an approximation[36] and refers to a single particle. What happens with larger systems? Nobody knows, but the first idea is to construct the total Hamiltonian as a sum of the Dirac Hamiltonians for individual particles plus their Coulombic interaction (the *Dirac–Coulomb approximation*). This is practised routinely nowadays for atoms and molecules. Most often we use the mean-field approximation (see Chapter 8) with the modification that each of the one-electron functions represents a four-component bispinor. Another approach is extremely pragmatic, maybe too pragmatic: we perform the non-relativistic calculations with a pseudopotential that mimics what is supposed to happen in a relativistic case.

Dirac–Coulomb approximation

---

[34] We have to remember, however, that the relativistic effects also propagate from the inner shells to the valence shell through the orthogonalization condition, that has to be fulfilled after the relativistic contraction. This is why the gold valence orbital 6*s* shrinks, which has an immediate consequence in the relativistic shortening of the bond length in Au_2, which we cited at the beginning of this chapter.

[35] I.e. the exact solution to the Dirac equation for the electron in the external electric field produced by the proton.

[36] Yet it is strictly invariant with respect to the Lorentz transformation.

3.6 BEYOND THE DIRAC EQUATION...

How reliable is the presented relativistic quantum theory? The Dirac or Klein–Gordon equations, as is usual in physics, describe only some aspects of reality. The fact that both equations are invariant with respect to the Lorentz transformation indicates *only* that the space-time *symmetry* properties are described correctly. The physical machinery represented by these equations is not so bad, since several predictions have been successfully made (antimatter, electron spin, energy levels of the hydrogen atom). Yet, in the latter case an assumption of the external field $V = -\frac{Ze^2}{r}$ is a positively desperate step, which in fact is unacceptable in a fair relativistic theory for the proton and the electron (and not only of the electron in the external field of the nucleus). Indeed, the proton and the electron move. At a given time their distance is equal to r, but such a distance might be inserted into the Coulombic law if the speed of light were infinite, because the two particles would feel their positions instantaneously. Since, however, any perturbation by a positional change of a particle needs time to travel to the other particle, we have to use another distance somehow taking this into account (Fig. 3.3). The same pertains, of course, to any pair of particles in a many-body system (the so-called *retarded potential*).

retarded potential

There is certainly a need for a more accurate theory.

3.6.1 THE BREIT EQUATION

Breit constructed a many-electron relativistic theory that takes into account such a retarded potential in an approximate way. Breit explicitly considered only the electrons of an atom, nucleus of which (similar to Dirac theory) created only an external field for the electrons. This ambitious project was only partly success-

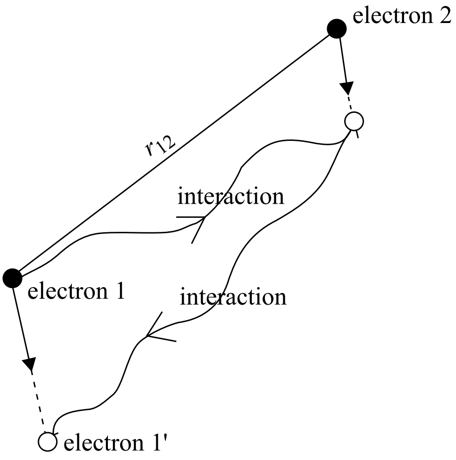


Fig. 3.3. Retardation of the interaction. The distance r_{12} of two particles in the interaction potential (as in Coulomb's law) is bound to represent an approximation, because we assume an instantaneous interaction. However, when the two particles catch sight of each other (which takes time) they are already somewhere else.

ful, because the resulting theory turned out to be approximate not only from the point of view of quantum theory (some interactions not taken into account) but also from the point of view of relativity theory (an *approximate* Lorentz transformation invariance).

Gregory Breit (1899–1981), American physicist, professor at the universities New York, Wisconsin, Yale, Buffalo. Breit with Eugene Wigner introduced the resonance states of particles, and with Condon created the proton–proton scattering theory.

For two electrons the Breit equation has the form (r_{12} stands for the distance between electron 1 and electron 2)

$$\left\{\hat{H}(1)+\hat{H}(2)+\frac{1}{r_{12}}-\frac{1}{2r_{12}}\left[\boldsymbol{\alpha}(1)\boldsymbol{\alpha}(2)+\frac{[\boldsymbol{\alpha}(1)\cdot\boldsymbol{r}_{12}][\boldsymbol{\alpha}(2)\cdot\boldsymbol{r}_{12}]}{r_{12}^2}\right]\right\}\Psi=E\Psi, \tag{3.69}$$

where (cf. eq. (3.54) with E replaced by the Hamiltonian)

$$\hat{H}(i)=q_i\phi(\boldsymbol{r}_i)+c\boldsymbol{\alpha}(i)\boldsymbol{\pi}(i)+\alpha_0(i)m_0c^2=-e\phi(\boldsymbol{r}_i)+c\boldsymbol{\alpha}(i)\boldsymbol{\pi}(i)+\alpha_0(i)m_0c^2$$

is the Dirac Hamiltonian for electron i pointed by vector $\boldsymbol{r}_i$, whereas the Dirac matrices for electron i: $\boldsymbol{\alpha}(i)=[\alpha_x(i),\alpha_y(i),\alpha_z(i)]$ and the corresponding operators $\pi_\mu(i)$ have been defined on p. 114, $\phi(\boldsymbol{r}_i)$ represents the scalar potential calculated at $\boldsymbol{r}_i$. The wavefunction Ψ represents a 16-component spinor (here represented by a square matrix of rank 4), because for each electron we would have the usual Dirac bispinor (four component) and the two-electron wavefunction depends on the Cartesian product of the components.[37]

The Breit Hamiltonian (in our example, for two electrons in an electromagnetic field) can be approximated by the following useful formula[38] known as the *Breit–Pauli Hamiltonian*

Breit–Pauli Hamiltonian

$$\hat{H}(1,2)=\hat{H}_0+\hat{H}_1+\cdots+\hat{H}_6, \tag{3.70}$$

where:

- $\hat{H}_0=\frac{\hat{p}_1^2}{2m_0}+\frac{\hat{p}_2^2}{2m_0}+V$ represents the familiar non-relativistic Hamiltonian.
- $\hat{H}_1=-\frac{1}{8m_0^3c^2}(\hat{p}_1^4+\hat{p}_2^4)$ comes from the velocity dependence of mass, more precisely from the Taylor expansion of eq. (3.38), p. 109, for small velocities.
- $\hat{H}_2=-\frac{e^2}{2(m_0c)^2}\frac{1}{r_{12}}[\hat{\boldsymbol{p}}_1\cdot\hat{\boldsymbol{p}}_2+\frac{\boldsymbol{r}_{12}\cdot(\boldsymbol{r}_{12}\cdot\hat{\boldsymbol{p}}_1)\hat{\boldsymbol{p}}_2}{r_{12}^2}]$ stands for the correction[39] that accounts in part for the above mentioned retardation. Alternatively, the term may be viewed as the interaction energy of two magnetic dipoles, each resulting from the orbital motion of an electron (*orbit–orbit term*).

orbit–orbit term

---

[37] In the Breit equation (3.69) the operators in {} act either by multiplying the 4 × 4 matrix Ψ by a function (i.e. each element of the matrix) or by a 4 × 4 matrix resulting from α matrices.

[38] H.A. Bethe, E.E. Salpeter, "*Quantum Mechanics of One- and Two-Electron Atoms*", Springer, 1977, p. 181.

[39] For non-commuting operators $\hat{\boldsymbol{a}}(\hat{\boldsymbol{a}}\cdot\hat{\boldsymbol{b}})\hat{\boldsymbol{c}}=\sum_{i,j=1}^{3}\hat{a}_i\hat{a}_j\hat{b}_j\hat{c}_i$.

spin–orbit coupling

- $\hat{H}_3 = \frac{\mu_B}{m_0 c}\{[\boldsymbol{\mathcal{E}}(\boldsymbol{r}_1) \times \hat{\boldsymbol{p}}_1 + \frac{2e}{r_{12}^3}\boldsymbol{r}_{12} \times \hat{\boldsymbol{p}}_2] \cdot \boldsymbol{s}_1 + [\boldsymbol{\mathcal{E}}(\boldsymbol{r}_2) \times \hat{\boldsymbol{p}}_2 + \frac{2e}{r_{12}^3}\boldsymbol{r}_{21} \times \hat{\boldsymbol{p}}_1] \cdot \boldsymbol{s}_2\}$ is the interaction energy of the electronic magnetic moments (resulting from the above mentioned orbital motion) with the spin magnetic dipole moments (*spin–orbit coupling*), μ_B stands for the Bohr magneton, and $\boldsymbol{\mathcal{E}}$ denotes the electric field vector. Since we have two orbital magnetic dipole moments and two spin orbital dipole moments, there are four spin–orbit interactions. The first term in square brackets stands for the spin–orbit coupling of the same electron, while the second term represents the coupling of the spin of one particle with the orbit of the second.

Darwin term

- $\hat{H}_4 = \frac{ie\hbar}{(2m_0c)^2}[\hat{\boldsymbol{p}}_1 \cdot \boldsymbol{\mathcal{E}}(\boldsymbol{r}_1) + \hat{\boldsymbol{p}}_2 \cdot \boldsymbol{\mathcal{E}}(\boldsymbol{r}_2)]$ is a non-classical term peculiar to the Dirac theory (also present in the one-electron Dirac Hamiltonian) called the *Darwin term*.

spin–spin

Fermi contact term

- $\hat{H}_5 = 4\mu_B^2\{-\frac{8\pi}{3}(\boldsymbol{s}_1 \cdot \boldsymbol{s}_2)\delta(\boldsymbol{r}_{12}) + \frac{1}{r_{12}^3}[\boldsymbol{s}_1 \cdot \boldsymbol{s}_2 - \frac{(\boldsymbol{s}_1 \cdot \boldsymbol{r}_{12})(\boldsymbol{s}_2 \cdot \boldsymbol{r}_{12})}{r_{12}^2}]\}$ corresponds to the spin dipole moment interactions of the two electrons (*spin–spin term*). The first term is known as the *Fermi contact term*, since it is non-zero only when the two electrons touch one another (see Appendix E, p. 951), whereas the second term represents the classical *dipole–dipole interaction* of the two electronic spins (cf. the multipole expansion in Appendix X, p. 1038 and Chapter 13), i.e. the interaction of the two spin magnetic moments of the electrons (with the factor 2, according to eq. (3.62), p. 122).

Zeeman term

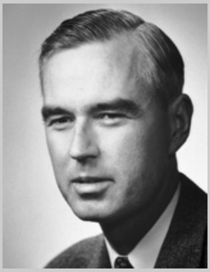
- $\hat{H}_6 = 2\mu_B[\boldsymbol{H}(\boldsymbol{r}_1) \cdot \boldsymbol{s}_1 + \boldsymbol{H}(\boldsymbol{r}_2) \cdot \boldsymbol{s}_2] + \frac{e}{m_0c}[\boldsymbol{A}(\boldsymbol{r}_1) \cdot \hat{\boldsymbol{p}}_1 + \boldsymbol{A}(\boldsymbol{r}_2) \cdot \hat{\boldsymbol{p}}_2]$ is known as the *Zeeman interaction*, i.e. the interaction of the spin (the first two terms) and the orbital (the second two terms) electronic magnetic dipole moments with the external magnetic field $\boldsymbol{H}$ (cf. eq. (3.62)).

The terms listed above are of prime importance in the theory of the interaction of matter with the electromagnetic field (e.g., in nuclear magnetic resonance).

3.6.2 A FEW WORDS ABOUT QUANTUM ELECTRODYNAMICS (QED)

The Dirac and Breit equations do not account for several subtle effects.[40] They are predicted by quantum electrodynamics, a many-particle theory.

Willis Eugene Lamb (b. 1913), American physicist, professor at Columbia, Stanford, Oxford, Yale and Tucson universities. He received the Nobel Prize in 1955 "*for his discoveries concerning the fine structure of the hydrogen spectrum*".



---

[40] For example, an effect observed in spectroscopy for the first time by Willis Lamb.

The QED energy may be conveniently developed in a series of $\frac{1}{c}$:

- in zero order we have the non-relativistic approximation (solution to the Schrödinger equation);
- there are no first order terms;
- the second order contains the Breit corrections;
- the third and further orders are called the *radiative corrections*.

radiative corrections

Radiative corrections

The radiative corrections include:

- *Interaction with the vacuum* (Fig. 3.4.a). According to modern physics the perfect vacuum does not just represent nothing. The electric field of the vacuum itself fluctuates about zero and these instantaneous fluctuations influence the motion of any charged particle. When a strong electric field operates in a vacuum, the latter undergoes a polarization (*vacuum polarization*), which means a *spontaneous creation of matter*, more specifically, of particle-antiparticle pairs.

vacuum polarization

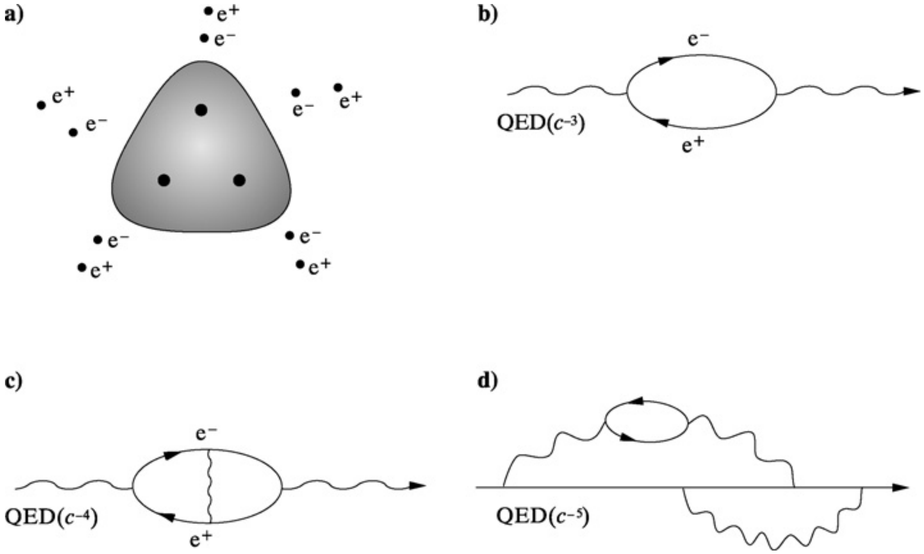


Fig. 3.4. (a) The electric field close to the proton (composed of three quarks) is so strong that it *creates matter and antimatter* (shown as electron–positron pairs). The three quarks visible in scattering experiments represent the valence quarks. (b) One of the radiative effects in the QED correction of the c^{-3} order (see Table 3.1). The pictures show the sequence of the events from left to the right. A photon (wavy line on the left) polarizes the vacuum and an electron–positron pair (solid lines) is created, and the photon vanishes. Then the created particles annihilate each other and a photon is created. (c) A similar event (of the c^{-4} order in QED), but during the existence of the electron–positron pair the two particles interact by exchange of a photon. (d) An electron (horizontal solid line) emits a photon, which creates an electron–positron pair, that annihilates producing another photon. Meanwhile the first electron emits a photon, then first absorbs the photon from the annihilation, and afterwards the photon emitted by itself earlier. This effect is of the order c^{-5} in QED.

The probability of this event (per unit volume and time) depends[41] (Fig. 3.4.a–d) on the particle mass m and charge q:

$$w = \frac{\mathcal{E}^2}{c\pi^2}\sum_{n=1}^{\infty}\frac{1}{n^2}\exp\left(-\frac{n\pi m^2}{|q\mathcal{E}|}\right), \tag{3.71}$$

where $\mathcal{E}$ is the electric field intensity. The creation of such pairs in a static electric field has never yet been observed, because we cannot yet provide sufficient $\mathcal{E}$. Even for the electron on the first Bohr orbit, the $|q\mathcal{E}|$ is small compared to m^2 (however, for smaller distances the exponent may be much smaller).

creation of matter

- *Interaction with virtual photons*. The electric field influences the motion of electron. What about its own electric field? Does it influence its motion as well? The latter effect is usually modelled by allowing the electron to emit photons and then to absorb them (*"virtual photons"*)[42] (Fig. 3.4.d).

The QED calculations performed to date have been focused on the energy. The first calculations of atomic susceptibilities (helium) within an accuracy including the c^{-2} terms were carried out independently[43] by Pachucki and Sapirstein[44] and by Cencek and coworkers,[45] and with accuracy up to c^{-3} (with estimation of the c^{-4} term) by Łach and coworkers (see Table 3.1). To get a flavour of what subtle effects may be computed nowadays, Table 3.1 shows the components of the first ionization energy and of the dipole polarizability (see Chapter 12) of the helium atom.

Comments to Table 3.1

- $\hat{H}_0$ denotes the result obtained from an accurate solution of the Schrödinger equation (i.e. the non-relativistic and finite nuclear mass theory). Today the solution of the equation could be obtained with greater accuracy than reported here. Imagine, that here the theory is limited by the *precision of our knowledge of the helium atom mass*, which is "only" 12 significant figures.
- The effect of the non-zero size of the nucleus is small, it is practically never taken into account in computations. If we enlarged the nucleus to the size of an apple, the first Bohr orbit would be 10 km from the nucleus. And still (sticking to our analogy) the electron is able to distinguish a point from an apple? Not quite. It sees the (tiny) difference because the electron knows the region close to the nucleus: it is there that it resides most often. Anyway the theory is able to compute such a tiny effect.

---

[41] C. Itzykson, J.-B. Zuber, *"Quantum Field Theory"*, McGraw-Hill, 1985, p. 193.

[42] As remarked by Richard Feynman (see Additional Literature in the present chapter, p. 140) for unknown reasons physics is based on the interaction of objects of spin $\frac{1}{2}$ (like electrons or quarks) mediated by objects of spin 1 (like photons, gluons or W particles).

[43] With identical result, that increases enormously the confidence one may place in such results.

[44] K. Pachucki, J. Sapirstein, *Phys. Rev. A* 63 (2001) 12504.

[45] W. Cencek, K. Szalewicz, B. Jeziorski, *Phys. Rev. Letters* 86 (2001) 5675.

Table 3.1. Contributions of various physical effects (non-relativistic, Breit, QED and beyond QED) to the ionization energy and the dipole polarizability α of the helium atom as well as comparison with the experimental values (all quantities in atomic units, i.e. $e = 1$, $\hbar = 1$, $m_0 = 1$, where m_0 denotes the rest mass of electron). The first column gives the symbol of the term in the Breit–Pauli Hamiltonian (3.70) as well as of the QED corrections given order by order (first corresponding to the electron–positron vacuum polarization (QED), then, beyond quantum electrodynamics, to other particle–antiparticle pairs (non-QED): μ, π, ...) split into several separate effects. The second column contains a short description of the effect. The estimated error (third column) is given in parentheses in the units of the last figure reported

Term	Physical interpretation	Ionization energy [MHz]	α [a.u.$\times 10^{-6}$][1]
$\hat{H}_0$	**Schrödinger equation**	**5 945 262 288.62(4)**	**1 383 809.986(1)**
δ	**non-zero size of the nucleus**	**−29.55(4)**	**0.022(1)**
$\hat{H}_1$	**p^4 term**	**1 233 305.45(1)**	**−987.88(1)**
$\hat{H}_2$(el-el)	electron–electron retardation (Breit interaction)	48 684.88(1)	−23.219(1)
$\hat{H}_2$(el-n)	electron–nucleus retardation (Breit interaction)	319.16(1)	−0.257(3)
$\hat{H}_2$	**Breit interaction (total)**	**49 004.04(1)**	**−23.476(3)**
$\hat{H}_3$	**spin–orbit**	0	0
$\hat{H}_4$(el-el)	electron–electron Darwin term	117 008.83(1)	−66.083(1)
$\hat{H}_4$(el-n)	electron–nucleus Darwin term	−1 182 100.99(1)	864.85(2)
$\hat{H}_4$	**Darwin term (total)**	**−1 065 092.16(1)**	**798.77(2)**
$\hat{H}_5$	**spin–spin (total)**	**−234 017.66(1)**	**132.166(1)**
$\hat{H}_6$	**spin-field**	0	0
QED(c^{-3})	vacuum polarization correction to electron–electron interaction	−72.48(1)	0.41(1)
QED(c^{-3})	vacuum polarization correction to electron–nucleus interaction	1 463.00(1)	−1.071(1)
QED(c^{-3})	Total vacuum polarization in c^{-3} order	1 390.52(1)	−1.030(1)
QED(c^{-3})	**vac.pol. + other c^{-3} QED correction**	**−40 483.98(5)**	**30.66(1)**
QED(c^{-4})	vacuum polarization	12.26(1)	0.009(1)
QED(c^{-4})	**Total c^{-4} QED correction**	**−834.9(2)**	**0.56(22)**
QED-h.o.	**Estimation of higher order QED correction**	**84(42)**	**−0.06(6)**
non-QED	**contribution of virtual muons, pions, etc.**	**0.05(1)**	**−0.004(1)**
$\sum$	**Theory (total)**	**5 945 204 223(42)**[2]	**1 383 760.79(23)**
	Experiment	**5 945 204 238(45)**[3]	**1 383 791(67)**[4]

[1] G. Łach, B. Jeziorski, K. Szalewicz, *Phys. Rev. Letters* 92 (2004) 233001.

[2] G.W.F. Drake, W.C. Martin, *Can. J. Phys*. 76 (1998) 679; V. Korobov, A. Yelkhovsky, *Phys. Rev. Letters* 87 (2001) 193003.

[3] K.S.E. Eikema, W. Ubachs, W. Vassen, W. Hogervorst, *Phys. Rev. A* 55 (1997) 1866.

[4] F. Weinhold, *J. Phys. Chem.* 86 (1982) 1111.

- *The term* p^4*and the total Darwin effect nearly cancel each other for unclear reasons*. This cancellation is being persistently confirmed also in other systems. Mysteriously enough, this pertains not only to the ionization energy, but also to the polarizability.
- After the above mentioned cancellation (of p^4 and Darwin terms), retardation becomes one of the most important relativistic effects. As seen from the Table, the effect is about a 100 times larger (both for the ionization energy and the polarizability) for the electron–electron retardation than for that of the nucleus–electron. This is quite understandable, because the nucleus represents a "massive rock" (it is about 7000 times heavier) in comparison to an electron, it moves slowly and in the nucleus–electron interaction only the electron contributes to the retardation effect. Two electrons make the retardation much more serious.
- Term $\hat{H}_3$ (spin–orbit coupling) is equal to zero for symmetry reasons (for the ground state).
- In the Darwin term, the nucleus–electron vs electron–electron contribution have reversed magnitudes: about 1 : 10 as compared to 100 : 1 in retardation). Again this time it seems intuitively correct. We have the sum of the particle–particle terms in the Hamiltonian $\hat{H}_4 = \frac{ie\hbar}{(2m_0c)^2}[\hat{\boldsymbol{p}}_1 \cdot \boldsymbol{\mathcal{E}}(\boldsymbol{r}_1) + \hat{\boldsymbol{p}}_2 \cdot \boldsymbol{\mathcal{E}}(\boldsymbol{r}_2)]$, where $\boldsymbol{\mathcal{E}}$ means an electric field created by two other particles on the particle under consideration. Each of the terms is proportional to $\nabla_i \nabla_i V = \Delta_i V = 4\pi q_i \delta(\boldsymbol{r}_i)$, where δ is the δ Dirac delta function (Appendix E, p. 951), and q_i denotes the charge of the particle "i". The absolute value of the nuclear charge is twice the electron charge.
- In term $\hat{H}_5$ the spin–spin relates to the electron–electron interaction because the helium nucleus has spin angular momentum of 0.
- The Coulombic interactions are modified by the polarization of vacuum (similar to the weaker interaction of two charges in a dielectric medium). Table 3.1 reports such corrections[46] to the electron–electron and the electron–nucleus interactions [QED(c^{-3})] taking into account that electron–positron pairs jump out from the vacuum. One of these effects is shown in Fig. 3.4.a. As seen from Table 3.1, the nucleus polarizes the vacuum much more easily (about ten times more that the polarization by electrons). Once again the larger charge of the nucleus makes the electric field larger and qualitatively explains the effect. Note that the QED corrections (corresponding to e-p creation) decrease quickly with their order. One of such higher order corrections is shown in Fig. 3.4.d.
- What about the creation of other (than e-p) particle-antiparticle pairs from the vacuum? From (3.71) we see that the larger the rest mass the more difficult it is to squeeze out the corresponding particle-antiparticle pair. And yet, we have some tiny effect (see non-QED entry) corresponding to the creation of such pairs as muon-antimuon (μ), pion-antipion[47] (π), etc. This means that the helium atom is composed of the nucleus and the two electrons only, when we look

---

[46] However, these effects represent a minor fraction of the total QED(c^{-3}) correction.

[47] Pions are π mesons, the subnuclear particles with mass comparable to that of the muon, a particle about 200 times more massive than an electron. Pions were discovered in 1947 by C.G. Lattes, G.S.P. Occhialini and C.F. Powell.

at it within a certain approximation. To tell the truth, the atom contains also photons, electrons, positrons, muons, pions, and whatever you wish, but with smaller and smaller probability. All that *silva rerum* has only a minor effect of the order of something like the seventh significant figure (both for the ionization potential and for the polarizability).

Summary

The beginning of the twentieth century has seen the birth and development of two revolutionary theories: relativity and quantum mechanics. These two theories turned out to be incompatible, and attempts were made to make them consistent. This chapter consists of two interrelated parts:

- introduction of the elements of relativity theory, and
- attempts to make quantum theory consistent with relativity (relativistic quantum mechanics).

ELEMENTS OF SPECIAL RELATIVITY THEORY

- If experiments are to be described in the same way in two laboratories that move with respect to the partner laboratory with constant velocities v and $-v$, respectively, then the *apparent forces have to vanish*. The same event is described in the two laboratories (by two observers) in the corresponding coordinate system (in one the event happens at coordinate x and time t, in the second – at x' and t'). A sufficient condition that makes the apparent forces vanish is based on linear dependence: $x' = Ax + Bt$ and $t' = Cx + Dt$, where A, B, C, D denote some constants.
- In order to put both observers on the same footing, we have to have $A = D$.
- The Michelson–Morley experiment has shown that each of the observers will note that in the partner's laboratory there is a *contraction* of the dimension pointing to the partner. As a consequence there is a time *dilation*, i.e. each of the observers will note that time flows slower in the partner's laboratory.
- Einstein assumed that in spite of this, any of the observers will measure the same speed of light, c, in his coordinate system.
- This leads to the *Lorentz transformation* that says where and when the two observers see the same event. The Lorentz transformation is especially simple after introducing the Minkowski space (x, ct):

$$\begin{bmatrix} x' \\ ct' \end{bmatrix} = \frac{1}{\sqrt{1-\frac{v^2}{c^2}}} \begin{Bmatrix} 1 & -\frac{v}{c} \\ -\frac{v}{c} & 1 \end{Bmatrix} \begin{bmatrix} x \\ ct \end{bmatrix}.$$

  None of the two coordinate systems is privileged (relativity principle).
- Finally, we derived Einstein's formula $E_{\text{kin}} = mc^2$ for the kinetic energy of a body with mass m (this depends on its speed with respect to the coordinate system where the mass is measured).

RELATIVISTIC QUANTUM DYNAMICS

- Fock, Klein and Gordon found the total energy for a particle using the Einstein formula for kinetic energy $E_{\text{kin}} = mc^2$, adding the potential energy and introducing the momentum[48] $p = mv$. After introducing an external electromagnetic field (characterized by the vector potential $\boldsymbol{A}$ and the scalar potential ϕ) they obtained the following relation among operators

$$\left[\frac{i\hbar\frac{\partial}{\partial t} - q\phi}{c}\right]^2 - \left[\left(-i\hbar\nabla - \frac{q}{c}\boldsymbol{A}\right)^2 + m_0^2c^2\right] = 0,$$

where m_0 denotes the rest mass of the particle.

- Paul Dirac factorized the left hand side of this equation by treating it as the difference of squares. This gave two continua of energy separated by a gap of width $2m_0c^2$. Dirac assumed that the lower (negative energy) continuum is fully occupied by electrons ("vacuum"), while the upper continuum is occupied by the single electron (our particle). If we managed to excite an electron from the lower continuum to the upper one, then in the upper continuum we would see an electron, while the hole in the lower continuum would have the properties of a positive electron (positron). This corresponds to the creation of the electron–positron pair from the vacuum.
- The Dirac equation for the electron has the form:

$$\left(i\hbar\frac{\partial}{\partial t}\right)\Psi = \left(q\phi + c\sum_{\mu=x,y,z}\alpha_\mu\pi_\mu + \alpha_0 m_0c^2\right)\Psi,$$

  where π_μ in the absence of magnetic field is equal to the momentum operator $\hat{p}_\mu$, $\mu = x, y, z$, while α_μ stand for the square *matrices* of the rank 4, which are related to the Pauli matrices (cf. introduction of spin, Chapter 1). In consequence, the wavefunction Ψ has to be a four-component vector composed of square integrable functions (*bispinor*).
- The Dirac equation demonstrated "pathological" behaviour when a numerical solution was sought. The very reason for this was the decoupling of the electron and positron equations. The exact separation of the negative and positive energy continua has been demonstrated by Barysz and Sadlej, but it leads to a more complex theory. Numerical troubles are often removed by an *ad hoc* assumption called *kinetic balancing*, i.e. fixing a certain relation among the bispinor components. By using this relation we prove that there are two *large* and two *small* (smaller by a factor of about $\frac{v}{2c}$) components of the bispinor.[49]
- The kinetic balance can be used to eliminate the small components from the Dirac equation. Then, the assumption $c = \infty$ (non-relativistic approximation) leads to the *Schrödinger equation* for a single particle.
- The Dirac equation for a particle in the electromagnetic field contains the interaction of the spin magnetic moment with the magnetic field. In this way spin angular momentum appears in the Dirac theory in a natural way (as opposed to the non-relativistic case, where it has had to be postulated).
- The problem of an electron in the external electric field produced by the nucleus (the hydrogen-like atom) has been solved exactly. It turned out that the relativistic corrections are important only for systems with heavy atoms.
- It has been demonstrated in a step-by-step calculation how to obtain an approximate solution of the Dirac equation for the hydrogen-like atom. One of the results is that the relativistic orbitals are contracted compared to the non-relativistic ones.

---

[48] They wanted to involve the momentum in the formula to be able to change the energy expression to an operator ($\boldsymbol{p} \to \hat{\boldsymbol{p}}$) according to the postulates of quantum mechanics.

[49] For solutions with negative energies this relation is reversed.

- Finally, the Breit equation has been given. The equation goes beyond the Dirac model, by taking into account the *retardation* effects. The Pauli–Breit expression for the Breit Hamiltonian contains several easily interpretable physical effects.
- Quantum electrodynamics (QED) provides an even better description of the system by adding *radiative* effects that take into account the interaction of the particles with the vacuum.

Main concepts, new terms

apparent forces (p. 93)
inertial system (p. 95)
Galilean transformation (p. 96)
Michelson–Morley experiment (p. 96)
length contraction (p. 100)
Lorentz transformation (p. 100)
velocity addition law (p. 103)
relativity principle (p. 104)
Minkowski space-time (p. 104)
time dilation (p. 105)
relativistic mass (p. 107)
Einstein equation (p. 108)
Klein–Gordon equation (p. 109)
Dirac electronic sea (p. 111)
Dirac vacuum (p. 112)
energy continuum (p. 112)
positron (p. 113)
anticommutation relation (p. 114)
Dirac equation (p. 115)
spinors and bispinors (p. 115)
kinetic balance (p. 119)
electron spin (p. 122)
Darwin solution (p. 123)
contraction of orbitals (p. 128)
retarded potential (p. 130)
Breit equation (p. 131)
spin–orbit coupling (p. 132)
spin–spin coupling (p. 132)
Fermi contact term (p. 132)
Zeeman effect (p. 132)
vacuum polarization (p. 133)
particle–antiparticle creation (p. 134)
virtual photons (p. 134)

From the research front

Dirac theory within the mean field approximation (Chapter 8) is routinely applied to molecules and allows us to estimate the relativistic effects even for large molecules. In the computer era, this means, that there are commercial programs available that allow anybody to perform relativistic calculations.

Much worse is the situation with more accurate calculations. The first estimation for molecules of relativistic effects beyond the Dirac approximation has been carried out by Janos Ladik[50] and then by Jeziorski and Kołos[51] while the first calculation of the interaction with the vacuum for molecules was done by Bukowski et al.[52] Besides the recent computation of the Lamb shift for the water molecule,[53] not much has been computed in this area.

Ad futurum...

In comparison with typical chemical phenomena, the relativistic effects in almost all instances, remain of marginal significance for biomolecules or for molecules typical of tradi-

---

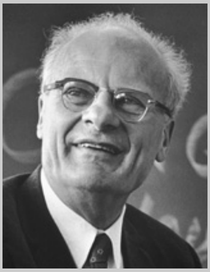
[50] J. Ladik, *Acta Phys. Hung.* 10 (1959) 271.

[51] The calculations were performed for the hydrogen molecular ion H_2^+, B. Jeziorski, W. Kołos, *Chem. Phys. Letters* 3 (1969) 677.

[52] R. Bukowski, B. Jeziorski, R. Moszyński, W. Kołos, *Int. J. Quantum Chem*. 42 (1992) 287.

[53] P. Pyykkö, K.G. Dyall, A.G. Császár, G. Tarczay, O.L. Polyansky, J. Tennyson, *Phys. Rev. A* 63 (2001) 24502.

Hans Albrecht Bethe (1906–2005), American physicist, professor at Cornell University, student of Arnold Sommerfeld. Bethe contributed to many branches of physics, e.g., crystal field theory, interaction of matter with radiation, quantum electrodynamics, structure and nuclear reactions of stars (for the latter achievement he received the Nobel Prize in 1967).



tional organic chemistry. In inorganic chemistry, these effects could however be much more important. Probably the Dirac–Coulomb theory combined with the mean field approach will for a few decades remain a satisfactory standard for the vast majority of researchers. At the same time there will be theoretical and computational progress for small molecules (and for atoms), where Dirac theory will be progressively replaced by quantum electrodynamics.

Additional literature

H. Bethe, E. Salpeter, "Quantum Mechanics of One- and Two-Electron Atoms", Springer, Berlin, 1957.

This book is absolutely exceptional. It is written by excellent specialists in such a competent way and with such care (no misprints), that despite the lapse of many decades it remains the fundamental and best source.

I.M. Grant, H.M. Quiney, "Foundations of the Relativistic Theory of Atomic and Molecular Structure", *Adv. At. Mol. Phys.*, 23 (1988) 37.

Very good review.

L. Pisani, J.M. André, M.C. André, E. Clementi, *J. Chem. Educ.*, **70**, 894–901 (1993), also J.M. André, D.H. Mosley, M.C. André, B. Champagne, E. Clementi, J.G. Fripiat, L. Leherte, L. Pisani, D. Vercauteren, M. Vracko, Exploring Aspects of Computational Chemistry: *Vol. I, Concepts*, Presses Universitaires de Namur, pp. 150–166 (1997), Vol. II, Exercises, Presses Universitaires de Namur, p. 249–272 (1997).

Fine article, fine book, written clearly, its strength is also in very simple examples of the application of the theory.

R.P. Feynman, "QED – The Strange Theory of Light and Matter", Princeton University Press, Princeton, 1988.

Excellent book written by one of the celebrities of our times in the style "*quantum electrodynamics not only for poets*".

Questions

1. In the Lorentz transformation the two coordinate systems:
   a) are both at rest; b) move with the same velocity; c) are related also by Galilean transformation; d) have x' and t' depending linearly on x and t.

2. The Michelson–Morley experiment has shown that when an observer in the coordinate system O measures a length in O' (both coordinate systems fly apart; $v' = -v$), then he obtains:

a) the same result that is obtained by an observer in O'; b) contraction of lengths along the direction of the motion; c) expansion of lengths along the direction of the motion; d) contraction of lengths in any direction.

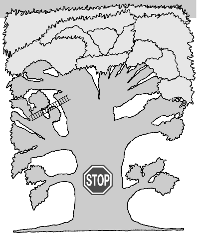
3. An observer in O measures the times a phenomenon takes in O and O' (both coordinate systems fly apart; $v' = -v$):
   a) the time of the phenomenon going on in O will be shorter; b) time goes with the same speed in O'; c) time goes more slowly in O' only if $|v| > \frac{c}{2}$; d) time goes more slowly in O' only if $|v| < \frac{c}{2}$.
4. In the Minkowski space, the distance of any event from the origin (both coordinate systems fly apart; $v' = -v$) is:
   a) equal to vt; b) equal to ct; c) the same for observers in O and in O'; d) equal to 0.
5. A bispinor represents:
   a) a two-component vector with functions as components; b) a two-component vector with complex numbers as components; c) a four-component vector with square integrable functions as components; d) a scalar square integrable function.
6. Non-physical results of numerical solutions to the Dirac equation appear because:
   a) the Dirac sea is neglected; b) the electron and positron have the same energies; c) the electron has kinetic energy equal to its potential energy; d) the electron has zero kinetic energy.
7. The Schrödinger equation can be deduced from the Dirac equation under the assumption that:
   a) $v = c$; b) v/c is small; c) all components of the bispinor have equal length; d) the magnetic field is zero.
8. In the Breit equation there is an approximate cancellation of:
   a) the retardation effect with the non-zero size of the nucleus effect; b) the retardation effect electron–electron with that of electron–nucleus; c) the spin–spin effect with the Darwin term; d) the Darwin term with the p^4 term.
9. Dirac's hydrogen atom orbitals when compared to Schrödinger's are:
   a) more concentrated close to the nucleus, but have a larger mean value of r; b) have a larger mean value of r; c) more concentrated close to the nucleus; d) of the same size, because the nuclear charge has not changed.
10. The Breit equation: a) is invariant with respect to the Lorentz transformation; b) takes into account the interaction of the magnetic moments of electrons resulting from their orbital motion; c) neglects the interaction of the spin magnetic moments; d) describes only a single particle.

Answers

1d, 2b, 3a, 4c, 5c, 6a, 7b, 8d, 9c, 10b

Chapter 4

EXACT SOLUTIONS – OUR BEACONS


WHERE ARE WE?

Where are we?

We are in the middle of the TREE trunk.

An example

Two chlorine atoms stay together – they form the molecule Cl_2. If we want to know its main mechanical properties, it would very quickly be seen that the two atoms have an equilibrium distance and any attempt to change this (in either direction) would be accompanied by work to be done. It looks like the two atoms are coupled together by a sort of spring. If one assumes that the spring satisfies Hooke's law,[1] the system is equivalent to a harmonic oscillator. If we require that no rotation in space of such a system is allowed, *the corresponding Schrödinger equation has the exact*[2] *analytical solution*.

What is it all about

---

[1] *And* if we limit ourselves to small displacements, see p. 239.

[2] Exact means ideal, i.e. without any approximation.

Short descriptions of exact solutions to the Schrödinger equations for the above model systems will be given.

Why is this important?

The Schrödinger equation is nowadays quite easy to solve with *a desired accuracy* for many systems. There are only a few systems for which the *exact* solutions are possible. These problems and solutions play an extremely important role in physics, since they represent kind of beacons for our navigation in science, when we deal with complex systems. Real systems may often be approximated by those for which exact solutions exist. For example, a real diatomic molecule is an extremely complex system, difficult to describe in detail and certainly does not represent a harmonic oscillator. Nevertheless, the main properties of diatomics follow from the simple harmonic oscillator model. When a chemist or physicist has to describe a complex system, he always first tries to simplify the problem,[3] to make it similar to one of the simple problems described in the present chapter. Thus, from the beginning *we know the (idealized) solution*. This is of prime importance when discussing the (usually complex) solution to a higher level of accuracy. If this higher level description differs dramatically from that of the idealized one, most often this indicates that there is an error in our calculations and nothing is more urgent than to find and correct it.

What is needed?

- The postulates of quantum mechanics (Chapter 1, necessary).
- Separation of the centre of mass motion (Appendix I on p. 971, necessary).
- Operator algebra (Appendix B on p. 895, necessary).

In the present textbook we assume that the reader knows most of the problems described in the present chapter from a basic course in quantum chemistry. This is why the problems are given in short – only the most important results, without derivation, are reported. On the other hand, such a presentation, in most cases, will be sufficient for our goals.

Classical works

The hydrogen atom problem was solved by Werner Heisenberg in "*Über quantentheoretischen Umdeutung kinematischer und mechanischer Beziehungen*" published in *Zeitschrift für Physik*, 33 (1925) 879. ★ Erwin Schrödinger arrived at an equivalent picture within his wave mechanics in "*Quantisierung als Eigenwertproblem. I.*" published in *Annalen der Physik*, 79 (1926) 361. Schrödinger also gave the solution for the harmonic oscillator in a paper (under almost same title) which appeared in *Annalen der Physik*, 79 (1926) 489. ★ The Morse

---

[3] One of the cardinal strategies of science, when we have to explain a strange phenomenon, is first to simplify the system and create a model or series of models (more and more simplified descriptions) that still exhibit the phenomenon. The first model to study should be as simple as possible, because it will shed light on the main machinery.

oscillator problem was solved by Philip McCord Morse in "*Diatomic Molecules According to the Wave Mechanics. II. Vibrational Levels*" in *Physical Review*, 34 (1929) 57.[4] ★ The tunnelling effect was first considered by Friedrich Hund in "*Zur Deutung der Molekelspektren*" published in *Zeitschrift für Physik*, 40 (1927) 742. ★ The Schrödinger equation for the harmonium[5] was first solved by Sabre Kais, Dudley R. Herschbach and Raphael David Levine in "*Dimensional Scaling as a Symmetry Operation*", which appeared in the *Journal of Chemical Physics*, 91 (1989) 7791.

4.1 FREE PARTICLE

The potential energy for a free particle is a constant (taken arbitrarily as zero): $V = 0$ and, therefore, energy E represents the kinetic energy only. The Schrödinger equation takes the form

$$-\frac{\hbar^2}{2m}\frac{\mathrm{d}^2\Psi}{\mathrm{d}x^2} = E\Psi$$

or in other words

$$\frac{\mathrm{d}^2\Psi}{\mathrm{d}x^2} + \kappa^2\Psi = 0$$

with $\kappa^2 = \frac{2mE}{\hbar^2}$. The constant κ in *this situation*[6] is a real number.

The special solutions to this equation are $\exp(i\kappa x)$ and $\exp(-i\kappa x)$. Their linear combination with arbitrary complex coefficients A' and B' represents the general solution:

$$\boxed{\Psi = A' \exp(i\kappa x) + B' \exp(-i\kappa x). \qquad (4.1)}$$

This is a de Broglie wave of wave length $\lambda = \frac{2\pi}{\kappa}$. Function $\exp(i\kappa x)$ represents the eigenfunction of the momentum operator:

$$\hat{p}_x \exp(i\kappa x) = -i\hbar\frac{\mathrm{d}}{\mathrm{d}x}\exp(i\kappa x) = -i\hbar i\kappa \exp(i\kappa x) = \kappa\hbar \exp(i\kappa x).$$

For eigenvalue $\hbar\kappa > 0$ the eigenfunction $\exp(i\kappa x)$ describes a particle moving towards $+\infty$. Similarly, $\exp(-i\kappa x)$ corresponds to a particle of the same energy, but moving in the opposite direction. The function $\Psi = A' \exp(i\kappa x) + B' \exp(-i\kappa x)$ is a superposition of these two states. A measurement of the momentum can give only two values: $\kappa\hbar$ with probability proportional to $|A'|^2$ or $-\kappa\hbar$ with probability proportional to $|B^{'}|^2$.

---

[4] Note the spectacular speed at which the scholars worked.

[5] A harmonic model of the helium atom.

[6] The kinetic energy is always positive.

4.2 PARTICLE IN A BOX

4.2.1 BOX WITH ENDS

The problem pertains to a single particle in a potential (Fig. 4.1.a)

$$\begin{aligned} V(x) &= 0 && \text{for } 0 \leqslant x \leqslant L, \\ V(x) &= \infty && \text{for other } x. \end{aligned}$$

Just because the particle will never go outside the section $0 \leqslant x \leqslant L$, therefore, the value of the wave function outside the section is equal to 0. It remains to find the function in $0 \leqslant x \leqslant L$.

Let us write down the Schrödinger equation for $0 \leqslant x \leqslant L$ with the Hamiltonian containing the kinetic energy only (since $V = 0$, one has $E \geqslant 0$)

$$-\frac{\hbar^2}{2m}\frac{d^2\Psi}{dx^2} = E\Psi. \tag{4.2}$$

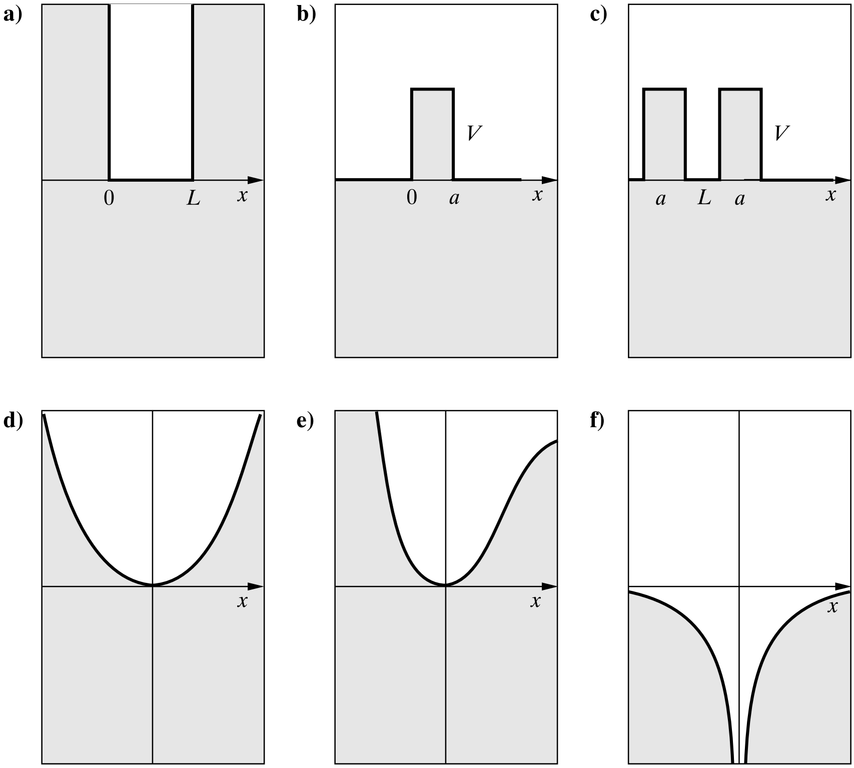


Fig. 4.1. The potential energy functions for a) particle in a box, b) single barrier, c) double barrier, d) harmonic oscillator, e) Morse oscillator, f) hydrogen atom.

The solution to this equation is given by (4.1), which may also be written as

$$\Psi = A\sin\kappa x + B\cos\kappa x \tag{4.3}$$

with

$$\kappa^2 = \frac{2mE}{\hbar^2}. \tag{4.4}$$

Now, the key is to recall (p. 74, Fig. 2.5), that the wave function has to be continuous and, therefore, two conditions have to be fulfilled: 1) $\Psi = 0$ for $x = 0$ and 2) $\Psi = 0$ for $x = L$. The first condition immediately gives $B = 0$, the second in this situation is equivalent to $\kappa L = n\pi$, for $n = 0, 1, \ldots$. *From this follows energy quantization*, because κ contains energy E. One obtains, therefore, the following solution (a standing wave[7]):

$$E_n = \frac{n^2h^2}{8mL^2}, \quad n = 1, 2, 3, \ldots, \tag{4.5}$$

$$\Psi_n = \sqrt{\frac{2}{L}}\sin\frac{n\pi}{L}x, \quad n = 1, 2, 3, \ldots, \tag{4.6}$$

because $n = 0$ has to be excluded as leading to the wave function equal to zero everywhere, while $n < 0$ may be safely excluded as leading to the same wave functions as[8] $n > 0$. Fig. 4.2 shows the wave functions for $n = 1, 2, 3$.

2D rectangular box

Let us consider a rectangular box (Fig. 4.3) with sides L_1 and L_2 and $V = 0$ inside and $V = \infty$ outside. We very easily obtain the solution to the Schrödinger equation after a straightforward separation of variables x and y leading to the two 1D Schrödinger equations.

The energy eigenvalue is equal to the sum of the energies for the 1D problems

$$E_n = \frac{h^2}{8m}\left(\frac{n_1^2}{L_1^2} + \frac{n_2^2}{L_2^2}\right), \tag{4.7}$$

while the wave function has form of the product

$$\Psi_{n_1n_2} = 2\sqrt{\frac{1}{L_1L_2}}\sin\frac{n_1\pi}{L_1}x \cdot \sin\frac{n_2\pi}{L_2}y, \tag{4.8}$$

where $n_1, n_2 = 1, 2, \ldots$

---

[7] Recall that any stationary state has a trivial time-dependence through the factor $\exp(-i\frac{E}{\hbar}t)$. A standing wave at any time t has a standing-still pattern of the nodes i.e. the points x with $\Psi = 0$.

[8] With the opposite sign, but it does not matter.

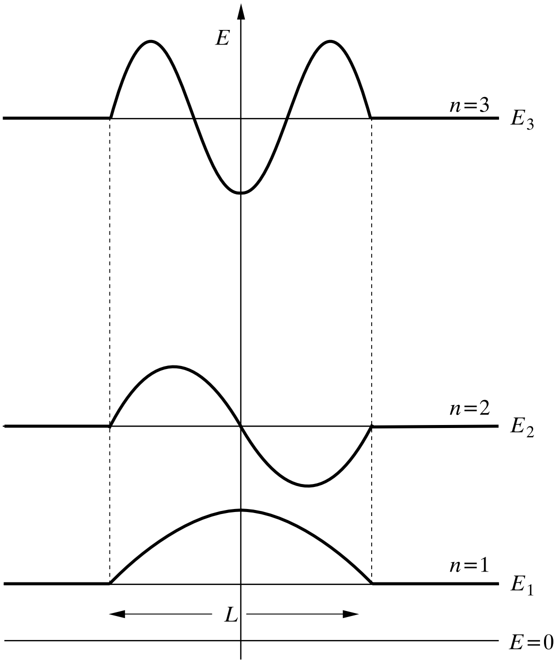


Fig. 4.2. The wave functions for the particle in a box corresponding to $n = 1, 2, 3$. Note the increasing number of nodes, when the energy E_i of the stationary state increases.

Example 1. *Butadiene naively*

The particle-in-box problem has more to do with chemistry than would appear at first glance.

In organic chemistry, we consider some molecules with conjugate double and single bonds, one of the simplest is butadiene: $= - =$.

What does this molecule have to do with the particle in a box? It seems nothing. First, we have not a single particle but 40 particles (10 nuclei and 30 electrons), second, where is this constant potential for the motion of the particle? Nowhere. Third, a molecule does not represent a one-dimensional but a three-dimensional object, and in addition, a curved one instead of a beautiful section of the x axis. It would seem that any attempt to apply such a primitive theory to our molecule is ridiculous and yet in such a difficult situation we will see the power of the exact solutions reported in the present chapter. All above objections are perfectly justified, but let us try to simplify our system a little.

In the molecule under study the CC bonds are "averaged", which facilitates the motion of the π electrons along the system (this notion will become clear in Chapter 8; the π electrons are loosely bound to the molecule, we may assume that other electrons are always rigidly bound and will therefore be ignored).

If

- we removed the π electrons from the molecule (and put them temporarily into a safe), and then
- "ground up" the remaining (positively charged) molecular core and distributed the ground mass uniformly along the x axis within a section of length L equal to

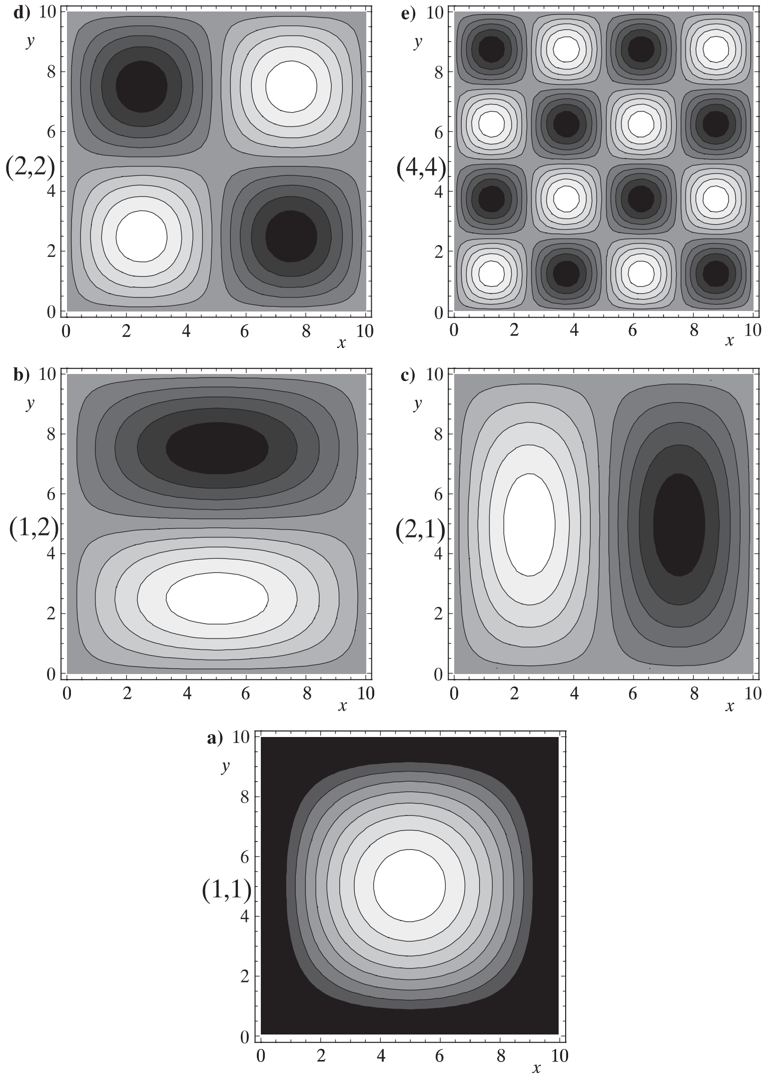


Fig. 4.3. Examples of the wave functions for a particle in a square box, the quantum numbers (n_1, n_2) correspond to: a) $(1, 1)$; b) $(1, 2)$; c) $(2, 1)$; d) $(2, 2)$; e) $(4, 4)$. The background colour corresponds to zero. In the case shown the higher the energy the more nodes in the wave function. This rule is not generally true. For example, in a rectangular box with $L_1 \gg L_2$ even a large increase of n_1 does not raise the energy too much, while introducing a lot of nodes. On the other hand, increasing n_2 by 1 raises the energy much more, while introducing only one extra node. A reader acquainted with hydrogen atom orbitals will easily recognize the resemblance of the above figures to some of them (cf. pp. 180–185), because of the rule mentioned above.

the length of the molecule (averaging the potential energy for a charged particle) to construct a highway for the π electrons

- added the first π electron from the safe, then this single electron would represent something similar to a particle in a box.[9] Assuming this simplified model we know all the details of the electron distribution, including the ground-state and excited-state wave functions (in the one-particle case called the orbitals). If we now took all the π electrons from the safe, added them one by one to the system, assuming that they would not see one another,[10] then taking into account the Pauli exclusion principle (described in more detail in Chapter 8) we would obtain information about the electron density distribution in the molecule. The idea we are describing is called the Free Electron Molecular Orbitals (FEMO) method.

FEMO method

In our example, the total electron density distribution (normalized to four π electrons, i.e. giving 4 after integration over x) is given as[11]

$$\rho(x) = 2\psi_1^2 + 2\psi_2^2 = 2\frac{2}{L}\sin^2\frac{\pi}{L}x + 2\frac{2}{L}\sin^2\frac{2\pi}{L}x = \frac{4}{L}\left(\sin^2\frac{\pi}{L}x + \sin^2\frac{2\pi}{L}x\right).$$

The function $\rho(x)$ is shown in Fig. 4.4.a.

> It is seen that:
> 1. $\rho(x)$ is the largest on the outermost bonds in the molecule, exactly where chemists put their two little lines to symbolize a double bond.
> 2. π-electron density, i.e. $\rho(x)$ is non-zero in the centre. This means that the bond over there is not strictly a single bond.

This key information about the butadiene molecule has been obtained at practically no cost from the simple FEMO model.

Of course, we cannot expect the description to reflect all the details of the charge distribution in the butadiene molecule, but one may expect this approach to be able to reflect at least some rough features of the π electron distribution. *If the results of more advanced calculations contradict the rough particle-in-box results, then we should take a closer look at them and search for an error.* This is the strength of the simple exact model systems. They play the role of the beacons – points of reference.

4.2.2 CYCLIC BOX

The 1D box described above is similar to a stick in which the particle can move. The butadiene molecule is rather similar to such a stick and, therefore, the 1D box models it quite well.

---

[9] Almost, because the potential is not quite constant (ends!). Also one might remove the particle from the box at the expense of a large but finite energy (ionization), which is not feasible for the particle in a box.

[10] As we will see in Chapter 8, this approximation is more realistic than it sounds.

[11] The student "i" is characterized by a probability density distribution $\rho_i(x)$ of finding him at coordinate x (we limit ourselves to a single variable, measuring his position, say, on his way from the dormitory to the university). If all students moved independently, the sum of their individual probability densities at point x_0, i.e. $\rho(x_0) = \sum_i \rho_i(x_0)$ would be proportional to the probability density of finding any student at x_0. The same pertains to electrons, when assumed to be independent.

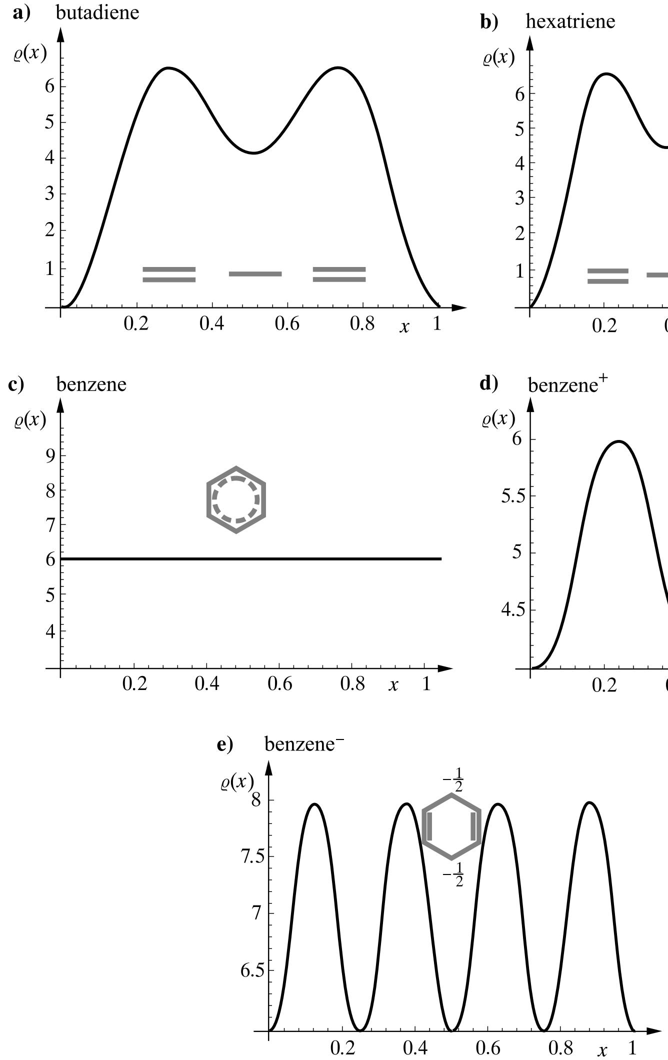


Fig. 4.4. π-electron density charge distributions for several molecules computed by the FEMO method. The length of each molecule L has been assumed to be equal 1. For other lengths the charge distributions are similar. The electron density for four electrons in butadiene (a) and of six electrons in hexatriene (b). The electron density maxima coincide with the positions chemists write as double bonds. *The six electron density distribution in the benzene molecule is peculiar, because it is constant along the perimeter of the molecule* (c). If we subtract an electron from benzene (d) or add an electron to it (e), then maxima and minima of the π electron density appear. If an electron is subtracted (d) there are two maxima (double bonds) and two π electron deficient regions denoted as having charge $+\frac{1}{2}$. After one electron π is added (e) then we obtain four maxima (two double bonds and two electron-rich regions denoted by charge $-\frac{1}{2}$).

And what can model the benzene molecule? In a crude approximation we may think of benzene as a stick with the two ends *joined* in such a way as to be unable to recognize where the union has taken place. Limiting ourselves to this effect,[12] we may use the solution given by (4.3) and impose appropriate boundary conditions. What could these boundary conditions be? The wave function at the two ends of the box has to be stitched together without leaving any trace of the seam. This is achieved by two boundary conditions: $\Psi(0) = \Psi(L)$ forcing the two wave function values to match and $\Psi'(0) = \Psi'(L)$ making the seam "invisible". The two conditions mean:

$$A\sin\kappa 0 + B\cos\kappa 0 = A\sin\kappa L + B\cos\kappa L,$$
$$A\kappa\cos\kappa 0 - B\kappa\sin\kappa 0 = A\kappa\cos\kappa L - B\kappa\sin\kappa L$$

or

$$B = A\sin\kappa L + B\cos\kappa L,$$
$$A = A\cos\kappa L - B\sin\kappa L.$$

To find a non-trivial solution the determinant of the coefficients at the unknown quantities A and B has to vanish:[13]

$$\begin{vmatrix} \sin\kappa L & \cos\kappa L - 1 \\ \cos\kappa L - 1 & -\sin\kappa L \end{vmatrix} = 0,$$

which is equivalent to

$$\cos\kappa L = 1.$$

The last condition gives $\kappa L = 2\pi n$, $n = 0, \pm 1, \pm 2, \ldots$. This immediately gives a formula for the energy very similar to that for the box with ends, but with the replacement $n \to 2n$:

$$E_n = \frac{(2n)^2 h^2}{8mL^2}, \tag{4.9}$$

where this time $n = 0, \pm 1, \pm 2, \ldots$

The corresponding wave functions are

$$\psi_0 = \sqrt{\frac{1}{L}} \quad \text{for } n = 0,$$
$$\psi_{n>0} = A\sin\frac{2\pi n}{L}x + B\cos\frac{2\pi n}{L}x,$$
$$\psi_{n<0} = -A\sin\frac{2\pi |n|}{L}x + B\cos\frac{2\pi |n|}{L}x.$$

---

[12] And neglecting such effects as the particular shape of the benzene (curvature, etc.).

[13] This is a set of homogeneous linear equations.

Since $\psi_{n>0}$ and $\psi_{n<0}$ correspond to the same energy, any combination of them also represents an eigenfunction of the Schrödinger equation corresponding to the same energy (Appendix B on p. 895). Taking therefore as the new wave functions (for $n \neq 0$) the normalized sum and difference of the above wave functions, we finally obtain the solutions to the Schrödinger equation

$$\Psi_0 \equiv \psi_0 = \sqrt{\frac{1}{L}} \quad \text{for } n = 0,$$

$$\Psi_{n>0} = \sqrt{\frac{2}{L}} \sin \frac{2\pi n}{L} x \quad \text{for } n > 0,$$

$$\Psi_{n<0} = \sqrt{\frac{2}{L}} \cos \frac{2\pi n}{L} x \quad \text{for } n < 0.$$

4.2.3 COMPARISON OF TWO BOXES: HEXATRIENE AND BENZENE

Let us take an example of two molecules: hexatriene and benzene (i.e. the cyclohexatriene). Let us assume for simplicity that the length of the hexatriene L is equal to the perimeter of the benzene.[14] Both molecules have 6 π electrons (any of them). The electrons doubly occupy (the Pauli exclusion principle) three one-electron wave functions corresponding to the lowest energies. Let us compute the sum of the electron energies[15] (in the units $\frac{h^2}{8mL^2}$, to have the formulae as compact as possible):

- HEXATRIENE: $E_{\text{heks}} = 2 \times 1 + 2 \times 2^2 + 2 \times 3^2 = 28$,
- BENZENE: $E_{\text{benz}} = 2 \times 0 + 2 \times 2^2 + 2 \times 2^2 = 16$.

We conclude, that 6 π electrons in the benzene molecule correspond to lower energy (i.e. is more stable) than the 6 π electrons in the hexatriene molecule. Chemists find this experimentally: the benzene ring with its π electrons survives in many chemical reactions, whereas this rarely happens to the π-electron system of hexatriene.

Our simple theory predicts the benzene molecule is more stable than the hexatriene molecule.

And what about the electronic density in both cases? We obtain (Fig. 4.4.b,c)

- HEXATRIENE: $\rho(x) = 2 \times \frac{2}{L}[\sin^2 \frac{\pi}{L} x + \sin^2 \frac{2\pi}{L} x + \sin^2 \frac{3\pi}{L} x]$,
- BENZENE: $\rho(x) = 2 \times \frac{1}{L} + 2 \times \frac{2}{L}[\sin^2 \frac{2\pi}{L} x + \cos^2 \frac{2\pi}{L} x] = \frac{6}{L}$.

---

[14] This is to some extent an arbitrary assumption, which simplifies the final formulae nicely. In such cases we have to be careful that the conclusions are valid.

[15] As will be shown in Chapter 8, this method represents an approximation.

This is an extremely interesting result.

The π-electron density is constant along the perimeter of the benzene molecule.

No single and double bonds – *all CC bonds are equivalent* (Fig. 4.4.c). Chemical experience led chemists already long time ago to the conclusion that all the C–C bonds in benzene are equivalent. This is why they decided to write down the benzene formula in the form of a regular hexagon with a circle in the middle (i.e. not to give the single and double bonds). The FEMO method reflected that feature in a naive way. Don't the π electrons see where the carbon nuclei are? Of course they do. We will meet some more exact methods in further chapters of this textbook, which give a more detailed picture. The π-electron density would be larger, closer to the nuclei, but *all CC bonds would have the same density distribution*, similar to the solution given by the primitive FEMO method. From (4.9) and the form of the wave functions it follows that this will happen not only for benzene, but also for all the systems with $(4n+2)$-electrons, $n = 1, 2, \ldots$, because of a very simple (and, therefore, very beautiful) reason that $\sin^2 x + \cos^2 x = 1$.

The addition or subtraction of an electron makes the distribution non-uniform (Fig. 4.4.d,e). Also in six π electron hexatriene molecule, uniform electron density is out of the question (Fig. 4.4.b). Note that the maxima of the density coincide with the double bonds chemists like to write down. However, even in this molecule, there is still a certain equalization of bonds, since the π electrons are also where the chemists write a single bond (although the π electron density is smaller over there[16]).

Again important information has been obtained at almost no cost.

4.3 TUNNELLING EFFECT

Is it possible to pass through a barrier with less energy than the barrier height? Yes.

4.3.1 A SINGLE BARRIER

Let us imagine a rectangular potential energy barrier (Fig. 4.1.b) for the motion of a particle of mass m: $V(x) = V$ for $0 \leqslant x \leqslant a$, with $V(x) = 0$ for other values of x (V is the barrier height). Let us assume that the particles go from left to right and that their energy E is smaller than V. This assumption will make it possible to study the most interesting phenomenon – tunnelling through the barrier.[17] In order to stress that $0 \leqslant E \leqslant V$ let us write:

$$E = V \sin^2 \beta. \tag{4.10}$$

---

[16] Where, in the classical picture, no π electron should be.

[17] Another interesting question would be what will happen if $E > V$. This question will be postponed for a moment.

The x axis will be divided in three parts:

region 1 $-\infty < x < 0$,
region 2 $0 \leqslant x \leqslant a$,
region 3 $a < x < \infty$.

In each of these regions the Schrödinger equation will be solved, then the solutions will be stitched together in such a way as to make it smooth at any boundary. The general solution for each region has the form[18] $\Psi(x) = Ae^{i\kappa x} + Be^{-i\kappa x}$, where A and B are the de Broglie wave amplitudes for motion to the right and to the left. The κ constant comes from the Schrödinger equation $\frac{\partial^2\Psi}{\partial x^2} + \kappa^2\Psi = 0$, where $\kappa^2 = \frac{2mE}{\hbar^2}$ for regions 1 and 3 and $\kappa^2 = \frac{2m(E-V)}{\hbar^2}$ for region 2. Therefore, the wave functions for each region is:

$$\Psi_1(x) = A_1 e^{\frac{ix\sqrt{2mE}}{\hbar}} + B_1 e^{-\frac{ix\sqrt{2mE}}{\hbar}}, \tag{4.11}$$

$$\Psi_2(x) = A_2 e^{\frac{-x\cot\beta\sqrt{2mE}}{\hbar}} + B_2 e^{\frac{x\cot\beta\sqrt{2mE}}{\hbar}}, \tag{4.12}$$

$$\Psi_3(x) = A_3 e^{\frac{ix\sqrt{2mE}}{\hbar}} + B_3 e^{-\frac{ix\sqrt{2mE}}{\hbar}}. \tag{4.13}$$

The second equation needs a little derivation, but using eq. (4.10) this is straightforward.

In regions 1 and 2 we may have the particle going right or left (reflection), hence in these regions A and B are non-zero. However, in region 3 we are sure that $B_3 = 0$, because there will be no returning particle (since no reflection is possible in region 3).

Now, the coefficients A and B are to be determined (with accuracy up to a multiplicative constant) in such a way as to ensure that the wave function sections match smoothly. This will be achieved by matching the function values and the first derivatives at each of the two boundaries.[19]

As the wave function has to be continuous for $x = 0$ and $x = a$, we obtain the following equations

$$A_1 + B_1 = A_2 + B_2,$$

$$A_2 \exp\left(-\frac{a \quad \cot\beta\sqrt{2mE}}{\hbar}\right) + B_2 \exp\left(+\frac{a \quad \cot\beta\sqrt{2mE}}{\hbar}\right)$$
$$= A_3 \exp\left(\frac{ia\sqrt{2mE}}{\hbar}\right).$$

---

[18] This is the free particle wave function. The particle has the possibility (and, therefore, certain probability) of going left or right.

[19] The second derivative is discontinuous, because of the form of the potential function $V(x)$ introduced.

The continuity of the first derivative at $x = 0$ and $x = a$ gives:

$$i(A_1 - B_1) = \cot(B_2 - A_2),$$
$$\cot\beta\left(-A_2\exp\left(-\frac{a\cot\beta\sqrt{2mE}}{\hbar}\right) + B_2\exp\left(+\frac{a\ \cot\beta\sqrt{2mE}}{\hbar}\right)\right)$$
$$= iA_3\exp\left(\frac{ia\sqrt{2mE}}{\hbar}\right).$$

After introducing the abbreviations:

$$k = \exp\left(\frac{a\cot\beta\sqrt{2mE}}{\hbar}\right) \quad \text{and} \quad N = (1-k^2)\cos 2\beta + i(1+k^2)\sin 2\beta,$$

we obtain the following ratios of all the coefficients and coefficient A_1:

$$\frac{B_1}{A_1} = \frac{k^2-1}{N}, \qquad \frac{A_2}{A_1} = \frac{k^2(1-\exp(-2i\beta))}{N},$$
$$\frac{B_2}{A_1} = \frac{(\exp(2i\beta)-1)}{N}, \qquad \frac{A_3}{A_1} = \frac{2ik\sin 2\beta\exp(-\frac{ia\sqrt{2mE}}{\hbar})}{N}.$$

A current in region 3 towards the positive direction of the x axis may be defined as the probability density $A_3^*A_3$ of the particle going right in region 3 times the velocity

$$\sqrt{\frac{2E}{m}} = \sqrt{\frac{2\frac{mv^2}{2}}{m}} = v$$

in this region. Therefore, the current passing through region 3 is equal to

current

$$A_3^*A_3\sqrt{\frac{2E}{m}} = \frac{4k^2A_1^*A_1\sin^2 2\beta}{|N|^2}\sqrt{\frac{2E}{m}}. \tag{4.14}$$

Therefore, the ratio of the current going right in 3 to the current going right in 1 is equal to: $D_{\text{single}} = \frac{|A_3|^2}{|A_1|^2}$ (the *barrier transmission coefficient*, in our case equal to the probability of passing the barrier):

transmission coefficient

$$\boxed{D_{\text{single}} = \frac{4k^2\sin^2(2\beta)}{NN^*}} \tag{4.15}$$

- This result is exact (for $0 \leqslant E \leqslant V$). Despite its conciseness the formula for D_{single} as a function of E, V looks quite complicated. What does this formula tell us? Below are some questions:
- D_{single} should increase when the particle energy E increases, but is the function $D_{\text{single}}(E)$ monotonic? Maybe there are some "magic" energies at which passage through the barrier becomes easier? Maybe this is what those guys in the movies use, when they go through walls.

The answer is given in Figs. 4.5.a–f. It has been assumed that the particle has the mass of an electron (1 a.u.). From Figs. 4.5.a–c for three barrier heights (V), it follows that the function is monotonic, i.e. the faster the particle the easier it is to pass the barrier – quite a banal result. There are no magic energies.

- How does the function $D_{\text{single}}(V)$ look with other parameters fixed? For example, whether it is easier to pass a low or a high barrier with the same energy, or are there some magic barrier heights. Figs. 4.5.a–c tell us that at a fixed E it is easier to pass a lower barrier and the function is monotonic, e.g., for $E = 0.5$ a.u. $\cong 13.5$ eV the transmission coefficient D_{single} is about 80% for $V = 0.5$, 40% for $V = 1$, and 10% for $V = 2$. No magic barrier heights.
- How does the transmission coefficient depend on the barrier width? From Figs. 4.5.d–f we see that $D_{\text{single}}(a)$ is also monotonic (no magic barrier widths) and dramatically drops, when the barrier width a increases. On the other hand the larger the kinetic energy of the projectile heading towards the barrier the better the chance to cross the barrier. For example, at electron energies of the order of 0.5 a.u. (at fixed $V = 1$ and $m = 1$) the barrier of width 2 a.u. $\simeq 1$ Å allows 6% of the particles to pass, while at energy 0.75 a.u. 18%, and at energy 1 a.u. 30% pass.

What does the wave function of the tunnelling particle look like? The answer is in Fig. 4.6. We see that:

- The real as well as the imaginary parts of the wave function are non-zero in the barrier, i.e. the particle penetrates the barrier.
- Both (real and imaginary) parts decay very rapidly (exponentially) for large penetrations.
- Since the barrier has a finite width, the wave function does not vanish completely within the barrier range. Thus, after leaving the barrier region we again have a wave with the same frequency but with a smaller amplitude than that before the barrier range.[20] This means that there is a *non-zero probability that the particle*

---

[20] The tunnelling of a particle is certainly a strange phenomenon and our imagination is too poor to figure out how this happens. However, as a matter of fact even in classical mechanics one may be taken by surprise. This happens when we have to do with the probability density distribution of configurations in an ensemble (as we often have to do in statistical thermodynamics and similarly in quantum mechanics). Then we may encounter the notion of "classical tunnelling" (J. Straub, "*New Developments in Theoretical Studies of Proteins*", ed. R. Elber, World Scientific, 1996), since the mean value of the kinetic energy is lower than the barrier and yet the system overcomes the barrier. Let us put this problem aside.

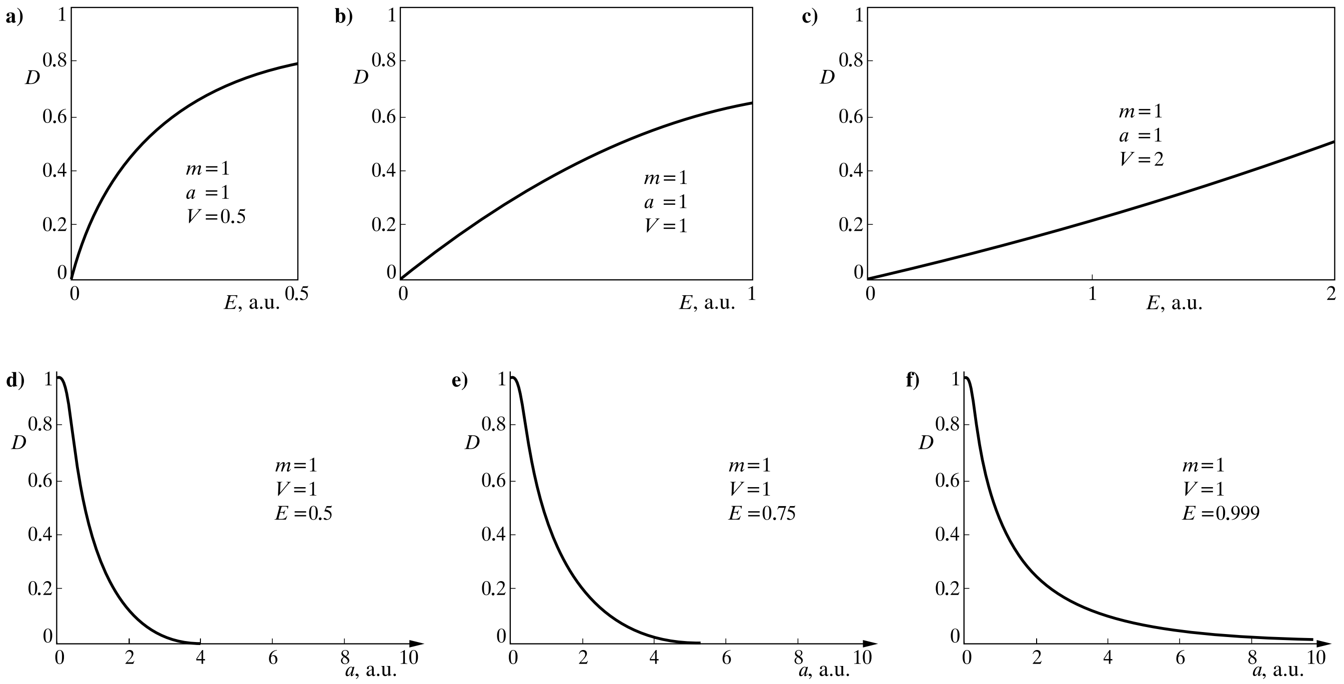


Fig. 4.5. A classical particle cannot tunnel through a barrier, while a quantum particle *can*. The figures show the transmission coefficient (tunnelling) of the electron having various energies (always lower than the barrier) and passing through a barrier of various heights and widths. Figs. a–c show, that the larger the energy the easier to tunnel, also the higher the barrier the harder to pass the barrier (at the same energy of the particle). Figs. d–f show the dependence of the transmission coefficient on the barrier width: the wider the barrier the harder to go through.

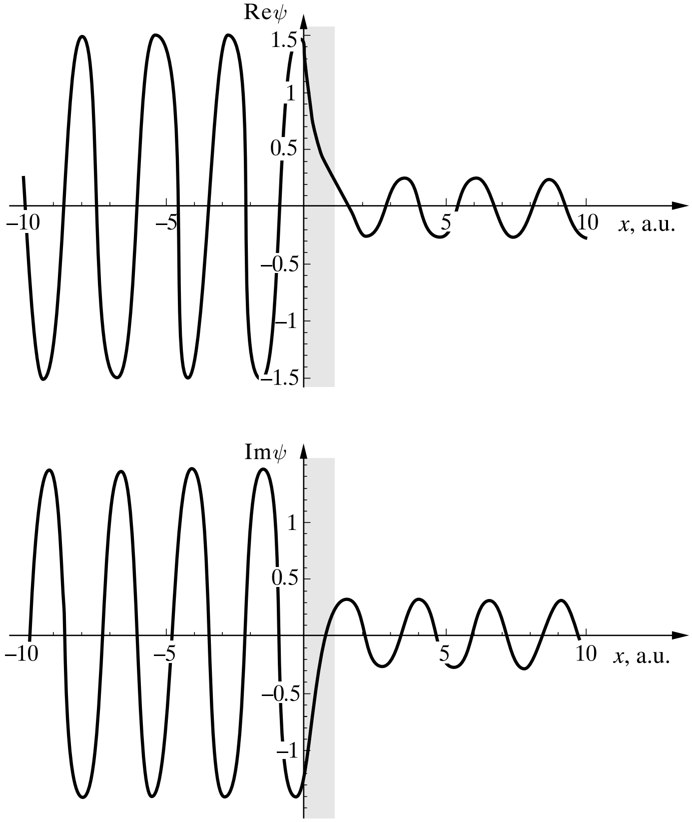


Fig. 4.6. Tunnelling of an electron ($m = 1$) with energy $E = 2.979$ a.u. through a single barrier of height $V = 5$ a.u., and width 1 a.u. The wave function plot (real and imaginary parts) corresponds to the following values of the coefficients $A_1 = 1$ (as a reference), $B_1 = 0.179 - 0.949i$, $A_2 = 1.166 - 0.973i$, $B_2 = 0.013 + 0.024i$, $A_3 = -0.163 - 0.200i$ and represents a wave.

reflects from the barrier and a non-zero probability that the particle passes through the barrier.[21]

4.3.2 THE MAGIC OF TWO BARRIERS...

Is there anything magic in life? Yes, there is. If we take *two* rectangular barriers of height V with a well between them (Fig. 4.1.c), then we have magic. This time we allow for any energy of the particle ($E > 0$).

How will the problem be solved?

We have five non-overlapping sections of the x axis. In each section the wave function will be assumed in the form $\Psi(x) = Ae^{i\kappa x} + Be^{-i\kappa x}$ with the corresponding

---

[21] This remains in general true even if $E > V$.

A and B coefficients, and with $\kappa^2 = \frac{2m(E-V)}{\hbar^2}$. In Section 5, however, the particle goes right and never left, hence $B_5 = 0$. Now, the other coefficients A and B will be determined by stitching the wave function nicely at each of the four boundaries in order to have it going smoothly through the boundary (the wave function values and the first derivative values have to be equal for the left and right section to meet at this boundary). In this way we obtain a set of eight linear equations with eight unknown ratios: $\frac{A_i}{A_1}$, $i = 2, 3, 4, 5$, and $\frac{B_i}{A_1}$, $i = 1, 2, 3, 4$. The most interesting ratio is A_5/A_1, because this coefficient determines the transmission coefficient through the two barriers. Using the program Mathematica,[22] we obtain an amazing result.

Transmission coefficient

Let us check how the transmission coefficient (in our case identical to the transmission probability) changes through two identical barriers of height $V = 5$, each of width $a = 1$, when increasing the impact energy E from 0 to $V = 5$. In general the transmission coefficient is very small. For example, for $E = 2$ the transmission coefficient through the *single* barrier (D_{single}) amounts to 0.028, that is the chance of transmission is about 3%, while the transmission coefficient through the double barrier (D_{double}) is equal to 0.00021, i.e. about 100 times smaller. It stands to reason, it is encouraging. It is fine that it is harder to cross two barriers than a single barrier.[23] And the story will certainly be repeated for other values of E. To be sure, let us scan the whole range $0 \leqslant E < V$. The result is shown in Fig. 4.7.

Magic energetic gates (resonance states)

There is something really exciting going on. In our case we have three energies $E \leqslant V$, at which the transmission coefficient D_{double} increases dramatically when compared to the neighbourhood. These are: 0.34, 1.364 and 2.979. Thus, there are three secret energetic gates for going through the double barrier! It is sufficient just to hit the right energy (resonance energy). Is the chance of passing two barriers large? Let us take a look. For all three resonances the transmission coefficient is equal to $D_{\text{double}} = 1$, but it drops down differently when going off resonance. Thus, there are three particle energies, for which the particle goes through the two barriers like a knife through butter, as if the barriers did not exist.[24] Moreover, as we can see for the third resonance, the transmission coefficient through the single barrier amounts to $D_{\text{single}} = 0.0669$ (i.e. only 7%), but through two barriers 100%! It looks as if it would be hard for a prisoner to pass through a single armoured prison door, but when the anxious prison governor built a second armoured door behind the first, the prisoner[25] disappeared through the two doors like a ghost.[26]

resonance

---

[22] See the Web Annex, the file Mathematica\ Dwiebar.ma.

[23] This is even more encouraging for a prison governor. Of course, a double wall is better than a single one!

[24] This news should be strictly confidential in penitentiary departments.

[25] Educated in quantum mechanics.

[26] There is experimental evidence for such resonance tunnelling through two energy barriers in semiconductors. One of the first reports on this topic was a paper by T.C.L.G. Sollner, W.D. Goodhue, P.E. Tannenwald, C.D. Parker, D.D. Peck, *Appl. Phys. Letters* 43 (1983) 588.

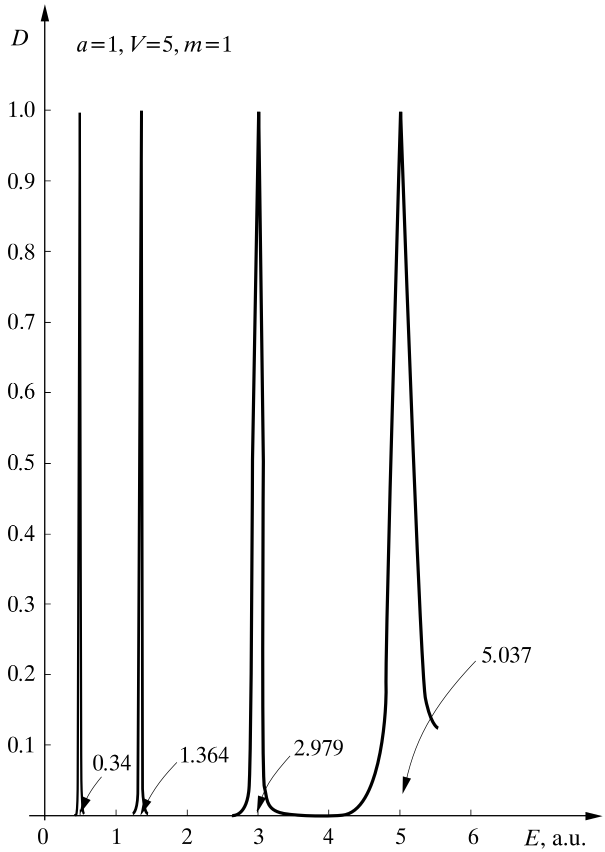


Fig. 4.7. The transmission coefficient (D) for a particle going through a potential double barrier (of height $V = 5$ a.u.) as a function of the particle impact energy E. We see some sudden increases of the transmission coefficient (resonance states).

What happens over there? Let us stress once more that the phenomenon is 100% of a quantum nature, because a classical particle would tunnel neither through the double nor through the single barrier. Why do we observe such dramatic changes in the transmission coefficient for the two barriers? We may have some suspicions. From the time the second barrier is created, a new situation appears: a well *between* the two barriers, something similar to the box discussed earlier.[27] A particle in a box has some peculiar energy values: the energies of the stationary states (cf. p. 146). In our situation all these states correspond to a continuum, but something magic might happen if the particle had just one of these energies. Let us calculate the stationary state energies *assuming* that $V = \infty$. Using the atomic units in the energy formula, we have $E_n = \frac{h^2}{8m}\frac{n^2}{L^2} = \frac{\pi^2}{L^2}\frac{n^2}{2}$. To simplify the formula even more let us take $L = \pi$. Finally, we have simply $E_n = \frac{n^2}{2}$. Hence, we might expect something strange for the energy E equal to $E_1 = \frac{1}{2}$, $E_2 = 2$, $E_3 = \frac{9}{2}$, $E_4 = 8$ a.u., etc. The last energy level, $E_4 = 8$, is already higher than the

---

[27] Note, however, that the box has finite well depth and final width of the walls.

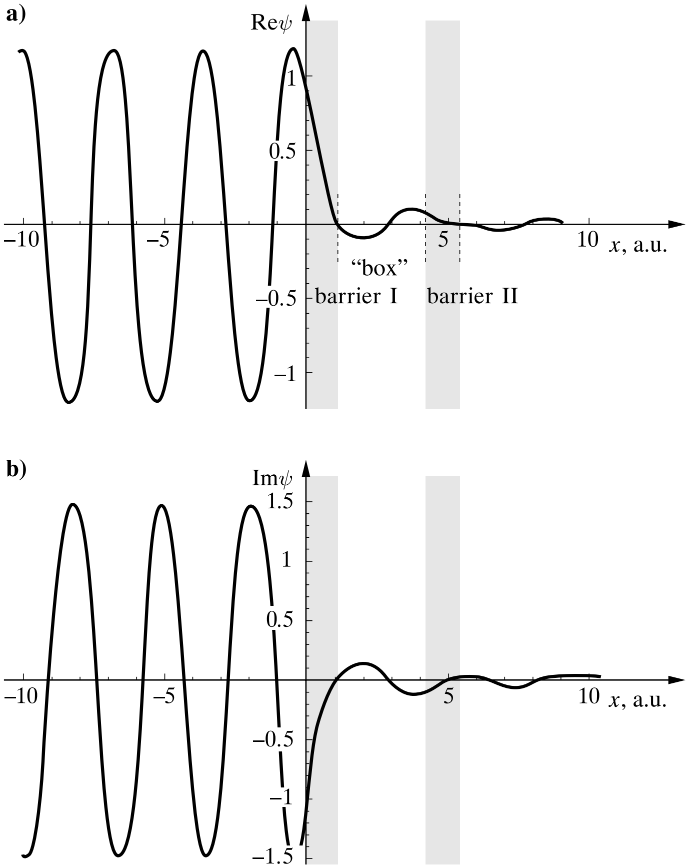


Fig. 4.8. Tunnelling of an electron with energy $E = 2$ a.u. through two barriers of height $V = 5$ and width $a = 1$, the barrier separation is $L = \pi$ (all quantities in a.u.). This is the off-resonance case. The real part of the wave function (a) oscillates before the first barrier, is reduced by an order of magnitude in the first barrier, between the barriers the function oscillates for ca. one period, decays in the second barrier and goes out of the barrier region with an amplitude representing about 5% of the starting amplitude. A similar picture follows from the imaginary part of the wave function (b).

barrier height. Note, however, that the resonance states obtained appear at quite different energies: 0.34, 1.364, 2.979.

Maybe this intuition nevertheless contains a grain of truth? Let us concentrate on E_1, E_2, E_3. One may expect that the wave functions corresponding to these energies are similar to the ground-state (nodeless), the first (single node) and second (two nodes) excited states of the particle in a box. What then happens to the nodes of the wave function for the particle going through two barriers? Here are the plots for the off-resonance (Fig. 4.8) and resonance (of the highest energy, Fig. 4.9) cases.

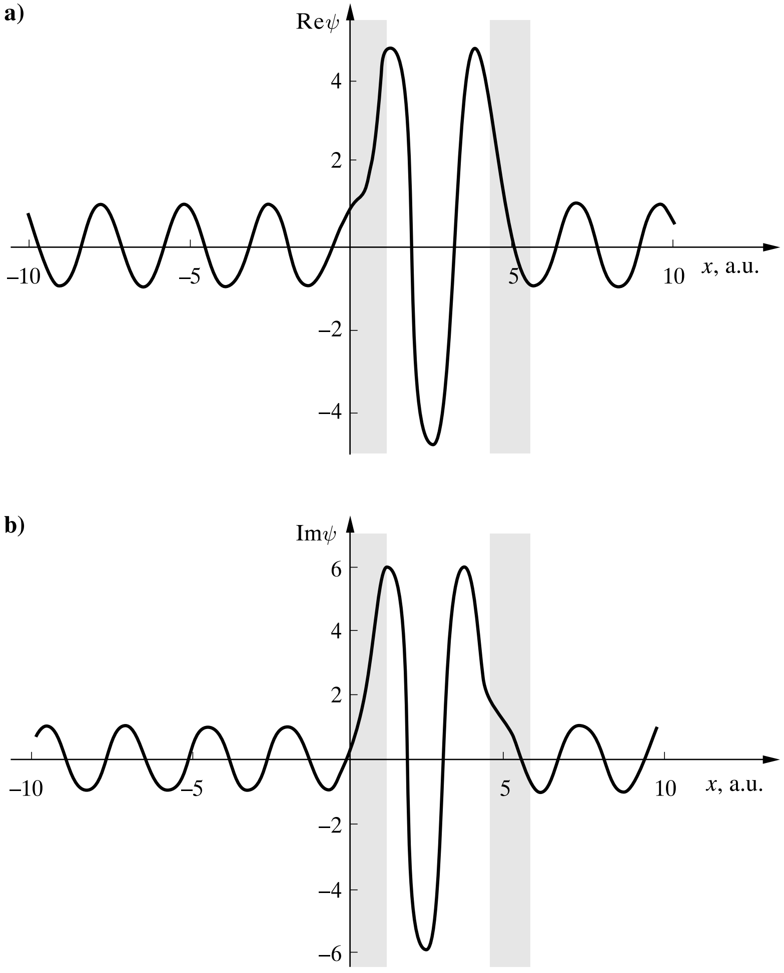


Fig. 4.9. Tunnelling of an electron with energy $E = 2.979$ a.u. through two barriers of height $V = 5$ and width $a = 1$, the barrier separation is $L = \pi$ (all quantities in a.u.). This is the resonance case. The real part of the wave function (a) oscillates before the first barrier with amplitude 1, *increases by a factor of about* 3.5 within the first barrier, between the barriers the function makes slightly more than about one period, decays in the second barrier and goes out of the barrier region with an amplitude representing about 100% of the starting amplitude. A similar picture follows from the imaginary part of the wave function (b).

These figures and similar figures for lower-energy resonances support the hypothesis: if an integer number of the half-waves of the wave function fit the region of the "box" between the barriers ("barrier-box-barrier"), in this case we may expect resonance – a secret gate to go through the barriers.[28] As we can see, in-

---

[28] As one can see in this case, contrary to what happened with a single barrier, the wave function does not vanish exponentially within the barriers.

deed we have been quite close to guessing the reason for the resonances. On the other hand, it turned out that the box length should include not only the box itself but also the barrier widths. Maybe to obtain the right resonance energies we simply have to adjust the box length? Since, instead of resonance at $E_1 = \frac{1}{2}$ we have resonance at energy 0.34, then we may guess that it is sufficient to change the box width L to $L' = \sqrt{\frac{0.5}{0.34}}L = 1.21L$, to make the first resonance energies match. Then, instead of $E_1 = \frac{1}{2}$, we have exactly the first resonance energy equal to $E_1' = 0.34$, an agreement forced on us. But later, instead of $E_2 = 2$ we obtain $E_2' = 1.36$, which agrees very well with the second resonance energy 1.364. Then, instead of $E_3 = 4.5$, we obtain $E_3' = 3.06$, a good approximation to 2.979, but evidently the closer the barrier energy the harder it is to obtain agreement.[29] The next resonance state is expected to occur at $E_4 = 8 \times 0.68 = 5.44$, but we have forgotten that this energy already exceeds the barrier height ($V = 5$ a.u.). We will come back to this state in a moment.

Close encounters of the third degree?

Let us consider the two barriers and an electron with *higher energy than the barrier* height V. What will happen? Well, we may say that this means the particle energy is sufficient to pass the barrier. Let us see.

Let us assume the barrier height $V = 5$ and the particle energy is equal to 5.5 a.u. We solve our equations and we obtain transmission coefficient equal to 0.138, hence the electron will bounce back with a probability of about 86%. How it did bounce off? Difficult to say.

Fig. 4.7 shows also the transmission coefficient also for energies higher than the barrier height. It turns out that at energy $E = 5.037$ a.u. (i.e. higher than the barrier height) another resonance state is hidden, which assures certainty (100%) of transmission (whereas the particle energies in the energetic neighbourhood lead to a considerable reflection rate as described above). We expected such behaviour for all $E > V$, but it turned out to be true for the resonance state. Let us recall that we have already predicted "by mistake" a box stationary state with energy $E_4 = 5.44$, higher than the barrier height V. This, and the number of the nodes within the barrier range seen in Fig. 4.10, tells us that indeed this is the state.[30]

What makes the difference between the resonance and off-resonance states for $E > V$? The corresponding wave functions (real and imaginary parts) are given in Figs. 4.10 and 4.11.

Thus, resonance states may also hide in that part of the continuum which has energy higher than the barriers (with a short life time, because such resonances are

---

[29] Note, please, that there is such a thing as resonance width, and that this width is different for each resonance. The most narrow resonance corresponds to the lowest energy, the widest to the highest energy. The width of resonances is related to the notion of the resonance life-time τ (τ is proportional to the inverse of the resonance width).

[30] It corresponds to a lower energy than we predicted (similar to the case of E_3). No wonder that due to finite well depth, the states corresponding to the upper part of the well "feel" the box is longer.

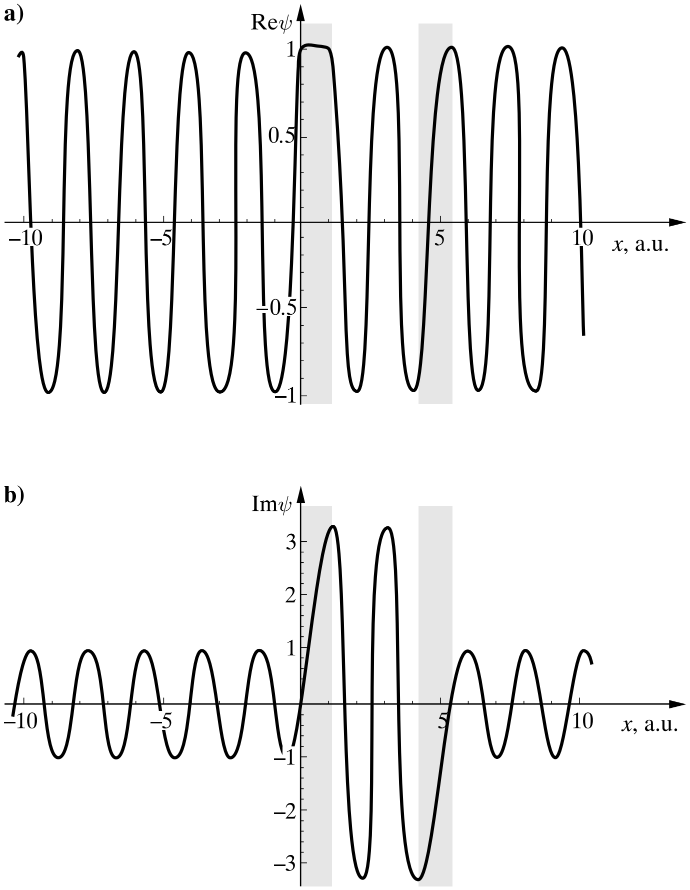


Fig. 4.10. The wave function for an electron with energy $E = 5.037$ a.u., i.e. *over the barrier* $V = 5$ (resonance case). As we can see the amplitude is nearly the same for the wave function before and after the barriers (this means the transmission coefficient of the order of 100%). The real part, and especially the imaginary part both wobble within the range of the barriers range, i.e. within section $(0, 5.14)$ (the imaginary part has a large amplitude). We may guess that the state is related to the three-node stationary state.

wide, cf. Fig. 4.7). They are also a reminder of the stationary states of the particle in a box longer than the separation of the barriers and infinite well depth.

4.4 THE HARMONIC OSCILLATOR

A one-dimensional harmonic oscillator is a particle of mass m, subject to force $-kx$, where the force constant $k > 0$, and x is the displacement of the particle from its equilibrium position ($x = 0$). This means the force pushes the particle

force constant

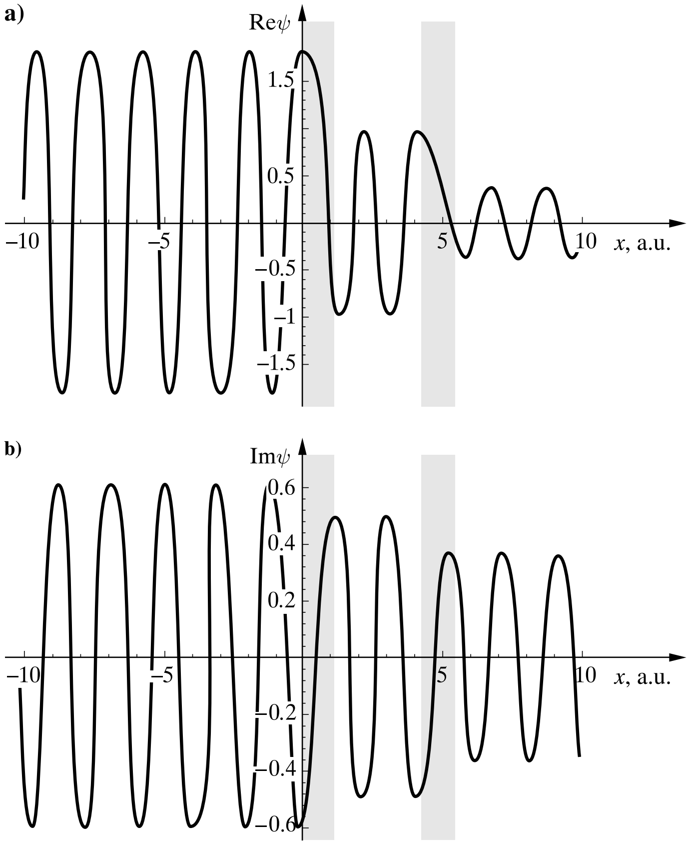


Fig. 4.11. The wave function for an electron in the off-resonance case ($E = 5.5$ a.u., i.e. *over* the barrier height $V = 5$). Despite the fact that $E > V$, the amplitude of the outgoing wave is considerably reduced after passing the range of the barriers $(0, 5.14)$. This means that the particle flying over the barriers will reflect from them.

always towards the origin, because it has a negative (positive) component for $x > 0$ ($x < 0$). The potential energy is given as a parabola $V = \frac{1}{2}kx^2$, Fig. 4.1.d.

The Schrödinger equation has the following solutions of class Q:

$$\boxed{\Psi_v(\xi) = N_v H_v(\xi) \exp\left(-\frac{\xi^2}{2}\right) \qquad (4.16)}$$

with energy

$$\boxed{E_v = h\nu(v + 1/2), \qquad (4.17)}$$

vibrational quantum number

where h is the Planck constant, $v = 0, 1, 2, \ldots$ is the *vibrational quantum number*, the variable ξ is proportional to the displacement x:

$$\xi = \sqrt{\alpha}x, \qquad \alpha = \sqrt{\frac{km}{\hbar^2}}, \qquad \nu = \frac{1}{2\pi}\sqrt{\frac{k}{m}}$$

frequency

Hermite polynomials

is the frequency of the classical vibration of a particle of mass m and a force constant k, H_v represent the Hermite polynomials[31] (of degree v) defined as[32]

$$H_v(\xi) = (-1)^v \exp\left(\xi^2\right) \frac{d^v \exp(-\xi^2)}{d\xi^v},$$

and N_v is the normalization constant, $N_v = \sqrt{(\frac{\alpha}{\pi})^{\frac{1}{2}} \frac{1}{2^v v!}}$.

> The harmonic oscillator finger print: it has an infinite number of energy levels, all non-degenerate, with constant separation equal to $h\nu$.

Note, that the oscillator energy is never equal to zero.

Fig. 4.12 shows what the wave functions for the one-dimensional harmonic oscillator look like. Fig. 4.13 also shows the plots for a *two-dimensional* harmonic oscillator (we obtain the solution by a simple separation of variables, the wave function is a product of the two wave functions for the harmonic oscillators with x and y variables, respectively).

The harmonic oscillator is one of the most important and beautiful models in physics. When almost nothing is known, except that the particles are held by some

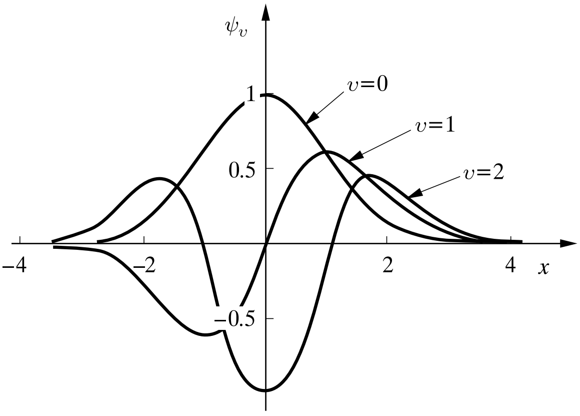


Fig. 4.12. Some of the wave functions Ψ_v for a one-dimensional oscillator. The number of nodes increases with the oscillation quantum number v.

---

[31] Charles Hermite was French mathematician (1822–1901), professor at the Sorbonne. The Hermite polynomials were defined half a century earlier by Pierre Laplace.

[32] $H_0 = 1$, $H_1 = 2\xi$, $H_2 = 4\xi^2 - 2$, etc.

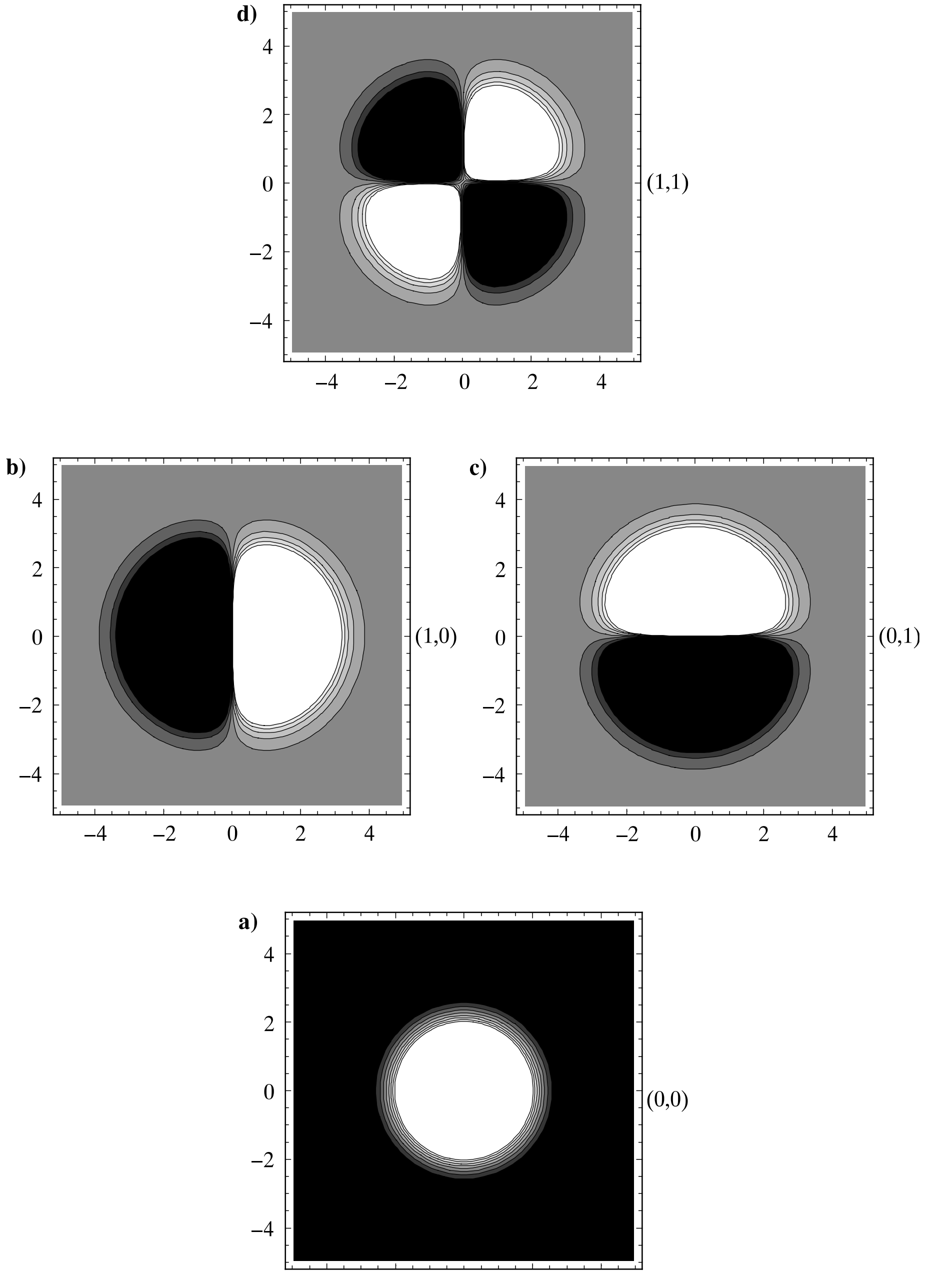


Fig. 4.13. A graphic representation of the 2D harmonic oscillator wave (isolines). The background colour corresponds to zero. Figs. a–i show the wave functions labelled by a pair of oscillation quantum numbers (v_1, v_2). The higher the energy the larger the number of node planes. A reader acquainted with the wave functions of the hydrogen atom will easily recognize a striking resemblance between these figures and the orbitals.

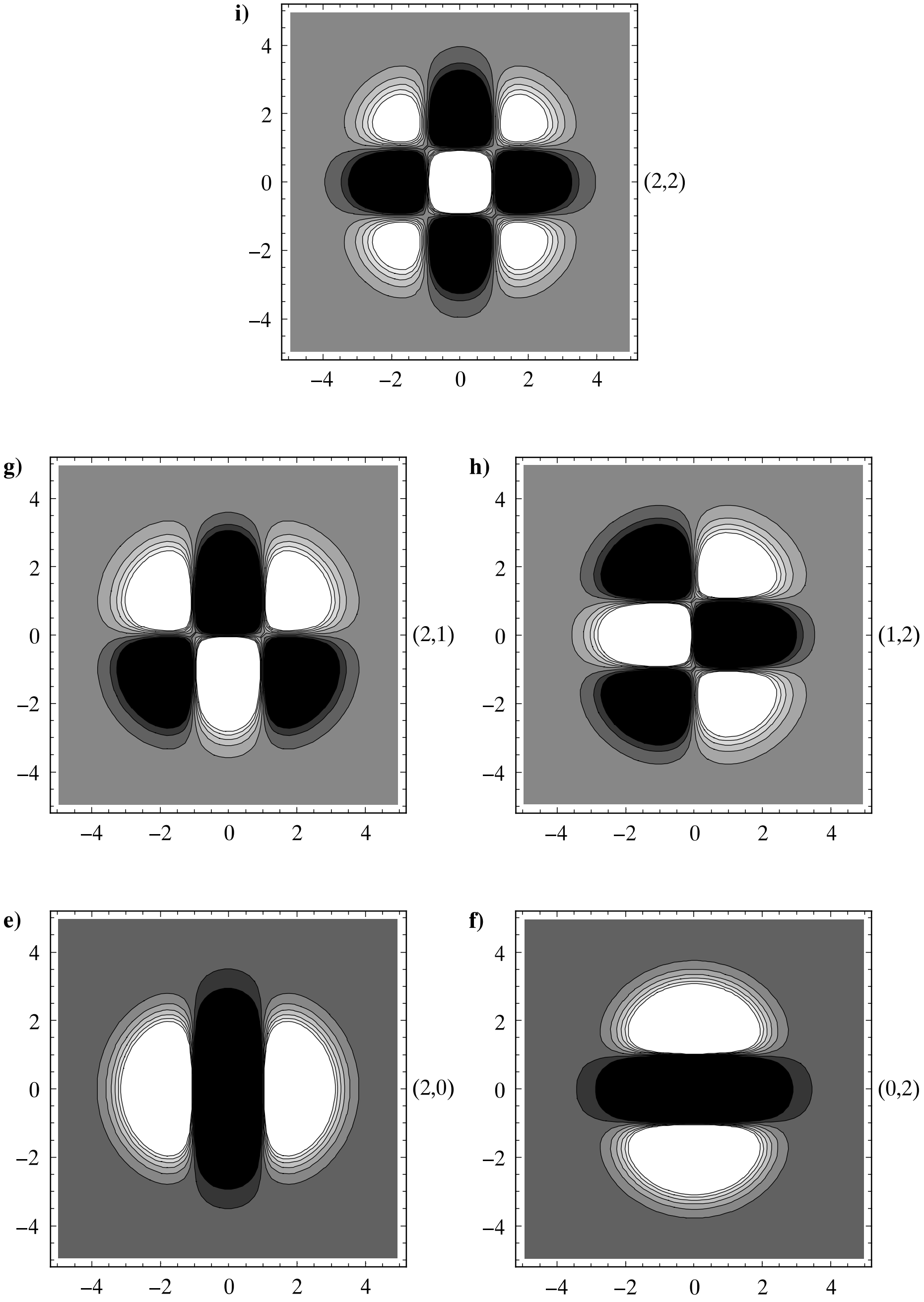


Fig. 4.13. *Continued*.

forces, then the first model to consider is the harmonic oscillator. This happened for the black body problem (Chapter 1), now it is the case with quantum dots,[33] string theory,[34] solvated electron,[35] and so on.

4.5 MORSE OSCILLATOR

4.5.1 MORSE POTENTIAL

Diatomic molecules differ from harmonic oscillators mainly in that they may dissociate. If we pull a diatomic molecule with internuclear distance R equal to the equilibrium distance R_e, then at the beginning, displacement $x = R - R_e$ is indeed proportional to the force applied, but afterwards the pulling becomes easier and easier. Finally, the molecule dissociates, i.e. we separate the two parts without any effort at all. This fundamental difference with respect to the harmonic oscillator is qualitatively captured by the potential proposed by Morse (parameter $\alpha > 0$):[36]

$$V(x) = De^{-\alpha x}\left(e^{-\alpha x} - 2\right). \tag{4.18}$$

As we can see (Fig. 4.14), D represents the well depth and, the parameter α decides its width. When the displacement $x = 0$, then the function attains the minimum $V = -D$, and when $x \to \infty$, then $V \to 0$.

The Morse oscillator will serve as a model of a diatomic molecule. In such a case $x = R - R_e$, where R_e means the length of the molecule which corresponds to the potential energy minimum. Besides the above mentioned advantage, the Morse oscillator differs from real diatomics mainly by two qualitative features. First, for $R = 0$ we obtain a *finite* potential energy for the Morse oscillator, second, the asymptotic behaviour of the Morse oscillator for $x \to \infty$ means exponential asymptotics, while the atomic and molecular systems at large distances interact as $\frac{1}{R^n}$.

The second derivative of $V(x)$ calculated at the minimum of the well represents the force constant k of the oscillator

$$k = 2\alpha^2 D. \tag{4.19}$$

Parabola $-D + \frac{1}{2}kx^2$ best approximates $V(x)$ close to $x = 0$ and represents the harmonic oscillator potential energy (with the force constant k). The Morse

---

[33] A part of the "nanotechnology": some atomic clusters are placed (*quantum dots*) on a solid surface, lines of such atoms (*nanowires*), etc. Such systems may exhibit unusual properties.

[34] Quarks interact through exchange of gluons. An attempt at separating two quarks leads to such a distortion of the gluon bond (string) that the string breaks down and separates into two strings with new quarks at their ends created from the distortion energy.

[35] Many polar molecules may lower their energy in a liquid by facing an extra electron with their positive pole of the dipole. This is the *solvated electron*.

[36] Philip McCord Morse (1903–1985) was American theoretical physicist.

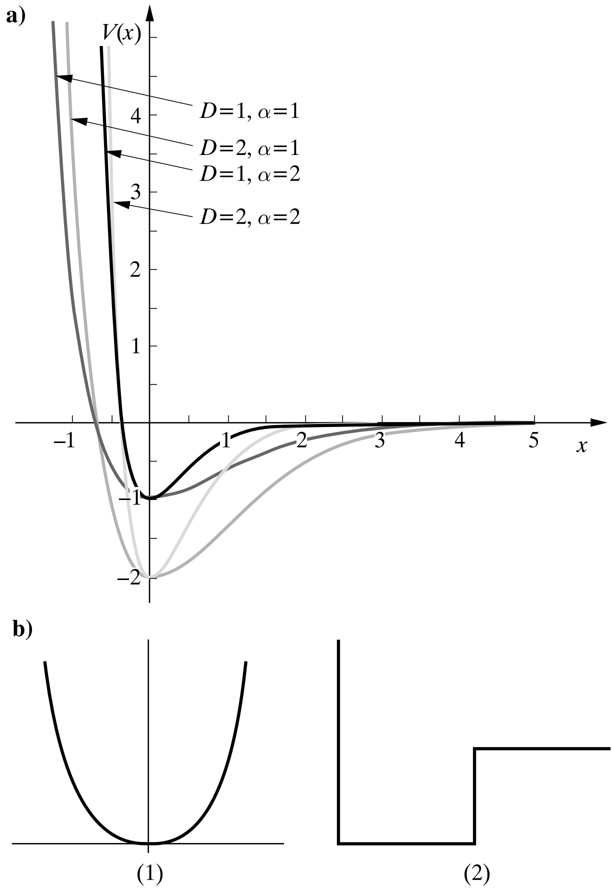


Fig. 4.14. (a) The Morse potential energy curves have the shape of a hook. How does the shape depend on the Morse parameters? The figures show the curves for $D = 1, 2$ and $\alpha = 1, 2$. As we can see D controls the well depth and α its width. (b) the Morse oscillator is a kind of compromise between the harmonic oscillator (b1) and a rectangular well (b2). Both potentials correspond to exact solutions of the Schrödinger equation. Model b2 gives the discrete spectrum as well as the continuum and the resonance states. The later ones are only very rarely considered for the Morse oscillator, but they play an important role in scattering phenomena (primarily in reactive collisions).

oscillator is hard to squeeze – the potential energy goes up faster than that of the harmonic oscillator with the same force constant k.

4.5.2 SOLUTION

One had to have courage to presume that analytical solution with such a potential energy exists. The solution was found by Morse. It represents a rare example of an exact solution to a non-linear problem. Exact solutions exist not only for the ground (vibrational quantum number $v = 0$) but also for all the excited states ($v = 1, 2, \ldots, v_{\max}$) belonging to the discrete spectrum. The energy levels are nondegenerate and are given by the formula:

$$E_v = -D + h\nu\left(v + \frac{1}{2}\right) - h\nu\left(v + \frac{1}{2}\right)^2 \beta, \quad v = 1, 2, \ldots, v_{\max}, \tag{4.20}$$

where using atomic units we obtain

$$h\nu = 2\alpha\left(\frac{D}{2\mu}\right)^{\frac{1}{2}}. \tag{4.21}$$

This formula follows from the parabolic approximation of the Morse potential (valid for small displacements x),[37] while

$$\beta = \alpha\left(\frac{1}{8\mu D}\right)^{\frac{1}{2}}, \tag{4.22}$$

where μ is the mass of the oscillating particle. When the Morse oscillator serves as a model of a diatomic molecule, μ stands for the reduced mass of both nuclei $\mu = (1/m_1 + 1/m_2)^{-1}$ (Appendix I on p. 971). As we can see, the energy of the oscillator never equals zero (similar to the harmonic oscillator) and that

> the separation between consecutive energy levels decreases.

The wave functions are slightly more complicated than those for the harmonic oscillator and are given by the formula:

$$\psi_v = N_v e^{-\frac{z}{2}} z^{b_v} L_v^{2b_v}(z), \tag{4.23}$$

where the normalization coefficient

$$N_v = \sqrt{\frac{2b_v v!}{\Gamma(2b_v + v + 1)}} \quad \text{with } \Gamma(z) = \int_0^\infty e^{-t} t^{z-1}\, \mathrm{d}t,$$

z is a real number related to displacement x by the formula $z = 2ae^{-\alpha x}$, while

$$a = \frac{\sqrt{2\mu D}}{\alpha}, \tag{4.24}$$

$$b_v = a - \frac{1}{2} - v > 0. \tag{4.25}$$

> The above condition gives maximum $v = v_{\max}$ and, therefore, $v_{\max} + 1$ is the number of eigenfunctions. Thus, we always have a finite number of energy levels.

---

[37] Let us recall that, for the harmonic oscillator $2\pi\nu = \sqrt{\frac{k}{\mu}}$, therefore, from (4.19) $h\nu = \hbar\alpha\sqrt{\frac{2D}{\mu}}$, while $\hbar = 1$ a.u.

L stands for the polynomial given by the formula

$$L_n^c(x) = \frac{1}{n!} e^x x^{-c} \frac{\mathrm{d}^n}{\mathrm{d}x^n}\left(e^{-x} x^{n+c}\right), \tag{4.26}$$

where $n = 0, 1, 2, \ldots$ is the polynomial degree.[38] A short exercise gives

$$\begin{aligned} &L_0^c(x) = 1, \\ &L_1^c(x) = (c+1) - x, \\ &L_2^c(x) = \frac{1}{2}x^2 - (c+2)x + \frac{1}{2}(c+1)(c+2), \\ &\ldots \end{aligned}$$

This means the number of nodes in a wave function is equal to v (as in the harmonic oscillator). The lowest energy level is, therefore, nodeless.

4.5.3 COMPARISON WITH THE HARMONIC OSCILLATOR

For very large well depths (D), the parameter β of eq. (4.22) becomes very small. This results in E_v approaching the corresponding formula for the harmonic oscillator $-D + h\nu(v + 1/2)$, and the energy levels become equidistant with the nearest neighbour separation equal to $h\nu$. The potential is highly anharmonic (of the "hook-type"), but the energy levels would be equidistant as in the harmonic oscillator. Is it possible? Yes, it is. The key is that, for small values of v, the term $-h\nu(v + 1/2)^2\beta$ does not yet enter into play and low-energy levels correspond to small amplitudes (x) of vibrations. For small x, the potential is close to parabolic[39] – as for the harmonic oscillator with force constant k.

4.5.4 THE ISOTOPE EFFECT

As we can see from eq. (4.21), $h\nu$ is large for narrow (large α) and deep (large D) wells, and for light oscillators (small μ). In particular, when we consider the ground states of two molecules that differ by an isotope substitution, the molecule with the heavier isotope (larger μ), would have lower energy than that corresponding to the light-isotope. This may be seen from the approximate formula for the energy $-D + \frac{1}{2}h\nu$ (*zero-vibration energy*).[40]

zero-vibration energy

This effect was also present in the harmonic oscillator. When β becomes larger this picture is modified. The larger ν, the larger the modification of the energies of the stationary states (see the last term in the formula for E_v).

Fig. 4.15 shows three different Morse curves and the calculated energy levels.

---

[38] Indeed, n-time derivation gives $e^{-x}x^{n+c}$ as a term with the highest power of x. Multiplication by $e^x x^{-c}$ gives x^n.

[39] As witnessed by a Taylor expansion of $V(x)$ for $x = 0$.

[40] We have to stress that V is almost identical for both molecules, as will be explained in Chapter 6. The energy difference comes, therefore, mainly from the zero-vibration (i.e. $v = 0$) energy difference.

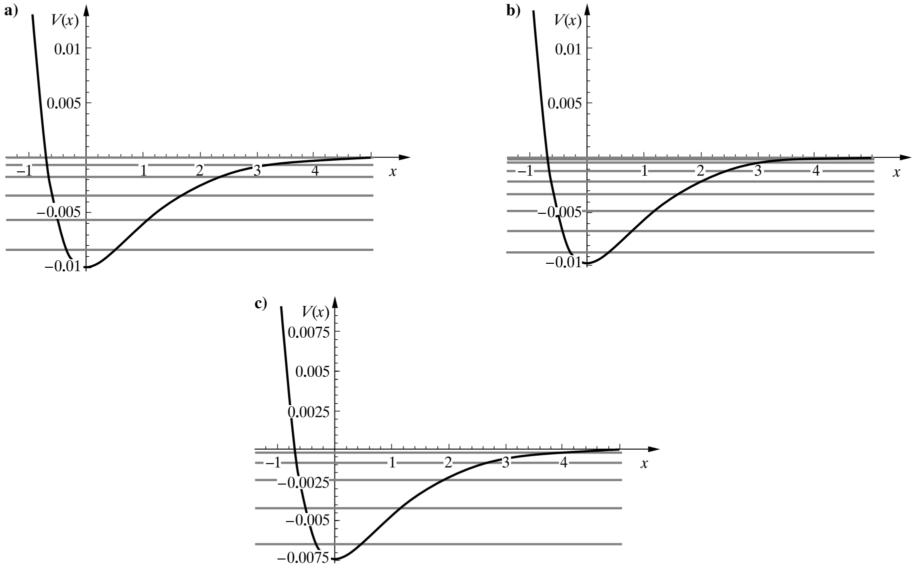


Fig. 4.15. The isotope effect and the effect of bond weakening. The Morse curve (a) corresponds to $D = 0.01$ a.u. and $\alpha = 1$. The calculated energy levels correspond to a proton mass $\mu = 1840$ a.u. The Morse curve (b) is identical, but an isotope substitution (deuteron instead of the proton) has been made. As a result we obtain a larger number of vibrational levels, the levels are closer each other and the system becomes more stable. The Morse curve (c) corresponds to $D = 0.008$ a.u., i.e. it is less deep by 20% with respect to curve (a). As we can see, the number of stationary states has diminished.

The two first curves are identical (Fig. 4.15 a and b) and illustrate the isotope effect. When calculating the energy levels in case of a (or b), the reduced mass of the proton (or deuteron) has been taken.[41] As we can see in the deuteron case, the number of energy levels has increased (from 6 to 9), the levels lowered and have became closer, when compared to the proton case.

ISOTOPE EFFECT (after substitution by a heavy isotope)
results mainly in decreasing the zero-vibration energy, as well as in lowering and condensing higher energy levels.

Eq. (4.21) shows that the ratio of the OD bond frequency to the OH bond frequency is equal to the square root of the inverse reduced masses of D-substituted

---

[41] Why these masses? Let us imagine that the oscillators are the molecules OH and OD. The reduced masses for these two cases are equal to 0.94 and 1.78 of the proton mass, respectively, which is *close* to the proton and deuteron masses, respectively (these have been used in the example). In the system of two nuclei (the heavy oxygen and the light hydrogen) the light atom (hydrogen) is mainly responsible for the oscillatory motion, which justifies the choice made.

and H-substituted compounds, which may be approximated as:[42]

$$\frac{\nu_{OD}}{\nu_{OH}} \cong \sqrt{\frac{m_H}{m_D}} \cong 0.7. \tag{4.27}$$

The red shift ($\nu_{OD} < \nu_{OH}$) in the IR spectrum represents one of the main characteristics of deuteration. The effect is used in spectroscopy to check whether a band is indeed the OH band. In such a case, the substance is dissolved in heavy water, and after a while the OH functional groups are replaced by OD functional groups. If the IR spectrum is registered again and compared with the previous one showing a red shift in agreement with (4.27), this proves that indeed we were concerned with an OH band.

4.5.5 BOND WEAKENING EFFECT

The condition $b_v > 0$ determines the number of vibrational levels, which may be accommodated by a potential well. This number is always finite. The key number, b_v, depends on a, whereas a is determined by μ, D and α. First of all, we can see that we may have a beautiful well which is unable to accommodate even a single vibrational energy level. This may happen if $b_0 < 0$, which, as seen from (4.25), requires $a \leqslant \frac{1}{2}$. Such a condition may be fulfilled because of a too shallow well (small D), or too light nuclei (small μ) or a too narrow well (large α). Even if in such a case there is no stationary vibrational state, such a shallow potential energy well may be detected in scattering experiments through the appearance of some resonance states. Such states are called *encounter complexes*.

encounter complexes

The third curve (Fig. 4.15.c) differs only by reducing the binding energy (D) with respect to the first curve, which in real situations is similar to a bond weakening. As we can see, the number of stationary states has decreased to 5. We may imagine that, in the extreme, the curve may become very shallow and unable to accommodate any vibrational level. In such a case, even if the binding energy (i.e. the well depth) is non-zero, the molecule will sooner or later dissociate.

4.5.6 EXAMPLES

Example 2. *Hydrogen molecule*

The hydrogen molecule has been investigated in detail. As will be seen in Chapters 6 and 10 the theory challenges some very subtle experiments. Let us approximate the most accurate theoretical potential energy curve[43] (as a function of the internuclear distance R) by a Morse curve.

Is such an approximation reasonable? Let us see. From Wolniewicz's calculations we may take the parameter $D = 109.52$ kcal/mol$= 38293\ \text{cm}^{-1}$, while the

---

[42] In our example from Fig. 4.15 the ratio equals 0.73, while simply taking the masses instead of the reduced masses, gives this ratio equal 0.71.

[43] L. Wolniewicz, *J. Chem. Phys.* 103 (1995) 1792.

parameter α is chosen in such a way as to reproduce the theoretical binding energy for $R = R_e + 0.4$ a.u.,[44] where $R_e = 1.4$ a.u. is the position of the minimum binding energy. It turns out that, say, "by chance" this corresponds to $\alpha = 1$. From eq. (4.24) and eq. (4.25) we obtain $a = 17.917$ and the allowed v are those satisfying the inequality $b_v = 17.417 - v > 0$. We expect, therefore, 18 energy levels with $v = 0, 1, \ldots, 17$ for H_2 and 25 energy levels for T_2 (in the last case $b_v = 24.838 - v > 0$). Accurate calculations of Wolniewicz give 14 vibrational levels for H_2, and 25 levels for T_2. Moreover, from eq. (4.21) we obtain for H_2: $h\nu = 0.019476$ a.u.$= 4274$ cm^{-1}, while from eq. (4.22) we have $\beta = 0.0279$. From these data one may calculate the energetic gap between the ground ($v = 0$) and the first excited state ($v = 1$) for H_2, $\Delta E_{0\to 1}$, as well as between the first and the second excited states, $\Delta E_{1\to 2}$. We get:

$$\Delta E_{0\to 1} = h\nu - h\nu\big[(1+1/2)^2 - (0+1/2)^2\big]\beta = h\nu(1-2\beta),$$

$$\Delta E_{1\to 2} = h\nu - h\nu\big[(2+1/2)^2 - (1+1/2)^2\big]\beta = h\nu(1-4\beta).$$

Inserting the calculated $h\nu$ and β gives $\Delta E_{0\to 1} = 4155$ cm^{-1} and $\Delta E_{1\to 2} =$ 3797 cm^{-1}. The first value agrees very well with the experimental value[45] 4161 cm^{-1}. Comparison of the second value with the measured value 3926 cm^{-1} is a little bit worse, although it is still not bad for our simple theory. The quantity D represents the *binding energy*, i.e. the energy difference between the well bottom and the energy of the dissociated atoms. In order to obtain the dissociation energy we have to consider that the system does not start from the energy corresponding to the bottom of the curve, but from the level with $v = 0$ and energy $\frac{1}{2}h\nu$, hence our estimation of the *dissociation energy* is $E_{\text{diss}} = D - \frac{1}{2}h\nu = 36156$ cm^{-1}, while the experimental value amounts to 36118 cm^{-1}.

binding energy

dissociation energy

Example 3. *Two water molecules*

The above example pertains to a chemical bond. Let us take, in the same way, a quite different situation where we have relatively weak intermolecular interactions, namely the hydrogen bond between two water molecules. The binding energy in such a case is of the order of $D = 6$ kcal/mol $= 0.00956$ a.u. $= 2097$ cm^{-1}, i.e. about twenty times smaller as before. To stay within a single oscillator model, let us treat each water molecule as a point-like mass. Then, $\mu = 16560$ a.u. Let us stay with the same value of $\alpha = 1$. We obtain (p. 171) $a = 17.794$ and hence $b_0 =$ 17.294, $b_1 = 16.294, \ldots, b_{17} = 0.294$, $b_{n>17} < 0$. Thus, (accidentally) we also have 18 vibrational levels.

This time from (4.21), we have $h\nu = 0.001074$ a.u $= 235$ cm^{-1}, and $\beta = 0.02810$ a.u., therefore $\Delta E_{0\to 1} = 222$ cm^{-1} and $\Delta E_{1\to 2} = 209$ cm^{-1}. These numbers have the same order of magnitude as those appearing in the experiments (cf. p. 303).

---

[44] This choice is of course arbitrary.

[45] I. Dabrowski, *Can. J. Phys.* 62 (1984) 1639.

4.6 RIGID ROTATOR

A rigid rotator is a system of two point-like masses, m_1 and m_2, with a constant distance R between them. The Schrödinger equation may be easily separated giving two equations, one for the centre-of-mass motion and the other for the relative motion of the two masses (see Appendix I on p. 971). We are interested only in the second equation, which describes the motion of a particle of mass equal to the reduced mass of the two particles μ, and position in space given by the spherical coordinates R, θ, ϕ, where $0 \leqslant R < \infty, 0 \leqslant \theta \leqslant \pi, 0 \leqslant \phi \leqslant 2\pi$. The kinetic energy operator is equal to $-\frac{\hbar^2}{2\mu}\Delta$, where the Laplacian Δ represented in the spherical coordinates is given in Appendix H on p. 969. Since R is a constant, the part of the Laplacian which depends on the differentiation with respect to R is absent.[46] In this way we obtain the equation (equivalent to the Schrödinger equation) for the motion of a particle on a sphere:

$$-\frac{\hbar^2}{2\mu R^2}\left\{\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\right)+\frac{1}{(\sin\theta)^2}\frac{\partial^2}{\partial\phi^2}\right\}Y=EY, \tag{4.28}$$

where $Y(\theta,\phi)$ is the wave function to be found, and E represents the energy. This equation may be rewritten as (Y is also an eigenfunction of $\hat{J}^2$):

$$\hat{J}^2 Y = 2\mu R^2 E Y, \tag{4.29}$$

where $\hat{J}^2$ is the square of the angular momentum operator.

Eq. (4.28) may be also written as

$$\frac{1}{Y}\left\{\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial Y}{\partial\theta}\right)+\frac{1}{(\sin\theta)^2}\frac{\partial^2 Y}{\partial\phi^2}\right\}=\lambda,$$

where $\lambda = -\frac{2\mu R^2}{\hbar^2}E$. The solution of the equation is known in mathematics as a *spherical harmonic*,[47] it exists if $\lambda = -J(J+1)$, $J = 0, 1, 2, \ldots$:

spherical harmonics

$$\boxed{Y_J^M(\theta,\phi) = N_{JM}\cdot P_J^{|M|}(\cos\theta)\cdot\frac{1}{2\pi}\exp(iM\phi)} \tag{4.30}$$

---

[46] This reasoning has a heuristic character, but the conclusions are correct. Removing an operator is a subtle matter. In the correct solution to this problem we have to consider the two masses with a variable distance R with the full kinetic energy operator and potential energy in the form of the Dirac delta function (see Appendix E on p. 951) $-\delta(R-R_0)$.

[47] There are a few definitions of the spherical harmonics in the literature (see E.O. Steinborn, K. Ruedenberg, *Advan. Quantum Chem.* 7 (1973) 1). The Condon–Shortley convention often is used, and is related to the definition given above in the following way: $Y_J^M = \varepsilon_M[Y_J^M]_{CS}$, $Y_J^J = (-1)^J[Y_J^J]_{CS}$, where $\varepsilon_M = i^{|M|+M}$.

where $N_{JM}=\sqrt{\frac{(2J+1)}{2}\frac{(J-|M|)!}{(J+|M|)!}}$ is the normalization coefficient, and P is the *associated Legendre polynomial*[48] defined as

associated Legendre polynomials

$$P_J^{|M|}(x)=\left(1-x^2\right)^{\frac{|M|}{2}}\frac{\mathrm{d}^{|M|}}{\mathrm{d}x^{|M|}}P_J(x) \tag{4.31}$$

with the *Legendre polynomial*

Legendre polynomial

$$P_J(x)=\frac{1}{2^J J!}\frac{\mathrm{d}^J}{\mathrm{d}x^J}\left(x^2-1\right)^J. \tag{4.32}$$

From the uniqueness of the solution (Fig. 2.5.g,h) it follows that M has to be an integer.[49] The solution exists if $J=0,1,2,3,\ldots$, and from the analysis of the associated Legendre polynomials, it follows that M cannot exceed[50] J, because otherwise $Y=0$. The energetic levels are given by

quantum number M

$$\boxed{E_J=J(J+1)\frac{\hbar^2}{2\mu R^2}\quad \text{for } J=0,1,2,\ldots \tag{4.33}}$$

It is seen that the lowest energy level ($J=0$) corresponds to $Y_0^0=\text{const}$ (the function is, of course, nodeless, Fig. 4.16.a). This means that all orientations of the rotator are of equal probability. The first excited state corresponds to $J=1$ and is triply degenerate, since $M=0,\pm1$. The corresponding wave functions are: $Y_1^0=\cos\theta$, $Y_1^1=\sin\theta\exp(i\phi)$, $Y_1^{-1}=\sin\theta\exp(-i\phi)$. The first function, being real, may be easily plotted (Fig. 4.16.b), while the second and the third are not (they are complex). Since they both correspond to the same eigenvalue of the Hamiltonian, their arbitrary linear combination is an equally good eigenfunction of this operator. We may take, therefore, Y_1^1 and Y_1^{-1} as $\psi_1=\frac{1}{2}(Y_1^1+Y_1^{-1})=\sin\theta\cos\phi$ and $\psi_2=\frac{1}{2i}(Y_1^1-Y_1^{-1})=\sin\theta\sin\phi$. Both functions are real, they are shown in Fig. 4.16.c,d. Note, that again we have the usual situation: the ground state is nodeless, the first excited state has a single node, etc.

quantum number J

Y_J^M is not only the eigenfunction of the Hamiltonian $\hat{H}$ and of the square of the angular momentum $\hat{J}^2$ (with the eigenvalue $J(J+1)\hbar^2$) but also of the z component of the angular momentum operator:

$$\hat{J}_z Y_J^M=M\hbar Y_J^M. \tag{4.34}$$

---

[48] Adrien Legendre (1752–1833), French mathematician, professor at the Ecole Normale Superieure – an elite school of France founded by Napoleon Bonaparte.

[49] Indeed, since ϕ is an angle, we have $\exp(iM\phi)=\exp[iM(\phi+2\pi)]$. Hence, $\exp(iM2\pi)=1$, and, therefore, $\cos(2\pi M)=1$ and $\sin(2\pi M)=0$. This is fulfilled only if M is an integer.

[50] $P_J(x)$ is a polynomial of J-th degree, while $\frac{\mathrm{d}^{|M|}}{\mathrm{d}x^{|M|}}$ in $P_J^{|M|}(x)$ decreases the degree by M. If $|M|$ exceeds J, then $P_J^{|M|}(x)$ automatically becomes equal to zero.

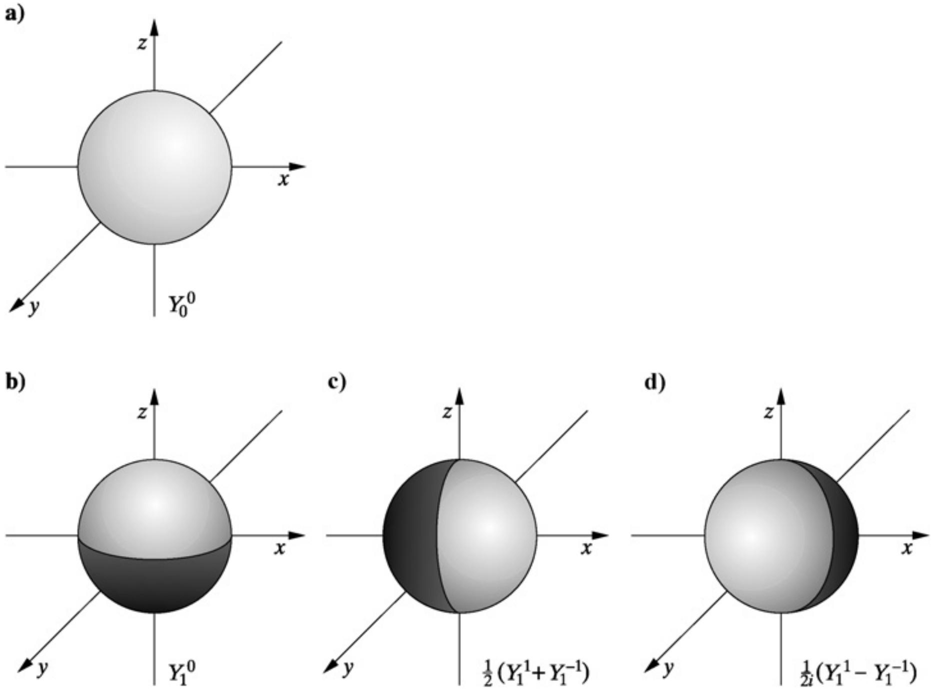


Fig. 4.16. A schematic representation of the nodes for rigid rotator wave functions: (a) ground state (nodeless), $J=0$ (b) triply degenerate first excited state (single node), $J=1$.

4.7 HYDROGEN-LIKE ATOM

We have two particles: an electron of mass m and charge $-e$ and a nucleus of mass M and charge $+Ze$. The Hamiltonian contains two kinetic energy operators and the Coulombic interaction $-Ze^2/r$, where r is the electron–nucleus separation. We have, therefore, 6 coordinates. In Appendix I on p. 971, it is shown how the centre-of-mass motion can be separated (we are not interested in this motion). There remain three coordinates, x, y, z, showing where the electron is with respect to the nucleus. The resulting Schrödinger equation contains a single kinetic energy operator of a particle of reduced mass μ (almost equal to the electron mass) with coordinates x, y, z, and Coulombic interaction of the electron and the nucleus (as before). Now, instead of x, y, z, we introduce the spherical coordinates r, θ, ϕ. Then, as in the class Q solution, we obtain

$$\boxed{\psi_{nlm}(r,\theta,\phi)=N_{nl}R_{nl}(r)Y_l^m(\theta,\phi) \qquad (4.35)}$$

where Y_l^m is identical to the solution (4.30) of a rigid rotator of length r, and the function R_{nl} has the following form in a.u. (N_{nl} is the normalization constant)

$$R_{nl}(r) = r^l L_{n+l}^{2l+1}\left(\frac{2Zr}{na_0}\right)\exp\left(-\frac{Zr}{na_0}\right), \tag{4.36}$$

where the *Bohr first orbit radius*

Bohr first orbit radius

$$a_0 = \frac{1}{\mu} \simeq 1 \text{ a.u.}, \tag{4.37}$$

where

principal quantum number $n = 1, 2, 3, \ldots,$
azimuthal quantum number $l = 0, 1, 2, \ldots, n-1,$
magnetic quantum number $m = -l, -l+1, \ldots, 0, \ldots, +l.$

and the *associated Laguerre polynomial* $L_\alpha^\beta(x)$ is defined as

associated Laguerre polynomial

$$L_\alpha^\beta(x) = \frac{\mathrm{d}^\beta}{\mathrm{d}x^\beta} L_\alpha(x), \tag{4.38}$$

while the *Laguerre polynomial* is given by[51]

$$L_\alpha(x) = \exp(x)\frac{\mathrm{d}^\alpha}{\mathrm{d}x^\alpha}\big[x^\alpha \exp(-x)\big]. \tag{4.39}$$

The one-electron wave functions (orbitals) of the hydrogen atom with $l = 0$ are traditionally denoted as ns: $\psi_{100} = 1\mathrm{s}$, $\psi_{200} = 2\mathrm{s}, \ldots$, with $l = 1$ as np, with $l = 2, 3, \ldots$ as nd, nf, ...

The wave functions ψ_{nlm} can be plotted in several ways. For example, the function $(nlm) = (100)$ or 1s given by the formula

$$\psi_{100}(r, \theta, \phi) = \sqrt{\frac{Z^3}{\pi}}\exp(-Zr), \tag{4.40}$$

and can be visualized in several alternative forms shown in Fig. 4.17.

We see that what the electron likes most is to sit on the nucleus. Indeed, if we chopped the space into tiny cubes, then computed the value of $(1s)^2$ in each cube (the function is real, therefore, the modulus is irrelevant), and multiplied

---

[51] Both $L_\alpha^\beta(x)$ and L_α are indeed polynomials of $\alpha - \beta$ degree. If $\beta > \alpha$, from (4.38) it follows that $L_\alpha^\beta = 0$.

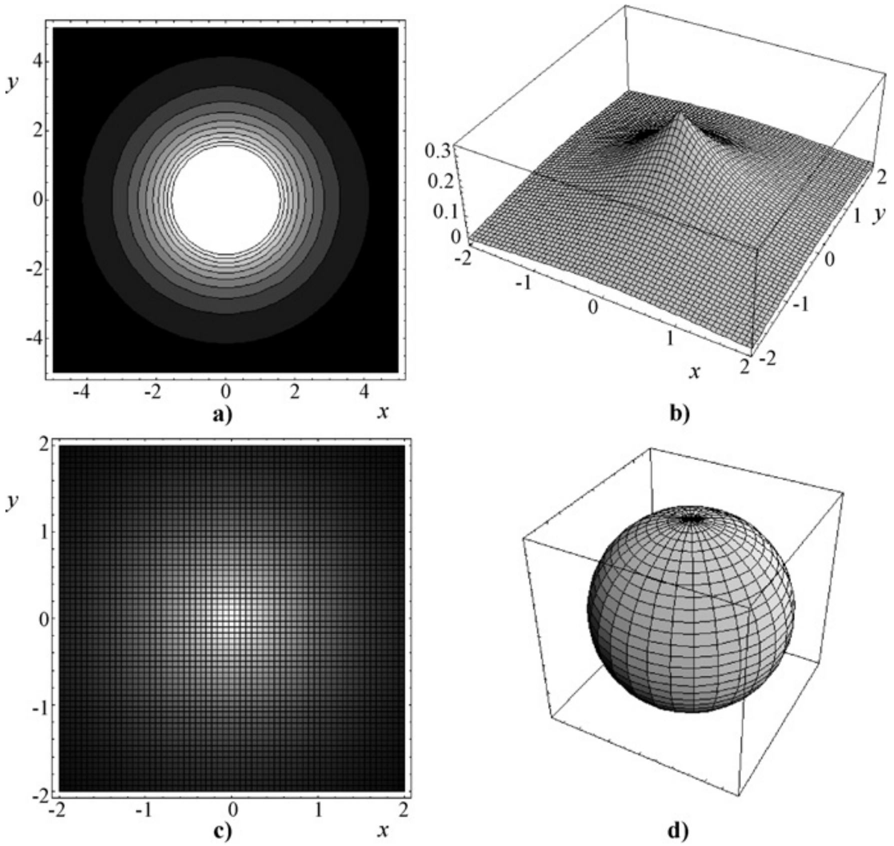


Fig. 4.17. Various ways of visualization the 1*s* hydrogen orbital, which is a function of electron position in 3D space (coordinates in Figs. (a)–(c) in a.u.). (a) Isolines of the $z = 0$ section of the wave function (going through the nucleus). Background colour means the value zero, white colour means a high value. This is a map of a mountain. The centre of Fig. (a) shows a large white plateau, which represents, however, an artifact. In fact (b), the section of the 1*s* orbital as a function of *r* represents a mountain with a sharp summit (a discontinuity of the first derivative). Fig. (c) is similar to (a), but instead of isolines we have a white mist with the highest concentration in the centre, disappearing exponentially with increasing distance *r*. Fig. (d) shows a spherically symmetric isosurface of the wave function.

the number obtained by the volume of the cube, the resulting number in each cube would have a meaning of the probability of finding the electron in a particular cube. Evidently, this number will be largest for the cube that contains the nucleus (the origin). We were told about the Bohr model[52] in school, about the orbits, and about the first Bohr orbit (corresponding to the atom ground state).

---

[52] Nobody is perfect. Geniuses also… Here is a story by John Slater: "… Brillouin delivered an interesting lecture concerning his relations. When he finished, Bohr stood up and attacked him with an inhuman fury. I have never heard any adult to scold another person in public with such an emotional engagement without any reason whatsoever. After this show I have decided that my antipathy with respect to Bohr dating since 1924 continues to be justified."

Do we relegate all this to fairy-tales? Not completely. If we changed the question: what is the *distance* at which the electron is most likely to be found, then the answer should indeed be as we were taught in school: the first Bohr orbit. This is easy to show by computing the *radial probability density* of finding the electron (i.e. integrating over all orientations, leaving the dependence on the distance): $\rho(r) = \int \mathrm{d}\theta\, \mathrm{d}\phi\, r^2 \sin\theta |\psi_{100}|^2 = 4Z^3 r^2 \exp -2Zr$, i.e. such a function computed for a given $r = r_0$ and multiplied by the volume $4\pi r_0^2\, \mathrm{d}r$ confined between two concentric spheres, one with radius r_0, the other with radius $r_0 + \mathrm{d}r$, gives the probability of finding the electron exactly between these spheres. The maximum of $\rho(r)$ corresponds exactly to $r = 1$ a.u. or the first Bohr orbit radius.[53]

Since the Hamiltonian commutes with the square of the total angular momentum operator $\hat{J}^2$ and with the operator of $\hat{J}_z$ (cf. Chapter 2 and Appendix F on p. 955), then the functions ψ_{nlm} are also the eigenfunctions of these two operators:

$$\hat{H}\psi_{nlm} = E_n \psi_{nlm}, \tag{4.41}$$

$$\hat{J}^2\psi_{nlm} = l(l+1)\hbar^2 \psi_{nlm}, \tag{4.42}$$

$$\hat{J}_z\psi_{nlm} = m\hbar \psi_{nlm}, \tag{4.43}$$

where in a.u.

$$E_n = -\frac{Z^2}{2n^2}\left(\frac{1}{1+\frac{1}{M_p}}\right), \tag{4.44}$$

with M_p representing the proton mass (in a.u.), i.e. about 1840. The content of the parenthesis in the last equation works out as 0.999457, almost 1, which would be obtained for an infinite mass of the nucleus.

Each of the energy levels is n^2-fold degenerate. Note, that the hydrogen atom energy depends solely on the principal quantum number n. The fact that the energy does not depend on the projection of the angular momentum $m\hbar$ is natural, because the space is isotropic and no direction is privileged. However, the fact that it does not depend on the length of the angular momentum $\sqrt{l(l+1)}\hbar$, is at first sight strange. The secret is in the Coulombic potential $\frac{1}{r}$ produced by the *point-like* nucleus and is connected with the notion of dynamic symmetry mentioned on p. 76. If we considered a non-point-like nucleus or were interested in the orbital $2s$ of such a quasi-hydrogen atom as lithium,[54] then the energy would depend on the quantum number l, e.g., would be different for orbitals $2s$ and $2p$.

The $2s$ orbital ($n = 2$, $l = 0$, $m = 0$) reads as (Fig. 4.18)

$$\psi_{2s}(r,\theta,\phi) = N_{2s}(Zr-2)\exp(-Zr/2) \tag{4.45}$$

---

[53] The computed maximum position does not coincide with the mean value of r (cf. Appendix H on p. 969) $\langle r\rangle = \langle\psi_{100}|r\psi_{100}\rangle = \int_0^\infty \mathrm{d}r\, r\rho(r) = \frac{3}{2}$ a.u.

[54] In which the nucleus is screened by a cloud of two 1s electrons. The 2s electron thinks it is in a hydrogen atom with a spacious nucleus of the size of the 1s orbital and an effective charge +1.

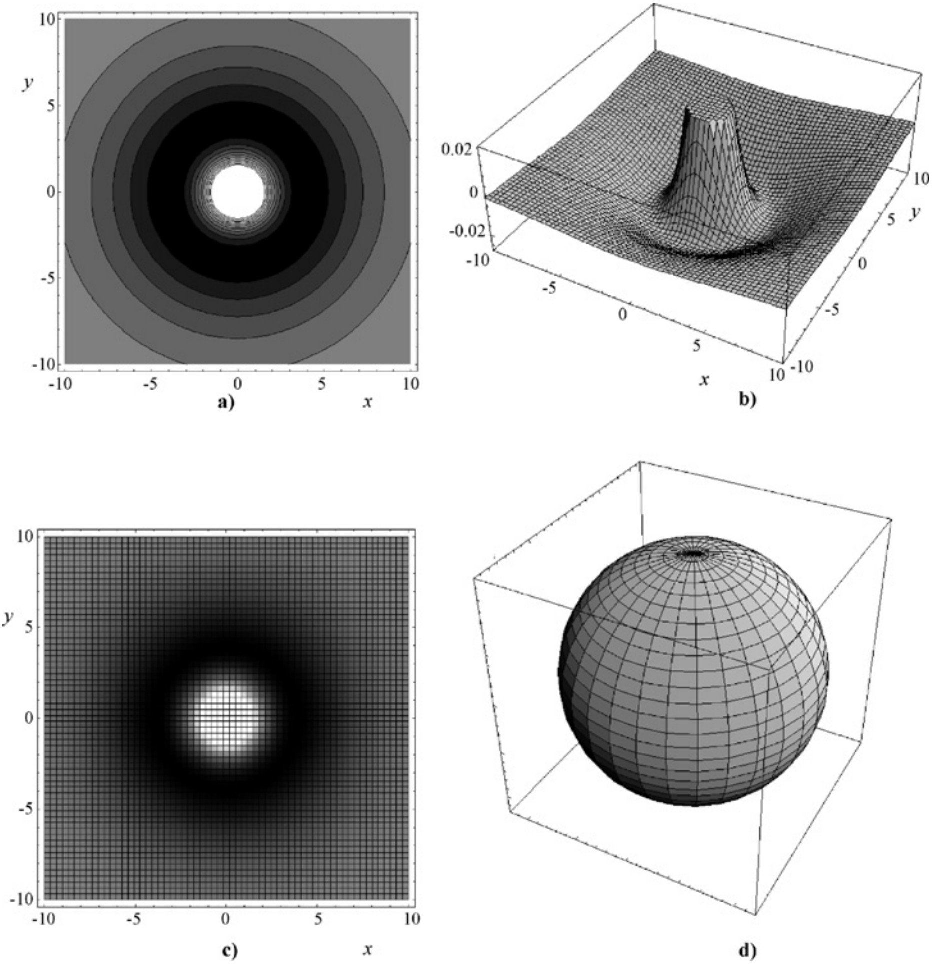


Fig. 4.18. Various graphical representations of the hydrogen 2*s* orbital (coordinates in Figs. (a)–(c) in a.u.). (a) Isolines of $z = 0$ section of the orbital. Gray means the value zero, white a high value, black a negative value. Note, that gray is not only at peripheries, but also around the centre. This means that 2*s* orbital exhibits a nodal sphere of radius 2 a.u. (see eq. (4.45)), that contains a little more than 5% of the electronic density (whereas for the 1*s* orbital such a sphere contains about 75% of electron cloud). The centre of the figure a shows a quite large white plateau, which represents, however, an artifact. In fact, (b) showing the section of 2*s* orbital as a function of *r* represents a mountain with a sharp peak (a discontinuity of the first derivative is shown incorrectly on the figure, instead a sharp summit one has an artifact plateau) with a depression at its base. Fig. (c) is similar to (a), but instead of isolines one has a white mist with the largest concentration in the centre, then taking the negative values (black mist) and finally disappearing exponentially with increasing distance *r*. Fig. (d) shows a spherically symmetric isosurface of the wave function (the sphere was shown as larger than 1*s* orbital, because 2*s* orbital decays more slowly than 1*s*).

with the normalization constant $N_{2s} = \frac{Z^{\frac{3}{2}}}{4\sqrt{2\pi}}$. A sphere of radius $2/Z$ (representing the nodal sphere) contains[55] only a little more than 5% of the total electronic density (independently of Z).[56]

The wave functions (orbitals) with $m \neq 0$ are difficult to draw, because they are complex. We may, however, plot the real part of ψ_{nlm} ($\mathrm{Re}\,\psi_{nlm}$) by taking the sum of ψ_{nlm} and ψ_{nl-m}, i.e. $2\mathrm{Re}\,\psi_{nlm}$ and the imaginary part of ψ_{nlm} ($\mathrm{Im}\,\psi_{nlm}$) from the difference of ψ_{nlm} and ψ_{nl-m} equal to $2i\,\mathrm{Im}\,\psi_{nlm}$. These functions are already real and can be easily plotted. In this way we obtain the orbitals $2p_x$ and $2p_y$ from the functions 211 and 21–1. The orbital 210 is identical to $2p_z$:

$$\begin{aligned}
2p_x &= N_{2p} x \exp(-Zr/2), \\
2p_y &= N_{2p} y \exp(-Zr/2), \\
2p_z &= N_{2p} z \exp(-Zr/2),
\end{aligned}$$

where an easy calculation (just five lines) gives the normalization constant $N_{2p} = ZN_{2s}$. The $2p$ orbitals are shown in Fig. 4.19.

Note, please, that a linear combination of eigenfunctions is not necessarily an eigenfunction. It does, if the functions mixed, correspond to the same eigenvalue. This is why $2p_x$ and $2p_y$ are the eigenfunctions of the Hamiltonian, of the square of the angular momentum operator, but are not eigenfunctions of $\hat{J}_z$.

Similarly we obtain the five real $3d$ orbitals. They can be easily obtained from eq. (4.35) and subsequently making them real by choosing $\mathrm{Re}\,\psi_{nlm}$ and $\mathrm{Im}\,\psi_{nlm}$. As a result we have the following normalized $3d$ orbitals ($N_{3d} = \frac{Z^{\frac{7}{2}}}{81}\sqrt{\frac{2}{\pi}}$)

$$\begin{aligned}
3d_{xy} &= N_{3d} xy \exp(-Zr/3), \\
3d_{xz} &= N_{3d} xz \exp(-Zr/3), \\
3d_{yz} &= N_{3d} yz \exp(-Zr/3), \\
3d_{x^2-y^2} &= \frac{1}{2} N_{3d} \left(x^2 - y^2\right) \exp(-Zr/3), \\
3d_{3z^2-r^2} &= \frac{1}{2\sqrt{3}} N_{3d} \left(3z^2 - r^2\right) \exp(-Zr/3).
\end{aligned}$$

The $3d$ orbitals are shown in Figs. 4.20 and 4.21. A summary of the hydrogen atomic orbitals is shown in Fig. 4.22.[57]

---

[55] See the Mathematica files for Chapter 4 in the Web Annex.

[56] A sphere of the same radius encloses about 75% of the electron density for the $1s$ orbital.

[57] A night bus ride might enrich us with some unexpected impressions. Of all atomic orbitals you may most easily "see" is the orbital $1s$. Just look through the condensation on a bus window at a single street lamp. You will see a gleam that decays to black night. You may also quite easily find a double lamp that will offer you a $2p$ orbital and sometimes have the chance to see some of the $3d$ orbitals. Once I have

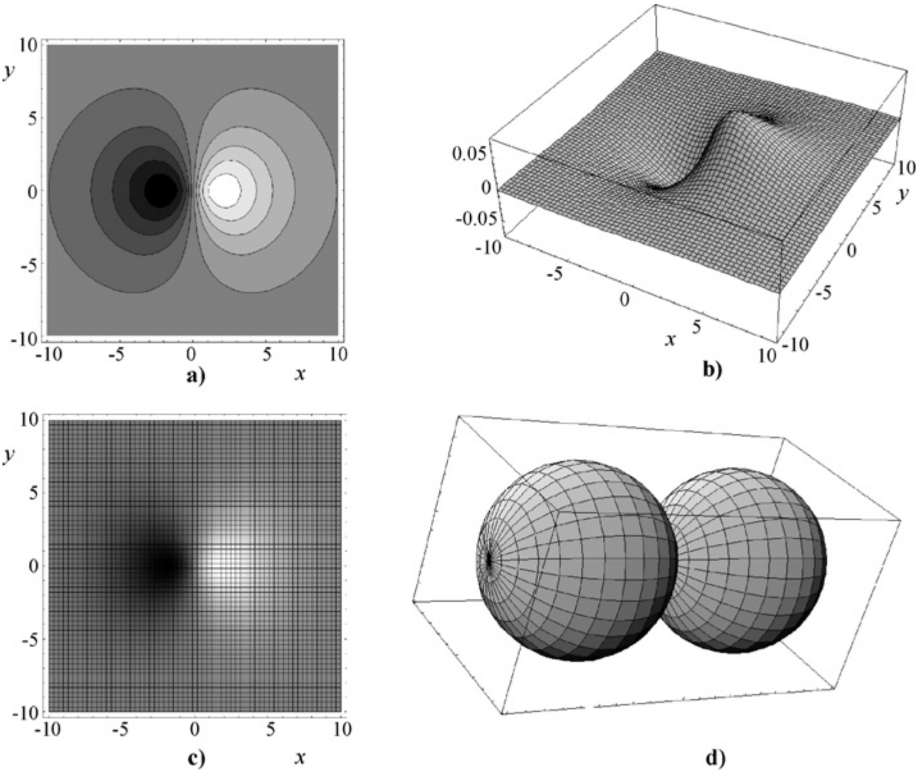


Fig. 4.19. Various graphical representations of the hydrogen $2p_x$ orbital (coordinates in Figs. (a)–(c) in a.u.). The two other $2p$ orbitals: $2p_y$, $2p_z$ look the same as $2p_x$, but are oriented along axes y and z. Note that for the hydrogen atom all four orbitals $2s$, $2p_x$, $2p_y$, $2p_z$ correspond to the same energy, and all have a single nodal surface. For $2s$ the surface (Fig. 4.18) is a sphere of radius 2, for the $2p_x$, $2p_y$ and $2p_z$ orbitals the nodal surfaces are the planes $x, y, z = 0$. (a) Isolines of the $z = 0$ section of the orbital. Gray means zero, white means a high value, black means a negative value. Note (Figs. (a), (b)), that the right-hand side of the orbital is positive, the left-hand side is negative. The maximum (minimum) value of the orbital is at $x = 2$ ($x = -2$) a.u. Fig. (c) is similar to (a), but instead of isolines we have a mist with the largest concentration (white) on the right and the smallest (and negative, black) concentration on the left. The orbital finally disappears exponentially with increasing distance r from the nucleus. Fig. (d) shows an isosurface of the absolute value of the angular part of the wave function $|Y_1^0|$. As for Y_1^0 itself, one of the lobes takes negative, the other positive values and they touch each other at the position of the nucleus. To obtain the orbital, we have to multiply this angular function by a spherically symmetric function of r. This means that an isosurface of the absolute value of the wave function will also have two lobes (for the wave function itself, one will be positive, the other negative), but they will not touch each other in full accordance with Fig. (a).

---

even found the $2s$ orbital, but I do not understand how it was possible. I was looking at a single lamp, which made an intense gleam in the centre, which gradually decayed and then again an annular gleam appeared that finally vanished. This is what the square of the $2s$ orbital looks like.

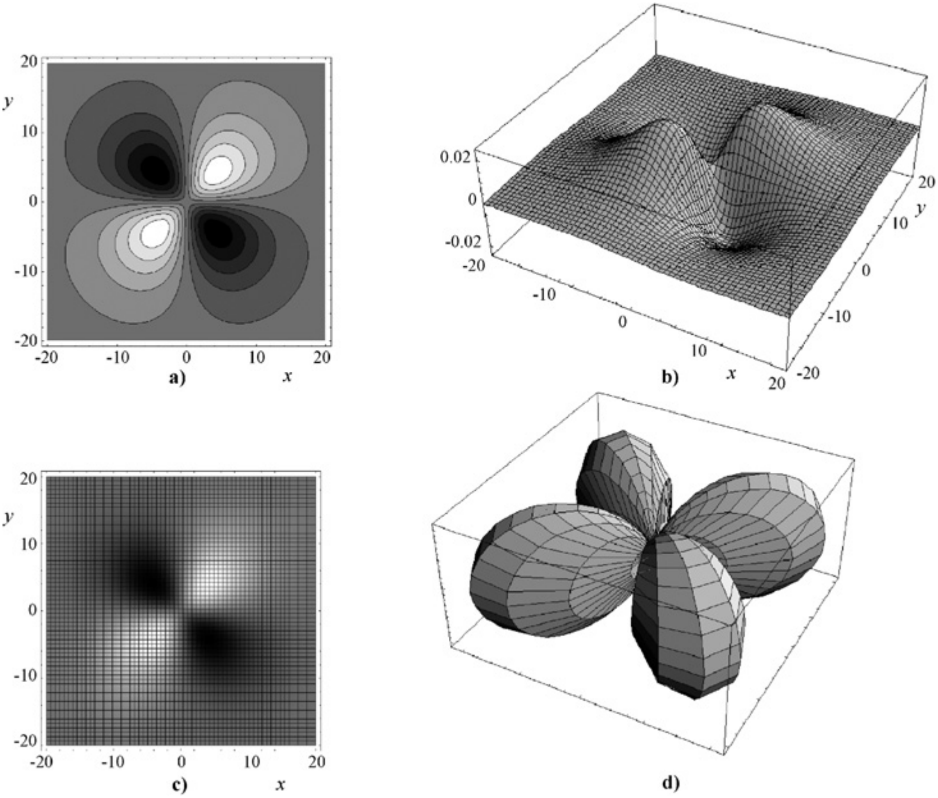


Fig. 4.20. Various graphical representations of the hydrogen $3d_{xy}$ orbital (coordinates in Figs. (a)–(c) in a.u.). The three other $3d$ orbitals: $3d_{yz}$, $3d_{xz}$ and $3d_{x^2-y^2}$ look the same as $3d_{xy}$, but are oriented in space according to their indices (in the last case along x and y axes). (a) Isolines of $z = 0$ section of the orbital. Gray means zero, white means a positive value, black means a negative value. Note (Figs. (a), (b)), that $3d$ orbitals are symmetric with respect to inversion. One may imagine the $z = 0$ section of $3d_{xy}$ as two hills and two valleys (b). Fig. (c) is similar to (a), but instead of isolines one has a white mist with the highest concentration on the North-East line and the smallest (and negative) concentration on the North-West line (black mist). The orbital finally disappears exponentially with increasing distance r from nucleus. Fig. (d) shows an isosurface of the absolute value of the angular part of the wave function: $|Y_2^2 - Y_2^{-2}|$. As for $Y_2^2 - Y_2^{-2}$ itself two of the lobes take the negative, the other two take positive values, they touch each other at nucleus. To obtain the orbital one has to multiply this angular function by a spherically symmetric function of r. This means that an isosurface of the absolute value of the wave function will have also four lobes (for the wave function itself two will be positive, the other two negative), but they will not touch in full accordance with Fig. (a).

4.8 HARMONIC HELIUM ATOM (HARMONIUM)

Two-electron systems already represent a serious problem for quantum chemistry, because the mutual correlation of electron motions must be carefully taken into account. As we will see in further chapters, such calculations are feasible, but the

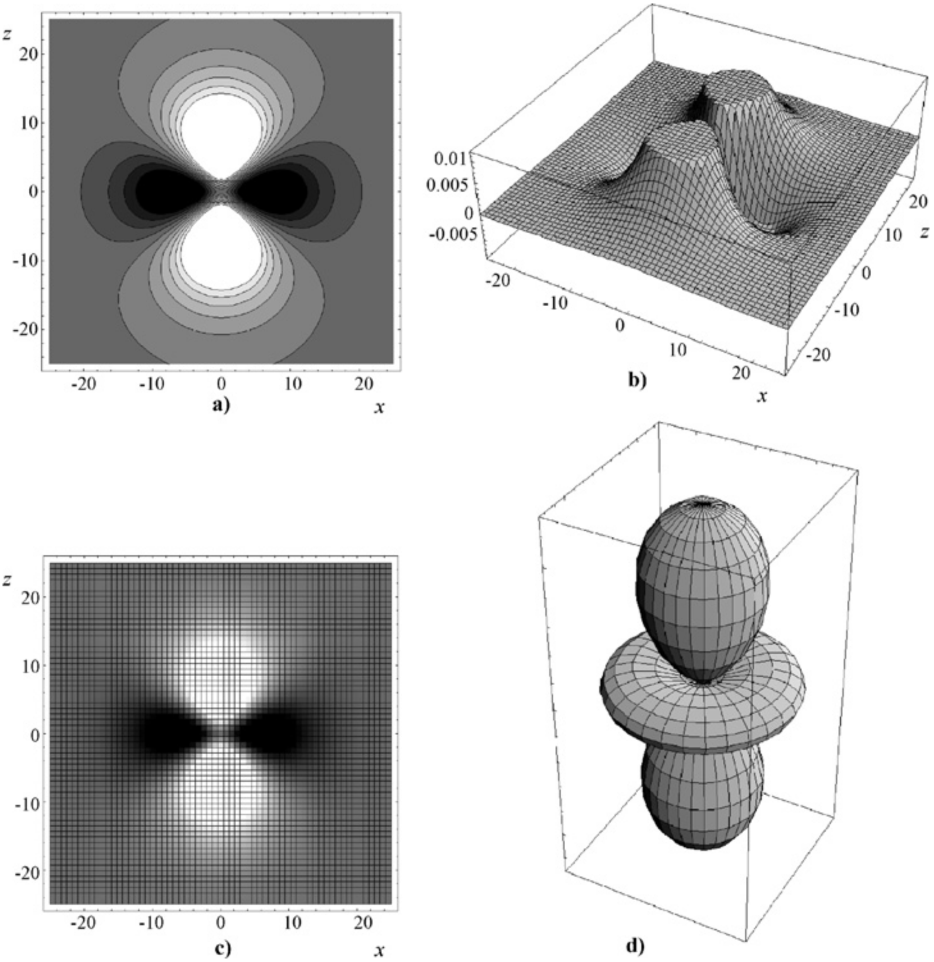


Fig. 4.21. Various graphical representations of the hydrogen $3d_{3z^2-r^2}$ orbital (coordinates in Figs. (a)–(c) in a.u.). The z axis is vertical, x axis is horizontal. (a) Isolines of the xz section of the orbital. Gray means zero, white means a high positive value, black means a negative value. Note (Figs. (a), (b)), that $3d_{3z^2-r^2}$ orbitals are symmetric with respect to inversion. We may imagine xz section of the $3d_{3z^2-r^2}$ as two hills and two valleys (b), the hills are higher than the depth of the valleys (the plateaus in Fig. (b) are artificial). Fig. (c) is similar to (a), but instead of isolines one has a mist with the highest concentration (white) on the North-South line and the smallest (and negative, black mist) concentration on the East-West line. The orbital finally disappears exponentially with increasing distance r from the nucleus. Fig. (d) shows an isosurface of the absolute value of the angular part of the wave function ($|Y_2^0|$). As for Y_2^0 itself there are two positive lobes and a negative ring, they touch each other at nucleus. To obtain the orbital, we have to multiply this angular function by a spherically symmetric function of r. This means that an isosurface of the absolute value of the wave function will have also two lobes along the z axis as well as the ring, but they will not touch in accordance with Fig. (a). The lobes along the z axis are positive, the ring is negative.

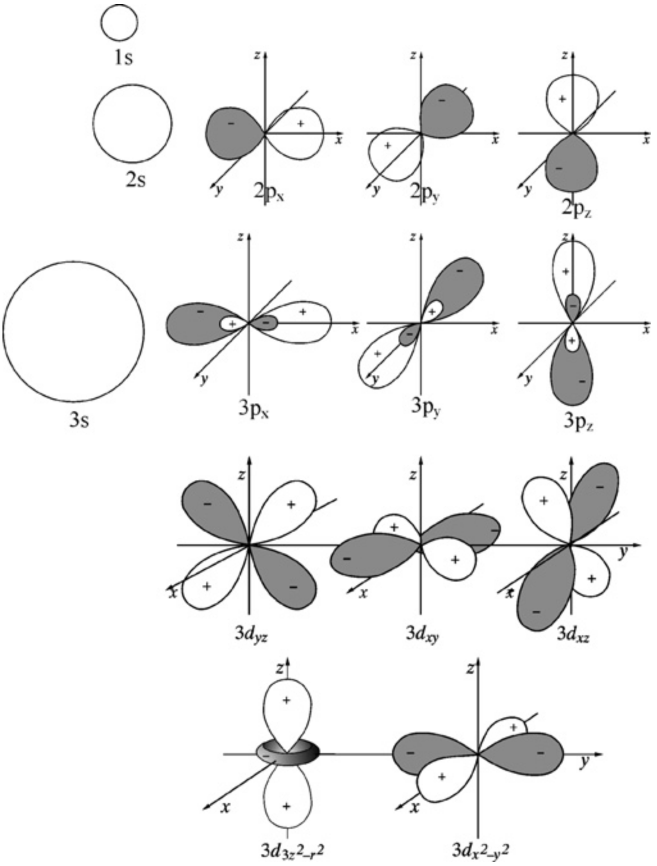


Fig. 4.22. Summary: schematic representation of $1s$, $2s$, $2p$, $3s$, $3p$, $3d$ orbitals of the hydrogen atom. $1s$, $2s$, $3s$ orbitals are spherically symmetric and increase in size; $1s$ has no node, $2s$ has one nodal sphere (not shown), $3s$ has two nodal spheres (not shown). The shadowed area corresponds to the "minus" sign of the orbital. The $2p$ orbitals have a single nodal plane (perpendicular to the orbital shape). $3p$ orbitals are larger than $2p$, and have a more complex nodal structure. Note that among $3d$ orbitals all except $3d_{3z^2-r^2}$ have identical shape, but differ by orientation in space. A peculiar form of $3d_{3z^2-r^2}$ becomes more familiar when one realizes that it simply represents a sum of two "usual" $3d$ orbitals. Indeed, $3d_{3z^2-r^2} \propto [2z^2 - (x^2 + y^2)]\exp(-Zr/3) \propto [(z^2 - x^2) + (z^2 - y^2)]\exp(-Zr/3) \propto (3d_{z^2-x^2} + 3d_{z^2-y^2})$.

wave functions are very complicated, e.g., may represent linear combinations of thousands of terms and still only be *approximations* to the exact solution to the Schrödinger equation. This is why people were surprised when Kais et al. showed that a two electron system has *an exact analytical solution*.[58]

Unfortunately, this wonderful two-electron system is (at least partially) nonphysical. It represents a strange helium atom, in which the two electrons (their distance denoted by r_{12}) interact through the Coulombic potential, but each is at-

---

[58] S. Kais, D.R. Herschbach, N.C. Handy, C.W. Murray, G.J. Laming, *J. Chem. Phys.* 99 (1993) 417.

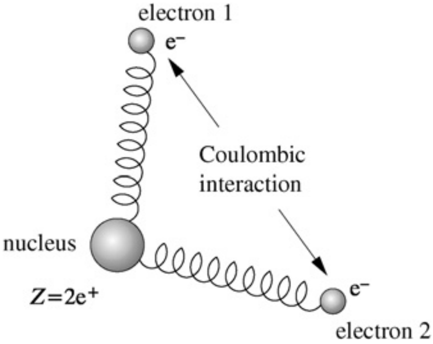


Fig. 4.23. The harmonic helium atom. The electrons repel by Coulombic forces and are attracted by the nucleus by a harmonic (non-Coulombic) force.

tracted to the nucleus by a harmonic spring (of equilibrium length 0 and force constant k, with electron–nucleus distances denoted by r_1 and r_2), Fig. 4.23.

The Hamiltonian of this problem (the adiabatic approximation and atomic units are used) has the form:

$$\hat{H} = -\frac{1}{2}\Delta_1 - \frac{1}{2}\Delta_2 + \frac{1}{2}k\left(r_1^2 + r_2^2\right) + \frac{1}{r_{12}}.$$

It is amazing in itself that the Schrödinger equation for this system has an analytical solution (for $k = \frac{1}{4}$), but it could be an extremely complicated analytical formula. It is a sensation that the solution is dazzlingly beautiful and simple

$$\psi(\boldsymbol{r}_1, \boldsymbol{r}_2) = N\left(1 + \frac{1}{2}r_{12}\right)\exp\left[-\frac{1}{4}\left(r_1^2 + r_2^2\right)\right],$$

where

$$|N|^2 = \frac{\pi^{\frac{3}{2}}}{8 + 5\sqrt{\pi}}.$$

The wave function represents the product of the two harmonic oscillator wave functions (Gaussian functions), but also an additional extremely simple correlation factor $(1 + \frac{1}{2}r_{12})$. As we will see in Chapter 13, exactly such a factor is required for the ideal solution. In this exact function there is nothing else, just what is needed.[59]

4.9 WHAT DO ALL THESE SOLUTIONS HAVE IN COMMON?

- In all the systems considered (except the tunnelling effect, where the wave function is non-normalizable), the stationary states are similar, the number of their nodes increasing with their energy (the nodeless function corresponds to the lowest energy).

---

[59] We might have millions of complicated terms.

- If the potential energy is a *constant* (particle in a box, rigid rotator), then the energy level (nearest-neighbour) distance *increases* with the energy.[60] The energy levels get closer for larger boxes, longer rotators, etc.
- *A parabolic* potential energy well (harmonic oscillator) reduces this tendency and the energy levels are *equidistant*. The distance decreases if the parabola gets wider (less restrictive).
- The Morse potential energy curve may be seen as a function that may be approximated (as the energy increases) by wider and wider parabolic sections. No wonder, therefore, that the energy level distance *decreases*. The number of energy levels is finite.[61]
- The Coulomb potential, such as that for the hydrogen atom, resembles vaguely the Morse curve. Yet its form is a little similar to the Morse potential (dissociation limit, but infinite depth). We expect, therefore, that the energy levels for the hydrogen-like atom will become closer and closer when the energy increases, and we are right. Is the number of these energy levels finite as for the Morse potential? This is a more subtle question. Whether the number is finite or not is decided by the asymptotics (the behaviour at infinity). The Coulomb potential makes the number infinite.

4.10 BEACONS AND PEARLS OF PHYSICS

Sometimes students, fascinated by handy computers available nowadays, tend to treat the simple systems described in this chapter as primitive and out of date. A Professor has taken them from the attic and after dusting off shows them to a class, whilst outside sophisticated computers, splendid programs and colourful graphs await. This is wrong. The simple systems considered in this chapter correspond to extremely rare *exact solutions of Schrödinger equation* and are, therefore, precious pearls of physics by themselves. Nobody will give a better solution, the conclusions are hundred percent sure. It is true that they all (except for the hydrogen atom) correspond to some idealized systems.[62] There is no such a thing as an unbreakable spring (e.g., harmonic oscillator) or a rotator, that does not change its length, etc. And yet these problems represent our firm ground or the beacons of our native land. After reading the present chapter we will be preparing our ship for a long voyage. When confronted with the surprises of new lands and trying to understand them

---

[60] In both cases the distance goes as the square of the quantum number.

[61] Such type of reasoning prepares us for confronting real situations. Practically, we will never deal with the abstract cases described in the present chapter, and yet in later years we may say something like this: "*look, this potential energy function is similar to case X in Chapter 4 of that thick boring book we have been forced to study. So the distribution of energy levels and wave functions has to be similar to those given there*".

[62] Like Platonic ideal solids.

the *only* points of reference or the beacons which tell us about *terra firma* will be the problems for which analytical solutions have been found.

Summary

Exact analytical solutions[63] to the Schrödinger equation play an important role as an organizer of our quantum mechanical experience. Such solutions have only been possible for some idealized objects. This is of great importance for the interpretation of approximate solutions for real systems. Another great feature of exact solutions is that they have an extremely wide range of applications: they are useful independently of whether we concentrate on an electron in an atom, in molecule, a nucleon in a nucleus or a molecule as an entity, etc.

The main features of the solutions are:

- **Free particle.** The particle may be described as the superposition of the state $\exp(i\kappa x)$, corresponding to the particle moving right (positive values of x), and the state $\exp(-i\kappa x)$, that corresponds to the particle moving left. Both states correspond to the same energy (and opposite momenta).
- **Particle in a box.** We consider first a particle in a 1D box, i.e. the particle is confined to section $[0, L]$ with potential energy (for a particle of mass m and coordinate x) equal to zero and ∞ outside the section. Such a potential forces the wave function to be nonzero only within the section $[0, L]$. We solve the elementary Schrödinger equation and obtain $\Psi = A\sin\kappa x + \cos\kappa x$, where $\kappa^2 = \frac{2mE}{\hbar^2}$. Quantization appears in a natural way from the condition of continuity for the wave function at the boundaries: $\Psi(0) = 0$ and $\Psi(L) = 0$. These two conditions give the expression for the energy levels $E_n = \frac{n^2h^2}{8mL^2}$ and for the wave functions $\Psi_n = \sqrt{\frac{2}{L}}\sin\frac{n\pi}{L}x$ with quantum number $n = 1, 2, \ldots$. Conclusion: the successive energy levels are more and more distant and the wave function is simply a section of the sine function (with 0 value at the ends).
- **Tunnelling effect.** We have a particle of mass m and a rectangular barrier (section $[0, a]$, width a and height V). Beyond this section the potential energy is zero. The particle comes from the negative x values and has energy $E < V$. A classical particle would be reflected from the barrier. However, for the quantum particle:
  - the transmission coefficient is non-zero,
  - the passage of a large energy particle is easier,
  - a narrower barrier means larger transmission coefficient,
  - the higher the barrier the smaller transmission coefficient.

The first feature is the most sensational, the others are intuitively quite acceptable. This is not the case for a particle tunnelling through two consecutive barriers. It turns out that (for a given interbarrier distance) *there are some "magic" energies of the particle (resonance energies), at which the transmission coefficient is particularly large*. The magic energies correspond to the stationary states that would be for a particle in a box a little longer than the interbarrier distance. The resonance states exist also for energies greater than barrier

---

[63] To distinguish from accurate solutions (i.e. received with a desired accuracy).

and have a transmission coefficients equal to 100%, whereas other energies may lead to reflection of the particle, even if they are larger than the barrier energy.

- **Harmonic oscillator.** A single particle of mass m attached to a harmonic spring (with force constant k) corresponds to potential energy $V = \frac{kx^2}{2}$. We obtain quantization of the energy: $E_v = h\nu(v + \frac{1}{2})$, where the vibrational quantum number $v = 0, 1, 2, \ldots$, and the angular frequency $\omega = 2\pi\nu = \sqrt{\frac{k}{m}}$. We see that *the energy levels are equidistant*, and their distance is larger for a larger force constant and smaller mass. The wave function[64] has the form of a Gaussian factor and a polynomial of degree v. The polynomial assures the proper number of nodes, while the Gaussian factor damps the plot to zero for large displacements from the particle equilibrium position. The harmonic oscillator may be viewed (Chapter 6) as equivalent (for small displacements) to two masses bound by a spring.
- **Morse oscillator.** The harmonic oscillator does not allow for the breaking of the spring connecting two particles, while the Morse oscillator admits dissociation. This is extremely important, because real diatomic molecules resemble the Morse rather than the harmonic oscillator. The solution for the Morse oscillator has the following features:
  - energy levels are non-degenerate,
  - their number is finite,
  - for large well depths the low energy levels tend to the energy levels of the harmonic oscillator (the levels are equidistant),
  - the higher the energy level the larger the displacement from the equidistant situation (the energy levels get closer),
  - the wave functions, especially those corresponding to deep-lying levels, are very similar to the corresponding ones of the harmonic oscillator,[65] but they do not exhibit the symmetry.[66]
- **Rigid rotator.** This is a system of two masses m_1 and m_2 that keeps their distance R fixed. After separating the centre-of-mass motion (Appendix I on p. 971) we obtain an equation of motion for a single particle of mass equal to the reduced mass μ moving on a sphere of radius R (position given by angles θ and ϕ). The energy is determined by the quantum number $J = 0, 1, 2, \ldots$ and is equal to $E_J = J(J+1)\frac{\hbar^2}{2\mu R^2}$. As we can see:
  - there is an infinite number of energy levels,
  - the separation of the energy levels increases with the energy (similar to the particle in a box problem),
  - the separation is smaller for larger masses,
  - the separation is smaller for longer rotators.

  The wave functions are the spherical harmonics $Y_J^M(\theta, \phi)$, which for low J are very simple, and for large J complicated trigonometric functions. The integer quantum number M satisfies the relation $|M| \leqslant J$. The energy levels are, therefore, $(2J+1)$-tuply degenerate.
- **Hydrogen-like atom.** We have an electron and a nucleus of charges $-e$ and $+Ze$, respectively, or -1 and $+Z$ in a.u. The wave function is labelled by three quantum numbers: principal $n = 1, 2, \ldots$, azimuthal $l = 0, 1, \ldots, (n-1)$ and magnetic $m = -l, (-l+$

---

[64] The energy levels are non-degenerate.

[65] Despite the fact, that the formula itself is very different.

[66] The wave functions for the harmonic oscillator are either even or odd with respect to the inversion operation ($x \to -x$).

1), …, 0, …, l. The energy in a.u. is given by the formula[67] $E_n = -Z^2/(2n^2)$. The wave function represents the product of a polynomial (of r), an exponential function decreasing with r and a spherical harmonic $Y_l^m(\theta, \phi)$, where r, θ, ϕ are the spherical coordinates of the electron, and the nucleus is at the origin. The wave functions that correspond to low energies are denoted by the symbols nl_m (with s for $l = 0$, p for $l = 1$, etc.): $1s$, $2s$, $2p_0$, $2p_1$, $2p_{-1}$, … The degeneracy of the n-th level is equal to n^2.

- **Harmonic helium atom.** In this peculiar helium atom the electrons are attracted to the nucleus by harmonic springs (of equal strength) of equilibrium length equal to zero. For $k = \frac{1}{4}$ an exact analytical solution exists. The exact wave function is a product of two Gaussian functions and a simple factor: $(1 + \frac{1}{2}r_{12})$ that correlates the motions of the two electrons.

Main concepts, new terms

free particle (p. 144)
particle in a box (p. 145)
box with ends (p. 145)
FEMO (p. 149)
cyclic box (p. 149)
tunnelling effect (p. 153)
current (p. 155)
transmission coefficient (p. 155)
resonance state (p. 155)
harmonic oscillator (p. 166)
Hermite polynomials (p. 166)
Morse oscillator (p. 169)
isotope effect (p. 172)
encounter complex (p. 174)
binding energy (p. 175)
dissociation energy (p. 175)
rigid rotator (p. 176)
spherical harmonics (p. 176)
Legendre polynomials (p. 176)
associated Legendre polynomials (p. 176)
hydrogen-like atom (p. 178)
Laguerre polynomials (p. 178)
associated Laguerre polynomials (p. 178)
correlation factors (p. 188)

From the research front

A field like that discussed in the present chapter seems to be definitely closed. We have been lucky enough to solve some simple problems, that could be solved, but others are just too complicated. This is not true. For several decades it has been possible to solve a series of non-linear problems, thought in the past to be hopeless. What decides success is: choice of the problem, quality of researchers, courage, etc.[68] It is worth noting that there are also attempts at a systematic search for promising systems to solve.

Ad futurum…

It seems that the number of exactly solvable problems will continue to increase, although the pace of such research will be low. If exactly solvable problems were closer and closer to practise of physics, it would be of great importance.

Additional literature

J. Dvořák and L. Skála, "Analytical Solutions of the Schrödinger Equation. Ground State Energies and Wave Functions", Collect. Czech. Chem. Commun., 63 (1998) 1161.

---

[67] An infinite number of levels.

[68] Already the Morse potential looks very difficult to manage, to say nothing about the harmonic helium atom.

Very interesting article with the character of a review. Many potentials,[69] leading to exactly solvable problems are presented in a uniform theoretical approach. The authors give also their own two generalizations.

Questions

1. Particle in a box. After doubling the box length the energy of the levels will:
   a) stay the same; b) decrease four times; c) increase twice; d) decrease twice.

2. Tunnelling of a particle through a system of two rectangular barriers of height V. $E > 0$ is the particle energy. The transmission coefficient as a function of E:
   a) does not depend on E; b) increases with E; c) has maxima; d) vanishes exponentially when E decreases.

3. Harmonic oscillator. The energy E of the lowest vibrational level of the H_2 molecule is equal to A, of the DH molecule is equal to B, of the TH molecule is equal to C. The following inequality holds:
   a) $A > B > C$; b) $C > A > B$; c) $B < C < A$; d) $A < B < C$.

4. Morse oscillator. The number of vibrational levels:
   a) is always larger than 1; b) does not depend on the well width, and depends on its depth; c) may be equal to zero for a non-zero well depth; d) may be equal to ∞.

5. Rigid rotator. The separation between neighbouring levels with quantum numbers J and $J+1$:
   a) increases linearly with J; b) increases with $J(J+1)$; c) decreases proportionally to $2J+1$; d) is constant.

6. The following spherical harmonics Y_J^M have the correct indices:
   a) $Y_{-1}^0, Y_1^2, Y_2^{-1}$; b) Y_2^1, Y_1^0, Y_1^5; c) Y_0^0, Y_0^1, Y_2^2; d) Y_3^3, Y_3^2, Y_3^{-2}.

7. In the ground-state hydrogen atom:
   a) the probability of finding the electron in a cube with its centre on the nucleus is equal to 0;
   b) the maximum probability density for finding the electron is on the nucleus;
   c) the probability density for finding the electron on the nucleus is equal to 0;
   d) the radial probability density has its maximum on the nucleus.

8. The following linear combination of the hydrogen atom orbitals ($\hat{H}$ is the Hamiltonian, $\hat{J}^2$ is the operator of the square of the angular momentum, $\hat{J}_z$ is the operator of the z-component of the angular momentum):
   a) $2s + 3 \cdot (2p_{-1})$ is an eigenfunction of $\hat{H}$ and $\hat{J}_z$;
   b) $3s - 1/3 \cdot 2p_z$ is an eigenfunction of $\hat{J}_z$;
   c) $2p_0 + 3p_{-1}$ is an eigenfunction of $\hat{H}$ and $\hat{J}^2$;
   d) $2s - 3s$ is an eigenfunction of $\hat{H}$.

9. Please choose the acceptable hydrogen atom orbitals ψ_{nlm}:
   a) $\psi_{200}, \psi_{100}, \psi_{220}$; b) $\psi_{100}, \psi_{2-10}, \psi_{211}$; c) $\psi_{320}, \psi_{52-1}, \psi_{210}$; d) $\psi_{-200}, \psi_{010}, \psi_{210}$.

---

[69] Among them six not discussed in the present textbook.

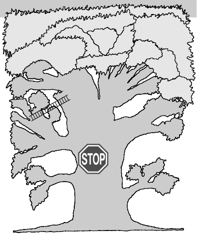
10. Harmonic helium atom. In this system the electrons:
    a) attract themselves harmonically, but interact by Coulombic force with the nucleus;
    b) interact harmonically with themselves and with the nucleus;
    c) oscillate between two positions;
    d) repel each other by Coulombic forces and are attracted harmonically by the nucleus.

Answers

1b, 2c, 3a, 4c, 5a, 6d, 7b, 8b, 9c, 10d

Chapter 5

TWO FUNDAMENTAL APPROXIMATE METHODS



WHERE ARE WE?

Where are we?

We are moving upwards in the central parts of the TREE trunk.

An example

We are interested in properties of the ammonia molecule in its ground and excited states, e.g., we would like to know the mean value of the nitrogen–hydrogen distance. Only quantum mechanics gives a method for calculation this value (p. 24): we have to calculate the mean value of an operator with the ground-state wave function. But where could this function be taken from? As a solution of the Schrödinger equation? Impossible. This equation is too difficult to solve (14 particles, cf. problems with exact solutions, Chapter 4).

The only possibility is somehow to obtain an *approximate* solution to this equation.

What is it all about

We need mathematical methods which *will allow us to obtain approximate solutions* of the Schrödinger equation. These methods are: the variational method and the perturbational approach.

Variational method (▲) **p. 196**

- Variational principle
- Variational parameters
- Ritz method

Perturbational method (△♦) **p. 203**

- Rayleigh–Schrödinger approach (△)
- Hylleraas variational principle (♦)
- Hylleraas equation (♦)
- Convergence of the perturbational series (♦)

Why is this important?

We have to know how to calculate wave functions. The exact wave function is definitely out of our reach, therefore in this chapter we will learn how to calculate the approximations.

What is needed?

- Postulates of quantum mechanics (Chapter 1, needed).

- Hilbert space (Appendix B, p. 895, necessary).
- Matrix algebra (Appendix A, p. 889, needed).
- Lagrange multipliers (Appendix N, on p. 997, needed).
- Orthogonalization (Appendix J, p. 977, occasionally used).
- Matrix diagonalization (Appendix K, p. 982, needed).
- Group theory (Appendix C, p. 903, occasionally used in this chapter).

Classical works

The variational method of linear combinations of functions was formulated by Walther Ritz in a paper published in *Zeitschrift für Reine und Angewandte Mathematik*, 135 (1909) 1. ★ The method was applied by Erwin Schrödinger in his first works "*Quantisierung als Eigenwertproblem*" in *Annalen der Physik*, 79 (1926) 361, *ibid*. 79 (1926) 489, *ibid.* 80 (1926) 437, *ibid.* 81 (1926) 109. Schrödinger also used the perturbational approach when developing the theoretical results of Lord Rayleigh for vibrating systems (hence the often used term Rayleigh–Schrödinger perturbation theory). ★ Egil Andersen Hylleraas, in *Zeitschrift der Physik*, 65 (1930) 209 showed for the first time that the variational principle may be used also for separate terms of the perturbational series.

5.1 VARIATIONAL METHOD

5.1.1 VARIATIONAL PRINCIPLE

variational function

Let us write the Hamiltonian $\hat{H}$ of the system under consideration[1] and take an *arbitrary* (variational) function Φ, which satisfies the following conditions:

- it depends on the same coordinates as the solution to the Schrödinger equation;
- it is of class Q, p. 73 (which enables it to be normalized).

We calculate the number ε that depends on Φ (i.e. ε is a functional of Φ)

$$\varepsilon[\Phi]=\frac{\langle\Phi|\hat{H}|\Phi\rangle}{\langle\Phi|\Phi\rangle}. \tag{5.1}$$

The variational principle states:

- $\varepsilon \geqslant E_0$, where E_0 is the *ground-state* energy of the system
- in the above inequality $\varepsilon = E_0$ happens, if and only if, Φ equals the exact ground-state wave function ψ_0 of the system, $\hat{H}\psi_0 = E_0\psi_0$.

---

[1] We focus here on the non-relativistic case (eq. (2.1)), where the lowest eigenvalue of $\hat{H}$ is bound from below ($> -\infty$). As we remember from Chapter 3, this is not fulfilled in the relativistic case (Dirac's electronic sea), and may lead to serious difficulties in applying the variational method.

Proof (expansion into eigenfunctions):

The unknown eigenfunctions $\{\psi_i\}$ of the Hamiltonian $\hat{H}$ represent a complete set (we may be assured of its orthonormality, see Appendix B on p. 895) in the Hilbert space of our system. This means that any function belonging to this space can be represented as a linear combination of the functions of this set

$$\Phi = \sum_{i=0}^{\infty} c_i \psi_i, \tag{5.2}$$

where c_i assure the normalization of Φ, i.e. $\sum_{i=0}^{\infty} |c_i|^2 = 1$, because

$$\langle \Phi | \Phi \rangle = \sum_{i,j} c_j^* c_i \langle \psi_j | \psi_i \rangle = \sum_{i,j} c_j^* c_i \delta_{ij} = \sum_i c_i^* c_i = 1.$$

Let us insert this into the expression for the mean value of the energy $\varepsilon = \langle \Phi | \hat{H} \Phi \rangle$

$$\begin{aligned}
\varepsilon - E_0 &= \langle \Phi | \hat{H} \Phi \rangle - E_0 = \left\langle \sum_{j=0}^{\infty} c_j \psi_j \middle| \hat{H} \sum_{i=0}^{\infty} c_i \psi_i \right\rangle - E_0 \\
&= \sum_{i,j=0}^{\infty} c_j^* c_i E_i \langle \psi_j | \psi_i \rangle - E_0 = \sum_{i,j=0}^{\infty} c_j^* c_i E_i \delta_{ij} - E_0 = \sum_{i=0}^{\infty} |c_i|^2 E_i - E_0 \cdot 1 \\
&= \sum_{i=0}^{\infty} |c_i|^2 E_i - E_0 \sum_{i=0}^{\infty} |c_i|^2 = \sum_{i=0}^{\infty} |c_i|^2 (E_i - E_0) \geqslant 0.
\end{aligned}$$

Note that the *equality* (in the last step) is satisfied only if $\Phi = \psi_0$. This therefore proves the variational principle (5.1): $\varepsilon \geqslant E_0$.

In several places in this book we will need similar proofs using Lagrange multipliers. This is why we will demonstrate how to prove the same theorem using this technique (Appendix N on p. 997).

Proof using Lagrange multipliers:

Take the functional

$$\varepsilon[\Phi] = \langle \Phi | \hat{H} \Phi \rangle. \tag{5.3}$$

We want to find a function that assures a minimum of the functional and satisfies the normalization condition

$$\langle \Phi | \Phi \rangle - 1 = 0. \tag{5.4}$$

We will change the function Φ a little (the change will be called "variation") and see, how this will change the value of the functional $\varepsilon[\Phi]$. In the functional we have, however, Φ and Φ^*. It seems that we have, therefore, to take into account in Φ^* the variation made in Φ. *However, in reality there is no need to do that:* it is sufficient to make the variation either in Φ or in Φ^* (the result does not depend on the choice[2]). *This makes the formulae simpler.* We decide to choose the variation of Φ^*, i.e. $\delta\Phi^*$.

Now we apply the machinery of the Lagrange multipliers (Appendix N on p. 997). Let us multiply eq. (5.4) by (for the time being) unknown Lagrange multiplier E and subtract afterwards from the functional ε, resulting in an auxiliary functional $G[\Phi]$

$$G[\Phi] = \varepsilon[\Phi] - E(\langle\Phi|\Phi\rangle - 1).$$

The variation of G (which is analogous to the differential of a function) represents a linear term in $\delta\Phi^*$. For an extremum the variation has to be equal to zero:

$$\delta G = \langle\delta\Phi|\hat{H}\Phi\rangle - E\langle\delta\Phi|\Phi\rangle = \langle\delta\Phi|(\hat{H} - E)\Phi\rangle = 0.$$

Since this has to be satisfied for *any* variation $\delta\Phi^*$, then it can follow only if

$$(\hat{H} - E)\Phi_{\text{opt}} = 0, \tag{5.6}$$

which means that the optimal $\Phi \equiv \Phi_{\text{opt}}$ is a solution of the Schrödinger equation[3] with E as the energy of the stationary state.

---

[2] Let us show this, because we will use it several times in this book. In all our cases the functional (which depends here on a single function $\phi(x)$, but later we will also deal with several functions in a similar procedure) might be rewritten as

$$\varepsilon[\phi] = \langle\phi|\hat{A}\phi\rangle, \tag{5.5}$$

where $\hat{A}$ is a Hermitian operator. Let us write $\phi(x) = a(x) + ib(x)$, where $a(x)$ and $b(x)$ are *real functions*. The change of ε is equal to

$$\begin{aligned}\varepsilon[\phi + \delta\phi] - \varepsilon[\phi] &= \big\langle a + \delta a + ib + i\delta b|\hat{A}(a + \delta a + ib + i\delta b)\big\rangle - \big\langle a + ib|\hat{A}(a + ib)\big\rangle \\ &= \big\langle\delta a + i\delta b|\hat{A}\phi\big\rangle + \big\langle\phi|\hat{A}(\delta a + i\delta b)\big\rangle + \text{quadratic terms} \\ &= \big\langle\delta a|\hat{A}\phi + (\hat{A}\phi)^*\big\rangle + i\big\langle\delta b|(\hat{A}\phi)^* - \hat{A}\phi\big\rangle + \text{quadratic terms.}\end{aligned}$$

The variation of a function only represents a linear part of the change, and therefore $\delta\varepsilon = \langle\delta a|\hat{A}\phi + (\hat{A}\phi)^*\rangle + i\langle\delta b|(\hat{A}\phi)^* - \hat{A}\phi\rangle$. At the extremum the variation has to equal zero *at any* variations of δa and δb. This may happen only if $\hat{A}\phi + (\hat{A}\phi)^* = 0$ and $(\hat{A}\phi)^* - \hat{A}\phi$. This means $\hat{A}\phi = 0$ or, *equivalently*, $(\hat{A}\phi)^* = 0$.

The first of the conditions would be obtained if in ε we made the variation in ϕ^* *only* (the variation in the extremum would then be $\delta\varepsilon = \langle\delta\phi|\hat{A}\phi\rangle = 0$), hence, from the arbitrariness of $\delta\phi^*$ we would get $\hat{A}\phi = 0$), the second, if we made the variation in ϕ *only* (then, $\delta\varepsilon = \langle\phi|\hat{A}\delta\phi\rangle = \langle\hat{A}\phi|\delta\phi\rangle = 0$ and $(\hat{A}\phi)^* = 0$) and the result is exactly the same. This is what we wanted to show: *we may vary either ϕ or ϕ^* and the result is the same.*

[3] In the variational calculus the equation for the optimum Φ, or the conditional minimum of a functional ε, is called the Euler equation. As one can see in this case the Euler equation is identical with the Schrödinger one.

Now let us multiply eq. (5.6) by Φ_{opt}^* and integrate. We obtain

$$\langle \Phi_{\text{opt}} | \hat{H} \Phi_{\text{opt}} \rangle - E \langle \Phi_{\text{opt}} | \Phi_{\text{opt}} \rangle = 0, \tag{5.7}$$

which means that the conditional minimum of $\varepsilon[\Phi]$ is $E = \min(E_0, E_1, E_2, \ldots) = E_0$ (the ground state). Indeed, eq. (5.7) may be written as the mean value of the Hamiltonian

$$\left\langle \frac{1}{\sqrt{\langle \Phi_{\text{opt}} | \Phi_{\text{opt}} \rangle}} \Phi_{\text{opt}} \middle| \hat{H} \frac{1}{\sqrt{\langle \Phi_{\text{opt}} | \Phi_{\text{opt}} \rangle}} \Phi_{\text{opt}} \right\rangle = E = \varepsilon \left[\frac{1}{\sqrt{\langle \Phi_{\text{opt}} | \Phi_{\text{opt}} \rangle}} \Phi_{\text{opt}} \right], \tag{5.8}$$

and the lowest possible eigenvalue E is E_0. Hence, for any other Φ we obtain a *higher* energy value, therefore $\varepsilon \geqslant E_0$.

The same was obtained when we expanded Φ into the eigenfunction series.

Variational principle for excited states

The variational principle (5.1) has been proved for an approximation to the ground-state wave function. What about excited states? If the variational function Φ is *orthogonal to exact solutions to the Schrödinger equation that correspond to all the states of lower energy than the state we are interested in*, the variational principle is still valid.[4] If the wave function k being sought represents the *lowest* state among those belonging to a given irreducible representation of the symmetry group of the Hamiltonian, then the orthogonality mentioned above is automatically guaranteed (see Appendix C on p. 903). For other excited states, the variational principle cannot be satisfied, except that function Φ does not contain lower-energy wave functions, i.e. is orthogonal to them, e.g., because the wave functions have been cut out of it earlier.

variational principle for excited states

Beware of mathematical states

We mentioned in Chapter 1 that not all solutions of the Schrödinger equation are acceptable. Only those are acceptable which satisfy the symmetry requirements with respect to the exchange of labels corresponding to identical particles (Postulate V). The other solutions are called *mathematical*. If, therefore, an incautious scientist takes a variational function Φ with a non-physical symmetry, the variational principle, following our derivation exactly (p. 197), will still be valid, but with respect to the *mathematical ground state*. The mathematical states may correspond to energy eigenvalues *lower* than the physical ground state (they are called the *underground* states, cf. p. 76). All this would end up as a catastrophe, because the mean value of the Hamiltonian would tend towards the non-physical underground mathematical state.

underground states

---

[4] The corresponding proof will only be slightly modified. Simply in the expansion eq. (5.2) of the variational function Φ, the wave functions ψ_i that correspond to lower energy states (than the state in which we are interested), will be absent. We will therefore obtain $\sum_{i=1} |c_i|^2 (E_i - E_k) \geqslant 0$, because state k is the lowest among all the states i.

5.1.2 VARIATIONAL PARAMETERS

The variational principle (5.1) may seem a little puzzling. We insert an *arbitrary* function Φ into the integral and obtain a result related to the ground state of the *system under consideration*. And yet the arbitrary function Φ may have absolutely nothing to do with the molecule we consider. The problem is that the integral contains the most important information about our system. The information resides in $\hat{H}$. Indeed, if someone wrote down the expression for $\hat{H}$, we would know right away that the system contains N electrons and M nuclei, we would also know the charges on the nuclei, i.e. the chemical elements of which the system is composed.[5] This is important information.

The variational method represents an application of the variational principle. The *trial wave function* Φ is taken in an analytical form (with the variables denoted by the vector $\boldsymbol{x}$ and automatically satisfying Postulate V). In the "key positions" in the formula for Φ we introduce the parameters $\boldsymbol{c} \equiv (c_0, c_1, c_2, \ldots, c_P)$, which we may change smoothly. The parameters play the role of tuning, their particular values listed in vector $\boldsymbol{c}$ result in a certain shape of $\Phi(\boldsymbol{x}; \boldsymbol{c})$. The integration in the formula for ε pertains to the variables $\boldsymbol{x}$, therefore the result depends uniquely on $\boldsymbol{c}$. Our function $\varepsilon(\boldsymbol{c})$ has the form

trial function

$$\varepsilon(c_0, c_1, c_2, \ldots, c_P) \equiv \varepsilon(\boldsymbol{c}) = \frac{\langle \Phi(\boldsymbol{x}; \boldsymbol{c}) | \hat{H} \Phi(\boldsymbol{x}; \boldsymbol{c}) \rangle}{\langle \Phi(\boldsymbol{x}; \boldsymbol{c}) | \Phi(\boldsymbol{x}; \boldsymbol{c}) \rangle}.$$

Now the problem is to find the minimum of the function $\varepsilon(c_0, c_1, c_2, \ldots, c_P)$.

In a general case the task is not simple, because what we are searching for is the *global* minimum. The relation

$$\frac{\partial \varepsilon(c_0, c_1, c_2, \ldots, c_P)}{\partial c_i} = 0 \quad \text{for } i = 0, 1, 2, \ldots, P,$$

therefore represents only a *necessary* condition for the global minimum.[6] This problem may be disregarded, when:

- the number of minima is small,
- in particular, when we use Φ with the linear parameters $\boldsymbol{c}$ (in this case we have a single minimum, see below).

The above equations enable us to find the optimum set of parameters $\boldsymbol{c} = \boldsymbol{c}_{\text{opt}}$. Then,

in a given class of the trial functions Φ the best possible approximation to ψ_0 is $\Phi(\boldsymbol{x}; \boldsymbol{c}_{\text{opt}})$, and the best approximation to E_0 is $\varepsilon(\boldsymbol{c}_{\text{opt}})$.

---

[5] And yet we would be unable to decide whether we have to do with matter or antimatter, or whether we have to perform calculations for the benzene molecule or for six CH radicals (cf. Chapter 2).

[6] More about global minimization may be found in Chapter 6.

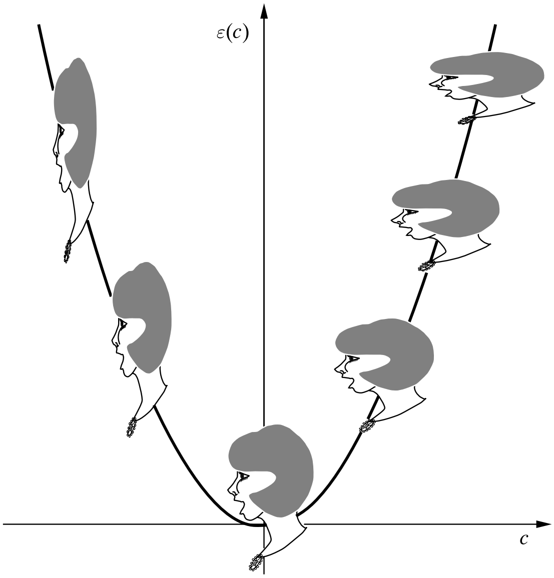


Fig. 5.1. The philosophy behind the variational method. A parameter c is changed in order to obtain the best solution possible. Any commentary would obscure the very essence of the method.

Fig. 5.1. shows the essence of the variational method.[7]

Let us assume that someone does not know that the hydrogen-like atom (the nucleus has a charge Z) problem has an exact solution.[8] Let us apply the simplest version of the variational method to see what kind of problem we will be confronted with.

An important first step will be to decide which class of trial functions to choose. We decide to take the following class[9] (for $c > 0$) $\exp(-cr)$ and after normalization of the function: $\Phi(r, \theta, \phi; c) = \sqrt{\frac{c^3}{\pi}} \exp(-cr)$. The calculation $\varepsilon[\Phi] = \langle\Phi|\hat{H}|\Phi\rangle$ is shown in Appendix H on p. 969. We obtain $\varepsilon(c) = \frac{1}{2}c^2 - Zc$. We very easily find the minimum of $\varepsilon(c)$ and the optimum c is equal to $c_{\text{opt}} = Z$, which, as we know from Chapter 4, represents the exact result. In practise (for atoms or molecules), we would never know the exact result. The optimal ε might then be obtained after many days of number crunching.[10]

---

[7] The variational method is used in everyday life. Usually we determine the target (say, cleaning the car), and then by trial, errors and corrections we approach the ideal, but never fully achieve it.

[8] For a larger system we will not know the exact solution either.

[9] A particular choice is usually made through scientific discussion. The discussion might proceed as follows.

The electron and the nucleus attract themselves, therefore they will be close in space. This assures many classes of trial functions, e.g., $\exp(-cr)$, $\exp(-cr^2)$, $\exp(-cr^3)$, etc., where $c > 0$ is a single variational parameter. In the present example we pretend not to know, which class of functions is most promising (i.e. which will give lower ε). Let us begin with class $\exp(-cr)$, and other classes will be investigated in the years to come. The decision made, we rush to do the calculations.

[10] For example, for $Z = 1$ we had to decide a starting value of c, say, $c = 2$; $\varepsilon(2) = 0$. Let us try $c = 1.5$, we obtain a lower (i.e. better) value $\varepsilon(1.5) = -0.375$ a.u., the energy goes down. Good direction, let us

5.1.3 RITZ METHOD[11]

basis functions

The Ritz method represents a special kind of variational method. The trial function Φ is represented as a *linear* combination of the *known* basis functions $\{\Psi_i\}$ with the (for the moment) *unknown* variational coefficients c_i

$$\Phi = \sum_{i=0}^{P} c_i \Psi_i. \tag{5.9}$$

Then

$$\varepsilon = \frac{\langle \sum_{i=0}^{P} c_i \Psi_i | \hat{H} \sum_{i=0}^{P} c_i \Psi_i \rangle}{\langle \sum_{i=0}^{P} c_i \Psi_i | \sum_{i=0}^{P} c_i \Psi_i \rangle} = \frac{\sum_{i=0}^{P} \sum_{j=0}^{P} c_i^* c_j H_{ij}}{\sum_{i=0}^{P} \sum_{j=0}^{P} c_i^* c_j S_{ij}} = \frac{A}{B}. \tag{5.10}$$

complete basis set

In the formula above $\{\Psi_i\}$ represents the chosen *complete basis set*.[12] The basis set functions are usually non-orthogonal, and therefore

$$\langle \Psi_i | \Psi_j \rangle = S_{ij}, \tag{5.11}$$

overlap matrix

where $\boldsymbol{S}$ stands for the *overlap matrix*, and the integrals

$$H_{ij} = \langle \Psi_i | \hat{H} \Psi_j \rangle \tag{5.12}$$

are the matrix elements of the Hamiltonian. Both matrices ($\boldsymbol{S}$ and $\boldsymbol{H}$) are calculated once and for all. The energy ε becomes a function of the linear coefficients $\{c_i\}$. The coefficients $\{c_i\}$ and the coefficients $\{c_i^*\}$ are not independent (c_i can be obtained from c_i^*). Therefore, as the linearly independent coefficients, we may treat either $\{c_i\}$ or $\{c_i^*\}$. When used for the minimization of ε, both choices would give the same. We decide to treat $\{c_i^*\}$ as variables. For each $k = 0, 1, \ldots, P$ we

---

try, therefore $c = 1.2$; $\varepsilon(1.2) = -0.48$ a.u. Indeed, a good direction. However, when we continue and take $c = 0.7$, we obtain $\varepsilon = -0.455$, i.e. a higher energy. We would continue this way and finally obtain something like $c_{\text{opt}} = 1.0000000$. We might be satisfied by 8 significant figures and decide to stop the calculations. We would never be sure, however, whether other classes of trial functions would provide still better (i.e. lower) energies. In our particular case this, of course, would never happen, because "accidentally" we have taken a class which contains the exact wave function.

[11] Walther Ritz was the Swiss physicist and a former student of Poincaré. His contributions, beside the variational approach, include perturbation theory, the theory of vibrations etc. Ritz is also known for his controversy with Einstein on the time flow problem ("time flash"), concluded by their joint article "*An agreement to disagree*" (W. Ritz, A. Einstein, *Phys. Zeit*. 10 (1909) 323).

[12] Such basis sets are available in the literature. A practical problem arises as to how many such functions should be used. In principle we should have used $P = \infty$. This, however, is unfeasible. We are restricted to a finite, *preferably small number*. And this is the moment when it turns out that some basis sets are more effective than others, that this depends on the problem considered, etc. This is how a new science emerges, which might facetiously be called basology.

have to have in the minimum

$$0 = \frac{\partial \varepsilon}{\partial c_k^*} = \frac{(\sum_{j=0}^{P} c_j H_{kj})B - A(\sum_{j=0}^{P} c_j S_{kj})}{B^2}$$

$$= \frac{(\sum_{j=0}^{P} c_j H_{kj})}{B} - \frac{A}{B}\frac{(\sum_{j=0}^{P} c_j S_{kj})}{B} = \frac{(\sum_{j=0}^{P} c_j (H_{kj} - \varepsilon S_{kj}))}{B},$$

what leads to the *secular equations*

secular equations

$$\left(\sum_{j=0}^{P} c_j (H_{kj} - \varepsilon S_{kj})\right) = 0 \quad \text{for } k = 0, 1, \ldots, P. \tag{5.13}$$

The unknowns in the above equation are the coefficients c_j and the energy ε. With respect to the coefficients c_j, eqs. (5.13) represent a homogeneous set of linear equations. Such a set has a non-trivial solution if the *secular determinant* is equal to zero (see Appendix A)

secular determinant

$$\det(H_{kj} - \varepsilon S_{kj}) = 0. \tag{5.14}$$

This happens however only for some particular values of ε satisfying the above equation. Since the rank of the determinant is equal $P + 1$, we therefore obtain $P + 1$ solutions ε_i, $i = 0, 1, 2, \ldots, P$. Due to the Hermitian character of operator $\hat{H}$, the matrix $\boldsymbol{H}$ will be also Hermitian. In Appendices J on p. 977 and L on p. 984, we show that the problem reduces to the diagonalization of some transformed $\boldsymbol{H}$ matrix (also Hermitian). This guarantees that all ε_i will be real. Let us denote the lowest ε_i as ε_0, to represent an approximation to the ground state energy.[13] The other ε_i, $i = 1, 2, \ldots, P$, will approximate the excited states of the system. We obtain an approximation to the i-th wave function by inserting the calculated ε_i into eq. (5.13), and then, after including the normalization condition, we find the corresponding set of c_i. The problem is solved.

5.2 PERTURBATIONAL METHOD

5.2.1 RAYLEIGH–SCHRÖDINGER APPROACH

The idea of the perturbational approach is very simple. We know everything about a non-perturbed problem. Then we slightly perturb the system and everything changes. If the perturbation is small, it seems there is a good chance that there

---

[13] Assuming we used the basis functions that satisfy Postulate V (on symmetry).

will be no drama: the wave function and the corresponding energy will change only a little (if the changes were large, the perturbational approach would simply be inapplicable). The whole perturbational procedure aims at finding these tiny changes with satisfactory precision.

Perturbational theory is notorious for quite clumsy equations. Unfortunately, there is no way round if we want to explain how to calculate things. *However, in practise only a few of these equations will be used – they will be highlighted in frames.*

unperturbed operator

Let us begin our story. We would like to solve the Schrödinger equation

$$\hat{H}\psi_k = E\psi_k, \tag{5.15}$$

and as a rule we will be interested in a single particular state k, most often the ground state ($k = 0$).

We apply a perturbational approach, when[14]

$$\hat{H} = \hat{H}^{(0)} + \hat{H}^{(1)},$$

perturbation

where the so called *unperturbed operator* $\hat{H}^{(0)}$ is "large", while the *perturbation operator* $\hat{H}^{(1)}$ is "small".[15] We assume that there is no problem whatsoever with solving the unperturbed Schrödinger equation

$$\hat{H}^{(0)}\psi_k^{(0)} = E_k^{(0)}\psi_k^{(0)}. \tag{5.16}$$

We assume that $\psi_k^{(0)}$ form an orthonormal set, which is natural.[16] We are interested in the fate of the wave function $\psi_k^{(0)}$ after the perturbation is switched on (when it changes to ψ_k). We choose the *intermediate normalization*, i.e.

intermediate normalization

$$\langle\psi_k^{(0)}|\psi_k\rangle = 1. \tag{5.17}$$

The intermediate normalization means that ψ_k, as a vector of the Hilbert space (see Appendix B on p. 895), has the normalized $\psi_k^{(0)}$ as the component along the unit basis vector $\psi_k^{(0)}$. In other words,[17] $\psi_k = \psi_k^{(0)} +$ terms orthogonal to $\psi_k^{(0)}$.

perturbational parameter

We are all set to proceed. First, we introduce the *perturbational parameter* $0 \leqslant \lambda \leqslant 1$ in Hamiltonian $\hat{H}$, making it, therefore, λ-dependent[18]

$$\hat{H}(\lambda) = \hat{H}^{(0)} + \lambda\hat{H}^{(1)}.$$

---

[14] We assume all operators are Hermitian.

[15] In the sense that the energy spectrum of $\hat{H}^{(0)}$ is only slightly changed after the perturbation $\hat{H}^{(1)}$ is switched on.

[16] We can always do that, because $\hat{H}^{(0)}$ is Hermitian (see Appendix B).

[17] The intermediate normalization is convenient, but not necessary. Although convenient for the derivation of perturbational equations, it leads to some troubles when the mean values of operators are to be calculated.

[18] Its role is technical. It will enable us to transform the Schrödinger equation into a sequence of perturbational equations, which must be solved one by one. Then the parameter λ *disappears from the theory*, because we put $\lambda = 1$.

When $\lambda = 0$, $\hat{H}(\lambda) = \hat{H}^{(0)}$, while $\lambda = 1$ gives $\hat{H}(\lambda) = \hat{H}^{(0)} + \hat{H}^{(1)}$. In other words, we tune the perturbation at will from 0 to $\hat{H}^{(1)}$. It is worth noting that $\hat{H}(\lambda)$ for $\lambda \neq 0, 1$ may not correspond to any physical system. It does not need to. We are interested here in a mathematical trick, we will come back to reality by putting $\lambda = 1$ at the end.

We are interested in the Schrödinger equation being satisfied for all values $\lambda \in [0, 1]$

$$\hat{H}(\lambda)\psi_k(\lambda) = E_k(\lambda)\psi_k(\lambda).$$

Now this is a key step in the derivation. We expect that both the energy and the wave function can be developed in a power series[19] of λ

$$E_k(\lambda) = E_k^{(0)} + \lambda E_k^{(1)} + \lambda^2 E_k^{(2)} + \cdots, \tag{5.18}$$

$$\psi_k(\lambda) = \psi_k^{(0)} + \lambda \psi_k^{(1)} + \lambda^2 \psi_k^{(2)} + \cdots, \tag{5.19}$$

where $E_k^{(i)}$ stand for some (unknown for the moment) coefficients, and $\psi_k^{(i)}$ represents the functions to be found. We expect the two series to converge (Fig. 5.2).

> In practise we calculate only $E_k^{(1)}, E_k^{(2)}$ and quite rarely $E_k^{(3)}$, and for the wave function, we usually limit the correction to $\psi_k^{(1)}$.

How are these corrections calculated?

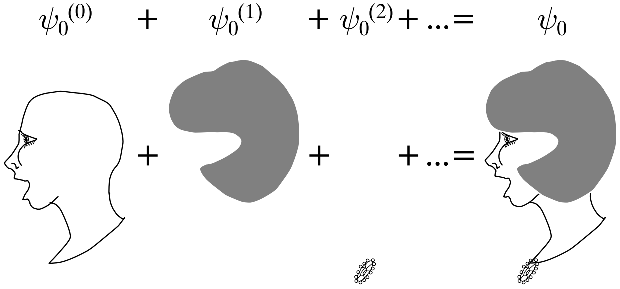


Fig. 5.2. Philosophy of the perturbational approach (optimistic version). The ideal ground-state wave function ψ_0 is constructed as a sum of a good zero-order approximation ($\psi_0^{(0)}$) and consecutive small corrections ($\psi_0^{(n)}$). The first-order correction ($\psi_0^{(1)}$) is still quite substantial, but fortunately the next corrections amount to only small cosmetic changes.

---

[19] It is in fact a Taylor series with respect to λ. The physical meaning of these expansions is the following: $E_k^{(0)}$ and $\psi_k^{(0)}$ are good approximations of $E_k(\lambda)$ and $\psi_k(\lambda)$. The rest will be calculated as a sum of small correction terms.

We insert the two perturbational series for $E_k(\lambda)$ and $\psi_k(\lambda)$ into the Schrödinger equation

$$\begin{aligned}&\big(\hat{H}^{(0)}+\lambda\hat{H}^{(1)}\big)\big(\psi_k^{(0)}+\lambda\psi_k^{(1)}+\lambda^2\psi_k^{(2)}+\cdots\big)\\&\quad=\big(E_k^{(0)}+\lambda E_k^{(1)}+\lambda^2E_k^{(2)}+\cdots\big)\big(\psi_k^{(0)}+\lambda\psi_k^{(1)}+\lambda^2\psi_k^{(2)}+\cdots\big)\end{aligned}$$

and, since the equation has to be satisfied for *any* λ belonging to $0\leqslant\lambda\leqslant 1$, this may happen only if

> the coefficients at the same powers of λ on the left- and right-hand sides are equal.

perturbational equations

This gives a sequence of an infinite number of perturbational equations to be satisfied by the unknown $E_k^{(n)}$ and $\psi_k^{(n)}$. *These equations may be solved consecutively allowing us to calculate $E_k^{(n)}$ and $\psi_k^{(n)}$ with larger and larger n.* We have, for example:

$$\begin{aligned}&\text{for }\lambda^0\text{:}\quad \hat{H}^{(0)}\psi_k^{(0)}=E_k^{(0)}\psi_k^{(0)}\\&\text{for }\lambda^1\text{:}\quad \hat{H}^{(0)}\psi_k^{(1)}+\hat{H}^{(1)}\psi_k^{(0)}=E_k^{(0)}\psi_k^{(1)}+E_k^{(1)}\psi_k^{(0)}\\&\text{for }\lambda^2\text{:}\quad \hat{H}^{(0)}\psi_k^{(2)}+\hat{H}^{(1)}\psi_k^{(1)}=E_k^{(0)}\psi_k^{(2)}+E_k^{(1)}\psi_k^{(1)}+E_k^{(2)}\psi_k^{(0)}\\&\dots\\&\text{etc.}^{20}\end{aligned}\tag{5.20}$$

Doing the same with the intermediate normalization (eq. (5.17)), we obtain

$$\big\langle\psi_k^{(0)}\big|\psi_k^{(n)}\big\rangle=\delta_{0n}.\tag{5.21}$$

The first of eqs. (5.20) is evident (the unperturbed Schrödinger equation does not contain any unknown). The second equation involves two unknowns, $\psi_k^{(1)}$ and $E_k^{(1)}$. To eliminate $\psi_k^{(1)}$ we will use the Hermitian character of the operators. Indeed, by making the scalar product of the equation with $\psi_k^{(0)}$ we obtain:

$$\begin{aligned}&\big\langle\psi_k^{(0)}\big|\big(\hat{H}^{(0)}-E_k^{(0)}\big)\psi_k^{(1)}+\big(\hat{H}^{(1)}-E_k^{(1)}\big)\psi_k^{(0)}\big\rangle\\&\quad=\big\langle\psi_k^{(0)}\big|\big(\hat{H}^{(0)}-E_k^{(0)}\big)\psi_k^{(1)}\big\rangle+\big\langle\psi_k^{(0)}\big|\big(\hat{H}^{(1)}-E_k^{(1)}\big)\psi_k^{(0)}\big\rangle\\&\quad=0+\big\langle\psi_k^{(0)}\big|\big(\hat{H}^{(1)}-E_k^{(1)}\big)\psi_k^{(0)}\big\rangle=0,\end{aligned}$$

i.e.

---

[20] We see the construction principle of these equations: we write down all the terms which give a given value of the sum of the upper indices.

the formula for the *first-order correction to the energy*

$$E_k^{(1)} = H_{kk}^{(1)}, \tag{5.22}$$

where we defined

$$H_{kn}^{(1)} = \langle \psi_k^{(0)} | \hat{H}^{(1)} | \psi_n^{(0)} \rangle. \tag{5.23}$$

first-order correction

Conclusion: the first order correction to the energy, $E_k^{(1)}$, represents the mean value of the perturbation with the unperturbed wave function of the state in which we are interested (usually the ground state).[21]

Now, from the perturbation equation (5.20) corresponding to $n = 2$ we have[22]

$$\begin{aligned} &\langle \psi_k^{(0)} | (\hat{H}^{(0)} - E_k^{(0)}) \psi_k^{(2)} \rangle + \langle \psi_k^{(0)} | (\hat{H}^{(1)} - E_k^{(1)}) \psi_k^{(1)} \rangle - E_k^{(2)} \\ &\quad = \langle \psi_k^{(0)} | \hat{H}^{(1)} \psi_k^{(1)} \rangle - E_k^{(2)} = 0, \end{aligned}$$

and hence

$$E_k^{(2)} = \langle \psi_k^{(0)} | \hat{H}^{(1)} \psi_k^{(1)} \rangle. \tag{5.24}$$

For the time being we cannot compute $E_k^{(2)}$, because we do not know $\psi_k^{(1)}$, but soon we will. In the perturbational equation (5.20) for λ^1 let us expand $\psi_k^{(1)}$ into the complete set of the basis functions $\{\psi_n^{(0)}\}$ with as yet unknown coefficients c_n:

$$\psi_k^{(1)} = \sum_{n\,(\neq k)} c_n \psi_n^{(0)}.$$

Note that because of the intermediate normalization (5.17) and (5.21), we did not take the term with $n = k$. We get

$$\left(\hat{H}^{(0)} - E_k^{(0)}\right) \sum_{n\,(\neq k)} c_n \psi_n^{(0)} + \hat{H}^{(1)} \psi_k^{(0)} = E_k^{(1)} \psi_k^{(0)},$$

---

[21] This is quite natural and we use such a perturbative estimation all the time. What it really says is: we do not know what the perturbation exactly does, but let us estimate the result by assuming *that all things are going on as they were before the perturbation was applied*. In the first-order approach, insurance estimates your loss by *averaging* over similar losses of others. A student score in quantum chemistry is often close to its *a posteriori* estimation from his/her other scores, etc.

[22] Also through a scalar product with $\psi_k^{(0)}$.

and then transform

$$\sum_{n\,(\neq k)} c_n\big(E_n^{(0)} - E_k^{(0)}\big)\psi_n^{(0)} + \hat{H}^{(1)}\psi_k^{(0)} = E_k^{(1)}\psi_k^{(0)}.$$

We find c_m by making the scalar product with $\psi_m^{(0)}$. Due to the orthonormality of functions $\{\psi_n^{(0)}\}$ we obtain

$$c_m = \frac{H_{mk}^{(1)}}{E_k^{(0)} - E_m^{(0)}},$$

first-order correction to wave function

which gives the following formula for the *first-order correction to the wave function*

$$\boxed{\psi_k^{(1)} = \sum_{n\,(\neq k)} \frac{H_{nk}^{(1)}}{E_k^{(0)} - E_n^{(0)}}\,\psi_n^{(0)}, \qquad (5.25)}$$

second-order energy

and then the formula for the *second-order correction to the energy*

$$\boxed{E_k^{(2)} = \sum_{n\,(\neq k)} \frac{|H_{kn}^{(1)}|^2}{E_k^{(0)} - E_n^{(0)}}. \qquad (5.26)}$$

From (5.25) we see that the contribution of function $\psi_n^{(0)}$ to the wave function deformation is large if the coupling between states k and n (i.e. $H_{nk}^{(1)}$) is large, and the closer in the energy scale these two states are.

The formulae for higher-order corrections become more and more complex. We will limit ourselves to the low-order corrections in the hope that the perturbational method converges fast (we will see in a moment how surprising the perturbational series behaviour can be) and further corrections are much less important.[23]

5.2.2 HYLLERAAS VARIATIONAL PRINCIPLE[24]

The derived formulae are rarely employed in practise, because we only very rarely have at our disposal all the necessary solutions of eq. (5.16). The eigenfunctions of the $\hat{H}^{(0)}$ operator appeared as a consequence of using them as the complete set of functions (e.g., in expanding $\psi_k^{(1)}$). There are, however, some numerical methods

---

[23] Some scientists have been bitterly disappointed by this assumption.

[24] See his biographic note in Chapter 10.

that enable us to compute $\psi_k^{(1)}$ using the complete set of functions $\{\phi_i\}$, which *are not* the eigenfunctions of $\hat{H}^{(0)}$.

Hylleraas noted[25] that the functional

$$\mathcal{E}[\tilde{\chi}] = \langle\tilde{\chi}|(\hat{H}^{(0)} - E_0^{(0)})\tilde{\chi}\rangle \tag{5.27}$$

$$+\langle\tilde{\chi}|(\hat{H}^{(1)} - E_0^{(1)})\psi_0^{(0)}\rangle + \langle\psi_0^{(0)}|(\hat{H}^{(1)} - E_0^{(1)})\tilde{\chi}\rangle \tag{5.28}$$

exhibits its minimum at $\tilde{\chi} = \psi_0^{(1)}$ and for this function the value of the functional is equal to $E_0^{(2)}$. Indeed, inserting $\tilde{\chi} = \psi_0^{(1)} + \delta\chi$ into eq. (5.28) and using the Hermitian character of the operators we have

$$\begin{aligned}
[\psi_0^{(1)} + \delta\chi] - [\psi_0^{(1)}] &= \langle\psi_0^{(1)} + \delta\chi|(\hat{H}^{(0)} - E_0^{(0)})(\psi_0^{(1)} + \delta\chi)\rangle \\
&\quad + \langle\psi_0^{(1)} + \delta\chi|(\hat{H}^{(1)} - E_0^{(1)})\psi_0^{(0)}\rangle \\
&\quad + \langle\psi_0^{(0)}|(\hat{H}^{(1)} - E_0^{(1)})(\psi_0^{(1)} + \delta\chi)\rangle \\
&= \langle\delta\chi|(\hat{H}^{(0)} - E_0^{(0)})\psi_0^{(1)} + (\hat{H}^{(1)} - E_0^{(1)})\psi_0^{(0)}\rangle \\
&\quad + \langle(\hat{H}^{(0)} - E_0^{(0)})\psi_0^{(1)} + (\hat{H}^{(1)} - E_0^{(1)})\psi_0^{(0)}|\delta\chi\rangle \\
&\quad + \langle\delta\chi|(\hat{H}^{(0)} - E_0^{(0)})\delta\chi\rangle = \langle\delta\chi|(\hat{H}^{(0)} - E_0^{(0)})\delta\chi\rangle \geqslant 0.
\end{aligned}$$

This proves the Hylleraas variational principle. The last equality follows from the first-order perturbational equation, and the last inequality from the fact that $E_0^{(0)}$ is assumed to be the lowest eigenvalue of $\hat{H}^{(0)}$ (see the variational principle).

What is the minimal value of the functional under consideration? Let us insert $\tilde{\chi} = \psi_0^{(1)}$. We obtain

$$\begin{aligned}
\mathcal{E}[\psi_0^{(1)}] &= \langle\psi_0^{(1)}|(\hat{H}^{(0)} - E_0^{(0)})\psi_0^{(1)}\rangle \\
&\quad + \langle\psi_0^{(1)}|(\hat{H}^{(1)} - E_0^{(1)})\psi_0^{(0)}\rangle + \langle\psi_0^{(0)}|(\hat{H}^{(1)} - E_0^{(1)})\psi_0^{(1)}\rangle \\
&= \langle\psi_0^{(1)}|(\hat{H}^{(0)} - E_0^{(0)})\psi_0^{(1)} + (\hat{H}^{(1)} - E_0^{(1)})\psi_0^{(0)}\rangle + \langle\psi_0^{(0)}|\hat{H}^{(1)}\psi_0^{(1)}\rangle \\
&= \langle\psi_0^{(1)}|0\rangle + \langle\psi_0^{(0)}|\hat{H}^{(1)}\psi_0^{(1)}\rangle = \langle\psi_0^{(0)}|\hat{H}^{(1)}\psi_0^{(1)}\rangle = E_0^{(2)}.
\end{aligned}$$

5.2.3 HYLLERAAS EQUATION

The first-order perturbation equation (p. 206, eq. (5.20)) after inserting

$$\psi_0^{(1)} = \sum_{j=1}^{N} d_j\phi_j \tag{5.29}$$

---

[25] E.A. Hylleraas, *Zeit. Phys.* 65 (1930) 209.

takes the form

$$\sum_{j=1}^{N} d_j(\hat{H}^{(0)} - E_0^{(0)})\phi_j + \big(\hat{H}^{(1)} - E_0^{(1)}\big)\psi_0^{(0)} = 0.$$

Making the scalar products of the left- and right-hand side of the equation with functions ϕ_i, $i = 1, 2, \dots$, we obtain

$$\sum_{j=1}^{N} d_j\big(\hat{H}_{ij}^{(0)} - E_0^{(0)} S_{ij}\big) = -\big(\hat{H}_{i0}^{(1)} - E_0^{(1)} S_{i0}\big) \quad \text{for } i = 1, 2, \dots, N,$$

where $\hat{H}_{ij}^{(0)} \equiv \langle\phi_i|\hat{H}^{(0)}\phi_j\rangle$, and the overlap integrals $S_{ij} \equiv \langle\phi_i|\phi_j\rangle$. Using the matrix notation we may write the *Hylleraas equation*

$$\big(\boldsymbol{H}^{(0)} - E_k^{(0)}\boldsymbol{S}\big)\boldsymbol{d} = -\boldsymbol{v}, \tag{5.30}$$

where the components of the vector $\boldsymbol{v}$ are $v_i = \hat{H}_{i0}^{(1)} - E_0^{(1)} S_{i0}$. All the quantities can be calculated and the set of N linear equations with unknown coefficients d_i remains to be solved.[26]

5.2.4 CONVERGENCE OF THE PERTURBATIONAL SERIES

The perturbational approach is applicable when the perturbation only slightly changes the energy levels, therefore not changing their order. This means that the unperturbed energy level separations have to be much larger than a measure of perturbation such as $\hat{H}_{kk}^{(1)} = \langle\psi_k^{(0)}|\hat{H}^{(1)}\psi_k^{(0)}\rangle$. However, even in this case we may expect complications.

The subsequent perturbational corrections need not be monotonically decreasing. However, if the perturbational series eq. (5.19) converges, for any $\varepsilon > 0$ we may choose such N_0 that for $N > N_0$ we have $\langle\psi_k^{(N)}|\psi_k^{(N)}\rangle < \varepsilon$, i.e. the vectors $\psi_k^{(N)}$ have smaller and smaller length in the Hilbert space.

asymptotic convergence

Unfortunately, perturbational series are often divergent in a sense known as *asymptotic convergence*. A divergent series $\sum_{n=0}^{\infty}\frac{A_n}{z^n}$ is called an *asymptotic series* of a function $f(z)$, if the function $R_n(z) = z^n[f(z) - S_n(z)]$, where $S_n(z) = \sum_{k=0}^{n}\frac{A_k}{z^k}$, satisfies the following condition: $\lim_{z\to\infty} R_n(z) = 0$ for any fixed n. In other words, the error of the summation, i.e. $[f(z) - S_n(z)]$ tends to 0 as $z^{-(n+1)}$ or faster.

Despite the fact that the series used in physics and chemistry are often asymptotic, i.e. divergent, we are able to obtain results of high accuracy with them provided we limit ourselves to appropriate number of terms. The asymptotic character

---

[26] We obtain the *same equation*, if in the Hylleraas functional eq. (5.28), the variational function χ is expanded as a linear combination (5.29), and then vary d_i in a similar way to that of the Ritz variational method described on p. 202.

of such series manifests itself in practise in such a way that the partial sums $S_n(z)$ stabilize and we obtain numerically a situation typical for convergence. For example, we sum up the consecutive perturbational corrections and obtain the partial sums changing on the eighth, then ninth, then tenth significant figures. This is a very good time to stop the calculations, publish the results, finish the scientific career and move on to other business. The addition of further perturbational corrections ends up in catastrophe, cf. Appendix X on p. 1038. It begins by an innocent, very small, increase in the partial sums, they just begin to change the ninth, then the eighth, then the seventh significant figure. Then, it only gets worse and worse and ends up by an explosion of the partial sums to ∞ and a very bad state of mind for the researcher (I did not dare to depict it in Fig. 5.2).

In perturbation theory we assume that $E_k(\lambda)$ and $\psi_k(\lambda)$ are analytical functions of λ (p. 205). In this *mathematical* aspect of the physical problem we may treat λ as a complex number. *Then the radius of convergence ρ of the perturbational series on the complex plane is equal to the smallest $|\lambda|$, for which one has a pole of $E_k(\lambda)$ or $\psi_k(\lambda)$.* The convergence radius ρ_k for the energy perturbational series may be computed as (if the limit exists[27])

$$\rho_k = \lim_{N\to\infty} \frac{|E_k^{(N)}|}{|E_k^{(N+1)}|}.$$

For physical reasons $\lambda = 1$ is most important. It is, therefore, desirable to have $\rho_k \geqslant 1$. Note (Fig. 5.3), that if $\rho_k \geqslant 1$, then the series with $\lambda = 1$ is convergent together with the series with $\lambda = -1$.

Let us take as the unperturbed system the harmonic oscillator (the potential energy equal to $\frac{1}{2}x^2$) in its ground state, and the operator $\hat{H}^{(1)} = -0.000001 \cdot x^4$ as its perturbation. In such a case the perturbation seems to be small[28] in comparison with the separation of the eigenvalues of $\hat{H}^{(0)}$. And yet the perturbational series carries the seed of catastrophe. It is quite easy to see why a catastrophe has to happen. After the perturbation is added, the potential becomes *qualitatively* different from $\frac{1}{2}x^2$. For large x, instead of going to ∞, it will tend to $-\infty$. The perturbation is not small at all, it is a monster. This will cause the perturbational series to diverge. How will it happen in practise? Well, in higher orders we have to calculate the integrals $\langle\psi_n^{(0)}|\hat{H}^{(1)}\psi_m^{(0)}\rangle$, where n, m stand for the vibrational quantum numbers. As we recall from Chapter 4 high-energy wave functions have large values for large x, where the perturbation changes as x^4 and gets larger and larger as x increases. This is why the integrals will be large. Therefore, the better we do our job (higher orders, higher-energy states) the faster we approach catastrophe.

Let us consider the opposite perturbation $\hat{H}^{(1)} = +0.000001 \cdot x^4$. Despite the fact that everything looks good (the perturbation does not qualitatively change the potential), the series will diverge sooner or later. It is bound to happen, because the

---

[27] If the limit does not exist, then nothing can be said about ρ_k.

[28] As a measure of the perturbation we may use $\langle\psi_0^{(0)}|\hat{H}^{(1)}\psi_0^{(0)}\rangle$, which means an integral of x^4 multiplied by a Gaussian function (cf. Chapter 4). Such an integral is easy to calculate and, in view of the fact that it will be multiplied by the (small) factor 0.000001, the perturbation will turn out to be small.

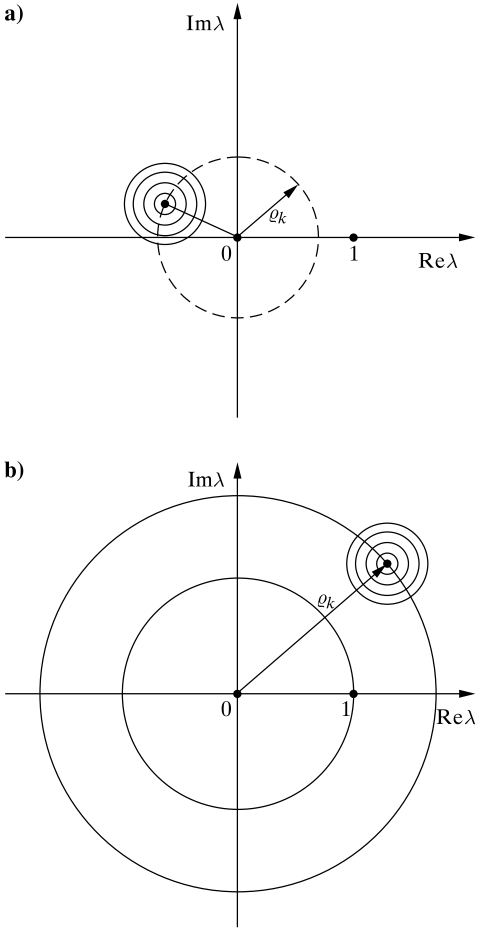


Fig. 5.3. The complex plane of the λ parameter. The physically interesting points are at $\lambda = 0, 1$. In perturbation theory we finally put $\lambda = 1$. Because of this the convergence radius ρ_k of the perturbational series has to be $\rho_k \geqslant 1$. However, *if any complex λ with $|\lambda| < 1$ corresponds to a pole of the energy, the perturbational series will diverge in the physical situation* ($\lambda = 1$). The figure shows the position of a pole by concentric circles. (a) The pole is too close ($\rho_k < 1$) and the perturbational series diverges; (b) the perturbational series converges, because $\rho_k > 1$.

convergence radius does not depend on the sign of the perturbation. A researcher might be astonished when the corrections begin to explode.

Quantum chemistry experiences with perturbational theories look quite consistent:

- low orders may give excellent results,
- higher orders often make the results worse.[29]

Summary

There are basically two numerical approaches to obtain approximate solutions to the Schrödinger equation, variational and perturbational. In calculations we usually apply variational methods, while perturbational is often applied to estimate some small physical effects.

---

[29] Even orders as high as 2000 have been investigated in the hope that the series will improve the results...

The result is that most concepts (practically all we know) characterizing the reaction of a molecule to an external field come from the perturbational approach. This leads to such quantities (see Chapter 12) as dipole moment, polarizability, hyperpolarizability, etc. The computational role of perturbational theories may, in this context, be seen as being of the second order.

- **Variational method**
  - The method is based on the variational principle, which says that, if for a system with Hamiltonian $\hat{H}$ we calculate the number $\varepsilon = \frac{\langle\Phi|\hat{H}\Phi\rangle}{\langle\Phi|\Phi\rangle}$, where Φ *stands for an arbitrary function*, then the number $\varepsilon \geqslant E_0$, with E_0 being the ground-state energy of the system. If it happens that $\varepsilon[\Phi] = E_0$, then there is only one possibility: Φ represents the exact ground-state wave function ψ_0.
  - The variational principle means that to find an approximate ground-state wave function we can use the *variational method*: minimize $\varepsilon[\Phi]$ by changing (varying) Φ. The minimum value of $\varepsilon[\Phi]$ is equal to $\varepsilon[\Phi_{\mathrm{opt}}]$ which approximates the ground-state energy E_0 and corresponds to Φ_{opt}, i.e. an approximation to the ground-state wave function ψ_0.
  - In practise the variational method consists from the following steps:
    - ∗ make a decision as to the *trial function class*, among which the $\Phi_{\mathrm{opt}}(\boldsymbol{x})$ will be sought[30]
    - ∗ introduce into the function the *variational parameters* $\boldsymbol{c} \equiv (c_0, c_1, \ldots, c_P)$: $\Phi(\boldsymbol{x}; \boldsymbol{c})$. In this way ε becomes a function of these parameters: $\varepsilon(\boldsymbol{c})$
    - ∗ minimize $\varepsilon(\boldsymbol{c})$ with respect to $\boldsymbol{c} \equiv (c_0, c_1, \ldots, c_P)$ and find the optimal set of parameters $\boldsymbol{c} = \boldsymbol{c}_{\mathrm{opt}}$
    - ∗ the value $\varepsilon(\boldsymbol{c}_{\mathrm{opt}})$ represents an approximation to E_0
    - ∗ the function $\Phi(\boldsymbol{x},; \boldsymbol{c}_{\mathrm{opt}})$ is an approximation to the ground-state wave function $\psi_0(\boldsymbol{x})$
  - The Ritz procedure is a special case of the variational method, in which the parameters $\boldsymbol{c}$ enter Φ *linearly*: $\Phi(\boldsymbol{x}; \boldsymbol{c}) = \sum_{i=0}^{P} c_i\Psi_i$, where $\{\Psi_i\}$ are some *known* basis functions that form (or more exactly, in principle form) the complete set of functions in the Hilbert space. This formalism leads to a set of homogeneous linear equations to solve ("secular equations"), from which we find approximations to the ground- and excited states energies and wave functions.
- **Perturbational method**

We assume that the solution to the Schrödinger equation for the *unperturbed system is known* ($E_k^{(0)}$ for the energy and $\psi_k^{(0)}$ for the wave function, usually $k = 0$, i.e. the ground state), but when a *small perturbation* $\hat{H}^{(1)}$ is added to the Hamiltonian, then the solution changes (to E_k and ψ_k, respectively) and is to be sought using the perturbational approach. Then the key assumption is: $E_k(\lambda) = E_k^{(0)} + \lambda E_k^{(1)} + \lambda^2 E_k^{(2)} + \cdots$ and $\psi_k(\lambda) = \psi_k^{(0)} + \lambda\psi_k^{(1)} + \lambda^2\psi_k^{(2)} + \cdots$, where λ is a parameter that tunes the perturbation. The goal of the perturbational approach is to compute corrections to the energy: $E_k^{(1)}, E_k^{(2)}, \ldots$ and to the wave function: $\psi_k^{(1)}, \psi_k^{(2)}, \ldots$. We assume that because the perturbation is small, only a few such corrections are to be computed, in particular,

$$E_k^{(1)} = \langle\psi_k^{(0)}|\hat{H}^{(1)}\psi_k^{(0)}\rangle, \quad E_k^{(2)} = \sum_{n\,(\neq k)} \frac{|H_{kn}^{(1)}|^2}{E_k^{(0)} - E_n^{(0)}}, \quad \text{where } H_{kn}^{(1)} = \langle\psi_k^{(0)}|\hat{H}^{(1)}\psi_n^{(0)}\rangle.$$

---

[30] $\boldsymbol{x}$ symbolizes the set of coordinates (space and spin, cf. Chapter 1).

Main concepts, new terms

variational principle (p. 196)
variational method (p. 196)
variational function (p. 196)
variational principle for excited states (p. 199)
underground states (p. 199)
variational parameters (p. 200)
trial function (p. 200)
Ritz method (p. 202)
complete basis set (p. 202)
secular equation (p. 203)
secular determinant (p. 203)
perturbational method (p. 203)
unperturbed system (p. 204)
perturbed system (p. 204)
perturbation (p. 204)
corrections to energy (p. 205)
corrections to wave function (p. 205)
Hylleraas functional (p. 209)
Hylleraas variational principle (p. 209)
Hylleraas equation (p. 210)
asymptotic convergence (p. 210)

From the research front

In practise, the Ritz variational method is used most often. One of the technical problems to be solved is the size of the basis set. Enormous progress in computation and software development now facilitate investigations which 20 years ago were absolutely beyond the imagination. The world record in quantum chemistry means a few billion expansion functions. To accomplish this quantum chemists have had to invent some powerful methods of applied mathematics.

Ad futurum…

The computational technique impetus we witness nowadays will continue in the future (maybe in a modified form). It will be no problem to find some reliable approximations to the ground-state energy and wave function for a molecule composed of thousands of atoms. We will be effective. We may, however, ask whether such effectiveness is at the heart of science. Would it not be interesting to know what these ten billion terms in our wave function are telling us about and what we could learn from this?

Additional literature

E. Steiner, "The Chemistry Maths Book", Oxford University Press, Oxford, 1996.

A very good textbook. We may find some useful information there about the secular equation.

W.H. Press, B.P. Flannery, S.A. Teukolsky, W.T. Vetterling, "Numerical Recipes. The Art of Scientific Computing", Cambridge University Press, 1986. p. 19–77, 274–326, 335–381.

Probably the best textbook in computational mathematics, some chapters are very closely related to the topics of this chapter (diagonalization, linear equations).

H. Margenau and G.M. Murphy, "The Mathematics of Physics and Chemistry", D. van Nostrand Co., 1956.

An excellent old book dealing with most mathematical problems which we may encounter in chemistry and physics, including the variational and perturbational methods.

J.O. Hirschfelder, W. Byers Brown, S.T. Epstein, "Recent Developments in Perturbation Theory", *Adv. Quantum Chem*. 1 (1964) 255.

A long article on perturbation theory. For many years obligatory for those working in the domain.

Questions

1. Variational method (Φ stands for the trial function, $\hat{H}$ the Hamiltonian, E_0 the exact ground-state energy, and ψ_0 the exact ground-state wave function, $\varepsilon = \frac{\langle\Phi|\hat{H}\Phi\rangle}{\langle\Phi|\Phi\rangle}$). If $\varepsilon = E_0$, this means that:
   a) $\psi_0 = \Phi$; b) $|\Phi|^2 = 1$; c) $\psi_0 \leqslant \Phi$; d) $\psi_0 = E_0\Phi$.

2. In the Ritz method (Φ stands for the trial function, $\hat{H}$ the Hamiltonian, E_0 the exact ground-state energy, ψ_0 the exact ground-state wave function, $\varepsilon = \frac{\langle\Phi|\hat{H}\Phi\rangle}{\langle\Phi|\Phi\rangle}$) the trial function Φ is always a linear combination of:
   a) orthonormal functions; b) unknown functions to be found in the procedure; c) eigenfunctions of $\hat{H}$; d) known functions.

3. A trial function used in the variational method for the hydrogen atom had the form: $\psi = \exp(-c_1 r) + c_2 \exp(-r/2)$. From a variational procedure we obtained:
   a) $c_1 = c_2 = 0$; b) $c_1 = 1, c_2 = 0$; c) $c_1 = 0, c_2 = 1$; d) $c_1 = 1, c_2 = 1$.

4. In the variational method applied to a molecule:
   a) we search an approximate wave function in the form of a secular determinant;
   b) we minimize the mean value of the Hamiltonian computed with a trial function;
   c) we minimize the trial function with respect to its parameters;
   d) we minimize the secular determinant with respect to the variational parameters.

5. In a variational method, four classes of trial functions have been applied and the total energy computed. The exact value of the energy is equal to −50.2 eV. Choose the best approximation to this value obtained in correct calculations:
   a) −48.2 eV; b) −50.5 eV; c) −45.3 eV; d) −43.0 eV.

6. In the Ritz method (M terms) we obtain approximate wave functions only for:
   a) the ground state; b) the ground state and M excited states; c) M states; d) one-electron systems.

7. In the perturbational method for the ground state ($k = 0$):
   a) the first-order correction to the energy is always negative;
   b) the second-order correction to the energy is always negative;
   c) the first-order correction to the energy is the largest among all the perturbational corrections;
   d) the first-order correction to the energy is $E_k^{(1)} = \langle\psi_k^{(0)}|\hat{H}^{(1)}\psi_k^{(1)}\rangle$, where $\psi_k^{(0)}$ stands for the unperturbed wave function, $\hat{H}^{(1)}$ is the perturbation operator and $\psi_k^{(1)}$ is the first-order correction to the wave function.

8. Perturbation theory [$\hat{H}, \hat{H}^{(0)}, \hat{H}^{(1)}$ stand for the total (perturbed), unperturbed and perturbation Hamiltonian operators, $\psi_k^{(0)}$ the normalized unperturbed wave function of state k corresponding to the energy $E_k^{(0)}$]. The first-order correction to energy $E_k^{(1)}$ satisfies the following relation:
   a) $E_k^{(1)} = \langle\psi_k^{(0)}|\hat{H}^{(1)}\psi_k^{(0)}\rangle$; b) $E_k^{(1)} = \langle\psi_k^{(0)}|\hat{H}^{(0)}\psi_k^{(0)}\rangle$; c) $E_k^{(1)} = \langle\psi_k^{(0)}|\hat{H}^{(0)}\psi_k^{(1)}t\rangle$; d) $E_k^{(1)} = \langle\psi_k^{(1)}|\hat{H}^{(0)}\psi_k^{(0)}\rangle$.

9. In perturbation theory:
   a) we can obtain accurate results despite the fact that the perturbation series diverges (converges asymptotically);

b) the first-order correction to the energy has to be larger than the second-order correction;
c) the wave function $\psi_k^{(0)}$ should not have any nodes;
d) $E_k^{(1)} > 0$.

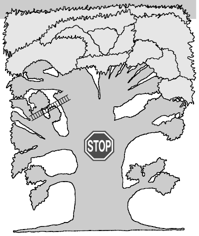
10. Perturbation theory [$\hat{H}, \hat{H}^{(0)}, \hat{H}^{(1)}$ stand for the total (perturbed), unperturbed and perturbation Hamiltonian operators, $\psi_k^{(0)}$ the normalized unperturbed wave function of state k corresponding to energy $E_k^{(0)}$]. The following equation is satisfied:
a) $\hat{H}^{(0)}\psi_k^{(0)} = E_k^{(1)}\psi_k^{(1)} + E_k^{(0)}\psi_k^{(1)}$; b) $\hat{H}^{(1)}\psi_k^{(0)} = E_k^{(1)}\psi_k^{(0)} + E_k^{(0)}\psi_k^{(1)}$; c) $\hat{H}^{(0)}\psi_k^{(1)} + \hat{H}^{(1)}\psi_k^{(0)} = E_k^{(1)}\psi_k^{(0)} + E_k^{(0)}\psi_k^{(1)}$; d) $\hat{H}^{(0)}\psi_k^{(1)} + \hat{H}^{(1)}\psi_k^{(0)} = E_k^{(1)}\psi_k^{(0)}$.

Answers

1a, 2d, 3b, 4b, 5a, 6c, 7b, 8a, 9a, 10c

Chapter 6

SEPARATION OF ELECTRONIC AND NUCLEAR MOTIONS



WHERE ARE WE?

Where are we?

We are on the *most important* branch of the TREE.

An example

A colleague shows us the gas phase absorption spectra of the hydrogen atom and of the hydrogen molecule recorded in the ultraviolet and visible (UV-VIS), infrared (IR) and microwave range. The spectrum of the hydrogen atom consists of separated narrow absorption lines. The hydrogen molecule spectrum is much more complex, instead of the absorption lines we have some absorption bands with a regular and mysterious structure. If the theory given in the previous chapters is correct, then it should explain why these bands appear and why the spectra have such a strange structure.

What is it all about

Nuclei are thousands times heavier than the electrons. As an example let us take the hydrogen atom. From the conservation of momentum law, it follows that the proton moves 1840 times slower than the electron. In a polyatomic system, while a nucleus moves a little, an electron travels many times through the molecule. It seems that a lot can be simplified when assuming electronic motion in a field created by immobile nuclei. This concept is behind what is called *adiabatic approximation*, in which the motions of the electrons and the nuclei are separated.[1] *Only after this approximation is introduced, can we obtain the fundamental concept of chemistry: the molecular structure in 3D space.*

The separation of the electronic and nuclear motions will be demonstrated in detail by taking the example of a diatomic molecule.

Why is it important?

The separation of the electronic and nuclear motions represents a fundamental approximation of quantum chemistry. Without this, chemists would lose their *basic model of the molecule*: the 3D structure with the nuclei occupying some positions in 3D space, with chemical bonds etc. This is why the present chapter occupies the *central* position on the TREE.

What is needed?

- Postulates of quantum mechanics (Chapter 1, needed).
- Separation of the centre-of-mass motion (Appendix I on p. 971, necessary).
- Rigid rotator (Chapter 4, necessary).
- Harmonic and Morse oscillators (Chapter 4, necessary).
- Conclusions from group theory (Appendix C, p. 903, advised).

---

[1] It does not mean that the electrons and the nuclei move independently. We obtain two coupled equations: one for the motion of the electrons in the field of the fixed nuclei, and the other for the motion of the nuclei in the potential averaged over the electronic positions.

Classical papers

John von Neumann (1903–1957) known as Jancsi (then Johnny) was the wunderkind of a top Hungarian banker (Jancsi showed off at receptions by reciting from memory all the phone numbers after reading a page of the phone book). He attended the same famous Lutheran High School in Budapest as Jenó Pál (who later used the name Eugene) Wigner. In 1926 von Neumann received his chemistry engineering diploma, and in the same year he completed his PhD in mathematics at the University of Budapest. He finally emigrated to the USA and founded the Princeton Advanced Study Institute. John von Neumann was a mathematical genius. He contributed to the mathematical foundations of quantum theory, computers, and game theory. Von Neumann made a strange offer of a professorship at the Advanced Study Institute to Stefan Banach from the John Casimir University in Lwów. He handed him a cheque with a handwritten figure “1” and asked Banach to add as many zeros as he wanted. “*This is not enough money to persuade me to leave Poland*” – answered Banach.

The conical intersection problem was first recognized by three young and congenial Hungarians: Janos (later John) von Neumann and Jenó Pál (later Eugene) Wigner in the papers “*Über merkwürdige diskrete Eigenwerte*” in *Physikalische Zeitschrift*, 30 (1929) 465 and “*Über das Verhalten von Eigenwerten bei adiabatischen Prozessen*” also published in *Physikalische Zeitschrift*, 30 (1929) 467, and later in a paper by Edward Teller published in the *Journal of Chemical Physics*, 41 (1937) 109. ★ A fundamental approximation (called the Born–Oppenheimer approximation) has been introduced in the paper “*Zur Quantentheorie der Molekeln*” by Max Born and Julius Robert Oppenheimer in *Annalen der Physik*, 84 (1927) 457, which follows from the fact that nuclei are much heavier than electrons. ★ Gerhard Herzberg was the greatest spectroscopist of the XX century, author of the fundamental three-volume work: “*Spectra of Diatomic Molecules*” (1939), “*Infrared and Raman Spectra of Polyatomic Molecules*” (1949) and “*Electronic Spectra of Polyatomic Molecules*” (1966).

Edward Teller (1908–2004), American physicist of Hungarian origin, professor at the George Washington University, the University of Chicago and the University of California. Teller left Hungary in 1926, received his PhD in 1930 at the University of Leipzig, and fled Nazi Germany in 1935. Teller was the project leader and the top brain behind the American hydrogen bomb project in Los Alamos, believing that this was the way to overthrow communism. The hydrogen bomb patent is owned by Edward Teller and Stanisław Ulam. Interrogated on Robert Oppenheimer’s possible contacts with Soviet Intelligence Service, he declared: “*I feel I would prefer to see the vital interests of this country in hands that I understand better and therefore trust more*”.

Eugene Paul Wigner (1902–1995), American chemist, physicist and mathematician of Hungarian origin, professor at the Princeton University (USA). At the age of 11 Wigner, a primary schoolboy from Budapest, was in a sanatorium in Austria with suspected tuberculosis. Lying for hours on a deck-chair reading books, he was seduced by the beauty of mathematics (fortunately, it turned out he did not have tuberculosis). In 1915 Wigner entered the famous Lutheran High School in Budapest. Fulfilling the wish of his father, who dreamed of having a successor in managing the familial tannery, Wigner graduated from the Technical University in Budapest as a chemist. In 1925, at the Technical University in Berlin he defended his PhD thesis on chemical kinetics "*Bildung und Zerfall von Molekülen*" under the supervision of Michael Polanyi, a pioneer in the study of chemical reactions. In 1926 Wigner left the tannery... Accidentally he was advised by his colleague von Neumann, to focus on group theory (where he obtained the most spectacular successes). Wigner was the first to understand the main features of the nuclear forces. In 1963 he won the Nobel Prize "*for his contributions to the theory of the atomic nucleus and elementary particles, particularly through the discovery and application of fundamental symmetry principles*".

★ The world's first computational papers using a rigorous approach to go beyond the Born–Oppenheimer approximation for molecules were two articles by Włodzimierz Kołos and Lutosław Wolniewicz, the first in *Acta Physica Polonica* 20 (1961) 129 entitled "*The Coupling between Electronic and Nuclear Motion and the Relativistic Effects in the Ground State of the* H_2 *Molecule*" and the second in *Physics Letters*, 2 (1962) 222 entitled "*A Complete Non-Relativistic Treatment of the* H_2 *Molecule*". ★ The discovery of the conical intersection and the funnel effect in photochemistry is attributed to Howard E. Zimmerman [*Journal of the American Chemical Society*, 88 (1966) 1566[2]] and to Josef Michl [*Journal of Molecular Photochemistry*, 243 (1972)]. Important contributions in this domain were also made by Lionel Salem and Christopher Longuet-Higgins.

Christopher Longuet-Higgins, professor at the University of Sussex, Great Britain, began his scientific career as a theoretical chemist. His main achievements are connected with conical intersection, as well as with the introduction of permutational groups in the theoretical explanation of the spectra of flexible molecules. Longuet-Higgins was elected the member of the Royal Society of London for these contributions. He turned to artificial intelligence at the age of 40, and in 1967 he founded the Department of Machine Intelligence and Perception at the University of Edinburgh. Longuet-Higgins investigated machine perception of speech and music. His contribution to this field was recognized by the award of an Honorary Doctorate in Music by Sheffield University.

---

[2] The term "funnel effect" was coined in this paper.

6.1 SEPARATION OF THE CENTRE-OF-MASS MOTION

6.1.1 SPACE-FIXED COORDINATE SYSTEM (SFCS)

Let us consider first a diatomic molecule with the nuclei labelled by a, b, and n electrons. Let us choose a Cartesian coordinate system in our laboratory (called the space-fixed coordinate system, SFCS) with the origin located at an arbitrarily chosen point and with arbitrary orientation of the axes.[3] The nuclei have the following positions: $\boldsymbol{R}_a = (X_a, Y_a, Z_a)$ and $\boldsymbol{R}_b = (X_b, Y_b, Z_b)$, while electron i has the coordinates x_i', y_i', z_i'.

We write the Hamiltonian for the system (Chapter 1):

$$\hat{\mathcal{H}} = -\frac{\hbar^2}{2M_a}\Delta_a - \frac{\hbar^2}{2M_b}\Delta_b - \sum_{i=1}^{n}\frac{\hbar^2}{2m}\Delta_i' + V, \tag{6.1}$$

where the first two terms stand for the kinetic energy operators of the nuclei (with masses M_a and M_b), the third term corresponds to the kinetic energy of the electrons (m is the electron mass, all Laplacians are in the space-fixed coordinate system), and V denotes the Coulombic potential energy operator (interaction of all the particles, nuclei–nuclei, nuclei–electrons, electrons–electrons; $\mathcal{Z}_a e$ and $\mathcal{Z}_b e$ are nuclear charges)

$$V = \frac{\mathcal{Z}_a\mathcal{Z}_b e^2}{R} - \mathcal{Z}_a\sum_i \frac{e^2}{r_{ai}} - \mathcal{Z}_b\sum_i \frac{e^2}{r_{bi}} + \sum_{i<j}\frac{e^2}{r_{ij}}. \tag{6.2}$$

When we are not interested in collisions of our molecule with a wall or similar obstruction, we may consider a separation of the motion of the centre-of-mass, then forget about the motion and focus on the rest, i.e. on the *relative* motion of the particles.

6.1.2 NEW COORDINATES

The total mass of the molecule is $M = M_a + M_b + mn$. The components of the centre-of-mass position vector are[4]

$$X = \frac{1}{M}\left(M_a X_a + M_b X_b + \sum_i m x_i'\right),$$

$$Y = \frac{1}{M}\left(M_a Y_a + M_b Y_b + \sum_i m y_i'\right),$$

$$Z = \frac{1}{M}\left(M_a Z_a + M_b Z_b + \sum_i m z_i'\right).$$

---

[3] For example, right in the centre of the Norwich market square.

[4] Do not mix the coordinate Z with the nuclear charge $\mathcal{Z}$.

Now, we decide to abandon this coordinate system (SFCS). Instead of the old coordinates, we will choose a new set of $3n+6$ coordinates (see Appendix I on p. 971, choice II):

- three centre-of-mass coordinates X, Y, Z,
- three components of the vector $\boldsymbol{R} = \boldsymbol{R}_a - \boldsymbol{R}_b$ that separates nucleus a from nucleus b,
- $3n$ electronic coordinates $x_i = x_i' - \frac{1}{2}(X_a + X_b)$ and similarly for y_i and z_i, for $i = 1, 2, \ldots, n$, which show the electron's position with respect to the *geometric centre*[5] of the molecule.

6.1.3 HAMILTONIAN IN THE NEW COORDINATES

The new coordinates have to be introduced into the Hamiltonian. To this end, we need the second derivative operators in the old coordinates to be expressed by the new ones. First (similarly as in Appendix I), let us construct the *first* derivative operators:

$$\begin{aligned}\frac{\partial}{\partial X_a} &= \frac{\partial X}{\partial X_a}\frac{\partial}{\partial X} + \frac{\partial Y}{\partial X_a}\frac{\partial}{\partial Y} + \frac{\partial Z}{\partial X_a}\frac{\partial}{\partial Z} + \frac{\partial R_x}{\partial X_a}\frac{\partial}{\partial R_x} + \frac{\partial R_y}{\partial X_a}\frac{\partial}{\partial R_y} + \frac{\partial R_z}{\partial X_a}\frac{\partial}{\partial R_z} \\ &\quad + \sum_i \frac{\partial x_i}{\partial X_a}\frac{\partial}{\partial x_i} + \sum_i \frac{\partial y_i}{\partial X_a}\frac{\partial}{\partial y_i} + \sum_i \frac{\partial z_i}{\partial X_a}\frac{\partial}{\partial z_i} \\ &= \frac{\partial X}{\partial X_a}\frac{\partial}{\partial X} + \frac{\partial R_x}{\partial X_a}\frac{\partial}{\partial R_x} + \sum_i \frac{\partial x_i}{\partial X_a}\frac{\partial}{\partial x_i} = \frac{M_a}{M}\frac{\partial}{\partial X} + \frac{\partial}{\partial R_x} - \frac{1}{2}\sum_i \frac{\partial}{\partial x_i}\end{aligned}$$

and similarly for the coordinates Y_a and Z_a. For the nucleus b the expression is a little bit different:

$$\frac{\partial}{\partial X_b} = \frac{M_b}{M}\frac{\partial}{\partial X} - \frac{\partial}{\partial R_x} - \frac{1}{2}\sum_i \frac{\partial}{\partial x_i}.$$

For the first derivative operator with respect to the coordinates of the electron i we obtain:

$$\begin{aligned}\frac{\partial}{\partial x_i'} &= \frac{\partial X}{\partial x_i'}\frac{\partial}{\partial X} + \frac{\partial Y}{\partial x_i'}\frac{\partial}{\partial Y} + \frac{\partial Z}{\partial x_i'}\frac{\partial}{\partial Z} + \frac{\partial R_x}{\partial x_i'}\frac{\partial}{\partial R_x} + \frac{\partial R_y}{\partial x_i'}\frac{\partial}{\partial R_y} + \frac{\partial R_z}{\partial x_i'}\frac{\partial}{\partial R_z} \\ &\quad + \sum_j \frac{\partial x_j}{\partial x_i'}\frac{\partial}{\partial x_j} + \sum_j \frac{\partial y_j}{\partial x_i'}\frac{\partial}{\partial y_j} + \sum_j \frac{\partial z_j}{\partial x_i'}\frac{\partial}{\partial z_j} \\ &= \frac{\partial X}{\partial x_i'}\frac{\partial}{\partial X} + \frac{\partial x_i}{\partial x_i'}\frac{\partial}{\partial x_i} = \frac{m}{M}\frac{\partial}{\partial X} + \frac{\partial}{\partial x_i}\end{aligned}$$

and similarly for y_i' and z_i'.

---

[5] If the origin were chosen in the centre of mass instead of the geometric centre, V becomes mass-dependent (J. Hinze, A. Alijah and L. Wolniewicz, *Pol. J. Chem.* 72 (1998) 1293), cf. also Appendix I, Example II. We want to avoid this.

Now, let us create the second derivative operators:

$$\frac{\partial^2}{\partial X_a^2}=\left(\frac{M_a}{M}\frac{\partial}{\partial X}+\frac{\partial}{\partial R_x}-\frac{1}{2}\sum_i\frac{\partial}{\partial x_i}\right)^2=\left(\frac{M_a}{M}\right)^2\frac{\partial^2}{\partial X^2}+\frac{\partial^2}{\partial R_x^2}+\frac{1}{4}\left(\sum_i\frac{\partial}{\partial x_i}\right)^2$$

$$+2\frac{M_a}{M}\frac{\partial}{\partial X}\frac{\partial}{\partial R_x}-\frac{\partial}{\partial R_x}\sum_i\frac{\partial}{\partial x_i}-\frac{M_a}{M}\frac{\partial}{\partial X}\sum_i\frac{\partial}{\partial x_i},$$

$$\frac{\partial^2}{\partial X_b^2}=\left(\frac{M_b}{M}\frac{\partial}{\partial X}-\frac{\partial}{\partial R_x}-\frac{1}{2}\sum_i\frac{\partial}{\partial x_i}\right)^2=\left(\frac{M_b}{M}\right)^2\frac{\partial^2}{\partial X^2}+\frac{\partial^2}{\partial R_x^2}+\frac{1}{4}\left(\sum_i\frac{\partial}{\partial x_i}\right)^2$$

$$-2\frac{M_b}{M}\frac{\partial}{\partial X}\frac{\partial}{\partial R_x}+\frac{\partial}{\partial R_x}\sum_i\frac{\partial}{\partial x_i}-\frac{M_b}{M}\frac{\partial}{\partial X}\sum_i\frac{\partial}{\partial x_i},$$

$$\frac{\partial^2}{\partial (x_i')^2}=\left(\frac{m}{M}\frac{\partial}{\partial X}+\frac{\partial}{\partial x_i}\right)^2=\left(\frac{m}{M}\right)^2\frac{\partial^2}{\partial X^2}+\frac{\partial^2}{\partial x_i^2}+2\frac{m}{M}\frac{\partial}{\partial X}\frac{\partial}{\partial x_i}.$$

After inserting all this into the Hamiltonian (6.1) we obtain the Hamiltonian expressed in the new coordinates:[6]

clamped nuclei Hamiltonian

$$\hat{\mathcal{H}}=-\frac{\hbar^2}{2M}\Delta_{XYZ}+\hat{H}_0+\hat{H}', \tag{6.3}$$

where the first term means the centre-of-mass kinetic energy operator, $\hat{H}_0$ is the *electronic Hamiltonian* (*clamped nuclei Hamiltonian*)

electronic Hamiltonian

$$\hat{H}_0=-\sum_i\frac{\hbar^2}{2m}\Delta_i+V, \tag{6.4}$$

while $\Delta_i\equiv\frac{\partial^2}{\partial x_i^2}+\frac{\partial^2}{\partial y_i^2}+\frac{\partial^2}{\partial z_i^2}$ and

$$\hat{H}'=-\frac{\hbar^2}{2\mu}\Delta_{\boldsymbol{R}}+\hat{H}'' \tag{6.5}$$

with $\Delta_{\boldsymbol{R}}\equiv\frac{\partial^2}{\partial R_x^2}+\frac{\partial^2}{\partial R_y^2}+\frac{\partial^2}{\partial R_z^2}$, where

$$\hat{H}''=\left[-\frac{\hbar^2}{8\mu}\left(\sum_i\nabla_i\right)^2-\frac{\hbar^2}{2}\left(\frac{1}{M_a}-\frac{1}{M_b}\right)\nabla_{\boldsymbol{R}}\sum_i\nabla_i\right],$$

and μ denotes the reduced mass of the two nuclei ($\mu^{-1}=M_a^{-1}+M_b^{-1}$).

---

[6] The potential energy also has to be expressed using the new coordinates.

The $\hat{H}_0$ does not contain the kinetic energy operator of the nuclei, but all the other terms (this is why it is called the electronic Hamiltonian): the first term stands for the kinetic energy operator of the electrons, and V means the potential energy corresponding to the Coulombic interaction of all particles. The first term in the operator $\hat{H}'$, i.e. $-\frac{\hbar^2}{2\mu}\Delta_{\boldsymbol{R}}$, denotes the kinetic energy operator of the nuclei,[7] while the operator $\hat{H}''$ couples the motions of the nuclei and electrons.[8]

6.1.4 AFTER SEPARATION OF THE CENTRE-OF-MASS MOTION

After separation of the centre-of-mass motion (the first term in eq. (6.3) is gone, see Appendix I on p. 971) we obtain the eigenvalue problem of the Hamiltonian

$$\hat{H} = \hat{H}_0 + \hat{H}'. \tag{6.6}$$

This is an exact result, fully equivalent to the Schrödinger equation.

6.2 EXACT (NON-ADIABATIC) THEORY

The total wave function that describes both electrons and nuclei can be proposed in the following form[9]

---

[7] What moves is a particle of reduced mass μ and coordinates R_x, R_y, R_z. This means that the particle has the position of nucleus a, whereas nucleus b is at the origin. Therefore, this term accounts for the vibrations of the molecule (changes in length of $\boldsymbol{R}$), as well as its rotations (changes in orientation of $\boldsymbol{R}$).

[8] The first of these two terms contains the reduced mass of the two nuclei, where ∇_i denotes the nabla operator for *electron* i, $\nabla_i \equiv \boldsymbol{i}\frac{\partial}{\partial x} + \boldsymbol{j}\frac{\partial}{\partial y} + \boldsymbol{k}\frac{\partial}{\partial z}$ with $\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k}$ being the unit vectors along the axes x, y, z. The second term is non-zero only for the heteronuclear case and contains the mixed product of nablas: $\nabla_R \nabla_i$ with $\nabla_{\boldsymbol{R}} = \boldsymbol{i}\frac{\partial}{\partial R_x} + \boldsymbol{j}\frac{\partial}{\partial R_y} + \boldsymbol{k}\frac{\partial}{\partial R_z}$ and R_x, R_y, R_z as the components of the vector $\boldsymbol{R}$.

[9] Where did such a form of the wave function come from?

If the problem were solved exactly, then the solution of the Schrödinger equation could be sought, e.g., by using the Ritz method (p. 202). Then we have to decide what kind of basis set to use. We could use two auxiliary complete basis sets: one that depended on the electronic coordinates $\{\bar{\psi}_k(\boldsymbol{r})\}$, and the second on the nuclear coordinates $\{\bar{\phi}_l(\boldsymbol{R})\}$. The complete basis set for the Hilbert space of our system could be constructed as a Cartesian product $\{\bar{\psi}_k(\boldsymbol{r})\} \times \{\bar{\phi}_l(\boldsymbol{R})\}$, i.e. all possible product-like functions $\bar{\psi}_k(\boldsymbol{r})\bar{\phi}_l(\boldsymbol{R})$. Thus, the wave function could be expanded in a series

$$\Psi(\boldsymbol{r}, \boldsymbol{R}) = \sum_{kl} c_{kl}\bar{\psi}_k(\boldsymbol{r})\bar{\phi}_l(\boldsymbol{R}) = \sum_k^{\mathcal{N}} \bar{\psi}_k(\boldsymbol{r})\Big[\sum_l c_{kl}\bar{\phi}_l(\boldsymbol{R})\Big] = \sum_k^{\mathcal{N}} \bar{\psi}_k(\boldsymbol{r}) f_k(\boldsymbol{R}),$$

where $f_k(\boldsymbol{R}) = \sum_l c_{kl}\bar{\phi}_l(\boldsymbol{R})$ stands for a to-be-sought coefficient depending on $\boldsymbol{R}$ (rovibrational function). If we had to do with complete sets, then both $\bar{\psi}_k$ and f_k should not depend on anything else, since a sufficiently long expansion of the terms $\bar{\psi}_k(\boldsymbol{r})\bar{\phi}_l(\boldsymbol{R})$ would be suitable to describe all possible distributions of the electrons and the nuclei.

However, we are unable to manage the complete sets, instead, we are able to take only a few terms in this expansion. We would like them to describe the molecule reasonably well, and at the same time to

$$\Psi(\boldsymbol{r},\boldsymbol{R})=\sum_k^{\mathcal{N}}\psi_k(\boldsymbol{r};R)f_k(\boldsymbol{R}), \tag{6.7}$$

where $\psi_k(\boldsymbol{r};R)$ are the eigenfunctions of $\hat{H}_0$

$$\hat{H}_0(R)\psi_k(\boldsymbol{r};R)=E_k^0(R)\psi_k(\boldsymbol{r};R) \tag{6.8}$$

that depend parametrically[10] on the internuclear distance R, and $f_k(\boldsymbol{R})$ are yet unknown rovibrational functions (describing the rotations and vibrations of the molecule).

Derivation

First, let us write down the Schrödinger equation with the Hamiltonian (6.6) and the wave function as in (6.7)

$$\left(\hat{H}_0+\hat{H}'\right)\sum_k^{\mathcal{N}}\psi_k(\boldsymbol{r};R)f_k(\boldsymbol{R})=E\sum_k^{\mathcal{N}}\psi_k(\boldsymbol{r};R)f_k(\boldsymbol{R}). \tag{6.9}$$

Let us multiply both sides by $\psi_l^*(\boldsymbol{r};R)$ and then integrate over the *electronic* coordinates $\boldsymbol{r}$ (which will be stressed by the subscript "e"):

$$\sum_k^{\mathcal{N}}\left\langle\psi_l\middle|\left(\hat{H}_0+\hat{H}'\right)(\psi_k f_k)\right\rangle_e=E\sum_k^{\mathcal{N}}\langle\psi_l|\psi_k\rangle_e f_k. \tag{6.10}$$

On the right-hand side of (6.10) we profit from the orthonormalization condition $\langle\psi_l|\psi_k\rangle_e=\delta_{kl}$, on the left-hand side we recall that ψ_k is an eigenfunction of $\hat{H}_0$

$$E_l^0 f_l+\sum_k^{\mathcal{N}}\left\langle\psi_l\middle|\hat{H}'(\psi_k f_k)\right\rangle_e=Ef_l. \tag{6.11}$$

Now, let us focus on the expression $\hat{H}'(\psi_k f_k)=-\frac{\hbar^2}{2\mu}\Delta_{\boldsymbol{R}}(\psi_k f_k)+\hat{H}''(\psi_k f_k)$, which we have in the integrand in eq. (6.11). Let us concentrate on the first of

---

have only a few, to be exact only one such term. If so, it would be reasonable to introduce a *parametric dependence of* the function $\bar{\psi}_k(\boldsymbol{r})$ on the position of the nuclei, which in our case of a diatomic molecule means the internuclear distance. *This is equivalent to telling someone how the electrons behave when the internuclear distances are such and such, and how they behave, when the distances are changed.*

[10] For each value of R we have a different formula for ψ_k.

these terms[11]

$$\begin{aligned}
-\frac{\hbar^2}{2\mu}\Delta_{\boldsymbol{R}}(\psi_k f_k) &= -\frac{\hbar^2}{2\mu}\nabla_{\boldsymbol{R}}\nabla_{\boldsymbol{R}}(\psi_k f_k) = -\frac{\hbar^2}{2\mu}\nabla_{\boldsymbol{R}}\big[\psi_k\nabla_{\boldsymbol{R}}f_k + (\nabla_{\boldsymbol{R}}\psi_k)f_k\big] \\
&= -\frac{\hbar^2}{2\mu}\big[\nabla_{\boldsymbol{R}}\psi_k\nabla_{\boldsymbol{R}}f_k + \psi_k\Delta_{\boldsymbol{R}}f_k + (\Delta_{\boldsymbol{R}}\psi_k)f_k + \nabla_{\boldsymbol{R}}\psi_k\nabla_{\boldsymbol{R}}f_k\big] \\
&= -\frac{\hbar^2}{2\mu}\big[2(\nabla_{\boldsymbol{R}}\psi_k)(\nabla_{\boldsymbol{R}}f_k) + \psi_k\Delta_{\boldsymbol{R}}f_k + (\Delta_{\boldsymbol{R}}\psi_k)f_k\big]. \qquad (6.12)
\end{aligned}$$

After inserting the result into $\langle\psi_l|\hat{H}'(\psi_k f_k)\rangle_e$ and recalling eq. (6.5) we have

$$\begin{aligned}
\langle\psi_l|\hat{H}'[\psi_k f_k]\rangle_e &= 2\Big(-\frac{\hbar^2}{2\mu}\Big)\langle\psi_l|\nabla_{\boldsymbol{R}}\psi_k\rangle_e\nabla_{\boldsymbol{R}}f_k + \langle\psi_l \mid \psi_k\rangle_e\Big(-\frac{\hbar^2}{2\mu}\Big)\Delta_{\boldsymbol{R}}f_k \\
&\quad + \Big\langle\psi_l\Big|\Big(-\frac{\hbar^2}{2\mu}\Big)\Delta_{\boldsymbol{R}}\psi_k\Big\rangle_e f_k + \langle\psi_l|\hat{H}''\psi_k\rangle_e f_k \\
&= (1-\delta_{kl})\Big(-\frac{\hbar^2}{\mu}\Big)\langle\psi_l|\nabla_{\boldsymbol{R}}\psi_k\rangle_e\nabla_{\boldsymbol{R}}f_k - \delta_{kl}\frac{\hbar^2}{2\mu}\Delta_{\boldsymbol{R}}f_k \\
&\quad + H_{lk}f_k, \qquad (6.13)
\end{aligned}$$

with

$$H'_{lk} \equiv \langle\psi_l|\hat{H}'\psi_k\rangle_e.$$

We obtain the following form of (6.11)

$$E_l^0 f_l + \sum_k^{\mathcal{N}}\Big[(1-\delta_{kl})\Big(-\frac{\hbar^2}{\mu}\Big)\langle\psi_l|\nabla_{\boldsymbol{R}}\psi_k\rangle_e\nabla_{\boldsymbol{R}}f_k - \delta_{kl}\frac{\hbar^2}{2\mu}\Delta_{\boldsymbol{R}}f_k + H'_{lk}f_k\Big] = Ef_l.$$

(we have profited from the equality $\langle\psi_k|\nabla_{\boldsymbol{R}}\psi_k\rangle_e = 0$, which follows from the differentiation of the normalization condition[12] for the function ψ_k)

Non-adiabatic nuclear motion

Grouping all the terms with f_l on the left-hand side we obtain a set of $\mathcal{N}$ equations

---

[11] We use the relation $\Delta_{\boldsymbol{R}} = (\nabla_{\boldsymbol{R}})^2$.

[12] We assume that the phase of the wave function $\psi_k(\boldsymbol{r}; R)$ does not depend on R, i.e. $\psi_k(\boldsymbol{r}; R) = \tilde{\psi}_k(\boldsymbol{r}; R)\exp(i\phi)$, where $\tilde{\psi}_k$ is a real function and $\phi \neq \phi(R)$. This immediately gives $\langle\psi_k|\nabla_{\boldsymbol{R}}\psi_k\rangle_e = \langle\tilde{\psi}_k|\nabla_{\boldsymbol{R}}\tilde{\psi}_k\rangle_e$, which is zero from differentiating the normalization condition. Indeed, the normalization condition: $\int\psi_k^2\,\mathrm{d}\tau_e = 1$. Hence, $\nabla_{\boldsymbol{R}}\int\psi_k^2\,\mathrm{d}\tau_e = 0$, or $2\int\psi_k\nabla_{\boldsymbol{R}}\psi_k\,\mathrm{d}\tau_e = 0$.
Without this approximation we will surely have trouble.

$$\left[-\frac{\hbar^2}{2\mu}\Delta_{\boldsymbol{R}}+E_l^0(R)+H'_{ll}(R)-E\right]f_l=-\sum_{k(\neq l)}^{\mathcal{N}}\Theta_{lk}f_k, \tag{6.14}$$

for $l=1,2,\dots,\mathcal{N}$ with the non-adiabatic coupling operators

$$\Theta_{lk}=-\frac{\hbar^2}{\mu}\langle\psi_l|\nabla_{\boldsymbol{R}}\psi_k\rangle_e\nabla_{\boldsymbol{R}}+H'_{lk}. \tag{6.15}$$

Note, that the operator H'_{lk} depends on the length of the vector $\boldsymbol{R}$, but not on its direction.[13]

Eq. (6.14) is equivalent to the Schrödinger equation.

Eqs. (6.14) and (6.15) have been derived under the assumption that ψ_k of eq. (6.7) satisfy (6.8). If instead of $\psi_k(\boldsymbol{r};R)$ we use a (generally non-orthogonal) complete set $\{\bar{\psi}_k(\boldsymbol{r};R)\}$ in (6.7), eqs. (6.14) and (6.15) would change to

$$\left[-\frac{\hbar^2}{2\mu}\Delta_{\boldsymbol{R}}+\bar{E}_l(R)+H'_{ll}(R)-E\right]f_l=-\sum_{k(\neq l)}^{\mathcal{N}}\Theta_{lk}f_k, \tag{6.16}$$

for $l=1,2,\dots,\mathcal{N}$ with the non-adiabatic coupling operators

$$\Theta_{lk}=-\frac{\hbar^2}{\mu}\langle\bar{\psi}_l|\nabla_{\boldsymbol{R}}\bar{\psi}_k\rangle_e\nabla_{\boldsymbol{R}}+H'_{lk}+\langle\bar{\psi}_l|\bar{\psi}_k\rangle_e\left(-\frac{\hbar^2}{2\mu}\Delta_{\boldsymbol{R}}\right) \tag{6.17}$$

and $\bar{E}_l(R)\equiv\langle\bar{\psi}_l|\hat{H}_0\bar{\psi}_l\rangle_e$.

6.3 ADIABATIC APPROXIMATION

If the curves $E_l^0(R)$ for different l are well separated on the energy scale, we may expect that the coupling between them is small, and therefore all Θ_{kl} for $k\neq l$ may be set equal to zero. This is called the adiabatic approximation. In this approximation we obtain from (6.14):

$$\left[-\frac{\hbar^2}{2\mu}\Delta_{\boldsymbol{R}}+E_l^0(R)+H'_{ll}(R)\right]f_l=Ef_l, \tag{6.18}$$

where the diagonal correction $H'_{ll}(R)$ is usually very small compared to $E_l^0(R)$. diagonal correction

In the adiabatic approximation the wave function is approximated by a product

---

[13] This follows from the fact that we have in $\hat{H}'$ (see eq. (6.5)) the products of nablas, i.e. scalar products. The scalar products do not change upon rotation, because both vectors involved rotate in the same way and the angle between them does not change.

$$\Psi \approx \psi_l(\boldsymbol{r}; R) f_l(\boldsymbol{R}). \tag{6.19}$$

The function $f_k(\boldsymbol{R})$ depends explicitly not only on R, but also on the direction of vector $\boldsymbol{R}$, and therefore will describe future vibrations of the molecule (changes of R) as well as its rotations (changes of the direction of $\boldsymbol{R}$).

A simple analogy

Let us stop for a while to catch the sense of the adiabatic approximation.

To some extent the situation resembles an attempt to describe a tourist (an electron) and the Alps (nuclei). Not only the tourist moves, but also the Alps, as has been quite convincingly proved by geologists.[14] The probability of encountering the tourist may be described by a "wave function" computed for a fixed position of the mountains (shown by a map bought in a shop). This is a very good approximation, because when the tourist wanders over hundreds miles, the beloved Alps move a tiny, tiny distance, so that the map seems to be perfect all the time. On the other hand the probability of having the Alps in a given configuration is described by the geologists' "wave function" f, saying for example, the probability that the distance between the Matterhorn and the Jungfrau is equal to R. When the tourist revisits the Alps after a period of time (say, a few million of years), the mountains will be changed (the new map bought in the shop will reflect this fact). The probability of finding the tourist may again be computed from the new wave function *valid for the new configuration of the mountains (a parametric dependence)*. Therefore, the probability of finding the tourist in the spot indicated by the vector $\boldsymbol{r}$ at a given configuration of the mountains $\boldsymbol{R}$ can be *approximated*[15] *by a product* of the probability of finding the mountains at this configuration $|f_l(\boldsymbol{R})|^2\, \mathrm{d}^3\boldsymbol{R}$ and the probability $|\psi_l(\boldsymbol{r}; \boldsymbol{R})|^2\, \mathrm{d}^3\boldsymbol{r}$ of finding the tourist in the position shown by the vector $\boldsymbol{r}$, when the mountains have this particular configuration $\boldsymbol{R}$. In the case of our molecule this means the adiabatic approximation (a product-like form), eq. (6.19).

This parallel fails in one important point: the Alps do not move in the potential created by tourists, the dominant geological processes are tourist-independent. As we will soon see, nuclear motion is dictated by the potential *which is the electronic energy*.

---

[14] The continental plates collide like billiard balls in a kind of quasi-periodic oscillation. During the current oscillation, the India plate which moved at record speed of about 20 cm a year hit the Euroasiatic plate. This is why the Himalayan mountains are so beautiful. The collision continues and the Himalayas will be even more beautiful. Europe was hit from the South by a few plates moving at about 4 cm a year, and this is why we have much lower Alps. While visiting the Atlantic coast of Maine (USA), I thought that the colour of the rocks was very similar to those I remembered from Brittany (France). That was it! Once upon a time the two coasts made a common continent. Later we had to rediscover America. The Wegener theory of continental plate tectonics, when created in 1911, was viewed as absurd, although the mountain *ranges* suggested that some plates were colliding.

[15] This is an approximation, because in the non-adiabatic, i.e. fully correct, approach the total wave function is a superposition of many such products, eq. (6.7), corresponding to various electronic and rovibrational wave functions.

6.4 BORN–OPPENHEIMER APPROXIMATION

In the adiabatic approximation, $H'_{ll} = \int \psi_l^* H' \psi_l \, \mathrm{d}\tau_e$ represents a small correction to $E_l^0(R)$. *Neglecting the correction results in the Born–Oppenheimer approximation*

$$H'_{ll} \cong 0.$$

Note that in the Born–Oppenheimer approximation the potential energy for the motion of the nuclei $E_l^0(R)$ is independent of the mass of the nuclei, whereas in the adiabatic approximation the potential energy $E_l^0(R) + H'_{ll}(R)$ depends on the mass.

Julius Robert Oppenheimer (1904–1967), American physicist, professor at the University of California in Berkeley and the California Institute of Technology in Pasadena, and at the Institute for Advanced Study in Princeton. In 1943–1945 Oppenheimer headed the Manhattan Project (atomic bomb).

From John Slater's autobiography: "*... Robert Oppenheimer was a very brilliant physics undergraduate at Harvard during the 1920s, the period when I was there on the faculty, and we all recognized that he was a person of very unusual attainments. Rather than going on for his graduate work at Harvard, he went to Germany, and worked with Born, developing what has been known as the Born–Oppenheimer approximation.*"

6.5 OSCILLATIONS OF A ROTATING MOLECULE

Our next step will be an attempt to separate rotations and oscillations within the adiabatic approximation. To this end the function $f_k(\boldsymbol{R}) = f_k(R, \theta, \phi)$ will be proposed as a *product* of a function Y which will account for rotations (depending on θ, ϕ), and a certain function $\frac{\chi_k(R)}{R}$ describing the oscillations, i.e. dependent on R

$$f_k(\boldsymbol{R}) = Y(\theta, \phi)\frac{\chi_k(R)}{R}. \tag{6.20}$$

No additional approximation is introduced. We say only that the isolated molecule vibrates absolutely independently of whether it is oriented towards the Capricorn or Taurus Constellations (space is isotropic). The function $\chi_k(R)$ is yet unknown, therefore dividing by R in (6.20) is meaningless.[16]

---

[16] In the case of polyatomics the function $f_k(\boldsymbol{R})$ may be more complicated, because some *vibrations* (e.g., a rotation of the CH_3 group) may contribute to the total angular momentum, which has to be conserved (this is related to space isotropy, cf. p. 63).

Now, we will try to separate the variables θ, ϕ from the variable R in eq. (6.18), i.e. to obtain two separate equations for them. First, let us define the quantity

$$U_k(R) = E_k^0(R) + H'_{kk}(R). \tag{6.21}$$

After inserting the Laplacian (in spherical coordinates, see Appendix H on p. 969) and the product (6.20) into (6.18) we obtain the following series of transformations

$$\left[-\frac{\hbar^2}{2\mu}\left(\frac{1}{R^2}\frac{\partial}{\partial R}R^2\frac{\partial}{\partial R} + \frac{1}{R^2\sin\theta}\frac{\partial}{\partial\theta}\sin\theta\frac{\partial}{\partial\theta} + \frac{1}{R^2\sin^2\theta}\frac{\partial^2}{\partial\phi^2}\right) + U_k(R)\right]Y\frac{\chi_k}{R}$$
$$= EY\frac{\chi_k}{R},$$
$$-\frac{\hbar^2}{2\mu}\left(\frac{Y}{R}\frac{\partial^2\chi_k}{\partial R^2} + \frac{\chi_k}{R}\frac{1}{R^2\sin\theta}\frac{\partial}{\partial\theta}\sin\theta\frac{\partial Y}{\partial\theta} + \frac{\chi_k}{R}\frac{1}{R^2\sin^2\theta}\frac{\partial^2 Y}{\partial\phi^2}\right) + YU_k(R)\frac{\chi_k}{R}$$
$$= EY\frac{\chi_k}{R},$$
$$-\frac{\hbar^2}{2\mu}\left(\frac{1}{\chi_k}\frac{\partial^2\chi_k}{\partial R^2} + \frac{1}{Y}\left(\frac{1}{R^2\sin\theta}\frac{\partial}{\partial\theta}\sin\theta\frac{\partial Y}{\partial\theta} + \frac{1}{R^2\sin^2\theta}\frac{\partial^2 Y}{\partial\phi^2}\right)\right) + U_k(R) = E,$$
$$-\left(\frac{R^2}{\chi_k}\frac{\partial^2\chi_k}{\partial R^2}\right) + \frac{2\mu}{\hbar^2}U_k(R)R^2 - \frac{2\mu}{\hbar^2}ER^2 = \frac{1}{Y}\left(\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\sin\theta\frac{\partial Y}{\partial\theta} + \frac{1}{\sin^2\theta}\frac{\partial^2 Y}{\partial\phi^2}\right).$$

Look! The left-hand side only depends on R, and the right-hand side only on θ and ϕ. Both sides equal each other *independently of the values of the variables*. This can only happen if each side is equal to a *constant* (λ), the same for each! Therefore, we have

$$-\left(\frac{R^2}{\chi_k}\frac{\partial^2\chi_k}{\partial R^2}\right) + \frac{2\mu}{\hbar^2}U_k(R)R^2 - \frac{2\mu}{\hbar^2}ER^2 = \lambda, \tag{6.22}$$

$$\frac{1}{Y}\left(\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\sin\theta\frac{\partial Y}{\partial\theta} + \frac{1}{\sin^2\theta}\frac{\partial^2 Y}{\partial\phi^2}\right) = \lambda. \tag{6.23}$$

Now, we are amazed to see that (6.23) *is identical* (cf. p. 176) *to that which appeared as a result of the transformation of the Schrödinger equation for a rigid rotator*, Y denoting the corresponding wave function. As we know from p. 176 mathematicians have proved that this equation has a solution only if $\lambda = -J(J+1)$, where $J = 0, 1, 2, \ldots$. *Since Y stands for the rigid rotator wave function, which we know very well, we now concentrate exclusively on the function χ_k*, which describes vibrations (changes in the length of $\boldsymbol{R}$).

After inserting the permitted values of λ into (6.22) we get

$$-\frac{\hbar^2}{2\mu}\left(\frac{\partial^2\chi_k}{\partial R^2}\right) + U_k(R)\chi_k - E\chi_k = -\frac{\hbar^2}{2\mu R^2}J(J+1)\chi_k.$$

Let us write this equation in the form of the eigenvalue problem for the unidimensional motion of a particle (we change the partial into the regular derivative) of mass μ

$$\left(-\frac{\hbar^2}{2\mu}\frac{d^2}{dR^2}+V_{kJ}\right)\chi_{kvJ}(R)=E_{kvJ}\chi_{kvJ}(R) \tag{6.24}$$

with potential energy (let us stress that $R>0$)

$$V_{kJ}(R)=U_k(R)+J(J+1)\frac{\hbar^2}{2\mu R^2} \tag{6.25}$$

centrifugal force effect

which takes the centrifugal force effect on the vibrational motion into account. The solution χ_k, as well as the total energy E, have been labelled by two additional indices: the rotational quantum number J (because the potential depends on it) and the numbering of the solutions $v=0,1,2,\ldots$.

The solutions of eq. (6.24) describe the vibrations of the nuclei. The function $V_{kJ}=E_k^0(R)+H'_{kk}(R)+J(J+1)\hbar^2/(2\mu R^2)$ plays the role of the potential energy curve for the motion of the nuclei.

The above equation, and therefore also

the very notion of the *potential energy curve for the motion of the nuclei* appears only after the adiabatic (the product-like wave function, and H'_{kk} preserved) or the Born–Oppenheimer (the product-like wave function, but H'_{kk} removed) approximations are applied.

If the $H'_{kk}(R)$ term were not present in $V_{kJ}(R)$ (as it is in the Born–Oppenheimer approximation), then the potential energy would not depend on the mass of the nuclei.

Therefore, in the Born–Oppenheimer approximation the potential energy is the same for H_2, HD and D_2.

It is worth noting that $V_{kJ}(R)$ only represents the *potential* energy of the motion of the nuclei. If $V_{kJ}(R)$ were a parabola (as it is for the harmonic oscillator), the system would never acquire the energy corresponding to the bottom of the parabola, because the harmonic oscillator energy levels (cf. p. 166) correspond to *higher* energy. The same pertains to V_{kJ} of a more complex shape.

6.5.1 ONE MORE ANALOGY

The fact that the electronic energy $E_k^0(R)$ plays the role of the potential energy for oscillations represents not only the result of rather complex derivations, but is also natural and understandable. The nuclei keep together thanks to the electronic "glue" (we will come back to this in Chapter 8). Let us imagine two metallic balls (nuclei) in a block of transparent glue (electronic cloud), Fig. 6.1.

If we were interested in the motion of the *balls*, we have to take the potential energy as well as the kinetic energy into account. The potential energy would depend on the distance R between the balls, in the way the glue's elastic energy depends on the stretching or squeezing of the glue to produce a distance between the balls equal to R. Thus, the potential energy for the motion of the balls (nuclei) has to be the potential energy of the glue (electronic energy).[17]

This situation corresponds to a non-rotating system. If we admit rotation, we would have to take the effect of centrifugal force on the potential energy of the gum into account. This effect is analogous to the second term in eq. (6.25) for $V_{kJ}(R)$.

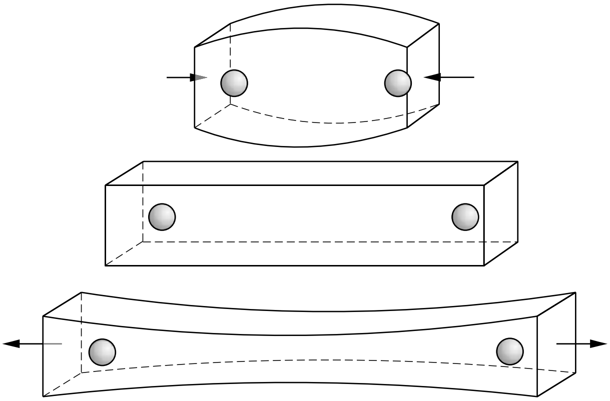


Fig. 6.1. Two metallic balls in a block of transparent glue. How they will vibrate? This will be dictated by the elastic properties of the glue.

---

[17] The adiabatic approximation is of more general importance than the separation of the electronic and nuclear motions. Its essence pertains to the problem of two coexisting time-scales in some phenomena: fast and slow scales. The examples below indicate that we have to do with an important and general philosophical approach:

- In Chapter 14 on chemical reactions, we will consider slow motion along a single coordinate, and fast motions along other coordinates (in the configurational space of the nuclei). "Vibrationally adiabatic" approximation will also be introduced, and the slow motion will proceed in the potential energy averaged over fast motions and calculated at each fixed value of the slow coordinate.
- Similar reasoning was behind vibrational analysis in systems with hydrogen bonds (Y. Marechal and A. Witkowski, *Theor. Chim. Acta* 9 (1967) 116. The authors selected a slow intermolecular motion proceeding in the potential energy averaged over fast intramolecular motions.

6.5.2 THE FUNDAMENTAL CHARACTER OF THE ADIABATIC APPROXIMATION – PES

In the case of a polyatomic molecule with N atoms ($N > 2$), V_{kJ} depends on $3N-6$ variables determining the configuration of the nuclei. The function $V_{kJ}(\boldsymbol{R})$ therefore represents a surface in $(3N-5)$-dimensional space (a *hypersurface*). This potential energy (hyper)surface $V_{kJ}(\boldsymbol{R})$, or PES, for the motion of the nuclei, represents one of the most important ideas in chemistry. *This concept makes possible contact with what chemists call the spatial "structure" of a molecule.*

potential energy surface (PES)

It is only because of the adiabatic approximation, that we may imagine the 3D shape of a molecule as a configuration of its nuclei (corresponding to a minimum of the electronic energy) bound by an electronic cloud, Fig. 6.2. This object moves and rotates in space, and in addition, the nuclei vibrate about their equilibrium positions with respect to other nuclei.

Without the adiabatic approximation, questions about the molecular 3D structure of the benzene molecule could only be answered in a very enigmatic way, e.g.,

- the molecule does not have any particular 3D shape,
- the motion of the electrons and nuclei are very complicated,
- correlations of motion of all the particles exist (electron–electron, nucleus–nucleus, electron–nucleus),
- these correlations are in general very difficult to elucidate.

Identical answers would be given, if we were to ask about the structure of the DNA molecule. Obviously, something is going wrong, we should expect more help from theory.

For the benzene molecule, we could answer the questions like: what is the mean value of the carbon–carbon, carbon–proton, proton–proton, electron–electron, electron–proton, electron–carbon distances in its ground and excited state. Note that because all identical particles are indistinguishable, for example, the carbon–proton distance pertains to any carbon and any proton, and so on. To discover that the benzene molecule is essentially a planar hexagonal object would be very difficult. What could we say about a protein? A pile of paper with such numbers would give us *the* true (non-relativistic though) picture of the benzene molecule, but it would be useless, just as a map of the world with 1:1 scale would be useless for a tourist. It is just too exact. If we relied on this, progress in the investigation of the molecular world would more or less stop. A radical approach in science, even if more rigorous, is very often less fruitful or fertile. Science needs models, simpler than reality but capturing the essence of it, which direct human thought towards much more fertile regions.

> The adiabatic approximation offers a *simple 3D model* of a molecule – *an extremely useful concept* with great interpretative potential.

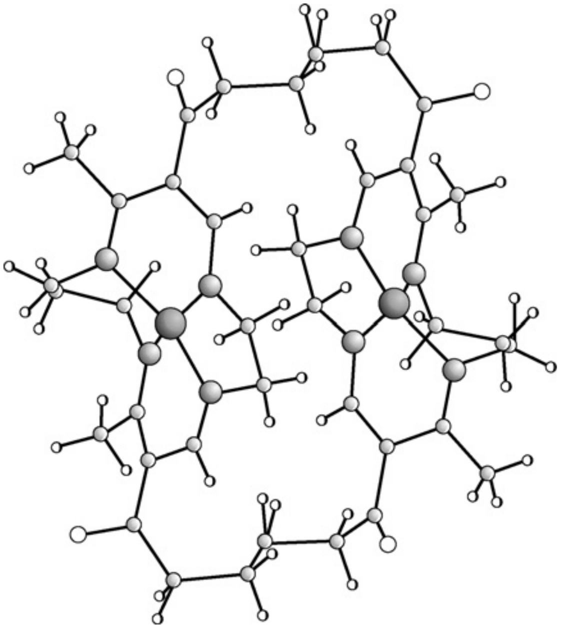


Fig. 6.2. A 3D *model* (called the "structure") of a molecule allows us to focus attention on spatial and temporal relations that are *similar* to those we know from the macroscopic world. Although the concept of "spatial structure" may occasionally fail, in virtually all cases in chemistry and physics we use a 3D molecular model which resembles what is shown in the figure for a particular molecule (using a 2D projection of the 3D model). There are "balls" and "connecting sticks". The balls represent atoms (of various sizes, the size characterizes the corresponding element), the sticks of different length are supposed to represent what are called "chemical bonds". What should be taken seriously and what not? First, the scale. The real molecule is about 100 000 000 times smaller than the picture in the figure. Second, the motion. This static model shows a kind of averaging over all the snapshots of the real oscillating atoms. In Chapters 8 and 11 we will see that indeed the atoms of which the molecule is composed keep together because of a pattern of interatomic chemical bonds (which characterizes the electronic state of the molecule) that *to some extent resemble* sticks. An atom in a molecule is never spherically symmetric (cf. Chapter 11), but can be *approximated* by its spherical core ("ball"). The particular molecule in the figure has two tetraazaanulene macrocycles that coordinate two Ni^{2+} ions (the largest spheres). The macrocycles are held together by two $-(CH_2)_4-$ molecular links. Note that any atom of a given type binds a certain number of its neighbours. The most important message is: *if such structural information offered by the 3D molecular model were not available, it would not be possible to design and carry out the complex synthesis of the molecule.* Courtesy of Professor B. Korybut-Daszkiewicz.

In the chapters to come, this model will gradually be enriched by introducing the notion of chemical bonds between *some* atoms, angles between consecutive chemical bonds, electronic lone pairs, electronic pairs that form the chemical bonds, etc. Such a model inspires our imagination.[18] This is the foundation of all chemistry, all organic syntheses, conformational analysis, most of spectroscopy etc. Without this beautiful model, progress in chemistry would be extremely difficult.

---

[18] Sometimes too much. We always have to remember that the useful model represents nothing more than a kind of better or worse pictorial representation of a more complex and unknown reality.

6.6 BASIC PRINCIPLES OF ELECTRONIC, VIBRATIONAL AND ROTATIONAL SPECTROSCOPY

6.6.1 VIBRATIONAL STRUCTURE

Eq. (6.24) represents the basis of molecular spectroscopy, and involves changing the molecular electronic, vibrational or rotational state of a diatomic molecule. Fig. 6.3 shows an example how the curves $U_k(R)$ may appear for three electronic states $k = 0, 1, 2$ of a diatomic molecule. Two of these curves ($k = 0, 2$) have a typical "hook-like" shape for bonding states, the third ($k = 1$) is also typical, but for repulsive electronic states.

It was assumed in Fig. 6.3 that $J = 0$ and therefore $V_{kJ}(R) = U_k(R)$. Next, eq. (6.24) was solved for $U_0(R)$ and a series of solutions χ_{kvJ} was found: χ_{000}, $\chi_{010}, \chi_{020}, \ldots$ with energies $E_{000}, E_{010}, E_{020}, \ldots$, respectively. Then, in a similar way, for $k = 2$, one has obtained the series of solutions: $\chi_{200}, \chi_{210}, \chi_{220}, \ldots$ with the corresponding energies $E_{200}, E_{210}, E_{220}, \ldots$ This means that these two electronic levels ($k = 0, 2$) have *a vibrational structure* ($v = 0, 1, 2, \ldots$), the corresponding vibrational levels are shown in Fig. 6.3. Any attempt to find the vibrational levels for the electronic state $k = 1$ would fail (we will return to this problem later).

vibrational structure

The pattern of the vibrational levels looks similar to those for the Morse oscillator (p. 173). The low levels are nearly equidistant, reminding us of the results

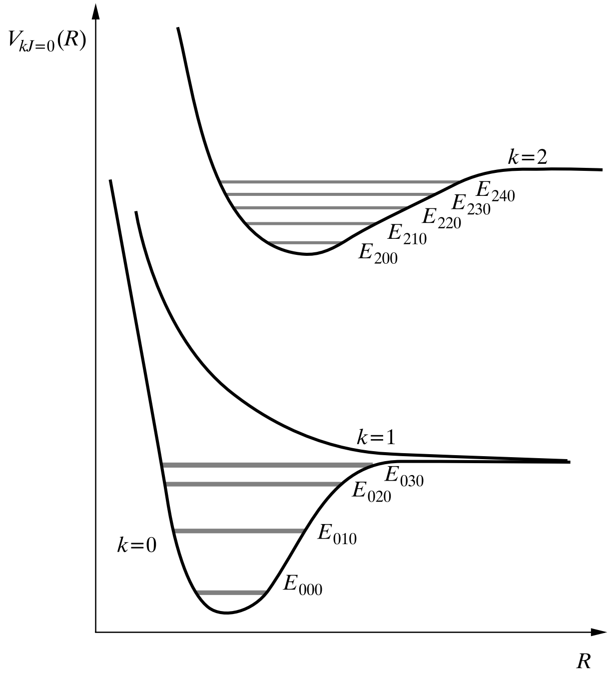


Fig. 6.3. The curves $V_{kJ}(R)$ for $J = 0$ $[V_{k0}(R) = U_k(R)]$ for the electronic states $k = 0, 1, 2$ of a diatomic molecule (scheme). The vibrational energy levels E_{kvJ} for $J = 0$ corresponding to these curves are also shown. The electronic state $k = 0$ has four, $k = 1$ has zero, and $k = 2$ has five vibrational energy levels.

for the harmonic oscillator. The corresponding wave functions also resemble those for the harmonic oscillator. Higher-energy vibrational levels are getting closer and closer, as for the Morse potential. This is a consequence of the anharmonicity of the potential – we are just approaching the dissociation limit where the $U_k(R)$ curves differ qualitatively from the harmonic potential.

6.6.2 ROTATIONAL STRUCTURE

What would happen if we took $J = 1$ instead of $J = 0$? This corresponds to the potential energy curves $V_{kJ}(R) = U_k(R) + J(J+1)\hbar^2/(2\mu R^2)$, in our case $V_{k1}(R) = U_k(R) + 1(1+1)\hbar^2/(2\mu R^2) = U_k(R) + \hbar^2/(\mu R^2)$ for $k = 0, 1, 2$. The new curves therefore represent the old curves plus the term $\hbar^2/(\mu R^2)$, which is the same for all the curves. This corresponds to a *small* modification of the curves for large R and a *larger* modification for small R. The potential energy curves just go up a little bit on the left.[19] Of course, this is why the solution of eq. (6.24) for these new curves will be similar to that which we had before, but this tiny shift upwards will result in a tiny shift upwards of all the computed vibrational levels. Therefore the levels E_{kv1} for $v = 0, 1, 2, \ldots$ will be a little higher than the corresponding E_{kv0} for $v = 0, 1, 2, \ldots$ (this pertains to $k = 0, 2$, there will be no vibrational states for $k = 1$). This means that each vibrational level v will have its own *rotational structure* corresponding to $J = 0, 1, 2, \ldots$.

rotational structure

Increasing J means that the potential energy curve becomes shallower.[20] It may happen that after a high-energy rotational excitation (to a large J) the potential energy curve will be so shallow, that no vibrational energy level will be possible. This means that the molecule will undergo dissociation due to the excessive centrifugal force. At some lower J's the molecule may accommodate all or part of the vibrational levels that exist for $J = 0$.

Example

Let us try this. An ideal experimental range for us would be a molecule with a Morse-like potential energy (p. 169), because here the problem is exactly solvable, yet preserves some important realistic features (e.g., dissociation). Unfortunately, even if we approximated $U_k(R) = E_k^0(R) + H'_{kk}(R)$ by a Morse curve, after adding the centrifugal term $J(J+1)\hbar^2/(2\mu R^2)$ the curve will no longer be of the

---

[19] With an accompanying small shift to the right the position of the minimum.

[20] It is interesting to note that the force constant corresponding to the curves with $J \neq 0$ may, in principle, increase with respect to what we had for $J = 0$. At least this is what happens when approximating the $U_k(R)$ curve by a parabola (harmonic oscillator). Then, due to the curvature $6J(J+1)\frac{\hbar^2}{2\mu R^4}$ (always positive) of the term $J(J+1)\frac{\hbar^2}{2\mu R^2}$, we have an increase of the force constant due to rotational excitation. We have no information about whether this effect applies to more realistic potentials. For sufficiently high rotational excitations, when the minimum position of $V_{kJ}(R)$ is getting larger, the force constant has to converge to zero.

Morse type. However, what we would get, would certainly resemble a Morse potential. Indeed, the resulting curve approaches zero at $R \to \infty$ (as do all the Morse curves), and at the equilibrium point, R_e will be shifted up by $J(J+1)\hbar^2/(2\mu R_e^2)$, which we may ensure by taking a Morse curve with a little lower dissociation energy $D \to D' = D - J(J+1)\hbar^2/(2\mu R_e^2)$. As seen from Fig. 4.14, we should not worry too much about what to take as the parameter α, which controls the well width. Let us keep it constant. We could take any example we want, but since in Chapter 4 we used the Morse potential for the hydrogen bond between two water molecules, we have already a feeling for what happens there, so therefore let us stick to this example.

The Morse parameters we used were: $D = 6$ kcal/mol $= 0.00956$ a.u. $= 2097$ cm^{-1}, $\mu = 16560$ a.u. and $\alpha = 1$, and we have got 18 vibrational levels for two vibrating point-like water molecules. Then we computed (p. 175) $h\nu = 0.001074$ a.u. $= 235$ cm^{-1}. Let us try something special and take a very high rotational excitation[21] $J = 40$. Let us calculate the new potential well depth D' assuming that $R_e = 5.6$ a.u., because this is what the hydrogen bond length amounts to. Then, we obtain $D' = 0.00798$ a.u., while $h\nu = 215$ cm^{-1} a bit smaller in comparison with the state with no rotational excitation ($h\nu = 235$ cm^{-1}).

Now, let us turn to the problem of the number of energy levels. The new $a' = 16.257$, and therefore the possible values of b_v are: $b_0 = 15.757, b_1 = 14.757, \ldots,$ so we have *16* allowed vibrational energy levels (and not 18 as we had before the rotational excitation). Two levels have vanished into thin air. We see that the molecule has been partially destabilized through rotational excitation. Higher excitations might result in dissociation.

Separation between energy levels

For molecules other than hydrides, the separation between rotational levels $(E_{kvJ+1} - E_{kvJ})$ is smaller by two to three orders of magnitude than the separation between vibrational levels $(E_{k,v+1,J} - E_{kvJ})$, and the later is smaller by one or two orders of magnitude when compared to the separation of the electronic levels $(E_{k+1,v,J} - E_{kvJ})$.

This is why electronic excitation corresponds to absorption of UV or visible light, vibrational excitation to absorption of infrared radiation, and a rotational excitation, to absorption of microwave radiation.

UV-VIS, IR spectra, microwave spectra

This is what we use in a microwave oven. Some chicken on a ceramic plate is irradiated by microwaves. This causes rotational excitation of the water molecules[22] always present in food. The "rotating" water molecules cause a transfer of kinetic

---

[21] At small excitations the effect is less visible.

[22] Such rotation is somewhat hindered in the solid phase.

energy to proteins, as would happen in traditional cooking. After removing the food from the microwave the chicken is hot, but the plate is cool (nothing to rotate in it).

In practice, our investigations always involve the absorption or emission spectra of a specimen from which we are trying to deduce the relative positions of the energy levels of the molecules involved. We may conclude that, in theoretical spectra computed in the centre-of-mass system, there will be allowed and forbidden energy intervals.[23] There is no energy levels in the forbidden intervals.[24] In the allowed intervals, any region corresponds to an electronic state, whose levels exhibit a pattern, i.e. clustering into vibrational series: one cluster corresponding to $v = 0$, the second to $v = 1$, etc. Within any cluster we have rotational levels corresponding to $J = 0, 1, 2, \ldots$ This follows from the fact that the distances between the levels with different k are large, with different v are smaller, and with different J are even smaller.

6.7 APPROXIMATE SEPARATION OF ROTATIONS AND VIBRATIONS

Vibrations cannot be exactly separated from rotations for a very simple reason: during vibrations the length R of the molecule changes, this makes the momentum of inertia $I = \mu R^2$ change and influences the rotation of the molecule[25] according to eq. (6.25), p. 231.

The separation is feasible only when making an approximation, e.g., when assuming the mean value of the momentum of inertia instead of the momentum itself. Such a mean value is close to $I = \mu R_e^2$, where R_e stands for the position of the minimum of the potential energy V_{k0}. So, we may decide to accept the potential (6.25) for the oscillations in the form[26]

$$V_{kJ}(R) \approx U_k(R) + J(J+1)\frac{\hbar^2}{2\mu R_e^2}.$$

Since the last term is a constant, this immediately gives the separation of the rotations from the vibrational equation (6.24)

$$\left(-\frac{\hbar^2}{2\mu}\frac{\mathrm{d}^2}{\mathrm{d}R^2} + U_k(R)\right)\chi_{kvJ}(R) = E'\chi_{kvJ}(R), \tag{6.26}$$

where the constant

$$E' = E_{kvJ} - E_{\mathrm{rot}},$$

---

[23] In a space-fixed coordinate system (see p. 971) we always have to do with a continuum of states (due to translations, see p. 61).

[24] Corresponding to bound states. The non-bound states densely fill the total energy scale above the dissociation limit of the ground state.

[25] Let us recall the energetic pirouette of charming dancer. Her graceful movements, stretching arms out or aligning them along the body, immediately translate into slow or fast rotational motion.

[26] Which looks reasonable for *small* amplitude oscillations only.

$$E_{\text{rot}} = J(J+1)\frac{\hbar^2}{2\mu R_e^2}. \tag{6.27}$$

rotational energy

Now, we may always write the potential $U_k(R)$ as a *number* $U_k(R_e)$ plus the "rest" labelled by $V_{\text{osc}}(R)$:

$$U_k(R) = U_k(R_e) + V_{\text{osc}}(R). \tag{6.28}$$

electronic energy

Then, it is appropriate to call $U_k(R_e)$ the *electronic energy* E_{el} (corresponding to the equilibrium internuclear distance in electronic state k), while the function $V_{\text{osc}}(R)$ stands, therefore, for the oscillation potential satisfying $V_{\text{osc}}(R_e) = 0$. After introducing this into eq. (6.26) we obtain the equation for oscillations (in general anharmonic)

$$\left(-\frac{\hbar^2}{2\mu}\frac{d^2}{dR^2} + V_{\text{osc}}(R)\right)\chi_{kvJ}(R) = E_{\text{osc}}\chi_{kvJ}(R),$$

vibrational energy

where the *vibrational energy* $E_{\text{osc}} = E' - E_{\text{el}}$, hence (after adding the translational energy – recalling that we have separated the centre-of-mass motion) we have the final approximation

$$\boxed{E_{kvJ} \approx E_{\text{trans}} + E_{\text{el}}(k) + E_{\text{osc}}(v) + E_{\text{rot}}(J), \tag{6.29}}$$

where the corresponding quantum numbers are given in parentheses: the electronic (k), the vibrational (v) and the rotational (J).

A quasi-harmonic approximation

The detailed form of $V_{\text{osc}}(R)$ is obtained from $U_k(R)$ of eq. (6.28) and therefore from the solution of the Schrödinger equation (6.8) with the clamped nuclei Hamiltonian. In principle there is no other way but to solve eq. (6.26) numerically. It is tempting, however, to get an idea of what would happen if a harmonic approximation were applied, i.e. when a harmonic spring was installed between both vibrating atoms. Such a model is very popular when discussing molecular vibrations. There is a unexpected complication though: such a spring cannot exist even in principle. Indeed, even if we constructed a spring that elongates according to Hooke's law, *one cannot ensure the same for shrinking*. It is true, that at the beginning, the spring may fulfil the harmonic law for shrinking too, but when $R \to 0_+$ the two nuclei just bump into each other and the energy goes to infinity instead of being parabolic. For the spring to be strictly harmonic, we have to admit $R < 0$, which is inconceivable. Fig. 6.4 shows the difference between the harmonic potential and the quasi-harmonic approximation for eq. (6.26).

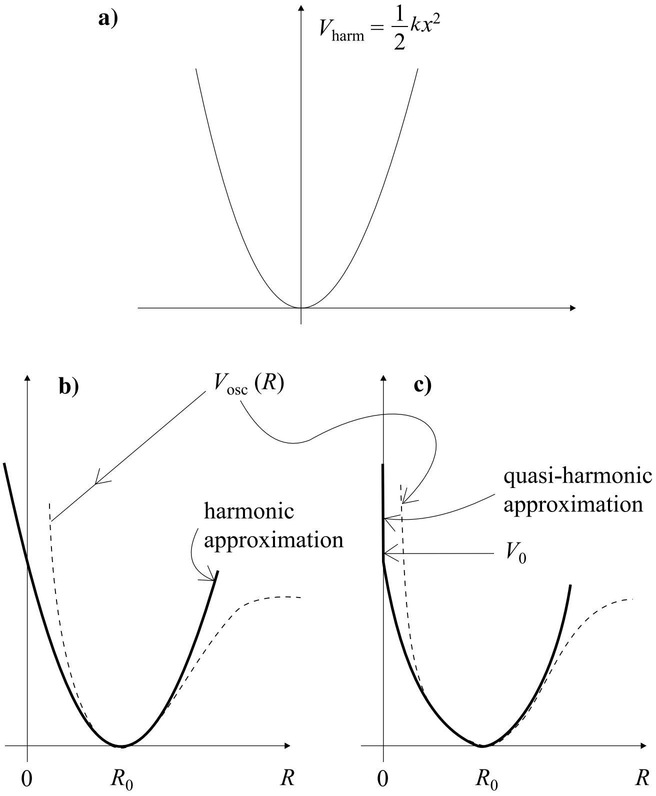


Fig. 6.4. The difference between harmonic and quasi-harmonic approximations for a diatomic molecule. (a) the potential energy for the harmonic oscillator (b) the harmonic approximation to the oscillator potential $V_{\text{osc}}(R)$ for a diatomic molecule is non-realistic, since at $R = 0$ (and at $R < 0$) the energy is finite, whereas it should go asymptotically to infinity when R tends to 0. (c) A more realistic (quasi-harmonic) approximation: the potential is harmonic up to $R = 0$, and for negative R it goes to infinity. The difference between the harmonic and quasi-harmonic approximations pertains to such high energies V_0 (high oscillation amplitudes), that practically it is of negligible importance. In cases (b) and (c), there is a range of small amplitudes where the harmonic approximation is applicable.

What do we do? Well, sticking to principles is always the best choice.[27] Yet, even in the case of the potential wall shown in Fig. 6.4c we have an analytical solution.[28] The solution is quite complex, but it gets much simpler assuming $\frac{V_0}{h\nu} \equiv \alpha \gg v$, where $v = 0, 1, 2, \ldots$ stands for the vibrational quantum number we are

---

[27] Let me stress once more that the problem appears when making the quasi-harmonic approximation, not in the real system we have.

[28] E. Merzbacher, "*Quantum Mechanics*", Wiley, New York, 2nd edition, 1970. The solution we are talking about has to be extracted from a more general problem in the reference above. The potential energy used in the reference also has its symmetric counterpart for $R < 0$. Hence, the solution needed here corresponds to the antisymmetric solutions in the more general case (only for such solutions where the wave function is equal to zero for $R = 0$).

going to consider, and $V_0 \equiv V_{\mathrm{osc}}(0)$. This means that we limit ourselves to those vibrational states that are much below V_0. This is quite satisfactory, because the hypothetical bump of the two nuclei would occur at vast (even unrealistic) V_0. In such a case the vibrational energy is equal to $E_v = h\nu(v' + \frac{1}{2})$, where the modified "quantum number" $v' = v + \varepsilon_v$ with a tiny modification

$$\varepsilon_v = \frac{1}{\sqrt{2\pi}}\frac{1}{v!}(4\alpha)^{v+\frac{1}{2}}\exp(-2\alpha).$$

The corresponding wave functions very much resemble those of the harmonic oscillator, except that for $R \leqslant 0$ they are equal to zero. The strictly harmonic approximation results in $\varepsilon_v = 0$, and therefore, $E_v = h\nu(v + \frac{1}{2})$, see Chapter 4.

Conclusion: the quasi-harmonic approximation means almost the same as the (less realistic) harmonic one.

6.8 POLYATOMIC MOLECULE

6.8.1 KINETIC ENERGY EXPRESSION

A similar procedure can be carried out for a polyatomic molecule.

Let us consider a space fixed Cartesian coordinate system (SFCS, see Appendix I on p. 971), and vector $\boldsymbol{R}_{CM}$ indicating the centre of mass of a molecule composed of M atoms, Fig. 6.5. Let us construct a Cartesian coordinate system (Body-Fixed Coordinate System, BFCS) with the origin in the centre of mass and the axes parallel to those of the SFCS (the third possibility in Appendix I).

In the BFCS an atom α of mass[29] M_α is indicated by the vector $\boldsymbol{r}_\alpha$, and, its equilibrium position[30] by $\boldsymbol{a}_\alpha$, the vector of displacement is $\boldsymbol{\xi}_\alpha = \boldsymbol{r}_\alpha - \boldsymbol{a}_\alpha$. If the molecule were *rigid and did not rotate* in the SFCS, then the velocity of the atom α would be equal to $\boldsymbol{V}_\alpha = \frac{\mathrm{d}}{\mathrm{d}t}(\boldsymbol{R}_{CM} + \boldsymbol{r}_\alpha) = \dot{\boldsymbol{R}}_{CM}$ (dots mean time derivatives), because the vector $\boldsymbol{r}_\alpha$, indicating the atom from the BFCS, would not change at all. If, in addition, the molecule, still preserving its rigidity, is *rotated* about its centre of mass with angular velocity $\boldsymbol{\omega}$ (the vector having the direction of the rotation axis, right-handed screw orientation, and length equal to the angular velocity in radians per second), then the velocity of the atom α would equal[31] $\boldsymbol{V}_\alpha = \dot{\boldsymbol{R}}_{CM} + (\boldsymbol{\omega} \times \boldsymbol{r}_\alpha)$. However, our molecule *is not rigid* and everything moves inside it (let us call these motions "vibrations"[32]). Note that *no restriction was made with respect to the*

---

[29] What this mass really means is still a unsolved problem. The essence of the problem is what electrons do when nuclei move. Besides the kinetic energy of the nuclei, we have to add the kinetic energy of the electrons that move together with the nuclei. We will leave this problem unsolved and treat M_α as the mass of the corresponding *atom*.

[30] We assume that such a position exists. If there are several equilibrium positions, we just choose one of them.

[31] $|\boldsymbol{\omega} \times \boldsymbol{r}_\alpha| = \omega r_\alpha \sin\theta$, where θ stands for the angle axis/vector $\boldsymbol{r}_\alpha$. If the atom α is on the rotation axis, this term vanishes ($\theta = 0$ or π). In other cases the rotation radius is equal to $r_\alpha \sin\theta$.

[32] Such a "vibration" may mean an oscillation of the OH bond, but also a rotation of the –CH_3 group or a large displacement of a molecular fragment.

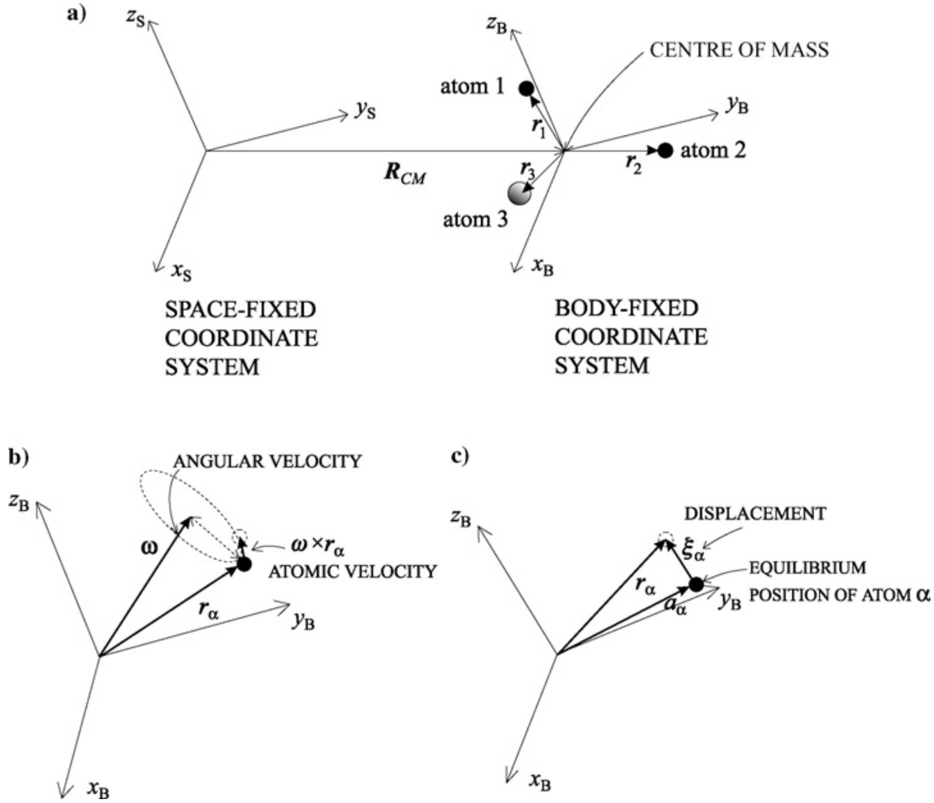


Fig. 6.5. Space- and Body-Fixed Coordinate Systems (SFCS and BFCS). (a) SFCS is a Cartesian coordinate system arbitrarily chosen in space (left). The origin of the BFCS is located in the centre of mass of the molecule (right). The centre of mass is shown by the vector $\boldsymbol{R}_{CM}$ from the SFCS. The nuclei of the atoms are indicated by vectors $\boldsymbol{r}_1, \boldsymbol{r}_2, \boldsymbol{r}_3 \ldots$ from the BFCS. Fig. (b) shows what happens to the velocity of atom α, when the system is rotating with the angular velocity given as vector $\boldsymbol{\omega}$. In such a case the atom acquires additional velocity $\boldsymbol{\omega} \times \boldsymbol{r}_\alpha$. Fig. (c) shows that if the molecule vibrates, then atomic positions $\boldsymbol{r}_\alpha$ differ from the equilibrium positions $\boldsymbol{a}_\alpha$ by the displacements $\boldsymbol{\xi}_\alpha$.

displacements $\boldsymbol{\xi}_\alpha$ – there could be some giant internal motions. Then, the velocity of the atom α with respect to the SFCS is

$$\boldsymbol{V}_\alpha = \dot{\boldsymbol{R}}_{CM} + (\boldsymbol{\omega} \times \boldsymbol{r}_\alpha) + \dot{\boldsymbol{\xi}}_\alpha. \tag{6.30}$$

When these velocities $\boldsymbol{V}_\alpha$ are inserted into the kinetic energy T of the molecule calculated in the SFCS, we get

$$T = \frac{1}{2}\sum_\alpha M_\alpha (\boldsymbol{V}_\alpha)^2 = \frac{1}{2}\big(\dot{\boldsymbol{R}}_{CM}\big)^2 \sum_\alpha M_\alpha + \frac{1}{2}\sum_\alpha M_\alpha (\boldsymbol{\omega} \times \boldsymbol{r}_\alpha)^2 + \frac{1}{2}\sum_\alpha M_\alpha \big(\dot{\boldsymbol{\xi}}_\alpha\big)^2$$
$$+ \dot{\boldsymbol{R}}_{CM} \cdot \left[\boldsymbol{\omega} \times \left(\sum_\alpha M_\alpha \boldsymbol{r}_\alpha\right)\right] + \dot{\boldsymbol{R}}_{CM} \cdot \sum_\alpha M_\alpha \dot{\boldsymbol{\xi}}_\alpha + \sum_\alpha M_\alpha (\boldsymbol{\omega} \times \boldsymbol{r}_\alpha) \cdot \dot{\boldsymbol{\xi}}_\alpha.$$

The first three ("diagonal") terms have a clear interpretation. These are: the kinetic energy of the centre of mass, the kinetic energy of rotation, and the kinetic energy of vibrations. The three further terms ("non-diagonal") denote the *roto-translational, vibro-translational and vibro-rotational couplings*.

6.8.2 SIMPLIFYING USING ECKART CONDITIONS

There is a little problem with the expression for the kinetic energy: we have a redundancy in the coordinates. Indeed, we have three coordinates for defining translation ($\boldsymbol{R}_{CM}$), three that determine rotation ($\boldsymbol{\omega}$) and on top of that M vectors $\boldsymbol{r}_\alpha$. Too many. Six are redundant. Using such coordinates would be very annoying, because we have to take into account that they are non-independent.[33]

We may impose six relations among the coordinates and in this way get rid of the redundancy. The first three relations are evident, because the origin of the BFCS is simply the centre of mass. Therefore,

$$\sum_\alpha M_\alpha \boldsymbol{r}_\alpha = \boldsymbol{0}, \tag{6.31}$$

also true when the atoms occupy equilibrium positions

$$\sum_\alpha M_\alpha \boldsymbol{a}_\alpha = \boldsymbol{0}.$$

Hence, we obtain a useful relation

$$\sum_\alpha M_\alpha (\boldsymbol{r}_\alpha - \boldsymbol{a}_\alpha) = \boldsymbol{0},$$

$$\sum_\alpha M_\alpha \boldsymbol{\xi}_\alpha = \boldsymbol{0},$$

which, after differentiation with respect to time, becomes first Eckart condition

first Eckart condition

$$\sum_\alpha M_\alpha \dot{\boldsymbol{\xi}}_\alpha = \boldsymbol{0}. \tag{6.32}$$

Inserting (6.31) *and* (6.32) *into the kinetic energy expression makes the roto-translational and vibro-translational couplings vanish.* Thus, we have

$$T = \frac{1}{2}\left(\dot{\boldsymbol{R}}_{CM}\right)^2 \sum_\alpha M_\alpha + \frac{1}{2}\sum_\alpha M_\alpha (\boldsymbol{\omega} \times \boldsymbol{r}_\alpha)^2 + \frac{1}{2}\sum_\alpha M_\alpha \left(\dot{\boldsymbol{\xi}}_\alpha\right)^2$$
$$+ \sum_\alpha M_\alpha (\boldsymbol{\omega} \times \boldsymbol{r}_\alpha) \cdot \dot{\boldsymbol{\xi}}_\alpha.$$

Noting that $\boldsymbol{r}_\alpha = \boldsymbol{a}_\alpha + \boldsymbol{\xi}_\alpha$ and using the relation[34] $(\boldsymbol{A} \times \boldsymbol{B}) \cdot \boldsymbol{C} = \boldsymbol{A} \cdot (\boldsymbol{B} \times \boldsymbol{C})$, we obtain immediately

$$T = \frac{1}{2}\left(\dot{\boldsymbol{R}}_{CM}\right)^2 \sum_\alpha M_\alpha + \frac{1}{2}\sum_\alpha M_\alpha (\boldsymbol{\omega} \times \boldsymbol{r}_\alpha)^2 + \frac{1}{2}\sum_\alpha M_\alpha \left(\dot{\boldsymbol{\xi}}_\alpha\right)^2$$
$$+ \boldsymbol{\omega} \cdot \sum_\alpha M_\alpha \left(\boldsymbol{a}_\alpha \times \dot{\boldsymbol{\xi}}_\alpha\right) + \boldsymbol{\omega} \cdot \sum_\alpha M_\alpha \left(\boldsymbol{\xi}_\alpha \times \dot{\boldsymbol{\xi}}_\alpha\right).$$

---

[33] And we would have to check all the time, whether their values are consistent.

[34] These are two ways of calculating the volume of the parallelepiped according to the formula: surface of the base times the height.

second Eckart condition

We completely get rid of the redundancy if the second *Eckart condition*[35] is imposed (equivalent to three conditions for the coordinates)

$$\sum_{\alpha} M_\alpha \big(\boldsymbol{a}_\alpha \times \dot{\boldsymbol{\xi}}_\alpha \big) = \mathbf{0}. \tag{6.33}$$

The condition means that *we do not want the internal motion to generate any angular momentum*.[36] This completes our final expression for the kinetic energy of a polyatomic molecule

$$\boxed{\begin{aligned} 2T = {} & \big(\dot{\boldsymbol{R}}_{CM}\big)^2 \sum_{\alpha} M_\alpha + \sum_{\alpha} M_\alpha (\boldsymbol{\omega} \times \boldsymbol{r}_\alpha)^2 + \sum_{\alpha} M_\alpha \big(\dot{\boldsymbol{\xi}}_\alpha\big)^2 \\ & + 2\boldsymbol{\omega} \cdot \sum_{\alpha} M_\alpha \big(\boldsymbol{\xi}_\alpha \times \dot{\boldsymbol{\xi}}_\alpha\big). \end{aligned}} \tag{6.34}$$

The kinetic energy in a space-fixed coordinate system (SFCS) is composed of:

- the kinetic energy of the centre of mass,
- the rotational energy of the whole molecule,
- the kinetic energy of the internal motions (“vibrations”),
- the last term, usually very small, is known as the Coriolis term.[37] The term couples the internal motions (“vibrations”) within the molecule with its rotation.

6.8.3 APPROXIMATION: DECOUPLING OF ROTATION AND VIBRATIONS

After the Eckart conditions are introduced, all the coordinates, i.e. the components of the vectors $\boldsymbol{R}_{CM}$, $\boldsymbol{\omega}$ and all $\boldsymbol{\xi}_\alpha$, *can be treated as independent*.

Since the Coriolis term is small, in the first approximation we may decide to neglect it. Also, when assuming *small vibrational amplitudes* $\boldsymbol{\xi}_\alpha$, which is a reasonable approximation in most cases, we may replace $\boldsymbol{r}_\alpha$ by the corresponding equilibrium positions $\boldsymbol{a}_\alpha$ in the rotational term of eq. (6.34): $\sum_\alpha M_\alpha(\boldsymbol{\omega} \times \boldsymbol{r}_\alpha)^2 \approx \sum_\alpha M_\alpha(\boldsymbol{\omega} \times \boldsymbol{a}_\alpha)^2$, which is analogous to eq. (6.27). After these two approximations have been made the kinetic energy represents the sum of the three *independent*

---

[35] Carl Eckart, professor at California Institute of Technology, contributed to the birth of quantum mechanics (e.g., C. Eckart, *Phys. Rev*. 28 (1926) 711).

[36] The problem is whether indeed we do not generate any momentum by displacing the nuclei from their equilibrium positions. A flexible molecule may have quite a number of different equilibrium positions (see Chapter 7). We cannot expect all of them to satisfy (6.33), where one of these equilibrium positions is treated as a reference. *Assuming* (6.33) means that we restrict the molecular vibrations to have only small amplitudes about a single equilibrium position.

[37] Gaspard Gustav de Coriolis (1792–1843), French engineer and mathematician, director of the Ecole Polytechnique in Paris. In 1829 Coriolis introduced the notion of work, the equivalence of work and energy, and also a coupling of rotation and vibrations.

terms (i.e. each depending on different variables)

$$T \cong \frac{1}{2}\left(\dot{\boldsymbol{R}}_{CM}\right)^2 \sum_\alpha M_\alpha + \frac{1}{2}\sum_\alpha M_\alpha(\boldsymbol{\omega} \times \boldsymbol{a}_\alpha)^2 + \frac{1}{2}\sum_\alpha M_\alpha\left(\dot{\boldsymbol{\xi}}_\alpha\right)^2, \tag{6.35}$$

the translational kinetic energy of the centre of mass $\frac{1}{2}(\dot{\boldsymbol{R}}_{CM})^2 \sum_\alpha M_\alpha$, the rotational energy $\frac{1}{2}\sum_\alpha M_\alpha(\boldsymbol{\omega} \times \boldsymbol{a}_\alpha)^2$ and the internal motion ("vibrational") kinetic energy $\frac{1}{2}\sum_\alpha M_\alpha(\dot{\boldsymbol{\xi}}_\alpha)^2$.

6.8.4 THE KINETIC ENERGY OPERATORS OF TRANSLATION, ROTATION AND VIBRATIONS

Eq. (6.35) may serve to construct the corresponding kinetic energy operator for a polyatomic molecule. There is no problem (see Chapter 1) with the translational term: $-\frac{\hbar^2}{2\sum_\alpha M_\alpha}\Delta_{\boldsymbol{R}_{CM}}$, the vibrational term will be treated in Chapter 7, p. 294.

There is a problem with the rotational term. A rigid body (the equilibrium atomic positions $\boldsymbol{a}_\alpha$ are used), e.g., the benzene molecule, rotates, but due to symmetry it may have some special axes characterizing its moment of inertia. The moment of inertia represents a tensor of rank 3 with the following components:

moment of inertia

$$\left\{\begin{array}{ccc} \sum_\alpha M_\alpha a_{x,\alpha}^2 & \sum_\alpha M_\alpha a_{x,\alpha} a_{y,\alpha} & \sum_\alpha M_\alpha a_{x,\alpha} a_{z,\alpha} \\ \sum_\alpha M_\alpha a_{x,\alpha} a_{y,\alpha} & \sum_\alpha M_\alpha a_{y,\alpha}^2 & \sum_\alpha M_\alpha a_{y,\alpha} a_{z,\alpha} \\ \sum_\alpha M_\alpha a_{x,\alpha} a_{z,\alpha} & \sum_\alpha M_\alpha a_{y,\alpha} a_{z,\alpha} & \sum_\alpha M_\alpha a_{z,\alpha}^2 \end{array}\right\},$$

to be computed in the BFCS (see Appendix I on p. 971). The diagonalization of the matrix (Appendix K on p. 982) corresponds to a certain rotation of the BFCS to a coordinate system *rotating with the molecule* (RMCS), and gives as the eigenvalues[38] I_{xx}, I_{yy}, I_{zz}. Then the classical expression for the kinetic energy of rotation takes the form[39]

spherical, symmetric, asymmetric tops

$$\frac{1}{2}\sum_\alpha M_\alpha(\boldsymbol{\omega} \times \boldsymbol{a}_\alpha)^2 = \frac{1}{2}\left(I_{xx}\omega_x^2 + I_{yy}\omega_y^2 + I_{zz}\omega_z^2\right) = \frac{J_x^2}{2I_{xx}} + \frac{J_y^2}{2I_{yy}} + \frac{J_z^2}{2I_{zz}},$$

where ω_x, ω_y, ω_z stand for the components of $\boldsymbol{\omega}$ in the RMCS, and J_x, J_y, J_z represent the components of angular momentum also computed in the RMCS. We recall from classical mechanics, that an expression for rotational motion results from the corresponding one for translational motion by replacing mass by moment of inertia, momentum by angular momentum and velocity by angular veloc-

---

[38] If $I_{xx} = I_{yy} = I_{zz}$, the body is called a spherical top (example: methane molecule); if $I_{xx} = I_{yy} \neq I_{zz}$, it is called a symmetric top (examples: benzene, ammonia molecules); if $I_{xx} \neq I_{yy} \neq I_{zz}$, then the top is asymmetric (example: water molecule).

[39] H. Goldstein, "*Classical Mechanics*", 2nd edition, Addison-Wesley, 1980.

ity. Therefore, the middle part of the above formula for kinetic energy represents an analogue of $\frac{mv^2}{2}$ and the last part is an analogue of $\frac{p^2}{2m}$.

It is not straightforward to write down the corresponding kinetic energy operator. The reason is that, in the above expression, we have curvilinear coordinates (because of the *rotation* from BFCS to RMCS[40]), whereas the quantum mechanical operators were introduced (Chapter 1) only for the Cartesian coordinates (p. 19). How do we write an operator expressed in some curvilinear coordinates q_i and the corresponding momenta p_i? Boris Podolsky solved this problem[41] and the result is:

$$\hat{T} = \frac{1}{2} g^{-\frac{1}{2}} \hat{\boldsymbol{p}}^T g^{\frac{1}{2}} \boldsymbol{G}^{-1} \hat{\boldsymbol{p}},$$

metric tensor

where $\hat{p}_i = -i\hbar \frac{\partial}{\partial q_i}$, $\boldsymbol{G}$ represents a symmetric matrix (*metric tensor*) of the elements g_{rs}, defined by the square of the length element $\mathrm{d}s^2 \equiv \sum_r \sum_s g_{rs}\, \mathrm{d}q_r\, \mathrm{d}q_s$, with $g = \det \boldsymbol{G}$ and g_{rs} being in general some functions of q_r.

6.8.5 SEPARATION OF TRANSLATIONAL, ROTATIONAL AND VIBRATIONAL MOTIONS

Eq. (6.35) represents approximate kinetic energy. To obtain the corresponding Hamiltonian we have to add the potential energy for the motion of the nuclei, U_k, to this energy where k labels the electronic state. The last energy depends uniquely on the variables $\boldsymbol{\xi}_\alpha$ that describe atomic vibrations and corresponds to the electronic energy $U_k(R)$ of eq. (6.28), except that instead of the variable R, which pertains to the oscillation, we have the components of the vectors $\boldsymbol{\xi}_\alpha$. Then, in full analogy with (6.28), we may write

$$U_k(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \ldots, \boldsymbol{\xi}_N) = U_k(\mathbf{0}, \mathbf{0}, \ldots, \mathbf{0}) + V_{k,\mathrm{osc}}(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \ldots, \boldsymbol{\xi}_N),$$

where the *number* $U_k(\mathbf{0}, \mathbf{0}, \ldots, \mathbf{0}) = E_{\mathrm{el}}$ may be called the *electronic energy in state* k, and $V_{k,\mathrm{osc}}(\mathbf{0}, \mathbf{0}, \ldots, \mathbf{0}) = 0$.

Since (after the approximations have been made) the translational, rotational and "vibrational" ("internal motion") operators depend on their own variables, after separation the total wave function represents a product of three eigenfunctions (translational, rotational and vibrational) and the total energy is the sum of the translational, rotational and vibrational energies (fully analogous with eq. (6.29))

$$\boxed{E \approx E_{\mathrm{trans}} + E_{\mathrm{el}}(k) + E_{\mathrm{rot}}(J) + E_{\mathrm{osc}}(v_1, v_2, \ldots, v_{3N-6}). \qquad (6.36)}$$

---

[40] The rotation is carried out by performing three successive rotations by what is known as Euler angles. For details see Fig. 14.3, also R.N. Zare, "*Angular Momentum*", Wiley, New York, 1988, p. 78.

[41] B. Podolsky, *Phys. Rev*. 32 (1928) 812.

6.9 NON-BOUND STATES

Repulsive potential

If we try to solve eq. (6.26) for oscillations with a repulsive potential, we would not find any solution of class Q. Among continuous, but non-square-integrable, functions we would find an infinite number of the eigenfunctions and the corresponding eigenvalues would form a continuum. These eigenvalues usually reflect the simple fact that the system has dissociated and its dissociation products may have any kinetic energy larger than the dissociation limit (i.e. having dissociated fragments with no kinetic energy), all energies measured in SFCS. Any collision of two fragments (that correspond to the repulsive electronic state) will finally result in the fragments flying off. Imagine that the two fragments are located at a distance R_0, with a corresponding total energy E, and that the system is allowed to relax according to the potential energy shown in Fig. 6.6.a. The system slides down the potential energy curve (the potential energy lowers) and, since the total energy is conserved its kinetic energy increases accordingly. Finally, the potential energy

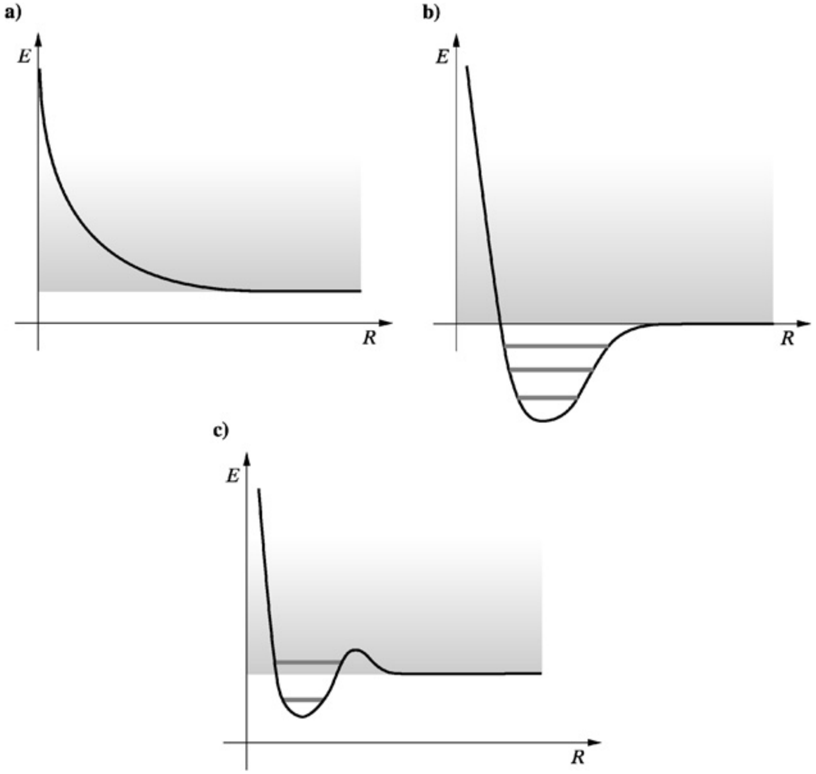


Fig. 6.6. Three different electronic states (R is the internuclear distance): (a) repulsive state (no vibrational states), (b) three bound (vibrational) states, (c) one bound vibrational state and one metastable vibrational state. A continuum of allowed states (shadowed area) is above the dissociation limit.

curve flattens, attaining $E_A + E_B$, where E_A denotes the internal energy of the fragment A (similarly for B). The final kinetic energy is equal to $E - (E_A + E_B)$ in SFCS.

"Hook-like" curves

Another typical potential energy curve is shown in Fig. 6.6.b, and has the shape of a hook. Solving (6.26) for such a curve usually[42] gives a series of bound states, i.e. with their wave functions (Fig. 6.7) concentrated in a finite region of space and exponentially vanishing on leaving it. Fig. 6.6 shows the three discrete energy levels found, and the continuum of states above the dissociation limit, similar to the curve in Fig. 6.6.a. The continuum has, in principle, the same origin as before (any kinetic energy of the fragments).

Thus, the overall picture is that a system may have some bound states, but above the dissociation limit it can also acquire any energy and the corresponding wave functions are non-normalizable (non-square-integrable).

Continuum

density of states

The continuum may have a quite complex structure. First of all, the number of states per energy unit depends, in general, on the position on the energy scale where this energy unit is located. Thus the continuum may be characterized by the *density of states* (the number of states per unit energy) as a function of energy. This may cause some confusion, because the number of continuum states in any energy section is infinite. The problem is, however, that the infinities differ, some are "more infinite than others". The continuum does not mean a banality of the states involved (Fig. 6.6.c). The continuum extends upward the dissociation limit irrespectively of what kind of potential energy curve one has for finite

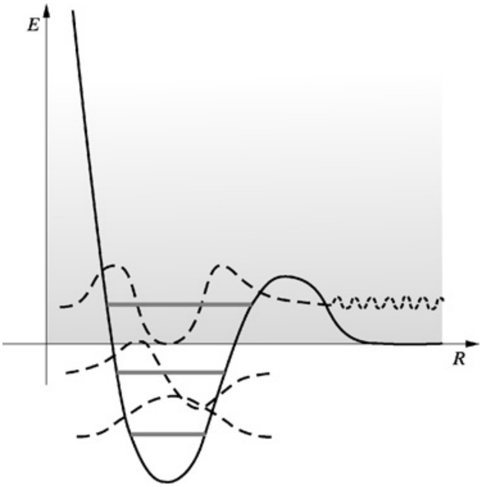


Fig. 6.7. The bound, continuum and resonance (metastable) states of an anharmonic oscillator. Two discrete bound states are shown (energy levels and wave functions) in the lower part of the figure. The continuum (shaded area) extends above the dissociation limit, i.e. the system may have any of the energies above the limit. There is one resonance state in the continuum, which corresponds to the third level in the potential energy well of the oscillator. Within the well, the wave function is very similar to the third state of the harmonic oscillator, but there are differences. One is that the function has some low-amplitude oscillations on the right-hand side. They indicate that the function is non-normalizable and that the system will sooner or later dissociate.

---

[42] For a sufficiently deep and large potential energy well.

values of R. In cases similar to that of Fig. 6.6.c the continuum will exist independently of how large and high the barrier is. But, the barrier may be so large that the system will have no idea about any "extra-barrier life", and therefore will have its "quasi-discrete" states with the energy higher than the dissociation limit. Yet, these states despite its similarity to bound states belong to the continuum (are non-normalizable). Such states are metastable and are called *resonances* (cf. p. 159), or *encounter complexes*. The system in a metastable state will sooner or later dissociate, but before this happens it may have a quite successful long life. Fig. 6.7 shows how the metastable and stationary states differ: the metastable ones do not vanish in infinity.

resonances

encounter complex

As shown in Fig. 6.8 rotational excitations may lead to a qualitative change of the potential energy curve for the motion of the nuclei. Rotational excitations lower the dissociation energy of the molecule. They may also create metastable vibrational states (vibrational resonances).

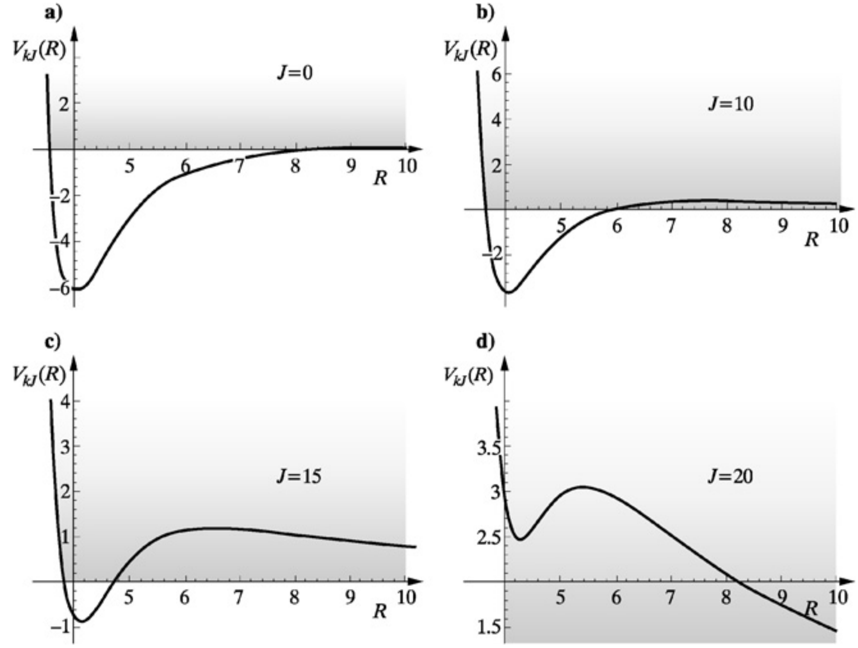


Fig. 6.8. Rotational excitation may lead to creating the resonance states. As an illustration a potential energy curve $V_{kJ}(R)$ of eq. (6.24) has been chosen that resembles what we would have for two water molecules bound by the hydrogen bond. Its first component $U_k(R)$ is taken in the form of the so called Lennard-Jones potential (cf. p. 287) $U_k(R) = \varepsilon_k[(\frac{R_{ek}}{R})^{12} - 2(\frac{R_{ek}}{R})^6]$ with the parameters for the electronic ground state ($k = 0$): $\varepsilon_0 = 6$ kcal/mol and $R_{e0} = 4$ a.u. and the corresponding reduced mass $\mu = 16560$ a.u. For $J = 0$ (a) the parameter ε_0 stands for the well depth, the R_{e0} denotes the position of the well minimum. Figs. (a), (b), (c), (d) correspond to $V_{kJ}(R) = U_k(R) + J(J+1)\hbar^2/(2\mu R^2)$ with $J = 0, 10, 15, 20$, respectively. The larger J the shallower the well: the rotation *weakens* the bond. Due to the centrifugal force a possibility of existence of the metastable resonance states appears. These are the "normal" vibrational states pushed up by the centrifugal energy beyond the energy of the dissociation limit. For $J = 20$ already *all* states (including the potential resonances) belong to the continuum.

Besides the typical continuum states that result from the fact that the dissociation products fly slower or faster, one may also have the continuum metastable or resonance states, that resemble the bound states.

The human mind wants to translate such situations into simple pictures, which help to "understand" what happens. Fig. 6.9 shows an analogy associated to astronomy: the Earth and the Moon are in a bound state, the Earth and an asteroid are in a "primitive" continuum-like state, but if it happens that an asteroid went around the Earth several times and then flew away into the Space, then one has to

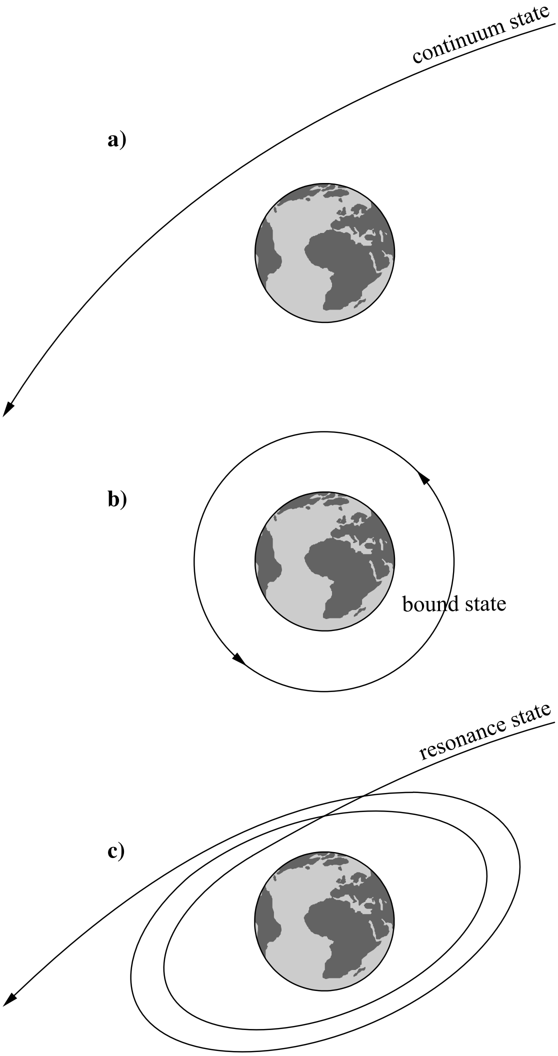


Fig. 6.9. Continuum, bound and resonance states – an analogy to the "states" of the Earth and an interacting body. (a) A "primitive" continuum state: an asteroid flies by the Earth and changes a little bit its trajectory. (b) A bound state: the Moon is orbiting around the Earth. (c) A resonance state: the asteroid was orbiting several times about the Earth and then flew away.

do with an analogue of a metastable or resonance state (characterized by a finite and non-zero life time).

The Schrödinger equation $\hat{H}\psi = E\psi$ is time-independent and, therefore, its solutions do not inform us about the sequence of events, but only all the possible events with their probability amplitudes. This is why the wave function for the metastable state of Fig. 6.7 exhibits oscillations at large x, it informs us about a possibility of dissociation.

Wave function "measurement"

Could we know a vibrational wave function in a given electronic and rotational state? It seemed that such a question could only be answered by quantum mechanical calculations. It turned out,[43] however, that the answer can also come from experiment. In this experiment three states are involved: the electronic ground state (G), an electronic excited state M, in particular its vibrational state – *this state will be measured*, and the third electronic state of a repulsive character (*REP*), see Fig. 6.10.

We excite the molecule from the ground vibrational state of G to a certain vibrational state ψ_v of M using a laser. Then the molecule undergoes a spontaneous *fluorescence transition* to *REP*. The electronic state changes so fast that the nuclei

fluorescence

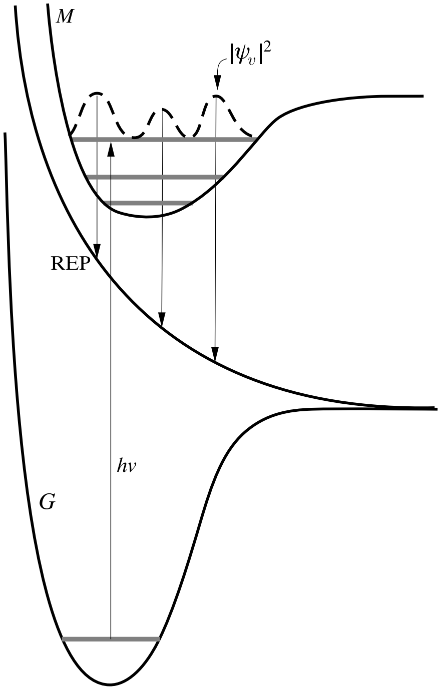


Fig. 6.10. A "measurement" of the wave function ψ_v, or more exactly of the corresponding probability density $|\psi_v|^2$. A molecule is excited from its electronic ground state G to a certain vibrational state ψ_v in the electronic excited state M. From M the molecule undergoes a fluorescence transition to the state *REP*. Since the *REP* state is of repulsive character the potential energy transforms into the kinetic energy (the total energy being preserved). By measuring the kinetic energy of the dissociation products one is able to calculate what their starting potential energy was, i.e. how high they were on the *REP* curve. This enables us to calculate $|\psi_v|^2$.

---

[43] W. Koot, P.H.P. Post, W.J. van der Zande, J. Los, *Zeit. Physik D* 10 (1988) 233. The experimental data pertain to the hydrogen molecule.

James Franck (1882–1964), German physicist, professor at the Kaiser Wilhelm Institut für Physikalische Chemie in Berlin, then at the University of Göttingen, from 1935 at the John Hopkins University in the USA, and then at the University of Chicago. Franck also participated in the Manhattan Project. As a freshman at the Department of Law at the University of Heidelberg he made the acquaintance of the student Max Born. Born persuaded him to resign from his planned career as a lawyer and choose chemistry, geology and then physics. In 1914 Franck and his colleague Gustav Hertz used electrons to bombard mercury atoms. The young researchers noted that electrons lose 4.9 eV of their kinetic energy after colliding with mercury atoms. This excess energy is then released by emitting a UV photon. This was the first experimental demonstration that atoms have the electronic energy levels foreseen by Niels Bohr. Both scientists obtained the Nobel Prize in 1925. The fact that, during the First World War, Franck was twice decorated with the Iron Cross was the reason that Franck was one of the few Jews whom the Nazis tolerated in academia.

Franck–Condon rule

have no time to move (*Franck–Condon rule*). Whatever falls (vertically, because of the Franck–Condon rule) on the *REP* state as a result of fluorescence, dissociates, because this state is repulsive. The kinetic energy of the dissociation products depends on the internuclear distance R when the fluorescence took place, i.e. on the length the system has to slide down the *REP*. How often suchan R occurs depends on $|\psi_v(R)|^2$. Therefore, investigating the kinetic energy of the dissociation products gives $|\psi_v|^2$.

Edward Condon, American physicist, one of the pioneers of quantum theory in the USA. In 1928 Condon and Gurney discovered the tunnelling effect. More widely known is his second great achievement – the Franck–Condon rule.

6.10 ADIABATIC, DIABATIC AND NON-ADIABATIC APPROACHES

Let us summarize the diabatic, adiabatic and non-adiabatic concepts, Fig. 6.11.

adiabatic states

Adiabatic case. Suppose we have a Hamiltonian $\hat{\mathcal{H}}(\boldsymbol{r};\boldsymbol{R})$ that depends on the electronic coordinates $\boldsymbol{r}$ and parametrically depends on the configuration of the nuclei $\boldsymbol{R}$. In practical applications, most often $\hat{\mathcal{H}}(\boldsymbol{r};\boldsymbol{R}) \equiv \hat{H}_0(\boldsymbol{r};\boldsymbol{R})$, the electronic clamped nuclei Hamiltonian corresponding to eq. (6.8) (generalized to polyatomic molecules). The eigenfunctions $\psi(\boldsymbol{r};\boldsymbol{R})$ and the eigenvalues $E_i(\boldsymbol{R})$ of the Hamiltonian $\hat{\mathcal{H}}(\boldsymbol{r};\boldsymbol{R})$ are called *adiabatic*, Fig. 6.11. If we take $\hat{\mathcal{H}} = \hat{H}_0(\boldsymbol{r};\boldsymbol{R})$, then in the adiabatic approximation (p. 227) the total wave function is represented by the

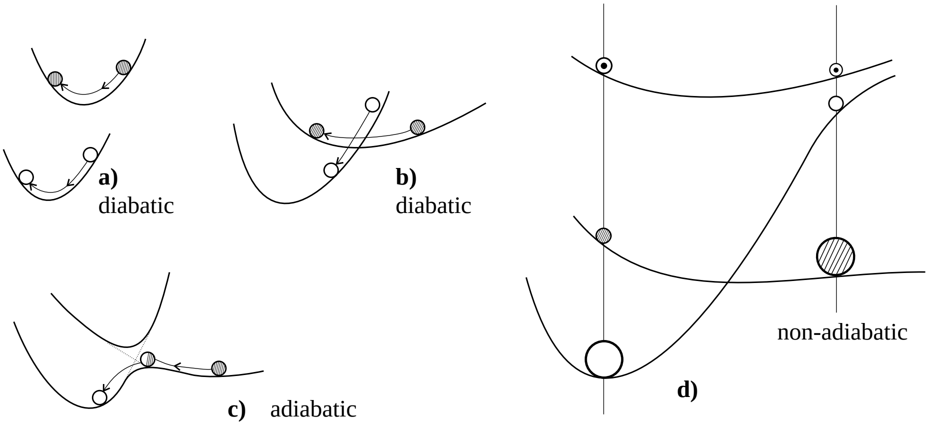


Fig. 6.11. The diabatic, adiabatic and non-adiabatic approaches to the motion of nuclei (a schematic view). (a) A state that preserves the chemical structure for any molecular geometry is called diabatic (e.g., is always ionic, or always covalent). The energies of these states are calculated as the mean values of the clamped nuclei Hamiltonian. In the lower-energy state, the system is represented by a white ball (say, the ionic state), in the second the system is represented by the black ball (say, covalent structure). These balls oscillate in the corresponding wells, preserving the chemical structure. (b) It may happen that two diabatic states cross. If the nuclear motion is fast, the electrons are unable to adjust and the nuclear motion may take place on the diabatic curves (i.e. the bond pattern does not change during this motion). Fig. (c) shows the adiabatic approach, where the diabatic states mix (mainly at a crossing region). Each of the adiabatic states is an eigenfunction of the clamped nuclei Hamiltonian, eq. (6.8). If the nuclear motion is slow, the electrons are able to adjust to it instantaneously and the system follows the lower adiabatic curve. The bond pattern changes qualitatively during this motion (black ball changes to white ball, e.g., the system undergoes a transition from covalent to ionic going through intermediate states shown as half-white and half-black ball). The total wave function is a product of the adiabatic electronic state and a rovibrational wave function. Finally, (d) pertains to the non-adiabatic approach. In this particular case, three diabatic curves come into play. The total wave function is the sum of three functions (their contributions are geometry-dependent, a larger ball means a larger contribution), each function is a product of a diabatic electronic state times a rovibrational wave function, eq. (6.7). The system is shown at two geometries. Changing the nuclear geometry, it is as if the system has moved on three diabatic surfaces at the same time. This motion is accompanied by changing the proportions (visualized by the size of the balls) of the electronic diabatic states composing it.

product

$$\Psi(\boldsymbol{r}, \boldsymbol{R}) = \psi(\boldsymbol{r}; \boldsymbol{R}) f(\boldsymbol{R}), \tag{6.37}$$

where $f(\boldsymbol{R})$ is a rovibrational wave function that describes the rotations and vibrations of the system.

diabatic states

Diabatic case. Imagine now a basis set $\bar{\psi}_i(\boldsymbol{r}; \boldsymbol{R})$, $i = 1, 2, 3, \dots, M$, of some particular electronic wave functions (we will call them *diabatic*) that also depend parametrically on $\boldsymbol{R}$. There are two reasons for considering such a basis set. The first is that we are going to solve the Schrödinger equation $\hat{\mathcal{H}}\Psi_i = E_i\Psi_i$ by using the Ritz method (Chapter 5) and we need a basis set of the expansion functions

$$\psi(\boldsymbol{r}; \boldsymbol{R}) \approx \sum_i^M c_i \bar{\psi}_i(\boldsymbol{r}; \boldsymbol{R}). \tag{6.38}$$

The second reason pertains to chemical interpretation: usually any of the diabatic wave functions are chosen as corresponding to a particular electronic distribution (chemical bond pattern) in the system,[44] and from (6.38) we may recognize what kind of chemical structure dominates ψ. Thus, using the diabatic basis, there is a chance of an insight into the chemistry going on in the system.[45]

The wave functions $\bar{\psi}_i$ are in general non-orthogonal (we assume them normalized). For each of them we may compute the mean value of the energy

$$\bar{E}_i(\boldsymbol{R}) = \langle \bar{\psi}_i \mid \hat{\mathcal{H}}(\boldsymbol{R})\bar{\psi}_i \rangle, \tag{6.39}$$

and we will call it the diabatic energy.

The key point is that we may compare the eigenvalues and eigenfunctions of $\hat{\mathcal{H}}(\boldsymbol{R})$, i.e. the adiabatic states with $\bar{E}_i$ and $\bar{\psi}_i$, respectively. If the diabatic states are chosen in a realistic way, they are supposed to be close to the adiabatic states for most configurations $\boldsymbol{R}$, Fig. 6.11.a,b,c. These relations will be discussed in a minute.

Non-adiabatic case. The diabatic states or the adiabatic states may be used to construct the basis set for the motion of the electrons *and* nuclei in the non-adiabatic approach. Such a basis function is taken as a product of the electronic (diabatic or adiabatic) wave function and of a rovibrational wave function that depends on $\boldsymbol{R}$. In a non-adiabatic approach the total wave function is a superposition of these product-like contributions [a generalization of eq. (6.7)]:

$$\Psi(\boldsymbol{r};\boldsymbol{R}) \approx \sum_k \bar{\psi}_k(\boldsymbol{r};\boldsymbol{R}) f_k(\boldsymbol{R}). \tag{6.40}$$

This sum means that in the non-adiabatic approach the motion of the system involves many potential energy surfaces at the same time, Fig. 6.11.d.

The diabatic and the adiabatic electronic states are simply two choices from the basis set in non-adiabatic calculations. If the sets were complete, the results would be identical. The first choice underlines the importance of the chemical bond pat-

---

[44] Let us take the example of the NaCl molecule: $\bar{\psi}_1$ may describe the ionic Na^+Cl^- distribution, while $\bar{\psi}_2$ may correspond to the covalent bond Na–Cl. The adiabatic wave function $\psi(\boldsymbol{r};\boldsymbol{R})$ of the NaCl molecule may be taken as a superposition of $\bar{\psi}_1$ and $\bar{\psi}_2$. The valence bond (VB) wave functions (VB structures) described in Chapter 10 may be viewed as diabatic states.

[45] Very important for chemical reactions, in which a chemical structure undergoes an abrupt change. In chemical reactions large changes of nuclear configuration are accompanied by motions of electrons, i.e. large changes in the chemical bond pattern (a qualitative change of c_i of eq. (6.38)). Such a definition leaves liberty in the choice of diabatic states. This liberty can be substantially reduced by the following. Let us take two adiabatic states that dissociate to different products, well separated on the energy scale. However, for some reason the two adiabatic energies are getting closer for some finite values of R. For each value of R we define a space spanned by the two adiabatic functions for that R. Let us find in this space two normalized functions that maximize the absolute value of the overlap integral with the two dissociation states. These two (usually non-orthogonal) states may be called diabatic.

tern and the interplay among such patterns. The second basis set highlights the order of the eigenvalues of $\hat{\mathcal{H}}(\boldsymbol{R})$ (the lower/higher-energy adiabatic state).[46]

6.11 CROSSING OF POTENTIAL ENERGY CURVES FOR DIATOMICS

6.11.1 THE NON-CROSSING RULE

Can the adiabatic curves cross when R changes?

To solve this problem in detail let us limit ourselves to the simplest situation: the two-state model (Appendix D). Let us consider a diatomic molecule and such an internuclear distance R_0 that the two electronic adiabatic states[47] $\psi_1(\boldsymbol{r}; R_0)$ and $\psi_2(\boldsymbol{r}; R_0)$) correspond to the non-degenerate (but close on the energy scale) eigenvalues of the clamped nuclei Hamiltonian $\hat{H}_0(R_0)$:

$$\hat{H}_0(R_0)\psi_i(\boldsymbol{r}; R_0) = E_i(R_0)\psi_i(\boldsymbol{r}; R_0), \quad i = 1, 2.$$

Since $\hat{H}_0$ is Hermitian and $E_1 \neq E_2$, we have the orthogonality of $\psi_1(\boldsymbol{r}; R_0)$ and $\psi_2(\boldsymbol{r}; R_0)$: $\langle\psi_1|\psi_2\rangle = 0$.

Now, we are interested in solving

$$\hat{H}_0(R)\psi(\boldsymbol{r}; R) = E\psi(\boldsymbol{r}; R)$$

for R in the vicinity of R_0 and ask, is it possible for the energy eigenvalues to cross?

The eigenfunctions of $\hat{H}_0$ will be sought as linear combinations of ψ_1 and ψ_2:

$$\psi(\boldsymbol{r}; R) = c_1(R)\psi_1(\boldsymbol{r}; R_0) + c_2(R)\psi_2(\boldsymbol{r}; R_0). \tag{6.41}$$

Note that for this distance R

$$\hat{H}_0(R) = \hat{H}_0(R_0) + V(R), \tag{6.42}$$

and $V(R)$ is small, because R is close to R_0 and $V(R_0) = 0$. Using the Ritz method (Chapter 5, Appendix D, case III), we arrive at two adiabatic solutions, and the corresponding energies are

$$E_\pm(R) = \frac{\bar{E}_1 + \bar{E}_2}{2} \pm \sqrt{\left(\frac{\bar{E}_1 - \bar{E}_2}{2}\right)^2 + |V_{12}|^2}, \tag{6.43}$$

---

[46] In polyatomic systems there is a serious problem with the adiabatic basis (this is why the diabatic functions are preferred). As we will see later (p. 264), the adiabatic electronic wave function is multi-valued, and the corresponding rovibrational wave function, having to compensate for this (because the total wave function must be single-valued), also has to be multi-valued.

[47] These states are adiabatic only for $R = R_0$, but when considering $R \neq R_0$ they may be viewed as diabatic (because they are not the eigenfunctions for that R).

where $V_{ij} \equiv \langle \psi_i | \hat{V}(R) \psi_j \rangle$ and

$$\bar{E}_i(R) = \big\langle \psi_i(\boldsymbol{r}; R_0) \big| \hat{H}_0(R) \psi_i(\boldsymbol{r}; R_0) \big\rangle = E_i(R_0) + V_{ii}(R). \tag{6.44}$$

The crossing of the energy curves at a given R means that $E_+ = E_-$, and from this it follows that the expression under the square root symbol has to equal zero. Since, however, the expression is the sum of two squares, the crossing needs *two* conditions to be satisfied simultaneously:

$$\bar{E}_1 - \bar{E}_2 = 0, \tag{6.45}$$

$$|V_{12}| = 0. \tag{6.46}$$

coupling of diabatic states

Two conditions, and a *single* changeable parameter R. If you adjust the parameter to fulfil the first condition, the second one is violated and vice versa. The crossing $E_+ = E_-$ may occur only when, for some reason, e.g., because of the symmetry, the *coupling constant* is automatically equal to zero, $|V_{12}| = 0$, for all R. Then, we have only a *single* condition to be fulfilled, and it can be satisfied by changing the parameter R, i.e. *crossing can occur*. The condition $|V_{12}| = 0$ is equivalent to

$$|H_{12}| \equiv \big\langle \psi_1 \big| \hat{H}_0(R) \psi_2 \big\rangle = 0,$$

because $\hat{H}_0(R) = \hat{H}_0(R_0) + \hat{V}$, and $[(H_0(R_0)]_{12} = 0$ due to the orthogonality of both eigenfunctions of $\hat{H}_0(R_0)$.

Now we will refer to group theory (see Appendix C, p. 903). The Hamiltonian represents a fully symmetric object, whereas the wave functions ψ_1 and ψ_2 are not necessarily fully symmetrical, because they may belong to other irreducible representations of the symmetry group. Therefore, in order to make the integral $|H_{12}| = 0$, it is sufficient that ψ_1 and ψ_2 transform according to *different* irreducible representations ("have different symmetries").[48] Thus, the adiabatic curves cannot cross if the corresponding wave functions have the same symmetry. What will happen if such curves are heading for something that looks like an inevitable crossing? Such cases are quite characteristic and look like an avoided crossing. The two curves look as if they repel each other and avoid the crossing.

If two states of a diatomic molecule have *the same* symmetry, then the corresponding potential energy curves *cannot* cross.

---

[48] H_{12} transforms according to the representation being the direct product of three irreducible representations: that of ψ_1, that of ψ_2 and that of $\hat{H}_0$ (the last is, however, fully symmetric, and therefore, does not count in this direct product). In order to have $H_{12} \neq 0$ this direct product, after decomposition into irreducible representations, *has to contain* a fully symmetric irreducible representation. This, however, is possible only when ψ_1 and ψ_2 transform according to *the same* irreducible representation.

6.11.2 SIMULATING THE HARPOONING EFFECT IN THE NaCl MOLECULE

Our goal now is to show, in an example, *what happens to adiabatic states (eigenstates of $\hat{\mathcal{H}}(R)$), if two diabatic energy curves (mean values of the Hamiltonian with the diabatic functions) do cross*. Although we are not aiming at an accurate description of the NaCl molecule (we prefer simplicity and generality), we will try to construct a toy (a model) that mimics this particular system.

The sodium atom has 11 electrons (the electronic configuration:[49] $1s^2 2s^2 2p^6 3s^1$), and the chlorine atom contains 17 electrons ($1s^2 2s^2 2p^6 3s^2 3p^5$). The solution of the Schrödinger equation for 28 electrons is difficult. But, we are not looking for trouble. Note that in the NaCl story the real film star is a single electron that goes from the sodium to the chlorine atom making Na^+ and Cl^- ions. The ions attract each other by the Coulombic force and form the familiar ionic bond. But wait a minute! There is a problem. Which is of lower energy: the two non-interacting atoms Na and Cl or the two non-interacting ions Na^+ and Cl^-? The ionization energy of sodium is $I = 495.8$ kJ/mol $= 0.1888$ a.u., whereas the electron affinity of chlorine is only $A = 349$ kJ/mol $= 0.1329$ a.u. This means that the NaCl molecule in its ground state dissociates into *atoms*, not ions.

To keep the story simple, let us *limit ourselves to the single electron* mentioned above.[50] First, let us define the two diabatic states (the basis set) of the system: the 3s orbital of Na (when the electron resides on Na; we have atoms) denoted by $|3s\rangle$ and the $3p_z$ orbital of Cl (when the electron is on Cl; we have ions, z is the axis of the molecule) $|3p\rangle$. Now, what about the Hamiltonian $\hat{\mathcal{H}}$? Well, a reasonable model Hamiltonian may be taken as[51]

$$\hat{\mathcal{H}}(\boldsymbol{r}; R) = -I|3s\rangle\langle 3s| - A|3p\rangle\langle 3p| - \frac{1}{R}|3p\rangle\langle 3p| + \exp(-R).$$

Indeed, the mean values of $\hat{\mathcal{H}}$ in the $|3s\rangle$ and $|3p\rangle$ states are equal to

$$\bar{E}_1(R) \equiv \mathcal{H}_{11} = \big\langle 3s\big|\hat{\mathcal{H}}(3s)\big\rangle = -I - AS^2 - \frac{1}{R}S^2 + \exp(-R),$$

$$\bar{E}_2(R) \equiv \mathcal{H}_{22} = \big\langle 3p\big|\hat{\mathcal{H}}(3p)\big\rangle = -IS^2 - A - \frac{1}{R} + \exp(-R),$$

where (assuming the diabatic functions to be real) the overlap integral

$$S \equiv \langle 3s|3p\rangle = \langle 3p|3s\rangle.$$

First of all, this Hamiltonian gives the correct energy limits $\bar{E}_1(R) = -I$ and

---

[49] What these configurations really mean is explained in Chapter 8.

[50] The other electrons in our approach will only influence the numerical values of the interaction parameters.

[51] $\boldsymbol{r}$ stands for the coordinates of the electron, for the diatomic molecule R replaces $\boldsymbol{R}$.

$\bar{E}_2(R) = -A$, when $R \to \infty$ (the electron binding energy by the sodium and by the chlorine atoms for dissociation into atoms and ions, respectively), which is already very important. The term $\exp(-R)$ mimics the repulsion of the inner shells of both atoms[52] and guarantees that the energies go up (they should do) as $R \to 0$. Note also that the $\bar{E}_1(R)$ and $\bar{E}_2(R)$ curves indeed mimic the approaching Na and Cl, and Na^+ and Cl^-, respectively, because in $\bar{E}_2(R)$ there is a Coulomb term $-\frac{1}{R}$, while in $\bar{E}_1(R)$ such an interaction practically disappears for large R. All this gives us a certain confidence that our Hamiltonian $\hat{\mathcal{H}}$ grasps the most important physical effects for the NaCl molecule. The resulting non-diagonal element of the Hamiltonian reads as:

$$\langle 3s|\hat{\mathcal{H}}(3p)\rangle \equiv \mathcal{H}_{12} = S\left[-I - A - \frac{1}{R} + \exp(-R)\right].$$

As to S, we could in principle calculate it by taking some approximate atomic orbitals, but our goal is less ambitious than that. Let us simply set $S = R\exp(-R/2)$. Why? Since $S = \langle 3s|3p\rangle = 0$, if $R \to \infty$ or if $R \to 0$, and $S > 0$ for other values of R, then at least our formula takes care of this. In addition, Figs. 6.12.a,b show that such a formula for S also gives a quite reasonable set of diabatic curves $\bar{E}_1(R)$ and $\bar{E}_2(R)$: both curves have single minimum, the minimum for the ionic curve is at about 5.23 a.u., close to the experimental value of 5.33 a.u., and the binding energy is 0.11 a.u. (0.13 for the adiabatic case, see below), also quite close the experimental value of 0.15 a.u. Thus, our model to a reasonable extent resembles the real NaCl molecule.

Our goal is the adiabatic energies computed using the diabatic basis chosen, eq. (6.38). Appendix D (general case) gives the eigenvalues [$E_+(R)$ and $E_-(R)$] and the eigenfunctions (ψ_+ and ψ_-). Figs. 6.12.c,d show the adiabatic compared to the diabatic curves. The avoided crossing at about 17.9 a.u. is the most important. If the two atoms begin to approach (Fig. 6.12.a, light gray) the energy does not change too much (flat energy curve), but if the ions do the same the energy goes down, because of Coulombic attraction (dark gray). Thus, the two adiabatic curves (that nearly coincide with the two diabatic curves, especially for large R) look as though they are going to cross each other (Figs. 6.12.a,b), but the two states have the same symmetry with respect to the molecular axis (note that $S \neq 0$) and, therefore, the crossing cannot occur, Fig. 6.12.d. As a result, the two curves avoid the crossing and, as shown in Fig. 6.12.c–f, the "atomic" curve switches to the "ionic" curve and *vice versa*. This switching means an electron jumping from Na to Cl and, therefore, formation of the ions Na^+ and Cl^- (then the ions approach fast – this is the *harpooning effect*, introduced to chemistry by Michael Polanyi). This jump occurs at long distances, of the order of 9 Å.

avoided crossing

harpooning effect

Is this jump inevitable?

---

[52] It prevents the two cores collapsing, cf. Chapter 13.

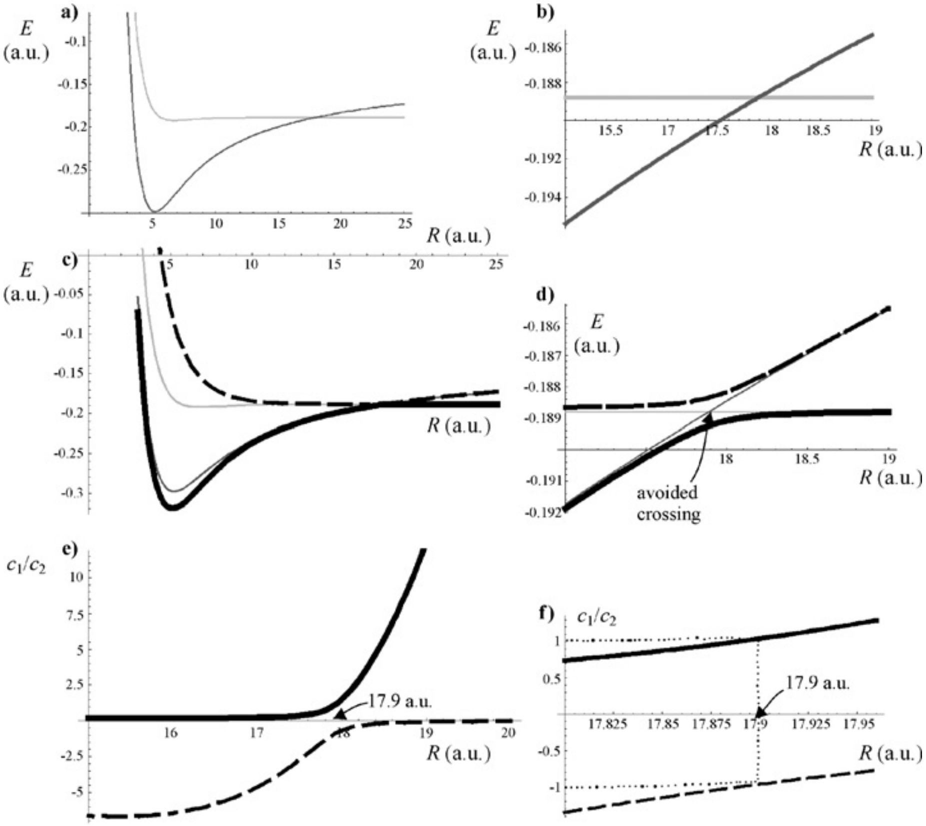


Fig. 6.12. A simple one-electron model of electron transfer in the NaCl molecule. R is the internuclear distance. (a) The mean values of the Hamiltonian with two *diabatic* states: one (light gray) being the $3s$ atomic orbital of the sodium atom (atomic curve), the second (dark gray) the $2p$ atomic orbital of the chlorine atom (ionic curve). The two diabatic curves intersect. (b) A closer view of the intersection. (c) The two diabatic curves [gray, as in (a,b)], and the two *adiabatic* curves (black): the lower-energy (solid), the higher-energy (dashed). Although the drawing looks like intersection, in fact the adiabatic curves "repel" each other, as shown in Fig. (d) (avoided crossing at 17.9 a.u.). (e) Each of the adiabatic states is a linear combination of two diabatic states (atomic and ionic) . The ratio c_1/c_2 of the coefficients for the lower-energy (solid line) and higher-energy (dashed line) states, c_1 is the contribution of the atomic function, c_2 – of the ionic function. As we can see, the lower-energy (higher-energy) adiabatic state is definitely atomic (ionic) for $R > 17.9$ a.u. and definitely ionic (atomic) for smaller R. (f) The ratio c_1/c_2 very close to the avoided crossing point. As we can see, at this point one of the adiabatic states is the sum, the other the difference of the two diabatic states.

If the electron is able to adapt instantaneously to the position of the nuclei (slow nuclear motion), the system follows the adiabatic curve and the electron jump occurs. If the nuclear motion is very fast, the system follows the diabatic curve and no electron transfer takes place. The electron transfer is more probable if the gap $2|\mathcal{H}_{12}|$ between $E_+(R)$ and $E_-(R)$ is large.

For large distances the adiabatic are practically identical with the diabatic states, except in the avoided crossing region, Figs. 6.12.c,d.

6.12 POLYATOMIC MOLECULES AND CONICAL INTERSECTION

Crossing for polyatomics

The non-crossing rule for a diatomic molecule has been based on eq. (6.43). To achieve the crossing we had to make two independent terms vanish with only one parameter (the internuclear distance R) able to vary. It is important to note that in the case of a polyatomic molecule *the formula would be the same*, but the number of parameters would be larger: $3N-6$ in a molecule with N nuclei. For $N=3$ one has already, therefore, three such parameters. No doubt even for a three-atomic molecule we would be able to make the two terms equal to zero and, therefore, achieve $E_+ = E_-$, i.e. the crossing of the two diabatic hypersurfaces.

Let us investigate this possibility, which, for reasons that will become clear later, is called conical intersection. We will approach this concept by a few steps.

Cartesian system of $3N$ coordinates ($\mathbf{O}_{3N}$)

All the quantities in eq. (6.43) depend on $n = 3N-6$ coordinates of the nuclei. These coordinates may be chosen in many different ways, the only thing we should bother about is that they have to determine the positions of N point objects. To begin, let us construct a Cartesian system of $3N$ coordinates ($\mathbf{O}_{3N}$). Let us locate (Fig. 6.13) nucleus 1 at the origin (in this way we eliminate three degrees of freedom connected with the translation of the system), nucleus 2 will occupy the point x_2 on the x axis, i.e. $y_2 = z_2 = 0$. In this way we have eliminated two rotations of the system. The total system may still be rotated about the x axis. This last possibility

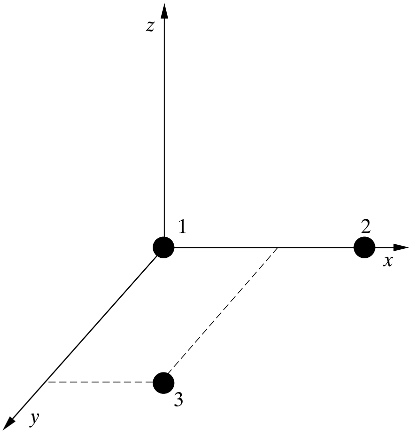


Fig. 6.13. The Cartesian coordinate system $\mathbf{O}_{3N}$ and the atoms 1, 2, 3 with their fixed positions.

can be eliminated when we decide to locate the nucleus 3 in the plane x, y (i.e. the coordinate $z_3 = 0$).

Thus six degrees of freedom have been eliminated from the $3N$ coordinates. The other nuclei may be indicated by vectors (x_i, y_i, z_i) for $i = 4, 5, \ldots, N$. As we can see there has been a lot of arbitrariness in these choices. By the way, if the molecule was diatomic, the third rotation need not be determined and the number of variables would be equal to $n = 3 \times 2 - 5 = 1$.

Cartesian system of $3N - 6$ coordinates ($\mathbf{O}_{3N-6}$)

This choice of coordinate system may be viewed a little differently. We may construct a Cartesian coordinate system with the origin at atom 1 and the axes x_2, x_3, y_3 and x_i, y_i, z_i for $i = 4, 5, \ldots, N$. Thus, we have a Cartesian coordinate system ($\mathbf{O}_{3N-6}$) with $3 + 3(N - 3) = 3N - 6 = n$ axes, which may be labelled (in the sequence given above) in a uniform way: $\bar{x}_i$, $i = 1, 2, \ldots, n$. A single point $\boldsymbol{R} = \bar{x}_1, \bar{x}_2, \ldots, \bar{x}_{3N-6})$ in this n-dimensional space determines the positions of all N nuclei of the system. If necessary all these coordinates may be expressed by the old ones, but we are not intending to make this expression.

Two special vectors in the $\mathbf{O}_{3N-6}$ space

Let us consider two functions $\bar{E}_1 - \bar{E}_2$ and V_{12} of the configuration of the nuclei $\boldsymbol{R} = (\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_{3N-6})$, i.e. with domain being the $\mathbf{O}_{3N-6}$ space. Now, let us construct two vectors in $\mathbf{O}_{3N-6}$:

$$\nabla(\bar{E}_1 - \bar{E}_2) = \sum_{i=1}^{3N-6} \boldsymbol{i}_i \frac{\partial(\bar{E}_1 - \bar{E}_2)}{\partial \bar{x}_i},$$

$$\nabla V_{12} = \sum_{i=1}^{3N-6} \boldsymbol{i}_i \frac{\partial V_{12}}{\partial \bar{x}_i},$$

where $\boldsymbol{i}_i$ labels the unit vector along axis $\bar{x}_i$.

Rotating $\mathbf{O}_{3N-6}$ to $\mathbf{O}'_{3N-6}$

We may introduce any coordinate system. We are free to do this because the system stays immobile, but our way of determining the nuclear coordinates changes. We will change the coordinate system in n-dimensional space once more. This new coordinate system is formed from the old one ($\mathbf{O}_{3N-6}$) by rotation.

> The rotation will be done in such a way as to make the plane determined by the two first axes ($\bar{x}_1$ and $\bar{x}_2$) of the old coordinate system coincide with the plane determined by the two vectors: $\nabla(\bar{E}_1 - \bar{E}_2)$ or $\nabla(V_{12})$.

Let us denote the coordinates in the rotated coordinate system by $\xi_i, i = 1, 2, \ldots, n$. The new coordinates can, of course, be expressed as some linear combinations of the old ones, but these details need not bother us. The most important thing is that we have the axes of the coordinates ξ_1 and ξ_2, which determine the same plane as the vectors $\nabla(\bar{E}_1 - \bar{E}_2)$ and ∇V_{12}. The directions $\nabla(\bar{E}_1 - \bar{E}_2)$ and ∇V_{12} need not be orthogonal, although they look this way in figures shown in the literature.[53]

Now we are all set to define the conical intersection.

6.12.1 CONICAL INTERSECTION

Why has this, a slightly weird, coordinate system been chosen? We see from the formula (6.43) for E_+ and E_- that ξ_1 and ξ_2 correspond to the fastest change of the first term and the second term under the square-root sign, respectively.[54]

> Any change of all other coordinates (being orthogonal to the plane $\xi_1\xi_2$) does not influence the value of the square root, i.e. does not change the difference between E_+ and E_- (although it changes the values of E_+ and E_-).

conical intersection subspace

Therefore, the hypersurface E_+ intersects with the hypersurface E_-, and their common part, i.e. the intersection set, are all those vectors of the n-dimensional space that fulfil the condition: $\xi_1 = 0$ and $\xi_2 = 0$. The intersection represents a $(n-2)$-dimensional subspace of the n-dimensional space of the nuclear configurations.[55] When we withdraw from the point $(0, 0, \xi_3, \xi_4, \ldots, \xi_{3N-6})$ by changing the coordinates ξ_1 and/or ξ_2, a difference between E_+ and E_- appears. For small increments $d\xi_1$ the changes in the energies E_+ and E_- are proportional to $d\xi_1$ and for E_+ and E_- differ in sign. This means that the hypersurfaces E_+ and E_- as functions of ξ_1 (at $\xi_2 = 0$ and fixed other coordinates) have the shapes shown in Fig. 6.14.a. For ξ_2 the situation is similar, but the cone may differ by its angle. From this it follows that

> two diabatic hypersurfaces intersect with each other (the intersection set represents the subspace of all vectors $(0, 0, \xi_3, \xi_4, \ldots, \xi_n)$) and split when we go out of the intersection point according to the *cone rule*, i.e. *linearly* when moving in the plane ξ_1, ξ_2 from the point $(0, 0)$.

---

[53] See: F. Bernardi, M. Olivucci, M.A. Robb, *Chem. Soc. Rev.* (1996) 321. The authors confirmed to me that the angle is often quite small.

[54] Let us take a scalar field V and calculate its value at the point $\boldsymbol{r}_0 + \boldsymbol{r}$, where we assume $|\boldsymbol{r}| \ll 1$. From the Taylor expansion we have with good accuracy, $V(\boldsymbol{r}_0 + \boldsymbol{r}) \cong V(\boldsymbol{r}_0) + (\nabla V)_{\boldsymbol{r}=\boldsymbol{r}_0} \cdot \boldsymbol{r} = V(\boldsymbol{r}_0) + |(\nabla V)_{\boldsymbol{r}=\boldsymbol{r}_0}| \cdot r \cos\theta$. We obtain the largest absolute value of the increment of V for $\theta = 0$ and $\theta = 180°$, i.e. along the vector $(\nabla V)_{\boldsymbol{r}=\boldsymbol{r}_0}$.

[55] If the axes ξ_1 and ξ_2 were chosen in another way *on the plane* determined by the vectors $\nabla(\bar{E}_1 - \bar{E}_2)$ and ∇V_{12}, the conical intersection would be described in a similar simple way. If, however, the axes were chosen outside the plane, it may happen that moving along *more than just two axes* the energy would split into E_+ and E_-. Our choice stresses that the intersection of E_+ and E_- represents a $(n-2)$-dimensional subspace.

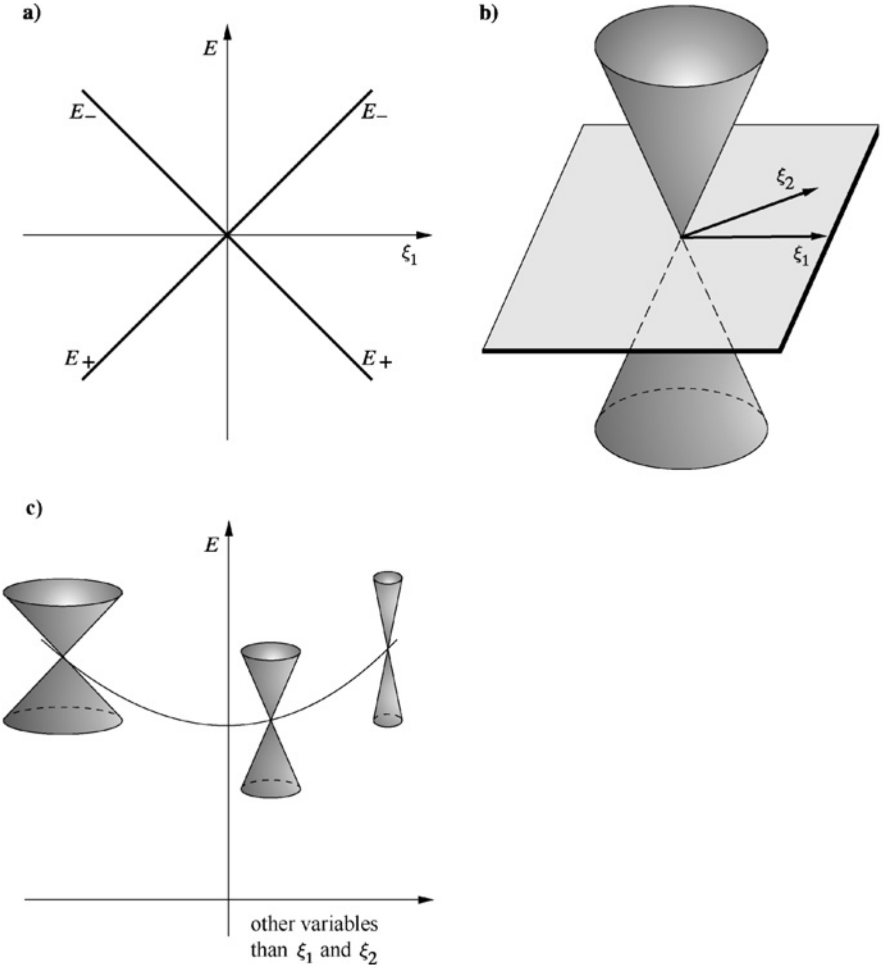


Fig. 6.14. Conical intersection: (a) a section of the cone along the ξ_1 axis; (b) the cone (variables ξ_1 and ξ_2); (c) the values of the other coordinates decide the cone opening angle as well as the intersection point energy.

This is called the *conical intersection*, Fig. 6.14.b. The cone opening angle is in general different for different values of the coordinates $\xi_3, \xi_4, \ldots, \xi_{3N-6}$ see Fig. 6.14.c.

conical intersection

The conical intersection plays a fundamental role in the theory of chemical reactions (Chapter 14). The lower (ground-state) as well as the higher (excited-state) adiabatic hypersurfaces are composed of two diabatic parts, which in polyatomics correspond to *different patterns of chemical bonds*. This means that the system, (point) when moving on the ground-state adiabatic hypersurface towards the join of the two parts, passes near the conical intersection point and overcomes the energy barrier. This is the essence of a chemical reaction.

6.12.2 BERRY PHASE

We will focus on the adiabatic wave functions close to the conical intersection. Our goal will be to show something strange, that

> when going around the conical intersection point in the configurational space, the electronic wave function changes its phase, and after coming back to the starting point this change results in the opposite sign of the function.

First let us prepare an itinerary in the configuration space around the conical intersection. We need a parameter, which will be an angle α, and will define our position during our trip around the point. Let us introduce some abbreviations in formula (6.43): $\delta \equiv \frac{\bar{E}_1-\bar{E}_2}{2}$, $h \equiv V_{12}$, and define α in the following way

$$\sin\alpha = \delta/\rho, \quad \cos\alpha = h/\rho, \quad \text{where } \rho = \sqrt{\delta^2 + h^2}.$$

We will move around the conical intersection within the plane given by the vectors $\nabla\delta$ and ∇h. The conical intersection point is defined by $|\nabla\delta| = |\nabla h| = 0$. Changing α from 0 to 2π we have to go, at a distance $\rho(\alpha)$, once through a maximum of h (say, in the direction of the maximum gradient ∇h), and once through its minimum $-h$ (the opposite direction). This is assured by $\cos\alpha = h/\rho$. Similarly, we have a single maximum and a single minimum of $\nabla\delta$ (as must be when going around), when assuming $\sin\alpha = \delta/\rho$. We do not need more information about our itinerary because what we are interested in is how the wave function changes after making a complete trip (i.e. 360°) around the conical intersection and returning to the starting point.

The adiabatic energies are given in (6.43) and the corresponding coefficients of the diabatic states are reported in Appendix D (the first, most general case):

$$\left(\frac{c_1}{c_2}\right)_{\pm} = \frac{1}{h}\left[\delta \pm \sqrt{\delta^2 + h^2}\right] = \tan\alpha \pm \frac{1}{\cos\alpha}.$$

Thus,

$$\frac{c_{1,+}}{c_{2,+}} = \frac{\sin\alpha + 1}{\cos\alpha} = \frac{(\sin\frac{\alpha}{2} + \cos\frac{\alpha}{2})^2}{\cos^2\frac{\alpha}{2} - \sin^2\frac{\alpha}{2}} = \frac{(\sin\frac{\alpha}{2} + \cos\frac{\alpha}{2})}{(\cos\frac{\alpha}{2} - \sin\frac{\alpha}{2})},$$

$$\frac{c_{1,-}}{c_{2,-}} = \frac{\sin\alpha - 1}{\cos\alpha} = \frac{-(\cos\frac{\alpha}{2} - \sin\frac{\alpha}{2})^2}{\cos^2\frac{\alpha}{2} - \sin^2\frac{\alpha}{2}} = -\frac{(\cos\frac{\alpha}{2} - \sin\frac{\alpha}{2})}{(\cos\frac{\alpha}{2} + \sin\frac{\alpha}{2})}.$$

To specify the coefficients in $\psi_+ = c_{1,+}\psi_1 + c_{2,+}\psi_2$ and $\psi_- = c_{1,-}\psi_1 + c_{2,-}\psi_2$ with ψ_1 and ψ_2 denoting the diabatic states, we have to take the two normalization conditions into account: $c_{1,+}^2 + c_{2,+}^2 = 1$, $c_{1,-}^2 + c_{2,-}^2 = 1$ and the orthogonality of

ψ_+ and ψ_- : $c_{1,+}c_{1,-} + c_{2,+}c_{2,-} = 0$. After a little algebra we get

$$c_{1,+} = \frac{1}{\sqrt{2}}\left(\cos\frac{\alpha}{2} + \sin\frac{\alpha}{2}\right),$$

$$c_{2,+} = \frac{1}{\sqrt{2}}\left(\cos\frac{\alpha}{2} - \sin\frac{\alpha}{2}\right).$$

$$c_{1,-} = -\frac{1}{\sqrt{2}}\left(\cos\frac{\alpha}{2} - \sin\frac{\alpha}{2}\right),$$

$$c_{2,-} = \frac{1}{\sqrt{2}}\left(\cos\frac{\alpha}{2} + \sin\frac{\alpha}{2}\right).$$

Now, let us make this journey by considering the wave functions ψ_+ and ψ_- at the angle α and at the angle $\alpha + 2\pi$. Note that $\cos\frac{\alpha+2\pi}{2} = \cos(\frac{\alpha}{2} + \pi) = -\cos\frac{\alpha}{2}$ and $\sin\frac{\alpha+2\pi}{2} = \sin(\frac{\alpha}{2} + \pi) = -\sin\frac{\alpha}{2}$. Therefore, both the electronic functions ψ_+ and ψ_- have to change their signs after the journey ("geometric" phase or Berry phase), i.e.

$$\psi_+(\alpha + 2\pi) = -\psi_+(\alpha) \quad \text{and} \quad \psi_-(\alpha + 2\pi) = -\psi_-(\alpha).$$

This is how the conical intersection is usually detected.

Since the total wave function has to be single-valued, this means the function that describes the motion of the nuclei (and multiplies the electronic function) has to compensate for that change, and has to undergo a change of sign.

The role of the conical intersection – non-radiative transitions and photochemical reactions

The conical intersection was underestimated in the past. However, photochemistry demonstrated that it happens much more frequently than expected.

Laser light may excite a molecule from its ground to an excited electronic state (Fig. 6.15). Let us assume that the nuclei in the electronic ground state have their positions characterized by point P′ in the configurational space (they vibrate in its neighbourhood, but let us ignore the quantum nature of these vibrations[56]).

---

[56] Electronic energy hypersurfaces represent the potential energy surface (PES) for the motion of the nuclei. In the quantum mechanical picture only some energies will be allowed: we will have the vibrational and rotational energy levels, as for diatomics. The same energy levels corresponding to E_+ may be close in the energy scale to those of E_-. Moreover, it may happen that the vibrational wave functions of two such levels may overlap significantly in space, which means that there is a significant probability that the system will undergo a transition from one to the other vibrational state. In short, in the quantum mechanical picture, the motion of the system is not necessarily bound to a single PES, but the two PESs are quite penetrable.

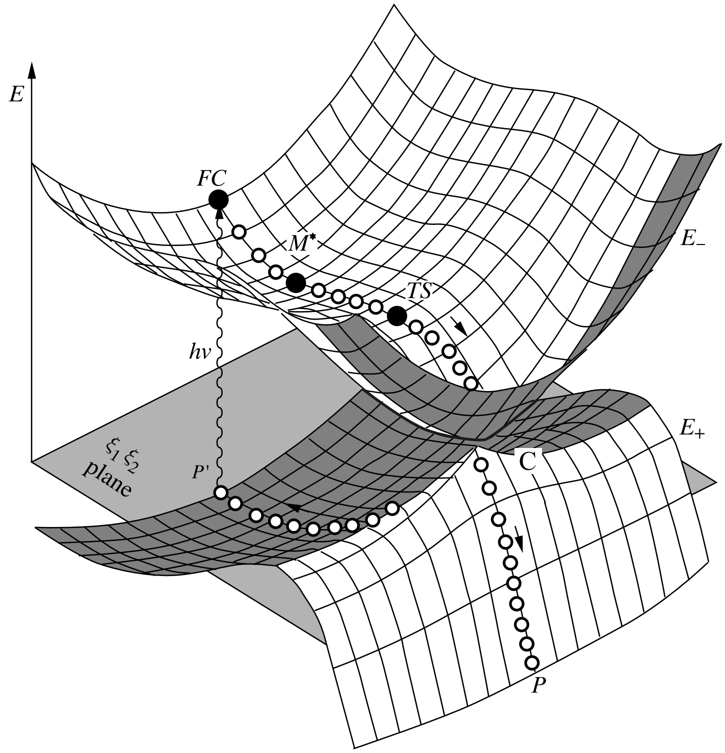


Fig. 6.15. The photochemical funnel effect. We can see two adiabatic surfaces (upper and lower), which resulted from intersection of two diabatic surfaces (white and gray). The lower surface corresponds to an electronic ground state, the upper to an excited electronic state. The molecule is excited from its ground state at the nuclear configuration P′ to the excited state (point FC) at the same nuclear configuration (according to the Franck–Condon rule). The point FC representing the system is usually located on a slope of the potential energy (corresponding to the excited state) and this is why it slides downhill towards the energy minimum M*. Its kinetic energy may be sufficient to go through M* and pass a barrier (saddle point) corresponding to the point TS. Then, the system inevitably slides down into the conical intersection point C ("funnel effect") and lands in the ground state surface (at the configuration of the conical intersection) with nearly 100% efficiency. The future of the system may correspond to different products: it may roll down to product P or slide back to product P′. Modified and adapted from F. Bernardi, M. Olivucci, M.A. Robb, *Chem. Soc. Rev*. (1996) 321.

The electronic excitation takes place so fast that the nuclei do not have enough time to move. Thus the positions of the nuclei in the excited state are identical to those in the ground state (Franck–Condon rule).

vertical transition

The point FC in Fig. 6.15 shows the very essence of the Franck–Condon rule – a *vertical transition*. The corresponding nuclear configuration may differ quite significantly from the nearest potential energy minimum M* in the excited state PES (E_-). In a few or a few tens of femtoseconds, the system slides down from P′ to the neighbourhood of M*, transforming its potential energy into kinetic energy. Usually point M* is separated from the conical intersection configuration C by a barrier with the corresponding potential energy saddle point TS ("transition state"). Behind the saddle point there is usually an energy valley[57] with a deep funnel ending in the conical intersection configuration. As soon as the system overcomes the barrier at TS, by going over it or by tunnelling, it will be sucked in by the conical intersection attractor with almost 100% probability.

funnel effect

> The system goes through the "funnel" to the electronic ground-state hypersurface.

Then the system will continue its path in the ground state PES, E_+. If its momentum is large enough, the system slides down along path P towards the nearest local minimum. If its momentum is small, the system may choose path P′. The P trajectory may mean a new reaction product, while P′ means returning to the original molecule.

Of course, the total energy has to be conserved. The non-radiative process described will take place if the system finds a way to dissipate its energy, i.e. to transfer an excess of electronic energy into the vibrational, rotational and translational degrees of freedom of its own or neighbouring molecules (e.g., of the solvent).[58]

energy dissipation

What will the energy in the plane $\xi_1\xi_2$ be, far away from the conical intersection point? Of course, there is no reason for the energy to change linearly. Instead we may expect a more complex landscape to emerge on the E_+ and E_- PESs, such as minima, saddle points, etc. shown in Fig. 6.15. We may ask whether we will find some other conical intersections in the ground-state PES. In general the answer is positive. There are at least two reasons for this.

> In the simplest case the conical intersection represents the dilemma of an atom C (approaching molecule AB): to attach either to A or B?

Thus any encounter of three atoms causes a conical intersection (see Chapter 14). In *each* case the important thing is a configuration of nuclei, where a small variation may lead to *distinct sets of chemical bonds* like in an equilateral triangle configuration of H_3. Similar "pivot points" may happen for four, five, six etc. atoms. Thus we will encounter not only the minima, maxima and saddle points, but also the conical intersection points when travelling in the ground-state PES.

The second reason is the permutational symmetry. Very often the system contains the same kinds of nuclei. Any exchange of the positions of such nuclei moves the point representing the system in configuration space to some distant regions, whereas the energy does not change at all. Therefore, any PES has to exhibit the

---

[57] On the excited state PES.

[58] The energy is usually distributed among the degrees of freedom in an unequal way.

corresponding permutational symmetry. All the details of PES will be repeated $N!$ times for a system with N identical nuclei. This will multiply the number of conical intersections.

More about conical intersection will be given in Chapter 14, when we will be equipped with the theoretical tools to describe how the electronic structure changes during chemical reactions.

6.13 BEYOND THE ADIABATIC APPROXIMATION...

6.13.1 MUON CATALYZED NUCLEAR FUSION

Andrei Dmitrievich Sakharov (1921–1989) Russian physicist, father of the Soviet hydrogen bomb. During the final celebration of the H bomb project Sakharov expressed his hope that the bombs would never be used. A Soviet general answered coldly that it was not the scientists' business to decide such things. This was a turning point for Sakharov and he began his fight against the totalitarian system.

The idea of muon induced fusion was conceived by Sakharov in 1945, in his first scientific paper, under the supervision of Tamm. In 1957 David Jackson realized that muons may serve as catalysts.

Some molecules look really peculiar, they may contain a muon instead of an electron. The muon is an unstable particle with the charge of an electron and mass equal to 207 electron masses.[59] For such a mass, assuming that nuclei are infinitely heavier than a muon looks like a very bad approximation. Therefore, the calculations need to be non-adiabatic. The first computations for muonic molecules were performed by Kołos, Roothaan and Sack in 1960.[60] The idea behind the project was muon catalyzed fusion of deuterium and tritium. This fascinating problem was proposed by the Russian physicist Andrei Sakharov. Its essence is as follows.

If an electron in the molecule dt^+ is replaced by a muon, immediately the dimension of the molecule decreases by a factor of about 200. How is this possible?

The radius of the first Bohr orbit in the hydrogen atom (see p. 179) is equal to $a_0 = \frac{\hbar^2}{\mu e^2}$. After introducing atomic units, this formula becomes $a_0 = \frac{1}{\mu}$, and when we take into account that the reduced mass $\mu \approx m$ (m stands for the electron mass) we get $a_0 \approx 1$. This approximation works for the electron, because in reality $\mu = 0.9995m$. If, in the hydrogen atom, instead an electron we have a muon, then μ would be equal about $250m$. This, however, means that such a "muon Bohr radius" would be about 250 times smaller. Nuclear forces begin to operate at such a small internuclear separation (strong interactions, Fig. 6.16.a), and are able to overcome

muon catalysis

---

[59] The muon was discovered in 1937 by C.D. Anderson and S.H. Neddermeyer. Its life time is about $2.2 \cdot 10^{-6}$ s. The muons belong to the leptons family (with the electron and τ particle, the later with mass equal to about 3640 electron masses). Nature created, for some unknown reasons, the "more massive electrons". When the nuclear physicist Isidor Rabi was told about the incredible mass of the τ particle, he dramatically shouted: "*Who ordered that?!*"

[60] W. Kołos, C.C.J. Roothaan, R.A. Sack, *Rev. Mod. Phys.* 32 (1960) 205.

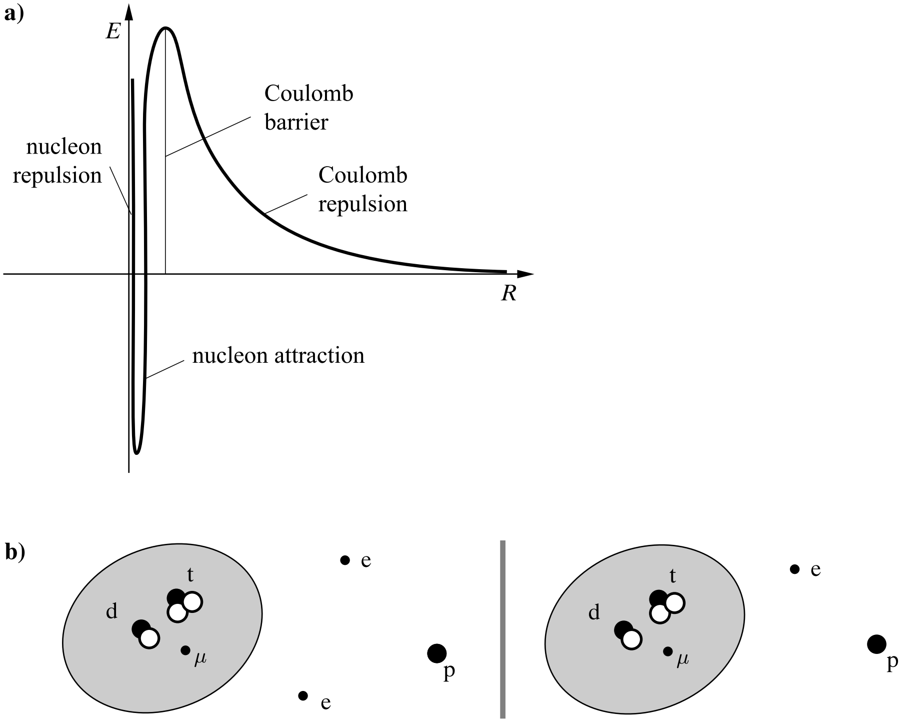


Fig. 6.16. (a) The interaction energy potential (E) of d and t as a function of the interparticle distance (R), with taking the nuclear forces into account (an outline). At large R, of the order of nanometers, we have Coulombic repulsion, at distances of the order of femtometers the internuclear attractive forces (called the strong interaction) are switched on and overcome the Coulombic repulsion. At a distance of a fraction of femtometer again we have a repulsion (b) "Russian dolls" (outline): the analogues of H_2 and H_2^+.

the Coulombic barrier and stick the nuclei together by *nuclear fusion*. The muon, however, is released, and may serve as a catalyst in the next nuclear reaction.

Deuteron and tritium bound together represent a helium nucleus. One muon may participate in about 200–300 such muon catalyzed fusion processes.[61] Everybody knows how much effort and money has been spent for decades (for the moment without success) to ignite the nuclear synthesis $\mathrm{d} + \mathrm{t} \rightarrow \mathrm{He}$. Muon catalyzed fusion might be an alternative solution. If the muon project were successful, humanity would have access to a practically unlimited source of energy. Unfortunately, theoretical investigations suggest that the experimental yield already achieved is about the maximum theoretical value.[62]

---

[61] The commercial viability of this process will not be an option unless we can demonstrate 900 fusion events for each muon. About 10 grams of deuterium and 15 g of tritium fusion would then be sufficient to supply the average person with electricity for life.

[62] More about this may be found in K. Szalewicz, S. Alexander, P. Froelich, S. Haywood, B. Jeziorski, W. Kołos, H.J. Monkhorst, A. Scrinzi, C. Stodden, A. Velenik, X. Zhao, in "*Muon Catalyzed Fusion*", eds. S.E. Jones, J. Rafelski, H.J. Monkhorst, AIP Conference Proceedings 181 (1989) 254.

6.13.2 "RUSSIAN DOLLS" – OR A MOLECULE WITHIN MOLECULE

Scrinzi and Szalewicz[63] carried out non-adiabatic calculations (p. 224) for a system of 6 particles: proton (p), deuterium (d), tritium (t), muon (μ) and two electrons (e) interacting by Coulombic forces (i.e. no nuclear forces assumed). It is not so easy to predict the structure of the system. It turns out that the resulting structure is a kind of "Russian doll"[64] (Fig. 6.16.b): the muon has acted according to its mass (see above) and created tdμ with a dimension of about 0.02 Å. This system may be viewed as a partly split nucleus of charge $+1$ or, alternatively, as a mini model of the hydrogen molecular ion (scaled at 1:200). The "nucleus" serves as a partner to the proton and both create a system similar to the hydrogen molecule, in which the two electrons play their usual binding role, and the internuclear distance is about 0.7 Å. It turns out that the non-zero dimension of the "nucleus" makes a difference, and the energies computed with and without an approximation of the point-like nucleus differ. The difference is tiny (about 0.20 meV), but it is there.

It is quite remarkable that such small effects are responsible for the fate of the total system. The authors observe that the relaxation of the "nucleus" dtμ (from the excited state to the ground state[65]) causes the ionization of the system: one of the electrons flies off. Such an effect may excite those who study this phenomenon. How is it possible? The "nucleus" is terribly small when seen by an electron orbiting far away. How could the electron detect that the nucleus has changed its state and that it has no future in the molecule? Here, however, our intuition fails. For the electron, the most frequently visited regions of the molecule are the nuclei. We will see this in Chapter 8, but even the 1s state of the hydrogen atom (p. 178, the maximum of the orbital is at the nucleus) suggests the same. Therefore, no wonder the electron *could* recognize that something has abruptly changed on one of the nuclei and (being already excited) it felt it was receiving much more freedom, so much that it could leave the molecule completely.

We may pose an interesting question, whether the "Russian doll" represents the global minimum of the particles system. We may imagine that the proton changes its position with the deuterium or tritium, i.e. new isomers (isotopomers[66]) appear.

---

[63] A. Scrinzi, K. Szalewicz, *Phys. Rev*. A 39 (1989) 4983.

[64] (((woman@ woman)@ woman)@)

[65] A. Scrinzi, K. Szalewicz, *Phys. Rev. A* 39 (1989) 2855. The dtμ ion is created in the rovibrational state $J = 1$, $v = 1$, and then the system spontaneously goes to the lower energy 01 or 00 states. The energy excess causes one electron to leave the system (ionization). This is an analogue of the Auger effect in spectroscopy.

[66] The situation is quite typical, although we very rarely think this way. Some people say that they observe *two different systems*, whereas others say, that they see *two states of the same system*. This begins with the hydrogen atom – it looks different in its $1s$ and $3p_z$ states. We can easily distinguish two different conformations of cyclohexane, two isomers of butane, and some chemists would say these are different substances. Going much further, N_2 and CO represent two different molecules, or is one of them nothing but an excited state of the other? However strange it may sound for a chemist, N_2 represents an excited state of CO, because we may imagine a nuclear reaction of the displacement of a proton from one nitrogen to the other (and the energy curve per nucleon as a function of the atomic

The authors did not study this question,[67] but investigated a substitution of the proton by deuterium and tritium (and obtained similar results).

Scrinzi and Szalewicz also performed some calculations for an analogue of H_2^+: proton, deuterium, tritium, muon and electron. Here the "Russian doll" looks wonderful (Fig. 6.16.c); it is a four-level object:

- the molecular ion (the analogue of H_2^+) is composed of *three* objects: the proton, the "split nucleus" of charge $+1$ and the electron;
- the "split nucleus" is also composed of *three* objects: d, t, μ (a *mini model of* H_2^+);
- the tritium is composed of *three* nucleons: the proton and the two neutrons;
- each of the nucleons is composed of *three* quarks (called the valence quarks).

Summary

- In the adiabatic approximation, the total wave function is approximated as a *product* $\Psi = \psi_k(\boldsymbol{r}; R) f_k(\boldsymbol{R})$ of the function $f_k(\boldsymbol{R})$, which describes the motion of the nuclei (vibrations and rotations) and the function $\psi_k(\boldsymbol{r}; R)$ that pertains to the motion of electrons (and depends parametrically on the configuration of the nuclei; here we give the formulae for a diatomic molecule). This approximation relies on the fact that the nuclei are thousands of times heavier than the electrons.
- The function $\psi_k(\boldsymbol{r}; R)$ represents an eigenfunction of the Hamiltonian $\hat{H}_0(R)$ of eq. (6.4), i.e. the Hamiltonian $\hat{H}$, in which the kinetic energy operator for the nuclei is assumed to be zero (the *clamped nuclei Hamiltonian*).
- The function $f_k(\boldsymbol{R})$ is a *product* of a spherical harmonic[68] Y_J^M that describes the *rotations* of the molecule (J and M stand for the corresponding quantum numbers) and a function that describes the *vibrations* of the nuclei.
- The diagram of the energy levels shown in Fig. 6.3 represents the basis of molecular spectroscopy. The diagram may be summarized in the following way:
  - the energy levels form some series separated by energy gaps, with no discrete levels. Each series corresponds to a single electronic state n, and the individual levels pertain to various vibrational and rotational states of the molecule in electronic state n;
  - within the series for a given electronic state, there are groups of energy levels, each group characterized by a distinct vibrational quantum number ($v = 0, 1, 2, \ldots$), and within the group the states of higher and higher energies correspond to the increasing rotational quantum number J;
  - the energy levels fulfil some general relations:
    - ∗ increasing n corresponds to an electronic excitation of the molecule (UV-VIS, *ultraviolet and visible spectrum*),
    - ∗ increasing v pertains to a vibrational excitation of the molecule, and requires the energy to be *smaller by one or two orders of magnitude* than an electronic excitation (IR, infrared spectrum).
    - ∗ increasing J is associated with energy *smaller by one or two orders of magnitude* than a vibrational excitation (microwaves).

---

mass is convex). Such a point of view is better for viewing each object as a "new animal": it enables us to see and use some relations among these animals.

[67] They focused their attention on tdμ.

[68] That is, of the eigenfunction for the rigid rotator.

- The electronic wave functions $\psi_k(\boldsymbol{r}; R)$ correspond to the energy eigenstates $E_k^0(R)$, which are functions of R. The energy curves[69] $E_k^0(R)$ for different electronic states k may cross each other, unless the molecule is diatomic and the two electronic states have the same symmetry.[70] In such a case we have what is known as an avoided crossing (see Fig. 6.12).
- For polyatomic molecules the energy hypersurfaces $E_k^0(\boldsymbol{R})$ can cross. The most important is *conical intersection* (Fig. 6.15) of the two (I and II) diabatic hypersurfaces, i.e. those that (each individually) preserve a given pattern of chemical bonds. This intersection results in two adiabatic hypersurfaces ("lower and upper"). Each of the adiabatic hypersurfaces consists of two parts: one belonging to I and the second to II. Using a suitable coordinate system in the configurational space, we obtain, *independence of the adiabatic hypersurface splitting of* $3N - 8$ *coordinates and dependence on two coordinates* (ξ_1 *and* ξ_2) *only*. The splitting begins by a linear dependence on ξ_1 and ξ_2, which gives a sort of cone (hence the name "conical intersection").
- Conical intersection plays a prominent role in the photochemical reactions, because the excited molecule slides down the upper adiabatic hypersurface to the funnel (just the conical intersection point) and then, with a yield close to 100% lands on the lower adiabatic hypersurface (assuming there is a mechanism for dissipation of the excess energy).

Main concepts, new terms

clamped nuclei Hamiltonian (p. 223)
non-adiabatic theory (p. 224)
adiabatic approximation (p. 227)
diagonal correction for the motion of the nuclei (p. 227)
Born–Oppenheimer approximation (p. 229)
potential energy curve (p. 231)
potential energy (hyper)surface (p. 233)
electronic-vibrational-rotational spectroscopy (p. 235)
non-bound states (p. 247)
non-bound metastable states (p. 247)
wave function "measurement" (p. 251)
diabatic curve (p. 253)
adiabatic curve (p. 253)
avoided crossing (p. 255)
non-crossing rule (p. 256)
harpooning effect (p. 257)
conical intersection (p. 262)
Berry phase (p. 264)
Franck–Condon rule (p. 266)
funnel effect (p. 266)
non-radiative transitions (p. 266)
photochemical reaction (p. 266)
muon catalyzed nuclear fusion (p. 268)
split nucleus effect (p. 270)

From the research front

Computing a reliable hypersurface of the potential energy (PES) for the motion of nuclei, represents an extremely difficult task for today's computers, even for systems of four atoms. In principle routine calculations are currently performed for three-atomic (and, of course, two-atomic) systems. The technical possibilities are discussed by J. Hinze, A. Alijah and L. Wolniewicz, *Pol. J. Chem.* 72 (1998) 1293, in which the most accurate calculations are also reported (for the H_3^+ system). Analysis of conical intersections is only occasionally carried out, because the problem pertains to mostly unexplored electronic excited states.

---

[69] As functions of R.

[70] That is, they transform according to the same irreducible representation.

Ad futurum...

The computational effort needed to calculate the PES for an N atomic molecule is proportional to 10^{3N-6}. This strong dependence suggests that, for the next 20 years, it would be rather unrealistic to expect high-quality PES computations for $N > 10$. However, experimental chemistry offers high-precision results for molecules with hundreds of atoms. It seems inevitable that it will be possible to freeze the coordinates of many atoms. There are good reasons for such an approach, because indeed most atoms play the role of spectators in chemical processes. It may be that limiting ourselves to, say, 10 atoms will make the computation of rovibrational spectra feasible.

Additional literature

J. Hinze, A. Alijah, L. Wolniewicz, "Understanding the Adiabatic Approximation; the Accurate Data of H_2 Transferred to H_3^+", *Pol. J. Chem.* 72 (1998) 1293.

The paper reports the derivation of the equation of motion for a *polyatomic* molecule. As the origin of the BFCS, unlike this chapter, the centre of mass was chosen.[71]

W. Kołos, "Adiabatic Approximation and its Accuracy", *Advan. Quantum Chem.* 5 (1970) 99.

Kołos was the No 1 expert in the domain.

F. Bernardi, M. Olivucci, M.A. Robb, "Potential Energy Surface Crossings in Organic Photochemistry", *Chem. Soc. Rev.* p. 321–328 (1996).

A review article by the top experts in conical intersection problems.

Questions

1. A diatomic homonuclear molecule, origin of the BFCS in the centre of the molecule, potential energy of the Coulombic interactions equals V. The total non-relativistic Hamiltonian is equal to:
   a) $\hat{H} = -\sum_i \frac{\hbar^2}{2m}\Delta_i - \frac{\hbar^2}{2\mu}\Delta_R$; b) $\hat{H} = -\sum_i \frac{\hbar^2}{2m}\Delta_i + V$; c) $\hat{H} = -\frac{\hbar^2}{2\mu}\Delta_R + V$; d) $\hat{H} = -\sum_i \frac{\hbar^2}{2m}\Delta_i + V - \frac{\hbar^2}{2\mu}\Delta_R - \frac{\hbar^2}{8\mu}(\sum_i \nabla_i)^2$.
2. A diatomic molecule, origin of the BFCS in the centre of the molecule. In the adiabatic approximation, the total wave function is in the form $\Psi = \psi_k(\boldsymbol{r}; R) f_k(\boldsymbol{R})$, where:
   a) $f_k(\boldsymbol{R})$ describes the translation of the molecule; b) $f_k(\boldsymbol{R})$ stands for a spherical harmonic describing the rotations of the molecule; c) $\psi_k(\boldsymbol{r}; R)$ denotes the eigenfunction of the clamped nuclei Hamiltonian; d) $\psi_k(\boldsymbol{r}; R)$ stands for the probability density of having the electrons with coordinates $\boldsymbol{r}$ and the nuclei at distance R.
3. A diatomic molecule, origin of the BFCS in the centre of the molecule, in the adiabatic approximation the total wave function is in the form $\Psi = \psi_k(\boldsymbol{r}; R) f_k(\boldsymbol{R})$. The potential energy for the vibrations of the molecule is equal to:
   a) $V + J(J+1)\frac{\hbar^2}{2\mu R^2}$; b) $\langle\psi_k|\hat{H}\psi_k\rangle + (2J+1)\frac{\hbar^2}{2\mu R^2}$; c) $\langle\psi_k|\hat{H}\psi_k\rangle + J(J+1)\frac{\hbar^2}{2\mu}$; d) $\langle\psi_k|\hat{H}\psi_k\rangle + J(J+1)\frac{\hbar^2}{2\mu R^2}$.

---

[71] We have chosen the centre of the *ab* bond.

4. The potential energy curves for the motion of the nuclei for a diatomic molecule
a) cross, if their derivatives differ widely; b) always cross; c) do not cross, if they correspond to the same irreducible representation of the symmetry group of the Hamiltonian; d) do not cross, if they correspond to different symmetry.

5. Please choose the *wrong* answer. The potential energy for the motion of the nuclei:
a) contains the eigenvalue of the clamped nuclei Hamiltonian;
b) does not change when the rotation excitation occurs;
c) represents the electronic energy when the Born–Oppenheimer approximation is used;
d) for bound states has to be convex as a function of the positions of the nuclei.

6. Please choose the *wrong* answer. As a result of the rotational excitation $J \to (J+1)$ of a molecule of length R:
a) the angular momentum increases;
b) the potential for vibrations changes;
c) the potential energy curve for the motion of the nuclei becomes shallower;
d) the potential energy increases by a term proportional to $(2J+1)$ and proportional to R^{-2}.

7. The potential energy hypersurface for the N-atomic molecule depends on the following number of variables:
a) $2N-8$; b) $3N-6$; c) $3N-5$; d) N.

8. At the conical intersection (Born–Oppenheimer approximation), the cone angle:
a) does not depend on the direction of leaving the conical intersection point; b) is different for the lower and for the higher cones; c) depends on the values of coordinates other than those along directions $\nabla(\bar{E}_1 - \bar{E}_2)$ and $\nabla(V_{12})$; d) is different for different isotopomers.

9. At the conical intersection the following directions in configurational space lead to splitting between E_+ and E_-:
a) $\nabla(\bar{E}_1 - \bar{E}_2)$ and $\nabla(V_{12})$; b) $\nabla(\bar{E}_1)$ and $\nabla(\bar{E}_2)$; c) $\nabla(\bar{E}_1 \cdot \bar{E}_2)$ and $\nabla(V_{12})$; d) $\nabla(\bar{E}_1 + \bar{E}_2)$ and $\nabla(V_{12})$.

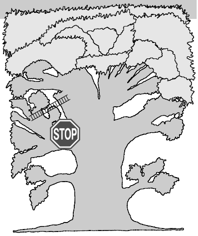
10. Please find the *wrong* answer. The adiabatic approximation:
a) is equivalent to the Born–Oppenheimer approximation; b) is related to the wave function in the form of a product of an electronic function and a function describing the motion of the nuclei; c) leads to the notion of the potential energy curve for the motion of the nuclei; d) is worse satisfied for molecules with muons instead of electrons.

Answers

1d, 2c, 3d, 4c, 5b, 6d, 7b, 8c, 9a, 10a

Chapter 7

MOTION OF NUCLEI



WHERE ARE WE?

Where are we?

We are on the most important side branch of the TREE.

An example

Which of conformations (Fig. 7.1) is more stable: the “boat” or “chair” of cyclohexane C_6H_{12}? How do particular conformations look in detail (symmetry, interatomic distances, bond angles), when the electronic energy as a function of the positions of the nuclei attains a minimum value? What will be the most stable conformation of the trimer: C_6H_{11}–$(CH_2)_3$–C_6H_{10}–$(CH_2)_3$–C_6H_{11}?

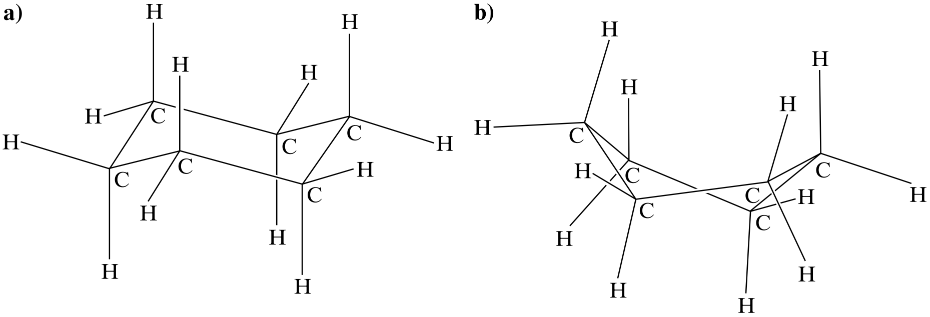


Fig. 7.1. The chair (a) and boat (b) conformations of cyclohexane. These geometries (obtained from arbitrary starting conformations) are optimized in the force field, which we will define in the present chapter. The force field indicates, in accordance with experimental results, that the chair conformation is the more stable (by about 5.9 kcal/mol). Thus we obtain all the details of the atomic positions (bond lengths, bond angles, etc.). Note that the chair conformation obtained exhibits D_{3d} symmetry, while the boat conformation corresponds to D_2 (the boat has somewhat warped planks, because of repulsion of the two upper hydrogen atoms).

What is it all about

As shown in Chapter 6, the solution of the Schrödinger equation in the adiabatic approximation can be divided into two tasks: the problem of electronic motion in the field of the clamped nuclei (this will be the subject of the next chapters) and the problem of *nuclear motion in the potential energy determined by the electronic energy*. The ground-state electronic energy $E_k^0(\boldsymbol{R})$ of eq. (6.8) (where $k = 0$ means the ground state) will be denoted in short as $V(\boldsymbol{R})$, where $\boldsymbol{R}$ represents the vector of the nuclear positions. The function $V(\boldsymbol{R})$ has quite a complex structure and exhibits many basins of stable conformations (as well as many maxima and saddle points).

The problem of the shape of $V(\boldsymbol{R})$, as well as of the nuclear motion on the $V(\boldsymbol{R})$ hypersurface, will be the subject of the present chapter. It will be seen that the electronic energy can be computed within sufficient accuracy as a function of $\boldsymbol{R}$ only for very simple systems (such as an atom plus a diatomic molecule), for which quite a lot of detailed information can be obtained.

In practice, for large molecules, we are limited to only some approximations to $V(\boldsymbol{R})$ called *force fields*. After accepting such an approximation we encounter the problem of geometry optimization, i.e. of obtaining the most stable molecular conformation. Such a conformation is usually identified with a minimum on the electronic energy hypersurface, playing the role of the potential energy for the nuclei (*local molecular mechanics*). In practice we have the problem of the huge number of such minima. The real challenge in such a case is finding the *most* stable structure, usually corresponding to the global minimum (*global molecular mechanics*) of $V(\boldsymbol{R})$.

Molecular mechanics does not deal with nuclear motion as a function of time as well as with the kinetic energy of the system. This is the subject of molecular dynamics, which means solving the Newton equation of motion for all the nuclei of the system interacting by potential energy $V(\boldsymbol{R})$. Various approaches to this question (of general importance) will be presented at the end of the chapter.

Why is this important?

In 2001 the Human Genome Project, i.e. the sequencing of human DNA, was announced to be complete. This represents a milestone for humanity and its importance will grow steadily over the next decades. In the biotechnology laboratories DNA sequences will continue to be translated at a growing rate into a multitude of the protein sequences of amino acids. Only a tiny fraction of these proteins (0.1 percent?) may be expected to crystallize and then their atomic positions will be resolved by X-ray analysis. The *function* performed by a protein (e.g., an enzyme) is of crucial importance, rather than its sequence. The function depends on the 3D shape of the protein. For enzymes the cavity in the surface, where the catalytic reaction takes place is of great importance. The complex catalytic function of an enzyme consists of a series of elementary steps such as: molecular recognition of the enzyme cavity by a ligand, docking in the enzyme active centre within the cavity, carrying out a particular chemical reaction, freeing the products and finally returning to the initial state of the enzyme. The function is usually highly selective (pertains to a particular ligand only), precise (high yield reaction) and reproducible. To determine the function we must first of all identify the active centre and understand how it works. This, however, is possible either by expensive X-ray analysis of the crystal, or by a much less expensive theoretical prediction of the 3D structure of the enzyme molecule with atomic resolution accuracy. That is an important reason for theory development, isn't it?

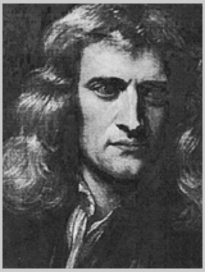
It is not necessary to turn our attention to large molecules only. Small ones are equally important: we are interested in predicting their structure and their conformation.

What is needed?

- Laplacian in spherical coordinates (Appendix H, p. 969, recommended).
- Angular momentum operator and spherical harmonics (Chapter 4, recommended).
- Harmonic oscillator (p. 166, necessary).
- Ritz method (Appendix L, p. 984, necessary).
- Matrix diagonalization (Appendix K, p. 982, necessary).
- Newton equation of motion (necessary).
- Chapter 8 (an exception: the Car–Parrinello method needs some results which will be given in Chapter 8, marginally important).
- Entropy, free energy, sum of states (necessary).

Classical works

There is no more classical work on dynamics than the monumental "*Philosophiae Naturalis Principia Mathematica*", Cambridge University Press, A.D. 1687 of Isaac Newton. ★ The idea of the force field was first presented by Mordechai Bixon and Shneior Lifson in *Tetrahedron* 23 (1967) 769 and entitled "*Potential Functions and Conformations in Cycloalkanes*".



Isaac Newton (1643–1727), English physicist, astronomer and mathematician, professor at Cambridge University, from 1672 member of the Royal Society of London, from 1699 Director of the Royal Mint – said to be merciless to the forgers. In 1705 Newton became a Lord. In the *opus magnum* mentioned above he developed the notions of space, time, mass and force, gave three principles of dynamics, the law of gravity and showed that the later pertains to problems that differ enormously in their scale (e.g., the famous apple and the planets). Newton is also a founder of differential and integral calculus (independently from G.W. Leibnitz). In addition Newton made some fundamental discoveries in optics, among other things he is the first to think that light is composed of particles.

★ The paper by Berni Julian Alder and Thomas Everett Wainwright "*Phase Transition for a Hard Sphere System*" in *Journal of Chemical Physics*, 27 (1957) 1208 is treated as the beginning of the molecular dynamics. ★ The work by Aneesur Rahman "*Correlations in the Motion of Atoms in Liquid Argon*" published in *Physical Review*, A136 (1964) 405 for the first time used a realistic interatomic potential (for 864 atoms). ★ The molecular dynamics of a small protein was first described in the paper by Andy McCammon, Bruce Gelin and Martin Karplus under the title "*Dynamics of folded proteins*", *Nature*, 267 (1977) 585. ★ The simulated annealing method is believed to have been used first by Scott Kirkpatrick, Charles D. Gellat and Mario P. Vecchi in a work "*Optimization by Simulated Annealing*", *Science*, 220 (1983) 671. ★ The Metropolis criterion for the choice of the current configuration in the Monte Carlo method was given by Nicolas Constantine Metropolis, Arianna W. Rosenbluth, Marshal N. Rosenbluth, Augusta H. Teller and Edward Teller in the paper "*Equations of State Calculations by Fast Computing Machines*" in *Journal of Chemical Physics*, 21 (1953) 1087. ★ The Monte Carlo method was used first by Enrico Fermi, John R. Pasta and Stanisław Marcin Ulam during their stay in Los Alamos (E. Fermi, J.R. Pasta, S.M. Ulam, "*Studies of Non-Linear Problems*", vol. 1, *Los Alamos Reports*, LA-1940). Ulam and John von Neumann are the discoverers of cellular automata.

7.1 ROVIBRATIONAL SPECTRA – AN EXAMPLE OF ACCURATE CALCULATIONS: ATOM – DIATOMIC MOLECULE

One of the consequences of adiabatic approximation is the idea of the potential energy hypersurface $V(\boldsymbol{R})$ for the motion of nuclei. To obtain the wave function for the motion of nuclei (and then to construct the total product-like wave function for the motion of electrons and nuclei) we have to solve the Schrödinger equation with $V(\boldsymbol{R})$ as the potential energy. This is what this hypersurface is for. We will find rovibrational (i.e. involving rotations and vibrations) energy levels and the corresponding wave functions, which will allow us to obtain rovibrational spectra (frequencies and intensities) to compare with experimental results.

7.1.1 COORDINATE SYSTEM AND HAMILTONIAN

Let us consider a diatomic molecule AB plus a weakly interacting atom C (e.g., H–H ... Ar or CO ... He), the total system in its electronic ground state. Let us centre the origin of the body-fixed coordinate system[1] (with the axes oriented as in the space-fixed coordinate system, see Appendix I, p. 971) in the centre of mass of AB. The problem involves therefore $3 \times 3 - 3 = 6$ dimensions.

However strange it may sound, six is too much for contemporary (otherwise impressive) computer techniques. Let us subtract one dimension by assuming that no vibrations of AB occur (rigid rotator). The five-dimensional problem becomes manageable. The assumption about the stiffness of AB now also pays off because we exclude right away two possible chemical reactions $C + AB \rightarrow CA + B$ and $C + AB \rightarrow CB + A$, and admit therefore only some limited set of nuclear configurations – only those that correspond to a weakly bound complex $C + AB$. This approximation is expected to work better when the AB molecule is stiffer, i.e. has a larger force constant (and therefore vibration frequency).[2]

Carl Gustav Jacob Jacobi (1804–1851), German mathematical genius, son of a banker, graduated from school at the age of 12, then associated with the universities of Berlin and Königsberg. Jacobi made important contributions to number theory, elliptic functions, partial differential equations, analytical mechanics.

Jacobi coordinates

We will introduce the *Jacobi coordinates* (Fig. 7.2, cf. p. 776): three components of vector $\boldsymbol{R}$ pointing to C from the origin of the coordinate system (the length R

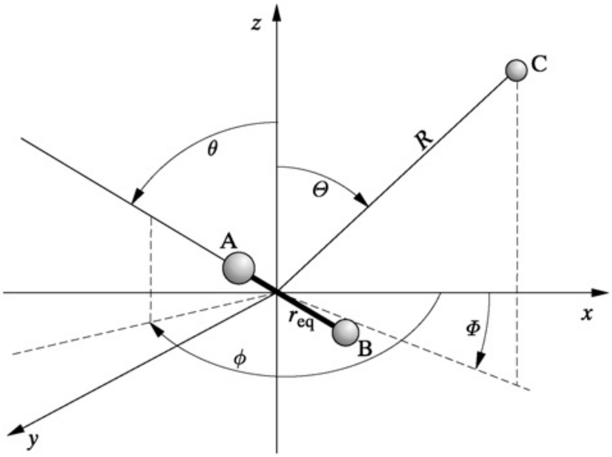


Fig. 7.2. The Jacobi coordinates for the C...AB system. The origin is in the centre of mass of AB (the distance AB is constant and equal to r_{eq}). The positions of atoms A and B are fixed by giving the angles θ, ϕ. The position of atom C is determined by three coordinates: R, Θ and Φ. Altogether we have 5 coordinates: $R, \Theta, \Phi, \theta, \phi$ or R, $\hat{\boldsymbol{R}}$ and $\hat{\boldsymbol{r}}$.

---

[1] Any coordinate system is equally good from the point of view of mathematics, but its particular choice may make the solution easy or difficult. In the case of a weak C ... AB interaction (our case) the proposed choice of the origin is one of the natural ones.

[2] A certain measure of this might be the ratio of the dissociation energy of AB to the dissociation energy of C...AB. The higher the ratio the better our model will be.

and angles Θ and Φ, both angles denoted by $\hat{\boldsymbol{R}}$) and the angles θ, ϕ showing the orientation $\hat{\boldsymbol{r}}$ of vector $\boldsymbol{r} = \overrightarrow{AB}$, altogether 5 coordinates – as there should be.

Now let us write down the Hamiltonian for the motion of the nuclei in the Jacobi coordinate system (with the stiff AB molecule with AB equilibrium distance equal to r_{eq}):[3]

$$\hat{H} = -\frac{\hbar^2}{2\mu R^2}\frac{\mathrm{d}}{\mathrm{d}R}R^2\frac{\mathrm{d}}{\mathrm{d}R} + \frac{\hat{l}^2}{2\mu R^2} + \frac{\hat{j}^2}{2\mu_{AB} r_{\mathrm{eq}}^2} + V,$$

where $\hat{l}^2$ denotes the operator of the square of the angular momentum of the atom C, $\hat{j}^2$ stands for the square of the angular momentum of the molecule AB,

$$\hat{l}^2 = -\hbar^2\left[\frac{1}{\sin\Theta}\frac{\partial}{\partial\Theta}\sin\Theta\frac{\partial}{\partial\Theta} + \frac{1}{\sin^2\Theta}\frac{\partial^2}{\partial\Phi^2}\right],$$

$$\hat{j}^2 = -\hbar^2\left[\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\sin\theta\frac{\partial}{\partial\theta} + \frac{1}{\sin^2\theta}\frac{\partial^2}{\partial\phi^2}\right],$$

μ is the reduced mass of C and the mass of (A + B), μ_{AB} denotes the reduced mass of A and B, V stands for the potential energy of the nuclear motion.

The expression for $\hat{H}$ is quite understandable. First of all, we have in $\hat{H}$ five coordinates, as there should be: R, two angular coordinates hidden in the symbol $\hat{\boldsymbol{R}}$ and two angular coordinates symbolized by $\hat{\boldsymbol{r}}$ – the four angular coordinates enter the operators of the squares of the two angular momenta. The first three terms in $\hat{H}$ describe the kinetic energy, V is the potential energy (the electronic ground state energy which depends on the nuclear coordinates). The kinetic energy operator describes the radial motion of C with respect to the origin (first term), the rotation of C about the origin (second term) and the rotation of AB about the origin (third term).

7.1.2 ANISOTROPY OF THE POTENTIAL V

How to figure out the shape of V? Let us first make a section of V. If we freeze the motion of AB,[4] the atom C would have (concerning the interaction energy) a sort of an energetic well around AB wrapping the AB molecule, caused by the C…AB van der Waals interaction, which will be discussed in Chapter 13. The bottom of the well would be quite distant from the molecule (van der Waals equilibrium distance), while the shape determined by the bottom points would resemble the shape of AB, i.e. would be a little bit elongated. The depth of the well would vary depending on orientation with respect to the origin.

---

[3] The derivation of the Hamiltonian is given in S. Bratož, M.L. Martin, *J. Chem. Phys.* 42 (1965) 1051.

[4] That is, fixed the angles θ and ϕ.

If V were isotropic, i.e. if atom C would have C…AB interaction energy independent[5] of $\hat{\boldsymbol{r}}$, then of course we might say that there is no coupling between the rotation of C and the rotation of AB. We would have then a conservation law separately for the first and the second angular momentum and the corresponding commutation rules (cf. Chapter 2 and Appendix F)

$$[\hat{H}, \hat{l}^2] = [\hat{H}, \hat{j}^2] = 0,$$
$$[\hat{H}, \hat{l}_z] = [\hat{H}, \hat{j}_z] = 0.$$

Therefore, the wave function of the total system would be the eigenfunction of $\hat{l}^2$ and $\hat{l}_z$ as well as of $\hat{j}^2$ and $\hat{j}_z$. The corresponding quantum numbers $l = 0, 1, 2, \ldots$ and $j = 0, 1, 2, \ldots$ that determine the squares of the angular momenta $\boldsymbol{l}^2$ and $\boldsymbol{j}^2$, as well as the corresponding quantum numbers $m_l = -l, -l+1, \ldots, l$ and $m_j = -j, -j+1, \ldots, j$ that determine the projections of the corresponding angular momenta on the z axis, would be legal[6] quantum numbers (full analogy with the rigid rotator, Chapter 4). The rovibrational levels could be labelled using pairs of quantum numbers: (l, j). In the absence of an external field (no privileged orientation in space) any such level would be $(2l+1)(2j+1)$-tuply degenerate, since this is the number of different projections of both angular momenta on the z axis.

7.1.3 ADDING THE ANGULAR MOMENTA IN QUANTUM MECHANICS

However, *V is not isotropic* (although the anisotropy is small). What then? Of all angular momenta, only the *total* angular momentum $\boldsymbol{J} = \boldsymbol{l} + \boldsymbol{j}$ is conserved (the conservation law results from the very foundations of physics, cf. Chapter 2).[7] Therefore, the vectors $\boldsymbol{l}$ and $\boldsymbol{j}$ when added to $\boldsymbol{J}$, would make all allowed angles: from minimum angle (the quantum number $J = l + j$), through smaller angles[8] and the corresponding quantum numbers $J = l + j - 1, l + j - 2$, etc., up to the angle 180°, corresponding to $J = |l - j|$). Therefore, the number of all possible values of J (each corresponding to a different energy) is equal to the number of projections of the *shorter*[9] of the vectors $\boldsymbol{l}$ and $\boldsymbol{j}$ on the longer one, i.e.

$$J = (l + j), (l + j - 1), \ldots, |l - j|. \tag{7.1}$$

For a given J there are $2J + 1$ projections of $\boldsymbol{J}$ on the z axis (because $|M_J| \leqslant J$); without any external field all these projections correspond to identical energy.

---

[5] I.e. the bottom of the well would be a sphere centred in the centre of mass of AB and the well depth would be independent of the orientation.

[6] We use to say "*good*".

[7] Of course, the momentum has also been conserved in the isotropic case, but in this case the energy was independent of the quantum number J (resulting from different angles between $\boldsymbol{l}$ and $\boldsymbol{j}$).

[8] The projections of the angular momenta are quantized.

[9] In the case of two vectors of the same length, the role of the shorter vector may be taken by either of them.

Please check that the number of all possible eigenstates is equal to $(2l+1)(2j+1)$, i.e. exactly what we had in the isotropic case. For example, for $l=1$ and $j=1$ the degeneracy in the isotropic case is equal to $(2l+1)(2j+1)=9$, while for anisotropic V we would deal with 5 states for $J=2$ (all of the same energy), 3 states corresponding to $J=1$ (the same energy, but different from $J=2$), a single state with $J=0$ (still another value of energy), altogether 9 states. This means that switching anisotropy on partially removed the degeneracy of the isotropic level (l, j) and gave the levels characterized by quantum number J.

7.1.4 APPLICATION OF THE RITZ METHOD

We will use the Ritz variational method (see Chapter 5, p. 202) to solve the Schrödinger equation. What should we propose as the expansion functions? It is usually recommended that we proceed systematically and choose first a complete set of functions depending on R, then a complete set depending on $\hat{\boldsymbol{R}}$ and finally a complete set that depends on the $\hat{\boldsymbol{r}}$ variables. Next, one may create the complete set depending on all five variables (these functions will be used in the Ritz variational procedure) by taking all possible products of the three functions depending on $R, \hat{\boldsymbol{R}}$ and $\hat{\boldsymbol{r}}$. There is no problem with the complete sets that have to depend on $\hat{\boldsymbol{R}}$ and $\hat{\boldsymbol{r}}$, as these may serve the spherical harmonics (the wave functions for the rigid rotator, p. 176) $\{Y_l^m(\Theta, \Phi)\}$ and $\{Y_{l'}^{m'}(\theta, \phi)\}$, while for the variable R we may propose the set of harmonic oscillator wave functions $\{\chi_v(R)\}$.[10] Therefore, we may use as the variational function:[11]

$$\Psi(R, \Theta, \Phi, \theta, \phi) = \sum c_{vlml'm'} \chi_v(R) Y_l^m(\Theta, \Phi) Y_{l'}^{m'}(\theta, \phi),$$

where c are the variational coefficients and the summation goes over v, l, m, l', m' indices. The summation limits have to be finite in practical applications, therefore the summations go to some maximum values of v, l and l' (m and m' vary from $-l$ to l and from $-l'$ to $+l'$). We hope (as always in quantum chemistry) that numerical results of a demanded accuracy will not depend on these limits. Then, as usual the Hamiltonian matrix is computed and diagonalized (see p. 982), and the eigenvalues E_J as well as the eigenfunctions ψ_{J,M_J} of the ground and excited states are found.

---

[10] See p. 164. Of course, our system does not represent any harmonic oscillator, but what counts is that the harmonic oscillator wave functions form a complete set (as the eigenfunctions of a Hermitian operator).

[11] The products $Y_l^m(\Theta, \Phi)\ Y_{l'}^{m'}(\theta, \phi)$ may be used to produce linear combinations that are automatically the eigenfunctions of $\hat{J}^2$ and $\hat{J}_z$, and have the proper parity (see Chapter 2). This may be achieved by using the Clebsch–Gordan coefficients (D.M. Brink, G.R. Satchler, "*Angular Momentum*", Clarendon, Oxford, 1975). The good news is that this way we can obtain a smaller matrix for diagonalization in the Ritz procedure, the bad news is that the matrix elements will contain more terms to be computed. The method above described will give the same result as using the Clebsch–Gordan coefficients, because the eigenfunctions of the Hamiltonian obtained within the Ritz method will automatically be the eigenfunctions of $\hat{J}^2$ and $\hat{J}_z$, as well as having the proper parity.

Each of the eigenfunctions will correspond to some J, M_J and to a certain parity. The problem is solved.

7.1.5 CALCULATION OF ROVIBRATIONAL SPECTRA

The differences of the energy levels provide the electromagnetic wave *frequencies* needed to change the stationary states of the system, the corresponding wave functions enable us to compute the *intensities* of the rovibrational transitions (which occur at the far-infrared and microwave wavelengths). When calculating the intensities to compare with experiments we have to take into account the Boltzmann distribution in the occupation of energy levels. The corresponding expression for the intensity $I(J'' \to J')$ of the transition from level J'' to level J' looks as follows:[12]

$$I\big(J'' \to J'\big) = (E_{J'} - E_{J''})\frac{\exp(\frac{E_{J''}-E_{J'}}{k_BT})}{Z(T)} \sum_{m,M_J',M_J''} \big|\langle\Psi_{J'M_J'}|\hat{\mu}_m|\Psi_{J''M_J''}\rangle\big|^2,$$

where:

partition function

- $Z(T)$ is the *partition function* (known from the statistical mechanics) – a function of the temperature T: $Z(T) = \sum_J (2J+1)\exp(-\frac{E_J}{k_BT})$, k_B is the Boltzmann constant
- $\hat{\mu}_m$ represents the dipole moment operator (cf. Appendix X)[13] $\hat{\mu}_0 = \hat{\mu}_z$, $\hat{\mu}_1 = \frac{1}{\sqrt{2}}(\hat{\mu}_x + i\hat{\mu}_y)$, $\hat{\mu}_{-1} = \frac{1}{\sqrt{2}}(\hat{\mu}_x - i\hat{\mu}_y)$
- the rotational state J'' corresponds to the vibrational state 0, while the rotational state J' pertains to the vibrational quantum number v, i.e. $E_{J''} \equiv E_{00J''}$, $E_{J'} \equiv E_{0vJ'}$ (index 0 denotes the electronic ground state)
- the integration is over the coordinates R, $\hat{\boldsymbol{R}}$ and $\hat{\boldsymbol{r}}$.

The dipole moment in the above formula takes into account that the charge distribution in the C…AB system depends on the nuclear configuration, i.e. on R,

---

[12] D.A. McQuarrie, "*Statistical Mechanics*", Harper&Row, New York, 1976, p. 471.

[13] The Cartesian components of the dipole moment operator read as

$$\hat{\mu}_x = \sum_{\alpha=1}^{M} Z_\alpha X_\alpha - \sum_{i=1}^{N}\langle\Psi_0^{el}|x_i|\Psi_0^{el}\rangle$$

and similarly for y and z, where Z_α denotes the charge (in a.u.) of the nucleus α, X_α denotes its x coordinate,

- Ψ_0^{el} denotes the electronic ground-state wave function of the system that depends parametrically on R, $\hat{\boldsymbol{R}}$ and $\hat{\boldsymbol{r}}$;
- $M = 3$, N stands for the number of electrons in C… AB;
- i is the electron index;
- the integration goes over the electronic coordinates.

Despite the fact that, for charged systems, the dipole moment operator $\hat{\boldsymbol{\mu}}$ depends on the choice of the origin of the coordinate system, the integral itself does not depend on such choice (good for us!). Why? Because these various choices differ by a constant vector (an example will be given in Chapter 13). The constant vector goes out of the integral and the corresponding contribution, depending on the choice of the coordinate system, gives 0, because of the orthogonality of the states.

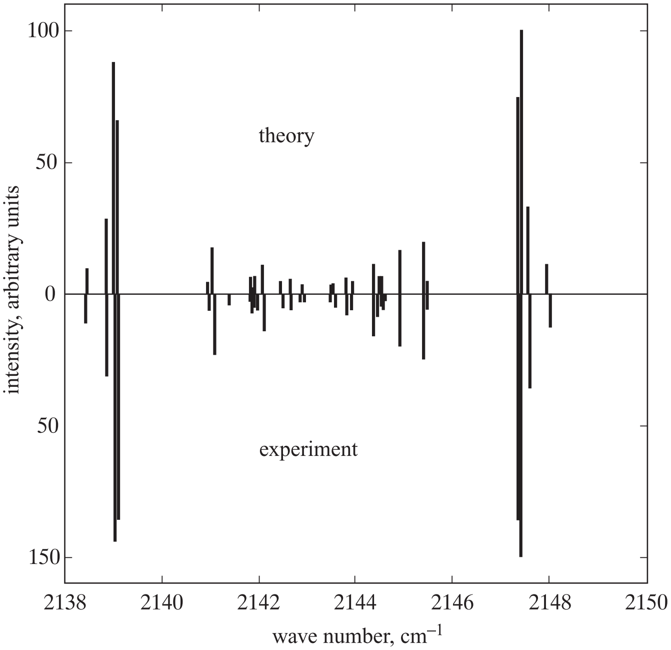


Fig. 7.3. Comparison of the theoretical and experimental intensities of the rovibrational transitions (in cm^{-1}) for the $^{12}C^{16}O \ldots {}^{4}He$ complex. Courtesy of Professor R. Moszyński.

$\hat{\boldsymbol{R}}$ and $\hat{\boldsymbol{r}}$, e.g., the atom C may have a net charge and the AB molecule may change its dipole moment when rotating.

Heijmen et al. carried out accurate calculations of the hypersurface V for a few atom-diatomic molecules, and then using the method described above the Schrödinger equation is solved for the nuclear motion. Fig. 7.3 gives a comparison of theory[14] and experiment[15] for the $^{12}C^{16}O$ complex with the ^{4}He atom.[16]

All the lines follow from the electric-dipole-allowed transitions [those for which the sum of the integrals in the formula $I(J'' \to J')$ is not equal to zero], each line is associated with a transition $(J'', l'', j'') \to (J', l', j')$.

7.2 FORCE FIELDS (FF)

The middle of the twentieth century marked the end of a long period of determining the building blocks of chemistry: chemical elements, chemical bonds, bond angles. The lists of these are not definitely closed, but future changes will be rather cosmetic than fundamental. This made it possible to go one step further and begin

---

[14] T.G.A. Heijmen, R. Moszyński, P.E.S. Wormer, A. van der Avoird, *J. Chem. Phys.* 107 (1997) 9921.

[15] C.E. Chuaqui, R.J. Le Roy, A.R.W. McKellar, *J. Chem. Phys.* 101 (1994) 39; M.C. Chan, A.R.W. McKellar, *J. Chem. Phys.* 105 (1996) 7910.

[16] Of course the results depend on the isotopes involved, even when staying within the Born–Oppenheimer approximation.

to rationalize the structure of molecular systems as well as to foresee the structural features of the compounds to be synthesized. The crucial concept is based on the adiabatic (or Born–Oppenheimer) approximation and on the theory of chemical bonds and resulted in the spatial structure of molecules. The great power of such an approach was proved by the construction of the DNA double helix model by Watson and Crick. The first DNA model was build from iron spheres, wires and tubes. This approach created problems: one of the founders of force fields, Michael Levitt, recalls[17] that a model of a tRNA fragment constructed by him with 2000 atoms weighted more than 50 kg.

James Dewey Watson, born 1928, American biologist, professor at Harvard University. Francis Harry Compton Crick (1916–2004), British physicist, professor at Salk Institute in San Diego. Both scholars won the 1962 Nobel Prize for "*their discoveries concerning the molecular structure of nucleic acids and its significance for information transfer in living material*". At the end of the historic paper J.D. Watson, F.H.C. Crick, *Nature*, 737 (1953) (of about 800 words) the famous enigmatic but crucial sentence appears: "*It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material*". The story behind the discovery is described in a colourful and unconventional way by Watson in his book "*Double Helix: A Personal Account of the Discovery of the Structure of DNA*".

The experience accumulated paid off by proposing some approximate expressions for electronic energy, which is, as we know from Chapter 6, the potential energy of the motion of the nuclei. This is what we are going to talk about.

Suppose we have a molecule (a set of molecules can be treated in a similar way). We will introduce the *force field,* which will be a *scalar* field – a function $V(\boldsymbol{R})$ of the nuclear coordinates $\boldsymbol{R}$. The function $V(\boldsymbol{R})$ represents a generalization (from one dimension to $3N-6$ dimensions) of the function $E_0^0(R)$ of eq. (6.8) on p. 225. The force acting on atom j occupying position x_j, y_j, z_j is computed as the components of the vector $\boldsymbol{F}_j = -\nabla_j V$, where

$$\nabla_j = \boldsymbol{i}\cdot\frac{\partial}{\partial x_j} + \boldsymbol{j}\cdot\frac{\partial}{\partial y_j} + \boldsymbol{k}\cdot\frac{\partial}{\partial z_j} \tag{7.2}$$

with $\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k}$ denoting the unit vectors along x, y, z, respectively.

FORCE FIELD

A force field represents a mathematical expression $V(\boldsymbol{R})$ for the electronic energy as a function of the nuclear configuration $\boldsymbol{R}$.

Of course, if we had to write down this scalar field in a 100% honest way, we have to solve (with an accuracy of about 1 kcal/mol) the electronic Schrödinger

---

[17] M. Levitt, *Nature Struct. Biol.* 8 (2001) 392.

equation (6.8) for every configuration $\boldsymbol{R}$ of the nuclei and take the eigenvalue [i.e. an analogue of $E_0^0(R)$] as $V(\boldsymbol{R})$. This would take so much time, even for small systems composed of a few atoms, that maybe even after some cursing, we would abandon this method with a great feeling of relief. Even if such a calculation required huge computation time, it would give results which would have been quite simple in their overall features (assuming that the molecule has a pattern of chemical bonds). It just would turn out that V would be characterized by the following:

- **Chemical bonds.** $V(\boldsymbol{R})$ would be obtained about its minimum, if any chemical bond between atoms X and Y had *a certain characteristic reference length* r_0 that would depend on the chemical species of the atoms X and Y. If the bond length were changed (shortened or elongated) to a certain value r, the energy would increase, first according to the harmonic law (with force constant k_{XY}) and then some deviations from the harmonic approximation begin.[18] A harmonic term of the kind $\frac{1}{2}k_{XY}(r-r_0)^2$ incorporated additively into V replaces the true anharmonic dependence by a harmonic approximation (assumption of small amplitudes) as if the two atoms had been bound by a harmonic spring (in the formula the atomic indices at symbols of distances have been omitted). The most important feature is that the same formula $\frac{1}{2}k_{XY}(r-r_0)^2$ is used for all chemical bonds $X{-}Y$, *independently of some particular chemical neighbourhood of a given* $X{-}Y$ *bond*. For example, one assumes that a non-distorted single C–C bond[19] has a characteristic reference length $r_0 = 1.523$ Å and a characteristic force constant $k_{XY} = 317\ \frac{\text{kcal}}{\text{mol Å}^2}$, similarly, some distinct parameters pertain to the C=C bond: $r_0 = 1.337$ Å, $k_{XY} = 690\ \frac{\text{kcal}}{\text{molÅ}^2}$ etc.[20,21]
- **Bond angles.** While preserving the distances r in the A–B and B–C bonds we may change the bond angle $\alpha =$ A–B–C, in this way changing the A...C distance. A corresponding change of V has to be associated with such a change. The energy has to increase when the angle α deviates from a characteristic reference value α_0. The harmonic part of such a change may be modelled by $\frac{1}{2}k_{XYZ}(\alpha-\alpha_0)^2$ (the indices for angles are omitted), which is equivalent to setting a corresponding harmonic spring for the bond angle and requires small amplitudes $|\alpha-\alpha_0|$. For example, for the angle C–C–C $\alpha_0 = 109.47°$ and $k_{XYZ} = 0.0099\frac{\text{kcal}}{\text{mol degree}^2}$, which means that to change the C...C distance by varying angle is about an order of magnitude easier than to change a CC bond length.

---

[18] These deviations from harmonicity (i.e. from the proportionality of force and displacement) are related to the smaller and smaller force needed to elongate the bond by a unit length and the larger and larger force needed to shorten the bond.

[19] That is, when all other terms in the force field equal zero.

[20] A CC bond involved in a conjugated single and double bonds (e.g., in benzene) also has its own parameters.

[21] A description of the popular MM2 force field is given in N.L. Allinger, *J. Am. Chem. Soc.* 99 (1977) 8127.

- **van der Waals interaction.** Two atoms X and Y, that do not form a chemical bond X–Y, as well as not participating in any sequence of bonds X–A–Y, still interact. There is nothing in the formulae introduced above that would prevent X and Y *collapsing* without any change of V. However, when two such atoms approach at a distance smaller than the sum of their radii (the van der Waals radii, see p. 742), then V had to increase very greatly.[22] On the other hand, at large interatomic distances the two atoms have to attract each other by the dispersion interaction vanishing as r^{-6} (cf. Chapter 13, p. 694). Hence, there is an equilibrium distance r_e, at which the interaction energy attains a minimum equal to $-\varepsilon$. These features of the interaction are captured by the widely used Lennard-Jones potential

John E. Lennard-Jones was professor of theoretical chemistry of the University of Cambridge, UK. The reader may find a historic picture of the theoretical chemistry team in *Intern. J. Quantum Chemistry*, S23 (1989), page XXXII.

$$V_{LJ}(X,Y) = \varepsilon\left[\left(\frac{r_e}{r}\right)^{12} - 2\left(\frac{r_e}{r}\right)^{6}\right],$$

where we skip for brevity the indices X, Y on the right-hand side. The Lennard-Jones potential given above is called LJ 12–6 (reflecting the powers involved). Sometimes other powers are used leading to other "LJ m–n" potentials.[23] Due to their simplicity, the LJ potentials are widely used, Fig. 7.4.

Lennard-Jones potential

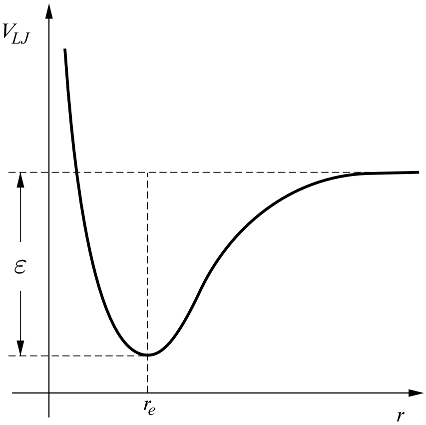


Fig. 7.4. The Lennard-Jones (LJ 12–6) potential. The parameter $\varepsilon > 0$ represents the depth of the potential well, while the parameter r_e denotes the minimum position. This r_e, corresponding to the *nonbonding* interaction of atoms X and Y, has no direct relation to the r_0 value pertaining to the chemical bond X–Y (discussed above; in order to keep the notation concise we do not distinguish between the two). The first is larger than the second by about an angstrom or so.

---

[22] A similar thing happens with cars: the repair cost increases very greatly, when the distance between two cars decreases below two thicknesses of the paint job.

[23] The power 12 has been chosen for two reasons: first, the power is sufficiently large to produce a strong repulsion when the two atoms approach each other, second, ... $12 = 6 \times 2$. The last reason makes the first derivative formula (i.e. the force) look more elegant than other powers do. A more elegant formula is often faster to compute and this is of practical importance.

- **Electrostatic interaction.** All the terms we have introduced to V so far do not take into account the fact that atoms carry net charges q_X and q_Y that have to interact electrostatically by Coulombic forces. To take this effect into account the electrostatic energy terms $q_X q_Y/r$ are added to V, where we assume the net charges q_X and q_Y are fixed (i.e. independent of the molecular conformation).[24]
- **Torsional interactions.** In addition to all the terms described above we often introduce to the force field a torsional term $A_{\text{X–Y–Z–W}}(1-\cos n\omega)$ for each torsional angle ω showing how V changes when a rotation ω about the chemical bond YZ, in the sequence X–Y–Z–W of chemical bonds, takes place (n is the multiplicity of the energy barriers per single turn[25]). Some rotational barriers already result from the van der Waals interaction of the X and W atoms, but in practice the barrier heights have to be corrected by the torsional potentials to reproduce experimental values.
- **Mixed terms.** Besides the above described terms one often introduces some *coupling (mixed) terms*, e.g., bond–bond angle etc. The reasoning behind this is simple. The bond angle force constant X–Y–Z has to depend on the bond-lengths X–Y and Y–Z, etc.

Summing up a simple force field might be expressed as shown in Fig. 7.5, where for the sake of simplicity the indices X, Y at r, r_0, as well as X, Y, Z at α, α_0, and X, Y, Z, W at ω have been omitted:

$$V=\sum_{X-Y}\frac{1}{2}k_{XY}(r-r_0)^2+\sum_{X-Y-Z}\frac{1}{2}k_{XYZ}(\alpha-\alpha_0)^2+\sum_{X...Y}V_{LJ}(X,Y)$$

$$+\sum_{X,Y}\frac{q_X q_Y}{r}+\sum_{\text{tors}}A_{\text{X–Y–Z–W}}(1-\cos n\omega)+\text{coupling terms (if any).}$$

Such simple formulae help us to figure out how the electronic energy looks as a function of the configuration of the nuclei. Our motivation is as follows:

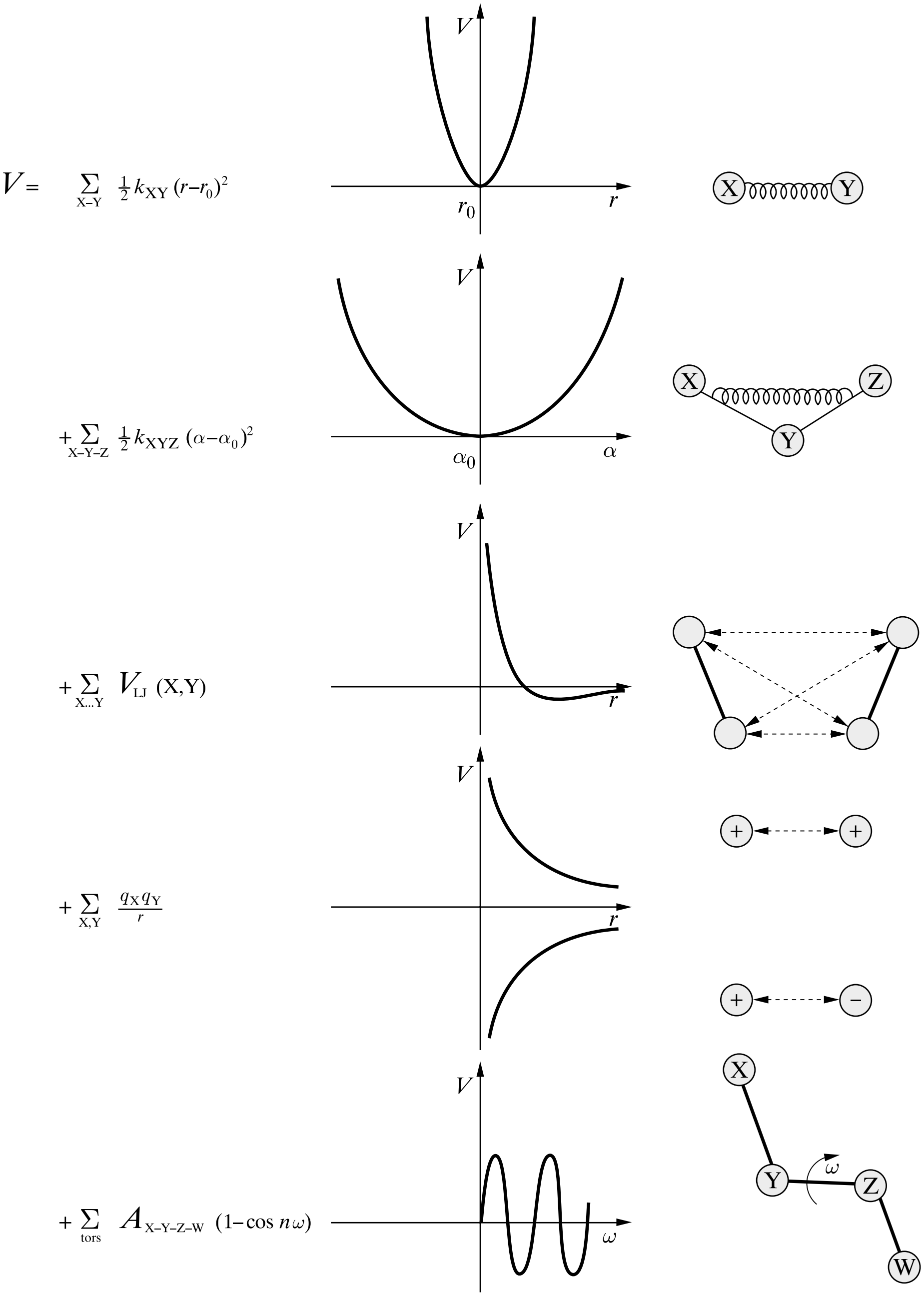
- economy of computation: *ab initio* calculations of the electronic energy for larger molecules would have been many orders of magnitude more expensive;

---

[24] In some force fields the electrostatic forces depend on the dielectric constant of the neighbourhood (e.g., solvent) despite the fact that this quantity has a macroscopic character and does not pertain to the nearest neighbourhood of the interacting atoms. If all the calculations had been carried out taking the molecular structure of the solvent into account as well as the polarization effects, no dielectric constant would have been needed. If this is not possible, then the dielectric constant effectively takes into account the polarization of the medium (including reorientation of the solvent molecules). The next problem is *how* to introduce the dependence of the electrostatic interaction of two atomic charges on the dielectric constant. In some of the force fields we introduce a brute force kind of damping, namely, the dielectric constant is introduced into the denominator of the Coulombic interaction as equal to the ... interatomic distance.

In second generation force fields (e.g., W.D. Cornell, P. Cieplak, C.I. Bayly, I.R. Gould, K.M. Merz Jr., D.M. Ferguson, D.C. Spellmeyer, T. Fox, J.W. Caldwell, P.A. Kollman, *J. Amer. Chem. Soc.* 117 (1995) 5179) we explicitly take into account the induction interaction, e.g., the dependence of the atomic electric charges on molecular conformations.

[25] For example, $n=3$ for ethane.



$$V = \sum_{X-Y} \tfrac{1}{2} k_{XY} (r-r_0)^2 + \sum_{X-Y-Z} \tfrac{1}{2} k_{XYZ} (\alpha-\alpha_0)^2 + \sum_{X...Y} V_{LJ}\,(X,Y) + \sum_{X,Y} \frac{q_X q_Y}{r} + \sum_{tors} A_{X-Y-Z-W}\,(1-\cos n\omega)$$

Fig. 7.5. The first force field of Bixon and Lifson in a mnemonic presentation.

- in addition, a force field gives $V(\boldsymbol{R})$ in the form of a simple formula for *any* positions $\boldsymbol{R}$ of the nuclei, while the calculation of the electronic energy would give us $V(\boldsymbol{R})$ numerically, i.e. for some *selected* nuclear configurations.

7.3 LOCAL MOLECULAR MECHANICS (MM)

7.3.1 BONDS THAT CANNOT BREAK

It is worth noting that the force fields correspond to a fixed (and unchangeable during computation) system of chemical bonds. The chemical bonds are treated as springs, most often satisfying Hooke's[26] law (harmonic), and therefore unbreakable.[27] Similarly, the bond angles are forced to satisfy Hooke's law. Such a force field is known as *flexible molecular mechanics*. To decrease the number of variables, we sometimes use *rigid molecular mechanics*,[28] in which the bond lengths and the bond angles are fixed at values close to experimental ones, but the torsional angles are free to change. The argument behind such a choice is that the frequencies associated with torsional motion are much lower than those corresponding to the bond angle changes, and much much lower than frequencies of the bond length vibrations. This means that a quantity of energy is able to make only tiny changes in the bond lengths, small changes in the bond angles and large changes in the torsional angles, i.e. the torsional variables determine the overall changes of the molecular geometry. Of course, the second argument is that a smaller number of variables means lower computational costs.

flexible MM

rigid MM

Molecular mechanics represents a method of finding a stable configuration of the nuclei by using a minimization of $V(\boldsymbol{R})$ with respect to the nuclear coordinates (for a molecule or a system of molecules).

The essence of molecular mechanics is that we roll the potential energy hypersurface slowly downhill from a starting point chosen (corresponding to a certain starting geometry of the molecule) to the "nearest" energy minimum corresponding to the final geometry of the molecule. The "rolling down" is carried out by a minimization procedure that traces point by point the trajectory in the configurational space, e.g., in the direction of the negative gradient vector calculated at any consecutive point. The minimization procedure represents a mechanism showing how to obtain the next geometry from the previous one. The procedure ends,

---

[26] Robert Hooke, British physicist and biologist (1635–1703).

[27] There are a few such force fields in the literature. They give similar results, as far as their main features are considered. The force field concept was able to clarify many observed phenomena, even fine effects. It may also fail as with anything confronting the real world.

[28] Stiff molecular mechanics was a very useful tool for Paul John Flory (1910–1985), American chemist, professor at the universities at Cornell and Stanford. Using such mechanics, Flory developed a theory of polymers that explained their physical properties. In 1974 he obtained the Nobel Prize "*for his fundamental achievements, both theoretical and experimental, in the physical chemistry of macromolecules*".

when the geometry ceases to change (e.g., the gradient vector has zero length[29]). The geometry attained is called the equilibrium or stable geometry. The rolling described above is more like a crawling down with large friction, since in molecular mechanics the kinetic energy is always zero and the *system is unable to go uphill*[30] of V.

A lot of commercial software[31] offers force field packets. For example, the Hyperchem package provides the force fields AMBER and MM2,[32] the program Insight offers the CVFF force field. Unfortunately, the results depend to quite a significant degree on the force field chosen. Even using the same starting geometry we may obtain final (equilibrium) results that differ very much one from another. Usually the equilibrium geometries obtained in one force field do not differ much from those from another one, but the corresponding energies may be very different. Therefore, the most stable geometry (corresponding to the lowest energy) obtained in a force field may turn out to be less stable in another one, thus leading to different predictions of the molecular structure.

A big problem in molecular mechanics is that the final geometry is very close to the starting one. We start from a boat (chair) conformation of cyclohexane and obtain a boat (chair) equilibrium geometry. The very essence of molecular mechanics however, is that when started from some, i.e. distorted boat (chair) conformation, we obtain the perfect, beautiful equilibrium boat (chair) conformation, which may be compared with experimental results. Molecular mechanics is extremely useful in conformational studies of systems with a small number of stable conformations, either because the molecule is small, rigid or its overall geometry is fixed. In such cases all or all "reasonable",[33] conformations can be investigated and those of lowest-energy can be compared with experimental results.

7.3.2 BONDS THAT CAN BREAK

Harmonic bonds cannot be broken and therefore molecular mechanics with harmonic approximation is unable to describe chemical reactions. When instead of harmonic oscillators we use Morse model (p. 169), then the bonds can be broken.

And yet we most often use the harmonic oscillator approximation. Why? There are a few reasons:

- the Morse model requires many computations of the exponential function, which is expensive[34] when compared to the harmonic potential;

---

[29] The gradient is also equal zero at energy maxima and energy saddle points. To be sure that a minimum really has been finally attained we have to calculate (at the particular point suspected to be a minimum) a Hessian, i.e. the matrix of the second derivatives of V, then diagonalize it (cf. p. 982) and check whether the eigenvalues obtained are all positive.

[30] Unless assuming too large a step (but this has to be considered as an error in the "art of computing").

[31] See the Web Annex.

[32] N.L. Allinger, *J. Am. Chem. Soc.* 99 (1977) 8127.

[33] A very dangerous word!

[34] Each time requires a Taylor expansion calculation.

- the Morse potential requires three parameters, while the harmonic model needs only two parameters;
- in most applications the bonds do not break and it would be very inconvenient to obtain breaking due, for instance, to a particular starting point;
- a description of chemical reactions requires not only the possibility of breaking bonds, but also a realistic, i.e. quantum chemical, computation of the charge distributions involved (cf. p. 308). The Morse potential would be too simplistic for such purposes.

7.4 GLOBAL MOLECULAR MECHANICS

7.4.1 MULTIPLE MINIMA CATASTROPHE

If the number of local minima is very large (and this may happen even for medium size molecules) or even "astronomic", then exploring the whole conformational space (all possible geometries) by finding all possible minima using a minimization procedure becomes impossible. Hence, we may postulate another procedure which may be called *global molecular mechanics* and could find the global minimum (the most stable conformation) starting from any point in the configurational space.

If the number of local minima is small, there is in principle no problem with using theory. Usually it turns out that the quantum mechanical calculations are feasible, often even at the *ab initio* level. A closer insight leads, however, to the conclusion that only some extremely accurate and expensive calculations would give the correct energy sequence of the conformers, and that only for quite small molecules with a dozen atoms. This means that for larger molecules we are forced to use molecular mechanics. For molecules with a few atoms we might investigate the whole conformational space by sampling it by a stochastic or systematic procedure, but this approach soon becomes prohibitive for larger molecules.

For such larger molecules we encounter difficulties which may only be appreciated by individuals who have made such computations themselves. We may say, in short, that virtually nothing helps us with the huge number of conformations to investigate. According to Schepens[35] the number of the conformations found is proportional to the time spent conducting the search. It is worth noting that this means catastrophe, because for a twenty amino acid oligopeptide the number of conformations is of the order[36] of 10^{20}, and for a hundred amino acids –

---

[35] Wijnand Schepens, PhD thesis, University of Gand, 2000.

[36] The difficulty of finding a particular conformation among 10^{20} conformations is a real horror. Maybe the example below will show what a severe problem has been encountered. A single grain of sand has a diameter of about 1 mm. Let us try to align 10^{20} of such sand grains side by side. What will the length of such a chain of grains be? Let us compute: 10^{20} mm $= 10^{17}$ m $= 10^{14}$ km. One light year is 300000 km/s × 3600 s × 24 × 365 $\simeq 10^{13}$ km. Hence, the length is about 10 light years, i.e. longer than the round trip from our Sun to the nearest star – Alpha Centauri. This is what the thing looks like.

10^{100}. Also methods based on molecular dynamics (cf. p. 304) do not solve the problem, since they could cover only a tiny fraction of the total conformational space.

7.4.2 IS IT THE GLOBAL MINIMUM WHICH COUNTS?

The goal of conformational analysis is to find those conformations of the molecule which are observed under experimental conditions. At temperatures close to 300 K the lowest-energy conformations prevail in the sample, i.e. first of all those corresponding the global minimum of the potential energy[37] V.

We may ask whether indeed the global minimum of the potential energy decides the observed experimental geometry. Let us neglect the influence of the solvent (neighbourhood). A better criterion would be the global minimum of the *free energy*, $E - TS$, where the entropic factor would also enter. A wide potential well means a higher density of vibrational states, a narrow well means a lower density of states (cf. eq. (4.21), p. 171; a narrow well corresponds to a large α). If the global minimum corresponds to a wide well, the well is additionally stabilized by the entropy,[38] otherwise it is destabilized.

free energy

For large molecules, there is a possibility that, due to the synthesis conditions, the molecule is trapped in a local minimum (*kinetic minimum*), different from the global minimum of the free energy (*thermodynamic minimum*), Fig. 7.6.

kinetic minimum

thermodynamic minimum

For the same reason that the diamonds (kinetic minimum) in your safe do not change spontaneously into graphite (thermodynamic minimum), a molecule imprisoned in the kinetic minimum may rest there for a very long time (when compared with experimental time). Nobody knows whether the native conformation of

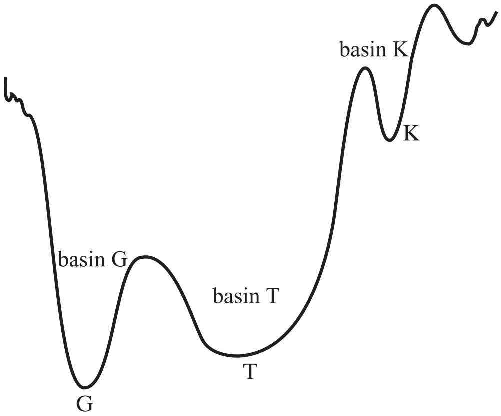


Fig. 7.6. Electronic energy $V(\boldsymbol{R})$ as function of the nuclear configuration $\boldsymbol{R}$. The basins of the *thermodynamic minimum* (T), of the *kinetic minimum* (K) and of the *global* minimum (G). The deepest basin (G) should not correspond to the thermodynamically most stable conformation (T). Additionally, the system may be caught in a kinetic minimum (K), from which it may be difficult to tunnel to the thermodynamic minimum basin. Diamond and fullerenes may serve as examples of K.

---

[37] Searching for the global minimum of V is similar to the task of searching for the lowest valley on Earth when starting from an arbitrary point on the surface.

[38] According the famous formula of Ludwig Boltzmann, entropy $S = k_B \ln \Omega(E)$, where Ω is the number of the states available for the system at energy E. The more states, the larger the entropy.

Christian Anfinsen obtained the Nobel Prize in 1972 "*for his work on ribonuclease, especially concerning the connection between the amino acid sequence and the biologically active conformation*". He made an important contribution showing that after denaturation (a large change of conformation) some proteins fold back spontaneously to their native conformation.

a protein corresponds to the thermodynamic or kinetic minimum.[39] Some experiments indicate the first, others the second possibility.

Despite these complications we generally assume in conformational analysis, that the global minimum and other low-energy conformations play the most important role. In living matter, taking a definite (native) conformation is sometimes crucial. It has been shown[40] that the native conformation of natural enzymes has much lower energy than those of other conformations (energy gap). Artificial enzymes with stochastic amino acid sequences do not usually have this property resulting in no well-defined conformation.

Global molecular mechanics is, in my opinion, one of the most important challenges in chemistry. Students need to look for an important research subject. This is such a subject.[41]

7.5 SMALL AMPLITUDE HARMONIC MOTION – NORMAL MODES

The hypersurface $V(\boldsymbol{R})$ has, in general (especially for large molecules), an extremely complex shape with many minima, each corresponding to a stable conformation. Let us choose *one* of those minima and ask *what kind of motion the molecule undergoes, when only small displacements from the equilibrium geometry are allowed*. In addition we assume that the potential energy for this motion is a *harmonic* approximation of the $V(\boldsymbol{R})$ in the neighbourhood of the minimum.[42] Then we obtain the *normal vibrations* or *normal modes*.

NORMAL MODES

A normal mode represents a harmonic oscillation (of a certain frequency) of all the atoms of the molecule about their equilibrium positions with the same phase for all the atoms (i.e. all the atoms attain their equilibrium position at the same time).

---

[39] It is clear if a protein were denatured very heavily (e.g., cooking chicken soup we could not expect the chicken to return to life).

[40] E.I. Shakanovich, A.M. Gutin, *Proc. Natl. Acad. Sci. USA* 90 (1993) 7195; A. Šali, E.I. Shakanovich, M. Karplus, *Nature* 369 (1994) 248.

[41] My own adventure with this topic is described in L. Piela, "*Handbook of Global Oprimization*", vol. 2, P.M. Pardalos, H.E. Romeijn, eds., Kluwer Academic Publishers, Boston, 2002.

[42] We may note *en passant* that a similar philosophy prevailed in science until quite recent times: take only the linear approximation and forget about non-linearities. It turned out, however, that the nonlinear phenomena (cf. Chapter 15) are really fascinating.

The number of such vibrations with non-zero frequencies is equal to $3N - 6$. A vibrational motion of the molecule represents a superposition of these individual normal modes.

7.5.1 THEORY OF NORMAL MODES

Suppose we have at our disposal an analytical expression for $V(\boldsymbol{R})$ (e.g., the force field), where $\boldsymbol{R}$ denotes the vector of the Cartesian coordinates of the N atoms of the system (it has $3N$ components). Let us assume (Fig. 7.7) that the function $V(\boldsymbol{R})$ has been minimized in the configurational space, starting from an initial position $\boldsymbol{R}_i$ and going downhill until a minimum position $\boldsymbol{R}_0$ has been reached, the $\boldsymbol{R}_0$ corresponding to one of many minima the V function may possess[43] (we will call the minimum the "closest" to the $\boldsymbol{R}_i$ point in the configurational space). All the points $\boldsymbol{R}_i$ of the configurational space that lead to $\boldsymbol{R}_0$ represent the *basin of the attractor*[44] $\boldsymbol{R}_0$.

From this time on, all other basins of the function $V(\boldsymbol{R})$ *have "disappeared from the theory" – only motion in the neighbourhood of* $\boldsymbol{R}_0$ is to be considered.[45] If someone is aiming to apply harmonic approximation and to consider small displacements from $\boldsymbol{R}_0$ (as we do), then it is a good idea to write down the Taylor expansion of V about $\boldsymbol{R}_0$ [hereafter instead of the symbols $X_1, Y_1, Z_1, X_2, Y_2, Z_2, \ldots$ for the atomic Cartesian coordinates we will use a slightly more uniform notation:

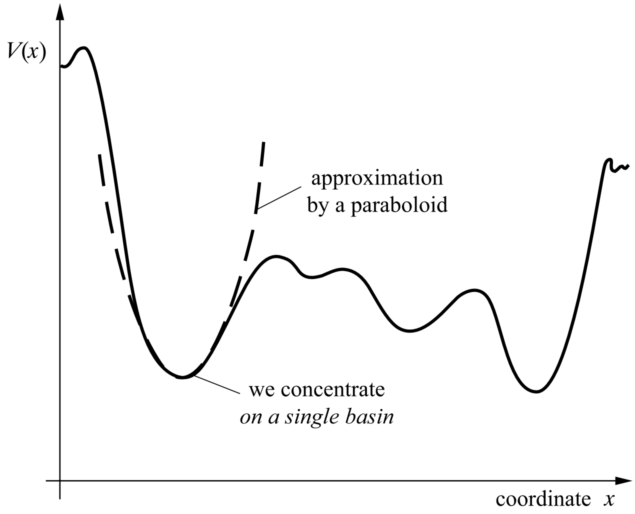


Fig. 7.7. A schematic (one-dimensional) view of the hypersurface $V(x)$ that illustrates the choice of a particular basin of V related to the normal modes to be computed. The basin chosen is then approximated by a paraboloid in $3N$ variables. This gives the $3N - 6$ modes with non-zero frequencies and 6 "modes" with zero frequencies.

---

[43] These are improper minima, because a translation or rotation of the system does not change V.

[44] The total configurational space consists of a certain number of such basins.

[45] For another starting conformation $\boldsymbol{R}_i$ we might obtain another minimum of $V(\boldsymbol{R})$. This is why the choice of $\boldsymbol{R}_i$ has to have a definite relation to that which is observed experimentally.

$\boldsymbol{R} = (X_1, X_2, X_3, X_4, X_5, X_6, \ldots, X_{3N})^T]$

$$V(\boldsymbol{R}_0 + \boldsymbol{x}) = V(\boldsymbol{R}_0) + \sum_i \left(\frac{\partial V}{\partial x_i}\right)_0 x_i + \frac{1}{2}\sum_{ij}\left(\frac{\partial^2 V}{\partial x_i \partial x_j}\right)_0 x_i x_j + \cdots, \tag{7.3}$$

where $\boldsymbol{x} = \boldsymbol{R} - \boldsymbol{R}_0$ is the vector with the *displacements* of the atomic positions from their equilibria ($x_i = X_i - X_{i,0}$ for $i = 1, \ldots, 3N$), while the derivatives are calculated at $\boldsymbol{R} = \boldsymbol{R}_0$.

In $\boldsymbol{R}_0$ all the first derivatives vanish. According to the harmonic approximation, the higher order terms denoted as "$+\cdots$" are neglected. In effect we have

$$V(\boldsymbol{R}_0 + \boldsymbol{x}) \cong V(\boldsymbol{R}_0) + \frac{1}{2}\sum_{ij}\left(\frac{\partial^2 V}{\partial x_i \partial x_j}\right)_0 x_i x_j. \tag{7.4}$$

In matrix notation we have $V(\boldsymbol{R}_0 + \boldsymbol{x}) = V(\boldsymbol{R}_0) + \frac{1}{2}\boldsymbol{x}^T \boldsymbol{V}'' \boldsymbol{x}$, where $\boldsymbol{V}''$ is a square matrix of the Cartesian *force constants*, $(\boldsymbol{V}'')_{ij} = (\frac{\partial^2 V}{\partial x_i \partial x_j})_0$.

force constant

The Newton equations of motion for all the atoms of the system can be written in matrix form as ($\ddot{\boldsymbol{x}}$ means the second derivative with respect to time t)

$$\boldsymbol{M}\ddot{\boldsymbol{x}} = -\boldsymbol{V}''\boldsymbol{x}, \tag{7.5}$$

where $\boldsymbol{M}$ is the diagonal matrix of the atomic masses (the numbers on the diagonal are: M_1, M_1, M_1, M_2, M_2, M_2, ...), because we calculate the force component along the axis k as

$$-\frac{\partial V}{\partial x_k} = -\frac{1}{2}\sum_j \left(\frac{\partial^2 V}{\partial x_k \partial x_j}\right)_0 x_j - \frac{1}{2}\sum_i \left(\frac{\partial^2 V}{\partial x_i \partial x_k}\right)_0 x_i$$
$$= -\sum_j \left(\frac{\partial^2 V}{\partial x_k \partial x_j}\right)_0 x_j = -(\boldsymbol{V}''\boldsymbol{x})_k.$$

We may use the relation $\boldsymbol{M}^{\frac{1}{2}}\boldsymbol{M}^{\frac{1}{2}} = \boldsymbol{M}$

$$\boldsymbol{M}^{\frac{1}{2}}\boldsymbol{M}^{\frac{1}{2}}\ddot{\boldsymbol{x}} = -\boldsymbol{M}^{\frac{1}{2}}\boldsymbol{M}^{-\frac{1}{2}}\boldsymbol{V}''\boldsymbol{M}^{-\frac{1}{2}}\boldsymbol{M}^{\frac{1}{2}}\boldsymbol{x}, \tag{7.6}$$

where $\boldsymbol{M}^{\frac{1}{2}}$ is a matrix similar to $\boldsymbol{M}$, but its elements are the square roots of the atom masses instead of the masses, while the matrix $\boldsymbol{M}^{-\frac{1}{2}}$ contains the inverse square roots of the masses. The last equation, after multiplying from the left by $\boldsymbol{M}^{-\frac{1}{2}}$, gives

$$\ddot{\boldsymbol{y}} = -\boldsymbol{A}\boldsymbol{y}, \tag{7.7}$$

where $\boldsymbol{y} = \boldsymbol{M}^{\frac{1}{2}}\boldsymbol{x}$ and $\boldsymbol{A} = \boldsymbol{M}^{-\frac{1}{2}}\boldsymbol{V}''\boldsymbol{M}^{-\frac{1}{2}}$.

Let us try to find the solution in the form[46]

$$\boldsymbol{y} = \boldsymbol{c}_1 \exp(+i\omega t) + \boldsymbol{c}_2 \exp(-i\omega t),$$

where the vectors $\boldsymbol{c}_i$ (of the dimension $3N$) of the complex coefficients are time independent. The coefficients $\boldsymbol{c}_i$ depend on the initial conditions as well as on the $\boldsymbol{A}$ matrix. If we say that at time $t = 0$ all the atoms are at equilibrium, i.e. $\boldsymbol{y}(t = 0) = \boldsymbol{0}$, then we obtain the relation $\boldsymbol{c}_1 = -\boldsymbol{c}_2$ leading to the formula

$$\boldsymbol{y} = \boldsymbol{L} \sin(\omega t), \tag{7.8}$$

where the vector[47] $\boldsymbol{L}$ and ω depend on the matrix $\boldsymbol{A}$. Vector $\boldsymbol{L}$ is determined only to the accuracy of a multiplication constant, because multiplication of $\boldsymbol{L}$ by any number does not interfere with satisfying (7.7).

When we insert the proposed solution (7.8) in (7.7), we immediately obtain, that ω and $\boldsymbol{L}$ have to satisfy the following equation

$$\left(\boldsymbol{A} - \omega^2 \mathbf{1}\right)\boldsymbol{L} = \boldsymbol{0}. \tag{7.9}$$

The values of ω^2 represent the eigenvalues,[48] while the $\boldsymbol{L}$ are the eigenvectors of the $\boldsymbol{A}$ matrix. There are $3N$ eigenvalues, and each of them corresponds to its eigenvector $\boldsymbol{L}$. This means that we have $3N$ *normal modes*, each mode characterized by its angular frequency $\omega = 2\pi\nu$ (ν is the frequency) and its vibration amplitudes $\boldsymbol{L}$. Hence, it would be natural to assign a normal mode index $k = 1, \ldots, 3N$ for ω and $\boldsymbol{L}$. Therefore we have

$$\boxed{\left(\boldsymbol{A} - \omega_k^2 \mathbf{1}\right)\boldsymbol{L}_k = 0 \qquad (7.10)}$$

The diagonalization of $\boldsymbol{A}$ (p. 982) is an efficient technique for solving the eigenvalue problem using commercial computer programs (diagonalization is equivalent to a rotation of the coordinate system, Fig. 7.8).

> This is equivalent to replacing V by a $3N$-dimensional paraboloid with origin at $\boldsymbol{R}_0$. The normal mode analysis means such a rotation of the coordinate system as will make the new axes coincide with the principal axes of the paraboloid.

---

[46] This form (with $\omega = a + ib$) allows for a constant solution ($a = b = 0$), an exponential growth or vanishing ($a = 0$, $b \neq 0$), oscillations ($a \neq 0$, $b = 0$), oscillatory growing or oscillatory vanishing ($a \neq 0$, $b \neq 0$). For $\boldsymbol{R}_0$ denoting a minimum, $\det \boldsymbol{A} > 0$ and this assures a solution with $a \neq 0$, $b = 0$.

[47] Equal to $2i\boldsymbol{c}_1$, but since $\boldsymbol{c}_1$ is unknown, as for the time being is $\boldsymbol{L}$, therefore we can say goodbye to $\boldsymbol{c}_1$ without feeling any discomfort whatsoever.

[48] $\boldsymbol{A}$ is a symmetric matrix, hence its eigenvalues ω^2 and therefore also $\omega = a + ib$ are real ($b = 0$). Whether ω are positive, negative or zero depends on the hypersurface V at $\boldsymbol{R}_0$, see Fig. 7.8.

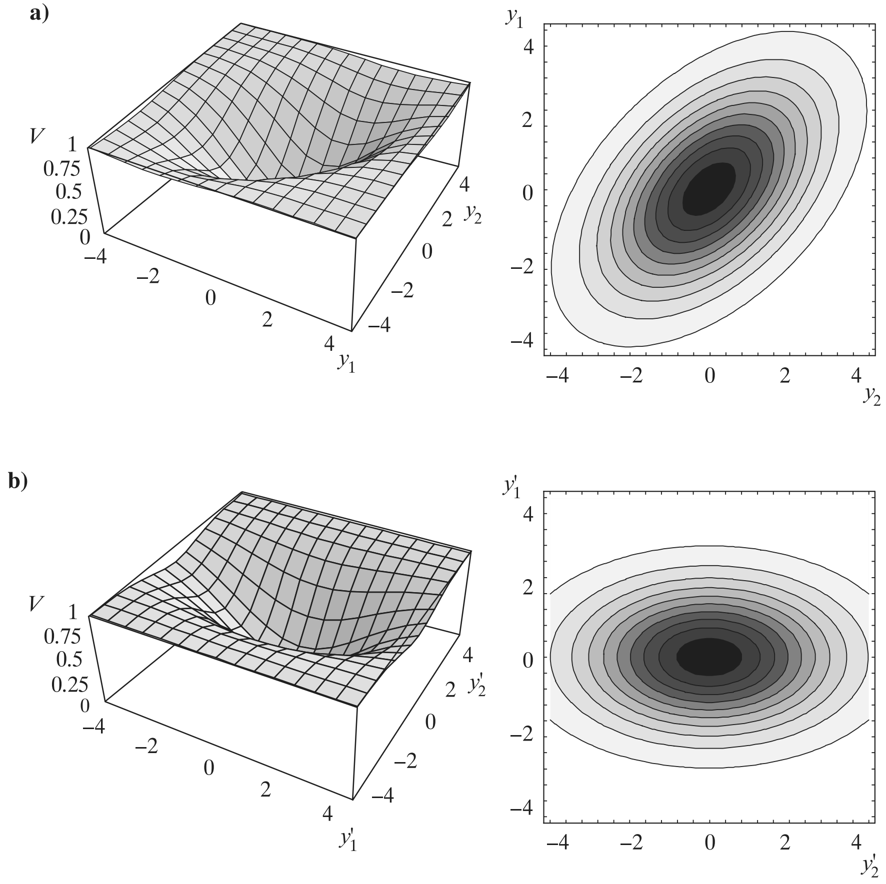


Fig. 7.8. (a) and (b) show the normal vibrations (normal modes) about a point $\boldsymbol{R}_0 = \mathbf{0}$ being *a minimum* of the potential energy function $V(\boldsymbol{R}_0 + \boldsymbol{y})$ of two variables $\boldsymbol{y} = (y_1, y_2)$. This function is first approximated by a quadratic function, i.e. a paraboloid $\tilde{V}(y_1, y_2)$. Computing the normal modes is equivalent to such a *rotation of the Cartesian coordinate system* (a), that the new axes (b) y_1' and y_2' become the principal axes of any section of $\tilde{V}$ by a plane $\tilde{V} = \text{const}$ (i.e. ellipses). Then, we have $\tilde{V}(y_1, y_2) = V(\boldsymbol{R}_0 = \mathbf{0}) + \frac{1}{2}k_1(y_1')^2 + \frac{1}{2}k_2(y_2')^2$. The problem then becomes equivalent to the two-dimensional harmonic oscillator (cf. Chapter 4) and separates into two independent one-dimensional oscillators (normal modes): one of angular frequency $\omega_1 = 2\pi\nu_1 = \sqrt{\frac{k_1}{m}}$ and the other with angular frequency $\omega_2 = 2\pi\nu_2 = \sqrt{\frac{k_2}{m}}$, where m is the mass of the oscillating particle. Figs. (c), (d) show what would happen, if $\boldsymbol{R}_\mathbf{0}$ corresponded not to a minimum, but to a maximum (c) or the saddle point (d). For a maximum (c) k_1 and k_2 in $\tilde{V}(y_1', y_2') = V(\mathbf{0}) + \frac{1}{2}k_1(y_1')^2 + \frac{1}{2}k_2(y_2')^2$ would be both *negative*, and therefore the corresponding normal "vibrations" would have had both imaginary frequencies, while for the saddle point (d) only one of the frequencies would be imaginary.

There will be six frequencies (five for a linear molecule) equal to zero. They are connected to the translation and rotation of the molecule in space: three translations along x, y, z and three rotations about x, y, z (two in the case of a linear

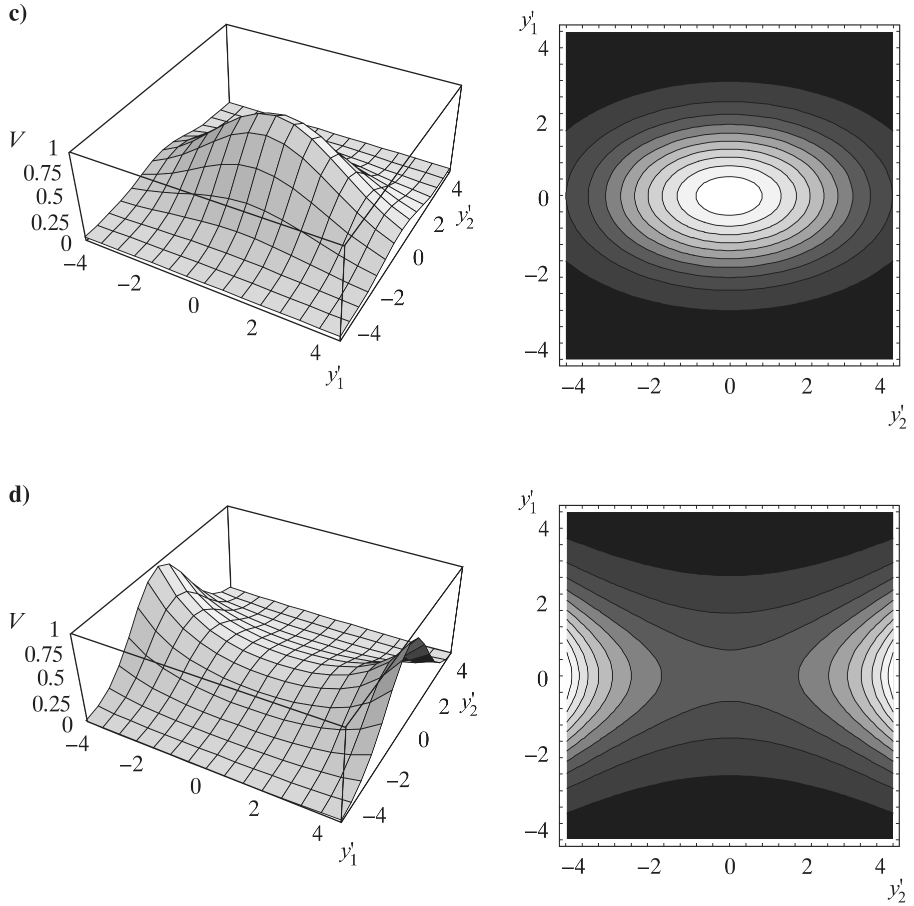


Fig. 7.8. *Continued.*

molecule). Such free translations/rotations do not change the energy and may be thought therefore to correspond to zero force constants.

If we are interested in what the particular atoms are doing, when a single mode l is active, then the displacements from the equilibrium position as a function of time are expressed as

$$\boldsymbol{x}_l = \boldsymbol{M}^{-\frac{1}{2}} \boldsymbol{y}_l = \boldsymbol{M}^{-\frac{1}{2}} \boldsymbol{L}_l \sin(\omega_l t). \tag{7.11}$$

A given atom participates in all vibrational modes. Even if any vibrational mode makes all atoms move, some atoms move more than others. It may happen that a particular mode changes mainly the length of *one of the chemical bonds (stretching mode)*, another mode moves another bond, another changes a particular *bond angle (bending mode)*, etc.

Table 7.1. Characteristic frequencies (wave numbers, in cm^{-1}) typical for some chemical bonds (stretching vibrations) and bond angles (bending vibrations). This is of great importance for chemical analysis.

Bond	Vibration	Wave number
C–H	stretching	2850–3400
H–C–H	bending	1350–1460
C–C	stretching	700–1250
C=C	stretching	1600–1700
C≡C	stretching	2100–2250
C=O	stretching	1600–1750
N–H	stretching	3100–3500
O–H	stretching	3200–4000

> This means that some chemical bonds or some functional groups may have *characteristic* vibration frequencies, which is of great importance for the identification of these bonds or groups in chemical analysis.

characteristic frequencies

In Table 7.1 typical ("*characteristic*") frequencies for some particular chemical bonds are reported. Note, that high frequencies correspond to light atoms (e.g., hydrogen). The wave numbers $\bar{\nu}$ are defined by the relation

$$\omega = 2\pi\nu = 2\pi\bar{\nu}c, \tag{7.12}$$

wave number

with c being the velocity of light and ν the frequency. The wave number is the number of the wave lengths covering a distance of 1 cm.

Example 1. *The water molecule*

The goal behind this example is to elaborate ideas associated with various bonds, their characteristic frequencies, and their applicability in chemical analysis.

The *single water* molecule has $3 \times 3 = 9$ normal modes. Six of them have the angular frequencies ω equal zero (they correspond to three free translations and three free rotations of the molecule in space). Three normal modes remain, the vectors $\boldsymbol{x}$ of eq. (7.11) for these modes can be described as follows (Fig. 7.9, the corresponding wave numbers have been given in parentheses[49]):

- one of the modes means a *symmetric* stretching of the two OH bonds ($\bar{\nu}_{\text{sym}} = 3894\ \text{cm}^{-1}$);

---

[49] J. Kim, J.Y. Lee, S. Lee, B.J. Mhin, K.S. Kim, *J. Chem. Phys.* 102 (1995) 310. This paper reports normal mode analysis for potential energy hypersurfaces computed by various methods of quantum chemistry. I have chosen the coupled cluster method (see Chapter 10) CCSD(T) as an illustration.

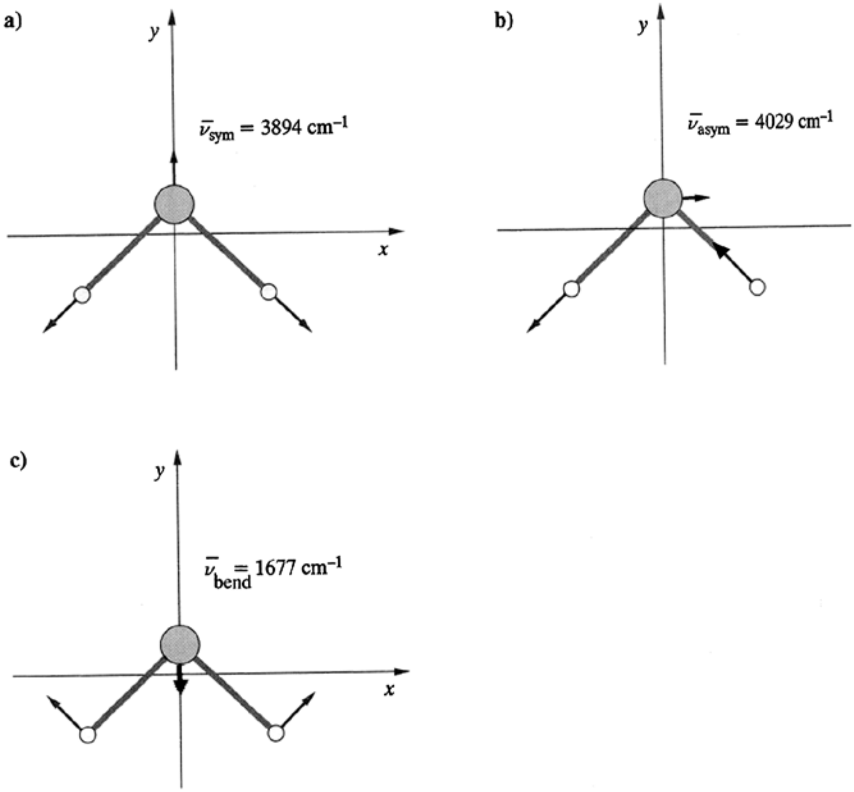


Fig. 7.9. The normal modes of the water molecule: (a) symmetric (b) antisymmetric (c) bending.

- the second mode corresponds to a similar, but *antisymmetric* motion, i.e. when one of the bonds shortens the other one elongates and *vice versa*[50] ($\bar{\nu}_{\mathrm{asym}} = 4029\ \mathrm{cm}^{-1}$);
- the third mode is called the bending mode and corresponds to an oscillation of the HOH angle about the equilibrium value ($\bar{\nu}_{\mathrm{bend}} = 1677\ \mathrm{cm}^{-1}$).

Example 2. *The water dimer*

Now let us take *two interacting water* molecules. First, let us ask how many minima we can find on the electronic ground-state energy hypersurface. Detailed calculations have shown that there are two such minima (Fig. 7.10). The global minimum corresponds to the configuration characteristic for the *hydrogen bond* (cf. p. 746). One of the molecules is a donor, the other is an acceptor of a proton, Fig. 7.10.a. A local minimum of smaller stability appears when one of the water molecules serves as a donor of two protons, while the other serves as an acceptor of them called the *bifurcated* hydrogen bond, Fig. 7.10.b.

hydrogen bond

---

[50] The shortening has the same value as the lengthening. This is a result of the harmonic approximation, in which both shortening and lengthening require the same energy.

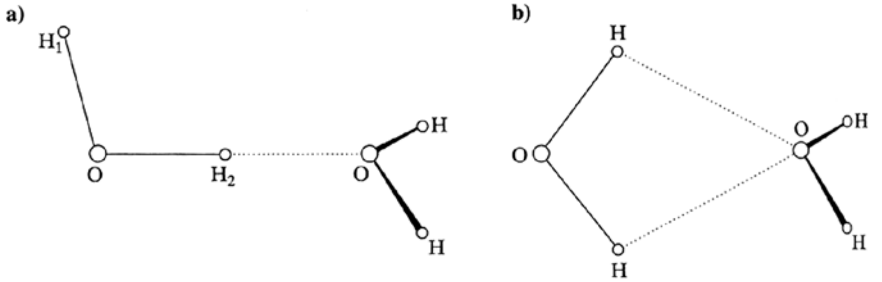


Fig. 7.10. The water dimer and the configurations of the nuclei that correspond to minima of the two basins of the potential energy V. The global minimum (a) corresponds to a single hydrogen bond O–H... O; the local minimum (b) corresponds to the bifurcated hydrogen bond.

bifurcated hydrogen bond

Now, *we decide* to focus on the global minimum potential well. We argue that for thermodynamic reasons, this particular well will be most often represented among water dimers. This potential energy well has to be approximated by a paraboloid. The number of degrees of freedom is equal to $6 \times 3 = 18$ and this is also the number of normal modes to be obtained. As in Example 1, six of them will have zero frequencies and the number of "true" vibrations is 12. This is the number of normal modes, each with its frequency ω_k and the vector $\boldsymbol{x}_k = \boldsymbol{M}^{-\frac{1}{2}} \boldsymbol{L}_k \sin(\omega_k t)$ that describes the atomic motion. The two water molecules, after forming the hydrogen bond, have not lost their individual features (in other words the OH vibration is *characteristic*). In dimer vibrations we will find the vibration frequencies of individual molecules changed a little by the water–water interaction. These modes should appear in pairs, but the two frequencies should differ (the role of the two water molecules in the dimer is different). The computed frequencies[51] are the following:

- two stretching vibrations with frequencies 3924 cm^{-1} (antisymmetric) and 3904 cm^{-1} (nearly antisymmetric), the higher frequency corresponds to the proton acceptor, the lower to the proton donor;
- two stretching vibrations with frequencies 3796 cm^{-1} (symmetric) and 3704 cm^{-1} (nearly symmetric), again the higher frequency corresponds to the proton acceptor, the lower to the proton donor;
- two bending vibrations with frequencies 1624 cm^{-1} (donor bending) and 1642 cm^{-1} (acceptor bending).

The proton acceptor has something attached to its heavy atom, the proton donor has something attached to the light hydrogen atom. Let us recall that in the harmonic oscillator, the reduced mass is relevant, which therefore is almost equal to the mass of the *light* proton. If something attaches to this atom, it means a considerable lowering of the frequency. This is why lower frequencies correspond to the proton donor.

Thus, among 12 modes of the dimer we have discovered six modes which are related to the individual molecules: 4 OH stretching and 2 HOH bending modes.

---

[51] R.J. Reimers, R.O. Watts, *Chem. Phys.* 85 (1984) 83.

Now, we have to identify the remaining 6 modes. These are the intermolecular vibrations (Fig. 7.10.a):

- stretching of the hydrogen bond O–H... O (the vibration of two water molecules treated as entities): 183 cm^{-1}
- bending of the hydrogen bond O–H... O in the plane of the figure: 345 cm^{-1}
- bending of the hydrogen bond O–H... O in the plane perpendicular to the figure: 645 cm^{-1}
- rocking of the hydrogen atom H_1 perpendicular to the figure plane: 115 cm^{-1}
- rocking of the second water molecule (the right-hand side of the figure) in the figure plane: 131 cm^{-1}
- rocking of the second water molecule (the right-hand side of the figure) about its symmetry axis: 148 cm^{-1}.

As we can see, the intermolecular interactions have made the "intramolecular" vibration frequencies decrease,[52] while the "intermolecular" frequencies have very low frequencies. The last effect is, of course, nothing strange, because a change of intermolecular distances does require a small expenditure of energy (which means small force constants). Note, that the simple Morse oscillator model considered in Chapter 4, p. 175, gave the correct order of magnitude of the intermolecular frequency of two water molecules (235 cm^{-1} as compared to the above, much more accurate, result 183 cm^{-1}).

7.5.2 ZERO-VIBRATION ENERGY

The computed minimum of V (using any method, either quantum-mechanical or force field) does not represent the energy of the system for exactly the same reason as the bottom of the parabola (the potential energy) does not represent the energy of the harmonic oscillator (cf. the harmonic oscillator, p. 166). The reason is the kinetic energy contribution.

If all the normal oscillators are in their ground states ($v_j = 0$, called the "zero-vibrations"), then the energy of the system is the energy of the bottom of the parabola V_{min} plus the zero-vibration energy (we assume no rotational contribution)

$$E = V_{\text{min}} + \frac{1}{2}\sum_j (h\nu_j). \tag{7.13}$$

It has been assumed that the vibrations are harmonic in the above formula. This assumption usually makes the frequencies higher by several percent (cf. p. 175).

Taking anharmonicity into account is a much more difficult task than normal mode analysis. Note (Fig. 7.11) that in such a case the position of the minimum

---

[52] This is how the hydrogen bonds behave. This, seemingly natural expectation after attaching an additional mass to a vibrating system is legitimate when assuming that the force constants have not increased. An interesting example of the opposite effect for a wide class of compounds has been reported by Pavel Hobza and Zdenek Havlas (P. Hobza, Z. Havlas, *Chem. Rev.* 100 (2000) 4253).

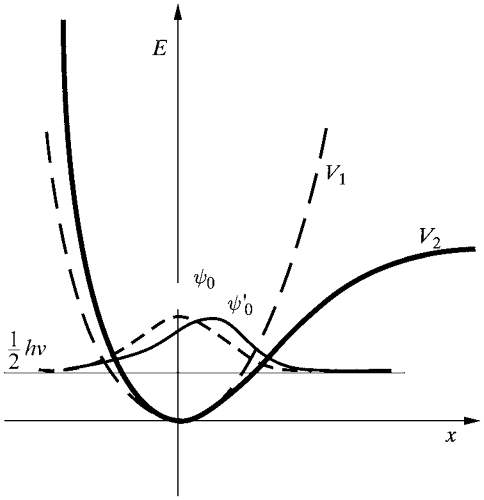


Fig. 7.11. The ground-state vibrational wave function ψ'_0 of the anharmonic oscillator (of potential energy V_2) is *asymmetric and shifted* towards positive values of the displacement when compared to the wave function ψ_0 for the harmonic oscillator with the same force constant (the potential energy V_1).

of V does not correspond to the mean value of the interatomic distance due to the asymmetry of V.

7.6 MOLECULAR DYNAMICS (MD)

In all the methods described above there is no such a thing as temperature. It looks as if all the experiments were made after freezing the lab to 0 K. It is difficult to tolerate such a situation.

7.6.1 THE MD IDEA

Molecular dynamics is a remedy. The idea is very simple.

If we knew the potential energy V as a function of the position ($\boldsymbol{R}$) of all the atoms (whatever force field has been used for the approximation[53]), then all the forces the atoms undergo could be easily computed. If $\boldsymbol{R} = (X_1, X_2, \ldots, X_{3N})^T$ denotes the coordinates of all the N atoms (X_1, X_2, X_3 are the x, y, z coordinates of atom 1, X_4, X_5, X_6 are the x, y, z of atom 2, etc.), then $-\frac{\partial V}{\partial X_1}$ is the x component of the force atom 1 undergoes, $-\frac{\partial V}{\partial X_2}$ is the y component of the same force, etc. When a force field is used, all this can be easily computed even analytically.[54] We had the identical situation in molecular mechanics, but there we were interested just in making these forces equal to zero (through obtaining the equilibrium geometry). In molecular dynamics we are interested in time t, the velocity of the atoms (in this way temperature will come into play) and the acceleration of the atoms.

---

[53] Cf. p. 288.

[54] That is, an analytical formula can be derived.

Our immediate goal is collecting the atomic positions as functions of time, i.e. the system trajectory.

The Newton equation tells us that, knowing the force acting on a body (e.g., an atom), we may compute the acceleration the body undergoes. We have to know the mass, but there is no problem with that.[55] Hence the i-th component of the acceleration vector is equal to

$$a_i = -\frac{\partial V}{\partial X_i} \cdot \frac{1}{M_i} \tag{7.14}$$

for $i = 1, 2, \ldots, 3N$ ($M_i = M_1$ for $i = 1, 2, 3$, $M_i = M_2$ for $i = 4, 5, 6$, etc.).

Now, let us assume that at $t = 0$ all the atoms have the initial coordinates $\boldsymbol{R}_0$ and the initial velocities[56] $\boldsymbol{v}_0$. Now we assume that the forces calculated act *unchanged* during a short period Δt (often 1 femtosecond or 10^{-15} s). We know what should happen to a body (atom) if under influence of a constant force during time Δt. Each atom undergoes a uniformly variable motion and the new position may be found in the vector

$$\boldsymbol{R} = \boldsymbol{R}_0 + \boldsymbol{v}_0 \Delta t + \boldsymbol{a}\frac{\Delta t^2}{2}, \tag{7.15}$$

and its new velocity in the vector

$$\boldsymbol{v} = \boldsymbol{v}_0 + \boldsymbol{a}\Delta t, \tag{7.16}$$

where the acceleration $\boldsymbol{a}$ is a vector composed of the acceleration vectors of all the N atoms

$$\begin{aligned} \boldsymbol{a} &= (\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots \boldsymbol{a}_N)^T, \\ \boldsymbol{a}_1 &= \left(-\frac{\partial V}{\partial X_1}, -\frac{\partial V}{\partial X_2}, -\frac{\partial V}{\partial X_3}\right) \cdot \frac{1}{M_1}, \\ \boldsymbol{a}_2 &= \left(-\frac{\partial V}{\partial X_4}, -\frac{\partial V}{\partial X_5}, -\frac{\partial V}{\partial X_6}\right) \cdot \frac{1}{M_2}, \quad \text{etc.} \end{aligned} \tag{7.17}$$

---

[55] We assume that what moves is the nucleus. In MD we do not worry about that the nucleus moves together with its electrons. To tell the truth both masses differ only by about 0.05%.

[56] Where could these coordinates be taken from? To tell the truth, almost from a "hat". "Almost" – because some essential things will be assumed. First, we may quite reasonably conceive the geometry of a molecule, because we know which atoms form the chemical bonds, their reasonable lengths, the reasonable values of the bond angles, etc. That is, however, not all we would need for larger molecules. What do we take as dihedral angles? This is a difficult case. Usually we take a conformation, which we could call as "reasonable". In a minute we will take a critical look at this problem. The next question is the velocities. Having nothing better at our disposal, we may use a random number generator, assuring however that the velocities are picked out according to the Maxwell–Boltzmann distribution suitable for a given temperature T of the laboratory, e.g., 300 K. In addition, we will make sure that the system does not rotate or flies off somewhere.

In this way we have our starting position and velocity vectors $\boldsymbol{R}_0$ and $\boldsymbol{v}_0$.

All on the right hand side of (7.15) and (7.16) is known. Therefore the new positions and the new velocities are easy to calculate.[57] Now, we may use the new positions and velocities as a start ones and repeat the whole procedure over and over. This makes it possible to go along the time axis in a step-like way in practice reaching even nanosecond times (10^{-9} sec), which means millions of such steps. The procedure described above simply represents the numerical integration of $3N$ differential equations. If $N = 2000$ then the task is impressive. It is so straightforward, because we are dealing with a numerical (not analytical) solution.[58]

7.6.2 WHAT DOES MD OFFER US?

MD trajectory

The computer simulation makes the system evolve from the initial state to the final one. The position $\boldsymbol{R}$ in $3N$-dimensional space becomes a function of time and therefore $\boldsymbol{R}(t)$ represents the *trajectory* of the system in the configurational space. A similar statement pertains to $\boldsymbol{v}(t)$. Knowing the trajectory means that we know the smallest details of the motion of all the atoms. Within the approximations used, we can therefore answer any question about this motion. For example we may ask about some mean values, like the mean value of the total energy, potential energy, kinetic energy, the distance between atom 4 and atom 258, etc. All these quantities may be computed at any step of the procedure, then added up and divided by the number of steps giving the mean values we require. In this way we may obtain the theoretical prediction of the mean value of the interatomic distance and then compare it to, say, the NMR result.

correlation and auto-correlation

In this way we may search for some correlation of motion of some atoms or groups of atoms, i.e. the *space correlation* ("when this group of atoms is shifted to the left, then another group is most often shifted to the right") or the *time correlation* ("when *this* thing happens to the functional group G_1, then after a time τ *that* most often takes place with another functional group G_2") or *time autocorrelation* ("when *this* happens to a certain group of atoms, then after time τ *the same* most often happens to the same group of atoms"). For example, is the x coordinate of atom 1, i.e. X_1 correlated to the coordinate y of atom 41, i.e. X_{122}, or are these two quantities absolutely independent? The answer to this question is given by the correlation coefficient $c_{1,122}$ calculated for M simulation steps in the following way:

$$c_{1,122} = \frac{\frac{1}{M}\sum_{i=1}^{M}(X_{1,i} - \langle X_1\rangle)(X_{122,i} - \langle X_{122}\rangle)}{\sqrt{(\frac{1}{M}\sum_{i=1}^{M}(X_{1,i} - \langle X_1\rangle)^2)(\frac{1}{M}\sum_{i=1}^{M}(X_{122,i} - \langle X_{122}\rangle)^2)}},$$

where $\langle X_1\rangle$ and $\langle X_{122}\rangle$ denote the mean values of the coordinates indicated, and the summation goes over the simulation steps. It is seen that any deviation from

---

[57] In practice we use a more accurate computational scheme called the *leap frog algorithm*.

[58] By the way, if somebody gave us the force field for galaxies (this is simpler than for molecules), we could solve the problem as easily as in our case. This is what astronomers often do.

independence means a non-zero value of $c_{1,122}$. What could be more correlated to the coordinate X_1 than the same X_1 (or $-X_1$)? Of course, absolutely nothing. In such a case (in the formula we replace $X_{122,i} \to X_{1,i}$ and $X_{122} \to X_1$), we obtain $c_{1,1} = 1$ or -1. Hence, c always belongs to $[-1, 1]$, $c = 0$ means independence, $c \pm 1$ means maximum dependence.

Does molecular dynamics have anything to do with reality?

If the described procedure were applied without any modification, then most probably we would have bad luck and our $\boldsymbol{R}_0$ would be located on a slope of the hypersurface V. Then, the solution of the Newton equations would reflect what happens to a point (representing the system) when placed on the slope – it would slide downhill. The point would go faster and faster and soon the vector $\boldsymbol{v}$ would not correspond to the room temperature, but, say, to 500 K. Of course, despite such a high temperature the molecule would not disintegrate, because this is not a real molecule but one operating with a force field that usually corresponds to unbreakable chemical bonds. Although the molecule will not fall apart,[59] such a large T has nothing to do with the temperature of the laboratory. This suggests that after some number of steps we should check whether the atomic velocities still correspond to the proper temperature. If not, it is recommended to scale all the velocities by multiplying them by such a factor in order to make them corresponding again to the desired temperature. For this reason, the only goal of the first part of a molecular dynamics simulation is called the “thermalization”, in which the error connected to the non-zero Δt is averaged and the system is forced stepwise (by scaling) to behave as what is called the canonical ensemble. The canonical ensemble preserves the number of molecules, the volume and the temperature (simulating contact with a thermostat at temperature T). In such a “thermalized” system total energy fluctuations are allowed.

thermalization

canonical ensemble

The thermalization done, the next (main) stage of molecular dynamics, i.e. the harvesting of data (trajectory) begins.

7.6.3 WHAT TO WORRY ABOUT?

- During simulation, the system has to have enough time to wander through all parts of the phase space[60] that are accessible in the experimental conditions (with which the simulation is to be compared). We are never sure that it happens. We have to check whether the computed mean values depend upon the simulation time. If they do not, then very probably everything is all right – we have a description of the equilibrium state.
- The results of the MD (the mean values) should not depend on the starting point, because it has been chosen arbitrarily. This is usually satisfied for small molecules and their collections. For large and flexible molecules we usually start

---

[59] This pertains to a single molecule bound by chemical bonds; a system of several molecules could fall apart.

[60] The Cartesian space of all atomic positions and momenta.

from the vector $\boldsymbol{R}_0$ found from X-ray determined atomic positions. Why? Because *after* the MD we will still stay close to this (all in all experimental) conformation. If the simulation started from another conformation, it would result in a conformation close to this new starting point. This is because even with the most powerful computers, simulation times are too short. In such a way we have a simulation of one conformation evolution rather than a description of the thermodynamic equilibrium.

- The simulation time in the MD is limited on one side by the power of computers and on the other side by the time step Δt, which is not infinitesimally small, and creates an error that cumulates during the simulation (as a result the total energy may vary too much and the system may be heading into non-physical regions of the phase space).

7.6.4 MD OF NON-EQUILIBRIUM PROCESSES

The thermalization is not always what we want. We may be interested in what happens, when a DNA molecule being anchored to a solid surface by one of its end functional groups is distorted by pulling the other end of the molecule. Such MD results may nowadays be compared to the corresponding experiment.

And yet another example. A projectile hits a wall. The projectile is being composed of Lennard-Jones atoms (with some ε_p and $r_{e,p}$, p. 287), we assume the same for the wall (for other values of the parameters, let us make the wall less resistant than the projectile: $\varepsilon_w < \varepsilon_p$ and $r_{e,w} > r_{e,p}$). Altogether we may have hundreds of thousands or even millions of atoms (i.e. millions of differential equations to solve). Now, we prepare the input $\boldsymbol{R}_0$ and $\boldsymbol{v}_0$ data. The wall atoms are assumed to have stochastic velocities drawn from the Maxwell–Boltzmann distribution for room temperature. The same for the projectile atoms, but additionally they have a constant velocity component along the direction pointing to the wall. At first, nothing particularly interesting happens – the projectile flies towards the wall with a constant velocity (while all the atoms of the system vibrate). Of course, the time the projectile hits the wall is the most interesting. Once the front part of the projectile touches the wall, the wall atoms burst into space in a kind of eruption, the projectile's tip loses some atoms as well, the spot on the wall hit by the projectile vibrates and sends a shock wave and concentric waves travelling within the wall. A violent (and instructive) movie.

Among more civil applications, we may think of the interaction of a drill and a steel plate, to plan better drills and better steel plates, as well as about other micro-tools which have a bright future.

7.6.5 QUANTUM-CLASSICAL MD

A typical MD does not allow for breaking bonds and the force fields which allow this give an inadequate, classical picture, so a quantum description is sometimes a must. The systems treated by MD are usually quite large, which excludes a full quantum-mechanical description.

For enzymes (natural catalysts) researchers proposed[61] joining the quantum and the classical description by making the precision of the description dependent on how far the region of focus is from the enzyme active centre (where the reaction the enzyme facilitates takes place). They proposed dividing the system (enzyme + solvent) into three regions:

- region I represents the active centre atoms,
- region II is the other atoms of the enzyme molecule,
- region III is the solvent.

Region I is treated as a quantum mechanical object and described by the proper time-dependent Schrödinger equation, region II is treated classically by the force field description and the corresponding Newton equations of motion, region III is simulated by a continuous medium (no atomic representation) with a certain dielectric permittivity.

The regions are coupled by their interactions: quantum mechanical region I is subject to the external electric field produced by region II evolving according to its MD as well as that of region III, region II feels the charge distribution changes region I undergoes through electrostatic interaction.

7.7 SIMULATED ANNEALING

The goal of MD may differ from simply calculating some mean values, e.g., we may try to use MD to find regions of the configurational space for which the potential energy V is particularly low.[62] From a chemist's point of view, this means trying to find a particularly stable structure (conformation of a single molecule or an aggregate of molecules). To this end, MD is sometimes coupled with an idea of Kirkpatrick et al.,[63] taken from an ancient method of producing metal alloys of exceptional quality (the famous steel of Damascus), and trying to find the minima of arbitrary functions.[64] The idea behind simulated annealing is extremely simple.

This goal is achieved by a series of heating and cooling procedures (called the simulation protocol). First, a starting configuration is chosen that, to the best of our knowledge, is of low energy and the MD simulation is performed at a high temperature T_1. As a result, the system (represented by a point $\boldsymbol{R}$ in the configuration space) rushes through a large manifold of configurations $\boldsymbol{R}$, i.e. wanders over

---

[61] P. Bała, B. Lesyng, J.A. McCammon, in "*Molecular Aspects of Biotechnology: Computational Methods and Theories*", Kluwer Academic Publishers, p. 299 (1992). A similar philosophy stands behind the Morokuma's ONIOM procedure: M. Svensson, S. Humbel, R.D.J. Froese, T. Matsubara, S. Sieber, K. Morokuma, *J. Phys. Chem.* 100 (1996) 19357.

[62] Like in global molecular mechanics.

[63] S. Kirkpatrick, C.D. Gellat Jr., M.P. Vecchi, *Science* 220 (1983) 671.

[64] I recommend a first-class book: W.H. Press, B.P. Flannery, S.A. Teukolsky, W.T. Vetterling, *Numerical Recipes The Art of Scientific Computing*, Cambridge Univ. Press, Cambridge.

a large portion of the hypersurface $V(\boldsymbol{R})$. Then, a lower temperature T_2 is chosen and the motion slows down, the visited portion of the hypersurface shrinks and hopefully corresponds to some regions of low values of V – the system is confined in a large superbasin (the basin composed of individual minima basins). Now the temperature is raised to a certain value $T_3 < T_1$, thus allowing the system eventually to leave the superbasin and to choose another one, maybe of lower energy. While the system explores the superbasin, the system is cooled again, this time to temperature $T_4 < T_2$, and so forth. Such a procedure does not give any guarantee of finding the global minimum of V, but there is a good chance of getting a configuration with much lower energy than the start. The method, being straightforward to implement, is very popular. Its successes are spectacular, although sometimes the results are disappointing. The highly prized swords made in ancient Damascus using annealing, prove that the metal atoms settle down in quasi-optimal positions forming a solid state of low energy – very difficult to break or deform.

7.8 LANGEVIN DYNAMICS

In the MD we solve Newton equations of motion for all atoms of the system. Imagine we have a large molecule in an aqueous solution (biology offers us important examples). We have no chance to solve Newton equations because there are too many of them (a lot of water molecules). What do we do then? Let us recall that we are interested in the macromolecule, the water molecules are interesting only as a medium that changes the conformation of the macromolecule. The changes may occur for many reasons, but the simplest is the most probable – just the fact that the water molecules in their thermal motion hit the atoms of the macromolecule. If so, their role is reduced to a source of chaotic strikes. The main idea behind Langevin dynamics is to ensure that the atoms of the macromolecule indeed feel some random hits from the surrounding medium *without taking this medium into consideration explicitly. This is the main advantage of the method.*

Paul Langevin (1872–1946), French physicist, professor at the College de France. His main achievements are in the theory of magnetism and in relativity theory. His PhD student Louis de Broglie made a breakthrough in quantum theory.

A reasonable part of this problem may be incorporated into the Langevin equation of motion:

$$M_i \ddot{X}_i = -\frac{\partial V}{\partial X_i} + F_i - \gamma_i M_i \dot{X}_i, \tag{7.18}$$

for $i = 1, 2, \ldots, 3N$, where besides the force $-\nabla V$ resulting from the potential energy V for the macromolecule alone, we also have an additional stochastic force $\boldsymbol{F}$, whose magnitude and direction are drawn keeping the force related to the temperature and assuring its isotropic character. The coefficient γ_i is a friction coefficient and the role of friction is proportional to atomic velocity.

The Langevin equations are solved in the same way as those of MD, with the additional stochastic force drawn using a random number generator.

7.9 MONTE CARLO DYNAMICS

Las Vegas, Atlantic City and Monte Carlo are notorious among upright citizens for day and night use of such random number generators as billiards, roulette or cards. Because of this, the idea and even the name of Monte Carlo has been accepted in mathematics, physics, chemistry and biology. The key concept is that a random number, when drawn successively many times, may serve to create a sequence of system snapshots.

All this began from an idea of the mathematician from Lwów, then in Poland (now Lviv in the Ukraine) Stanisław Marcin Ulam.

Perhaps an example will best explain the Monte Carlo method. I have chosen the methodology introduced to the protein folding problem by Andrzej Koliński

Stanisław Ulam (1909–1984), first associated with the University of Lwów, then professor at the Harvard University, University of Wisconsin, University of Colorado, Los Alamos National Laboratory. In Los Alamos Ulam solved the most important bottleneck in hydrogen bomb construction by suggesting that pressure is the most important factor and that sufficient pressure could be achieved by using the *atomic* bomb as a detonator. Using this idea and an idea of Edward Teller about further amplification of the ignition effect by implosion of radiation, both scholars designed the hydrogen bomb. They both own the US patent for H-bomb production.

According to the *Ulam Quarterly Journal* (http://www.ulam.usm.edu/editor.html), Ulam's contribution to science includes logic, set theory, measure theory, probability theory, computer science, topology, dynamic systems, number theory, algebra, algebraic and arithmetic geometry, mathematical biology, control theory, mathematical economy and mathematical physics. He developed and coined the name of the Monte Carlo method, and also the cellular automata method (described at the end of this Chapter). Stanisław Ulam wrote a very interesting autobiography "*Adventures of a Mathematician*".

The picture below shows one of the "magic places" of international science, the *Szkocka Café*, Akademicka street, Lwów, now a bank at Prospekt Szewczenki 27, where, before the World War II, young Polish mathematicians, among them the mathematical genius Stefan Banach, made a breakthrough thereafter called the "Polish school of mathematics".

and Jeffrey Skolnick.[65] In a version of this method we use a simplified model of the real protein molecule, a polymer composed of monomeric peptide units …HN–CO–CHR–…, as a chain of point-like entities HN–CO–CH from which protrude points representing various side chains R. The polymer chain goes through the vertices of a crystallographic lattice (the side chain points can also occupy only the lattice vertices), which excludes a lot of unstable conformations and enable us to focus on those chemically relevant. The lattice representation speeds computation by several orders of magnitude.

The reasoning goes as follows. The non-zero temperature of the water the protein is immersed in makes the molecule acquire random conformations all the time. The authors assumed that a given starting conformation is modified by a series of random micro-modifications. The micro-modifications allowed have to be chosen so as to obey three rules, these have to be:

- chemically/physically acceptable;
- always local, i.e. they have to pertain to a small fragment of the protein, because in future we would like to simulate the kinetics of the protein chain (how a conformational change evolves);
- able to transform any conformation into any other conformation of the protein.

This way we are able to modify the molecular conformation, but we want the protein to move, i.e. to have the *dynamics* of the system, i.e. a sequence of molecular conformations, each one derived from the previous one in a physically acceptable way.

To this end we have to be able to write down the energy of any given conformation. This is achieved by giving the molecule an energy award if the configuration corresponds to intramolecular energy gain (e.g., trans conformation, the possibility of forming a hydrogen bond or a hydrophobic effect, see Chapter 13), and an energy penalty for intramolecular repulsion (e.g., cis conformation, or when two fragments of the molecule are to occupy the same space). It is, in general, better if the energy awards and penalties have something to do with experimental data for final structures, e.g., can be deduced from crystallographic data.[66]

Now we have to let the molecule move. We start from an arbitrarily chosen conformation and calculate its energy E_1. Then, a micro-modification, or even a series of micro-modifications (this way the calculations go faster), is drawn from the micro-modifications list and applied to the molecule. Thus a new conformation is obtained with energy E_2. Now the most important step takes place. We decide to *accept* or to *reject* the new conformation according to the *Metropolis criterion*,[67]

Metropolis criterion

---

[65] J. Skolnick, A. Koliński, *Science* 250 (1990) 1121.

[66] The Protein Data Bank is the most famous. This Data Basis may serve to form what is called the statistical interaction potential. The potential is computed from the frequency of finding two amino acids close in space (e.g., alanine and serine; there are 20 natural amino acids) in the Protein Data Bank. If the frequency is large, we deduce an attraction has to occur between them, etc.

[67] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, E. Teller, *J. Chem. Phys*. 21 (1953) 1087.

which gives the probability of the acceptance as:

$$P_{1\to 2} = \begin{cases} 1 & \text{if } E_2 \leqslant E_1, \\ a = \exp\left(-\frac{(E_2-E_1)}{k_B T}\right) & \text{if } E_2 > E_1. \end{cases}$$

Well, we have a *probability* but what we need is a clear decision: *to be or not to be in state* 2. This is where the Monte Carlo spirit comes in, see Fig. 7.12. By using a random number generator we draw a random number u from section $[0, 1]$ and… compare it with the number a. If $u \leqslant a$, then we accept the new conformation, otherwise conformation 2 is rejected (and we forget about it). The whole procedure is repeated over and over again: drawing micro-modifications → a new conformation → comparison with the old one by the Metropolis criterion → accepting (the new conformation becomes the current one) or rejecting (the old conformation remains the current one), etc.

The Metropolis criterion is one of those mathematical tricks a chemist has to know about. Note that the algorithm always accepts the conformation 2 if $E_2 \leqslant E_1$ and therefore will have a tendency to lower the energy of the current conformation. On the other hand, when $E_2 > E_1$ the algorithm may decide to increase the energy by accepting the higher energy conformation 2. If $\frac{(E_2-E_1)}{k_B T} > 0$ is small, the algorithm accepts the new conformation very easily (Fig. 7.12.a), at a given $E_2 - E_1$ the easier the higher the temperature. On the other hand, an attempt at a very high jump (Fig. 7.12.b) in energy may be successful in practice only at very high temperatures. The algorithm prefers higher energy conformations to the same extent as the Boltzmann distribution. Thus, grinding the mill of the algorithm on and on

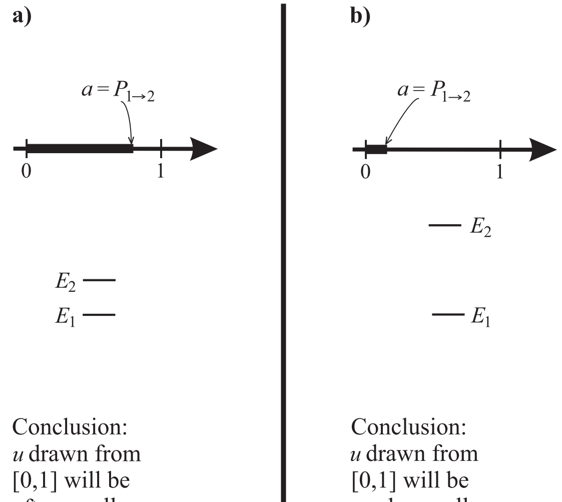


Fig. 7.12. Metropolis algorithm. (a) If E_2 is only a little higher than E_1, then the Metropolis criterion often leads to accepting the new conformation (of energy E_2). (b) On the other hand if the energy difference is large, then the new conformation is accepted only rarely. If the temperature increases, the acceptance rate increases too.

(sometimes it takes months on the fastest computers of the world) and calculating statistics of the number of accepted conformations as a function of energy, we arrive at the Boltzmann distribution as it should be in thermodynamic equilibrium.

Thus as the mill grinds we can make a film. The film would reveal how the protein molecule behaves at high temperature: the protein acquires practically any new conformation generated by the random micro-modifications and it looks as if the molecule is participating in a kind of rodeo. However, we decide the temperature. Thus let us decide to lower the temperature. Until a certain temperature we will not see any qualitative change in the rodeo, but at a sufficiently low temperature we can recognize that something has happened to the molecule. From time to time (time is proportional to the number of simulation steps) some local structures typical of the secondary structures of proteins (the α-helices and the zig-zag type β-strands, the latter like to bind together laterally by hydrogen bonds) emerge and vanish, emerge again etc.

> When the temperature decreases, at a certain critical value, T_{crit}, all of a sudden a stable structure emerges (an analog of the so called native structure, i.e. the one ensuring the molecule can perform its function in nature).

critical temperature

The structure vibrates a little, especially at the ends of the protein, but further cooling does not introduce anything new. The native structure exhibits a unique secondary structure pattern along the polymeric chain (i.e. definite sections of the α and β structures) which packs together into a unique *tertiary structure*. In this way a highly probable scenario for the coil-globular *phase transition* was demonstrated for the first time by Koliński and Skolnick. It seems that predicting the 3D structure of globular proteins uniquely from the sequence of amino acids (an example is shown in Fig. 7.15[68]), will be possible in the near future.

coil-globular transition

7.10 CAR–PARRINELLO DYNAMICS

Despite the fact that the present textbook assumes that the reader has completed a basic quantum chemistry course, the author (according to his declaration in the Introduction) does not profit from this very extensively. Car–Parrinello dynamics is an exception. It positively belongs to the present chapter, while borrowing heavily from the results of Chapter 8. If the reader feels uncomfortable with this, this section may just be omitted.

We have already listed some problems associated with the otherwise nice and powerful MD. We have also mentioned that the force field parameters (e.g., the net atomic charges) do not change when the conformation changes or when two mole-

---

[68] This problem is sometimes called "the second genetic code" in the literature. This name reflects the final task of obtaining information about protein function from the "first genetic code" (i.e. DNA) information that encodes protein production.

cules approach, whereas everything has to change. Car and Parrinello[69] thought of a remedy in order to make the parameters change "in flight".

Let us assume the one-electron approximation.[70] Then the total electronic energy $E_0^0(\boldsymbol{R})$ is (in the adiabatic approximation) not only a function of the positions of the nuclei, but also a functional of the spinorbitals $\{\psi_i\}$: $V = V(\boldsymbol{R}, \{\psi_i\}) \equiv E_0^0(\boldsymbol{R})$.

> The function $V = V(\boldsymbol{R}, \{\psi_i\})$ will be minimized with respect to the positions $\boldsymbol{R}$ of the nuclei and the spinorbitals $\{\psi_i\}$ depending on the electronic coordinates.

If we are going to change the spinorbitals, we have to take care of their orthonormality at all stages of the change.[71] For this reason Lagrange multipliers appear in the equations of motion (Appendix N). We obtain the following set of Newton equations for the motion of M nuclei

$$M_I \ddot{X}_I = -\frac{\partial V}{\partial X_I} \quad \text{for } I = 1, \ldots, 3M$$

and an equation of motion for each spinorbital (each corresponding to the evolution of one electron probability density in time)

$$\mu \ddot{\psi}_i = -\hat{F}\psi_i + \sum_{j=1}^{N} \Lambda_{ij}\psi_j, \quad i = 1, 2, \ldots, N, \tag{7.19}$$

where $\mu > 0$ is a *fictitious parameter*[72] for the electron, $\hat{F}$ is a Fock operator (see Chapter 8, p. 341), and Λ_{ij} are the Lagrange multipliers to assure the orthonormality of the spinorbitals ψ_j.

Both equations are quite natural. The first (Newton equation) says that a nucleus has to move in the direction of the force acting on it ($-\frac{\partial V}{\partial X_I}$) and the larger the force and the smaller the mass, the larger the acceleration achieved. Good! The left hand side of the second equation and the first term on the right hand side say the following: let the spinorbital ψ_i change in such a way that the orbital energy has a tendency to go down (in the sense of the mean value). How on earth does this follow from the equations? From a basic course in quantum chemistry (this will be repeated in Chapter 8) we know, that the orbital energy may be computed as the mean value of the operator $\hat{F}$ with the spinorbital ψ_i, i.e. $\langle\psi_i|\hat{F}\psi_i\rangle$. To focus our attention, let us assume that $\delta\psi_i$ is localized in a small region of space (see Fig. 7.13).

---

[69] R. Car, M. Parrinello, *Phys. Rev. Letters* 55 (1985) 2471.

[70] The approximation will be described in Chapter 8 and consists of assuming the wave function in the form of a single Slater determinant built of orthonormal spinorbitals. Car and Parrinello gave the theory for the density functional theory (DFT). As will be seen in Chapter 11, a single determinant function is also considered.

[71] Because the formulae they satisfy are valid under this condition.

[72] We may call it "mass". In practical applications μ is large, usually taken as a few hundreds of the electron mass, because this assures the agreement of theory and experiment.

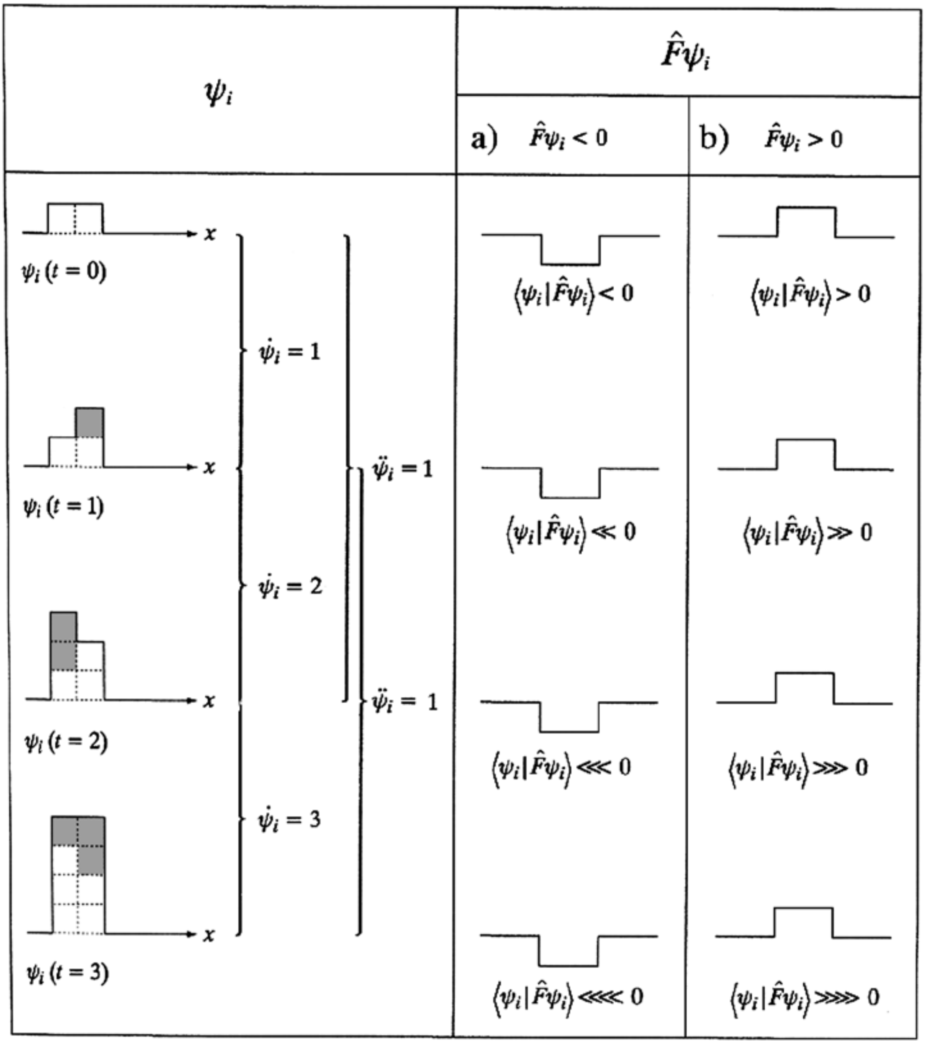


Fig. 7.13. A scheme showing why the acceleration $\ddot{\psi}_i$ of the spinorbital ψ_i has to be of the same sign as that of $-\hat{F}\psi_i$. Time (arbitrary units) goes from up ($t=0$) downwards ($t=3$) where the time step is $\Delta t = 1$. On the left hand side the changes (localized in 1D space, x axis) of ψ_i are shown in a schematic way (in single small square units). It is seen that the velocity of the change is not constant and the corresponding acceleration is equal to 1. Now let us imagine for simplicity that function $\hat{F}\psi_i$ has its non-zero values precisely where $\psi_i \neq 0$ and let us consider two cases: a) $\hat{F}\psi_i < 0$ and b) $\hat{F}\psi_i > 0$. In such a situation we may easily foresee the sign of the mean value of the energy $\langle\psi_i|\hat{F}\psi_i\rangle$ of an electron occupying spinorbital ψ_i. In situation a) the conclusion for changes of ψ_i is: *keep that way* or, in other words, even *increase* the acceleration $\ddot{\psi}_i$ making it proportional to $-\hat{F}\psi_i$. In b) the corresponding conclusion is: *suppress* these changes or in other words *decrease the acceleration* e.g., making it negative as $-\hat{F}\psi_i$. Thus, in *both* cases we have $\mu\ddot{\psi}_i = -\hat{F}\psi_i$, which agrees with eq. (7.19). In both cases there is a trend to lower orbital energy $\varepsilon_i = \langle\psi_i|\hat{F}\psi_i\rangle$.

From Fig. 7.13, it is seen that it would be desirable to have the acceleration $\ddot{\psi}_i$ with the same sign as $-\hat{F}\psi_i$. This is equivalent to increase the changes that lower the corresponding orbital energy, and to suppress the changes that make it higher. The ψ_i spinorbitals obtained in the numerical integration have to be corrected for orthonormality, as is assured by the second term in (7.19).

The prize for the elegance of the Car–Parrinello method is the computation time, which allows one to treat systems currently up to a few hundreds of atoms (while MD may even deal with a million of atoms). The integration interval has to be decreased by a factor of 10 (i.e. 0.1 fs instead of 1 fs), which allows us to reach simulation times of the order of 10–100 picoseconds instead of (in classical MD) nanoseconds.

7.11 CELLULAR AUTOMATA

Another powerful tool for chemists is the cellular automata method invented by John (in his Hungarian days Janos) von Neumann[73] and Stanisław Marcin Ulam (under the name of "cellular spaces"). The cellular automata are mathematical models in which space and time both have a granular structure (similar to Monte Carlo simulations on lattices, in MD only time has such a structure). A cellular automaton consists of a periodic lattice of cells (nodes in space). In order to describe the system locally, we assume that every cell has its "state" representing a vector of N components. Each component is a Boolean variable, i.e. a variable having a logical value (e.g., "0" for "*false*" and "1" for "*true*").

A cellular automaton evolves using some *propagation and collision (or actualization) rules* that are always of a *local character*. The local character means that (at a certain time step t and a certain cell) the variables change their values depending only on what happened at *the* cell and at its *neighbours* at time step $t-1$. The propagation rules dictate what would happen next (for each cell) with variables on the cell and on the nearest neighbour cells for each cell independently. But this may provoke a collision of the rules, because a Boolean variable on a cell may be forced to change by the propagation rules related to two or more cells. We need a unique decision and this comes from the collision, or actualization, rules.

For physically relevant states, the propagation and collision rules for the behaviour of such a set of cells as time goes on, may mirror what would happen with a physical system. This is why cellular automata are appealing. Another advantage is that due to the locality mentioned above, the relevant computer programs may be effectively parallelized, which usually significantly speeds up computations. The most interesting cellular automata are those for which the rules are of a non-linear character (cf. Chapter 15).

---

[73] His short biography is in Chapter 6.

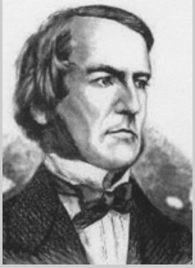
Example 3. *Gas lattice model*

One of the simplest examples pertains to a lattice model of a gas. Let the lattice be regular two-dimensional (Fig. 7.14).

Propagation rules:

There are a certain number of point-like particles of equal mass which may occupy the nodes (cells) only and have unit velocities pointing either in North–South or East–West directions, thus reaching the next row or column after a unit of time. We assign each cell a state which is a four-dimensional vector of Boolean variables. The first component tells us whether there is a particle moving North on the node (Boolean variables take 0 or 1), the second moving East, the third South and the fourth West. There should be no more than one particle going in one direction at the node, therefore a cell may correspond to 0, 1, 2, 3, 4 particles. Any particle is shifted by one unit in the direction of the velocity vector (in unit time).

George Boole (1815–1864), British mathematician and logician. Despite the fact that he was self-taught, he became professor of Mathematics at Queen's College in Cork and a member of the Royal Society. In 1854 Boole wrote his Opus Magnum "*An Investigation of the Laws of Thought*", creating a domain of mathematical logic. The logic was treated there as a kind of algebra.



Collision rules:

If two particles are going to occupy the same state component at the same cell, the two particles are annihilated and a new pair of particles is created with drawn

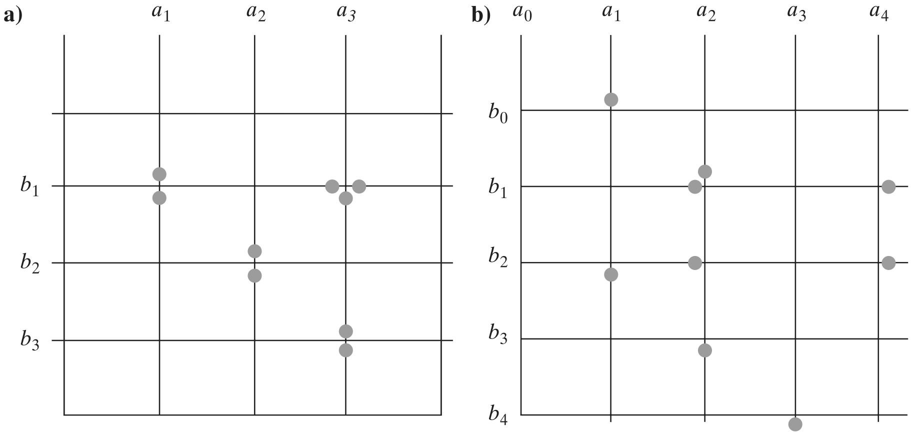


Fig. 7.14. Operation of a cellular automaton – a model of gas. The particles occupy the lattice nodes (cells). Their displacement from the node symbolizes which direction they are heading in with the velocity equal to 1 length unit per 1 time step. On the left scheme (a) the initial situation is shown. On the right scheme (b) the result of the one step propagation and one step collision is shown. Collision only take place in one case (at a_3b_2) and the collision rule has been applied (of the lateral outgoing). The game would become more dramatic if the number of particles were larger, if the walls of the box as well as the appropriate propagation rules (with walls) were introduced.

positions and velocities. Any two particles which meet at a node with opposite velocities acquire the velocities that are opposite to each other and perpendicular to the old ones (the "lateral outgoing", see Fig. 7.14).

This primitive model has nevertheless an interesting property. It turns out that such a system attains a thermodynamic equilibrium state. No wonder that this approach with more complex lattices and rules became popular. Using the cellular automata we may study an extremely wide range of phenomena, such as turbulent flow of air along a wing surface, electrochemical reactions, etc. It is a simple and powerful tool of general importance.

Summary

- A *detailed* information about a molecule (in our case: a three atom complex C...AB) may be obtained making use of the potential energy hypersurface for the nuclear motion computed as the ground-state electronic energy (however, even in this case simplified: the AB distance has been frozen). After constructing the basis functions appropriate for the 5 variables and applying the Ritz method of solving the Schrödinger equation we obtain the rovibrational levels and corresponding wave functions for the system. This allows us to compute the IR and microwave spectrum, and as it turns out, this agrees very well with the experimental data, which confirms the high quality of the hypersurface and of the basis set used.
- We may construct an approximation to the potential energy hypersurface for the motion of the nuclei by designing what is called a *force field*, or a simple expression for the electronic energy as a function of the position of the nuclei. Most often in proposed force fields we assume harmonicity of the chemical bonds and bond angles ("springs"). The hypersurface obtained often has a complex shape with many local minima.
- *Molecular mechanics* (should have the adjective "local") represents
  - choice of the starting configuration of the nuclei (a point in the configuration space),
  - sliding slowly downhill from the point (configuration) to the "nearest" local minimum, which corresponds to a stable conformation with respect to small displacements in the configurational space.
- *Global molecular mechanics* means
  - choice of the starting configuration of the nuclei,
  - finding the global (the lowest-energy) minimum, i.e. the most stable configuration of the nuclei.

While the local molecular mechanics represents a standard procedure, the global one is still *in statu nascendi*.

- Any of the potential energy minima can be approximated by a paraboloid. Then, for N nuclei, we obtain $3N - 6$ *normal modes* (i.e. harmonic and having the same phase) of the molecular vibrations. This represents important information about the molecule, because it is sufficient to calculate the IR and Raman spectra (cf. p. 903). Each of the normal modes makes all the atoms move, but some atoms move more than others. It often happens that a certain mode is dominated by the vibration of a particular bond or functional group and therefore the corresponding frequency is *characteristic* for this bond or functional group, which may be very useful in chemical analysis.

- Molecular mechanics does not involve atomic kinetic energy, *molecular dynamics* (MD) does. MD represents a method of solving the Newton equations of motion[74] for all the atoms of the system. The forces acting on each atom at a given configuration of the nuclei are computed (from the potential energy V assumed to be known[75]) as $\boldsymbol{F}_j = -\nabla_j V$ for atoms $j = 1, 2, \ldots, N$. The forces known, we calculate the acceleration vector, and from that the velocities and the new positions of the atoms after a unit time. The system starts to evolve, as time goes on. Important ingredients of the MD procedure are:
  - choice of starting conformation,
  - choice of starting velocities,
  - thermalization at a given temperature (with velocity adjustments to fulfil the appropriate Maxwell–Boltzmann distribution),
  - harvesting the system trajectory,
  - conclusions derived from the trajectory.
- In MD (also in the other techniques listed below) there is the possibility of applying a sequence (protocol) of cooling and heating periods in order to achieve a low-energy configuration of the nuclei (*simulated annealing*). The method is very useful and straightforward to apply.
- Besides MD, there are other useful techniques describing the motion of the system:
  - Langevin dynamics that allows the surrounding solvent to be taken into account, inexpensively.
  - Monte Carlo dynamics – a powerful technique basing on drawing and then accepting/rejecting random configurations by using the *Metropolis criterion*. The criterion says that if the energy of the new configuration is lower, the configuration is accepted, if it is higher, it is accepted with a certain probability.
  - Car–Parrinello dynamics allows for the electron structure to be changed "in flight", when the nuclei move.
  - cellular automata – a technique of general importance, which divides the total system into cells. Each cell is characterized by its state being a vector with its components being Boolean variables. There are propagation rules that change the state, as time goes on, and collision rules, which solve conflicts of the propagation rules. Both types of rules have a local character. Cellular automata evolution may have many features in common with thermodynamic equilibrium.

Main concepts, new terms

Jacobi coordinate system (p. 279)
angular momenta addition (p. 281)
rovibrational spectrum (p. 283)
dipole moment (p. 283)
sum of states (p. 283)
force field (p. 284)
Lennard-Jones potential (p. 287)
torsional potential (p. 288)
molecular mechanics (p. 290)
global optimization (p. 292)
global minimum (p. 292)
kinetic minimum (p. 293)
thermodynamic minimum (p. 293)
free energy (p. 293)
entropy (p. 293)
normal modes (p. 294)
characteristic frequency (p. 300)
molecular dynamics (p. 304)
spatial correlation (p. 306)
time correlation (p. 306)
autocorrelation (p. 306)
thermalization (p. 307)

---

[74] We sometimes say: integration.

[75] Usually it is a force field.

simulated annealing (p. 309)	Metropolis algorithm (p. 312)
cooling protocol (p. 309)	Car–Parrinello algorithm (p. 314)
Langevin dynamics (p. 310)	cellular automata(p. 317)
Monte Carlo dynamics (p. 311)	Boolean variables (p. 317)

From the research front

The number of atoms taken into account in MD may nowadays reach a million. The real problem is not the size of the system, but rather its complexity and the wealth of possible structures, with too large a number to be investigated. Some problems may be simplified by considering a quantum-mechanical part in the details and a classical part described by Newton equations. Another important problem is to predict the 3D structure of proteins, starting from the available amino acid sequence. Every two years from 1994 a CASP (Critical Assessment of techniques for protein Structure Prediction) has been organized in California. CASP is a kind of scientific competition, in which theoretical laboratories (knowing only the amino acid sequence) make blind predictions about 3D protein structures about to be determined in experimental laboratories, see Fig. 7.15. Most of the theoretical methods are based on the similarity of the sequence to a sequence from the Protein Data Bank of the 3D structures, only some of the methods are related to chemical-physics.[76]

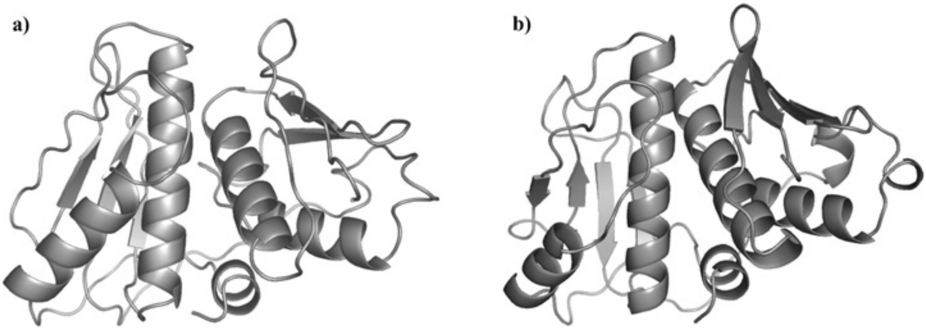


Fig. 7.15. One of the target proteins in the 2004 CASP6 competition. The 3D structure (in ribbon representation) obtained for the putative nitroreductase, one of the 1877 proteins of the bacterium *Thermotoga maritima*, which lives in geothermal marine sediments. The energy expression which was used in theoretical calculations takes into account the physical interactions (such as hydrogen bonds, hydrophobic interactions, etc., see Chapter 13) as well as an empirical potential deduced from representative protein experimental structures deposited in the Brookhaven Protein Data Bank (no bias towards the target protein). The molecule represents a chain of 206 amino acids, i.e. about 3000 heavy atoms. Both theory (CASP6 blind prediction) and experiment (carried out within CASP6 as well) give the target molecule containing five α-helices and two β-pleated sheets (wide arrows). These secondary structure elements interact and form the unique (native) tertiary structure, which is able to perform its biological function. (a) predicted by A. Kolinski (to be published) by the Monte Carlo method, and (b) determined experimentally by X-ray diffraction. Both structures in atomic resolution differ (rms) by 2.9 Å. Reproduced by courtesy of Professor Andrzej Koliński.

---

[76] More details, e.g., in http://predictioncenter.llnl.gov/casp6/Casp6.html

Ad futurum…

The maximum size of the systems investigated by the MM and MD methods will increase systematically to several million atoms in the near future. A critical problem will be the choice of the system to be studied as well as the question to be asked. Very probably non-equilibrium systems will become more and more important, e.g., concerning impact physics, properties of materials subject to various stresses, explosions, self-organization (see Chapter 13), and first of all chemical reactions. At the same time the importance of MD simulations of micro-tools of dimensions of tens of thousands Å will increase.

Additional literature

A.R. Leach, "Molecular Modelling. Principles and Applications", Longman, 1996.
A "Bible" of theoretical simulations.

M.P. Allen, D.J. Tildesley, "Computer Simulations of Liquids", Oxford Science Publications, Clarendon Press, Oxford, 1987.
A book with a more theoretical flavour.

Questions

1. A three-atomic system C…AB with the Hamiltonian

$$\hat{H} = -\frac{\hbar^2}{2\mu R^2}\frac{\mathrm{d}}{\mathrm{d}R}R^2\frac{\mathrm{d}}{\mathrm{d}R} + \frac{\hat{l}^2}{2\mu R^2} + \frac{\hat{j}^2}{2\mu_{AB}r_{\mathrm{eq}}^2} + V.$$

   The symbols $R, r_{\mathrm{eq}}, \hat{j}^2$ denote:
   a) $R =$ CA distance, $r_{\mathrm{eq}} =$ AB distance, $\hat{j}^2$ operator of the square of the angular momentum of C with respect to the centre of mass of AB;
   b) $R =$ AB distance, $r_{\mathrm{eq}} =$ distance of C from the centre of mass of AB, $\hat{j}^2$ operator of the square of the angular momentum of C with respect to the centre of mass of AB;
   c) $R =$ distance of C from the centre of mass of AB, $r_{\mathrm{eq}} =$ AB distance, $\hat{j}^2$ operator of the square of the angular momentum of AB with respect to the centre of mass of AB;
   d) $R =$ distance of C from the centre of mass of AB, $r_{\mathrm{eq}} =$ AB distance, $\hat{j}^2$ operator of the square of the angular momentum of C with respect to the centre of mass of AB.

2. A force field represents an approximation to:
   a) the ground-state electronic energy as a function of nuclear configuration; b) vibrational wave function as a function of nuclear configuration; c) potential energy of nuclear repulsion; d) electric field produced by the molecule.

3. Frequencies of the normal modes:
   a) pertain to a particular potential energy minimum and correspond to a quadratic dependence of the potential on the displacement from the equilibrium; b) do not depend on the local minimum; c) take into account a small anharmonicity of the oscillators; d) are identical to H_2 and HD, because both PESs are identical.

4. The Lennard-Jones potential corresponds to

$$V = \varepsilon\left[\left(\frac{r_0}{r}\right)^{12} - 2\left(\frac{r_0}{r}\right)^{6}\right],$$

where:
a) ε is the dissociation energy, r_0 is the well depth of V; b) $(\frac{dV}{dr})_{r=r_0} = -\varepsilon$; c) ε represents the dissociation energy, r_0 is the distance for which $V = 0$; d) ε is the well depth, r_0 is the position of the minimum of V.

5. In equation $(\boldsymbol{A} - \omega_k^2 \mathbf{1})\boldsymbol{L}_k = \mathbf{0}$ for the normal modes ($\boldsymbol{M}$ is the diagonal matrix of the atomic masses):
   a) ω_k^2 may be imaginary; b) ω_k^2 represents an eigenvalue of $\boldsymbol{M}^{-\frac{1}{2}}\boldsymbol{V}''\boldsymbol{M}^{-\frac{1}{2}}$, where $\boldsymbol{V}''$ is the Hessian computed at the minimum of the potential energy; c) the vector $\boldsymbol{L}_k$ is the k-th column of the Hessian $\boldsymbol{V}''$; d) ω_k is an eigenvalue of $\boldsymbol{M}^{-\frac{1}{2}}\boldsymbol{V}''\boldsymbol{M}^{-\frac{1}{2}}$, where $\boldsymbol{V}''$ is the Hessian computed in the minimum of the potential energy.

6. The most realistic set of the wave numbers (cm^{-1}) corresponding to vibrations of the chemical bonds: C–H, C–C, C=C, respectively, is:
   a) 2900, 1650, 800; b) 800, 2900, 1650; c) 1650, 800, 2900; d) 2900, 800, 1650.

7. The goal of the simulated annealing in MD is:
   a) to lower the temperature of the system; b) to find the most stable structure; c) to adjust the atomic velocities to the Maxwell–Boltzmann distribution; d) thermalization for a given temperature.

8. In the Metropolis algorithm within the Monte Carlo method (for temperature T) a new configuration is accepted:
   a) on condition that its energy is higher; b) always; c) always, when its energy is lower, and sometimes when its energy is higher; d) only when its energy is higher than kT.

9. In the Langevin MD the solvent molecules:
   a) are treated on the same footing as the solute molecules; b) cause a resistance to the molecular motion and represent a source of random forces; c) cause a resistance to the molecular motion and represent the only source of forces acting on the atoms; d) cause a friction proportional to the acceleration and represent a source of random forces.

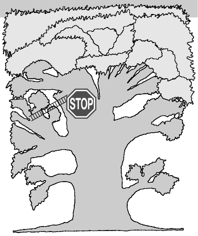
10. In Car–Parrinello dynamics:
    a) when the nuclei move the atomic net charges change; b) we minimize the conformational energy in a given force field; c) nuclei and electrons move according to the same equations of motion; d) nuclei move while the electronic charge distribution is “frozen”.

Answers

1c, 2a, 3a, 4d, 5b, 6d, 7b, 8c, 9b, 10a

Chapter 8

ELECTRONIC MOTION IN THE MEAN FIELD: ATOMS AND MOLECULES



WHERE ARE WE?

Where are we?

We are in the upper part of the main trunk of the TREE.

An example

What is the electronic structure of atoms? How do atoms interact in a molecule? Two *neutral* moieties (say, hydrogen atoms) attract each other with a force *of a similar order of magnitude* to the Coulombic forces between two ions. This is quite surprising. What pulls these neutral objects to one another? These questions are at the foundations of chemistry.

What is it all about

The Born–Oppenheimer (or adiabatic) approximation is the central point of this book (note its position in the TREE). Thanks to the approximation, we can consider *separately* two *coupled* problems concerning molecules:

- the motion of the electrons at fixed positions of the nuclei (to obtain the electronic energy),
- the motion of nuclei in the potential representing the electronic energy of the molecule (see Chapter 7).

From now on we will concentrate on the *motion of the electrons at fixed positions of the nuclei* (the Born–Oppenheimer approximation, p. 229).

To solve the corresponding eq. (6.18), we have at our disposal the variational and the perturbation methods. The latter one should have a reasonable starting point (i.e. an unperturbed system). This is not the case in the problem we want to consider at the moment. Thus, only the variational method remains. If so, a class of trial functions should be proposed. In this chapter the trial wave function will have a very specific form, bearing significant importance for the theory. We mean here what is called Slater determinant, which is composed of molecular orbitals. At a certain level of approximation, each molecular orbital is a "home" for two electrons. We will now learn on how to get the optimum molecular orbitals (Hartree–Fock method). Despite some quite complex formulas, which will appear below, the main idea behind them is extremely simple. It can be expressed in the following way.

Let us consider the road traffic, the cars (electrons) move at fixed positions of buildings (nuclei). The motion of the cars proves to be very complex (as it does for the electrons) and therefore the problem is extremely difficult. How can such a motion be described in an approximate way? To describe such a complex motion one may use the so called mean field approximation (paying the price of poor quality). In the mean field approximation method we focus on the motion of *one car only* considering its motion in such way that the car avoids *those streets that are usually most jammed*. In this chapter we will treat the electrons in a similar manner (leaving the difficulties of considering the correlation of the motions of the electrons to Chapter 10). Now, the electrons will not feel the true electric field of the other electrons (as it should be in a precise approach), but rather their *mean* electric field, i.e. averaged over their motions.

Translating it into quantum mechanical language, the underlying assumptions of the mean field method for the N identical particles (here: electrons) are as follows:

- there is a certain "effective" *one-particle* operator $\hat{F}(i)$ of an identical mathematical form for all particles $i = 1, 2, \ldots, N$, which has the eigenfunctions φ_k, i.e. $\hat{F}\varphi_k = \varepsilon_k \varphi_k$, such that
- $\langle\Psi|\hat{H}\Psi\rangle \approx \langle\tilde{\Psi}|\hat{H}^{ef}\tilde{\Psi}\rangle$, where $\tilde{\Psi}$ is an approximate wave function (to the exact wave function Ψ, both functions normalized) for the total system, $\hat{H}$ is the electronic Hamiltonian (in the clamped nuclei approximation, Chapter 6), and $\hat{H}^{ef} = \sum_{i=1}^{N} \hat{F}(i)$. In such a case the eigenvalue equation $\hat{H}^{ef} \prod_{i=1}^{N} \varphi_i(i) = E_0 \prod_{i=1}^{N} \varphi_i(i)$ holds, and the approximate total energy is equal to $E_0 = \sum_{i=1}^{N} \varepsilon_k$, as if the particles were independent.

Any mean field method needs to solve two problems:

- How should $\tilde{\Psi}$ be constructed using N eigenfunctions φ_k?
- What is the form of the one-particle effective operator $\hat{F}$?

These questions will be answered in the present chapter.

Such effectively independent, yet interacting particles, are called quasiparticles or – as we sometimes use to say – bare particles dressed up by the interaction with others.

It is worth remembering that the mean field method bears several different names in chemistry:

- one-determinant approximation,
- one-electron approximation,
- one-particle approximation,
- molecular orbital method,
- independent-particle approximation,
- mean field approximation,
- Hartree–Fock method,
- self-consistent field method (as regards practical solutions).

It will be shown how the mean field method implies that mile-stone of chemistry: the periodic table of chemical elements.

Next, we will endeavour to understand *why* two atoms create a chemical bond, and also what affects the ionization energy and the electron affinity of a molecule.

Then, still within the molecular orbital scheme, we will show how we can reach a localized description of a molecule, with chemical bonds between *some* atoms, with the inner electronic shells, and the lone electronic pairs. The last terms are elements of a rich and very useful language commonly used by chemists.

Why this is important?

Contemporary quantum chemistry uses better methods than the mean field, described in this chapter. We will get to know them in Chapters 10 and 11. Yet all these methods *start from the mean field approximation* and in most cases they only perform cosmetic changes in energy and in electron distribution. For example, the methods described here yield about 99% of the total energy of a system.[1] There is one more reason why this chapter is important. Methods beyond the one-electron approximation are – computationally – very time-consuming (hence they may be applied only to small systems), while the molecular orbital approach is the "daily bread" of quantum chemistry. It is a sort of standard method, and the standards have to be learnt.

What is needed?

- Postulates of quantum chemistry (Chapter 1, necessary).
- Operator algebra, Hermitian operators (Appendix B, p. 895, necessary).
- Complete set of functions (Chapter 1, necessary).
- Hilbert space (Appendix B, p. 895, recommended).
- Determinants (Appendix A, p. 889, absolutely necessary).
- Slater–Condon rules (Appendix M, p. 986, only the results are needed).
- Lagrange multipliers (Appendix N, p. 997, necessary).
- Mulliken population analysis (Appendix S, p. 1015, occasionally used).

Classical works

This chapter deals with the basic theory explaining electronic structure of atoms and molecules. This is why we begin by Dimitrii Ivanovich Mendeleev who discovered in 1865, when writing his book "*Osnovy Khimii*" ("*Principles of Chemistry*"), St Petersburg, Tovarishchestvo Obshchestvennaya Polza, 1869–71, his famous periodic table of elements – one of the greatest human achievements. ★ Gilbert Newton Lewis in the paper "*The Atom and the Molecule*" published in the *Journal of the American Chemical Society*, 38 (1916) 762 and Walter Kossel in an article "*Über die Molekülbildung als Frage des Atombaus*" published in *Annalen der Physik*, 49 (1916) 229, introduced such important theoretical tools as the octet rule and stressed the importance of the noble gas electronic configurations. ★ As soon as quantum mechanics was formulated in 1926, Douglas R. Hartree published several papers in the *Proceedings of the Cambridge Philosophical Society*, 24 (1927) 89, 24 (1927) 111, 26 (1928) 89, entitled "*The Wave Mechanics of an Atom with a Non-Coulomb Central Field*", containing the computations for atoms such large as Rb and Cl. These were *self-consistent ab initio*[2] computations…, and the wave function was assumed to be the *product* of spinor-

---

[1] In physics and chemistry we are seldom interested in the total energy. The energy differences of various states are of importance. Sometimes such precision is not enough, but the result speaks for itself.

[2] That is, derived from the first principles of (non-relativistic) quantum mechanics! Note, that these young people worked incredibly fast (no e-mail, no PCs).

bitals. ★ The LCAO approximation (for the solid state) was introduced by Felix Bloch in his PhD thesis "*Über die Quantenmechanik der Elektronen in Kristallgittern*", University of Leipzig, 1928, and three years later Erich Hückel used this method to describe the first molecule (benzene) in a publication "*Quantentheoretische Beitrage zum Benzolproblem. I. Die Elektronenkonfiguration des Benzols*", which appeared in *Zeitschrift für Physik*, 70 (1931) 203. ★ Vladimir Fock introduced the antisymmetrization of the spinorbital product in his publication "*Näherungsmethode zur Lösung des quantenmechanischen Mehrkörperproblems*" in *Zeitschrift für Physik*, 61 (1930) 126 and *ibid*. 62 (1930) 795. ★ John Slater proposed the idea of the multi-configurational wave function ("*Cohesion in Monovalent Metals*", *Physical Review*, 35 (1930) 509). ★ The Hartree–Fock method in the LCAO approximation was formulated by Clemens C.J. Roothaan in his work "*New Developments in Molecular Orbital Theory*" published in the *Reviews of Modern Physics*, 23 (1951) 69, and, independently, by George G. Hall in a paper "*The Molecular Orbital Theory of Chemical Valency*" in *Proceedings of the Royal Society (London)*, A205 (1951) 541. ★ The physical interpretation of the orbital energies in the Hartree–Fock method was given by Tjalling C. Koopmans in his *only* quantum chemical paper "*On the Assignment of Wave Functions and Eigenvalues to the Individual Electron of an Atom*" published in *Physica*, 1 (1933/1934) 104. ★ The first localized orbitals (for the methane molecule) were computed by Charles A. Coulson despite the difficulties of war time (*Transactions of the Faraday Society*, 38 (1942) 433). ★ Hideo Fukutome, first in *Progress in Theoretical Physics*, 40 (1968) 998, and then in several following papers, analyzed general solutions for the Hartree–Fock equations from the symmetry viewpoint, and showed exactly eight classes of such solutions.

In the previous chapter the motion of the nuclei was considered. In the Born–Oppenheimer approximation (Chapter 6) the motion of the nuclei takes place in the potential, which is the electronic energy of a system (being a function of the nuclei position, $\boldsymbol{R}$, in the configurational space). The electronic energy $E_k^0(\boldsymbol{R})$ is an eigenvalue given in eq. (6.8) (adapted to the polyatomic case, hence $R \to \boldsymbol{R}$): $\hat{H}_0\psi_k(\boldsymbol{r};\boldsymbol{R}) = E_k^0(\boldsymbol{R})\psi_k(\boldsymbol{r};\boldsymbol{R})$. We will now deal exclusively with this equation, i.e. we will consider the electronic motion at fixed positions of the nuclei (clamped nuclei). Thus, our goal is two-fold: we are interested in what the electronic structure looks like and in how the electronic energy depends on the positions of the nuclei.[3]

Any theoretical method applicable to molecules may be also used for atoms, albeit very accurate wave functions, even for simple atoms, are not easy to calculate.[4] In fact for atoms we know the solutions quite well only in the mean field approximation, i.e. the atomic orbitals. Such orbitals play an important role as building blocks of many-electron wave functions.

---

[3] In the previous chapter the ground-state electronic energy $E_0^0(\boldsymbol{R})$ was denoted as $V(\boldsymbol{R})$.

[4] If an atom is considered in the Born–Oppenheimer approximation, the problem is even simpler, the electronic equation also holds; we can then take, e.g., $\boldsymbol{R} = \boldsymbol{0}$. People still try to compute correlated wave functions (i.e. beyond the mean field approximation, see Chapter 10) for heavier atoms. Besides, relativistic effects (see Chapter 3) play increasingly important roles for such atoms. Starting with magnesium, they are larger than the correlation corrections. Fortunately, the relativistic corrections for atoms are largest for the inner electronic shells, which are the least important for chemists.

8.1 HARTREE–FOCK METHOD – A BIRD'S EYE VIEW

Douglas R. Hartree (1897–1958) was born and died in Cambridge. He was a British mathematician and physicist, professor at Manchester University, and then professor of mathematical physics at Cambridge. Until 1921 his interest was in the development of numerical methods for anti-aircraft artillery (he had some experience from the 1st World War), but a lecture by Niels Bohr has completely changed his career. Hartree immediately started investigating atoms. He used the atomic wave function in the form of the spinorbital *product*. Hartree learnt to use machines to solve differential equations while in Boston, and then he built one for himself at Cambridge. The machine was invented by Lord Kelvin, and constructed by Vannevar Bush in the USA. The machine integrated equations using a circle which rolled on a rotating disc. Later the first electronic computer, ENIAC, was used, and Hartree was asked to come and help to compute missile trajectories. An excerpt from "*Solid State and Molecular Theory*", Wiley, London, 1975 by John C. Slater: "*Douglas Hartree was very distinctly of the matter-of-fact habit of thought that I found most congenial. The hand-waving magical type of scientist regarded him as a "mere computer". Yet he made a much greater contribution to our knowledge of the behaviour of real atoms than most of them did. And while he limited himself to atoms, his demonstration of the power of the self-consistent field for atoms is what has led to the development of that method for molecules and solids as well*".

Before introducing the detailed formalism of the Hartree–Fock method, let us first look at its principal features. It will help us to understand our mathematical goal.

First of all, the positions of the nuclei are frozen (Born–Oppenheimer approximation) and then we focus on the wave function of N *electrons*. Once we want to move nuclei, we need to repeat the procedure from the beginning (for the new position of the nuclei).

Vladimir A. Fock (1898–1974), Russian physicist, professor at the Leningrad University (Sankt Petersburg), led investigations on quantum mechanics, gravity theory, general relativity theory, and in 1930, while explaining atomic spectra, invented the *antisymmetrization* of the spinorbitals product.

8.1.1 SPINORBITALS

Although this comparison is not precise, the electronic wave function for a molecule is built of segments, as a house is constructed from bricks.

The electronic wave function of a molecule containing N electrons depends on $3N$ Cartesian coordinates of the electrons and on their N spin coordinates (for each electron, its $\sigma = \frac{1}{2}$ or $-\frac{1}{2}$). Thus, it is a function of position in $4N$-dimensional space. This function will be created out of simple "bricks", i.e. *molecular spinorbitals*. Each of those will be a function of the coordinates of *one* electron only:

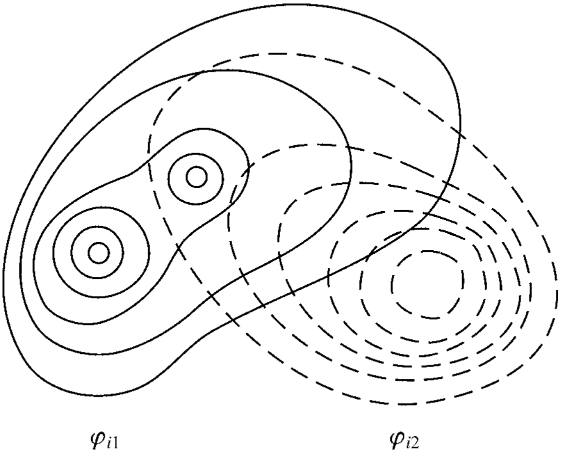


Fig. 8.1. According to eq. (8.1) a spinorbital is a mixture of α and β orbital components: $\varphi_{i1}(\boldsymbol{r})$ and $\varphi_{i2}(\boldsymbol{r})$, respectively. Figure shows two sections of such a spinorbital (z denotes the Cartesian axis perpendicular to the plane of the page): section $z = 0, \sigma = \frac{1}{2}$ (solid isolines) and section $z = 0$, $\sigma = -\frac{1}{2}$ (dashed isolines). In practical applications most often a restricted form of spinorbitals is used: either $\varphi_{i1} = 0$ or $\varphi_{i2} = 0$, i.e. a spinorbital is taken as an orbital part times spin function α or β.

three Cartesian coordinates and one spin coordinate (cf. Chapter 1). A spinorbital is therefore a function of the coordinates in the 4D space,[5] and in the most general case a normalized spinorbital reads as (Fig. 8.1)

$$\boxed{\phi_i(\boldsymbol{r}, \sigma) = \varphi_{i1}(\boldsymbol{r})\alpha(\sigma) + \varphi_{i2}(\boldsymbol{r})\beta(\sigma), \qquad (8.1)}$$

where the orbital components φ_{i1} and φ_{i2} (square-integrable functions) that depend on the position $\boldsymbol{r}$ of the electron can adopt *complex* values, while the spin functions α and β, which depend on the spin coordinate σ, are defined in Chapter 1, p. 28. In the vast majority of quantum mechanical calculations the spinorbital ϕ_i is a *real* function, and φ_{i1} and φ_{i2} are such that either $\varphi_{i1} = 0$ or $\varphi_{i2} = 0$. *Yet for the time being we do not introduce any significant*[6] *restrictions for the spinorbitals.* Spinorbital ϕ_i will adopt different complex values for various spatial coordinates as well as for a given value[7] of the spin coordinate σ.

8.1.2 VARIABLES

Thus the variables, on which the wave function depends, are as follows:

$x_1,\ y_1,\ z_1,\ \sigma_1$ or briefly 1,
$x_2,\ y_2,\ z_2,\ \sigma_2$ or briefly 2,
....
$x_N,\ y_N,\ z_N,\ \sigma_N$ or briefly N,

where x_i, y_i, z_i are the Cartesian coordinates and σ_i is the spin coordinate of electron i.

The true wave function ψ (i.e. the eigenfunction of the Hamiltonian $\hat{H}$) belongs (see Fig. 8.2) to the set Ω which is the common part of the following sets:

---

[5] The analogy of a house and bricks fails here, because both the house and the bricks come from the same 3D space.

[6] The normalization condition does not reduce the generality of the approach.

[7] That is, we put $\sigma = \frac{1}{2}$ or $\sigma = -\frac{1}{2}$.

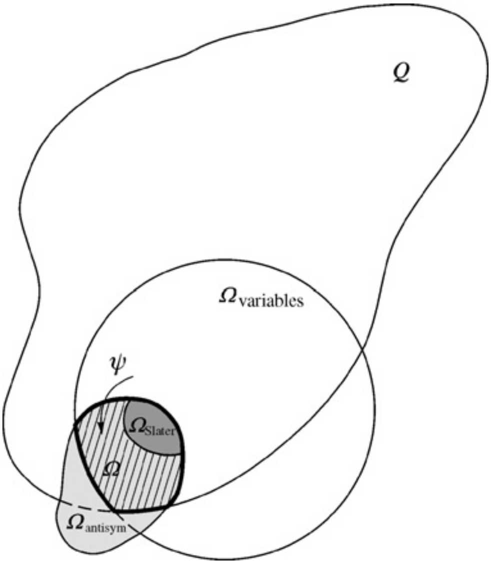


Fig. 8.2. Diagram of the sets, among which the solution ψ of the Schrödinger equation is sought. The Q set is the one of all square-integrable functions, $\Omega_{\text{variables}}$ is the set of the functions with variables as those of the solution of Schrödinger equation, ψ, and Ω_{antisym} is the set of the functions which are antisymmetric with respect to the exchange of coordinates of any two electrons. The solutions of the Schrödinger equation, ψ, will be sought in the common part of these three sets: $\psi \in \Omega = Q \cap \Omega_{\text{variables}} \cap \Omega_{\text{antisym}}$. The Ω_{Slater} represents the set of single Slater determinants built of normalizable spinorbitals. The exact wave function ψ always belongs to $\Omega - \Omega_{\text{Slater}}$.

- set Q of all square-integrable functions,
- set $\Omega_{\text{variables}}$ of all the functions dependent on the above mentioned variables,
- set Ω_{antisym} of all the functions which are antisymmetric with respect to the mutual exchange of the coordinates of any two electrons (p. 33).

$$\psi \in \Omega - \Omega_{\text{Slater}} \quad \text{with } \Omega = Q \cap \Omega_{\text{variables}} \cap \Omega_{\text{antisym}}.$$

John C. Slater (1901–1976), American physicist, for 30 years a professor and dean at the Physics Department of the Massachusetts Institute of Technology, then at the University of Florida Gainesville and the Quantum Theory Project at this school. His youth was in the stormy period of the intense development of quantum mechanics, and he participated vividly in it. For example, in 1926–1932 he published articles on the ground state of the helium atom, on the screening constants (Slater orbitals), on the antisymmetrization of the wave function (Slater determinant), and on the algorithm for calculating the integrals (the Slater–Condon rules). In this period he made the acquaintance of John Van Vleck, Robert Mulliken, Arthur Compton, Edward Condon and Linus Pauling. In Europe (Zurich and Leipzig) he exchanged ideas with Werner Heisenberg, Friedrich Hund, Peter Debye, Felix Bloch, Douglas Hartree, Eugene Wigner, Albert Einstein, Erich Hückel, Edward Teller, Nevil Mott, and John Lennard-Jones. The frequency of the appearance of his name in this book is the best testament to his great contribution to quantum chemistry.

8.1.3 SLATER DETERMINANTS

There should be something in the theory which assures us that, if we renumber the electrons, no theoretical prediction will change. The postulate of the *antisymmetric* character of the wave function with respect to the *exchange* of the coordinates of *any* two electrons, certainly ensures this (Chapter 1, p. 28). The solution of the Schrödinger equation for a given stationary state of interest should be sought amongst *such* functions.

A Slater determinant is a function of the coordinates of N electrons, which *automatically* belongs to Ω:

$$\psi = \frac{1}{\sqrt{N!}} \begin{vmatrix} \phi_1(1) & \phi_1(2) & \dots & \phi_1(N) \\ \phi_2(1) & \phi_2(2) & \dots & \phi_2(N) \\ \dots & \dots & \dots & \dots \\ \phi_N(1) & \phi_N(2) & \dots & \phi_N(N) \end{vmatrix},$$

where ϕ_i are the orthonormal[8] one-electron[9] functions, i.e. molecular spinorbitals. The Slater determinants form a subset $\Omega_{\text{Slater}} \subset \Omega$.

A Slater determinant carries two important attributes of the exact wave function:

- Suppose we want to calculate the probability density that two electrons with the same spin coordinate σ are in the same place, i.e. such that two electrons have *all* their coordinates (spatial and spin ones) identical. If so, then the two columns of the above mentioned determinant are identical. And this means that the determinant becomes equal to zero.[10] From this and from the continuity of the wave function we may conclude that:

  > electrons of the same spin cannot approach each other.

- Let us now imagine two electrons with opposite values of their spin coordinate σ. If these two electrons take the same position in space, the Slater determinant will not vanish, because in the general case there is nothing that forces $\phi_i(1)$ to be equal to $\phi_i(2)$, when $1 \equiv (\boldsymbol{r}_1, \sigma = \frac{1}{2})$ and $2 \equiv (\boldsymbol{r}_1, \sigma = -\frac{1}{2})$ for

---

[8] It is *most often* so, and then the factor standing in front of the determinant ensures normalization. The spinorbitals could be *non-normalized* (but, if they are to describe a stationary state, they should be square-integrable). They also do not need to be mutually orthogonal, but certainly they *need to be linearly independent*. Any attempt to insert the linearly-dependent functions in the determinant will have a "tragic outcome" – we will get 0. It comes from the properties of the determinant (if a row is a linear combination of the others, the determinant is zero). It also follows that if we have a set of non-orthogonal spinorbitals in a Slater determinant, we could orthogonalize them by making the appropriate linear combinations. This would multiply the original Slater determinant by an irrelevant constant. This is why *it is no loss of generality to require the spinorbitals to be orthonormal*.

[9] In the theory of the atomic nucleus, the determinant wave function for the nucleons (fermions) is also used.

[10] Indeed, this is why we exist. Two objects built out of fermions (e.g., electrons) cannot occupy the same position in space. If it were not so, our bodies would sink in the ground.

$i = 1, 2, \ldots$. From this, and from the continuity of the wave function, we conclude that:

electrons of opposite spins can approach each other.

8.1.4 WHAT IS THE HARTREE–FOCK METHOD ALL ABOUT?

The Hartree–Fock method is a *variational* one (p. 196) and uses the variational wave function in the form of a *single Slater determinant*.

In other words we seek (among the Ω_{Slater} set of trial functions) the determinant (ψ_{HF}), which results in the lowest mean value of the Hamiltonian.

In this case the mathematical form of the spinorbitals undergoes variation – change $\varphi_{i1}(\boldsymbol{r})$ as well as $\varphi_{i2}(\boldsymbol{r})$ in eq. (8.1) (however you want) to try to lower the mean value of the Hamiltonian as much as possible. The output determinant which provides the *minimum* mean value of the Hamiltonian is called the Hartree–Fock function. The Hartree–Fock function is an *approximation* of the true wave function (which satisfies the Schrödinger equation $H\psi = E\psi$), because the former is indeed the optimal solution, but *only among single Slater determinants*. The Slater determinant is an antisymmetric function, but an antisymmetric function does not necessarily need to take the shape of a Slater determinant.

Taking the variational wave function in the form of one determinant means an automatic limitation to the subset Ω_{Slater} for searching for the optimum wave function. In fact, we should search the optimum wave function in the set $\Omega - \Omega_{\text{Slater}}$. Thus it is an *approximation* for the solution of the Schrödinger equation, with no chance of representing the exact result.

The true solution of the Schrödinger equation is never a single determinant.

Why are Slater determinants used so willingly? There are two reasons for this:

- a determinant is a kind of "template".[11] Whatever you put inside, the result (if not zero) is antisymmetric *by definition*, i.e. it automatically satisfies one of the postulates of quantum mechanics.
- it is constructed out of simple "bricks" – the one-electron functions (spinorbitals).

The Slater determinants built out of the complete set of spinorbitals do form the complete set.

---

[11] An interesting analogy to the history of algebra appears here. The matrix (lat. *matrix*) took its name from the printing stamp, because the latter indeed served the inventor of matrix algebra, James Joseph Sylvester (1814–1897), for automatically "cutting out" the determinants.

Because of this, the true wave function can take the form of a linear combination of the determinants (we will discuss this later in Chapter 10).

8.2 THE FOCK EQUATION FOR OPTIMAL SPINORBITALS

8.2.1 DIRAC AND COULOMB NOTATIONS

The spatial *and* spin coordinate integrals (ϕ are the spinorbitals, φ are the orbitals) in the *Dirac notation* will be denoted with angle brackets $\langle\,\rangle$ ($\hat{h}$ denotes a one-electron operator and r_{12} – the distance between electrons 1 and 2, $\mathrm{d}V_1 = \mathrm{d}x_1\,\mathrm{d}y_1\,\mathrm{d}z_1$, $\mathrm{d}V_2 = \mathrm{d}x_2\,\mathrm{d}y_2\,\mathrm{d}z_2$), for the one-electron integrals:

$$\langle i|\hat{h}|j\rangle \equiv \sum_{\sigma_1}\int \mathrm{d}V_1\,\phi_i^*(1)\hat{h}\phi_j(1) \equiv \int \mathrm{d}\tau_1\,\phi_i^*(1)\hat{h}\phi_j(1), \tag{8.2}$$

and for the two-electron integrals:

$$\begin{aligned}\langle ij|kl\rangle &\equiv \sum_{\sigma_1}\sum_{\sigma_2}\int \mathrm{d}V_1\int \mathrm{d}V_2\,\phi_i^*(1)\phi_j^*(2)\frac{1}{r_{12}}\phi_k(1)\phi_l(2)\\ &\equiv \int \mathrm{d}\tau_1\,\mathrm{d}\tau_2\,\phi_i^*(1)\phi_j^*(2)\frac{1}{r_{12}}\phi_k(1)\phi_l(2).\end{aligned} \tag{8.3}$$

The integrals over the spatial (only) coordinates will be denoted by round brackets ( ), for the one-electron integrals:

$$(i|\hat{h}|j) \equiv \int \mathrm{d}V_1\,\varphi_i^*(1)\hat{h}(1)\varphi_j(1), \tag{8.4}$$

and for the two-electron integrals:

$$(ij|kl) \equiv \int \mathrm{d}V_1\int \mathrm{d}V_2\,\varphi_i^*(1)\,\varphi_j^*(2)\frac{1}{r_{12}}\varphi_k(1)\varphi_l(2). \tag{8.5}$$

This is called Dirac notation (of the integrals).[12]

8.2.2 ENERGY FUNCTIONAL

mean value of the Hamiltonian

Applying the first Slater–Condon rule[13] we get the following equation for *the mean value of Hamiltonian* (without nuclear repulsion) calculated using the normalized

---

[12] Sometimes one uses *Coulomb notation* $(ij|kl)_{\text{Dirac}} \equiv (ik|jl)_{\text{Coulomb}}$, also $\langle ij|kl\rangle_{\text{Dirac}} \equiv \langle ik|jl\rangle_{\text{Coulomb}}$. Coulomb notation emphasizes the physical interpretation of the two electron integral, as the energy of the Coulombic interaction of two charge distributions $\varphi_i^*(1)\varphi_k(1)$ for electron 1 and $\varphi_j^*(2)\varphi_l(2)$ for electron 2. Dirac notation for the two-electron integrals emphasizes the two-electron functions "bra" and "ket" from the general Dirac notation (p. 19). *In the present book we will consequently use Dirac notation* (both for integrals using spinorbitals, and for those using orbitals, the difference being emphasized by the type of bracket). Sometimes the self-explaining notation $\langle i|\hat{h}|j\rangle \equiv \langle\phi_i|\hat{h}|\phi_j\rangle$, etc. will be used.

[13] Appendix M, p. 986; please take a look at this rule (you may leave out its derivation).

Slater one-determinant function ψ, i.e. the energy functional $E[\psi]$

$$E[\psi] = \langle\psi|\hat{H}|\psi\rangle = \sum_{i=1}^{N}\langle i|\hat{h}|i\rangle + \frac{1}{2}\sum_{i,j=1}^{N}\big(\langle ij|ij\rangle - \langle ij|ji\rangle\big), \tag{8.6}$$

where the indices symbolize the spinorbitals, and the symbol $\hat{h}$

$$\hat{h}(1) = -\frac{1}{2}\Delta_1 - \sum_{a=1}^{M}\frac{Z_a}{r_{a1}} \tag{8.7}$$

is the one-electron operator (in atomic units) of the kinetic energy of the electron plus the operator of the nucleus–electron attraction (there are M nuclei).

8.2.3 THE SEARCH FOR THE CONDITIONAL EXTREMUM

We would like to find such spinorbitals ("the best ones"), that *any* change in them leads to an *increase* in energy $E[\psi]$. But the changes of the spinorbitals need to be such that the above formula still holds, and it would hold only by assuming the *orthonormality* of the spinorbitals. This means that there are some constraints for the changed spinorbitals:

$$\langle i|j\rangle - \delta_{ij} = 0 \quad \text{for } i,j = 1,2,\ldots,N. \tag{8.8}$$

Thus we seek the *conditional* minimum. We will find it using the Lagrange multipliers method (Appendix N, p. 997). In this method the equations of the constraints multiplied by the Lagrange multipliers are added to the original function which is to be minimized. Then we minimize the function as if the constraints did not exist.

conditional minimum

We do the same for the functionals. The necessary condition for the minimum is that the variation[14] of

stationary points

$$E - \sum_{ij} L_{ij}\big(\langle i|j\rangle - \delta_{ij}\big)$$

is equal zero.

The variation of a functional is defined as the *linear* part of the functional change coming from a change in the function which is its argument.

---

[14] However, this is not a sufficient condition, because the vanishing of the differential for certain values of independent variables happens not only for minima, but also for maxima and saddle points (*stationary points*).

Variation is an analogue of the differential (the differential is just the linear part of the function's change). Thus we calculate the linear part of a change (variation):

$$\delta\Bigl(E - \sum_{ij} L_{ij}\langle i|j\rangle\Bigr) = 0 \tag{8.9}$$

using the (yet) undetermined Lagrange multipliers L_{ij} and we set the variation equal to zero.[15]

The stationarity condition for the energy functional

It is sufficient to vary *only the functions complex conjugate to the spinorbitals* or only the spinorbitals (cf. p. 197), yet the result is always the same. We decide the first.

Substituting $\phi_i^* \rightarrow \phi_i^* + \delta\phi_i^*$ in (8.6) (and retaining only linear terms in $\delta\phi_i^*$ to be inserted into (8.9)) the variation takes the form (the symbols δi^* and δj^* mean $\delta\phi_i^*$ and $\delta\phi_j^*$)

$$\sum_{i=1}^{N}\Bigl(\langle\delta i|\hat{h}|i\rangle + \frac{1}{2}\sum_{ij}\bigl(\langle\delta i, j|ij\rangle + \langle i, \delta j|ij\rangle - \langle\delta i, j|ji\rangle - \langle i, \delta j|ji\rangle - 2L_{ij}\langle\delta i|j\rangle\bigr)\Bigr) = 0. \tag{8.10}$$

Now we will try to express this in the form:

$$\sum_{i=1}^{N}\langle\delta i|\ldots\rangle = 0.$$

Since the δi^* may be arbitrary, the equation $|\ldots\rangle = 0$ (called the Euler equation in variational calculus), results. This will be our next goal.

Noticing that the sum indices and the numbering of electrons in the integrals are arbitrary we have the following equalities

$$\sum_{ij}\langle i, \delta j|ij\rangle = \sum_{ij}\langle j, \delta i|ji\rangle = \sum_{ij}\langle\delta i, j|ij\rangle,$$

$$\sum_{ij}\langle i, \delta j|ji\rangle = \sum_{ij}\langle j, \delta i|ij\rangle = \sum_{ij}\langle\delta i, j|ji\rangle,$$

and after substitution in the expression for the variation, we get

$$\sum_{i}\Bigl(\langle\delta i|\hat{h}|i\rangle + \frac{1}{2}\sum_{j}\bigl(\langle\delta i, j|ij\rangle + \langle\delta i, j|ij\rangle - \langle\delta i, j|ji\rangle - \langle\delta i, j|ji\rangle - 2L_{ij}\langle\delta i|j\rangle\bigr)\Bigr) = 0. \tag{8.11}$$

---

[15] Note that $\delta(\delta_{ij}) = 0$.

Let us rewrite this equation in the following manner:

$$\sum_i \left\langle \delta i \middle| \left(\hat{h}\phi_i(1) + \sum_j \left(\int \mathrm{d}\tau_2 \frac{1}{r_{12}} \phi_j^*(2)\phi_j(2)\phi_i(1) \right.\right.\right.$$
$$\left.\left.\left. - \int \mathrm{d}\tau_2 \frac{1}{r_{12}} \phi_j^*(2)\phi_i(2)\phi_j(1) - L_{ij}\phi_j(1) \right)\right)\right\rangle_1 = 0, \quad (8.12)$$

where $\langle \delta i|\ldots\rangle_1$ means integration over coordinates of electron 1 and $\mathrm{d}\tau_2$ refers to *the spatial coordinate integration and spin coordinate summing for electron 2*. The above must be true for *any* $\delta i^* \equiv \delta\phi_i^*$, which means that *each individual term* in parentheses needs to be equal to zero:

$$\hat{h}\phi_i(1) + \sum_j \left(\int \mathrm{d}\tau_2 \frac{1}{r_{12}} \phi_j^*(2)\phi_j(2) \cdot \phi_i(1) - \int \mathrm{d}\tau_2 \frac{1}{r_{12}} \phi_j^*(2)\phi_i(2) \cdot \phi_j(1) \right)$$
$$= \sum_j L_{ij}\phi_j(1). \quad (8.13)$$

The Coulombic and exchange operators

Let us introduce the following linear operators:

Coulombic and exchange operators

a) *two Coulombic operators: the total operator* $\hat{J}(1)$ and *the orbital operator* $\hat{J}_j(1)$, defined *via* their action on an arbitrary function $u(1)$ of the coordinates of electron 1

$$\hat{J}(1)u(1) = \sum_j \hat{J}_j(1)u(1) \quad (8.14)$$

$$\hat{J}_j(1)u(1) = \int \mathrm{d}\tau_2 \frac{1}{r_{12}} \phi_j^*(2)\phi_j(2)u(1) \quad (8.15)$$

b) and similarly, *two exchange operators: the total operator* $\hat{K}(1)$ and *the orbital operator* $\hat{K}_j(1)$

$$\hat{K}(1)u(1) = \sum_j \hat{K}_j(1)u(1) \quad (8.16)$$

$$\hat{K}_j(1)u(1) = \int \mathrm{d}\tau_2 \frac{1}{r_{12}} \phi_j^*(2)u(2)\phi_j(1). \quad (8.17)$$

Then eq. (8.13) takes the form

$$\left(\hat{h}(1) + \hat{J}(1) - \hat{K}(1)\right)\phi_i(1) = \sum_j L_{ij}\phi_j(1). \quad (8.18)$$

The equation is nice and concise except for one thing. It would be even nicer if the right-hand side were proportional to $\phi_i(1)$ instead of being a linear combination of all the spinorbitals. In such a case the equation would be similar to the eigenvalue problem and we would like it a lot. It would be similar but not identical, since the operators $\hat{J}$ and $\hat{K}$ include the *sought* spinorbitals ϕ_i. Because of this, the equation would be called the pseudo-eigenvalue problem.

8.2.4 A SLATER DETERMINANT AND A UNITARY TRANSFORMATION

How can we help? Let us notice that we do not care too much about the spinorbitals themselves, because these are *by-products* of the method which is to give the optimum mean value of the Hamiltonian, and the corresponding N-electron wave function. We can choose some other spinorbitals, such that the mean value of the Hamiltonian does not change and the Lagrange multipliers matrix is diagonal. Is this at all possible? Let us see.

Let us imagine the linear transformation of spinorbitals ϕ_i, i.e. in matrix notation:

$$\boldsymbol{\phi}' = \boldsymbol{A}\boldsymbol{\phi}, \tag{8.19}$$

where $\boldsymbol{\phi}$ and $\boldsymbol{\phi}'$ are vertical vectors containing components ϕ_i. A vertical vector is uncomfortable for typography, in contrast to its transposition (a horizontal vector), and it is easier to write the transposed vector: $\boldsymbol{\phi}'^T = [\,\phi'_1,\quad \phi'_2,\quad \dots,\quad \phi'_N\,]$ and $\boldsymbol{\phi}^T = [\,\phi_1,\quad \phi_2,\quad \dots,\quad \phi_N\,]$. If we construct the determinant built of spinorbitals $\boldsymbol{\phi}'$ and not of $\boldsymbol{\phi}$, an interesting chain of transformations will result:

$$\begin{aligned}
&\frac{1}{\sqrt{N!}}\begin{vmatrix} \phi'_1(1) & \phi'_1(2) & \dots & \phi'_1(N) \\ \phi'_2(1) & \phi'_2(2) & \dots & \phi'_2(N) \\ \dots & \dots & \dots & \dots \\ \phi'_N(1) & \phi'_N(2) & \dots & \phi'_N(N) \end{vmatrix} \\
&= \frac{1}{\sqrt{N!}}\begin{vmatrix} \sum_i A_{1i}\phi_i(1) & \dots & \sum_i A_{1i}\phi_i(N) \\ \sum_i A_{2i}\phi_i(1) & \dots & \sum_i A_{2i}\phi_i(N) \\ \dots & \dots & \dots \\ \sum_i A_{Ni}\phi_i(1) & \dots & \sum_i A_{Ni}\phi_i(N) \end{vmatrix} \qquad (8.20) \\
&= \det\left\{ \boldsymbol{A}\frac{1}{\sqrt{N!}}\begin{bmatrix} \phi_1(1) & \phi_1(2) & \dots & \phi_1(N) \\ \phi_2(1) & \phi_2(2) & \dots & \phi_2(N) \\ \dots & \dots & \dots & \dots \\ \phi_N(1) & \phi_N(2) & \dots & \phi_N(N) \end{bmatrix} \right\} \\
&= \det \boldsymbol{A} \cdot \frac{1}{\sqrt{N!}}\begin{vmatrix} \phi_1(1) & \phi_1(2) & \dots & \phi_1(N) \\ \phi_2(1) & \phi_2(2) & \dots & \phi_2(N) \\ \dots & \dots & \dots & \dots \\ \phi_N(1) & \phi_N(2) & \dots & \phi_N(N) \end{vmatrix}. \qquad (8.21)
\end{aligned}$$

We have therefore obtained our initial Slater determinant multiplied by a *number*: $\det \boldsymbol{A}$. Thus, provided that $\det \boldsymbol{A}$ is not zero,[16]

> the new wave function would provide the same mean value of the Hamiltonian.

The only problem from such a transformation is loss of the normalization of the wave function. Yet we may even preserve the normalization. Let us choose such a matrix $\boldsymbol{A}$, that $|\det \boldsymbol{A}| = 1$. This condition will hold if $\boldsymbol{A} = \boldsymbol{U}$, where $\boldsymbol{U}$ is a *unitary* matrix.[17] This means that

> if a unitary transformation $\boldsymbol{U}$ is performed on the orthonormal spinorbitals (when $\boldsymbol{U}$ is real, we call $\boldsymbol{U}$ an *orthogonal* transformation), then the new spinorbitals ϕ' are also orthonormal.

This is why a unitary transformation is said to represent a rotation in the Hilbert space: the mutually orthogonal and perpendicular vectors do not lose these features upon rotation.[18] This can be verified by a direct calculation:

$$\begin{aligned}\langle \phi_i'(1)|\phi_j'(1)\rangle &= \left\langle \sum_r U_{ir}\phi_r(1)\middle|\sum_s U_{js}\phi_s(1)\right\rangle \\ &= \sum_{rs} U_{ir}^* U_{js}\langle \phi_r(1)|\phi_s(1)\rangle = \sum_{rs} U_{ir}^* U_{js}\delta_{rs} \\ &= \sum_r U_{ir}^* U_{jr} = \delta_{ij}.\end{aligned}$$

Thus, in the case of a unitary transformation even the normalization of the total one-determinant wave function is preserved; at worst the phase χ of this function will change (while $\exp(i\chi) = \det \boldsymbol{U}$), and this factor does not change either $|\psi|^2$ or the mean value of the operators.

8.2.5 INVARIANCE OF THE $\hat{J}$ AND $\hat{K}$ OPERATORS

How does the Coulombic operator change upon a unitary transformation of the spinorbitals? Let us see,

$$\hat{J}(1)'\chi(1) = \int d\tau_2 \frac{1}{r_{12}} \sum_j \phi_j'^*(2)\phi_j'(2)\chi(1)$$

---

[16] The $\boldsymbol{A}$ transformation thus cannot be singular (see Appendix A, p. 889).

[17] For a unitary transformation $\boldsymbol{U}\boldsymbol{U}^\dagger = \boldsymbol{U}^\dagger\boldsymbol{U} = \mathbf{1}$. The matrix $\boldsymbol{U}^\dagger$ arises from $\boldsymbol{U}$ *via* the exchange of rows and columns (this does not influence the value of the determinant), and *via* the complex conjugation of all elements (and this gives $\det \boldsymbol{U}^\dagger = (\det \boldsymbol{U})^*$). Finally, since $(\det \boldsymbol{U})(\det \boldsymbol{U}^\dagger) = 1$ we have $|\det \boldsymbol{U}| = 1$.

[18] Just as three fingers held at right angles do not cease to be of the same length (normalization) after rotation of your palm and continue to be orthogonal.

$$= \int \mathrm{d}\tau_2 \frac{1}{r_{12}} \sum_j \sum_r U_{jr}^* \phi_r^*(2) \sum_s U_{js} \phi_s(2) \chi(1)$$

$$= \int \mathrm{d}\tau_2 \frac{1}{r_{12}} \sum_{r,s} \Big(\sum_j U_{js} U_{jr}^* \Big) \phi_r^*(2) \phi_s(2) \chi(1)$$

$$= \int \mathrm{d}\,\tau_2 \frac{1}{r_{12}} \sum_{r,s} \Big(\sum_j U_{rj}^\dagger U_{js} \Big) \phi_r^*(2) \phi_s(2) \chi(1)$$

$$= \int \mathrm{d}\tau_2 \frac{1}{r_{12}} \sum_{r,s} \delta_{sr} \phi_r^*(2) \phi_s(2) \chi(1)$$

$$= \int \mathrm{d}\tau_2 \frac{1}{r_{12}} \sum_r \phi_r^*(2) \phi_r(2) \chi(1) = \hat{J}(1) \chi(1).$$

The operator $\hat{J}(1)'$ proves to be *identical* with the operator $\hat{J}(1)$. Similarly we may prove *the invariance of the operator K*.

> The operators $\hat{J}$ and $\hat{K}$ are invariant with respect to any unitary transformation of the spinorbitals.

In conclusion, while deriving the new spinorbitals from a unitary transformation of the old ones, we do not need to worry about $\hat{J}$ and $\hat{K}$ since they remain the same.

8.2.6 DIAGONALIZATION OF THE LAGRANGE MULTIPLIERS MATRIX

Eq. (8.18) may be written in matrix form:

$$\big[\hat{h}(1) + \hat{J}(1) - \hat{K}(1)\big] \boldsymbol{\phi}(1) = \boldsymbol{L} \boldsymbol{\phi}(1), \tag{8.22}$$

where $\boldsymbol{\phi}$ is a column of spinorbitals. Transforming $\boldsymbol{\phi} = \boldsymbol{U}\boldsymbol{\phi}'$ and multiplying the Fock equation by $\boldsymbol{U}^\dagger$ (where $\boldsymbol{U}$ is a unitary matrix), we obtain

$$\boldsymbol{U}^\dagger \big[\hat{h}(1) + \hat{J}(1) - \hat{K}(1)\big] \boldsymbol{U} \boldsymbol{\phi}(1)' = \boldsymbol{U}^\dagger \boldsymbol{L} \boldsymbol{U} \phi(1)', \tag{8.23}$$

because $\hat{J}$ and $\hat{K}$ did not change upon the transformation.

> The $\boldsymbol{U}$ matrix *may* be chosen such that $\boldsymbol{U}^\dagger \boldsymbol{L} \boldsymbol{U}$ is the *diagonal* matrix.

Its diagonal elements[19] will now be denoted as ε_i. Because $\hat{h}(1) + \hat{J}(1) - \hat{K}(1)$ is a linear operator we get equation

---

[19] Such diagonalization is possible because $\boldsymbol{L}$ is a Hermitian matrix (i.e. $\boldsymbol{L}^\dagger = \boldsymbol{L}$), and each Hermitian matrix may be diagonalized *via* the transformation $\boldsymbol{U}^\dagger \boldsymbol{L} \boldsymbol{U}$ with the unitary matrix $\boldsymbol{U}$. Matrix $\boldsymbol{L}$ is indeed Hermitian. It is clear when we write the complex conjugate of the variation $\delta(E - \sum_{ij} L_{ij} \langle i|j\rangle) = 0$. This gives $\delta(E - \sum_{ij} L_{ij}^* \langle j|i\rangle) = 0$, because E is real, and after the change of the summation indices $\delta(E - \sum_{ij} L_{ji}^* \langle i|j\rangle) = 0$. Thus, $L_{ij} = L_{ji}^*$, i.e. $\boldsymbol{L} = \boldsymbol{L}^\dagger$.

$$\boldsymbol{U}^\dagger \boldsymbol{U}\big(\hat{h}(1) + \hat{J}(1) - \hat{K}(1)\big)\boldsymbol{\phi}(1)' = \boldsymbol{U}^\dagger \boldsymbol{L}\boldsymbol{U}\phi(1)' \tag{8.24}$$

or alternatively

$$\big(\hat{h}(1) + \hat{J}(1) - \hat{K}(1)\big)\boldsymbol{\phi}(1)' = \boldsymbol{\varepsilon}\boldsymbol{\phi}(1)', \tag{8.25}$$

where $\varepsilon_{ij} = \varepsilon_i \delta_{ij}$.

8.2.7 THE FOCK EQUATION FOR OPTIMAL SPINORBITALS (GENERAL HARTREE–FOCK METHOD – GHF)

We leave out the "prime" to simplify the notation[20] and write *the Fock equation* for a single spinorbital:

THE FOCK EQUATION IN THE GENERAL HARTREE–FOCK METHOD (GHF)

$$\hat{F}(1)\phi_i(1) = \varepsilon_i \phi_i(1), \tag{8.26}$$

where the Fock operator $\hat{F}$ is

$$\hat{F}(1) = \hat{h}(1) + \hat{J}(1) - \hat{K}(1). \tag{8.27}$$

These ϕ_i are called *canonical spinorbitals*, and are the solution of the Fock equation, ε_i is the *orbital energy* corresponding to the spinorbital ϕ_i. It is indicated in brackets that both the Fock operator and the molecular spinorbital depend on the coordinates of one electron only (exemplified as electron 1).[21]

Fock operator

canonical spin-orbitals

orbital energy

---

[20] This means that we finally forget about ϕ' (we pretend that they have never appeared), and we will deal only with such ϕ as correspond to the *diagonal* matrix of the Lagrange multipliers.

[21] The above derivation looks more complex than it really is. The essence of the whole machinery will now be shown as exemplified by two coupled (bosonic) harmonic oscillators, with the Hamiltonian $\hat{H} = \hat{T} + \hat{V}$ where $\hat{T} = -\frac{\hbar^2}{2m_1}\frac{\partial^2}{\partial x_1^2} - \frac{\hbar^2}{2m_2}\frac{\partial^2}{\partial x_2^2}$ and $V = \frac{1}{2}kx_1^2 + \frac{1}{2}kx_2^2 + \lambda x_1^4 x_2^4$, with $\lambda x_1^4 x_2^4$ as the coupling term. Considering the bosonic nature of the particles (the wave function is symmetric, see Chapter 1), we will use $\psi = \phi(1)\phi(2)$ as a variational function, where ϕ is a normalized spinorbital. The expression for the mean value of the Hamiltonian takes the form

$$\begin{aligned}
E[\phi] &= \langle\psi|\hat{H}\psi\rangle = \langle\phi(1)\phi(2)|(\hat{h}(1) + \hat{h}(2))\phi(1)\phi(2)\rangle + \lambda\langle\phi(1)\phi(2)|x_1^4 x_2^4 \phi(1)\phi(2)\rangle \\
&= \langle\phi(1)\phi(2)|\hat{h}(1)\phi(1)\phi(2)\rangle + \langle\phi(1)\phi(2)|\hat{h}(2)\phi(1)\phi(2)\rangle + \lambda\langle\phi(1)|x_1^4\phi(1)\rangle\langle\phi(2)|x_2^4\phi(2)\rangle \\
&= \langle\phi(1)|\hat{h}(1)\phi(1)\rangle + \langle\phi(2)|\hat{h}(2)\phi(2)\rangle + \lambda\langle\phi(1)|x_1^4\phi(1)\rangle\langle\phi(2)|x_2^4\phi(2)\rangle \\
&= 2\langle\phi|\hat{h}\phi\rangle + \lambda\langle\phi|x^4\phi\rangle^2,
\end{aligned}$$

where one-particle operator $\hat{h}(i) = -\frac{\hbar^2}{2m_i}\frac{\partial^2}{\partial x_i^2} + \frac{1}{2}kx_i^2$.

The change of E, because of the variation $\delta\phi^*$, is $E[\phi + \delta\phi] - E[\phi] = 2\langle\phi + \delta\phi|\hat{h}\phi\rangle + \lambda\langle\phi + \delta\phi|x^4\phi\rangle^2 - [2\langle\phi|\hat{h}\phi\rangle + \lambda\langle\phi|x^4\phi\rangle^2] = 2\langle\phi|\hat{h}\phi\rangle + 2\langle\delta\phi|\hat{h}\phi\rangle + \lambda\langle\phi|x^4\phi\rangle^2 + 2\lambda\langle\delta\phi|x^4\phi\rangle\langle\phi|x^4\phi\rangle + \lambda\langle\delta\phi|x^4\phi\rangle^2 - [2\langle\phi|\hat{h}\phi\rangle + \lambda\langle\phi|x^4\phi\rangle^2]$.

Unrestricted Hartree–Fock method (UHF)

The GHF method derived here is usually presented in textbooks as the *unrestricted Hartree–Fock method* (UHF). Despite its name, UHF is not a fully unrestricted method (as the GHF is). In the UHF we assume (cf. eq. (8.1)):

- orbital components φ_{i1} and φ_{i2} are *real* and
- *there is no mixing of the spin functions* α and β, i.e. either $\varphi_{i1}=0$ and $\varphi_{i2}\neq 0$ or $\varphi_{i1}\neq 0$ and $\varphi_{i2}=0$.

8.2.8 THE CLOSED-SHELL SYSTEMS AND THE RESTRICTED HARTREE–FOCK (RHF) METHOD

Double occupation of the orbitals and the Pauli exclusion principle

When the number of electrons is even, the spinorbitals are usually formed out of orbitals in a very easy (and simplified with respect to eq. (8.1)) manner, by multiplication of each orbital by the spin functions[22] α or β:

$$\phi_{2i-1}(\boldsymbol{r},\sigma)=\varphi_i(\boldsymbol{r})\alpha(\sigma) \tag{8.28}$$

$$\phi_{2i}(\boldsymbol{r},\sigma)=\varphi_i(\boldsymbol{r})\beta(\sigma), \quad i=1,2,\ldots,\frac{N}{2}, \tag{8.29}$$

where – as it can be clearly seen – there are twice as few occupied orbitals φ as occupied spinorbitals ϕ (*occupation means that a given spinorbital appears in the Slater determinant*[23]) (see Fig. 8.3). Thus we introduce an artificial *restriction* for spinorbitals (some of the consequences will be described on p. 369). This is why the method is called the Restricted Hartree–Fock.

> There are as many spinorbitals as electrons, and therefore there can be a maximum of two electrons per orbital.

If we wished to occupy a given orbital with more than two electrons, we would need once again to use the spin function α or β when constructing the spinorbitals,

---

Its linear part, i.e. the variation, is $\delta E=2\langle\delta\phi|\hat{h}\phi\rangle+2\lambda\langle\delta\phi|x^4\phi\rangle\langle\phi|x^4\phi\rangle$. The variation $\delta\phi^*$ has, however, to ensure the normalization of ϕ, i.e. $\langle\phi|\phi\rangle=1$. After multiplying by the Lagrange multiplier 2ε, we get the extremum condition $\delta(E-2\varepsilon\langle\phi|\phi\rangle)=0$, i.e. $2\langle\delta\phi|\hat{h}\phi\rangle+2\lambda\langle\delta\phi|x^4\phi\rangle\langle\phi|x^4\phi\rangle-2\varepsilon\langle\delta\phi|\phi\rangle=0$. This may be rewritten as $2\langle\delta\phi|[\hat{h}+\lambda\bar{x}^4x^4-\varepsilon]\phi\rangle=0$, where $\bar{x}^4=\langle\phi|x^4\phi\rangle$, which gives ($\delta\phi^*$ is arbitrary) the Euler equation $[\hat{h}+\lambda\bar{x}^4x^4-\varepsilon]\phi=0$, i.e. the analogue of the Fock equation (8.27): $\hat{F}\phi=\varepsilon\phi$ with the operator $\hat{F}=[\hat{h}+\lambda\bar{x}^4x^4]$. Let us emphasize that the operator $\hat{F}$ is a one-particle operator, *via* the notation $\hat{F}(1)\phi(1)=\varepsilon\phi(1)$, while $\hat{F}(1)=[\hat{h}(1)+\lambda\bar{x}^4x_1^4]$.

It is now clear what the mean field approximation is: the two-particle problem is reduced to a *single*-particle one (denoted as number 1), and the influence of the second particle is *averaged over its positions* ($\bar{x}^4=\langle\phi|x^4\phi\rangle=\langle\phi(2)|x_2^4\phi(2)\rangle$).

[22] It is not necessary, but quite comfortable. This means: $\phi_1=\varphi_1\alpha$, $\phi_2=\varphi_1\beta$, etc.

[23] And only this. When the Slater determinant is written, the electrons lose their identity – they are not anymore distinguishable.

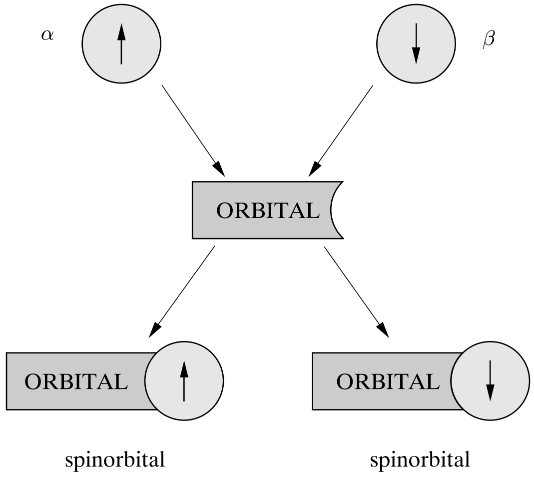


Fig. 8.3. Construction of a spinorbital in the RHF method (i.e. a function x, y, z, σ) as a *product* of an orbital (a function of x, y, z) and one of the two spin functions $\alpha(\sigma)$ or $\beta(\sigma)$.

i.e. repeating a spinorbital. This would imply two identical rows in the Slater determinant, and the wave function would equal zero. This cannot be accepted. The above rule of maximum double occupation is called the Pauli exclusion principle.[24] Such a formulation of the Pauli exclusion principle requires two concepts: the postulate of the antisymmetrization of the electronic wave function, p. 28, and double orbital occupancy. The first of these is of fundamental importance, the second is of a technical nature.[25]

We often assume the double occupancy of orbitals within what is called the *closed shell*. The latter term has an approximate character (Fig. 8.4). It means that for the studied system, there is a large energy difference between HOMO and LUMO orbital energies.

closed shell

HOMO is the Highest Occupied Molecular Orbital, and LUMO is the Lowest Unoccupied Molecular Orbital. The unoccupied molecular orbitals are called *virtual orbitals*.

---

[24] From "*Solid State and Molecular Theory*", Wiley, London, 1975 by John Slater: "… I had a seminar about the work which I was doing over there – the only lecture of mine which happened to be in German. It has appeared that not only Heisenberg, Hund, Debye and young Hungarian PhD student Edward Teller were present, but also Wigner, Pauli, Rudolph Peierls and Fritz London, all of them on their way to winter holidays. Pauli, of course, behaved in agreement with the common opinion about him, and disturbed my lecture saying that "*he had not understood a single word out of it*", but Heisenberg has helped me to explain the problem. (…) Pauli was extremely bound to his own way of thinking, similar to Bohr, who did not believe in the existence of photons. Pauli was a warriorlike man, a kind of dictator…".

[25] The concept of orbitals, occupied by electron pairs, exists only in the mean field method. We will leave this idea in the future, and the Pauli exclusion principle should survive as a postulate of the antisymmetry of the electronic wave function (more generally speaking, of the wave function of fermions).

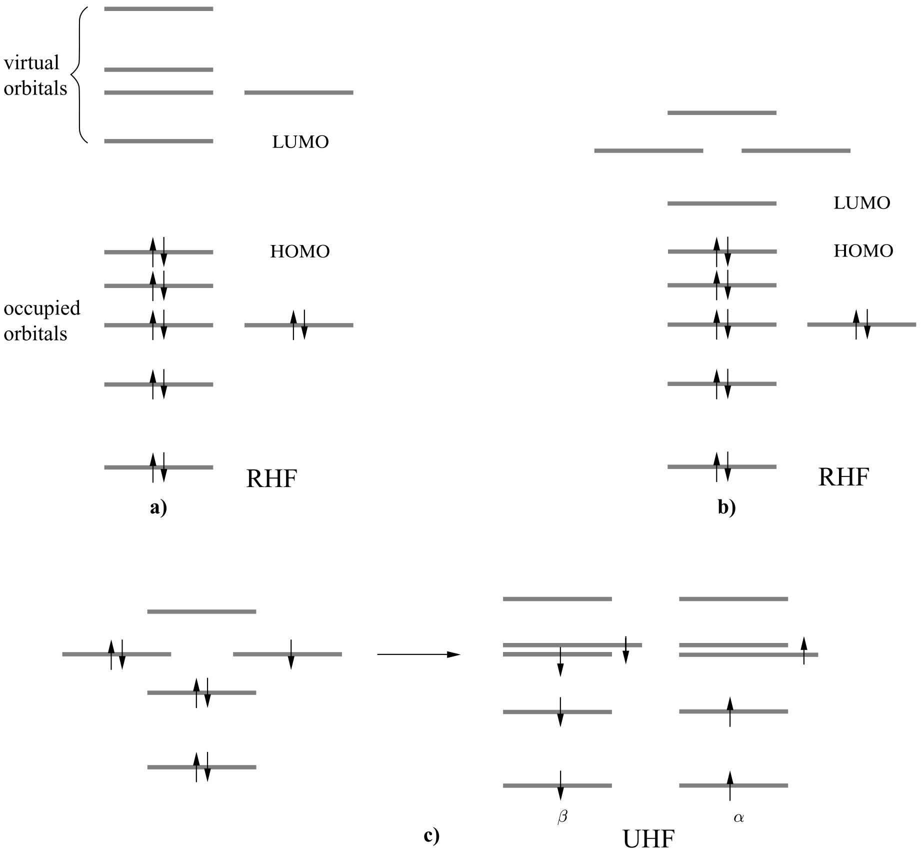


Fig. 8.4. The closed (a), poorly closed (b) and open (c) shell. The figure shows the occupancy of the molecular orbitals together with the corresponding spin functions (spin up and spin down for α and β functions): in case (a) and (b) the double occupancy of the lowest lying orbitals (on the energy scale) has been assumed; in the case (c) there is also an attempt to doubly occupy the orbitals (left-hand side), but a dilemma appears about which spinorbitals should be occupied. For example, in Fig. (c) we have decided to occupy the β spinorbital ("spin down"), but there is also a configuration with the α spinorbital ("spin up") of the same energy. This means that we need to use a scheme which allows different orbitals for different spins, e.g., UHF. The UHF procedure gives different orbitals energies for the α and β spins. One possibility is shown on the right-hand side of Fig. (c).

A CLOSED SHELL

A closed shell means that the HOMO is doubly occupied as are all the orbitals which are equal or lower in energy. The occupancy is such that the mathematical form of the Slater determinant does not depend on the spatial orientation of the x, y, z axis. Using group theory nomenclature (Appendix C), this function transforms according to *fully symmetric irreducible representation* of the symmetry group of the electronic Hamiltonian.

If a shell is not closed, it is called "open".[26] We assume that there is a unique assignment for which molecular spinorbitals[27] within a closed shell are occupied in the ground state. The concept of the closed shell is approximate because it is not clear what it means when we say that the HOMO–LUMO energy distance[28] is large or small.[29]

We need to notice that HOMO and LUMO have somewhat different meanings. As will be shown on p. 393, $-\varepsilon_{\mathrm{HOMO}}$ represents an approximate ionization energy, i.e. binding energy of an electron interacting with the $(N-1)$-electron system, while $-\varepsilon_{\mathrm{LUMO}}$ is an approximate electron affinity energy, i.e. energy of an electron interacting with the N-electron system.

The Fock equations for a closed shell (RHF method) can be derived in a very similar way as in the GHF method. This means the following steps:

- we write down the expression for the mean value of the Hamiltonian as a functional of the *orbitals* (the summation extends over all the *occupied orbitals*, there are $N/2$ of them, as will be recalled by the upper limit denoted by MO):[30] $E = 2\sum_i^{\mathrm{MO}}(i|\hat{h}|i) + \sum_{i,j}^{\mathrm{MO}}[2(ij|ij) - (ij|ji)]$;
- we seek the conditional minimum of this functional (Lagrange multipliers method) allowing for the variation of the *orbitals* which takes their orthonormality into account $\delta E = 2\sum_i^{\mathrm{MO}}(\delta i|\hat{h}|i) + \sum_{i,j}^{\mathrm{MO}}[2(\delta ij|ij) - (\delta ij|ji) + 2(i\delta j|ij) - (i\delta j|ji)] - \sum_{i,j}^{\mathrm{MO}} L'_{ij}(\delta i|j) = 0$;

---

[26] Sometimes we use the term semi-closed shell, if it is half-occupied by the electrons and we are interested in the state bearing maximum spin. In this case the Slater determinant is a good approximation. The reasons for this is, of course, the uniqueness of electron assignment to various spinorbitals. *If there is no uniqueness (as in the carbon atom), then the single-determinant approximation cannot be accepted.*

[27] The adjective "molecular" is suggested even for calculations for an atom. In a correct theory of electronic structure, the number of nuclei present in the system should not play any role. Thus, from the point of view of the computational machinery, an atom is just a molecule with one nucleus.

[28] The decision to occupy only the lowest energy MOs (so called *Aufbau Prinzip*; a name left over from the German origins of quantum mechanics) is accepted under the *assumption* that the total energy differences are sufficiently well approximated by the differences in the orbital energies.

[29] Unless the distance is zero. The helium atom, with the two electrons occupying the $1s$ orbital (HOMO), is a $1s^2$ shell of impressive "closure", because the HOMO–LUMO energy difference calculated in a good quality basis set (6-31G**, see p. 364) of atomic orbitals is of the order of 62 eV. On the other hand, the HOMO–LUMO distance is zero for the carbon atom, because in the ground state 6 electrons occupy the $1s, 2s, 2p_x, 2p_y$ and $2p_z$ orbitals. There is room for 10 electrons, and we only have six. Hence, the occupation (configuration) in the ground state is $1s^22s^22p^2$. Thus, both HOMO and LUMO are the $2p$ orbitals, with zero energy difference. If we asked for a single sentence describing why carbon compounds play a prominent role in Nature, it should be emphasized that, for carbon atoms, the HOMO–LUMO distance is equal to zero and that the orbital levels ε_{2s} and ε_{2p} are close in energy.

On the other hand, the beryllium atom is an example of a closed shell, which is not very tightly closed. Four electrons are in the lowest lying configuration $1s^22s^2$, but the orbital level $2p$ (LUMO) is relatively close to $2s$ (HOMO) (10 eV for the 6-31G** basis set is not a small gap, yet it amounts much less than that of the helium atom).

[30] And not spinorbitals; see eqs. (M.17) and (M.18).

- we derive the Euler equation for this problem from $(\delta i|\ldots)=0$. In fact it is the Fock equation expressed in *orbitals*[31]

$$\boxed{\hat{F}(1)\varphi_i(1)=\varepsilon_i\varphi_i(1), \qquad (8.30)}$$

closed-shell Fock operator

where φ are the *orbitals. The Fock operator* is defined for the closed shell, as

$$\boxed{\hat{F}(1)=\hat{h}(1)+2\hat{\mathcal{J}}(1)-\hat{\mathcal{K}}(1), \qquad (8.31)}$$

where the first term (see eq. (8.7)) is the sum of the kinetic energy operator of electron 1 and the operator of the interaction of this electron with the nuclei in the molecule, the next two terms, i.e. Coulombic $\hat{\mathcal{J}}$ and exchange $\hat{\mathcal{K}}$ operators, are connected with the potential energy of the interaction of electron 1 with all electrons in the system, and they are defined (slightly differently than before for $\hat{J}$ and $\hat{K}$ operators[32]) *via* the action on any function (χ) of the position of electron 1:

$$\begin{aligned}2\hat{\mathcal{J}}(1)\chi(1)&=\sum_{i=1}^{\mathrm{MO}}2\hat{\mathcal{J}}_i(1)\chi(1)=\sum_{i=1}^{\mathrm{MO}}2\int \mathrm{d}V_2\,\frac{1}{r_{12}}\varphi_i^*(2)\varphi_i(2)\chi(1)\\&\equiv 2\sum_{i}^{\mathrm{MO}}\int \mathrm{d}V_2\,\frac{1}{r_{12}}\left|\varphi_i(2)\right|^2\chi(1), &(8.32)\\\hat{\mathcal{K}}(1)\chi(1)&=\sum_{i=1}^{\mathrm{MO}}\hat{\mathcal{K}}_i(1)\chi(1)=\sum_{i=1}^{\mathrm{MO}}\int \mathrm{d}V_2\,\frac{1}{r_{12}}\varphi_i^*(2)\chi(2)\varphi_i(1), &(8.33)\end{aligned}$$

where integration is now exclusively over the spatial coordinates[33] of electron 2. Factor 2 multiplying the Coulombic operator results (as the reader presumably guessed) from the *double* occupation of the orbitals.

Interpretation of the Coulombic operator

The Coulombic operator is nothing else but a calculation of the Coulombic potential (with the opposite sign as created by all the *electrons*, Fig. 8.5) at the position of

---

[31] After a suitable unitary transformation of orbitals, analogous to what we have done in GHF case.

[32] Because we have orbitals here, and not spinorbitals.

[33] Simply, the summation over the spin coordinates has already been done when deriving the equation for the mean value of the Hamiltonian.

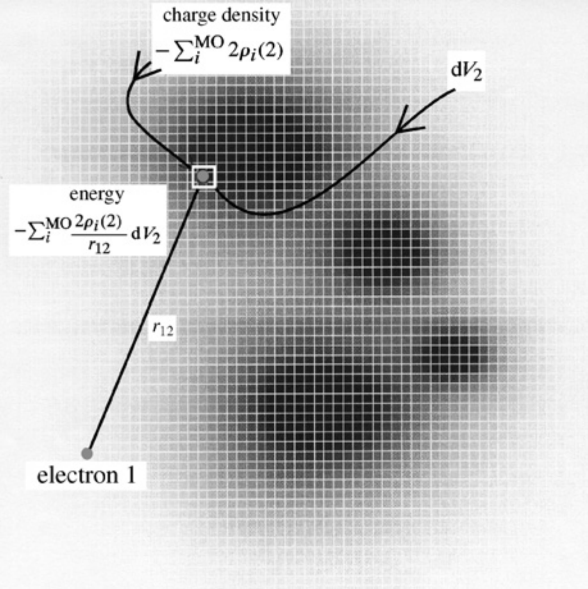


Fig. 8.5. Point-like electron 1 interacts with the total electron density (shown as electron cloud with density $\sum_i^{\mathrm{MO}} 2\rho_i(2)$). To compute the interaction energy the total electron density is chopped into small cubes. The interaction energy of electron 1 with one of such cubes of volume $\mathrm{d}V_2$ containing charge $-\sum_i^{\mathrm{MO}} 2\rho_i(2)\,\mathrm{d}V_2$ is calculated according to the Coulomb law: charge × charge divided by their distance: $\frac{-1\times(-1)\sum_i^{\mathrm{MO}} 2\rho_i(2)\,\mathrm{d}V_2}{r_{12}}$ or, alternatively, as charge -1 times electric potential produced by a single cube at electron 1. The summation over all cubes gives $\int \frac{\sum_i^{\mathrm{MO}} 2\rho_i(2)}{r_{12}}\,\mathrm{d}V_2 = 2\hat{\mathcal{J}}$.

electron 1. Indeed, such a potential coming from an electron occupying molecular orbital φ_i is equal to

$$\int \frac{\rho_i(2)}{r_{12}}\,\mathrm{d}V_2, \tag{8.34}$$

where $\rho_i(2) = \varphi_i(2)^*\varphi_i(2)$ is the probability density of finding electron 2 described by orbital φ_i. If we take into account that the orbital φ_i is occupied by *two* electrons, and that the number of the doubly occupied molecular orbitals is $N/2$, then the electrostatic potential calculated at the position of the electron 1 is

$$\int \frac{\sum_i^{\mathrm{MO}} 2\rho_i(2)}{r_{12}}\,\mathrm{d}V_2 = \sum_i^{\mathrm{MO}} 2\hat{\mathcal{J}}_i = 2\hat{\mathcal{J}}(1).$$

The same expression also means an interaction of two *elementary* charges 1 and 2, one of each represented by a diffused cloud with a given charge density distribution $\rho(2) = \sum_i^{\mathrm{MO}} 2\rho_i(2)$.

Integration in the formula for the operator $\hat{\mathcal{J}}$ is a consequence of the approximation of independent particles.

This approximation means that, in the Hartree–Fock method, the electrons do not move in the electric field of the other point-like electrons, but in the *mean static field of all the electrons* represented by electron cloud ρ. It is as if a driver (one of the electrons) in Paris did not use the position of other cars, but a map showing only the traffic intensity *via* the probability density cloud. The driver would then have a diffuse image of other vehicles,[34] and could not satisfactorily optimize the position towards other cars (it means higher energy for the molecule under study).

mean field

THE MEAN FIELD
This is typical for all the mean field methods. In these methods, *instead of watching the motion of other objects in detail*, *we average* these motions, and the problem simplifies (obviously, we pay the price of lower quality).

However, this trick is ingenious and worth remembering.[35]

Coulombic self-interaction

There is a problem with this. From what we have said, it follows that the electron 1 uses the "maps" of *total* electron density, i.e. including its own contribution to the density.[36] This looks strange though. Let us take a closer look, maybe something has been missed in our reasoning. Note first of all that the repulsion of electron 1 (occupying, say, orbital k) with the electrons, which is visible in the Fock operator, reads as $(\varphi_k|(2\hat{\mathcal{J}}-\hat{\mathcal{K}})\varphi_k)$ and not as $(\varphi_k|(2\hat{\mathcal{J}})\varphi_k)$. Let us write it down in more details:

$$
\begin{aligned}
\big(\varphi_k|(2\hat{\mathcal{J}}-\hat{\mathcal{K}})\varphi_k\big) &= \int \mathrm{d}V_1 \big|\varphi_k(1)\big|^2 \sum_{i=1}^{\mathrm{MO}} 2 \int \mathrm{d}V_2 \frac{1}{r_{12}} \varphi_i^*(2)\varphi_i(2) \\
&\quad - \sum_{i=1}^{\mathrm{MO}} \int \mathrm{d}V_1 \varphi_k(1)^* \varphi_i(1) \int \mathrm{d}V_2 \frac{1}{r_{12}} \varphi_i^*(2)\varphi_k(2) \\
&= \int \mathrm{d}V_1 \big|\varphi_k(1)\big|^2 \sum_{i=1}^{\mathrm{MO}} 2 \int \mathrm{d}V_2 \frac{1}{r_{12}} \varphi_i^*(2)\varphi_i(2)
\end{aligned}
$$

---

[34] An effect similar to the action of fog or alcohol. Both lead to miserable consequences.

[35] We use it every day, although we do not call it a mean field approach. Indeed, if we say: "*I will visit my aunt at noon, because it is easier to travel out of rush hours*", or "*I avoid driving through the centre of town, because of the traffic jams*", in practice we are using the mean field method. We average the motions of all citizens (including ourselves!) and we get a "map" (temporal or spatial), which allows us to optimize *our own* motion. The motion of our fellow-citizens *disappears*, and we obtain a *one-body* problem.

[36] Exactly as happens with real city traffic maps.

$$
\begin{aligned}
&-\int \mathrm{d}V_1 \varphi_k(1)^* \varphi_k(1) \int \mathrm{d}V_2 \frac{1}{r_{12}} \varphi_k^*(2)\varphi_k(2) \\
&-\sum_{i(\neq k)}^{\mathrm{MO}} \int \mathrm{d}V_1 \varphi_k(1)^* \varphi_i(1) \int \mathrm{d}V_2 \frac{1}{r_{12}} \varphi_i^*(2)\varphi_k(2) \\
&= \iint \mathrm{d}V_1 \mathrm{d}V_2 \frac{1}{r_{12}} \rho_k(1)\big[\rho(2)-\rho_k(2)\big] - \sum_{i(\neq k)}^{\mathrm{MO}} (ki|ik),
\end{aligned}
$$

where $\rho_k = |\varphi_k(1)|^2$, i.e. the distribution of electron 1 interacts electrostatically with all the *other* electrons,[37] i.e. with the distribution $[\rho(2)-\rho_k(2)]$ with ρ denoting the total electron density $\rho = \sum_{i=1}^{\mathrm{MO}} 2|\varphi_i|^2$ and $-\rho_k$ excluding from it the self-interaction energy of the electron in question. Thus, the Coulombic and exchange operators together ensure that an electron interacts electrostatically with other electrons, not with itself.

Electrons with parallel spins repel less

There is also an exchange remainder $-\sum_{i(\neq k)}^{\mathrm{MO}}(ki|ik)$, which is just a by-product of the antisymmetrization of the wave function (i.e. the Slater determinant), which tells us that in the Hartree–Fock picture electrons of the same spin functions[38] repel less. What??? As shown at the beginning of the present chapter, two electrons of the same spin cannot occupy the same point in space, and therefore (from the continuity of the wave function) they avoid each other. It is as if they repelled each other, because of the Pauli exclusion principle, *in addition* to their Coulombic repulsion. Is there something wrong in our result then? No, everything is OK. The necessary antisymmetric character of the wave function says simply that the same spins should keep apart. However, when the electrons described by the same spin functions keep apart, this obviously means *their Coulombic repulsion is weaker than that of electrons of opposite spins*.[39] This is what the term $-\sum_{i(\neq k}^{\mathrm{MO}}(ki|ik)$ really means.

Hartree method

The exchange operator represents a (non-intuitive) result of the antisymmetrization postulate for the total wave function (Chapter 1) and it has no classical inter-

---

[37] The fact that the integration variables pertain to electron 2 is meaningless, it is just a definite integration and the name of the variable does not count at all.

[38] When deriving the total energy expression (Appendix M), only those exchange terms survived, which correspond to the parallel spins of the interacting electrons. Note also, that for real orbitals (as in the RHF method), every exchange contribution $-(ki|ik) \equiv -\int \mathrm{d}V_1\, \varphi_k(1)\varphi_i(1) \int \mathrm{d}V_2\, \frac{1}{r_{12}}\varphi_i(2)\varphi_k(2)$ means a repulsion, because this is a self-interaction of the cloud $\varphi_k\varphi_i$.

[39] Note that in the Hamiltonian, the Coulombic repulsion of the electrons is spin-independent. This suggests that when trying to improve the description (by going beyond the Hartree–Fock approximation), we have to worry more about correlation of electrons with the opposite spin functions (e.g., those occupying the same orbital).

pretation. If the variational wave function were the *product* of the spinorbitals[40] (Douglas Hartree did this in the beginning of the quantum chemistry)

$$\phi_1(1)\phi_2(2)\phi_3(3)\cdots\phi_N(N),$$

then we would get the corresponding Euler equation, which in this case is called the Hartree equation

$$\hat{F}_{\text{Hartree}}(1)\phi_i(1) = \varepsilon_i\phi_i(1),$$

$$\hat{F}_{\text{Hartree}}(1) = \hat{h}(1) + \sum_{j(\neq i)}^{N} \hat{J}_j(1),$$

where $\hat{F}_{\text{Hartree}}$ corresponds to the Fock operator. Note that there is no self-interaction there.

8.2.9 ITERATIVE PROCEDURE FOR COMPUTING MOLECULAR ORBITALS: THE SELF-CONSISTENT FIELD METHOD

The following is a typical technique for solving the Fock equation.

First, we meet the difficulty that in order to solve the Fock equation we should first ... know its solution. Indeed, the Fock equation is not an eigenvalue problem, but a pseudo-eigenvalue problem, because the Fock operator depends on the *solutions* (obviously, unknown). Regardless of how strange it might seem, we deal with this situation quite easily using an iterative approach. This is called the self-consistent field method (SCF). In this method (Fig. 8.6) we

SCF iterations

- assume at the beginning (zero-th iteration) a certain shape of molecular orbitals;[41]
- introduce these orbitals to the Fock operator, thus obtaining a sort of "caricature" of it (the zero-order Fock operator);
- solve the eigenvalue problem using the above "Fock operator" and get the molecular orbitals of the first iteration;
- repeat the process until the shape of the orbitals does not change in the next iteration, i.e. until the Fock equations are solved.[42]

---

[40] Such a function is not legal – it does not fulfil the antisymmetrization postulate. This illegal character (caused by a lack of the Pauli exclusion principle) would sometimes give unfortunate consequences: e.g., more than two electrons would occupy the 1*s* orbital, etc.

[41] These are usually the any-sort "orbitals", although recently, because of the *direct SCF* idea (we calculate the integrals whenever they are needed, i.e. at each iteration), an effort is made to save computational time per iteration and therefore to provide as good-quality a starting function as possible. We may obtain it *via* an initial calculation with some two-electron integrals neglected.

[42] Using our driver analogy, we may say that at the beginning the driver has *false* maps of the probability density (thus the system energy is high – in our analogy the car repair costs are large). The next iterations (repair costs effectively teach all the drivers) improve the map, the final energy decreases, and at the very end we get the best map possible. The mean energy is the lowest possible (within the mean field method). A further energy lowering is only possible beyond the Hartree–Fock approximation, i.e. outside of the mean field method, which for the drivers means not using maps. A suspicious

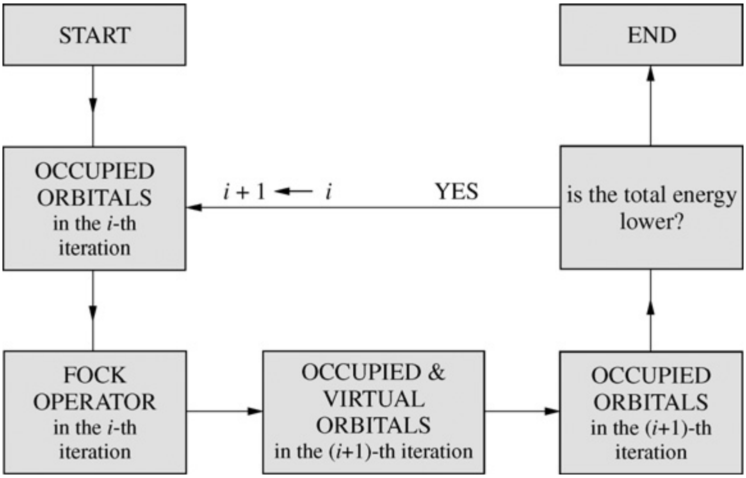


Fig. 8.6. Iterative solution of the Fock equation (the self-consistent field method, SCF). We:
- start from any set of occupied orbitals (zeroth iteration),
- insert them to the Fock operator,
- solve the Fock equation,
- obtain the molecular orbitals of the first approximation,
- choose those of the lowest energy as the occupied ones and if your criterion of the total energy is not satisfied, repeat the procedure.

8.3 TOTAL ENERGY IN THE HARTREE–FOCK METHOD

In Appendix M, p. 986, we derived the following expressions for the mean value of the Hamiltonian using the normalized determinant (without a constant additive term for the nuclear repulsion energy V_{nn}, SMO means summation over the spinorbitals $i = 1, \dots, N$; in the RHF method, the MO summation limit means summation over the orbitals $i = 1, \dots, N/2$)

$$\begin{aligned} E'_{\mathrm{HF}} &= \sum_{i}^{\mathrm{SMO}} \langle i|\hat{h}|i\rangle + \frac{1}{2}\sum_{i,j=1}^{\mathrm{SMO}} \big[\langle ij|ij\rangle - \langle ij|ji\rangle\big] \\ &\equiv \sum_{i}^{\mathrm{SMO}} h_{ii} + \frac{1}{2}\sum_{i,j=1}^{\mathrm{SMO}} [J_{ij} - K_{ij}]. \end{aligned} \tag{8.35}$$

---

person (scientist) should be careful, because our solution *may depend on the starting point used*, i.e. from the initial, completely arbitrary orbitals. Besides, the iteration process does not necessarily need to be convergent. But it appears that the solutions in the Hartree–Fock method are usually independent on the zero-th order MOs, and convergence problems are very rare. This is surprising. This situation is much worse for better-quality computations, where the AOs of small exponents are included (*diffuse orbitals*). Then we truly meet the problem already described (p. 292) of searching for the global energy minimum among a multitude of local ones.

If double occupancy is assumed (i.e. the flexibility of the variational wave function is *restricted*) we may transform this expression in the following way

$$
\begin{aligned}
&E'_{\text{RHF}}(\text{double occupancy})\\
&\quad=\sum_{i}^{\text{MO}}\big(\langle i\alpha|\hat{h}|i\alpha\rangle+\langle i\beta|\hat{h}|i\beta\rangle\big)\\
&\qquad+\frac{1}{2}\sum_{i}^{\text{MO}}\sum_{j}^{\text{SMO}}\big[\langle i\alpha,j|i\alpha,j\rangle-\langle i\alpha,j|j,i\alpha\rangle+\langle i\beta,j|i\beta,j\rangle-\langle i\beta,j|ji\beta\rangle\big]\\
&\quad=2\sum_{i}^{\text{MO}}(i|\hat{h}|i)+\frac{1}{2}\sum_{i}^{\text{MO}}\sum_{j}^{\text{MO}}\big[\langle i\alpha,j\alpha|i\alpha,j\alpha\rangle+\langle i\alpha,j\beta|i\alpha,j\beta\rangle-\langle i\alpha,j\alpha|j\alpha,i\alpha\rangle\\
&\qquad\qquad-\langle i\alpha,j\beta|j\beta,i\alpha\rangle+\langle i\beta,j\alpha|i\beta,j\alpha\rangle+\langle i\beta,j\beta|i\beta,j\beta\rangle\\
&\qquad\qquad-\langle i\beta,j\alpha|j\alpha,i\beta\rangle-\langle i\beta,j\beta|j\beta,i\beta\rangle\big]\\
&\quad=2\sum_{i}^{\text{MO}}(i|\hat{h}|i)+\frac{1}{2}\sum_{i}^{\text{MO}}\sum_{j}^{\text{MO}}\big[4(ij|ij)-2(ij|ji)\big]\\
&\quad=2\sum_{i}^{\text{MO}}(i|\hat{h}|i)+\sum_{i}^{\text{MO}}\sum_{j}^{\text{MO}}\big[2(ij|ij)-(ij|ji)\big].
\end{aligned}
$$

This finally gives

$$
E'_{\text{RHF}}=2\sum_{i}^{\text{MO}}(i|\hat{h}|i)+\sum_{i,j}^{\text{MO}}\big[2(ij|ij)-(ij|ji)\big]\equiv 2\sum_{i}^{\text{MO}}h_{ii}+\sum_{i,j}^{\text{MO}}[2\mathcal{J}_{ij}-\mathcal{K}_{ij}]. \tag{8.36}
$$

Given the equality $\langle i|\hat{h}|i\rangle=(i|\hat{h}|i)$, these integrals have been written here as h_{ii}. The *Coulombic and exchange* integrals expressed in spinorbitals are denoted J_{ij} and K_{ij} and expressed in orbitals as $\mathcal{J}_{ij}$ and $\mathcal{K}_{ij}$.

Both formulae (8.35) and (8.36) may give different results, because in the first, no double occupancy is assumed (we will discuss this further on p. 372).

The additive constant corresponding to the internuclear repulsion (it is constant, since the nuclei positions are frozen)

$$
V_{nn}=\sum_{a<b}\frac{Z_aZ_b}{R_{ab}}, \tag{8.37}
$$

has not yet been introduced and thus the full Hartree–Fock energy is

$$
E_{\text{RHF}}=E'_{\text{RHF}}+V_{nn}. \tag{8.38}
$$

Note that

$$\sum_{i,j}^{\text{MO}}[2\mathcal{J}_{ij} - \mathcal{K}_{ij}] \equiv \sum_{i,j}^{\text{MO}}\big[2(ij|ij) - (ij|ji)\big] = \left\langle \psi_{\text{HF}} \middle| \sum_{i<j} \frac{1}{r_{ij}} \middle| \psi_{\text{HF}} \right\rangle = V_{ee}. \tag{8.39}$$

Hence, V_{ee} is the mean electronic repulsion energy in our system.[43]

It is desirable (interpretation purposes) to include the orbital energies in the formulae derived. Let us recollect that the orbital energy ε_i is the mean value of the Fock operator for orbital i, i.e. the energy of an effective electron described by this orbital. Based on formulae (8.31)–(8.33), this can be expressed as (i stands for the molecular orbital)

$$\varepsilon_i = h_{ii} + \sum_{j}^{\text{MO}}[2\mathcal{J}_{ij} - \mathcal{K}_{ij}], \tag{8.40}$$

and this in turn gives an elegant expression for the Hartree–Fock electronic energy

$$E'_{\text{RHF}} = \sum_{i}^{\text{MO}}[h_{ii} + \varepsilon_i]. \tag{8.41}$$

From eqs. (8.36), (8.40) and (8.39), the total electronic energy may be expressed as

$$E'_{\text{RHF}} = \sum_{i=1}^{\text{MO}} 2\varepsilon_i - V_{ee}. \tag{8.42}$$

It can be seen that the total electronic energy (i.e. E'_{RHF}) *is not the sum of the orbital energies of electrons* $\sum_i 2\varepsilon_i$.

And we would already expect full additivity, since the electrons in the Hartree–Fock method are treated as independent. Yet "independent" does not mean "non-interacting". The reason for the non-additivity is that for *each* electron we need to calculate its effective interaction with *all* the electrons, hence we would get too much repulsion.[44] Of course, the total energy, and not the sum of the orbital energies, is the most valuable. Yet in many quantum chemical problems we interpret orbital energy lowering as energetically profitable. And it turns out that such an interpretation has an approximate justification. Works by Fraga, Politzer and Rue-

---

[43] Please recall that $\langle\psi_{\text{HF}}|\hat{H}|\psi_{\text{HF}}\rangle = E_{\text{RHF}}$ and V_{ee} is, therefore, the Coulombic interaction of electrons.

[44] For example, the interaction of electron 5 and electron 7 is calculated *twice*: as the interaction $\frac{1}{r_{57}}$ and $\frac{1}{r_{75}}$

denberg[45] show that at the equilibrium geometry of a molecule, the formula

$$E_{\mathrm{RHF}} = E'_{\mathrm{RHF}} + V_{nn} \approx \frac{3}{2}\sum_{i=1}^{\mathrm{MO}} 2\epsilon_i, \tag{8.43}$$

works with 2%–4% precision, and even better results may be obtained by taking a factor of … 1.55 instead of $\frac{3}{2}$.

8.4 COMPUTATIONAL TECHNIQUE: ATOMIC ORBITALS AS BUILDING BLOCKS OF THE MOLECULAR WAVE FUNCTION

atomic orbital

exponential basis sets

We have to be careful because the term “atomic orbital” is used in quantum chemistry with a double meaning. These are: (i) orbitals of the mean field for a particular atom, or (ii) functions localized in the space about a given centre. We nearly always use what are known as *exponential basis sets*:[46]

$$g(\boldsymbol{r}) = f(x,y,z)\exp(-\zeta r^n),$$

where $f(x,y,z)$ is a polynomial. Such an atomic orbital is localized (centred) around $(0,0,0)$. The larger the exponent ζ, the more effective is this centring.

For $n=1$, we have what is called the STO – *Slater-Type Orbitals*, and for $n=2$ the GTO – *Gaussian Type Orbitals*.

8.4.1 CENTRING OF THE ATOMIC ORBITAL

Atomic orbital $g(\boldsymbol{r})$ may be shifted by a vector $\boldsymbol{A}$ in space [translation operation $\hat{\mathcal{U}}(\boldsymbol{A})$, see Chapter 2] to result in the new function $\hat{\mathcal{U}}(\boldsymbol{A})g(\boldsymbol{r}) = g(\hat{T}^{-1}(\boldsymbol{A})\boldsymbol{r}) = g(\hat{T}(-\boldsymbol{A})\boldsymbol{r}) = g(\boldsymbol{r}-\boldsymbol{A})$, because $\hat{T}^{-1}(\boldsymbol{A}) = \hat{T}(-\boldsymbol{A})$. Hence the orbital centred at a given point (indicated by a vector $\boldsymbol{A}$) is (Fig. 8.7):

$$g(\boldsymbol{r}-\boldsymbol{A}) = f(x-A_x, y-A_y, z-A_z)\exp\bigl[-\zeta|\boldsymbol{r}-\boldsymbol{A}|^n\bigr]. \tag{8.44}$$

Different centring of the atomic orbitals is used in practice, although if the complete set of the orbitals were at our disposal, then it might be centred in a *single* point.

---

[45] S. Fraga, *Theor. Chim. Acta* 2 (1964) 406; P. Politzer, *J. Chem. Phys.* 64 (1976) 4239; K. Ruedenberg, *J. Chem. Phys.* 66 (1977) 375.

[46] Atomic orbitals (the first meaning) are usually expressed as linear combinations of atomic orbitals in the second meaning. There may be, but does not need to be, a nucleus in the centre. *If* we set an atomic nucleus at this point, we emphasize the important fact that an electron will reside close to the nucleus.

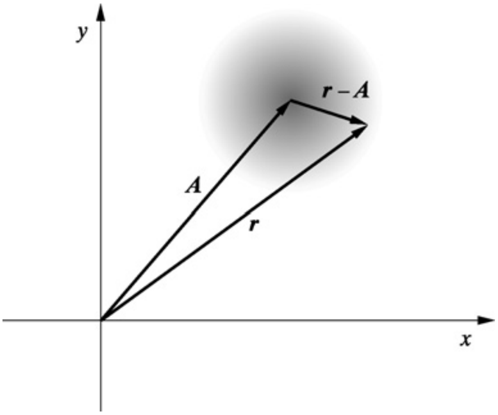


Fig. 8.7. The AO $g(\boldsymbol{r})$ centred at the point shown by vector $\boldsymbol{A}$, means the creation of the orbital $g(\boldsymbol{r}-\boldsymbol{A})$. A linear combination of such orbitals can describe any smooth function of the position in space, of any degree of complexity.

> It is more economic, however, to allow using the incomplete set and the possibility of AO centred in various points of space.

We could construct a molecular orbital of any complexity *exclusively* using the orbitals $g(\boldsymbol{r}) = \exp(-\zeta|\boldsymbol{r}-\boldsymbol{A}|^n)$, i.e. the $f(x,y,z) = \text{const}$, colloquially known as the 1*s* orbitals. It is clear that we could do it in any "hole-repairing" (plastering-like) procedure.[47] But why do we not do it like this? The reason is simple: the number of such atomic orbitals that we would have to include in the calculations would be too large. Instead, chemists allow for higher-order polynomials $f(x,y,z)$. This makes for more efficient "plastering", because, instead of spherically symmetric objects (1*s*), we can use orbitals $g(\boldsymbol{r})$ of virtually any shape (*via* an admixture of the *p*, *d*, *f*, ... functions). For example, how a rugby-ball shaped orbital can be achieved is shown in Fig. 8.8.

8.4.2 SLATER-TYPE ORBITALS (STO)

The Slater-type orbitals[48] differ from the atomic orbitals of the hydrogen atom (see p. 178). The first difference is that the radial part is simplified in the STOs.

---

[47] Frost even derived the method of FSGO – *Floating Spherical Gaussian Orbitals*, A.A. Frost, *J. Chem. Phys.* 47 (1967) 3707), i.e. Gaussian type orbitals of variationally-chosen positions. Their number is truly minimal – equal to the number of occupied MOs.

[48] We will distinguish two similar terms here: Slater-type orbitals and Slater orbitals. The latter is reserved for special Slater-type orbitals, in which the exponent is easily computed by considering the effect of the screening of nucleus by the internal electronic shells. The screening coefficient is calculated according to the Slater rules, see below. Slater formulated these rules by watching the orbitals delivered by his young coworkers from the computing room. As he writes: "when the boys were computing", he noticed that we can quite easily *predict* the orbital shape without any calculation. It is enough to introduce screening, which could easily be predicted from numerical experience.

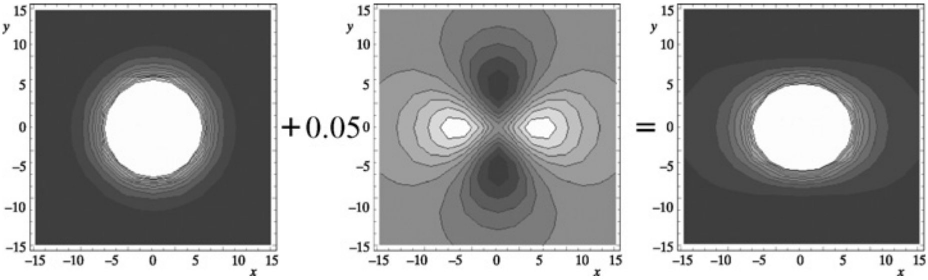


Fig. 8.8. An example of function (xy section, in a.u.) modelling by a linear combination of AOs. If a tiny admixture of the $3d_{x^2-y^2}$ function is added to the spherically symmetric $1s$ orbital (a football ball, both orbitals with 0.5 orbital exponent). We will get shrinking in one direction, and elongation in the other (the dimension in the third direction is unchanged), i.e. a flattened rugby ball. In our case the tiny admixture means 0.05. If the admixture were of the $2p$ type, the ball would look more like an egg. As we see, nearly everything can be simulated like this. This is essence of the LCAO method.

We put r^k, where k is a natural number or zero, instead of Laguerre polynomials[49] given on p. 179. The second difference is in the orbital exponent, which has no constraint except that it has to be positive.[50]

The STOs have a great advantage: they decay with distance from the centre in a similar way to the "true" orbitals – let us recall the exponential vanishing of the hydrogen atom orbitals (see Chapter 4).[51] STOs would be fine, but finally we have

---

[49] This means that the radial part of a STO has no nodes. Because of this, STOs of the same angular dependence, in contrast to the hydrogen-like atom orbitals, are not orthogonal.

[50] Otherwise the orbital would not be square-integrable. To get a rough idea of how the atomic orbitals for a particular atom look, Slater orbitals have been proposed: $1s, 2s, 2p, \ldots$. They are Slater-type orbitals with $\zeta = \frac{Z-\sigma}{n}$, where Z stands for the nuclear charge, σ tells us how other electrons screen (i.e. effectively diminish) the charge of the nucleus ($\sigma = 0$ for an atom with a single electron), and n is the principal quantum number. The key quantity σ, is calculated for each orbital of an atom using simple rules of thumb (designed by Prof. Slater after examining his students' computer outputs). We focus on the electron occupying the orbital in question, and we try to see what it sees. The electron sees that the nucleus charge is screened by its fellow electrons. The Slater rules are as follows:

- write down the electronic configuration of an atom grouping the orbitals in the following way: $[1s][2s2p][3s3p][3d]\ldots$,
- electrons to the right give zero contribution,
- *other* electrons in the same group contribute 0.35 each, except $[1s]$ which contributes 0.30,
- for an electron in an $[nsnp]$ group each electron in the $n-1$ group contributes 0.85, for lower groups each contributes 1.0 and for the $[nd]$ or $[nf]$ groups, all electrons in groups to the left contribute 1.0.

Example: The carbon atom. Configuration in groups: $[1s^2][2s^2 2p^2]$. There will be two σ's: $\sigma_{1s} = 0.30$, $\sigma_{2s} = \sigma_{2p} = 3 \cdot 0.35 + 2 \cdot 0.85 = 2.75$. Hence, $\zeta_{1s} = \frac{6-0.30}{1} = 5.70$, $\zeta_{2s} = \zeta_{2p} = \frac{6-2.75}{2} = 1.625$. Hence, $1s_C = N_{1s} \exp(-5.70r)$, $2s_C = N_{2s} \exp(-1.625r)$, $2p_{x,C} = N_{2p} x \exp(-1.625r)$, $2p_{y,C} = N_{2p} y \exp(-1.625r)$, $2p_{z,C} = N_{2p} z \exp(-1.625r)$.

[51] It has been proved that *each* of the Hartree–Fock orbitals has *the same* asymptotic dependence on the distance from the molecule (N.C. Handy, M.T. Marron, H.J. Silverstone, *Phys. Rev.* 180 (1969) 45), i.e. $\text{const} \cdot \exp(-\sqrt{-2\varepsilon_{\max}}r)$, where $\varepsilon_{\max}$ is the orbital energy of HOMO. Earlier, people thought the orbitals decay as $\exp(-\sqrt{-2\varepsilon_i}r)$, where ε_i is the orbital energy expressed in atomic units. The last formula, as is easy to prove, holds for the atomic orbitals of hydrogen atoms (see p. 178). R. Ahlrichs,

to compute a large number of the integrals needed.[52] And here is a real problem. Since the Hamiltonian contains the electron–electron interactions, integrals appear with, in general, four atomic orbitals (of different centres). These integrals are difficult to calculate, and are therefore excessively computer time-consuming.

8.4.3 GAUSSIAN-TYPE ORBITALS (GTO)

If the exponent in eq. (8.44) is equal to $n = 2$, we are dealing with Gaussian Type Orbitals (GTO).

The most important among them are 1s-type orbitals:

$$\chi_p \equiv G_p(\boldsymbol{r}; \alpha_p, \boldsymbol{R}_p) = \left(\frac{2\alpha_p}{\pi}\right)^{\frac{3}{4}} \exp\left(-\alpha_p |\boldsymbol{r} - \boldsymbol{R}_p|^2\right), \tag{8.45}$$

where α_p is the orbital exponent, $\boldsymbol{R}_p$ is the vector indicating the centre of the orbital, and the factor standing before the expression is the normalization constant. Why are 1s-type orbitals so important? Because we may construct "everything" (even s, p, d-like orbitals) out of them using proper linear combinations. For example, the difference of two 1s orbitals, centred at $(a, 0, 0)$ and $(-a, 0, 0)$, is similar to the $2p_x$ orbital (Fig. 8.9).

> The most important reason for the great progress of quantum chemistry in recent years is replacing the Slater-type orbitals, formerly used, by Gaussian-type orbitals as the expansion functions.

Orbital size

Each orbital extends to infinity and it is impossible to measure its extension using a ruler. Still, the α_p coefficient may allow comparison of the sizes of various orbitals. And the quantity

$$\rho_p = (\alpha_p)^{-\frac{1}{2}} \tag{8.46}$$

may be called (which is certainly an exaggeration) the *orbital radius* of the orbital χ_p, because[53]

orbital radius

$$\int_0^{\rho_p}\int_0^{\pi}\int_0^{2\pi} \chi_p^2 \,\mathrm{d}\tau = 4\pi \int_0^{\rho_p} \chi_p^2 r^2 \,\mathrm{d}r = 0.74, \tag{8.47}$$

M. Hoffmann-Ostenhoff, T. Hoffmann-Ostenhoff, J.D. Morgan III, *Phys. Rev.* A23 (1981) 2106 have shown that at a long distance r from an atom or a molecule, the square root of the *ideal* electron density satisfies the inequality: $\sqrt{\rho} \leqslant C(1+r)^{\frac{(Z-N+1)}{\sqrt{2\varepsilon}}-1} \exp[-(2\varepsilon)]$, where ε is the first *ionization potential*, Z is the sum of the nuclear charges, N is the number of electrons, and C is a constant.

[52] The number of necessary integrals may reach billions.

[53] See, e.g., I.S. Gradshteyn, J.M. Rizhik, "*Table of Integrals, Series, and Products*", Academic Press, Orlando, 1980, formula 3.381.

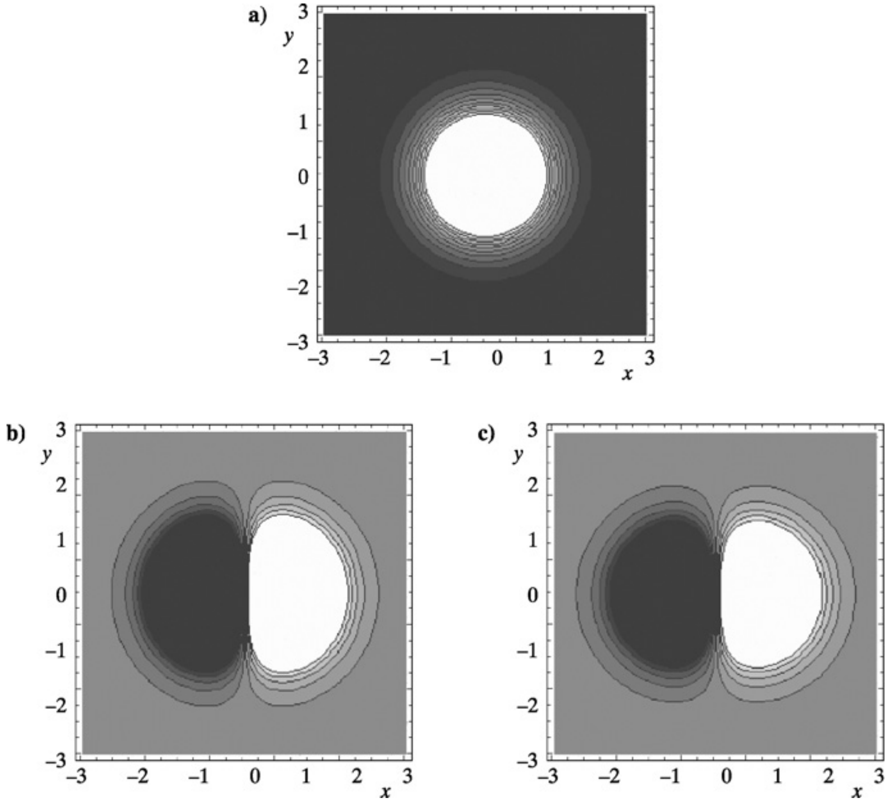


Fig. 8.9. Two spherically symmetric Gaussian-type orbitals (xy section, in a.u.) of the "1s-type" $G(\boldsymbol{r}; 1, \mathbf{0})$ (a) are used to form the difference orbital (b): $G(\boldsymbol{r}; 1, -0.5\boldsymbol{i}) - G(\boldsymbol{r}; 1, +0.5\boldsymbol{i})$, where $\boldsymbol{i}$ is the unity vector along the x axis. For comparison (c) the Gaussian-type p_x orbital is shown: $xG(\boldsymbol{r}; 1, \mathbf{0})$. It can be seen that the spherical orbitals may indeed simulate the $2p$ ones. Similarly, they can model the spatial functions of arbitrary complexity.

where the integration over r goes through the inside of a sphere of radius ρ_p. This gives us an idea about the part of space in which the orbital has an important amplitude. For example, the 1s hydrogen atom orbital can be approximated as a linear combination of three 1s GTOs (here centred on the origin of the coordinate system; such a popular approximation is abbreviated to STO-3G):[54]

$$\begin{aligned} 1s \approx\, & 0.64767G_1(\boldsymbol{r}; 0.151374, \mathbf{0}) + 0.40789G_2(\boldsymbol{r}; 0.681277, \mathbf{0}) \\ & + 0.07048G_3(\boldsymbol{r}; 4.50038, \mathbf{0}) \end{aligned} \tag{8.48}$$

which corresponds to the following radii ρ of the three GTOs: 2.57, 1.21 and 0.47 a.u.

---

[54] S. Huzinaga, *J. Chem. Phys.* 42 (1965) 1293.

Product of GTOs

> The product of two Gaussian-type 1s orbitals (even if they have different centres) is a single Gaussian-type 1s orbital.[55]

The case of GTOs other than 1s does not give any trouble, but the result is slightly different. The product of the exponential factors is, of course, the 1s-type GTO, shown above. The polynomials of x, y, z standing in both GTOs multiplied by each other (recall the dependence of the polynomial on the orbital centring,

---

[55] It is the most important feature of GTOs (along with the square dependence in the exponent).

Let us take two (not normalized) GTOs 1s: $\exp(-a(\boldsymbol{r}-\boldsymbol{A})^2)$ and $\exp(-b(\boldsymbol{r}-\boldsymbol{B})^2)$, the first centred on the point shown by vector $\boldsymbol{A}$, the second – by vector $\boldsymbol{B}$. It will be shown that their product is the Gaussian-type orbital

$$\exp\bigl(-a(\boldsymbol{r}-\boldsymbol{A})^2\bigr)\exp\bigl(-b(\boldsymbol{r}-\boldsymbol{B})^2\bigr)=N\exp\bigl(-c(\boldsymbol{r}-\boldsymbol{C})^2\bigr),$$

with parameters $c=a+b$, $\boldsymbol{C}=(a\boldsymbol{A}+b\boldsymbol{B})/(a+b)$, $N=\exp[-\frac{ab}{a+b}(\boldsymbol{A}-\boldsymbol{B})^2]$.

Vector $\boldsymbol{C}$ shows the centre of the new Gaussian-type orbital. It is identical to the centre of mass position, where the role of mass is played by the orbital exponents a and b.

Here is the proof:

$$\begin{aligned}\textit{Left side} &= \exp\bigl(-ar^2+2a\boldsymbol{rA}-aA^2-br^2+2b\boldsymbol{rB}-bB^2\bigr)\\ &= \exp\bigl(-(a+b)r^2+2\boldsymbol{r}(a\boldsymbol{A}+b\boldsymbol{B})\bigr)\exp\bigl[-\bigl(aA^2+bB^2\bigr)\bigr]\\ &= \exp\bigl(-cr^2+2c\boldsymbol{Cr}\bigr)\exp\bigl[-\bigl(aA^2+bB^2\bigr)\bigr],\\ \textit{Right side} &= N\exp\bigl(-c(\boldsymbol{r}-\boldsymbol{C})^2\bigr)=N\exp\bigl[-c\bigl(r^2-2\boldsymbol{Cr}+C^2\bigr)\bigr]=\textit{Left side},\end{aligned}$$

if $N=\exp(cC^2-aA^2-bB^2)$. It is instructive to transform the expression for N, which is a kind of amplitude of the Gaussian-type orbital originating from the multiplication of two GTOs. So,

$$\begin{aligned}N &= \exp\bigl[(a+b)C^2-aA^2-bB^2\bigr]=\exp\left(\frac{(a^2A^2+b^2B^2+2ab\boldsymbol{AB})}{(a+b)}-aA^2-bB^2\right)\\ &= \exp\left(\frac{1}{a+b}\bigl(a^2A^2+b^2B^2+2ab\boldsymbol{AB}-a^2A^2-abA^2-b^2B^2-abB^2\bigr)\right)\\ &= \exp\left(\frac{1}{a+b}\bigl(2ab\boldsymbol{AB}-abA^2-abB^2\bigr)\right)\\ &= \exp\left(\frac{ab}{a+b}\bigl(2\boldsymbol{AB}-A^2-B^2\bigr)\right)=\exp\left(\frac{-ab}{a+b}(\boldsymbol{A}-\boldsymbol{B})^2\right).\end{aligned}$$

This is what we wanted to show.

It is seen that if $\boldsymbol{A}=\boldsymbol{B}$, then amplitude N is equal to 1 and the GTO with the $a+b$ exponent results (as it should). The amplitude N strongly depends on the distance $|\boldsymbol{A}-\boldsymbol{B}|$ between two centres. If the distance is large, the N is very small, which gives the product of two distant GTOs as practically zero (in agreement with common sense). It is also clear that if we multiply two strongly contracted GTOs ($a, b \gg 1$) of different centres, the "GTO-product" is again small. Indeed, let us take, e.g., $a=b$. We get $N=\exp\{[-a/2][\boldsymbol{A}-\boldsymbol{B}]^2\}$.

formula (8.44)), can always be presented as a certain polynomial of x', y', z' taken versus the new centre $\boldsymbol{C}$. Hence, in the general case,

the product of any two Gaussian-type orbitals is a linear combination of Gaussian-type orbitals.

Integrals

If somebody wanted to perform alone[56] quantum chemical calculations, they would immediately face integrals to compute, the simplest among them being the $1s$-type. Expressions for these integrals are given in Appendix P on p. 1004.

8.4.4 LINEAR COMBINATION OF ATOMIC ORBITALS (LCAO) METHOD

Algebraic approximation

LCAO MO

Usually we apply the self-consistent field approach with the LCAO method; this is then the SCF LCAO MO.[57] In the SCF LCAO MO method, each molecular orbital is presented as a linear combination of atomic orbitals χ_s

$$\varphi_i(1) = \sum_s^M c_{si}\chi_s(1) \tag{8.49}$$

where the symbol (1) emphasizes that each of the atomic orbitals, and the resulting molecular orbital, depend on the spatial coordinates of *one* electron only (*say*, electron 1). The coefficients c_{si} are called the LCAO coefficients.

The approximation, in which the molecular orbitals are expressed as linear combinations of the atomic orbitals, is also called the *algebraic approximation*.[58]

---

[56] That is, independent of existing commercial programs, which only require the knowledge of how to push a few buttons.

[57] *Linear Combination of Atomic Orbitals – Molecular Orbitals*. This English abbreviation helped Polish quantum chemists in totalitarian times (as specialists in "MO methods", MO standing for the mighty "citizen police" which included the secret police). It was independently used by Professors Wiktor Kemula (University of Warsaw) and Kazimierz Gumiński (Jagiellonian University). A young coworker of Prof. Gumiński complained, that despite much effort he still could not get the official registered address in Cracow, required for employment at the university. The Professor wrote a letter to the officials, and asked his coworker to deliver it in person. The reaction was immediate: "*Why didn't you show this to us earlier?!*".

[58] It was introduced in solid state theory by Felix Bloch (his biography is on p. 435), and used in chemistry for the first time by Hückel.

Erich Hückel (1896–1980), German physicist, professor at the universities in Stuttgart and Marburg, student of Bohr and Debye. Erich Hückel, presumably inspired by his brother Walter, an eminent organic chemist, created a simplified version of the Hartree–Fock method, which played a major role in linking the quantum theory with chemistry. Even today, although this tool is extremely simplistic and has been superseded by numerous and much better computational schemes, Hückel theory is valued as an initial insight into the electronic structure of some categories of molecules and solids.

Curiosity: these people liked to amuse themselves with little rhymes. Felix Bloch has translated a poem by Walter Hückel from German to English. It does not look like a great poetry, but deals with the famous Erwin (Schrödinger) and his mysterious function ψ:

"*Erwin with his* ψ *can do*
Calculations quite a few.
But one thing has not been seen,
Just what does ψ *really mean*".

Why is it so useful? Imagine we do not have such a tool at our disposal. Then we are confronted with defining a function that depends on the position in 3D space and has a quite complex shape (Fig. 8.10). If we want to do it accurately, we should provide the function values at many points in space, say for a large grid with a huge number of nodes, and the memory of our PC will not stand it. Also, in such an approach one would not make use of the fact that the function is smooth. We find our way through by using atomic orbitals. For example, even if we wrote that a molecular orbital is in fact *a single* atomic orbital (we can determine the latter by giving *only four* numbers: three coordinates of the orbital centre and the orbital exponent), although very primitive, this would carry a lot of physical intuition (truth…): (i) the spatial distribution of the probability of finding the electron is concentrated in some small region of space, (ii) the function decays exponentially when we go away from this region, etc.

"Blocks" of molecular orbitals φ_i are constructed out of "primary building blocks" – the one-electron functions χ_s (in the jargon called *atomic orbitals*), which

atomic orbitals (AO)

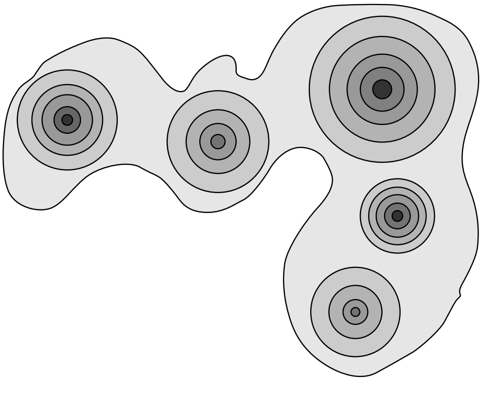


Fig. 8.10. The concept of a molecular orbital (MO) as a linear combination of atomic orbitals (LCAO), a section view. From the point of view of mathematics, it is an expansion in a series of a complete set of functions. From the viewpoint of physics, it is just recognizing that when an electron is close to nucleus a, it should behave in a similar way as that required by the atomic orbital of atom a. From the point of view of a bricklayer, it represents the construction of a large building from soft and mutually interpenetrating bricks.

are required to fill two basic conditions:

- they need to be square-integrable,
- they need to form the complete set, i.e. "everything" can be constructed from this set (any smooth square-integrable function of x, y, z),

and several *practical* conditions:

- they should be *effective*, i.e. each single function should include a part of the physics of the problem (position in space, decay rate while going to ∞, etc.),
- should be "flexible", i.e. their parameters should influence their shape to a large extent,
- the resulting *integrals should be easily computable* (numerically and/or analytically), see p. 360.

In *computational practice*, unfortunately, we fulfil the second set of conditions only to some extent: the set of orbitals taken into calculations (i.e. *the basis set*) is always limited, because computing time means money, etc. In some calculations for crystals, we also remove the first set of conditions (e.g., we often use *plane waves*: $\exp(i\boldsymbol{k}\cdot\boldsymbol{r})$, and these are not square-integrable).

plane waves

Interpretation of LCAO. If in Fig. 8.10, we take the linear combination of five atomic orbitals and provide a reasonable choice of their centres, the exponents and the weights of the functions, we will get quite a good approximation of the ideal orbital. We account for the advantages as follows: instead of providing a huge number of function values at the grid nodes, we master the function using only $5 \times 5 = 25$ numbers.[59]

The idea of LCAO MO is motivated by the fact that the molecular orbital should consist of spatial sections (atomic orbitals), because in a molecule in the vicinity of a given atom, an electron should be described by an atomic orbital of *this* atom. The essence of the LCAO approach is just the connection (unification) of such sections. But only some AOs are important in practice. This means that the main effort of constructing MOs is connected to precise shaping and polishing, by inclusion of more and more of the necessary AOs.[60]

Effectiveness of AOs mixing

When could we expect that two normalized AOs will have comparable LCAO coefficients in a low-energy MO? Two rules hold (both can be deduced from eq. (D.1)) for the *mixing effectiveness* of the AOs, obtained from numerical experience:

mixing effectivity AO

EFFECTIVENESS OF AO MIXING

- AOs must correspond to comparable energies (in the meaning of the mean value of the Fock operator),
- AOs must have large overlap integral.

---

[59] Three coordinates of the centre, the exponent and the coefficient c_{si} standing at AO altogether give five parameters per one AO.

[60] Which plays the role of the filling mass, because we aim for a beautiful shape (i.e. ideal from the point of view of the variational method) for the MOs.

Let us see what we obtain as the orbital energies[61] (in a.u.) for several important atoms:

	$1s$	$2s$	$2p$	$3s$	$3p$
H	**−0.5**	–	–	–	–
C	−11.34	**−0.71**	**−0.41**	–	–
N	−15.67	−0.96	**−0.51**	–	–
O	−20.68	−1.25	**−0.62**	–	–
F	−26.38	−1.57	**−0.73**	–	–
Cl	−104.88	−10.61	−8.07	−1.07	**−0.51**

Now, which orbitals will mix effectively when forming methane? The hydrogen atom offers the $1s$ orbital with energy −0.5. As we can see from the table, there is no possibility of effectively mixing with the carbon $1s$ orbital, while the $2s$ and $2p$ are very good candidates. Note that

> the orbital energies of all the outer-most (the so called valence) orbitals are similar for all the elements (highlighted as bold in the table), and therefore they are able to mix effectively, i.e. to lower energy by forming chemical bonds.

This is why chemistry is mainly the science of outer shell orbitals.

AO basis set

The mathematical meaning of LCAO. From mathematical point of view, formula (8.49) represents a expansion of an unknown function φ_i in a series of the known functions χ_s, which belong to a certain complete set, thus M should be equal ∞. In real life, we need to truncate this series, i.e. use some limited M.

8.4.5 BASIS SETS OF ATOMIC ORBITALS

> BASIS SET
> The set of the AOs $\{\chi_s\}$ used in the LCAO expansion is called a basis set.

The choice of the basis set functions χ (the incomplete set) is one of the most important *practical* (numerical) problems of quantum chemistry. Yet, because it is of a technical character, we will just limit ourselves to a few remarks.

Although atomic functions do not need to be *atomic* orbitals (e.g., they may be placed *in-between* nuclei), in most cases they are centred directly on the nuclei[62] of the atoms belonging to the molecule under consideration. If M is small (in the less precise calculations), the Slater atomic orbitals discussed above are often used as the expansion functions χ_s; for larger M (in more accurate calculations), the

---

[61] J.B. Mann, "*Atomic Structure Calculations. I. Hartree–Fock Energy Results for the Elements H through Lr*", Report LA-3690 (Los Alamos National Laboratory, 1967).

[62] It is about the choice of the local coordinate system at the nucleus.

relation between χ_s and the orbitals of the isolated atoms is lost, and χ_s are chosen based on the numerical experience gathered from the literature.[63]

8.4.6 THE HARTREE–FOCK–ROOTHAAN METHOD (SCF LCAO MO)

Clemens C.J. Roothaan (b. 1916), American physicist, professor at the University of Chicago. He became interested in this topic, after recognizing that in the literature people write about the effective one-electron operator, but he could not find its mathematical expression.

George G. Hall (b. 1925), Irish physicist, professor of Mathematics at the University of Nottingham. His scientific achievements are connected to localized orbitals, ionization potentials, perturbation theory, solvation and chemical reactions.

The Hartree–Fock (HF) equations are nonlinear differential-integral equations, which can be solved by appropriate numerical methods. For example, in the case of atoms and diatomics the orbitals may be obtained in a numerical form.[64] High accuracy at long distances from the nuclei is their great advantage. However, the method is very difficult to apply for larger systems.

A solution is the use of the LCAO MO method (algebraization of the Fock equations). It leads to simplification of the computational scheme of the Hartree–Fock method.[65] If the LCAO expansion is introduced to the expression for the total energy, then formula (8.41) (together with $\varepsilon_i = (i|\hat{F}|i)$) gives:

$$E'_{\mathrm{HF}} = \sum_i \big[h_{ii} + (i|\hat{F}|i)\big] = \sum_{i=1}^{\mathrm{MO}} \sum_{rs} c^*_{ri} c_{si} \big[(r|\hat{h}|s) + (r|\hat{F}|s)\big]$$
$$\equiv \frac{1}{2} \sum_{rs} P_{sr}[h_{rs} + F_{rs}], \tag{8.50}$$

bond-order matrix

where $\boldsymbol{P}$ in the RHF method is called the *bond-order matrix*,

---

[63] For those who love such problems, we may recommend the article by S. Wilson “*Basis Sets*” in the book “*Ab initio Methods in Quantum Chemistry*”, ed. by K.P. Lawley, 1987, p. 439. In fact this knowledge is a little magic. Certain notations describing the quality of basis sets are in common use. For example, the symbol 6-31G* means that the basis set uses GTOs (G), the hyphen divides two electronic shells (here K and L, see p. 381). The K shell is described by a single atomic orbital, which is a certain linear combination (a “contracted orbital”) of six GTOs of the 1s type, and the two digits, 31, pertain to the L shell and denote two contracted orbitals for each valence orbital (2s,2p$_x$,2p$_y$,2p$_z$), one of these contains three GTOs, the other one GTO (the latter is called “contracted”, with a bit of exaggeration). The starlet corresponds to d functions used additionally in the description of the L shell (called polarization functions).

[64] J. Kobus, *Adv. Quantum Chem.* 28 (1997) 1.

[65] The LCAO approximation was introduced to the Hartree–Fock method, independently, by C.C.J. Roothaan, *Rev. Modern Phys.* 23 (1951) 69 and G.G. Hall, *Proc. Royal Soc.* A205 (1951) 541.

$$\boxed{P_{sr} = 2\sum_{i}^{\text{MO}} c_{ri}^{*} c_{si},}$$

and the summation goes over all the occupied MOs. The symbols h_{rs} and F_{rs}, introduced here, are the matrix elements of the corresponding operators. In consequence, a useful expression for the total energy in the HF method may be written as

$$E_{\text{HF}} = \frac{1}{2}\sum_{rs}^{\text{AO}} P_{sr}(h_{rs} + F_{rs}) + \sum_{a<b} \frac{Z_a Z_b}{R_{ab}}, \tag{8.51}$$

where the first summation goes over the atomic orbitals (AO). For completeness, we also give the expression for F_{rs}

$$F_{rs} = \big(r\big|\hat{h} + 2\hat{\mathcal{J}} - \hat{\mathcal{K}}\big|s\big) = h_{rs} + \sum_{i}^{\text{MO}} \big[2(ri|si) - (ri|is)\big], \tag{8.52}$$

where i is the index of a MO, and r and s denote the AOs.

Expressing everything in AOs we obtain:

$$\begin{aligned} F_{rs} &= h_{rs} + \sum_{i}^{\text{MO}} \sum_{pq}^{\text{AO}} c_{pi}^{*} c_{qi} \big[2(rp|sq) - (rp|qs)\big] \\ &= h_{rs} + \sum_{pq}^{\text{AO}} P_{qp}\left[(rp|sq) - \frac{1}{2}(rp|qs)\right], \end{aligned} \tag{8.53}$$

where the summation goes over the AOs. We will use these formulae in the future.

In the SCF LCAO MO method, the Fock equations (complicated differential-integral equations) are solved in a very simple way. From (8.49) and (8.30) we have

$$\hat{F}\sum_{s} c_{si}\chi_s = \varepsilon_i \sum_{s} c_{si}\chi_s. \tag{8.54}$$

Making the scalar product with χ_r for $r = 1, 2, \ldots, M$ we obtain

$$\sum_{s} (F_{rs} - \varepsilon_i S_{rs}) c_{si} = 0. \tag{8.55}$$

This is equivalent to the *Roothaan matrix equation*:[66]

Roothaan matrix equation

$$\boxed{\boldsymbol{F}\boldsymbol{c} = \boldsymbol{S}\boldsymbol{c}\boldsymbol{\varepsilon} \qquad (8.56)}$$

where $\boldsymbol{S}$ is the matrix of the overlap integrals $\langle \chi_r|\chi_s\rangle$ involving the AOs, $\boldsymbol{\varepsilon}$ is the diagonal matrix of the orbital energies[67] ε_i, and $\boldsymbol{F}$ is the Fock operator matrix.

---

[66] Left-hand side: $\sum_s F_{rs}c_{si}$, right-hand side: $\sum_{s,l} S_{rs}c_{sl}\varepsilon_{li} = \sum_{s,l} S_{rs}c_{sl}\delta_{li}\varepsilon_i = \sum_s S_{rs}c_{si}\varepsilon_i$. Comparison of both sides of the equation gives the desired result.

[67] In fact some approximations to them. Their values approach the orbital energies, when the basis set of AOs gets closer to the complete basis set.

Each of these matrices is square (of the rank M). $\boldsymbol{F}$ depends on $\boldsymbol{c}$ (and this is why it is a *pseudo*-eigenvalue equation).

The Hartree–Fock–Roothaan matrix equation is solved iteratively:

a) we assume an initial $\boldsymbol{c}$ matrix (i.e. also an initial $\boldsymbol{P}$ matrix; often in the zero-th iteration we put $\boldsymbol{P} = \boldsymbol{0}$, as if there were no electron repulsion),
b) we find the $\boldsymbol{F}$ matrix using matrix $\boldsymbol{P}$,
c) we solve the Hartree–Fock–Roothaan equation (see Appendix L, p. 984) and obtain the M MOs, we choose the $N/2$ occupied orbitals (those of lowest energy),
d) we obtain a new $\boldsymbol{c}$ matrix, and then a new $\boldsymbol{P}$, etc.,
e) we go back to a).

The iterations are terminated when the total HF energy (more liberal approach) or the coefficients $\boldsymbol{c}$ (less liberal one) change less than the assumed threshold values. Both these criteria (ideally fulfilled) may be considered as a sign that the output orbitals are already *self-consistent*. Practically, these are never the exact solutions of the Fock equations, because a *limited* number of AOs was used, while expansion to the complete set requires the use of an infinite number of AOs (the total energy in such a case would be called the *Hartree–Fock limit energy*).

Hartree–Fock limit

After finding the MOs (hence, also the HF function) in the SCF LCAO MO approximation, we may calculate the total energy of the molecule as the mean value of its Hamiltonian. We need only the *occupied* orbitals, and not the *virtual* ones for this calculation.

The Hartree–Fock method only takes care of the total energy *and completely ignores the virtual orbitals*, which may be considered as a kind of by-product.

8.4.7 PRACTICAL PROBLEMS IN THE SCF LCAO MO METHOD

Size of the AO basis set

NUMBER OF MOs

The number of MOs obtained from the SCF procedure is always equal to the number of the AOs used. Each MO consists of various contributions of the same basis set of AOs (the apparent exception is when, *due to symmetry*, the coefficients at some AOs are equal to zero).

For double occupancy, M needs to be larger or equal to $N/2$. Typically we are forced to use large basis sets ($M \gg N/2$), and then along the occupied orbitals we get $M - N/2$ unoccupied orbitals, which are also called *virtual orbitals*. Of course, we should aim at the best *quality* MOs (i.e. how close they are to the solutions of the Fock equations), and avoiding large M (computational effort is proportional to M^4), but in practice a better basis set often means a larger M. The variational

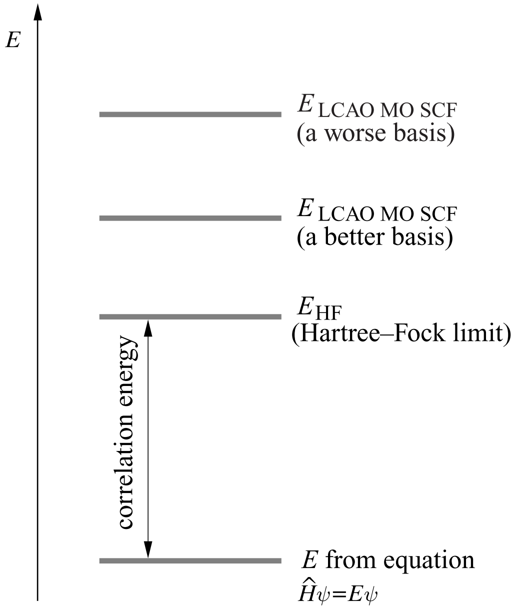


Fig. 8.11. The Hartree–Fock method is variational. The better the wave function, the lower the mean value of the Hamiltonian. An extension of the AO basis set (i.e. adding new AOs) *has to* lower the energy, and the ideal solution of the Fock equations gives the "Hartree–Fock limit". The ground-state eigenvalue of the Hamiltonian is thus *always* lower than the limit.

principle implies the ordering of the total energy values obtained in different approximations (Fig. 8.11).

It is required that as large a basis set as possible is used (mathematics: we approach the complete set), but we may also ask if a basis set dimension may be *decreased* freely (economy!). Of course, the answer is no! The absolute limit M is equal to half the number of the electrons, because only then can we create M spinorbitals and write the Slater determinant. However, in quantum chemistry rather misleadingly, we call the minimal basis set the basis set resulting from inner shell and valence orbitals in the corresponding atoms. For example, the minimum basis set for a water molecule is: $1s$, $2s$ and three $2p$ orbitals of oxygen and two $1s$ orbitals of hydrogen atoms, seven AOs in total (while the truly minimal basis would contain only $10/2 = 5$ AOs).

minimal basis set

"Flip-flop"

The M MOs result from each iteration. We order them using the increasing orbital energy ε criterion, and then we use the $N/2$ orbitals of the lowest orbital energy in the Slater determinant – we call it the occupation of MOs by electrons. We might ask why we make the lowest lying MOs occupied? *The variational principle does not hold for orbital energies.* And yet we do so (not trying all possible occupations), and only very rarely we get into trouble. The most frequent trouble is that the criterion of orbital energy leads to the occupation of one set of MOs in odd iterations, and another set of MOs in even ones (typically both sets differ by including/excluding one of the two MOs that are neighbours on the energy scale) and the energy re-

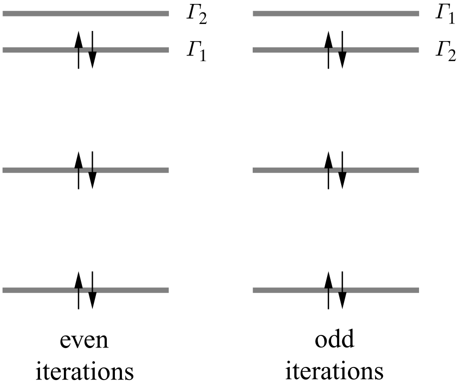


Fig. 8.12. A difficult case for the SCF method ("flip-flop"). We are sure that the orbitals exchange in subsequent iterations, because they differ in symmetry (Γ_1, Γ_2).

sulting from the odd iterations is different from that of the even ones.[68] Such behaviour of the Hartree–Fock method is indeed annoying[69] (Fig. 8.12).

Dilemmas of the AOs centring

Returning to the total energy issue, we should recall that in order to decrease the total energy, we may move the nuclei (so far frozen during the HF procedure). This is called the geometry optimization. Practically all calculations should be repeated for each nuclear geometry during such optimization.[70] And there is one more subtlety. As was said before, the AOs are most often centred on the nuclei. When the nuclei are moved, the question arises whether a nucleus should pull its AOs to a new place, or not.[71] If *not*, then this "slipping off" the nuclei will significantly increase the energy (independent of, whether the geometry is improved or not). If *yes*, then in fact we use different basis sets for each geometry, hence in each case we search for the solution in a slightly different space (because it is spanned by other basis sets). People use the second approach. It is worth notifying that the problem would disappear if the basis set of AOs were complete.

geometry optimization

The problem of AO centring is a bit shameful in quantum chemistry. Let us consider the LCAO approximation and a real molecule such as Na_2CO_3. As mentioned above, the LCAO functions have to form a complete set. But which functions? Since they have to form a complete set, they may be chosen as the eigenfunc-

---

[68] "Flip-flop" is the common name for this sort of behaviour.

[69] There are methods for mastering this rodeo by using the matrix $\boldsymbol{P}$ in the k-th iteration, not taken from the previous iteration (as usual), but as a certain linear combination of $\boldsymbol{P}$ from the $k-1$ and $k-2$ iterations. When the contribution of $\boldsymbol{P}$ from the $k-2$ iteration is large, in comparison with that from the $k-1$ iteration, it corresponds to a gentle attempt at quietening the nervous stallion.

[70] Let us take an example of CH_4. First, we set any starting geometry, say, a square-like planar. Now, we try to change the configuration to make it out-of-plane (the energy goes down). Taking the HCH angles as all equal (tetrahedral configuration) once more lowers the total energy computed. Putting all the CH bonds of equal length gives even lower energy. Finally, by trying different CH bond lengths we arrive at the optimum geometry (for a given AO basis set). In practice, such geometry changes are made automatically by computing the gradient of total energy. The geometry optimization is over.

[71] Even if the AOs were off the nuclei, we would have the same dilemma.

tions of a certain Hermitian operator (e.g., the energy operator for the harmonic oscillator or the energy operator for the hydrogen atom or the uranium atom). We decide, and we are free to choose. In addition to this freedom, we add another freedom, that of the centring. Where should the eigenfunctions (of the oscillator, hydrogen or uranium atom) of the complete set be centred, i.e. positioned in space? Since it is the complete set, *each way of centring is OK by definition*. It really looks like this if we hold to principles.

But in practical calculations, we never have the complete set at our disposal. We always need to limit it to a certain finite number of functions, and it does not represent any complete set. Depending on our computational resources, we limit the number of functions. We usually try to squeeze the best results from our time and money. How do we do it? We apply our physical intuition to the problem, believing that it will pay off. First of all, intuition suggests the use of functions for some atom which is present in the molecule, and not those of the harmonic oscillator, or the hydrogen or uranium atom, which are absent from our molecule. And here we meet another problem. *Which* atom, because we have Na, C and O in Na_2CO_3. It appears that

> the solution close to optimum is to take as a basis set *the beginnings* of several complete sets – each of them centred on one of the atoms.

So, we could centre the $1s$, $2s$, $2p$, $3s$ orbitals on both Na atoms, and the $1s$, $2s$, $2p$ set on the C and O atoms.[72]

8.5 BACK TO FOUNDATIONS...

8.5.1 WHEN DOES THE RHF METHOD FAIL?

The reason for any Hartree–Fock method failure can be only one thing: the wave function is approximated as a *single* Slater determinant. All possible catastrophes come from this. And we might even deduce when the Hartree–Fock method is not appropriate for description of a particular real system. First, let us ask when a single determinant would be OK? Well, if out of all determinants which could be constructed from a certain spinorbital basis set, only its energy (i.e. the mean value of Hamiltonian for this determinant) were close to the true energy of the molecule. In such a case, only this determinant would matter in the linear combination of

---

[72] This is nearly everything, except for a small paradox, that if we are moderately poor (reasonable but not extensive basis sets), then our results will be good, but if we became rich (and we perform high-quality computations using very large basis sets for each atom) then we would get into trouble. This would come from the fact that our basis set starts to look like six distinct complete sets. Well, that looks too good, doesn't it? We have an *overcomplete* set, and trouble must come. The overcompleteness means that any orbital from one set is already representable as a linear combination of another complete set. You would see strange things when trying to diagonalize the Fock matrix. No way! Be sure that you would be begging to be less rich.

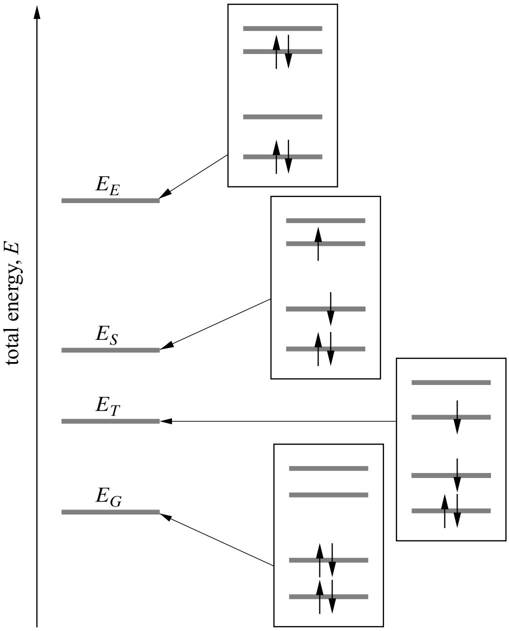


Fig. 8.13. In exact theory there is no such a thing as molecular orbitals. In such a theory we would only deal with the many-electron states and the corresponding energies of the molecule. If, nevertheless, we decided to *stick to the one-electron approximation*, we would have the MOs and the corresponding orbital energies. These one-electron energy levels can be occupied by electrons (0,1 or 2) in various ways (the meaning of the occupation is given on p. 342), and a many-electron wave function (a Slater determinant) corresponds to each occupation. This function gives a certain mean value of the Hamiltonian, i.e. the total energy of the molecule. *In this way one value of the total energy of the molecule corresponds to a diagram of orbital occupation*. The case of the S and T states is somewhat more complex than the one shown here, and we will come back to it on p. 390.

determinants,[73] and the others would have negligible coefficients. It could be so,[74] if the energies of the occupied orbitals were much lower than those of the virtual ones ("*Aufbau Prinzip*", p. 380). Indeed, various electronic states of different total energies may be approximately formed while the orbitals scheme is occupied by electrons (Fig. 8.13), and if the virtual levels are at high energies, the total energy calculated from the "excited determinant" (replacement: occupied spinorbital → virtual spinorbital) would also be high.

In other words, the danger for the RHF method is when the energy difference between HOMO and LUMO is small. For example, RHF completely fails to describe metals properly[75] Always, when the HOMO–LUMO gap is small, expect bad results.

Incorrect description of dissociation by the RHF method

An example is provided by the H_2 molecule at long internuclear distances.

bonding orbital

In the simplest LCAO MO approach, two electrons are described by the *bonding orbital* (χ_a and χ_b are 1*s* orbitals centred on the H nuclei, *a* and *b*, respectively)

---

[73] The Slater determinants form the complete set, p. 334. In the configuration interaction method (which will be described in Chapter 10) the electronic wave function is expanded using Slater determinants.

[74] We shift here from the total energy to the one-electron energy, i.e. to the orbital picture.

[75] It shows up as strange behaviour of the total energy per metal atom, which exhibits poorly-decaying oscillations with an increasing of numbers of atoms. In addition, the exchange interactions, notorious for fast (exponential) decay as calculated by the Hartree–Fock method, are of a long-range character (see Chapter 9).

$$\varphi_{\text{bond}} = \frac{1}{\sqrt{2(1+S)}}(\chi_a + \chi_b), \tag{8.57}$$

but there is another orbital, an *antibonding* one

antibonding orbital

$$\varphi_{\text{antibond}} = \frac{1}{\sqrt{2(1-S)}}(\chi_a - \chi_b). \tag{8.58}$$

These names stem from the respective energies. For the bonding orbital:

$$E_{\text{bond}} = \frac{H_{aa} + H_{ab}}{1+S} < H_{aa},$$

and for the antibonding orbital

$$E_{\text{antibond}} = \frac{H_{aa} - H_{ab}}{1-S} > H_{aa}.$$

These approximate formulae are obtained if we accept that the molecular orbital satisfies a sort of "Schrödinger equation" using an effective Hamiltonian (say, an analogue of the Fock operator): $\hat{H}_{ef}\varphi = E\varphi$ and after introducing notation: the overlap integral $S = (\chi_a|\chi_b)$, $H_{aa} = (\chi_a|\hat{H}_{ef}\chi_a)$, *the resonance integral*[76] $H_{ab} = H_{ba} = (\chi_a|\hat{H}_{ef}\chi_b) < 0$. The resonance integral H_{ab}, and the overlap integral S, decay exponentially when the internuclear distance R increases.

resonance integral

INCORRECT DISSOCIATION LIMIT OF THE HYDROGEN MOLECULE

Thus we have obtained the *quasi-degeneracy* (a near degeneracy of two orbitals) for long distances, while we need to occupy *only one* of these orbitals (bonding one) in the HF method. The antibonding orbital is treated as virtual, and as such, is *completely ignored*. However, as a matter of fact, for long distances R, it corresponds to the same energy as the bonding energy.

We have to pay for such a huge drawback. And the RHF method pays, for its result significantly deviates (Fig. 8.14) from the exact energy for large R values (tending to the energy of the two isolated hydrogen atoms). This effect is known as an "incorrect dissociation of a molecule" in the RHF method (here exemplified by the hydrogen molecule). The failure may be explained in several ways and we have presented one point of view above.

incorrect dissociation

If one bond is broken and another is formed in a molecule, the HF method does not need to fail. It appears that RHF performs quite well in such a situation, because two errors of similar magnitude (Chapter 10) cancel each other.[77]

---

[76] This integral is negative. It is its sign which decides the energy effect of the chemical bond formation (because H_{aa} is nearly equal to the energy of an electron in the H atom, i.e. $-\frac{1}{2}$ a.u.).

[77] Yet the description of the transition state (see Chapter 14) is then of lower quality.

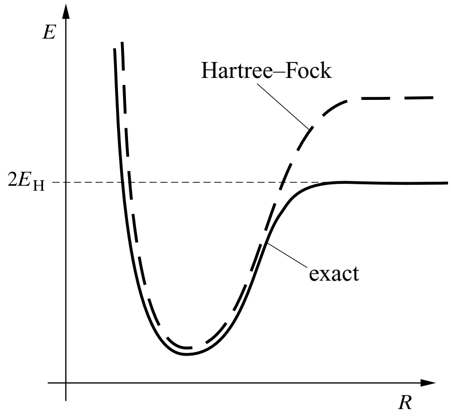


Fig. 8.14. Incorrect dissociation of H_2 in the molecular orbital (i.e. HF) method. The wave function in the form of one Slater determinant leads to dissociation products, which are neither atoms, nor ions (they should be two ground-state hydrogen atoms with energy $2E_H = -1$ a.u.).

8.5.2 FUKUTOME CLASSES

Symmetry dilemmas and the Fock operator

We have derived the general Hartree–Fock method (GHF, p. 341) providing completely free variations for the spinorbitals taken from formula (8.1). As a result, the Fock equation of the form (8.26) was derived.

We then decided to limit the spinorbital variations *via our own* condition of the double occupancy of the molecular orbitals as the real functions. This has led to the RHF method and to the Fock equation in the form (8.30).

The Hartree–Fock method is a complex (nonlinear) procedure. Do the HF solutions have any symmetry features as compared to the Hamiltonian ones? This question may be asked both for the GHF method, and also for any *spinorbital constraints* (e.g., the RHF constraints). The following problems may be addressed:

- Do the output orbitals belong to the irreducible representations of the symmetry group (Appendix C on p. 903) of the Hamiltonian? Or, if we set the nuclei in the configuration corresponding to symmetry group G, will the canonical orbitals transform according to some irreducible representations of the G group? Or, still in other words, does the Fock operator exhibits the symmetry typical of the G group?
- Does the same apply to electron density?
- Is the Hartree–Fock determinant an eigenfunction of the $\hat{S}^2$ operator?[78]
- Is the probability density of finding a $\sigma = \frac{1}{2}$ electron equal to the probability density of finding a $\sigma = -\frac{1}{2}$ electron at any point of space?

Instabilities in the Hartree–Fock method

stability of solutions

The above questions are connected to the *stability of the solutions*. The HF solution is stable if any change of the spinorbitals leads to a *higher* energy than the one

---

[78] For $\hat{S}_z$ it is always an eigenfunction.

found before. We may put certain conditions for spinorbital changes. *Relaxing the condition of double occupancy may take various forms*, e.g., the paired orbitals may be equal but *complex*, or all orbitals may be *different* real functions, or we may admit them as different complex functions, etc. Could the energy increase along with this gradual orbital constraints removal? No, an energy increase is, of course, impossible, because of the variational principle, the energy might, however, remain constant or decrease.

The general answer to this question (the character of the energy change) cannot be given since it depends on various things, such as the molecule under study, interatomic distances, the AOs basis set, etc. However, as shown by Fukutome[79] using a group theory analysis, there are exactly eight situations which *may* occur. Each of these leads to a characteristic shape of the set of occupied orbitals, which is given in Table 8.1. We may pass the borders between these eight classes of GHF method solutions while changing various parameters.[80]

The Fukutome classes may be characterized according to total spin as a function of position in space:

- The first two classes RHF (TICS) and CCW correspond to identical electron spin densities for α and β electrons at any point of space (total spin density equal to zero). This implies double orbital occupancy (the orbitals are real in RHF, and complex in CCW).
- The further three classes ASCW, ASDW and ASW are characterized by the non-vanishing spin density keeping a certain direction (hence A = *axial*). The popular ASDW, i.e. UHF (no. 4) class is worth mentioning.[81] We will return to the UHF function in a moment.
- The last three classes TSCW, TSDW, TSW correspond to spin density with a total non-zero spin, where direction in space varies in a complex manner.[82]

UHF method

The Fukutome classes allow some of the posed questions to be answered:

- The resulting RHF MOs *may* belong (and most often do) to the irreducible symmetry representations (Appendix C in p. 903) of the Hamiltonian. But this is not necessarily the case.
- In the majority of calculations, the RHF electron density shows (at molecular geometry close to the equilibrium) spatial symmetry identical with the point symmetry group (the nuclear configuration) of the Hamiltonian. But the RHF method may also lead to *broken symmetry solutions*. For example, a system composed of the equidistant H atoms uniformly distributed on a circle shows bond alternation, i.e. symmetry breaking of the BOAS type.[83]

broken symmetry

BOAS

---

[79] A series of papers by H. Fukutome starts with the article in *Prog. Theor. Phys.* 40 (1968) 998 and the review article *Int. J. Quantum Chem.* 20 (1981) 955. I recommend a beautiful paper by J.-L. Calais, *Adv. Quantum Chem.* 17 (1985) 225.

[80] In the space of the parameters it is something like a phase diagram for a phase transition.

[81] UHF, i.e. *Unrestricted Hartree–Fock*.

[82] See J.-L. Calais, *Adv. Quantum Chem.* 17 (1985) 225.

[83] BOAS stands for the *Bond-Order Alternating Solution*. It has been shown, that the translational symmetry is broken and that the symmetry of the electron density distribution in polymers exhibits a unit

Table 8.1. Fukutome classes (for φ_{i1} and φ_{i2} see eq. (8.1))

Class	Orbital components $\begin{bmatrix}\varphi_{11} & \varphi_{21} & \cdots & \varphi_{N1}\\ \varphi_{12} & \varphi_{22} & \cdots & \varphi_{N2}\end{bmatrix}$	Remarks	Name
1	$\begin{bmatrix}\varphi_1 & 0 & \varphi_2 & 0 & \ldots & \varphi_{N/2} & 0\\ 0 & \varphi_1 & 0 & \varphi_2 & \cdots & 0 & \varphi_{N/2}\end{bmatrix}$	φ_i real	RHF≡TICS[1]
2	$\begin{bmatrix}\varphi_1 & 0 & \varphi_2 & 0 & \ldots & \varphi_{N/2} & 0\\ 0 & \varphi_1 & 0 & \varphi_2 & \cdots & 0 & \varphi_{N/2}\end{bmatrix}$	φ_i complex	CCW[2]
3	$\begin{bmatrix}\varphi_1 & 0 & \varphi_2 & 0 & \ldots & \varphi_{N/2} & 0\\ 0 & \varphi_1^* & 0 & \varphi_2^* & \cdots & 0 & \varphi_{N/2}^*\end{bmatrix}$	φ_i complex	ASCW[3]
4	$\begin{bmatrix}\varphi_1 & 0 & \varphi_2 & 0 & \ldots & \varphi_{N/2} & 0\\ 0 & \chi_1 & 0 & \chi_2 & \cdots & 0 & \chi_{N/2}\end{bmatrix}$	φ, χ real	UHF≡ ASDW[4]
5	$\begin{bmatrix}\varphi_1 & 0 & \varphi_2 & 0 & \ldots & \varphi_{N/2} & 0\\ 0 & \chi_1 & 0 & \chi_2 & \cdots & 0 & \chi_{N/2}\end{bmatrix}$	φ, χ complex	ASW[5]
6	$\begin{bmatrix}\varphi_1 & \chi_1 & \varphi_2 & \chi_2 & \cdots & \varphi_{N/2} & \chi_{N/2}\\ -\chi_1^* & \varphi_1^* & -\chi_2^* & \varphi_2^* & \cdots & -\chi_{N/2}^* & \varphi_{N/2}^*\end{bmatrix}$	φ, χ complex	TSCW[6]
7	$\begin{bmatrix}\varphi_1 & \chi_1 & \varphi_2 & \chi_2 & \cdots & \varphi_{N/2} & \chi_{N/2}\\ \tau_1 & \kappa_1 & \tau_2 & \kappa_2 & \cdots & \tau_{N/2} & \kappa_{N/2}\end{bmatrix}$	$\varphi, \chi, \tau, \kappa$ real	TSDW[7]
8	$\begin{bmatrix}\varphi_1 & \chi_1 & \varphi_2 & \chi_2 & \cdots & \varphi_{N/2} & \chi_{N/2}\\ \tau_1 & \kappa_1 & \tau_2 & \kappa_2 & \cdots & \tau_{N/2} & \kappa_{N/2}\end{bmatrix}$	$\varphi, \chi, \tau, \kappa$ complex	TSW[8]

[1] Also, according to Fukutome, TICS, i.e. *Time-reversal-Invariant Closed Shells.*
[2] *Charge Current Waves.*
[3] *Axial Spin Current Waves.*
[4] *Axial Spin Density Waves.*
[5] *Axial Spin Waves.*
[6] *Torsional Spin Current Waves.*
[7] *Torsional Spin Density Waves.*
[8] *Torsional Spin Waves.*

triplet instability

- The RHF function is always an eigenfunction of the $\hat{S}^2$ operator (and, of course, of the $\hat{S}_z$). This is no longer true, when extending beyond the RHF method (*triplet instability*).
- The probability densities of finding the $\sigma = \frac{1}{2}$ and $\sigma = -\frac{1}{2}$ electron coordinate are different for the majority of Fukutome classes ("*spin waves*").

Example: Triplet instability

The wave function in the form of a Slater determinant is always an eigenfunction of the $\hat{S}_z$ operator, and if in addition double occupancy is assumed (RHF) then it is also an eigenfunction of the $\hat{S}^2$ operator, as exemplified by the hydrogen molecule in Appendix Q on p. 1006.

---

cell twice as long as that of the nuclear pattern [J. Paldus, J. Čižek, *J. Polym. Sci., Part C* 29 (1970) 199, also J.-M. André, J. Delhalle, J.G. Fripiat, G. Hennico, J.-L. Calais, L. Piela, *J. Mol. Struct. (Theochem)* 179 (1988) 393]. The BOAS represents a feature related to the Jahn–Teller effect in molecules and to the Peierls effect in the solid state (see Chapter 9).

And what about the UHF method?

Let us study the two electron system, where the RHF function (the TICS Fukutome class) is:

$$\psi_{\mathrm{RHF}} = \frac{1}{\sqrt{2}} \begin{vmatrix} \phi_1(1) & \phi_1(2) \\ \phi_2(1) & \phi_2(2) \end{vmatrix},$$

and both spinorbitals have a *common real orbital part* φ: $\phi_1 = \varphi\alpha$, $\phi_2 = \varphi\beta$.

Now we allow for a *diversification* of the orbital part (keeping the functions *real*, i.e. staying within the ASDW Fukutome class, usually called UHF in quantum chemistry) for both spinorbitals. We proceed slowly from the closed-shell situation, using as the orthonormal spinorbitals:

$$\phi_1' = N_-(\varphi - \delta)\alpha, \qquad \phi_2' = N_+(\varphi + \delta)\beta,$$

where δ is a small real correction to the φ function, and N_+ and N_- are the normalization factors.[84] The electrons hate each other (Coulomb law) and may thank us for giving them separate apartments: $\varphi + \delta$ and $\varphi - \delta$. We will worry about the particular mathematical shape of δ in a minute. For the time being let us see what happens to the UHF function:

$$\begin{aligned}
\psi_{\mathrm{UHF}} &= \frac{1}{\sqrt{2}} \begin{vmatrix} \phi_1'(1) & \phi_1'(2) \\ \phi_2'(1) & \phi_2'(2) \end{vmatrix} \\
&= \frac{1}{\sqrt{2}} N_+ N_- \begin{vmatrix} [\varphi(1) - \delta(1)]\alpha(1) & [\varphi(2) - \delta(2)]\alpha(2) \\ \phi_2'(1) & \phi_2'(2) \end{vmatrix} \\
&= \frac{1}{\sqrt{2}} N_+ N_- \left\{ \begin{vmatrix} \varphi(1)\alpha(1) & \varphi(2)\alpha(2) \\ \phi_2'(1) & \phi_2'(2) \end{vmatrix} - \begin{vmatrix} \delta(1)\alpha(1) & \delta(2)\alpha(2) \\ \phi_2'(1) & \phi_2'(2) \end{vmatrix} \right\} \\
&= \frac{1}{\sqrt{2}} N_+ N_- \left\{ \begin{aligned} &\begin{vmatrix} \varphi(1)\alpha(1) & \varphi(2)\alpha(2) \\ [\varphi(1) + \delta(1)]\beta(1) & [\varphi(2) + \delta(2)]\beta(2) \end{vmatrix} - \\ &\begin{vmatrix} \delta(1)\alpha(1) & \delta(2)\alpha(2) \\ [\varphi(1) + \delta(1)]\beta(1) & [\varphi(2) + \delta(2)]\beta(2) \end{vmatrix} \end{aligned} \right\} \\
&= \frac{1}{\sqrt{2}} N_+ N_- \left\{ \begin{aligned} &\begin{vmatrix} \varphi(1)\alpha(1) & \varphi(2)\alpha(2) \\ \varphi(1)\beta(1) & \varphi(2)\beta(2) \end{vmatrix} + \begin{vmatrix} \varphi(1)\alpha(1) & \varphi(2)\alpha(2) \\ \delta(1)\beta(1) & \delta(2)\beta(2) \end{vmatrix} - \\ &\begin{vmatrix} \delta(1)\alpha(1) & \delta(2)\alpha(2) \\ \varphi(1)\beta(1) & \varphi(2)\beta(2) \end{vmatrix} - \begin{vmatrix} \delta(1)\alpha(1) & \delta(2)\alpha(2) \\ \delta(1)\beta(1) & \delta(2)\beta(2) \end{vmatrix} \end{aligned} \right\} \\
&= N_+ N_- \psi_{\mathrm{RHF}} \\
&\quad + \frac{1}{\sqrt{2}} N_+ N_- \big\{ \big[\varphi(1)\delta(2) - \varphi(2)\delta(1)\big]\big[\alpha(1)\beta(2) + \alpha(1)\beta(2)\big] \big\} \\
&\quad - \frac{1}{\sqrt{2}} N_+ N_- \begin{vmatrix} \delta(1)\alpha(1) & \delta(2)\alpha(2) \\ \delta(1)\beta(1) & \delta(2)\beta(2) \end{vmatrix}.
\end{aligned}$$

---

[84] Such a form is not fully equivalent to the UHF method, in which a general form of real orbitals is allowed.

The first and last functions are singlets ($S_z = 0, S = 0$), while the second function represents a *triplet state* ($S_z = 0$, $S = 1$), Appendix Q on p. 1006. Thus a small diversification of the orbital functions leads to some *triplet (second term) and singlet (third term) admixtures* to the original singlet function $N_+N_-\psi_{\text{RHF}}$ (called *triplet contamination*). The former is proportional to δ and the latter to δ^2. Now the total wave function is no longer an eigenfunction of the $\hat{S}^2$ operator. How is this possible? If one electron has a spin coordinate of $\frac{1}{2}$ and the second one of $-\frac{1}{2}$, aren't they paired? Well, not necessarily, because one of the triplet functions (which describes the *parallel* configuration of both spins[85]) is $[\alpha(1)\beta(2) + \alpha(1)\beta(2)]$.

ASDW

Is the resulting UHF energy (calculated for such a function) lower than the corresponding RHF energy (calculated for ψ_{RHF}), i.e. is the RHF solution unstable towards ASDW-type spinorbitals changes (no. 4 in the Table of Fukutome classes)?

It depends on a particular situation. A while before, we promised to consider what the δ function should look like for the hydrogen molecule. In the RHF method, both electrons occupy the same molecular orbital φ. If we assume within the UHF method that whenever one electron is close to the a nucleus, the second one prefers to be closer to b, this would happily be accepted by the electrons, since they repel each other (the mean value of the Hamiltonian would decrease, this is welcome). Taking the $\delta = \varepsilon\tilde{\varphi}$ function (where $\tilde{\varphi}$ is the antibonding orbital, and $\varepsilon > 0$ is a small coefficient) would have such consequences. Indeed, the sum $\varphi + \delta = \varphi + \varepsilon\tilde{\varphi}$ takes larger absolute value preferentially at one of the nuclei[86] (Fig. 8.15). Since both orbitals correspond to electrons with opposite spins, there will be some net spin on each of the nuclei. This nicely justifies the name of Axial Spin Density Wave (ASDW) Fukutome gave to the UHF method.

AMO method

A similar reasoning pertaining function $\varphi - \delta = \varphi - \varepsilon\tilde{\varphi}$ results in opposite preferences for the nuclei. Such a particular UHF method, which uses virtual orbitals $\tilde{\varphi}$ to change RHF orbitals, carries the friendly name of the AMO approach.[87]

Now,

$$\begin{aligned}
\psi_{\text{UHF}} &= N_+N_-\psi_{\text{RHF}} \\
&\quad + \frac{1}{\sqrt{2}}N_+N_-\varepsilon\big\{\big[\varphi(1)\tilde{\varphi}(2) - \varphi(2)\tilde{\varphi}(1)\big]\big[\alpha(1)\beta(2) + \alpha(1)\beta(2)\big]\big\} \\
&\quad - \frac{1}{\sqrt{2}}N_+N_-\varepsilon^2 \begin{vmatrix} \tilde{\varphi}(1)\alpha(1) & \tilde{\varphi}(2)\alpha(2) \\ \tilde{\varphi}(1)\beta(1) & \tilde{\varphi}(2)\beta(2) \end{vmatrix} \\
&= N_+N_-\big[\psi_{\text{RHF}} + \varepsilon\sqrt{2}\psi_{T''} - \varepsilon^2\psi_E\big],
\end{aligned}$$

---

[85] To call them parallel is an exaggeration, since they form an angle 70.5° (see Chapter 1, p. 28), but this is customary in physics and chemistry.

[86] In our example, the approximate bonding orbital is $\varphi = \frac{1}{\sqrt{2}}(1s_a + 1s_b)$, and $\tilde{\varphi} = \frac{1}{\sqrt{2}}(1s_a - 1s_b)$, hence $\varphi + \varepsilon\tilde{\varphi} = \frac{1}{\sqrt{2}}[(1+\varepsilon)1s_a + (1-\varepsilon)1s_b]$, while $\varphi - \varepsilon\tilde{\varphi} = \frac{1}{\sqrt{2}}[(1-\varepsilon)1s_a + (1+\varepsilon)1s_b]$. Thus one of the new orbitals has a larger amplitude at nucleus a, while the other one has it at nucleus b (as we had initially planned).

[87] *Alternant Molecular Orbitals*; P.-O. Löwdin is its author, *Symp. Mol. Phys.*, Nikko (Tokyo Maruzen), 1954, p. 13, also R. Pauncz, "*Alternant Molecular Orbitals*", Saunders, Philadelphia, 1967.

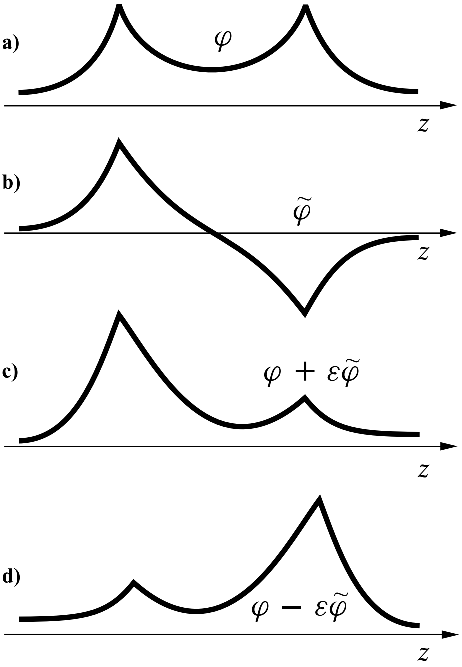


Fig. 8.15. The effect of mixing the bonding orbital φ (Fig. a) with the antibonding orbital $\tilde{\varphi}$ (Fig. b). A small admixture (c) of $\tilde{\varphi}$ to the orbital φ leads to an increase of the probability amplitude of the resulting orbital at the left nucleus, while a subtraction of $\tilde{\varphi}$ (d) leads to a larger probability amplitude of the resulting orbital at the right nucleus. Thus it results in partial separation of the spins $\frac{1}{2}$ and $-\frac{1}{2}$.

where the following notation is used for normalized functions: ψ_{RHF} for the ground state of the energy E_{RHF}, $\psi_{T''}$ for the triplet state of the energy E_T, and ψ_E for the singlet state with a doubly occupied antibonding orbital that corresponds to the energy E_E.

Let us calculate the mean value of the Hamiltonian using the ψ_{UHF} function. Because of the orthogonality of the spin functions (remember that the Hamiltonian is independent of spin) we have $\langle\psi_{\text{RHF}}|\hat{H}\psi_{T''}\rangle = \langle\psi_{\text{RHF}}|\psi_{T''}\rangle = 0$ and obtain (with accuracy up to ε^2 terms)

Per-Olov Löwdin (1916–2000), Swedish chemist and physicist, student of Pauli, professor at the University of Uppsala (Sweden), founder and professor of the Quantum Theory Project at Gainesville University (Florida, USA), very active in organizing the scientific life of the international quantum chemistry community

$$\bar{E}_{\text{UHF}} \approx \frac{\langle\psi_{\text{RHF}}|\hat{H}\psi_{\text{RHF}}\rangle + 2\varepsilon^2\langle\psi_{T''}|\hat{H}\psi_{T''}\rangle - 2\varepsilon^2\langle\psi_{\text{RHF}}|\hat{H}\psi_E\rangle}{\langle\psi_{\text{RHF}}|\psi_{\text{RHF}}\rangle + 2\varepsilon^2\langle\psi_{T''}|\psi_{T''}\rangle}$$

$$= \frac{E_{\text{RHF}} + 2\varepsilon^2 E_T - 2\varepsilon^2(\varphi\varphi|\tilde{\varphi}\tilde{\varphi})}{1 + 2\varepsilon^2} \approx E_{\text{RHF}} + 2\varepsilon^2\big[(E_T - E_{\text{RHF}}) - (\varphi\varphi|\tilde{\varphi}\tilde{\varphi})\big],$$

where the Taylor expansion and the III Slater–Condon rule have been used (p. 986): $\langle\psi_{\text{RHF}}|\hat{H}\psi_E\rangle = (\varphi\varphi|\tilde{\varphi}\tilde{\varphi}) > 0$. The last integral is greater than zero, because it corresponds to the Coulombic self-repulsion of a certain charge distribution.

It is now clear that everything depends on the sign of the square bracket. If $E_T \gg E_{RHF}$, then the spatial diversification of the opposite spin electrons (connected with the stabilization of $-2\varepsilon^2(\varphi\varphi|\tilde{\varphi}\tilde{\varphi})$) will not pay because in such a case $\overline{E}_{UHF} \gg E_{RHF}$. However, if the E_T is close to the ground state energy, then the *total energy*

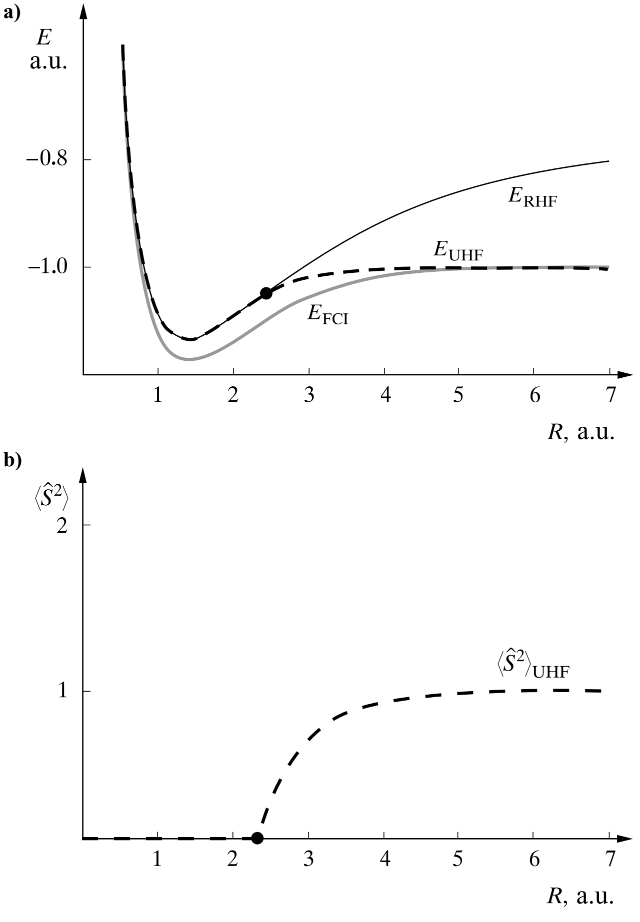


Fig. 8.16. (a) The mean value of Hamiltonian (E) calculated by the RHF and UHF methods. The lowest curve (E_{FCI}) corresponds to the accurate result (called the full configuration interaction method, see Chapter 10). (b) The mean value of the $\hat{S}^2$ operator calculated by the RHF and UHF methods. The energies $E_{RHF}(R)$ and $E_{UHF}(R)$ are identical for internuclear distances $R < 2.30$ a.u. For larger R values the two curves separate, and the RHF method gives an incorrect description of the dissociation limit, while the UHF method still gives a correct dissociation limit. For $R < 2.30$ a.u., the RHF and UHF wave functions are identical, and they correspond to a singlet, while for $R > 2.30$ the UHF wave function has a triplet contamination. T. Helgaker, P. Jørgensen, J. Olsen, "*Molecular Electronic Structure Theory*", Wiley, Chichester, © 2000, reproduced with permission of John Wiley and Sons Ltd.

will decrease upon the addition of the triplet state, i.e. the RHF solutions will be unstable towards the AMO-type change of the orbitals.

This is the picture one obtains in numerical calculations for the hydrogen molecule (Fig. 8.16). At short distances between the atoms (up to 2.30 a.u.) the interaction is strong and the triplet state is of high energy. Then the variational principle does not allow the triplet state to contribute to the ground state and the UHF and the RHF give the same result. But beyond the 2.30 a.u. internuclear distance, the triplet admixture results in a small stabilization of the ground state and the UHF energy is lower than the RHF. For very long distances (when the energy difference between the singlet and triplet states is very small), the energy gain associated with the triplet component is very large.

We can see from Fig. 8.16.b the drama occurring at $R = 2.30$ a.u. for the mean value of the $\hat{S}^2$ operator. For $R < 2.30$ a.u. the wave function preserves the singlet character, for larger R the triplet addition increases fast, and at $R = \infty$ the mean value of the square of the total spin $\hat{S}^2$ is equal to 1, i.e. half-way between the $S(S+1) = 0$ result for the singlet ($S = 0$) and the $S(S+1) = 2$ result for the triplet ($S = 1$), since the UHF determinant is exactly $50\% : 50\%$ singlet and triplet mixture. Thus, *one determinant (UHF) is able to describe properly the dissociation of the hydrogen molecule in its ground state (singlet), but at the expense of a large spin contamination (triplet admixture).*

RESULTS OF THE HARTREE–FOCK METHOD

8.6 MENDELEEV PERIODIC TABLE OF CHEMICAL ELEMENTS

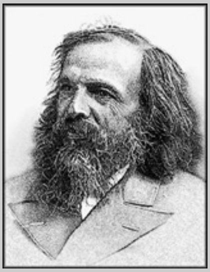
8.6.1 SIMILAR TO THE HYDROGEN ATOM – THE ORBITAL MODEL OF ATOM

The Hartree–Fock method gives an approximate wave function for the atom of *any* chemical element from the Mendeleev periodic table *(orbital picture)*. The Hartree–Fock method stands behind the *orbital model of atoms*. The model says essentially that a single Slater determinant can describe the atom to an accuracy that in most cases satisfies chemists. To tell the truth, the orbital model is in principle false,[88] but it is remarkable that nevertheless the conclusions drawn from it agree with experiment, at least qualitatively. It is quite exciting that

> the electronic structure of all elements can be generated to a reasonable accuracy using the *Aufbau Prinzip*, i.e. a certain scheme of filling the atomic orbitals of the hydrogen atom.

---

[88] Because the contributions of other Slater determinants (configurations) is not negligible (see Chapter 10).



Dimitrii Ivanovich Mendeleev (1834–1907), Russian chemist, professor at the University in Petersburg, and later controller of the Russian Standards Bureau of Weights and Measures (after he was expelled from the University by the tsarist powers for supporting a student protest). He was born in Tobolsk, as the youngest of fourteen children of a headmaster. In 1859 young Mendeleev – thanks to a tsarist scholarship – went to Paris and Heidelberg, where he worked with Robert Bunsen and Gustav Kirchhoff. After getting his Ph.D. in 1865, he became at 32 professor of Chemistry at the University in Sankt Petersburg. Since he had no good textbook, he started to write his own ("Principles of chemistry"). This is when he discovered one of the major human generalizations (1869): the periodicity law of chemical elements.

In 1905 he was nominated for the Nobel Prize, but lost by one vote to Henri Moissan, the discoverer of fluorine. The Swedish Royal Academy thus lost *its* chance, because in a year or so Mendeleev died. Many scientists have had similar intuition as had Mendeleev, but it was Mendeleev who completed the project, who organized the known elements in the Table, and who predicted the existence of unknown elements. The following example shows how difficult it was for science to accept the Periodic Table. In 1864 John Newlands presented to The Royal Society in London his work showing similarities of the light elements, occurring for each eighth element with increasing atomic mass. The President of the meeting, quite amused by these considerations, suggested: "*haven't you tried to organize them according to the alphabetic order of their names?*".

Thus, the simple and robust orbital model serves chemistry as a "work horse". Let us take some examples. All the atoms are build on a similar principle. A nodeless spherically symmetric atomic orbital (called $1s$) of the lowest orbital energy, next, the second lowest (and also the spherically symmetric, one radial node) is called $2s$, etc. Therefore, when filling orbital energy states by electrons some electronic shells are formed: K ($1s^2$), L ($2s^2 2p^6$), ..., where the maximum for shell orbital occupation by electrons is shown.

The very foundations of a richness around us (its basic building blocks being atoms in the Mendeleev Periodic Table) result from a very simple idea, that the proton and electron form a stable system called the hydrogen atom.

8.6.2 YET THERE ARE DIFFERENCES...

The larger the atomic number, the more complex the electronic structure. For neutral atoms the following occupation scheme applies.

Aufbau Prinzip

The Aufbau Prinzip relies on a scheme of orbital energies, Fig. 8.17. We cannot however expect that all nuances of atomic stabilities and of the ions corresponding to them might be deduced from a single simple rule like the Aufbau Prinzip, and not from the hard work of solving the Schrödinger equation (plus also the relativistic effects, Chapter 3) individually for each particular system.

electronic configuration

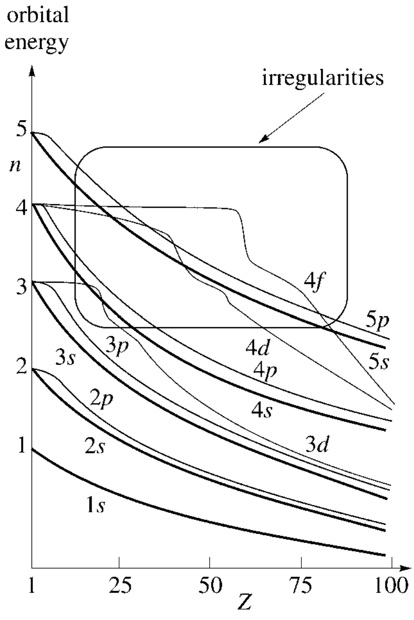


Fig. 8.17. A diagram of the order (in an energy scale) of the orbital energies as functions of the atomic number Z. This diagram, together with the Aufbau Prinzip, allows to write down the electronic configurations of atoms and explains the physical and chemical properties of chemical elements (adapted from P. Atkins, "*Physical Chemistry*", sixth ed., Oxford University Press, Oxford, 1998).

From Fig. 8.17 can see that:

- the orbital energy depends not only on the principal quantum number n, but also on the angular quantum number[89] l, and the larger the l, the higher the energy,
- since for large n the *Aufbau Prinzip* is not always valid, the levels of a given n overlap in the energy scale with the $n' = n + 1$ levels.

Even so, the consecutive occupation of the electronic shells by electrons leads to a quasi-periodicity (sometimes called the periodicity) of the electronic configurations, and in consequence a quasi-periodicity of all chemical and physical properties of the elements.

Example 1. ***Noble gases.*** The atoms He, Ne, Ar, Kr, Xe, Rn have a remarkable feature, that all the subshells below and including ns np subshell are fully occupied.

	configuration	number of electrons
He:	$1s^2$	2
Ne:	$1s^2 2s^2 2p^6$	$10 = 2 + 8$
Ar:	$1s^2 2s^2 2p^6 3s^2 3p^6$	$18 = 2 + 8 + 8$
Kr:	$1s^2 2s^2 2p^6 3s^2 3p^6 3d^{10} 4s^2 4p^6$	$36 = 2 + 8 + 8 + 18$
Xe:	$1s^2 2s^2 2p^6 3s^2 3p^6 3d^{10} 4s^2 4p^6 4d^{10} 5s^2 5p^6$	$54 = 2 + 8 + 8 + 18 + 18$
Rn:	$1s^2 2s^2 2p^6 3s^2 3p^6 3d^{10} 4s^2 4p^6 4d^{10} 5s^2 5p^6 4f^{14} 5d^{10} 6s^2 6p^6$	$86 = 2 + 8 + 8 + 18 + 18 + 32$

---

[89] If the nucleus were large, then orbitals of different l would have different orbital energies. This explains the energy differences for the $s, p, d, \ldots$ levels, because the outer shell electrons move in the field of the nucleus shielded by the inner shell electrons (thus, in a field of something that can be seen as a large pseudo-nucleus).

According to the discussion on p. 363, what chemistry is all about is the outermost occupied orbitals that participate in forming chemical bonds. The noble gases stand out from other elements by completing their electronic shells, no wonder then that they are distinguished by very special chemical properties. The noble gases do not form chemical bonds.[90]

Example 2. ***Alkali metals.*** The atoms Li, Na, K, Rb, Cs, Fr have the following dominant electronic configurations (the inner shells have been abbreviated by reporting the corresponding noble gas atom configuration):

	inner shells	valence configuration
Li	[He]	$2s^1$
Na	[Ne]	$3s^1$
K	[Ar]	$4s^1$
Rb	[Kr]	$5s^1$
Cs	[Xe]	$6s^1$
Fr	[Rn]	$7s^1$

No wonder that the elements Li, Na, K, Rb, Cs, Fr exhibit similar chemical and physical properties. Let us take any property we want, e.g., what will we get if the element is thrown into water. Lithium is a metal that reacts slowly with water, producing a colourless basic solution and hydrogen gas. Sodium is a metallic substance, and with water is a very dangerous spectacle (wild dancing flames). It reacts rapidly with water to form a colourless basic solution and hydrogen gas. The other alkali metals are even more dangerous. Potassium is a metal as well, and reacts very rapidly with water giving a colourless basic solution and hydrogen gas. Rubidium is a metal which reacts very rapidly with water producing a colourless basic solution and hydrogen gas. Cesium metal reacts rapidly with water. The result is a colourless solution and hydrogen gas. Francium is very scarce and expensive, and probably no one has tried its reaction with water. *We may however expect, with very high probability, that if the reaction were made, it would be faster than that with cesium and that a basic solution would be produced*.

However maybe all elements react rapidly with water to form a colourless basic solution and hydrogen gas? Well, this is not true. The noble gases do not. Helium does not react with water. Instead it dissolves slightly in it to the extent of about 8.61 cm^3/kg at 293 K. Also neon does not react with water, but it does dissolve in it – just about 10.5 cm^3/kg at 293 K. Argon, krypton, xenon and radon also do not react with water. They dissolve in it to the extent of 33.6, 59.4, 108.1 and 230 cm^3/kg at 293 K, respectively. It is clear that these elements form a family that does not react with water at 293 K, but instead dissolves (slightly) in water.[91] The reason is that all these elements have closed (i.e. fully occupied) shells, whereas a chemical reaction needs the opening of closed shells (see Chapter 14).

---

[90] We have to add though, that the closed shells of the noble gases can be opened either in extreme physical conditions or by using aggressive compounds. Then, they may form chemical bonds.

[91] Note, that the concentration increases monotonically.

Example 3. ***Halogens.*** Let us see whether there are other families. Let us concentrate on atoms which have p^5 as the outer-most configuration. Using our scheme of orbital energies we produce the following configurations with this property: $[\mathrm{He}]2s^22p^5$ with 9 electrons, i.e. F, $[\mathrm{Ne}]3s^23p^5$ with 17 electrons, i.e. Cl, $[\mathrm{Ar}]3d^{10}4s^24p^5$ with 35 which corresponds to Br, $[\mathrm{Kr}]4d^{10}5s^25p^5$ with 53 electrons which is iodine, $[\mathrm{Xe}]4f^{14}5d^{10}6s^26p^5$ means 85 electrons, i.e. astatine, or At. Are these elements similar? What happens to halogens in contact with water? Maybe they react very rapidly with water producing a colourless basic solution and hydrogen gas like the alkali metals, or do they just dissolve in water like the noble gases? Let us see.

Fluorine reacts with water to produce oxygen, O_2, and ozone O_3. This is strange in comparison with alkali metals. Next, chlorine reacts with water to produce hypochlorite, OCl^-. Bromine and iodine do a similar thing producing hypobromite OBr^- and hypoiodite OI^-. Nothing is known about the reaction of astatine with water. Apart from the exceptional behaviour of fluorine,[92] there is no doubt we have a family of elements. This family is different from the noble gases and from the alkali metals.

Thus, the families show evidence that elements differ widely among families, but much less within a family, with rather small (and often monotonic) changes within it. This is what (quasi) periodicity is all about. The families are called *groups* (usually columns) in the Mendeleev Table.

group

The Mendeleev Periodic Table represents a kind of compass in chemistry. Instead of having a sort of wilderness, where all the elements exhibit their unique physical and chemical properties as *deus ex machina*, we obtain *understanding* that the animals are in a zoo, and are not unrelated, that there are some families, which follow from similar structure and occupancy of the outer electronic shells. Moreover, it became clear that there are cages in the zoo waiting for animals yet to be discovered. The animals could have been described in detail *before they were actually found by experiment*. This periodicity pertains not only to the chemical and physical properties of elements, but also to all parameters that appear in theory and are related to atoms, molecules and crystals.

8.7 THE NATURE OF THE CHEMICAL BOND

As shown on p. 371, the MO method explains the nature of the chemical bond *via* the argument that the orbital energy in the molecule is lower than that in the isolated atom. But why is this so? Which interactions decide bond formation? Do they have their origin in quantum or in classical mechanics?

To answer these questions, we will analyze the simplest case: chemical bonding in a molecular ion H_2^+. It seems that quantum mechanics is not required here: we

---

[92] For light elements the details of the electronic configuration play a more important role. For example, hydrogen may also be treated as an alkali metal, but its properties differ widely from the properties of the other members of this family.

deal with one repulsion and two attractions. No wonder there is bonding, since the net effect is one attraction. But the same applies, however, to the dissociated system (the hydrogen atom and the proton). Thus, the story is becoming more subtle.

8.7.1 H_2^+ IN THE MO PICTURE

Let us analyze chemical bonding as viewed by the poor version of the MO method (only two 1*s* hydrogen atom orbitals are used in the LCAO expansion, see Appendix R on p. 1009). Much can be seen thanks to such a poor version. The mean kinetic energy of the (only) electron of H_2^+, residing on the bonding MO $\varphi = [2(1+S)]^{-1/2}(a+b)$, is given as ($a$ and b denote the atomic 1*s* orbitals centred, respectively, on the a and b nuclei)

$$\bar{T} \equiv (\varphi|\hat{T}\varphi) = \frac{T_{aa}+T_{ab}}{1+S}, \tag{8.59}$$

where S is the overlap integral $S = (a|b)$, and

$$T_{aa} = \left(a\middle|-\frac{1}{2}\Delta\middle|a\right) = T_{bb},$$

$$T_{ab} = \left(a\middle|-\frac{1}{2}\Delta\middle|b\right) = T_{ba}.$$

The non-interacting hydrogen atom and the proton have the mean kinetic energy of the electron equal to T_{aa}. The kinetic energy change is thus

$$\Delta T = \bar{T} - T_{aa} = \frac{T_{ab} - ST_{aa}}{1+S}. \tag{8.60}$$

The denominator is always positive, and the numerator (as known from computational experience) is negative for any internuclear distance. This means that the kinetic energy of the electron decreases upon molecule formation.[93] Hence,

> kinetic energy stabilizes the molecule but not the atom.

Let us note (please recall the a and b functions are the eigenfunctions of the hydrogen atom Hamiltonian), that $T_{ab} = E_H S - V_{ab,b}$ and $T_{aa} = E_H - V_{aa,a}$, where

---

[93] This agrees with intuition, which suggests that an electron now has more space for penetration ("larger box", see p. 145), and the energy levels in the box (potential energy is zero in the box, therefore we mean kinetic energy here) decrease, when the box dimension increases. This example shows that some abstract problems which can be solved exactly (here the particle in the box), serve as a beacon for more complex problems.

E_H is the ground state energy of the H atom,[94] and

$$V_{ab,b} = V_{ab,a} = -\left(a\left|\frac{1}{r_b}\right|b\right),$$
$$V_{aa,a} = -\left(a\left|\frac{1}{r_a}\right|a\right).$$

Now, ΔT can be presented as

$$\Delta T = -\frac{V_{ab,a} - SV_{aa,a}}{1+S}, \tag{8.61}$$

because the terms with E_H cancel each other. In this way the change in kinetic energy of the electron when a molecule is formed may be *formally* presented as the integrals describing the potential energy.

Now let us calculate the change in the mean potential energy. The mean potential energy of the electron (the nucleus–nucleus interaction will be added later) equals to

$$\bar{V} = (\varphi|V|\varphi) = \left(\varphi\left|-\frac{1}{r_a} - \frac{1}{r_b}\right|\varphi\right) = \frac{(V_{aa,a} + V_{aa,b} + 2V_{ab,a})}{1+S} \tag{8.62}$$

while in the hydrogen atom it was equal to $V_{aa,a}$. The difference, ΔV, is

$$\Delta V = \frac{(-SV_{aa,a} + 2V_{ab,a} + V_{aa,b})}{1+S}. \tag{8.63}$$

We can see that when the change in *total* electronic energy $\Delta E_{\mathrm{el}} = \Delta T + \Delta V$ is calculated, some kinetic energy terms will cancel the corresponding potential energy terms, and *potential energy will dominate during bond formation*:

$$\Delta E_{\mathrm{el}} = \frac{V_{ab,a} + V_{aa,b}}{1+S}. \tag{8.64}$$

To obtain the change, ΔE, in the total energy of the system during bond formation, we have to add the term $1/R$ describing the nuclear repulsion

$$\Delta E = \frac{V_{ab,a}}{1+S} + \frac{V_{aa,b}}{1+S} + \frac{1}{R}. \tag{8.65}$$

This formula is identical (because $V_{ab,a} = V_{ab,b}$) to the difference in orbital energies in the molecule H_2^+ and in the hydrogen atom, as given in Appendix R on p. 1009.

---

[94] For example, $T_{ab} = (a| -\frac{1}{2}\Delta|b) = (a| -\frac{1}{2}\Delta - \frac{1}{r_b} + \frac{1}{r_b}|b) = E_H S + (a|\frac{1}{r_b}|b) = E_H S - V_{ab,b}$.

The formula can be easily interpreted. Let us first consider the electron density described by the φ orbital: $\varphi^2 = [2(1+S)]^{-1}(a^2+b^2+2ab)$. Let us note that the density can be divided into the part close to nucleus a, that close to nucleus b, and that concentrated in the bonding region[95]

$$\varphi^2 = \rho_a + \rho_b + \rho_{ab}, \tag{8.66}$$

where $\rho_a = [2(1+S)]^{-1}a^2$, $\rho_b = [2(1+S)]^{-1}b^2$, $\rho_{ab} = [(1+S)]^{-1}ab$. It can be seen,[96] that the charge associated with ρ_a is $[-2(1+S)]^{-1}$, the charge connected with the nucleus b is the same, and the overlap charge ρ_{ab} is $-S/(1+S)$. Their sum gives $-2/[2(1+S)] - 2S/[2(1+S)] = -1$ (the unit electronic charge). The formula for ΔE may also be written as (we use symmetry: the nuclei are identical, and the a and b orbitals differ only in their centres):

$$\Delta E = \frac{V_{ab,a}}{[2(1+S)]} + \frac{V_{ab,b}}{[2(1+S)]} + \frac{V_{aa,b}}{[2(1+S)]} + \frac{V_{bb,a}}{[2(1+S)]} + \frac{1}{R}. \tag{8.67}$$

Now it is clear that this formula exactly describes the Coulombic interaction (Fig. 8.18.a,b):

- of the electron cloud from the a atom (with density $\frac{1}{2}\rho_{ab}$) with the b nucleus, and vice versa (the first two terms of the expression),
- of the electron cloud of density ρ_a with the b nucleus (third term),
- of the electron cloud of density ρ_b with the a nucleus (fourth term),
- of the a and b nuclei (fifth term).

If we consider *classically* a proton approaching a hydrogen atom, the only terms for the total interaction energy are (Fig. 8.18.c):

$$\Delta E_{\text{class}} = V_{aa,b} + \frac{1}{R}. \tag{8.68}$$

The difference between ΔE and ΔE_{class} only originates from the difference in electron density, calculated quantum mechanically and classically, cf. Fig. 8.18.b,c. The ΔE_{class} is a *weak* interaction (especially for long distances), and tends to $+\infty$ for small R, because[97] of the $1/R$ term. This can be understood because ΔE_{class} is the difference between two Coulombic interactions: of a point charge with a spherical charge cloud, and of the respective two point charges (called *penetration energy*). ΔE contains two more terms in comparison with ΔE_{class}: $V_{ab,a}/[2(1+S)]$ and $V_{ab,b}/[2(1+S)]$, and both decrease exponentially to $V_{aa,a} = -1$ a.u., when R decreases to zero. Thus these terms are not important for long distances, stabilize the molecule for intermediate distances (and provide the main contribution to the chemical bond energy), and are dominated by the $1/R$ repulsion for small distances.

penetration energy

---

[95] Function $a(1)b(1)$ has the highest value in the middle of the bond.

[96] After integrating of ρ_a.

[97] $V_{aa,a}$ is finite.

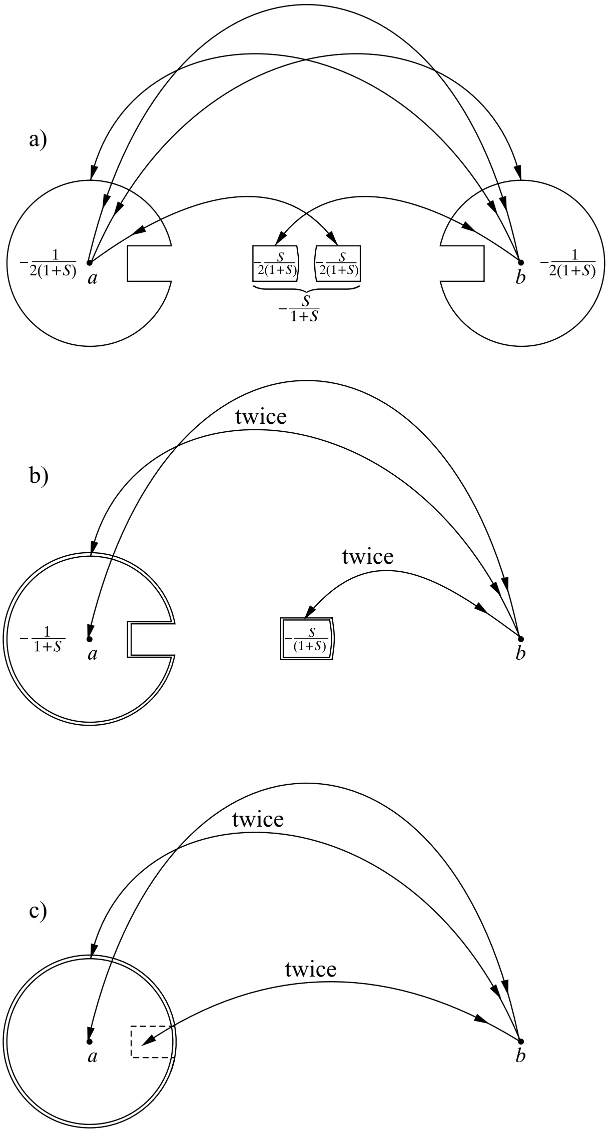


Fig. 8.18. The nature of the chemical bond in the H_2^+ molecule (schematic interpretation):
— (a) *The quantum picture of the interaction*. The total electron density $\varphi^2 = \rho_a + \rho_b + \rho_{ab}$, consists of three electronic clouds $\rho_a = [2(1+S)]^{-1}a^2$ bearing the $-\frac{1}{2(1+S)}$ charge concentrated close to the a nucleus, a similar cloud $\rho_b = [2(1+S)]^{-1}b^2$ concentrated close to the b nucleus and the rest (the total charge is -1) $\rho_{ab} = [(1+S)]^{-1}ab$ bearing the charge of $-2\frac{S}{2(1+S)}$, concentrated in the middle of the bond. The losses of the charge on the a and b atoms have been shown schematically, since the charge in the middle of the bond originates from these losses. The interactions have been denoted by arrows: there are all kinds of interactions of the fragments of one atom with the fragments of the second one.
— (b) *The quantum picture – summary* (we will need it in just a moment). This scheme is similar to (a), but it has been emphasized that the attraction of ρ_a by nucleus b is the same as the attraction of ρ_b by nucleus a, hence they were both presented as one interaction of nucleus b with charge of $-2\rho_a$ at a (hence the double contour line in the figure). In this way two of the interaction arrows have disappeared as compared to (a).
— (c) *The classical picture of the interaction between the hydrogen atom and a proton*. The proton (nucleus b) interacts with the electron of the a atom, bearing the charge of $-1 = -2\frac{1}{2(1+S)} - 2\frac{S}{2(1+S)}$ and with nucleus a. Such division of the electronic charge indicates that it consists of two fragments ρ_a [as in (b)] and of two $-\frac{S}{2(1+S)}$ charges [i.e. similar to (b), but *centred in another way*]. The only difference as compared to (b) is, that in the classical picture nucleus b interacts with two *quite distant* electronic charges (put in the vicinity of nucleus a), while in the quantum picture [schemes (a) and (b)] the same charges attract themselves at short distance.

In the quantum case, for the electron charge cloud connected with the a nucleus, a^2 is decreased by a charge of $S/(1+S)$, *which shifts to the halfway point towards nucleus b*. In the classical case, there is no charge shift – the whole charge is close to a. In both cases there is the nucleus–nucleus and the nucleus–electron interaction. The first is identical, but the latter is completely different in both cases. Yet even in the latter interaction, there is something in common: the interaction of the nucleus with the major part of the electron cloud, with charge $-[1-S/(1+S)] = -1/(1+S)$. The difference in the cases is the interaction with the remaining part of the electron cloud,[98] the charge $-S/(1+S)$.

In the classical view this cloud is located close to *distant* nucleus a, in the quantum view it is in the *middle of the bond*. The latter is much better for bonding. This interaction, of the (negative) electron cloud ρ_{ab} in the middle of the bond with the positive nuclei, stabilizes the chemical bond.

8.7.2 CAN WE SEE A CHEMICAL BOND?

If a substance forms crystals, it may be subjected to X-ray analysis. Such an analysis is quite exceptional, since it is one of very few techniques (which include neutronography and nuclear magnetic resonance spectroscopy), which can show atomic positions in space. More precisely, the X-ray analysis shows electronic density maps, because the radiation sees electrons, not nuclei. The inverse is true in neutronography. If we have the results of X-ray and neutron scattering, we can subtract the electron density of atoms (positions shown by neutron scattering) from the electron density of the molecular crystal (shown by X-ray scattering). This difference would be a consequence of the chemical bonding (and to a smaller extent of the intermolecular interactions). This method is called X–N or X–Ray minus Neutron Diffraction.[99] Hence differential maps of the crystal are possible, where we can see the shape of the "additional" electron density at the chemical bond, or the shape of the electron deficit (negative density) in places where the interaction is antibonding.[100]

---

[98] This simple interpretation gets more complex when further effects are considered, such as contributions to the energy due to the polarization of the spherically symmetric atomic orbitals or the exponent dependence of the 1*s* orbitals (i.e. the dimensions of these orbitals) on the internuclear distance. When there are several factors at play (some positive, some negative) and when the final result is of the order of a single component, then *we* decide which component carries responsibility for the outcome. The situation is similar to that in Parliament, when two MPs from a small party are blamed for the result of a vote (the party may be called the balancing party) while perhaps 200 others who also voted in a similar manner are left in peace.

[99] There is also a pure X-ray version of this method. It uses the fact that the X-ray reflections obtained at large scattering angles see only the spherically symmetric part of the atomic electron density, similarly to that which we obtain from neutron scattering.

[100] R. Boese, *Chemie in unserer Zeit* 23 (1989) 77; D. Cremer, E. Kraka, *Angew. Chem*. 96 (1984) 612.

From the differential maps we can estimate (by comparison with standard substances):

1) the strength of a chemical bond *via* the value of the positive electron density at the bond,
2) the deviation of the bond electron density (perpendicular intersection) from the cylindrical symmetry, which gives information on the π character of the chemical bond,
3) the shift of the maximum electron density towards one of the atoms which indicates the *polarization* of the bond,
4) the shift of the maximum electron density away from the straight line connecting the two nuclei, which indicates bent (banana-like) bonding.

bond polarization

This opens up new possibilities for comparing theoretical calculations with experimental data.

8.8 EXCITATION ENERGY, IONIZATION POTENTIAL, AND ELECTRON AFFINITY (RHF APPROACH)

8.8.1 APPROXIMATE ENERGIES OF ELECTRONIC STATES

Let us consider (within the RHF scheme) the simplest closed-shell system with both electrons occupying the same orbital φ_1. The Slater determinant, called ψ_G (G from the *ground state*) is built from two spinorbitals $\phi_1 = \varphi_1\alpha$ and $\phi_2 = \varphi_1\beta$. We also have the virtual orbital φ_2, corresponding to orbital energy ε_2, and we may form two other spinorbitals from it. We are now interested in the energies of *all the possible excited states* which can be formed from this pair of orbitals. These states will be represented as Slater determinants, built from φ_1 and φ_2 orbitals with the appropriate electron occupancy. We will also assume that excitations do not deform the φ orbitals (which is, of course, only partially true). Now all possible states may be listed by occupation of the ε_1 and ε_2 orbital levels, see Table 8.2.

Table 8.2. All possible occupations of levels ε_1 and ε_2

level	function	ψ_G	ψ_T	$\psi_{T'}$	ψ_1	ψ_2	ψ_E
ε_2		–	α	β	β	α	$\alpha\beta$
ε_1		$\alpha\beta$	α	β	α	β	–

E is a doubly excited electronic state, T and T' are two of three possible triplet states of the same energy. If we require that any state should be an eigenfunction of the $\hat{S}^2$ operator (it also needs to be an eigenfunction of $\hat{S}_z$, but this condition is fortunately fulfilled by all the functions listed above), it appears that only ψ_1 and

ψ_2 are illegal. However, their combinations:

$$\psi_S = \frac{1}{\sqrt{2}}(\psi_1 - \psi_2) \tag{8.69}$$

$$\psi_{T''} = \frac{1}{\sqrt{2}}(\psi_1 + \psi_2). \tag{8.70}$$

are legal. The first describes the singlet state, and the second the triplet state (the third function missing from the complete triplet set).[101] This may be easily checked by inserting the spinorbitals into the determinants, then expanding the determinants, and separating the spin part. For ψ_S, the spin part is typical for the singlet, $\alpha(1)\beta(2) - \alpha(2)\beta(1)$, for T, T' and T'' the spin parts are, respectively, $\alpha(1)\alpha(2)$, $\beta(1)\beta(2)$ and $\alpha(1)\beta(2)+\alpha(2)\beta(1)$. This is expected for triplet functions with components of total spin equal to $1, -1, 0$, respectively (Appendix Q).

Now let us calculate the mean values of the Hamiltonian using the states mentioned above. Here we will use the Slater–Condon rules (p. 986), which soon[102] produce in the MO representation:

$$E_G = 2h_{11} + \mathcal{J}_{11}, \tag{8.71}$$

$$E_T = h_{11} + h_{22} + \mathcal{J}_{12} - \mathcal{K}_{12}, \tag{8.72}$$

(for all three components of the triplet)

$$E_S = h_{11} + h_{22} + \mathcal{J}_{12} + \mathcal{K}_{12}, \tag{8.73}$$

$$E_E = 2h_{22} + \mathcal{J}_{22}, \tag{8.74}$$

where $h_{ii} = (\varphi_i|\hat{h}|\varphi_i)$, and $\hat{h}$ is a one-electron operator, the same as that appearing in the Slater–Condon rules, and *explicitly* shown on p. 335, $\mathcal{J}_{ij}$ and $\mathcal{K}_{ij}$ are two two-electron integrals (Coulombic and exchange): $\mathcal{J}_{ij} = (ij|ij)$ and $\mathcal{K}_{ij} = (ij|ji)$.

The orbital energies of a molecule (calculated for the state with the doubly occupied φ_1 orbital) are:

$$\varepsilon_i = (\varphi_i|\hat{F}|\varphi_i) = (\varphi_i|\hat{h} + 2\hat{\mathcal{J}} - \hat{\mathcal{K}}|\varphi_i), \tag{8.75}$$

where

$$\hat{\mathcal{J}}(1)\chi(1) = \int \mathrm{d}V_2\, \varphi_1^*(2)\varphi_1(2)\frac{1}{r_{12}}\chi(1), \tag{8.76}$$

---

[101] Let us make a convention, that in the Slater determinant $\frac{1}{\sqrt{2}}\det|\phi_1(1)\phi_2(2)|$, the spinorbitals are organized according to increasing orbital energy. This is important because only then are the signs in formulae (8.69) and (8.70) valid.

[102] For E_G the derivation of the final formula is given on p. 352 (E'_{RHF}). The other derivations are simpler.

$$\hat{\mathcal{K}}(1)\chi(1) = \int dV_2\, \varphi_1^*(2)\chi(2)\frac{1}{r_{12}}\varphi_1(1). \tag{8.77}$$

Thus, we get:

$$\varepsilon_1 = h_{11} + \mathcal{J}_{11}, \tag{8.78}$$

$$\varepsilon_2 = h_{22} + 2\mathcal{J}_{12} - \mathcal{K}_{12}. \tag{8.79}$$

Now, the energies of the electronic states can be expressed in terms of orbital energies:

$$E_G = 2\varepsilon_1 - \mathcal{J}_{11}, \tag{8.80}$$

$$E_T = \varepsilon_1 + \varepsilon_2 - \mathcal{J}_{11} - \mathcal{J}_{12} \tag{8.81}$$

(for the ground singlet state and for the three triplet components of the common energy E_T). The distinguished role of φ_1 (in E_T) may be surprising (since the electrons reside on φ_1 and φ_2), but φ_1 is indeed distinguished, because the ε_i values are derived from the Hartree–Fock problem with the *only* occupied orbital φ_1. So we get:

$$E_S = \varepsilon_1 + \varepsilon_2 - \mathcal{J}_{11} - \mathcal{J}_{12} + 2\mathcal{K}_{12}, \tag{8.82}$$

$$E_E = 2\varepsilon_2 + \mathcal{J}_{22} - 4\mathcal{J}_{12} + 2\mathcal{K}_{12}. \tag{8.83}$$

Now it is time for conclusions.

8.8.2 SINGLET OR TRIPLET EXCITATION?

The Jabłoński diagram plays an important role in molecular spectroscopy (Fig. 8.19). It shows three energy levels: the ground state (G), the first excited singlet state (S), and the metastable in-between state. Later on researchers identified this metastable state as the lowest triplet (T).[103]

Let us compute the energy difference between the singlet and triplet states:

Aleksander Jabłoński (1898–1980), Polish theoretical physicist, professor at the John Casimirus University in Vilnius, then at the Nicolaus Copernicus University in Toruń, studied photoluminescence problems.

$$\boxed{E_T - E_S = -2\mathcal{K}_{12} < 0} \tag{8.84}$$

This equation says that

> a molecule always has lower energy in the excited triplet state than in the excited singlet state (both states resulting from the use of the same orbitals),

---

[103] A. Jabłoński, *Nature* 131 (1933) 839; G.N. Lewis, M. Kasha, *J. Am. Chem. Soc.* 66 (1944) 2100.

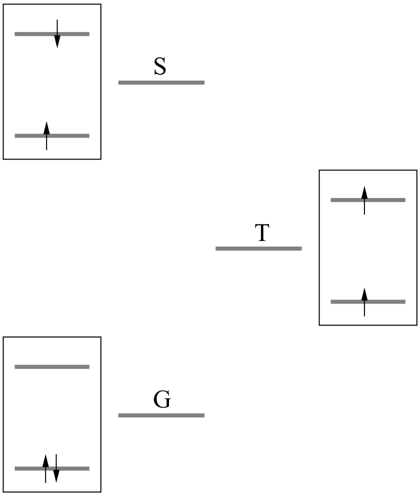


Fig. 8.19. The Jabłoński diagram. The ground state is G. The energy of the singlet excited state (S) is *higher* than the energy of the corresponding triplet state (T; that resulting from use of the same orbitals).

because $\mathcal{K}_{12} = (\varphi_1(1)\varphi_2(2)|\frac{1}{r_{12}}|\varphi_2(1)\varphi_1(2))$ is always positive being the interaction of two identical charge distributions (interpretation of an integral, real functions assumed). This rule holds firmly for the energy of the two lowest (singlet and triplet) states.

8.8.3 HUND'S RULE

The difference between the energies of the ground and triplet states is:

$$\boxed{E_T - E_G = (\varepsilon_2 - \varepsilon_1) - \mathcal{J}_{12}. \qquad (8.85)}$$

This result has a simple interpretation. The excitation of a single electron (to the triplet state) costs some energy ($\varepsilon_2 - \varepsilon_1$), but (since $\mathcal{J}_{12} > 0$) there is also an energy gain ($-\mathcal{J}_{12}$) connected with the removal of the (mutually repulsing) electrons from the "common apartment" (orbital φ_1) to the two separate "apartments" (φ_1 and φ_2). Apartment φ_2 is admittedly on a higher floor ($\varepsilon_2 > \varepsilon_1$), but if $\varepsilon_2 - \varepsilon_1$ is small, then it may still pay to move.

In the limiting case, if $\varepsilon_2 - \varepsilon_1 = 0$, the system prefers to put electrons in separate orbitals and with the same spins (according to the empirical Hund rule, Fig. 8.20).

Friedrich Hermann Hund (1896–1997), professor of theoretical physics at the Universities in Rostock, in Leipzig (1929–1946), Jena, Frankfurt am Main, and finally Göttingen, where in his youth he had worked with Born and Franck. He applied quantum theory to atoms, ions and molecules and discovered his famous empirical rule in 1925 (biography in German: *Intern. J. Quantum Chem.* S11 (1977) 6).

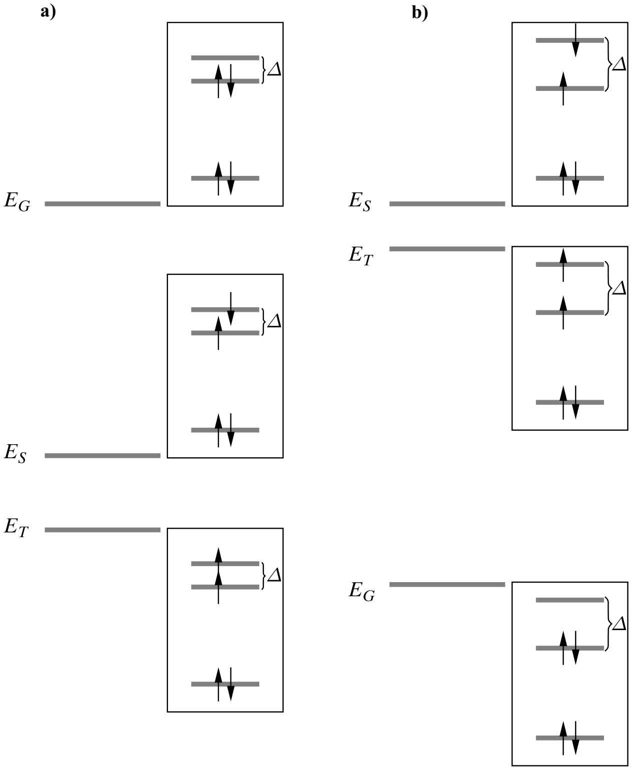


Fig. 8.20. Energy of each configuration (E_G, E_T, E_S; left side of the pictures (a) and (b)) corresponds to an electron occupation of the orbital energy levels (shown in boxes). Two electrons of the HOMO face a dilemma:

— is it better for one of them (fortunately, they are not distinguishable…) to make a sacrifice and move to the upper-floor apartment (then they can avoid each other), Fig. (a);

— or is it better to occupy a common apartment on the lower floor (…but electrons do not like each other), Fig. (b).

If the upper floor is not too high in the energy scale (small Δ, Fig. (a)), then each of the electrons occupies a separate apartment and they feel best having their spins parallel (triplet state). But when the upper floor energy is very high (large Δ, Fig. b), then both electrons are forced to live in the same apartment, and in that case they have to have antiparallel spins (this ensures lower energy).

The Hund's rule pertains to case (a) in its extreme form ($\Delta = 0$). When there are several orbitals of the same energy and there are many possibilities for their occupation, then the state with the lowest energy is such that the electrons each go to a separate orbital, and the alignment of their spins is "parallel" (see p. 32).

8.8.4 IONIZATION POTENTIAL AND ELECTRON AFFINITY (KOOPMANS RULE)

The ionization potential of the molecule M is defined as the minimum energy needed for an electron to detach from the molecule. The electron affinity energy of the molecule M is defined as the minimum energy for an electron detachment from

M^-. Let us assume again naively, that *during these operations the molecular orbitals and the orbital energies do not undergo any changes*. In fact, of course, everything changes, and the computations should be repeated for each system separately (the same applies in the previous section for excitations).

In our two-electron system, which is a model of any closed-shell molecule,[104] the electron removal leaves the molecule with one electron only, and its energy has to be

$$E_+ = h_{11}. \tag{8.86}$$

However,

$$h_{11} = \varepsilon_1 - \mathcal{J}_{11}. \tag{8.87}$$

This formula looks like trouble! After the ionization there is only a single electron in the molecule, while here some electron–electron repulsion (integral $\mathcal{J}$) appears! But everything is fine, because we still use the two-electron problem as a reference, and ε_1 relates to the two-electron problem, in which $\varepsilon_1 = h_{11} + \mathcal{J}_{11}$. Hence,

IONIZATION ENERGY

The ionization energy is equal to the negative of the orbital energy of an electron:

$$E_+ - E_G = -\varepsilon_1. \tag{8.88}$$

To calculate the electron affinity energy we need to consider a determinant as large as 3 × 3, but this proves easy if the useful Slater–Condon rules (Appendix M) are applied. Rule number I gives (we write everything using the spinorbitals, then note that the three spinorbitals are derived from two orbitals, and then sum over the spin variables):

$$E_- = 2h_{11} + h_{22} + \mathcal{J}_{11} + 2\mathcal{J}_{12} - \mathcal{K}_{12}, \tag{8.89}$$

and introducing the orbital energies we get

$$E_- = 2\varepsilon_1 + \varepsilon_2 - \mathcal{J}_{11}, \tag{8.90}$$

which gives

$$E_- - E_G = \varepsilon_2 \tag{8.91}$$

Hence,

ELECTRON AFFINITY

The electron affinity is the difference of the energies of the system without an electron and one representing an anion, $E_G - E_- = -\varepsilon_2$. It is equal approximately to the negative energy of the virtual orbital on which the electron lands.

---

[104] Koopmans theorem applies for this case.

A comment on Koopmans theorem

Tjalling Charles Koopmans (1910–1985), American econometrist of Dutch origin, professor at Yale University (USA), introduced mathematical procedures of linear programming to economics, and received the Nobel Prize for in 1975 "*for work on the theory of optimum allocation of resources*".

The MO approximation is, of course, a rough approximation to reality. So is Koopmans theorem, which proves to be poorly satisfied for most molecules. But these approximations are often used for practical purposes. This is illustrated by a certain quantitative relationship, derived by Grochala et al.[105]

The authors noted, that a very simple relationship holds surprisingly well for the equilibrium *bond lengths* R of four objects: the ground state M_0 of the closed shell molecule, its excited triplet state M_T, its radical–cation $M^{+\cdot}$, and radical–anion $M^{-\cdot}$:

$$R(M_T) = R(M^{-\cdot}) + R(M^{+\cdot}) - R(M_0).$$

The above relationship is similar to that pertaining to the corresponding energies

$$E(M_T) = E(M^{-\cdot}) + E(M^{+\cdot}) - E(M_0),$$

which may be deduced, basing on certain approximations, from Koopmans theorem,[106] or from the Schrödinger equation while neglecting the two-electron operators (i.e. Coulomb and exchange). The difference between these two expressions is, however, fundamental: the latter holds for the four species at *the same* nuclear geometry, while the former describes the geometry *changes* for the "relaxed" species.[107] The first equation proved to be satisfied for a variety of molecules: ethylene, cyclobutadiene, divinylbenzene, diphenylacetylene, *trans*-N_2H_2, CO, CN^-, N_2, and NO^+. It also inspired Andreas Albrecht to derive general inequalities, holding for any one-electron property. The first equation, inspired by Koopmans theorem, was analyzed in detail within density functional theory[108] (described in Chapter 11). It is not yet clear, if it would hold beyond the one-electron approximation, or for experimental bond lengths (these are usually missing, especially for polyatomic molecules).

---

[105] W. Grochala, A.C. Albrecht, R. Hoffmann, *J. Phys. Chem. A* 104 (2000) 2195.

[106] Let us check it using the formulae derived by us: $E(M_T) = \varepsilon_1 + \varepsilon_2 - \mathcal{J}_{11} - \mathcal{J}_{12}$, and $E(M^{-\cdot}) + E(M^{+\cdot}) - E(M_0) = [2\varepsilon_1 + \varepsilon_2 - \mathcal{J}_{11}] + [\varepsilon_1 - \mathcal{J}_{11}] - [2\varepsilon_1 - \mathcal{J}_{11}] = \varepsilon_2 + \varepsilon_1 - \mathcal{J}_{11}$. The equality is obtained after neglecting $\mathcal{J}_{12}$ as compared to $\mathcal{J}_{11}$.

[107] If we assume that a geometry change in these states induces an energy increase proportional to the square of the change, and that the curvature of all these parabolas is identical, then the above relationship would be easily proved. The problem is that these states have significantly different force constants, and the curvature of parabolas strongly varies among them.

[108] P.W. Ayers, R.G. Parr, *J. Phys. Chem. A* 104 (2000) 2211.

8.9 LOCALIZATION OF MOLECULAR ORBITALS WITHIN THE RHF METHOD

The canonical MOs derived from the RHF method are usually delocalized over the whole molecule, i.e. their amplitudes are significant for all atoms in the molecule. This applies, however, mainly to high energy MOs, which exhibit a similar AO amplitude for most atoms. Yet the canonical MOs of the inner shells are usually very well localized. The canonical MOs are occupied, as usual, by putting two electrons on each low lying orbital (the Pauli exclusion principle).

> The picture obtained is in contrast to chemical intuition, which indicates that the electron pairs are localized within the chemical bonds, free electron pairs and inner atomic shells. The picture which agrees with intuition may be obtained after the localization of the MOs.

The localization is based on making new orbitals as linear combinations of the canonical MOs, a fully legal procedure (see p. 338). Then, the determinantal wave function, as shown on p. 338, expressed in the new spinorbitals, takes the form $\psi' = (\det \boldsymbol{A})\psi$. For obvious reasons, the total energy will not change in this case. If linear transformation applied is an orthogonal transformation, i.e. $\boldsymbol{A}^T\boldsymbol{A} = \mathbf{1}$, or a unitary one, $\boldsymbol{A}^\dagger\boldsymbol{A} = \mathbf{1}$, then the new MOs preserve orthonormality (like the canonical ones) as shown on p. 339. We emphasize that we can make *any non-singular*[109] *linear transformation* $\boldsymbol{A}$, not only orthogonal or unitary ones. This means something important, namely

> the solution in the Hartree–Fock method depends on the *space* spanned by the occupied orbitals (i.e. on the set of all linear combinations which can be formed from the occupied MOs), and not on the orbitals only.

> The new orbitals do not satisfy the Fock equations (8.30), these are satisfied by canonical orbitals only.

The localized orbitals (being some other orthonormal basis set in the space spanned by the canonical orbitals) satisfy the Fock equation (8.18) with the off-diagonal Lagrange multipliers.

Can a chemical bond be defined in a polyatomic molecule?

Unfortunately, the view to which chemists get used, i.e. the chemical bonds between pairs of atoms, lone electron pairs, inner shells, can be derived in an infinite

---

[109] For any singular matrix $\det \boldsymbol{A} = 0$, and this should not be allowed (p. 339).

number of ways (because of the arbitrariness of transformation $\boldsymbol{A}$), and in each case the effects of localization ... vary. Hence,

> we cannot uniquely define the chemical bond in a polyatomic molecule.

It is not a tragedy, however, because what really matters is the probability density, i.e. the square of the complex modulus of the *total* many-electron wave function. The concept of the (localized or delocalized) molecular orbitals represents simply an attempt to divide this total density into various spatially separated although overlapping parts, each belonging to a single MO. It is similar to dividing an apple into N parts. The freedom of such a division is unlimited. For example, we could envisage that each part would have the dimension of the apple ("delocalized orbitals"), or an apple would be simply cut axially, horizontally, concentrically etc. into N equal parts, forming an analogue of the localized orbitals. Yet each time the full apple could be reconstructed from these parts.

As we will soon convince ourselves, the problem of defining a chemical bond in a polyatomic molecule is not so hopeless, because various methods lead to essentially the same results.

Now let us consider some practical methods of localization. There are two categories of these: internal and external.[110] In the external localization methods we plan where the future MOs will be localized, and the localization procedure only slightly alters our plans. This is in contrast with the internal methods where certain general conditions are imposed that induce automatically localization of the orbitals.

8.9.1 THE EXTERNAL LOCALIZATION METHODS

Projection method

This is an amazing method,[111] in which *we* first construct some *arbitrary*[112] (but linearly independent[113]) orbitals χ_i of the bonds, lone pairs, and the inner shells, the total number of these being equal to the number of the occupied MOs. Now let us project them on the space of the occupied HF molecular orbitals $\{\varphi_j\}$ using the projection operator $\hat{P}$:

$$\hat{P}\chi_i \equiv \left(\sum_j^{\mathrm{MO}} |\varphi_j\rangle\langle\varphi_j|\right)\chi_i. \tag{8.92}$$

---

[110] Like medicines.

[111] A. Meunier, B. Levy, G. Berthier, *Theoret. Chim. Acta* 29 (1973) 49.

[112] This is the beauty of the projection method.

[113] A linear dependence cannot be allowed. If this happens then we need to change the set of functions χ_i.

Table 8.3. Influence of the initial approximation on the final localized molecular orbitals in the projection method (the LCAO coefficients for the CH_3F molecule)

Function χ for the CF bond				The localized orbital of the CF bond				
$2s$(C)	$2p$(C)	$2s$(F)	$2p$(F)	$2s$(C)	$2p$(C)	$2s$(F)	$2p$(F)	$1s$(H)
0.300	0.536	0.000	−0.615	0.410	0.496	−0.123	−0.654	−0.079
0.285	0.510	0.000	−0.643	0.410	0.496	−0.131	−0.655	−0.079
0.272	0.487	0.000	−0.669	0.410	0.496	−0.138	−0.656	−0.079
0.260	0.464	0.000	−0.692	0.410	0.496	−0.144	−0.656	−0.079
0.237	0.425	0.000	−0.730	0.410	0.496	−0.156	−0.658	−0.079

The projection operator is used to create the new orbitals

$$\varphi_i' = \sum_j^{\text{MO}} \langle \varphi_j | \chi_i \rangle \varphi_j. \tag{8.93}$$

The new orbitals φ_i', as linearly independent combinations of the occupied canonical orbitals φ_j, span the space of the canonical occupied HF orbitals $\{\varphi_j\}$. They are in general non-orthogonal, but we may apply the Löwdin orthogonalization procedure (symmetric orthogonalization, see Appendix J, p. 977).

Do the final localized orbitals depend on the starting χ_i in the projection method? The answer[114] is in Table 8.3. The influence is small.

8.9.2 THE INTERNAL LOCALIZATION METHODS

Ruedenberg method: the maximum interaction energy of the electrons occupying a MO

The basic concept of this method was given by Lennard-Jones and Pople,[115] and applied by Edmiston and Ruedenberg.[116] It may be easily shown that for a given geometry of the molecule the functional $\sum_{i,j=1}^{\text{MO}} \mathcal{J}_{ij}$ is invariant with respect to *any* unitary transformation of the orbitals:

$$\sum_{i,j=1}^{\text{MO}} \mathcal{J}_{ij} = \text{const}. \tag{8.94}$$

The proof is very simple and similar to the one on p. 340, where we derived the invariance of the Coulombic and exchange operators in the Hartree–Fock method. Similarly, we can prove another invariance

$$\sum_{i,j=1}^{\text{MO}} \mathcal{K}_{ij} = \text{const}'. \tag{8.95}$$

---

[114] B. Lévy, P. Millié, J. Ridard, J. Vinh, *J. Electr. Spectr.* 4 (1974) 13.

[115] J.E. Lennard-Jones, J.A. Pople, *Proc. Roy. Soc. (London)* A202 (1950) 166.

[116] C. Edmiston, K. Ruedenberg, *Rev. Modern Phys.* 34 (1962) 457.

This further implies that

maximization of $\sum_{i=1}^{\mathrm{MO}} \mathcal{J}_{ii}$,

which is the very essence of the localization criterion, is equivalent to the *minimization* of the off-diagonal elements

$$\sum_{i<j}^{\mathrm{MO}} \mathcal{J}_{ij}. \tag{8.96}$$

This means that to localize the molecular orbitals we try to make them as small as possible, because then the Coulombic repulsion $\mathcal{J}_{ii}$ will be large.

It may be also expressed in another way, given that

$$\sum_{i,j}^{\mathrm{MO}} \mathcal{K}_{ij} = \mathrm{const}' = \sum_{i}^{\mathrm{MO}} \mathcal{K}_{ii} + 2\sum_{i<j}^{\mathrm{MO}} \mathcal{K}_{ij} = \sum_{i}^{\mathrm{MO}} \mathcal{J}_{ii} + 2\sum_{i<,j}^{\mathrm{MO}} \mathcal{K}_{ij}.$$

Since we maximize the $\sum_{i}^{\mathrm{MO}} \mathcal{J}_{ii}$, then simultaneously

we minimize the sum of the exchange contributions

$$\sum_{i<j}^{\mathrm{MO}} \mathcal{K}_{ij}. \tag{8.97}$$

Boys method: the minimum distance between electrons occupying a MO

In this method[117] we minimize the functional[118]

$$\sum_{i}^{\mathrm{MO}} (ii|r_{12}^2|ii), \tag{8.98}$$

where the symbol $(ii|r_{12}^2|ii)$ denotes an integral similar to $\mathcal{J}_{ii} = (ii|ii)$, but instead of the $1/r_{12}$ operator, we have r_{12}^2. Functional (8.98) is invariant with respect to any unitary transformation of the molecular orbitals.[119] Since the integral $(ii|r_{12}^2|ii)$ represents the definition of the mean square of the distance between two electrons described by $\varphi_i(1)\varphi_i(2)$, the Boys criterion means that we try to obtain the localized orbitals as small as possible (small orbital *dimensions*), i.e. localized in

---

[117] S.F. Boys, in "*Quantum Theory of Atoms, Molecules and the Solid State*", P.O. Löwdin, ed., Academic Press, New York, 1966, p. 253.

[118] Minimization of the interelectronic distance is in fact similar in concept to the maximization of the Coulombic interaction of two electrons in the same orbital.

[119] We need to represent the orbitals as components of a vector, the double sum as two scalar products of such vectors, then transform the orbitals, and show that the matrix transformation in the integrand results in a unit matrix.

some small volume in space. The method is similar to the Ruedenberg criterion of the maximum interelectron repulsion. The detailed technique of localization will be given in a moment. The integrals (8.98) are trivial. Indeed, using Pythagoras' theorem, we get the sum of three simple one-electron integrals of the type:

$$\begin{aligned}&\big(i(1)i(2)\big|(x_2-x_1)^2\big|i(1)i(2)\big)\\&\quad=\big(i(2)\big|x_2^2\big|i(2)\big)+\big(i(1)\big|x_1^2\big|i(1)\big)-2\big(i(1)\big|x_1\big|i(1)\big)\big(i(2)\big|x_2\big|i(2)\big)\\&\quad=2\big(i\big|x^2\big|i\big)-2\big(i\big|x\big|i\big)^2\end{aligned}$$

8.9.3 EXAMPLES OF LOCALIZATION

Despite the freedom of the localization criterion choice, the results are usually similar. The orbitals of the CC and CH bonds in ethane, obtained by various approaches, are shown in Table 8.4.

Let us try to understand Table 8.4. First note the similarity of the results of various localization methods. The methods are different, the starting points are different, and yet we get almost the same in the end. It is both striking and important that

Table 8.4. The LCAO coefficients of the localized orbitals of ethane in the antiperiplanar conformation [P. Millié, B. Lévy, G. Berthier, in: "*Localization and Delocalization in Quantum Chemistry*", ed. O. Chalvet, R. Daudel, S. Diner, J.P. Malrieu, Reidel Publish. Co., Dordrecht (1975)] . Only the nonequivalent atomic orbitals have been shown in the table (four significant digits) for the CC and one of the equivalent CH bonds [with the proton H(1), Fig. 8.21]. The z axis is along the CC′ bond. The localized molecular orbitals corresponding to the carbon inner shells 1s are not listed

	The projection method	Minimum distance method	Maximum repulsion energy
	CC′ bond		
$1s$(C)	−0.0494	−0.1010	−0.0476
$2s$(C)	0.3446	0.3520	0.3505
$2p_z$(C)	0.4797	0.4752	0.4750
$1s$(H)	−0.0759	−0.0727	−0.0735
	CH bond		
$1s$(C)	−0.0513	−0.1024	−0.0485
$2s$(C)	0.3397	0.3373	0.3371
$2p_z$(C)	−0.1676	−0.1714	−0.1709
$2p_x$(C)	0.4715	0.4715	0.4715
$1s$(C′)	0.0073	0.0081	0.0044
$2s$(C′)	−0.0521	−0.0544	−0.054
$2p_z$(C′)	−0.0472	−0.0503	−0.0507
$2p_x$(C′)	−0.0082	−0.0082	−0.0082
$1s$(H1)	0.5383	0.5395	0.5387
$1s$(H2)	−0.0942	−0.0930	−0.0938
$1s$(H3)	−0.0942	−0.0930	−0.0938
$1s$(H4)	0.0580	0.0584	0.0586
$1s$(H5)	−0.0340	−0.0336	−0.0344
$1s$(H6)	−0.0340	−0.0336	−0.0344

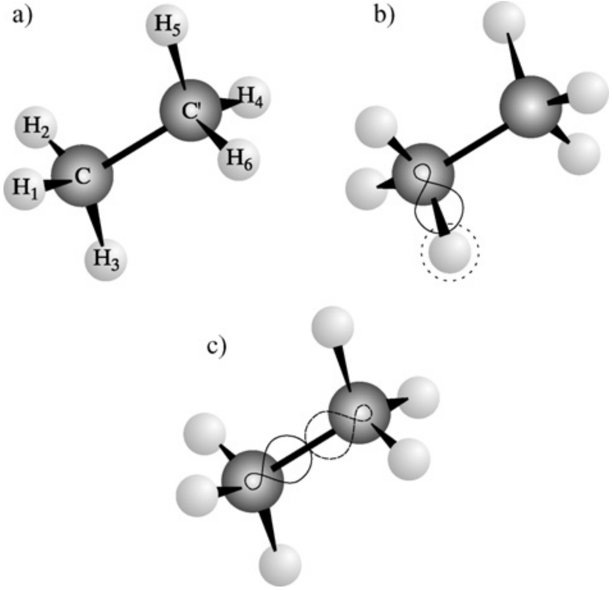


Fig. 8.21. The ethane molecule in the antiperiplanar configuration (a). The localized orbital of the CH bond (b) and the localized orbital of the CC′ bond (c). The carbon atom hybrid forming the CH bond is quite similar to the hybrid forming the CC bond.

the results of various localizations are similar to one another, and in practical terms (not theoretically) we can speak of the unique definition of a chemical bond in a polyatomic molecule.

Nobody would reject the statement that a human body is composed of the head, the hands, the legs, etc. Yet a purist (i.e. theoretician) might get into troubles defining, e.g., a hand (where does it end up?). Therefore, purists would claim that it is impossible to define a hand, and as a consequence there is no such a thing as hand – it simply does not exist. This situation is quite similar to the definition of the chemical bond between two atoms in a polyatomic molecule.

It can be seen that some localized orbitals are concentrated mainly in *one* particular bond between two atoms. For example, in the CC bond orbital, the coefficients at the $1s$ orbitals of the hydrogen atom are small (-0.08). Similarly, the $2s$ and $2p$ orbitals of *one* carbon atom and *one* (the closest) hydrogen atom, dominate the CH bond orbital. Of course, localization is never complete. The oscillating "tails" of the localized orbital may be found even in distant atoms. They assure the mutual orthogonality of the localized orbitals.

8.9.4 COMPUTATIONAL TECHNIQUE

Let us take as an example the maximization of the electron interaction within the same orbital (Ruedenberg method):

$$I = \sum_i^{\text{MO}} \mathcal{J}_{ii} = \sum_i^{\text{MO}} (ii|ii). \tag{8.99}$$

Suppose we want to make an orthogonal transformation (i.e. a rotation in the Hilbert space, Appendix B) of – so far only two – orbitals:[120] $|i\rangle$ and $|j\rangle$, in order to maximize I. The rotation (an orthogonal transformation which preserves the orthonormality of the orbitals) can be written as

$$\begin{aligned}\big|i'(\vartheta)\big\rangle &= |i\rangle\cos\vartheta + |j\rangle\sin\vartheta,\\ \big|j'(\vartheta)\big\rangle &= -|i\rangle\sin\vartheta + |j\rangle\cos\vartheta,\end{aligned}$$

where ϑ is an angle measuring the rotation (we are going to find the optimum angle ϑ). The contribution from the changed orbitals to I, is

$$I(\vartheta) = \big(i'i'\big|i'i'\big) + \big(j'j'\big|j'j'\big). \tag{8.100}$$

Then,[121]

$$\begin{aligned}I(\vartheta) = I(0)\left(1 - \frac{1}{2}\sin^2 2\vartheta\right) &+ \big(2(ii|jj) + (ij|ij)\big)\sin^2 2\vartheta\\ &+ \big((ii|ij) - (jj|ij)\big)\sin 4\vartheta,\end{aligned} \tag{8.101}$$

where $I(0) = (ii|ii) + (jj|jj)$ is the contribution of the orbitals before their rotation.

Requesting that $\frac{\mathrm{d}I(\vartheta)}{\mathrm{d}\vartheta} = 0$, we easily get the condition for optimum $\vartheta = \vartheta_{\mathrm{opt}}$:

$$\begin{aligned}&-2I(0)\sin 2\vartheta_{\mathrm{opt}}\cos 2\vartheta_{\mathrm{opt}} + \big(2(ii|jj) + (ij|ij)\big)4\sin 2\vartheta_{\mathrm{opt}}\cos 2\vartheta_{\mathrm{opt}}\\ &\quad + \big((ii|ij) - (jj|ij)\big)4\cos 4\vartheta_{\mathrm{opt}} = 0,\end{aligned} \tag{8.102}$$

and hence

$$\mathrm{tg}(4\vartheta_{\mathrm{opt}}) = 2\frac{(ij|jj) - (ii|ij)}{2(ii|jj) + (ij|ij) - \frac{1}{2}I(0)}. \tag{8.103}$$

The operation described here needs to be performed for all pairs of orbitals, and then repeated (iterations) until the numerator vanishes for each pair, i.e.

$$(ij|jj) - (ii|ij) = 0. \tag{8.104}$$

The value of the numerator for each pair of orbitals is thus the criterion for whether a rotation is necessary for this pair or not. The matrix of the full orthogonal transformation represents the product of the matrices of these successive rotations.

The same technique of successive 2 × 2 rotations applies to other localization criteria.

---

[120] The procedure is an iterative one. First we rotate one pair of orbitals, then we choose another pair and make another rotation etc., until the next rotations do not introduce anything new.

[121] Derivation of this formula is simple and takes one page.

8.9.5 THE σ, π, δ BONDS

Localization of the MOs leads to the orbitals corresponding to chemical bonds (as well as lone pairs and inner shells). In the case of a bond orbital, a given localized MO is in practice dominated by the AOs of *two atoms only*, those, which create the bond.[122] According to the discussion on p. 371, the larger the overlap integral of the AOs the stronger the bonding. The energy of a molecule is most effectively decreased if the AOs are oriented in such a way as to maximize their overlap integral, Fig. 8.22. We will now analyze the kind and the mutual orientation of these AOs.

As shown in Fig. 8.23, the orbitals σ, π, δ (either canonical or not) have the following features:

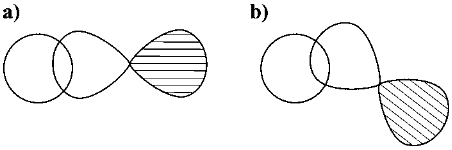


Fig. 8.22. Maximization of the AO overlap requests position (a), while position (b) is less preferred.

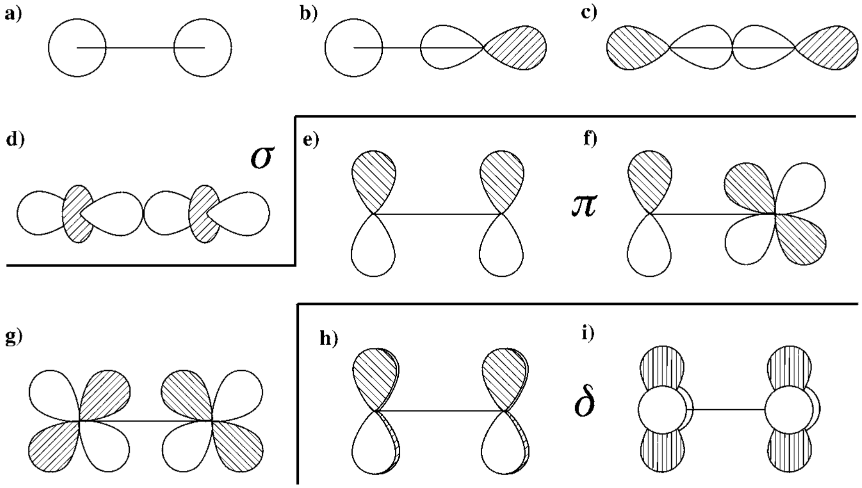


Fig. 8.23. Symmetry of the MOs results from the mutual arrangement of those AOs of both atoms which have the largest LCAO coefficients. Figs. (a–d) show the σ type bonds, (e–g) the π type bonds, and (h,i) the δ type bonds. The σ bond orbitals have no nodal plane (containing the nuclei), the π orbitals have one such plane, the δ ones – two such planes. If the z axis is set as the bond axis, and the x axis is set as the axis perpendicular to the bonding and lying in the plane of the figure, then the cases (b–i) correspond (compare Chapter 4) to the overlap of the following AOs: (b): s with p_z, (c): p_z with p_z, (d): $3\mathrm{d}_{3z^2-r^2}$ with $3\mathrm{d}_{3z^2-r^2}$, (e): p_x with p_x, (f): p_x with $3\mathrm{d}_{xz}$, (g): $3\mathrm{d}_{xz}$-$3\mathrm{d}_{xz}$, (h): $3\mathrm{d}_{xy}$ with $3\mathrm{d}_{xy}$, (i): $3\mathrm{d}_{x^2-y^2}$ with $3\mathrm{d}_{x^2-y^2}$. The figures show such atomic orbitals which correspond to the *bonding* MOs. To get the corresponding *antibonding* MOs, we need to change the sign of *one* of the two AOs.

---

[122] That is, they have the largest absolute values of LCAO coefficients.

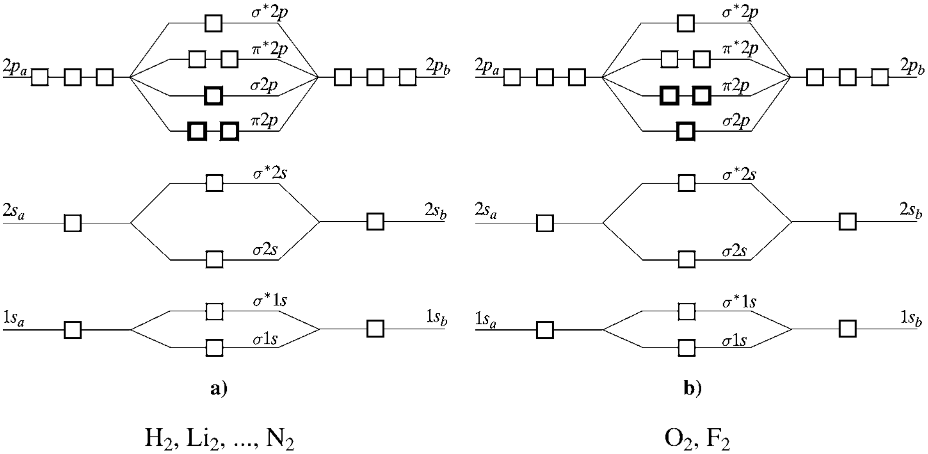


Fig. 8.24. Scheme of the bonding and antibonding MOs in homonuclear diatomics from H_2 through F_2. This scheme is better understood after we recall the rules of effective mixing of AOs, p. 362. All the orbital energies become lower in this series (due to increasing of the nuclear charge), but lowering of the bonding π orbitals leads to changing the *order* of the orbital energies, when going from N_2 to F_2. The sequence of orbital energies (schematically) for the molecules (a) from H_2 through N_2 and (b) for O_2 and F_2.

- the σ-type orbital has *no nodal plane going through the nuclei*,
- the π-type orbital has *one such a nodal plane*,
- the δ-type orbital has *two such nodal planes*.

If a MO is antibonding, then a little star is added to its symbol, e.g., σ^*, π^*, etc. Usually we also give the orbital quantum number (in order of increasing energy), e.g., $1\sigma, 2\sigma, \ldots$ etc. For homonuclear diatomics additional notation is used (Fig. 8.24) showing the main atomic orbitals participating in the MO, e.g., $\sigma 1s = 1s_a + 1s_b$, $\sigma^* 1s = 1s_a - 1s_b$, $\sigma 2s = 2s_a + 2s_b$, $\sigma^* 2s = 2s_a - 2s_b$, etc.

The very fact that the π and δ molecular orbitals have zero value at the positions of the nuclei (the region most important for lowering the potential energy of electrons) suggests that they are bound to be of higher energy than the σ ones, and they are indeed.

8.9.6 ELECTRON PAIR DIMENSIONS AND THE FOUNDATIONS OF CHEMISTRY

What are the dimensions of the electron pairs described by the localized MOs? Well, but how to define such dimensions? All orbitals extend to infinity, so you cannot measure them easily, but some may be more diffuse than others. It also depends on the molecule itself, the role of a given MO in the molecular electronic structure (the bonding orbital, lone electron pair or the inner shell), the influence of neighbouring atoms, etc. These are fascinating problems, and the issue is at the heart of structural studies of chemistry.

Several concepts may be given to calculate the dimensions of the molecular orbitals mentioned above. For example, we may take the integrals $(ii|r_{12}^2|ii) \equiv \langle r^2 \rangle$ calculated within the Boys localization procedure, and use them to measure the square of the dimension of the (normalized) molecular orbital φ_i. Indeed, $\langle r^2 \rangle$ is the mean value of the interelectronic distance for a two-electron state $\varphi_i(1)\varphi_i(2)$, and $\rho_i(\text{Boys}) = \sqrt{\langle r^2 \rangle}$ may be viewed as an estimate of the φ_i orbital dimension. Or, we may do a similar thing by the Ruedenberg method, by noting that the Coulombic integral $\mathcal{J}_{ii}$, calculated in atomic units, is nothing more than the mean value of the *inverse of the distance* between two electrons described by the φ_i orbital. In this case, the dimension of the φ_i orbital may be proposed as $\rho_i(\text{Ruedenberg}) = \frac{1}{\mathcal{J}_{ii}}$. Below, the calculations are reported, in which the concept of $\rho_i(\text{Boys})$ is used. Let us compare the results for CH_3OH and CH_3SH (Fig. 8.25) in order to see, what makes these two molecules so different,[123] Table 8.5.

Interesting features of both molecules can be deduced from Table 8.5. The most fundamental is whether formally the same chemical bonds (say, the CH ones) are indeed similar for both molecules. A purist approach says that each molecule is a New World, and thus these are two different bonds by definition. Yet chemical intuition says that some local interactions (in the vicinity of a given bond) should mainly influence the bonding. If such local interactions are similar, the bonds should turn out similar as well. Of course, the purist approach is formally right, but the New World is quite similar to the Old World, because of local interactions. If chemists desperately clung to purist theory, they would know some 0.01% or so of what they now know about molecules. *It is of fundamental importance for chemistry that we do not study particular cases, case by case, but derive general rules.* Strictly speaking, these rules are false from the very beginning, for they are valid to some extent only, but they enable chemists to understand, to operate, and to be efficient, otherwise there would be no chemistry at all.

The periodicity of chemical elements discovered by Mendeleev is another fundamental idea of chemistry. It has its source in the shell structure of atoms. Following on, we can say that the compounds of sulphur with hydrogen should be

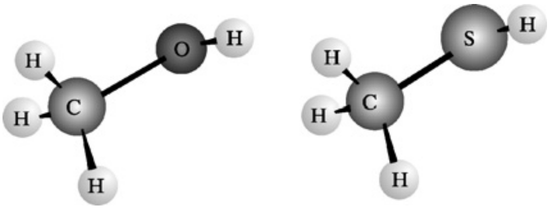


Fig. 8.25. Methanol (CH_3OH) and methanethiol (CH_3SH).

---

[123] Only those who have carried out experiments in person with methanethiol (knowns also as methyl mercaptan), or who have had neighbours (even distant ones) involved in such experiments, understand how important the difference between the OH and SH bonds really is. In view of the theoretical results reported, I am sure they also appreciate the blessing of theoretical work. According to the Guinness book of records, CH_3SH is the most smelly substance in the Universe.

Table 8.5. The dimensions of the electron pairs, i.e. $\sqrt{\langle r^2\rangle}$ (a.u.) for CH_3OH and CH_3SH according to Csizmadia.[a] "Core" means the 1*s* orbital of the atom indicated

CH_3OH		CH_3SH	
core O	0.270	core S (1*s*)	0.148
core C	0.353	core C	0.357
		S (L shell)	0.489
			0.489
			0.483
			0.484
CO	1.499	CS	2.031
CH_1[b]	1.576	CH_1[b]	1.606
$CH_{2,3}$[b]	1.576	$CH_{2,3}$[b]	1.589
OH	1.353	SH	1.847
lone $\mathrm{pair}_{1,2}$[c]	1.293	lone $\mathrm{pair}_{1,2}$[c]	1.886

[a] I.G. Csizmadia, in "*Localization and Delocalization in Quantum Chemistry*", ed. by O. Chalvet and R. Daudel, D. Reidel Publ. Co., Dordrecht, 1975.

[b] Different electron pair dimensions originate from their different positions vs the OH or SH group.

[c] There are two lone pairs in the molecule.

similar to the compounds of oxygen with hydrogen, because sulphur and oxygen have analogous electronic configuration of the valence electrons (i.e. those of the highest energies), and they differ only in the inner shells (O: [He]$2s^2 2p^4$ as compared to S: [Ne]$3s^2 3p^4$).

Take a look of Table 8.5. Note that:

1. The dimension of the electron lone pair localized on the 1*s* orbital of the sulphur atom is twice as small as the dimension of a similar pair of the 1*s* orbital of the oxygen atom. Nothing special. The electrons occupying the 1*s* orbital of S experience a strong electric field of the nucleus charged +16, while the charge of the O nucleus is only +8. Let us note that the core of the carbon atom is even larger, because it is controlled by an even less charged nucleus[124] (+6).
2. The dimension of the electron pair of the 1*s* orbital of the carbon atom (core C) for CH_3OH is very similar to that of the corresponding orbital for CH_3SH (0.353 vs 0.357).[125] This means that the influence of the S atom (as compared to the oxygen atom) on the 1*s* orbital of the *neighbouring atom* is small. The *local character of the interactions* is thus the most decisive.
3. The influence of the S and O atoms on the CH bonds of the methyl group is only slightly larger. For example, in CH_3OH one of the CH bond localized orbital has the dimension of 1.576 a.u., while in CH_3SH 1.606 a.u.
4. The three CH bonds in methanol are very similar to each other (the numbers in Table 8.5 are identical), yet only two of them are strictly equivalent

---

[124] The mean value of the nucleus–electron distance can be easily computed as $\langle 1s|r|1s\rangle = \frac{3}{2}\frac{1}{Z}$, where Z is the charge of the nucleus. The results discussed are consistent with such a simple calculation.

[125] Even these small changes may be detected experimentally by removal of electrons from the molecules by monochromatic X-ray radiation and subsequent measurement of the kinetic energy of the removed electrons. Those which were more strongly bound, move slower.

due to symmetry. It is even more interesting that the CH bonds in CH_3SH are also similar to them, although the differences between the various CH bonds of methanethiol, and between the corresponding CH bonds in methanol and methanethiol, are clearer. So, *even despite the different atomic environment, the chemical bond preserves its principal and individual features*.

5. Let me apologize for a banality: CH_3SH differs from CH_3OH in that the O atom is replaced by the S atom. No wonder then that large differences in the close vicinity of the O and S atoms are easily noticeable. The dimensions of the electron pairs at the S atom (lone pairs and the SH and CS bonds) are always larger than the corresponding pair at the O atom. The differences are at the 30% level. The sulphur atom is simply larger than the oxygen atom, indicating that the electrons are more loosely bound when we go down within a given group of the periodic table.

These conclusions are instructive and strongly encouraging, because *we see a locality in chemistry, and therefore chemistry is easier than it might be* (e.g., CH bonds have similar properties in two different molecules). On the other hand, we may play a subtle game with local differences on purpose, by making suitable chemical substitutions. In this way we have a possibility of *tuning the chemical and physical properties of materials, which is of prime importance in practical applications*.

8.9.7 HYBRIDIZATION

The localized orbitals may serve to present the idea of a *hybrid atomic orbital*. A given localized orbital φ of a bond represents a linear combination of the atomic orbitals of *mainly two atoms* – the partners which form the chemical bond, say a and b. If so, then (for each localized bond orbital) all the atomic orbitals of atom a may be added together with their specific LCAO coefficients,[126] and the same can be done for atom b. These two sums represent two *normalized hybrid atomic orbitals* χ_a and χ_b multiplied by the resulting coefficients c_a and c_b and together form the approximate[127] bond orbital:

$$\varphi \approx c_a h_a + c_b h_b$$

with the corresponding LCAO expansions

$$h_a = \sum_{j \in a} c_{ji} \chi_j,$$

$$h_b = \sum_{j \in b} c_{ji} \chi_j.$$

---

[126] That serve to express the localized orbital through the atomic basis set.

[127] The "tails" of the localized orbital, i.e. its amplitudes on other atoms, have been neglected.

Such a definition of the hybrid orbitals is not unique, since the localized orbitals used are also not unique. However, as shown above, this ambiguity is of secondary importance. The advantage of such an approach to hybridization is that:

- It can be determined for any configuration of the nuclei, e.g., for the tetrahedral as well as for any other configuration of CH_4, etc.
- The definition is applicable at any LCAO basis set used.
- It gives a clear message that *all* the atoms in a molecule are hybridized (why not?), e.g., the carbon atom in the methane molecule as well as all the hydrogen atoms. The only difference between these two hybridizations is that the χ_a for the carbon atom does not resemble any of the χ_j in $\sum_{j\in a} c_{ji}\chi_j$ (because of comparable values[128] of $|c_{ji}|$ meaning an effective mixing of the atomic orbitals), while the χ_b for the hydrogen atom is *dominated by the single atomic orbital* $1s_b$, which may be treated as a lack of hybridization.[129]

sp^3 hybridization (tetrahedral)

How will the hybridization in the optimized geometry of methane look? Well, among five doubly occupied localized molecular orbitals, four[130] protrude from the carbon nucleus towards one of the hydrogens (four hydrogens form a regular tetrahedron) and will have only some marginal amplitudes on the three other hydrogens. If we neglect these "tails" on the other atoms and the contributions of the atomic orbitals other than $2s$ and $2p$ (i.e. their c_{ji}'s) of the carbon atom (also eliminating from the MO the $1s$ orbital of the partner hydrogen atom), we obtain the following normalized hybrid carbon orbitals:

$$h_i = \frac{1}{\sqrt{1+\lambda_i^2}}\left[(2s) + \lambda_i(2p_i)\right]$$

for $i = 1, 2, 3, 4$. If we force the four hybrids to be equivalent, then this means $\lambda_i = \lambda$. Forcing the hybrids to be mutually orthogonal:[131]

$$\langle h_i|h_j\rangle = \frac{1}{1+\lambda^2}\left[1 + \lambda^2\langle 2p_i|2p_j\rangle\right] = \frac{1}{1+\lambda^2}\left[1 + \lambda^2\cos\theta_{ij}\right] = 0,$$

we obtain as the $2s$ and $2p$ mixing ratio

$$\lambda = \sqrt{\frac{-1}{\cos\theta_{ij}}}. \qquad (8.105)$$

---

[128] Mainly of $2s_a$ and $2p_a$, which have the highest values of the LCAO coefficients.

[129] The reason why the carbon atom (and some other atoms such as N, O, etc.) is effectively hybridized, while the hydrogen atom not, is that the $2s$ and $2p$ orbital energy levels in those atoms are close in the energy scale, while the energy difference between the $1s$ hydrogen orbital energy and higher-energy hydrogen orbitals is larger.

[130] The fifth will be composed mainly of the $1s$ carbon orbital.

[131] "Orthogonal" also means "absolutely independent".

Since, for the tetrahedral configuration $\theta_{ij} = 109°28'$, hence, from eq. (8.105): $\cos 109°28' = -\frac{1}{3}$ and $\lambda_i = \sqrt{3}$. Therefore the orthogonal hybrids on the carbon atom (Figs. 8.26.h and 8.27.a) read as:

$$h_i(sp^3) = \frac{1}{2}\big[(2s) + \sqrt{3}(2p_i)\big],$$

where $2s$ and $2p_i$ are the normalized carbon atomic orbitals[132] with i denoting the direction of the hybrid, one of the four directions from the carbon atom towards the tetrahedrally located hydrogen atoms.[133]

sp^2 hybridization (trigonal)

If we tried to find the lowest-energy configuration of ethylene (C_2H_4), it would correspond to a planar structure (Fig. 8.27.b) of D_{2h} symmetry. After analyzing the localized molecular orbitals, it would turn out that three hybrids protrude from each carbon nucleus, their directions lying in the molecular plane (say, xy). These hybrids form angles very close to 120°.

For the trigonal hybridization (i.e. pure sp^2 hybridization, with the $\theta_{ij} = 120°$ angles) we obtain from (8.105) $\lambda = \sqrt{2}$, and, therefore, the three orthogonal normalized sp^2 hybrids are:

$$h_i(sp^2) = \frac{1}{\sqrt{3}}\big[(2s) + \sqrt{2}(2p_i)\big],$$

where the directions $i = 1, 2, 3$ form the Mercedes logo on a plane.

sp hybridization (digonal)

Such hybridization is said to occur in acetylene: HCCH, which after optimization of the Hartree–Fock energy, corresponds to the linear symmetric configuration. According to this explanation, each carbon atom exposes two hybrids (Fig. 8.27c): one towards its carbon and one towards its hydrogen partner. These hybrids use

---

[132] Say, the Slater Type Orbitals (STOs), p. 355.

[133] Such orientation of the (normalized) $2p_i$'s may be achieved by the following choices (just look at the vortices of a cube with the carbon atom at its centre and the four directions forming the tetrahedron):

$$2p_1 = \frac{1}{\sqrt{3}}(2p_x + 2p_y + 2p_z),$$

$$2p_2 = \frac{1}{\sqrt{3}}(2p_x - 2p_y - 2p_z),$$

$$2p_3 = \frac{1}{\sqrt{3}}(-2p_x + 2p_y - 2p_z),$$

$$2p_4 = \frac{1}{\sqrt{3}}(-2p_x - 2p_y + 2p_z).$$

The normalization of the above functions is obvious, since the $2p_x, 2p_y, 2p_z$ are orthogonal.

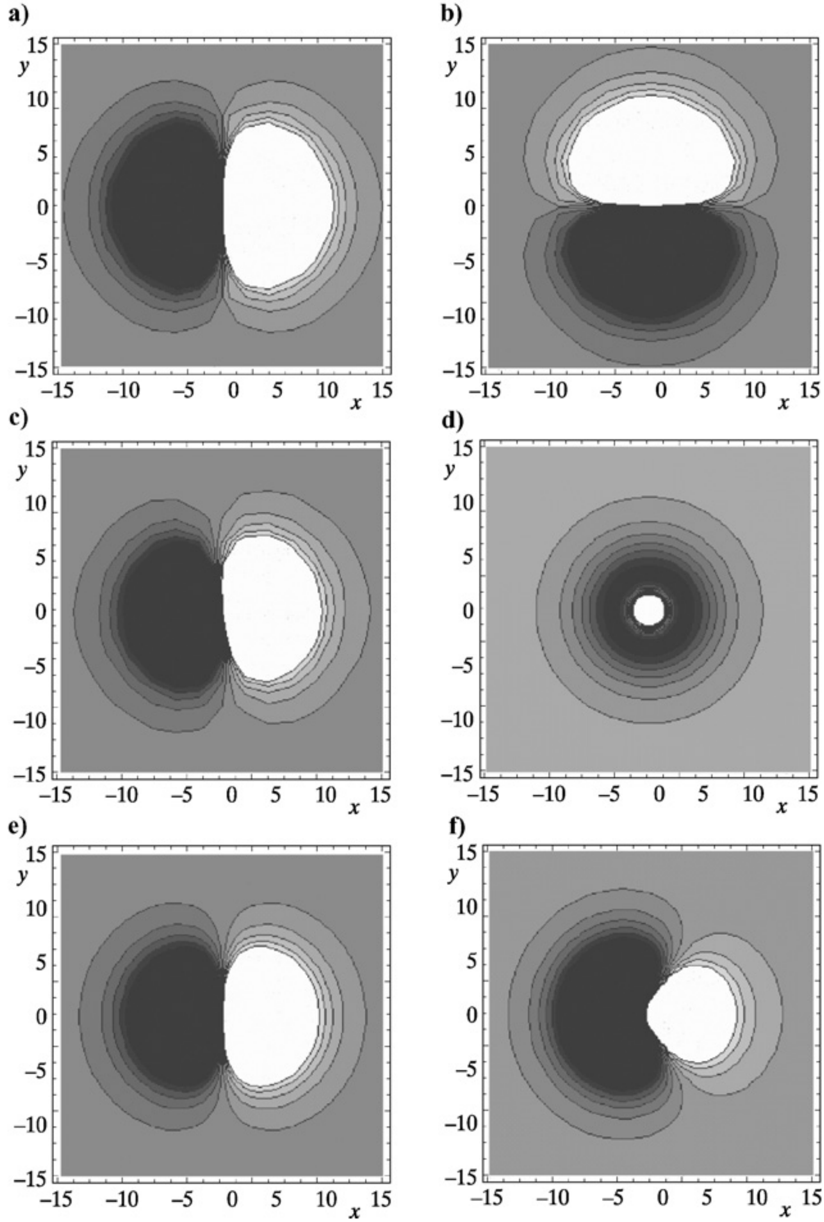


Fig. 8.26. The Slater-type orbitals shown as contours of the section at $z = 0$. The background corresponds to the zero value of the orbital, the darker regions to the negative, the brighter to the positive value of the orbital. (a) $2p_x$ and (b) $2p_y$, and their linear combination (c) equal to $\cos 5° 2p_x + \sin 5° 2p_y$, which is also a $2p$ orbital, but rotated by 5° with respect to the $2p_x$ orbital. In (d) and (e) we show the normalized $2s$ and $2p$ orbitals, which will now be mixed in various proportions: (f) the $1:1$ ratio, i.e. the sp hybridization, (g) the $1:\sqrt{2}$ ratio, i.e. the sp^2 hybridization, and (h) the $1:\sqrt{3}$ ratio, i.e. the sp^3 hybridization.

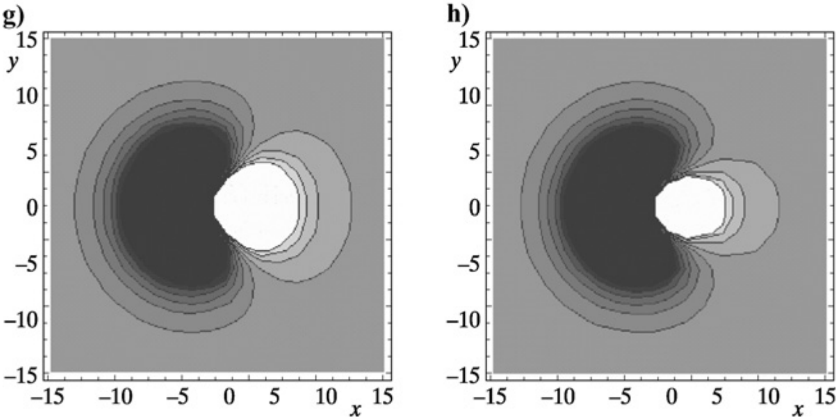


Fig. 8.26. *Continued.*

the two carbon $2s$ and the two carbon $2p_z$, and together with the two $1s$ orbitals of the hydrogens, form the two HC σ bonds and one CC σ bond. This means that each carbon atom has two electrons left, which occupy its $2p_x$ and $2p_y$ orbitals (perpendicular to the molecular axis). The $2p_x$ orbitals of the two carbon atoms form the doubly occupied π_x bonding localized molecular orbital and the same happens to the $2p_y$ orbitals. In this way the carbon atoms form the C≡C triple bond composed of one σ and two π (i.e. π_x and π_y) bonds.

The angle between the two equivalent orthonormal hybrids should be $\theta_{ij} = 180°$, then the mixing ratio will be determined by $\lambda = 1$. Two such hybrids are therefore[134] $h_i(sp) = \frac{1}{\sqrt{2}}[(2s) + (2p_i)]$, and making the two opposite directions explicit: $h_1(sp) = \frac{1}{\sqrt{2}}[(2s) + (2p_z)]$ and $h_2(sp) = \frac{1}{\sqrt{2}}[(2s) - (2p_z)]$.

Is hybridization concept of any value?

The general chemistry textbook descriptions of hybridization for methane, ethylene and acetylene usually start from the electronic configuration of the carbon atom: $1s^2 2s^2 2p^2$. Then it is said that, according to valence bond theory (VB, see Chapter 10), this configuration predicts CH_2 as the carbon hydride (bivalent carbon atom) with the CH bonds forming the right angle.[135] This differs very much from the way the methane molecule looks in reality (regular tetrahedron and tetravalent carbon). If the carbon atom were excited (this might happen at the expense of future energy gains and is known as "promotion") then the configuration might look like $1s^2 2s^1 2p_x^1 2p_y^1 2p_z^1$. The textbooks usually go directly to the mixing of the valence atomic orbitals $2s, 2p_x 2p_y 2p_z$ to form four equivalent sp^3 hybrids,

---

[134] This cannot be exact (cf. the ethylene case), because the two hybrids must not be equivalent. One corresponds to the CC, the other to the CH bond.

[135] Because $2p^2$ means, say, $2p_x^1 2p_y^1$ and these singly occupied atomic orbitals form the two CH bonds with two $1s$ hydrogen orbitals.

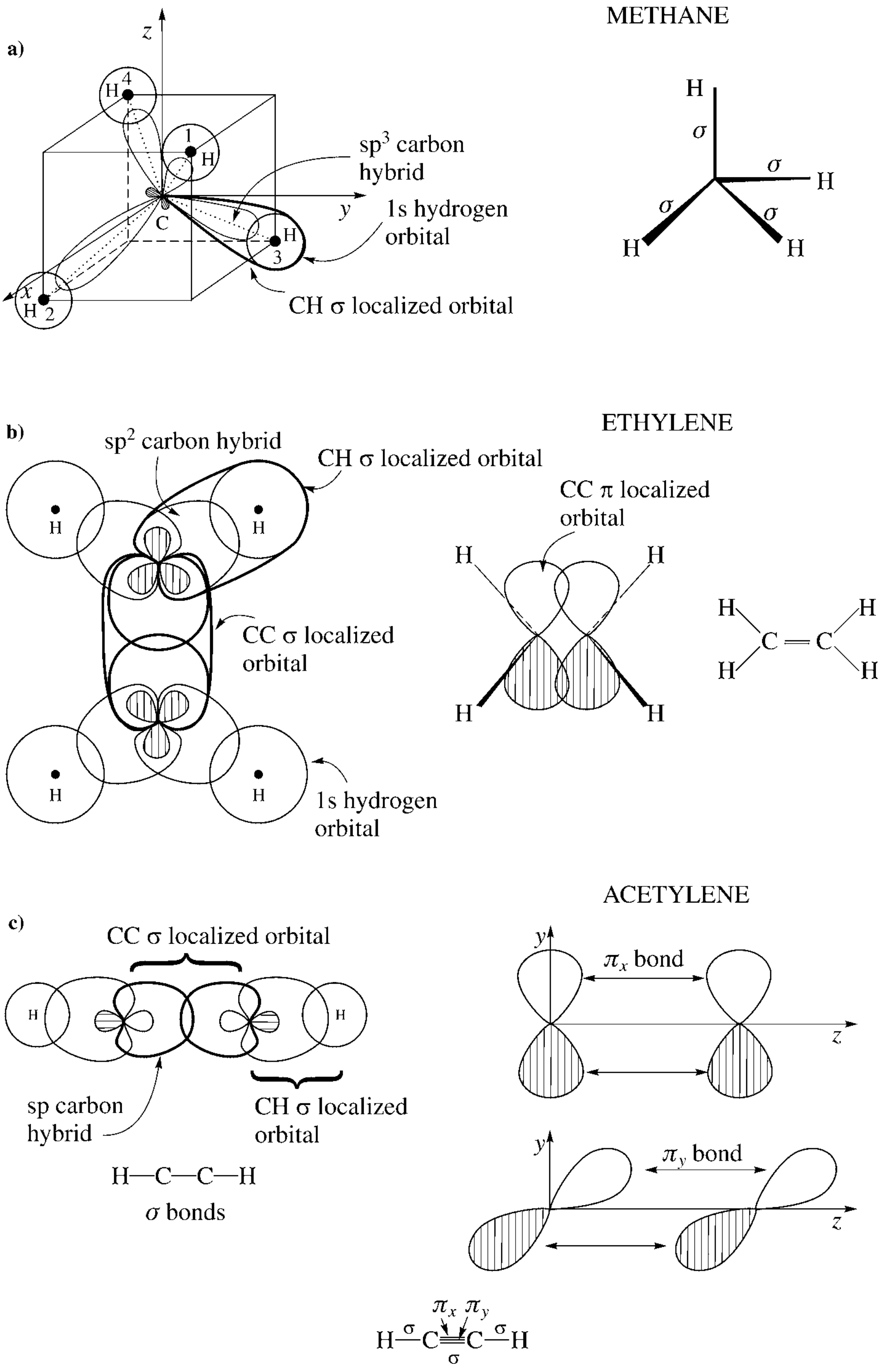

a)
METHANE
z
y
x
C
H 1
H 2
H 3
H 4
sp^3 carbon hybrid
1s hydrogen orbital
CH σ localized orbital
H
σ
b)
ETHYLENE
sp^2 carbon hybrid
CH σ localized orbital
CC π localized orbital
CC σ localized orbital
1s hydrogen orbital
H
H–C=C–H
c)
ACETYLENE
CC σ localized orbital
sp carbon hybrid
CH σ localized orbital
H—C—C—H
σ bonds
π_x bond
π_y bond
y
z
π_x π_y
H$\overset{\sigma}{—}$C≡C$\overset{\sigma}{—}$H
σ

Fig. 8.27. (a) The sp^3 hybridization in the methane molecule in its tetrahedral equilibrium geometry [that corresponds to the minimum of the ground-state electronic energy $E_0^0(\boldsymbol{R})$, see p. 229]. There are four doubly occupied CH σ localized molecular orbitals and one that is essentially the doubly occupied $1s$ carbon atomic orbital. Each of the CH molecular orbital (of the *nearly* cylindrical symmetry) is composed mainly of the carbon hybrid and the hydrogen $1s$ atomic orbital. The figure shows a scheme of the four carbon hybrids called the sp^3 hybrids. (b) An example of the nearly perfect sp^2 hybridization of the carbon atoms in the ethylene (C_2H_4), which is perfectly planar in its ground electronic state (D_{2h} symmetry). Such a hybridization is only approximate, because the CCH angle has to differ from the HCH angle, both slightly deviate from 120°. The localized molecular orbitals are the following (occupied by altogether 16 electrons):
– two essentially $1s$ carbon orbitals,
– four CH orbitals and one CC orbital having the *nearly* cylindrical symmetry (i.e. σ type),
– one bond orbital being antisymmetric with respect to the reflection in the molecular plane (i.e. of the π symmetry).
(c) An example of the sp hybridization: the acetylene molecule. The Hartree–Fock geometry optimization gives the lowest-energy linear configuration: HCCH. The localization gives seven localized doubly occupied molecular orbitals:
— two of them are essentially the $1s$ carbon orbitals,
— two represent the cylindrical CH orbitals (σ),
— one cylindrical CC σ orbital,
— two CC orbitals that are of π symmetry (perpendicular to each other).

which lead directly to the tetrahedral hydride: the methane. Note, however, that being in less of a rush, we would draw the conclusion that the $1s^2 2s^1 2p_x^1 2p_y^1 2p_z^1$ configuration leads to four *non*-equivalent CH bonds in the CH_4 hydride.[136] Only equivalent mixing (hybridization) gives the correct picture. When aiming at ethylene or acetylene the reasoning changes, because some orbitals are left without mixing. We assume sp^2 (one orbital left) or sp (two orbitals left) hybridizations, respectively, which leads to the correct compounds and (almost) correct structures. *It looks as if when we know what we are aiming for, we decide what mixes and what does not.* This seems to be not fair.

Let us check how important the role of hybridization is in the formation of chemical bonds in methane. Let us imagine four scientists performing Hartree–Fock computations for methane in its tetrahedral configuration[137] of nuclei. They use four LCAO basis sets. Professor A believes that in this situation it is important to remember sp^3 hybridization and uses the following basis set (first go the $1s$ orbitals for the hydrogen atoms, then the carbon orbitals):

$$\text{A: } 1s_{H1}, 1s_{H2}, 1s_{H3}, 1s_{H4}, 1s_C, h_1(sp^3), h_2(sp^3), h_3(sp^3), h_4(sp^3).$$

Student B did not read anything about hybridization and just uses the common orbitals:

$$\text{B: } 1s_{H1}, 1s_{H2}, 1s_{H3}, 1s_{H4}, 1s_C, 2s_C, 2p_{x,C}, 2p_{y,C}, 2p_{z,C}.$$

---

[136] Three CH bonds would form right angles (because of $2p_x^1, 2p_y^1, 2p_z^1$), one CH bond however (formed by $2s^1$ together with the corresponding $1s$ hydrogen orbital) would have a quite different character. This contradicts what we get from experiments.

[137] Or any other one.

Students C and D are not the brightest, they have mixed up the hybridization for methane with that for ethylene and acetylene and used the following basis sets:

$$\text{C: } 1s_{H1}, 1s_{H2}, 1s_{H3}, 1s_{H4}, 1s_C, 2p_{x,C}, h_1(sp^2), h_2(sp^2), h_3(sp^2).$$

$$\text{D: } 1s_{H1}, 1s_{H2}, 1s_{H3}, 1s_{H4}, 1s_C, 2p_{x,C}, 2p_{y,C}, h_1(sp), h_2(sp).$$

Who of these scientists will obtain the lowest total energy, i.e. the best approximation to the wave function?

Well, we could perform these calculations, but it is a waste of time. Indeed, each of the scientists used different basis sets, but they all used *the same space* spanned by the AOs. This is because all these hybrids are linear combinations of the orbitals of student B. All the scientists are bound to obtain the same total energy, the same molecular orbitals[138] and the same orbital energies.

Hybridization is useful

Is hybridization a useless concept then? No, it is not. It serves as a first indicator (when calculations are not yet performed) of what happens to a local atomic electronic structure, if the atomic configuration is tetrahedral, trigonal, etc. For example, the trigonal hybrids describe the main features of the electronic configuration in the benzene molecule, Fig. 8.28.

Let us take a slightly more complicated example of what is known as a peptide bond (of great importance in biology), Fig. 8.29.

It is important to remember that we always start from some chemical intuition[139] and use the structural formula given in Fig. 8.29.a. Most often we do not even consider other possibilities (isomers), like those shown in Fig. 8.29.b. Now, we try to imagine what kind of local electronic structure we have around the particular atoms. Let us start from the methyl, i.e. –CH_3 functional groups. Of course, such a group resembles methane, except that one carbon hybrid extends to another atom (not hydrogen). Thus, we expect hybridization over there close to sp^3 one (with all consequences, i.e. angles, etc.). Next, we have the carbon atom that is believed[140] to make the double bond with the oxygen atom. The double bond means an ethylene-like situation, i.e. both atoms should have hybridizations similar to sp^2.

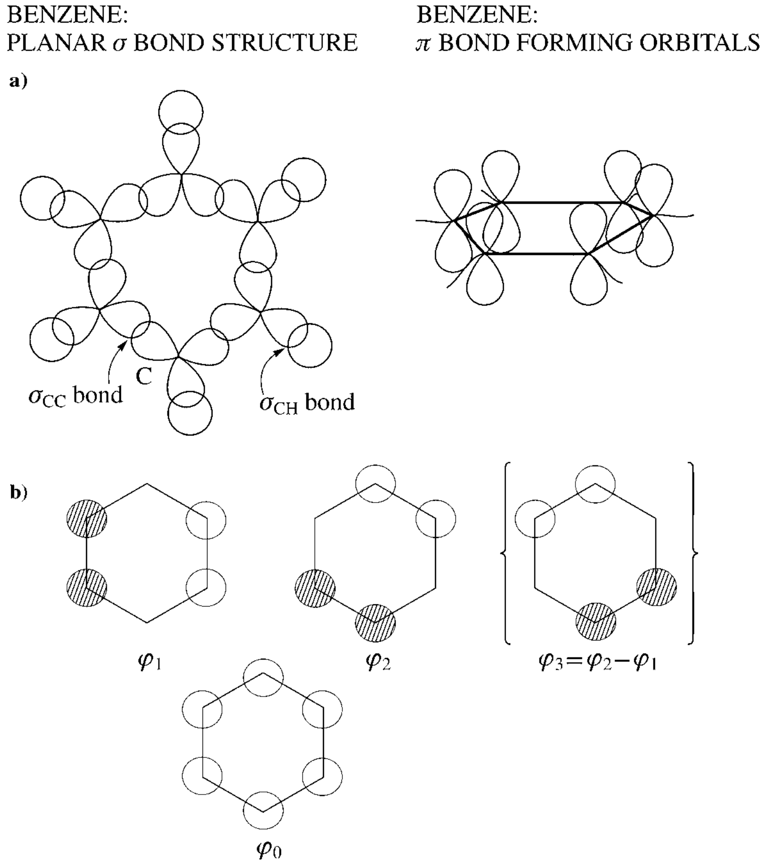
---

[138] Although the LCAO coefficients will be, of course, different, because the expansion functions are different. The orbital plots will be the same.

[139] Based on the vast experience of chemists

[140] Here we rely on the concept of what is known as the *valency* of atoms, i.e. the number of bonds a given atom is able to form with its neighbours. The valency is equal to the number of valence electrons or valence holes, e.g., the valency of the carbon atom is four (because its electron configuration is $K2s^2 2p^2$, four valence electrons), of the oxygen atom is two (because its electron configuration is $K2s^2 2p^4$, two valence holes). An element may have several valencies, because of the opening several electronic shells.

We are making several assumptions based on chemical intuition or knowledge. The reason is that we want to go quickly without performing any computations. *This ambiguity disappears, if we make real computations, e.g., using the Hartree–Fock method. Then the chemical bonds, hybrids etc. are obtained as a result of the computations.*



BENZENE: PLANAR σ BOND STRUCTURE

BENZENE: π BOND FORMING ORBITALS

Fig. 8.28. The benzene molecule. The hybridization concept allows us to link the actual geometry of a molecule with its electronic structure (a). The sp^2 hybrids of the six carbon atoms form the six σ CC bonds and the structure is planar. Each carbon atom thus uses two out of its three sp^2 hybrids, the third one lying in the same plane protrudes towards a hydrogen atom and forms the σ CH bond. In this way, each carbon atom uses its three valence electrons. The fourth one resides on the $2p$ orbital that is perpendicular to the molecular plane. The six $2p$ orbitals form six π molecular orbitals, out of which three are doubly occupied and three are empty. The doubly occupied ones are shown in Fig. (b). The φ_0 of the lowest-energy is an all-in-phase linear combination of the $2p$ atomic orbitals (only their upper lobes are shown). The φ_1 and φ_2 correspond to higher and to the same energy, and have a single node (apart from the node plane of the AOs). The φ_3 orbital that apparently completes all combinations of single-node molecular orbitals is redundant (that is why it is in parentheses), because the orbital represents a linear combination of the φ_1 and φ_2.

Let us begin from the oxygen atom, Fig. 8.29.c. The sp^2 means three hybrids (planar configuration) protruding from the O atom. One of them will certainly bind to a similar one protruding from the carbon atom (OC σ bond), it therefore needs a only single electron from the oxygen. The oxygen atom has 6 valence electrons, therefore there remain five to think of. Four of them will occupy the other two hybrids protruding into space (nothing to bind; they are lone pairs). Hence there

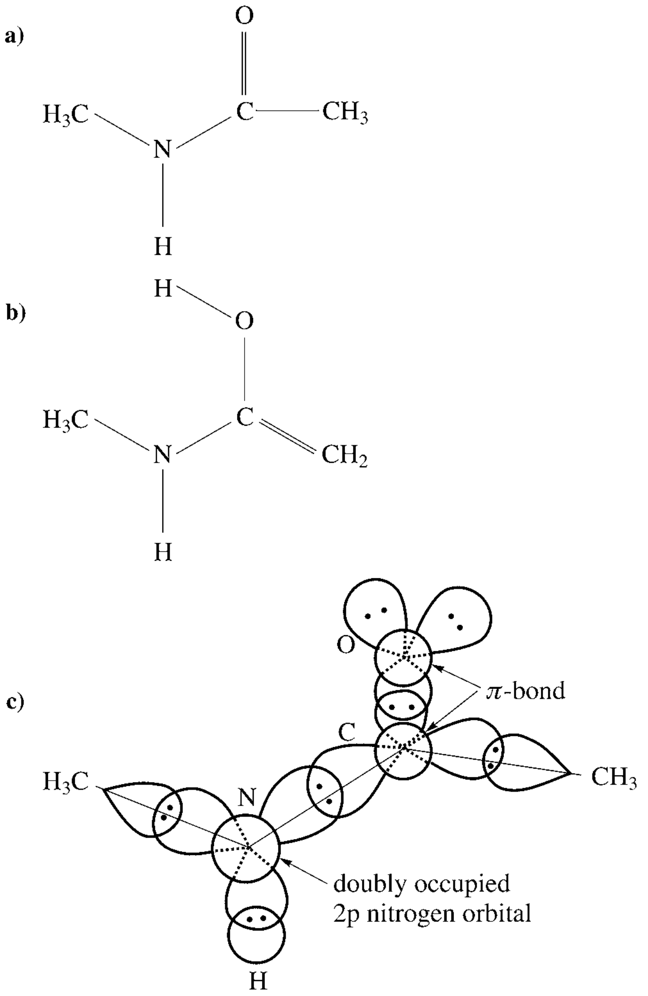


Fig. 8.29. How does the hybridization concept help? The figure shows the all important (proteins) example of the peptide bond. (a) We assume a certain pattern of the chemical bonds (this choice is knowledge based) ignoring other possibilities, such as the isomers shown in (b). Apart from the methyl groups (they have the familiar tetrahedral configuration) the molecule is planar. Usually in chemistry, knowing the geometry, we make a conjecture pertaining to the hybridization of particular atoms. This leads to the electron count for each atom: the electrons left are supposed to participate in bonds with other atoms. In the example shown, the sp^2 hybridization is assumed for the central carbon and for the nitrogen and oxygen atoms (c). A π bonding interaction of the nitrogen, carbon and oxygen should therefore stabilize the planarity of the system, which is indeed an experimental fact.

is one electron left. This is very good, because it will participate in the OC π bond. Let us go to the partner carbon atom. It is supposed to make a double bond with the oxygen. Hence it is reasonable to ascribe to it an ethylene-like hybridization as well. Out of four valence electrons for carbon, two are already used up by the σ and π CO bonds. Two other sp^2 hybrids remain that, of course, accommodate the two electrons and therefore are able to make two σ bonds: one with $-CH_3$ and one with the nitrogen atom. Then we go to the nitrogen atom. It has three substituents in most cases in the (almost) planar configuration (we know this from experiment). To make the analysis simple, we assume an sp^2 ideal hybridization. The nitrogen atom has five valence electrons. Three of them will go to form the σ NC, NH, $N{-}CH_3$ bonds. Note, that although the configuration at N is assumed to be planar, this plane may not coincide with the analogous plane on the carbon atom. Finally, we predict the last two valence electrons of the nitrogen will occupy the $2p$ orbital perpendicular to the plane determined by the substituents of the nitrogen. Note that the $2p$ orbital could overlap (making a bonding effect) with the analogous $2p$ orbital of the carbon atom *provided that the two planes will coincide*. This is why we could expect the planarity of the O–C–N–H, known as peptide bond. This bond plays a prominent role in proteins, because it is responsible for making the chain of amino acid residues. It is an experimental fact that *deviations of the peptide bond from planarity are very small*.

The value of the analyses, as is given above, is limited to qualitative predictions. Of course, computations would give us a much more precise picture of the molecule. In such computations the orbitals would be more precise, or would not be present at all, because, to tell the truth, there is no such thing as orbitals. We badly need to interpret the numbers, to communicate them to others in a understandable way, to say whether we understand these numbers or they are totally unexpected. Reasoning like that given above has a great value as part of our understanding of chemistry, of speaking about chemistry, of predicting and of discussing structures. This is why we need hybridization. Moreover, if our calculations were performed within the VB method (in its simplest formulation; the details of the method will be explained in Chapter 10), then the lowest energy would be obtained by Professor A (who assumed the sp^3 hybridization), because the energy gain over there is very much connected to the overlap of the atomic orbitals forming the basis, and the overlap with the $1s$ hydrogen orbitals is the best for the basis set of Professor A. The other people would get high total energies, because of poor overlap of their atomic orbitals with the $1s$ hydrogen orbitals.

8.10 A MINIMAL MODEL OF A MOLECULE

It is easy to agree that our world is a complex business. It would be great, however, to understand how the world is operating. Answers look more and more complex as we go from crude to more and more accurate theories. Therefore, we would like to consider a simpler world (say, a model of our real world), which

- would work to very good accuracy, i.e. resembled the real world quite well,
- would be based on such simple rules that we could understand it in detail.

We could explain these rules to anybody who were interested. Not only could we predict almost anything, but we ourselves could be confident that we understand *most* of chemistry, because it is based on *several simple rules*. Moreover, why worry about details? Most often we want just to grasp the essence of the problem. On top of that, if this essence were free, only sometimes would we be interested in a more detailed (and expensive) picture.

Is this utopia or can such a model of chemistry be built?

Well, it seems that theoretical chemistry nowadays offers such a model describing chemical structures.

The model is based on the following basic simplifications of the real world:

- The *non-relativistic approach*, i.e. the speed of light is assumed to be infinite, which leads to the Schrödinger equation (Chapter 2).
- The Born–Oppenheimer approximation (Chapter 6) that separates the motion of the nuclei from the motion of the electrons. This approximation allows us to introduce the concept of the *3D structure* of the molecule: the heavy nuclear framework of the molecule kept together by "electronic glue" moves in space (translation), and at the same time rotates in space.
- The mean-field approximation of the present Chapter offers us the *orbital model* of the electronic structure of molecules within the Restricted Hartree–Fock approach. In this picture the electrons are described by the doubly occupied molecular orbitals. Localization of the orbitals gives the doubly occupied *inner shell, lone pair and bond* molecular orbitals. The first and second are sitting on atoms, the latter on chemical bonds. Not all atoms are bound to each other, but instead the *molecule has a pattern of chemical bonds*.
- These bonds are traditionally and formally represented as *single*: e.g., C–C; *double*, e.g., C=C or *triple*, e.g., C≡C, although some intermediate situations usually take place. The total number of these formal bonds of a given atom is equal to its valency. This helps a lot in selecting the chemical bond pattern, which afterwards may be checked against experiment (e.g., bond distances).[141] In most cases a single bond is of the σ type, a double one is composed of one σ and one π, a triple bond means one σ and two π bonds (cf. p. 403).
- The minimal model of a molecule may explain most of the chemical reactions, if besides the closed-shell configuration (double occupancy of the molecular orbitals, including HOMO) we consider excited configurations corresponding to electron transfer(s) from the HOMO to LUMO orbital (see Chapter 14).
- The bonds behave very much like *springs of a certain strength and length*,[142] and therefore, apart from the translational and rotational motion, the atoms vibrate about their equilibrium positions.[143] As to the structural problems (not chemical reactions), these vibrations may be treated as harmonic.
- For the 3D shape of our model molecule, most chemical structures can be correctly predicted using the Hartree–Fock model. The main features of the 3D structure can be also predicted (without any calculation) by using the concept of the *minimum repulsion energy of the electrons pairs*. Within the molecular orbital model, such repulsion is given by eq. (8.96).

---

[141] For some molecules this procedure is not unique, i.e. several chemical bond patterns may be conceived ("resonance structures", cf. the valence bond method in Chapter 10). In such cases the real electronic structure corresponds to an averaging of all of them.

[142] Both depend first of all of the elements making the bond, also a single bond is the weakest and longest, the triple is the strongest and shortest.

[143] The model of molecule visualized in virtually all popular computer programs shows spherical atoms and chemical bonds as shining rods connecting them. First of all, atoms are not spherical, as is revealed by Bader analysis (p. 573) or atomic multipole representations (Appendix S). Second, a chemical bond resembles more a "rope" (higher values) of electronic density than a cylindrical rod. The "rope" is not quite straight and is slimiest at a critical point (see p. 575). Moreover, the rope, when cut perpendicularly, has a circular cross section for pure σ bonds, and an oval cross section for the double bond σ and π (cf. Fig. 11.1).

8.10.1 VALENCE SHELL ELECTRON PAIR REPULSION (VSEPR)

The underlying assumptions of VSEPR[144] are as follows:

- Atoms in a molecule are bound by chemical bonds residing in a space between the bounded atoms. The pattern of such bonds has to be assumed. Each chemical bond represents an electron bonding pair (see the present chapter).
- Some atoms may possess electron pairs that do not participate in a chemical bond pattern (inner shells, lone pairs, see the present chapter).
- The bonding pairs as well as the lone pairs around any atom of the molecule adopt positions in space (on a sphere) such as to minimize pair–pair Coulombic interactions, i.e. they try to be as far away as possible, cf. (8.96).
- The lone pairs repel more than the bonding pairs, and the repulsion bond pair – lone pair is in-between.
- Multiple bonds occupy more space than single bonds.

The total electronic energy in the Restricted Hartree–Fock model is given by eq. (8.36). It is worth stressing, that at a *fixed* geometry of the molecule, the minimization of the electron pair repulsion (by redefinition of the orbitals through a unitary transformation) given by eq. (8.94) does not lead to any change of the total electronic repulsion energy (including self-interaction), which stays invariant. However, when considering variations of geometry (which is at the heart of VSEPR) it is plausible, that smaller electron repulsion (i.e. a smaller const in (8.94)) represents a factor that stabilizes the structure. For small changes of geometry, self-interaction, i.e. $2\sum_i^{\mathrm{MO}} h_{ii} + \sum_i^{\mathrm{MO}} \mathcal{J}_{ii}$ is not supposed to change very much in eq. (8.36), because each term is connected to a particular *localized* orbital, which is not expected to change much when changing the *inter*bond angles. What should change most in (8.36), are the interactions of *different* localized orbitals, because their distances are affected. These interactions are composed of the Coulombic and exchange contributions. The exchange contribution of two different localized orbitals is small, because the orbitals overlap only by their "tails". Hence, *minimization of the interpair Coulombic interactions of* eq. (8.96) *as a function of the geometry of the molecule can be viewed as a rationalization for VSEPR.* Note also, that in each h_{ii} there is an attraction of the electrons occupying the localized orbital i with all the nuclei. This term is responsible for the VSEPR rule that lone pairs repel more strongly than bond pairs.[145]

In the VSEPR method the resulting structure depends on the calculated number of electron pairs around the central atom of the molecule.[146] The resulting geometry is given by Table 8.6

---

[144] R.J. Gillespie, R.S. Nyholm, *Quart. Rev. Chem. Soc.* 11 (1957) 339.

[145] The Coulombic interaction of electron pairs is damped by those nuclei, which are immersed in the electron cloud.

[146] If several atoms may be treated as central, it is necessary to perform the VSEPR procedure for every such atom.

Table 8.6.

Number of electron pairs	Geometry
2	linear
3	trigonal planar
4	tetrahedral
5	trigonal bipyramidal
6	octahedral

Example 4. *Water molecule.* First, some guesses before using the minimal model. The hydrogen atom has a single electron and, therefore valency one, the oxygen atom has valency two (two holes in the valence shell). *We expect*, therefore, that the compounds of the two elements will have the following chemical bond patterns (that saturate their valencies): H–O–H, H–O–O–H, etc. Now our minimal model comes into play. Even quite simple Hartree–Fock calculations show that the system H–O–O–H is less stable than H–O–H + O. Thus, the minimal model predicts, in accordance with what we see in the oceans, that the H_2O compound called water is the most stable.

Now, what can we say about the 3D structure of the water molecule?

Let us take the VSEPR as a first indication. The central atom is oxygen, the number of its valence electrons is six. To this number, we add the number of electrons brought by two hydrogens: $6 + 2 = 8$. Therefore, the number of the electron pairs is $\frac{8}{2} = 4$. According to the above table, oxygen has a tetrahedral arrangement (the angle $109°28'$) of its four electron pairs. Two of them are lone pairs, two are bonding pairs with the hydrogens. Since, as the VSEPR model says, the lone pairs repel more strongly than the bonding pairs, we expect the angle between the lone pairs to be larger than $109°28'$, and the HOH bond angle to be *smaller than* $109°28'$.

Let us see what the minimal model is able to tell us about the geometry of the water molecule. The model (STO 6-31G** basis set, geometry optimization) predicts correctly that there are two *equivalent* OH chemical bonds (and there is no H–H bond[147]) of length $R_{\mathrm{OH}} = 0.943$ Å, whereas experiment gives the result $R_{\mathrm{OH}} = 0.957$ Å. The model predicts, also in accordance with experiment, that the molecule is non-linear (!): the minimum energy HOH angle is 106.0° (the Hartree–Fock limit corresponds to 105.3°), while the experimental HOH angle is 104.5°. *The minimal model is usually able to predict the bond lengths within an accuracy of about* 0.01 Å, *and bond angles to an accuracy of about* 1°.

The minimal model (within the STO 6-31G** basis set) predicts three harmonic vibrational frequencies of the water molecule: antisymmetric stretching 4264 cm^{-1}, symmetric stretching 4147 cm^{-1} and bending 1770 cm^{-1}. It is not easy, though, to predict the corresponding experimental frequencies. We measure the energy differences between consecutive vibrational levels (see Chapter 6, p. 235), which are

---

[147] In agreement with common knowledge in chemistry.

not equal each other (due to anharmonicity). We may, however, deduce these experimental values as they would have been if the bottom of the well were perfectly quadratic (harmonic approximation), they are the following: 3942, 3832, 1648 cm^{-1}, respectively. Similarly to this case, *the minimal model systematically predicts vibrational frequencies that are 7–8% larger than experimental values.* This is not too bad by itself. In practical applications we often take this systematic error into account and correct the calculated frequencies by a scaling factor, thus predicting the frequencies to good accuracy.

Example 5. ***Chlorine trifluoride*** **ClF_3.** It is not easy to tell what kind of structure we will have. Well, it is easy with VSEPR. The central atom will be chlorine. It has 7 valence electrons. Each fluorine contributes one electron. Thus, altogether the chlorine has $7 + 3 = 10$ electrons, i.e. five electron pairs. This means a trigonal bipyramide in VSEPR. However, this does not tell us where the lone pairs and where the fluorine atoms are. Indeed, there are two physically distinct positions in such a bipyramide: the axial and the equatorial, Fig. 8.30.

This corresponds to the interactions of the (lone or bond) electron pairs forming 90°, 120° and 180°. There are $5 \cdot 4/2 = 10$ such interactions. There are three isomers (a,b,c) possible that differ in interaction energy (L-L or lone pair – lone pair, b-L or bond pair – lone pair, b-b or bond–bond), Fig. 8.30.

Isomer	90°	120°	180°
a	2 b-b 4 b-L	1 L-L 2 b-L	1 b-b
b	2 b-b 3 b-L 1 L-L	1 b-b 2 b-L	1 b-L
c	6 b-L	3 b-b	1 L-L

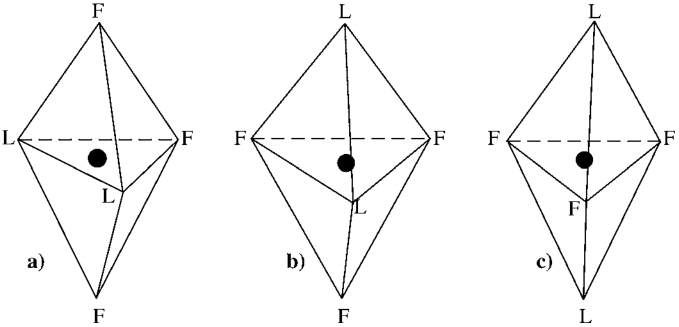


Fig. 8.30. The trigonal bipyramide has two physically distinct positions: three equatorial and two axial. In the ClF_3 we have two lone pairs (L) and three F atoms as candidates for these positions. There are three isomers that differ in energy: (a) having the two lone pairs in equatorial positions – this gives a planar T-shaped molecule (b) having one lone pair equatorial and one axial – this gives a non-planar molecule with two F–Cl–F angles equal to 90°, and one F–Cl–F angle equal to 120° (c) having two lone pairs axial – this gives a planar molecule with F–Cl–F angles equal to 120°. All the isomers have 6 interactions of electron pairs (lone or bond) at 90°, 3 interactions at 120° and one interaction at 180°.

Definitely, the 90° interaction of electron pairs is the most important, because of the shortest L-L distance. In the first approximation, let us look at the 90° interactions only. If we subtract from the energy of each isomer the same number: 3 b-L, then it remains the following

Isomer	90°
a	2 b-b 1 b-L
b	2 b-b 1 L-L
c	3 b-L

According to VSEPR, the L-L repulsion is the strongest, then the b-L follows and the weakest is the b-b repulsion. Now it is clear that the isomer a is of the lowest energy. Therefore, we predict a planar T-like structure with the F_{axial}–Cl–$F_{\text{equatorial}}$ angle equal to 90°. Since the lone pairs take more volume than the bond pairs, the T-shape is a little squeezed. Experiment indeed gives a weird-looking, planar T-shaped molecule, with the F_{axial}–Cl–$F_{\text{equatorial}}$ angle equal to 87.5°.

Summary

- The Hartree–Fock procedure is a *variational* method. The variational function takes the form of *a single Slater determinant* ψ built of orthonormal molecular spinorbitals:

$$\psi = \frac{1}{\sqrt{N!}}\begin{vmatrix} \phi_1(1) & \phi_1(2) & \dots & \phi_1(N) \\ \phi_2(1) & \phi_2(2) & \dots & \phi_2(N) \\ \dots & \dots & \dots & \dots \\ \phi_N(1) & \phi_N(2) & \dots & \phi_N(N) \end{vmatrix}$$

- A molecular spinorbital $\phi_i(1)$ is a one-electron function of the coordinates of electron 1, i.e. of x_1, y_1, z_1, σ_1. In the RHF method, it is the product $\varphi_i(x_1, y_1, z_1)\alpha(\sigma_1)$ or $\varphi_i(x_1, y_1, z_1)\beta(\sigma_1)$ of a real *molecular orbital* $\varphi_i(x_1, y_1, z_1)$ and of the *spin function* $\alpha(\sigma_1)$ or $\beta(\sigma_1)$, respectively. In the general HF method (GHF), a spinorbital is a complex function, which depends both on $\alpha(\sigma_1)$ and $\beta(\sigma_1)$. The UHF method uses, instead, real orbitals, which are all different and are multiplied either by α or β ("different orbitals for different spins").
- Minimization of the mean value of the Hamiltonian, $E = \frac{\langle\psi|\hat{H}\psi\rangle}{\langle\psi|\psi\rangle}$, with respect to the orthonormal spinorbitals ϕ_i (GHF) leads to equations for *optimum spinorbitals* (Fock equations): $\hat{F}(1)\phi_i(1) = \varepsilon_i\phi_i(1)$, where the Fock operator $\hat{F}$ is $\hat{F}(1) = \hat{h}(1) + \hat{J}(1) - \hat{K}(1)$, the Coulombic operator is defined by

$$\hat{J}(1)u(1) = \sum_j \hat{J}_j(1)u(1) \quad \text{and} \quad \hat{J}_j(1)u(1) = \int \mathrm{d}\tau_2 \frac{1}{r_{12}}\phi_j^*(2)\phi_j(2)u(1),$$

and the exchange operator by

$$\hat{K}(1)u(1) = \sum_j \hat{K}_j(1)u(1) \quad \text{and} \quad \hat{K}_j(1)u(1) = \int \mathrm{d}\tau_2 \frac{1}{r_{12}}\phi_j^*(2)u(2)\phi_j(1).$$

- In the Restricted Hartree–Fock method (RHF) for closed shell systems, we assume *double orbital occupancy*, i.e. we form two spinorbitals out of each molecular orbital (by multiplying either by α or β).
- The Fock equations are solved by an iterative approach (with *an arbitrary* starting point) and as a result we obtain approximations to:
  - the total energy,
  - the wave function (the optimum Slater determinant),
  - the canonical molecular orbitals (spinorbitals),
  - the orbital energies.
- Use of the LCAO expansion leads to the Hartree–Fock–Roothaan equations $\boldsymbol{F}\boldsymbol{c} = \boldsymbol{S}\boldsymbol{c}\boldsymbol{\varepsilon}$. Our job is then to find the LCAO coefficients $\boldsymbol{c}$. This is achieved by transforming the matrix equation to the form of the eigenvalue problem, and to diagonalize the corresponding Hermitian matrix. The canonical molecular orbitals obtained are linear combinations of the atomic orbitals. The lowest-energy orbitals are occupied by electrons, those of higher energy are called virtual and are left empty.
- Using the H_2^+ example, we found that a chemical bond results from an electron density flow towards the bond region. This results from a superposition of atomic orbitals due to the variational principle.
- In the simplest MO picture:
  - The *excited triplet state has lower energy than the corresponding excited singlet state*.
  - In case of orbital degeneracy, the system prefers *parallel electron spins* (Hund's rule).
  - The ionization energy is equal to the negative of the *orbital energy of the removed* electron. The electron affinity is equal to the negative of the *orbital energy corresponding to the virtual orbital accommodating the added* electron (Koopmans theorem).
- The canonical MOs for closed-shell systems (the RHF method) may – completely legally – be transformed to orbitals localized in the chemical bonds, lone pairs and inner shells.
- There are many methods of localization. The most important ones are: the projection method, the method of minimum distance between two electrons from the same orbital (Boys approach), and the method of maximum interaction of electrons from the same orbital (Ruedenberg approach).
- Different localization methods lead to sets of localized molecular orbitals which are slightly different but their general shape is very similar.
- The molecular orbitals (localized as well as canonical) can be classified as to the number of nodal surfaces going through the nuclei. A σ bond orbital has no nodal surface at all, a π bond orbital has a single nodal surface, a δ bond orbital has two such surfaces.
- The localization allows comparison of the molecular fragments of different molecules. It appears that the features of the MO localized on the AB bond relatively weakly depend on the molecule in which this bond is found. This is a strong argument and a true source of experimental tactics in chemistry.
- Localization may serve to determine hybrids.
- In everyday practice, chemists use a minimal model of molecules that enables them to compare the geometry and vibrational frequencies with experiment. This model assumes that the speed of light is infinite (non-relativistic effects only), the Born–Oppenheimer approximation is valid (i.e. the molecule has a 3D structure), the nuclei are bound by chemical bonds and vibrate (often harmonic vibrations are assumed), the molecule moves (translation) and rotates as a whole in space.
- In many cases we can successfully predict the 3D structure of a molecule by using a very simple tool: the Valence Shell Electron Pair Repulsion concept.

Main concepts, new terms

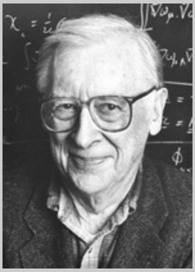
molecular spinorbital (p. 330)
Slater determinant (p. 332)
energy functional (p. 335)
conditional extremum (p. 336)
Lagrange multipliers (p. 336 and p. 997)
variation of spinorbital (p. 336)
Coulombic operator (p. 337)
exchange operator (p. 337)
invariance with respect to a unitary transformation (p. 340)
General Hartree–Fock method (GHF) (p. 341)
Unrestricted Hartree–Fock method (UHF) (p. 342)
Restricted Hartree–Fock method (RHF) (p. 342)
molecular orbital (p. 342)
occupied orbital (p. 343)
virtual orbital (p. 343)
HOMO (p. 343)
LUMO (p. 343)
closed shell (p. 344)
mean field (p. 348)
orbital centring (p. 354)
Slater-type orbital (p. 355)
Slater orbital (p. 356)
Gaussian-type orbital (p. 357)
atomic orbital size (p. 357)
LCAO (p. 360)
atomic basis set (p. 363)
Hartree–Fock–Roothaan method (p. 364)
bonding orbital (p. 371)
antibonding orbital (p. 371)
instability (p. 372)
Fukutome classes (p. 372)
Mendeleev Periodic Table (p. 379)
electronic shells (p. 381)
electronic configuration (p. 381)
chemical bond (p. 383)
penetration energy (p. 386)
Jabłoński diagram (p. 391)
Hund's rule (p. 392)
Koopmans theorem (p. 393)
orbital localization (p. 396)
σ, π, δ – molecular orbitals (p. 403)
electronic pair dimension (p. 404)
hybrids (p. 408)
tetrahedral hybridization (p. 408)
trigonal hybridization (p. 409)
digonal hybridization (p. 409)
minimal model of a molecule (p. 417)
Valence Shell Electron Pair Repulsion (VSEPR) (p. 419)

From the research front

The Hartree–Fock method belongs to a narrow 2–3-member class of standard methods of quantum chemistry. It is the source of basic information about the electronic ground state of a molecule. It also allows for geometry optimization. At present, the available computational codes limit the calculations to the systems built of several hundreds of atoms. Moreover, the programs allow calculations to be made by clicking the mouse. The Hartree–Fock method is always at their core. The GAUSSIAN is one of the best known programs. It is the result of many years of coding by several tens of quantum chemists working under John Pople. Pople was given Nobel Prize in 1998 mainly for this achievement. To get a flavour of the kind of data needed, I provide below a typical data set necessary for GAUSSIAN to perform the Hartree–Fock computations for the water molecule:

```
#HF/STO-3G opt freq pop
water, the STO-3G basis set
0 1
O
H1 1 r12
H2 1 r12 2 a213
r12=0.96
a213=104.5
```

John Pople (1925–2004), British mathematician and one of the founders of the modern quantum chemistry. His childhood was spent in difficult war time in England (every day 25 mile train journeys, sometimes under bombing). He came from a lower middle class family (drapers and farmers), but his parents were ambitious for the future of their children. At the age of twelve John developed an intense interest in mathematics. He entered Cambridge University after receiving a special scholarship. John Pople made important contributions to theoretical chemistry. To cite a few: proposing semiempirical methods – the famous PPP method for π electron systems, the once very popular CNDO approach for all-valence calculations, and finally the monumental joint work on GAUSSIAN – a system of programs that constitutes one of most important computational tools for quantum chemists. John Pople received the Nobel prize in 1998 "*for his development of computational methods in quantum chemistry*" sharing it with Walter Kohn.



The explanatory comments, line by line:

- #HF/STO-3G opt freq pop is a command which informs GAUSSIAN that the computations are of the Hartree–Fock type (HF), that the basis set used is of the STO-3G type (each STO is expanded into three GTOs), that we want to optimize geometry (opt), compute the harmonic vibrational frequencies (freq) and perform the charge population analysis for the atoms (known as Mulliken population analysis, see Appendix S, p. 1015);
- just a comment line;
- 0 1 means that the total charge of the system is equal to 0, and the singlet state is to be computed (1);
- O means that the first atom in the list is oxygen;
- H1 1 r12 means that the second atom in the list is hydrogen named H1, distant from the first atom by r12;
- H2 1 r12 2 a213 means that the third atom in the list is hydrogen named H2, distant from atom number 1 by r12, and forming the 2-1-3 angle equal to a213;
- r12=0.96 is a starting OH bond length in Å;
- a213=104.5 is a starting angle in degrees.

Similar inputs are needed for other molecules. The initial geometry is to some extent arbitrary, and therefore in fact it cannot be considered as real input data. The only true information is the number and charge (kind) of the nuclei, the total molecular charge (i.e. we know how many electrons are in the system), and the multiplicity of the electronic state to be computed. The basis set issue (STO-3G) is purely technical, and gives information about the quality of the results.

Ad futurum...

Along with the development of computational technique, and with progress in the domain of electronic correlation, the importance of the Hartree–Fock method as a source of information about total energy, or total electron density, will most probably decrease. Simply, much larger molecules (beyond the HF level) will be within the reach of future computers. Yet Hartree–Fock calculations will be still carried out, and their results will be carefully analyzed. There are at least two reasons for it:

- Hartree–Fock calculations are most often the necessary step before more precise computations are performed.
- Hartree–Fock computations result in the molecular orbital model: the molecular orbitals and the orbital energies scheme ("minimal model"), and thus they provide the *conceptual framework for the molecule*. It is the sort of model, which may be discussed, thought of, and used to search for explanation of physical and chemical phenomena. So far such a possibility does not exist for advanced methods, where often we obtain very good results, but it is extremely difficult to get an idea *why* they agree so well with experiments.[148]

Additional literature

A. Szabo, N.S. Ostlund, "Modern Quantum Chemistry", McGraw-Hill, New York, 1989, p. 108–231.

Excellent book.

T. Helgaker, P. Jørgensen, J. Olsen, "Molecular Electronic-Structure Theory", Wiley, Chichester, 2000, p. 433–513.

Very good book. It is a contemporary compendium of computational quantum chemistry.

"Localization and Delocalization in Quantum Chemistry", Ed. by O. Chalvet, R. Daudel, S. Diner, J.-P. Malrieu, D. Reidel Publish. Co., Dordrecht, 1975.

A set of the very interesting articles by the leading quantum chemists.

Questions

1. The HF method for the N electron system leads to the wave function:
   a) which depends on all coordinates of N electrons and satisfies the Schrödinger equation;
   b) in a form of one Slater determinant, that depends on all coordinates of N electrons and which is an eigenfunction of the Fock operator $\hat{F}(1)$;
   c) in a form of one Slater determinant, which does not satisfy the Schrödinger equation;
   d) in a form of the product of molecular spinorbitals.

2. The canonical orbitals of a molecule:
   a) represent the minimum basis set of the atoms of a molecule;
   b) do not satisfy the Fock equation and give the lowest-energy Slater determinant;
   c) satisfy the Fock equation and give the lowest-energy Slater determinant;
   d) are the localized orbitals for the chemical bonds of the molecule.

3. The localized orbitals:
   a) are the orthonormal orbitals localized on an atom and satisfying the Fock equation for this atom;
   b) do not satisfy the Fock equation (8.30) and give the lowest-energy Slater determinant;
   c) satisfy the Fock equation (8.30) and give the lowest-energy Slater determinant;
   d) are the atomic orbitals which satisfy the Fock equations for the atoms.

---

[148] The fact of solving the Schrödinger equation, unfortunately, in most cases does not instruct us on the nature of physical phenomena.

4. The orbital energy in a molecule:
   a) is the energy of the electron that occupies the highest-energy atomic orbital;
   b) is the mean energy per one electron in the molecule;
   c) is the sum of the energies of two electrons described by a given molecular orbital;
   d) is the mean value of the Fock operator computed with the corresponding canonical orbital.

5. The Fock operator contains (among other operators) the operator:
   a) of the kinetic energy of all electrons;
   b) of the Coulombic repulsion of all the electrons;
   c) of the kinetic energy of the electrons, and the two-electron Coulombic operator;
   d) of the kinetic energy of an electron and of its electrostatic repulsion with the density distribution of all the electrons.

6. In the LCAO MO method, each MO is:
   a) a function of the electron position in 3D space;
   b) a linear combination of the hybrids generated from the valence atomic orbitals;
   c) a product of the AOs for the atoms of a molecule;
   d) a linear combination of the orbitals of electron lone pairs.

7. In the Hartree–Fock method (closed shell, $U =$ sum of the orbital energies for the doubly occupied orbitals), the total electronic energy of a system is:
   a) $2U - V_{ee}$; b) $2U - 2V_{ee}$; c) $U - V_{ee}$; d) $2U + 2V_{ee}$.

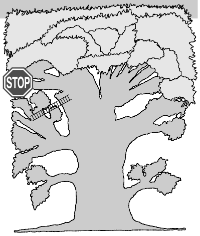
8. Localization of the molecular orbitals (MOs) is performed to:
   a) divide the MOs into bonding and antibonding orbitals;
   b) modify the total electron density, to make it closer to experimental results;
   c) decrease the total energy of a system;
   d) get the MOs of the bonds, electron lone pairs and inner shells.

9. Find the *false* statement:
   a) the UHF method always gives some spin contamination;
   b) $E_{\text{GHF}} \leqslant E_{\text{RHF}}$; c) $E_{\text{GHF}} \leqslant E_{\text{UHF}}$; d) $E_{\text{UHF}} \leqslant E_{\text{RHF}}$.

10. The MO method for the hydrogen molecule (R is the internuclear distance):
    a) cannot be applied for large R values;
    b) properly describes the dissociation of the molecule;
    c) shows large differences for HOMO and LUMO for large R values;
    d) takes the electronic correlation into account.

Answers

1c, 2c, 3b, 4d, 5d, 6a, 7a, 8d, 9a, 10a



WHERE ARE WE?

Chapter 9

ELECTRONIC MOTION IN THE MEAN FIELD: PERIODIC SYSTEMS

Where are we?

We are on the upper left branch of the TREE.

An example

Polyacetylene is an example of new technologically interesting materials[1] and represents a *practically* infinite polymeric chain:[2] …–CH=CH–CH=CH–CH=CH–CH=CH–… There is no such a thing in Nature as an infinite system. Yet, if we examine larger and larger portions of a homogeneous material, we come to the idea that such quantities as energy per stoichiometric unit, electron excitation energy, vibrational frequencies, etc. depend less and less on system size. This means that a boundary-region (polymer ends, crystal surface) contribution to these quantities becomes negligible. Therefore, these (known as *intensive*) quantities attain limit values identical to those for an infinite system. *It pays to investigate the infinite system, because we can use its translational symmetry*.

Herman Staudinger (1881–1965), German polymer chemist, professor at the University of Freiburg, received the Nobel Prize in 1953 "*for his discoveries in the field of macromolecular chemistry*". However strange it may sound now, as late as 1926 the concept of polymers was unthinkable in chemistry. It will be encouraging for PhD students to read that a professor advised Staudinger in the late 1920s: "*Dear colleague, leave the concept of large molecules well alone: organic molecules with a molecular weight above 5000 do not exist. Purify your products, such as rubber, then they will crystallise and prove to be lower molecular substances.*"

---

[1] The discovery of conducting polymers was highlighted by the Nobel Prize 2000 for Hideki Shirakawa (who synthesized a crystalline form of poliacetylene) as well as Allan G. MacDiarmid and Allan J. Heeger, who increased its electric conductivity by 18 orders of magnitude by doping the crystal with some electron acceptors and donors. This incredible increase is probably the largest known to humanity in any domain of experimental sciences (H. Shirakawa, E.J. Louis, A.G. MacDiarmid, C.K. Chiang, A.J. Heeger, *Chem. Soc. Chem. Commun*. 578 (1977)).

[2] That is, a macromolecule. The concept of polymer was introduced to chemistry by Herman Staudinger.

We would like to know whether polyacetylene represents a metal, a semiconductor or an insulator, and how its properties change upon doping. For the time being we have at our disposal the Hartree–Fock method, but it seems to be useless here, because the number of electrons and nuclei is infinite. We may cut out a finite section from the infinite polyacetylene chain and to saturate the resulting dangling bonds by hydrogen atoms, e.g., $CH_2=CH–CH=CH–CH=CH_2$. Calculations for such a molecule of manageable size will not give us the expected answers, because we will first meet bankruptcy. We take pains to compute integrals involving atomic orbitals, but the vast majority of the most essential of them are already in our pocket. It would certainly pay to take into account the translational symmetry of the infinite system.

What is it all about

If a motif (e.g., a cluster of atoms) associated with a unit cell is *regularly translated along three different directions in space*, we obtain an infinite periodic structure (with translational symmetry).

When applying the Hartree–Fock method to such periodic infinite objects one is forced to exploit the translational symmetry of the system, e.g., in calculating integrals. It would indeed be prodigal to compute the integrals many times, the equality of which is *guaranteed* by translational symmetry. When translational symmetry is taken into account, the problem reduces to the calculation of interaction of a single unit cell (reference labelled by 0) with all other unit cells, the nearest neighbour cells being most important. The infinite size of the system is hidden in the plethora of points (to be taken into account) in what is known as the First Brillouin Zone (FBZ). The FBZ represents a unit cell in what is called inverse lattice (associated with a given lattice reflecting the translation symmetry).

The electronic orbital energy becomes a function of the FBZ points and we obtain what is known as band structure of the energy levels. This band structure decides the electronic properties of the system (insulator, semiconductor, metal). We will also show how to carry out the mean field (Hartree–Fock) computations on infinite periodic systems. The calculations require infinite summations (interaction of the reference unit cell with the infinite crystal) to be made. This creates some mathematical problems, which will be also described in the present chapter.

Why is this important?

The present chapter is particularly important for those readers who are interested in solid state physics and chemistry. Others may treat it as exotic and, if they decide they do not like exotic stuff, may go directly to other chapters.

The properties of a polymer or a crystal sometimes differ very widely from those of the atoms or molecules of which they are built. *The same* substance may form *different* periodic structures, which have *different* properties (e.g., graphite and diamond). The properties of periodic structures[3] could be computed by extrapolation of the results obtained for larger and larger clusters of the atoms from which the substance is composed. This avenue is however non-economic. It is easier to carry out quantum mechanical calculations for an infinite system,[4] than for a large cluster.[5]

What is needed?

- Operator algebra (Appendix B, p. 895, necessary).
- Translation operator (Appendix C, p. 903, necessary).
- Hartree–Fock method (Chapter 8, necessary).

---

[3] Also aperiodic but homogeneous.

[4] The surface effects can be neglected and the units the system is composed of, turn out to be equivalent.

[5] Sometimes we may be interested in a particular cluster, not in an infinite system. Then it may turn out that it is more economic to perform the calculations for the infinite system, and use the results in computations for the clusters (e.g., R.A. Wheeler, L. Piela, R. Hoffmann, *J. Am. Chem. Soc.* 110 (1988) 7302).

- Multipole expansion (Appendix X, p. 1038, advised).
- Matrix diagonalization (Appendix K, p. 982, advised).

Classical works

At the age of 23, Felix Bloch published an article "*Über die Quantenmechanik der Elektronen in Kristallgittern*" in *Zeitschrift für Physik*, 52 (1928) 555 (only two years after Schrödinger's historic publication) on the translation symmetry of the wave function. This result is known as the Bloch theorem. This was the first application of LCAO expansion. ★ A book appeared in 1931 by Leon Brillouin entitled *Quantenstatistik* (Springer Verlag, Berlin, 1931), in which the author introduced some of the fundamental notions of band theory. ★ The first *ab initio* calculations for a polymer were carried out by Jean-Marie André in a paper "*Self-Consistent Field Theory for the Electronic Structure of Polymers*" published in the *Journal of the Chemical Physics*, 50 (1969) 1536.

9.1 PRIMITIVE LATTICE

Let us imagine an *infinite crystal*, e.g., a system that exhibits the *translational symmetry* of charge distribution (nuclei and electrons). The translational symmetry will be fully determined by three (linearly independent) basis vectors:[6] $\boldsymbol{a}_1$, $\boldsymbol{a}_2$ and $\boldsymbol{a}_3$ having the property that $\boldsymbol{a}_i$ beginning at any atom, extends to identically located atom of the crystal. The lengths of the basis vectors $\boldsymbol{a}_1$, $\boldsymbol{a}_2$ and $\boldsymbol{a}_3$ are called the *lattice constants* along the three periodicity axes.[7]

basis

lattice constant

There are many such basis sets possible. Any basis vectors choice is acceptable from the point of view of mathematics. For economic reasons we choose one of the possible vector sets that give the *least volume parallelepiped*[8] with sides $\boldsymbol{a}_1$, $\boldsymbol{a}_2$ and $\boldsymbol{a}_3$. This parallelepiped (arbitrarily shifted in space,[9] Fig. 9.1) represents *our* choice of the *unit cell*,[10] which together with its content (motif) is to be translationally repeated.[11]

unit cell

---

[6] Not necessarily perpendicular though; they determine the periodicity axes.

[7] As shown on p. 372, a symmetry of the nuclear framework does not guarantee the same symmetry of the electronic charge distribution computed using a mean field method. We may have cope with the period doubling as compared to the period of the nuclear framework (cf. BOAS, p. 8.5.2). If this happens, then we should choose such lattice constants that ensure the periodicity of both nuclear and electron distributions.

[8] Yes, because multiplicity of $\boldsymbol{a}_i$ would also lead to unit cells that, when repeated, would reproduce the whole crystal. We are, however, interested in the smallest unit cell.

[9] The choice of the origin of the coordinate system is arbitrary, the basis vectors are determined within the accuracy of an arbitrary translation.

[10] An example of a jigsaw puzzle shows that other choices are possible as well. A particular choice may result from its convenience. This freedom will be used on p. 438.

[11] The motif can be ascribed to the unit cell (i.e. chosen) in many different ways provided that after putting the cells together, we get the same original infinite crystal. Let me propose disregarding this problem for the time being (as well as the problem of the choice of the unit cell) and to think of the unit cell as a space-fixed *parallelepiped with the motif that has been enclosed in it*. We will come back to this complex problem at the end of the present chapter.

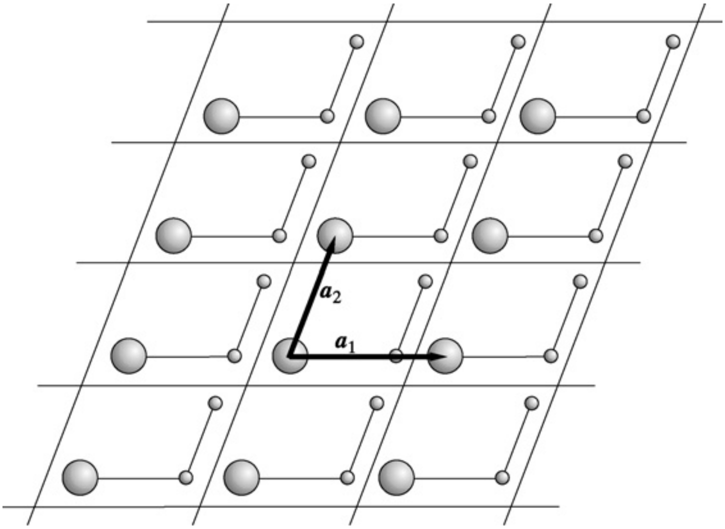


Fig. 9.1. Periodicity in 2D. We choose the unit cell (the parallelogram with vectors $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$) and its content (motif) in such a way as to reproduce the whole infinite crystal by repeating the unit cells through its translation vectors $\boldsymbol{R}_i = n_1\boldsymbol{a}_1 + n_2\boldsymbol{a}_2$ with integers n_1, n_2. In 3D instead of the parallelogram, we would have a parallelepiped, which would be repeated by translation vectors $\boldsymbol{R}_i = n_1\boldsymbol{a}_1 + n_2\boldsymbol{a}_2 + n_3\boldsymbol{a}_3$ with integers n_1, n_2, n_3.

motif

Let us now introduce the *space of translation vectors* $\boldsymbol{R}_i = \sum_{j=1}^{3} n_{ij}\boldsymbol{a}_j$, where n_{ij} are *arbitrary integer numbers* (cf. Appendix B, p. 895).

> The points indicated by all the translation vectors ("lattice vectors") are called the *crystallographic lattice* or the primitive lattice or simply the lattice.

translation operator

Let us introduce the *translation operators* $\hat{T}(\boldsymbol{R}_i)$ defined as translations of a *function*, on which the operator acts, by vector $\boldsymbol{R}_i$ (cf. Chapter 2 and Appendix C on p. 903):

$$\hat{T}(\boldsymbol{R}_i)f(\boldsymbol{r}) = f(\boldsymbol{r} - \boldsymbol{R}_i). \tag{9.1}$$

The function $f(\boldsymbol{r}) \equiv f(\boldsymbol{r} - \boldsymbol{0})$ is centred in the neighbourhood of the origin of the coordinate system, while the function $f(\boldsymbol{r} - \boldsymbol{R}_i)$ is centred on the point shown by vector $\boldsymbol{R}_i$.

The crystal periodicity is reflected by the following property of the potential energy V for an electron (V depends on its position in the crystal):

$$V(\boldsymbol{r}) = V(\boldsymbol{r} - \boldsymbol{R}_i), \tag{9.2}$$

for any $\boldsymbol{R}_i$. The equation simply says that the infinite crystal looks exactly the same close to the origin O, as it does at the point shown by vector $\boldsymbol{R}_i$.

It is easy to see that the operators $\hat{T}(\boldsymbol{R}_i)$ form a *group* (Appendix C, p. 903) with respect to their multiplication as the group operation.[12,13] In Chapter 2 it was shown that the Hamiltonian is invariant with respect to any translation of a molecule. For infinite systems, the proof looks the same for the kinetic energy operator, the invariance of V is guaranteed by eq. (9.2). Therefore, the effective one-electron Hamiltonian commutes with any translation operator:

$$\hat{H}\hat{T}(\boldsymbol{R}_i) = \hat{T}(\boldsymbol{R}_i)\hat{H}.$$

9.2 WAVE VECTOR

If $\hat{T}(\boldsymbol{R}_i)$ commutes with the Hamiltonian, its eigenfunctions also represent the eigenfunctions of the translation operator[14] (cf. Chapter 2, p. 69, also Appendix C on p. 903), i.e. in this case $\hat{H}\psi = E\psi$ and $\hat{T}(\boldsymbol{R}_j)\psi(\boldsymbol{r}) = \psi(\boldsymbol{r} - \boldsymbol{R}_j) = \lambda_{\boldsymbol{R}_j}\psi(\boldsymbol{r})$. The symmetry of V requires the equality of the probability densities

$$\left|\psi(\boldsymbol{r} - \boldsymbol{R}_j)\right|^2 = \left|\psi(\boldsymbol{r})\right|^2, \tag{9.7}$$

for any lattice vector $\boldsymbol{R}_j$, which gives $|\lambda_{\boldsymbol{R}_j}|^2 = 1$, and therefore we may write:

$$\lambda_{\boldsymbol{R}_j} = \exp(-i\theta_{\boldsymbol{R}_j}), \tag{9.8}$$

where $\theta_{\boldsymbol{R}_j}$ will be found in a moment.[15]

---

[12] Indeed, first a product of such operators represents a translational operator:

$$\begin{aligned}\hat{T}(\boldsymbol{R}_1)\hat{T}(\boldsymbol{R}_2)f(\boldsymbol{r}) &= \hat{T}(\boldsymbol{R}_1)f(r - \boldsymbol{R}_2) = f(\boldsymbol{r} - \boldsymbol{R}_1 - \boldsymbol{R}_2) = f\big(\boldsymbol{r} - (\boldsymbol{R}_1 + \boldsymbol{R}_2)\big)\\ &= \hat{T}(\boldsymbol{R}_1 + \boldsymbol{R}_2)f(\boldsymbol{r})\end{aligned}$$

therefore:

$$\hat{T}(\boldsymbol{R}_1)\hat{T}(\boldsymbol{R}_2) = \hat{T}(\boldsymbol{R}_1 + \boldsymbol{R}_2). \tag{9.3}$$

The second requirement is to have a unity operator. This role is played by $\hat{T}(\boldsymbol{0})$, since

$$\hat{T}(\boldsymbol{0})f(\boldsymbol{r}) = f(\boldsymbol{r} + \boldsymbol{0}) = f(\boldsymbol{r}). \tag{9.4}$$

The third condition is the existence [for every $\hat{T}(\boldsymbol{R}_i)$] of the inverse operator, which in our case is $\hat{T}(-\boldsymbol{R}_i)$, because:

$$\hat{T}(\boldsymbol{R}_i)\hat{T}(-\boldsymbol{R}_i) = \hat{T}(\boldsymbol{R}_i - \boldsymbol{R}_i) = \hat{T}(\boldsymbol{0}). \tag{9.5}$$

The group is Abelian (i.e. the operations commute), since:

$$\hat{T}(\boldsymbol{R}_1)\hat{T}(\boldsymbol{R}_2) = \hat{T}(\boldsymbol{R}_1 + \boldsymbol{R}_2) = \hat{T}(\boldsymbol{R}_2 + \boldsymbol{R}_1) = \hat{T}(\boldsymbol{R}_2)\hat{T}(\boldsymbol{R}_1). \tag{9.6}$$

[13] Besides the translational group, the crystal may also exhibit what is called the *point group*, associated with rotations, reflections in planes, inversion, etc., and the *space group* that results from the translational group and the point group. In such cases, a smaller unit cell may be chosen, because the whole crystal is reproduced not only by translations, but also by other symmetry operations. In the present textbook, we will concentrate on the translational symmetry group only.

[14] The irreducible representations of an Abelian group are one-dimensional. In our case (translation group) this means that there is no degeneracy, and that an eigenfunction of the Hamiltonian is also an eigenfunction of all the translation operators.

[15] The exponent sign is arbitrary, we use "−" following a widely used convention.

From equation $\hat{T}(\boldsymbol{R}_j)\psi(\boldsymbol{r}) = \lambda_{\boldsymbol{R}_j}\psi(\boldsymbol{r})$ it follows that

$$\lambda_{\boldsymbol{R}_j}\lambda_{\boldsymbol{R}_l} = \lambda_{\boldsymbol{R}_j+\boldsymbol{R}_l}, \tag{9.9}$$

because

$$\hat{T}(\boldsymbol{R}_j + \boldsymbol{R}_l)\psi(\boldsymbol{r}) = \lambda_{\boldsymbol{R}_j+\boldsymbol{R}_l}\psi(\boldsymbol{r} - \boldsymbol{R}_j - \boldsymbol{R}_l). \tag{9.10}$$

On the other hand

$$\begin{aligned}\hat{T}(\boldsymbol{R}_j + \boldsymbol{R}_l)\psi(\boldsymbol{r}) &= \hat{T}(\boldsymbol{R}_j)\hat{T}(\boldsymbol{R}_l)\psi(\boldsymbol{r}) = \lambda_{\boldsymbol{R}_l}\hat{T}(\boldsymbol{R}_j)\psi(\boldsymbol{r} - \boldsymbol{R}_l)\\ &= \lambda_{\boldsymbol{R}_j}\lambda_{\boldsymbol{R}_l}\psi(\boldsymbol{r} - \boldsymbol{R}_j - \boldsymbol{R}_l).\end{aligned}$$

Since this relation has to be satisfied for any $\boldsymbol{R}_j$ and $\boldsymbol{R}_l$, it is therefore sufficient to have

$$\theta_{\boldsymbol{R}_j} = \boldsymbol{k} \cdot \boldsymbol{R}_j, \tag{9.11}$$

because a multiplication of λ by λ corresponds to adding the exponents, which results in adding vectors $\boldsymbol{R}$, which we need to have. The dot product $\boldsymbol{k} \cdot \boldsymbol{R}_j$ for simplicity will also be written as $\boldsymbol{k}\boldsymbol{R}_j$.

CONCLUSION:

The eigenfunctions of the one-electron Hamiltonian and the translation operators correspond to the following eigenvalues of the translation operator: $\lambda_{\boldsymbol{R}_j} = \exp(-i\boldsymbol{k}\boldsymbol{R}_j)$,

where the wave vector $\boldsymbol{k}$ characterizes the function, not the direction of $\boldsymbol{R}_j$. In other words, any one-electron wave function (crystal orbital), which is the eigenfunction of the one-electron Hamiltonian could be labelled by its corresponding vector $\boldsymbol{k}$, i.e. $\psi(\boldsymbol{r}) \rightarrow \psi_{\boldsymbol{k}}(\boldsymbol{r})$.

Bloch theorem

BLOCH THEOREM

The value of such a function in the point shifted by the vector $\boldsymbol{R}_j$ is equal to:

$$\psi_{\boldsymbol{k}}(\boldsymbol{r} - \boldsymbol{R}_j) = \exp(-i\boldsymbol{k}\boldsymbol{R}_j)\psi_{\boldsymbol{k}}(\boldsymbol{r}) \tag{9.12}$$

The above equality is called the *Bloch theorem*.

Felix Bloch (1905–1983), American physicist of Swiss origin, from 1936–1971 professor at Stanford University. Bloch contributed to the electronic structure of metals, superconductivity, ferromagnetism, quantum electrodynamics and the physics of neutrons. In 1946, independently of E.M. Purcell, he discovered the nuclear magnetic resonance effect. Both scientists received the Nobel Prize in 1952 "*for the development of new methods for nuclear magnetic precision measurements and the discoveries in connection therewith*".

This relation represents a necessary condition to be fulfilled by the eigenfunctions for a perfect periodic structure (crystal, layer, polymer). This equation differs widely from eq. (9.2) for potential energy. Unlike potential energy, which does not change at all upon a lattice translation, the wave function undergoes a change of its phase acquiring the factor $\exp(-i\boldsymbol{k}\boldsymbol{R}_j)$.

Any linear combination of functions labelled by the same $\boldsymbol{k}$ represents an eigenfunction of any lattice translation operator, and corresponds to the same $\boldsymbol{k}$. Indeed, from the linearity of the translation operator

$$\begin{aligned}\hat{T}(\boldsymbol{R}_l)\big(c_1\phi_{\boldsymbol{k}}(\boldsymbol{r})+c_2\psi_{\boldsymbol{k}}(\boldsymbol{r})\big) &= c_1\phi_{\boldsymbol{k}}(\boldsymbol{r}-\boldsymbol{R}_l)+c_2\psi_{\boldsymbol{k}}(\boldsymbol{r}-\boldsymbol{R}_l)\\ &= c_1\exp(-i\boldsymbol{k}\boldsymbol{R}_l)\phi_{\boldsymbol{k}}(\boldsymbol{r})+c_2\exp(-i\boldsymbol{k}\boldsymbol{R}_l)\psi_{\boldsymbol{k}}(\boldsymbol{r})\\ &= \exp(-i\boldsymbol{k}\boldsymbol{R}_l)\big(c_1\phi_{\boldsymbol{k}}(\boldsymbol{r})+c_2\psi_{\boldsymbol{k}}(\boldsymbol{r})\big).\end{aligned}$$

Let us construct the following function (called a *Bloch function*) from a function $\chi(\boldsymbol{r})$, that in future will play the role of an atomic orbital:

$$\phi(\boldsymbol{r})=\sum_j \exp(i\boldsymbol{k}\boldsymbol{R}_j)\chi(\boldsymbol{r}-\boldsymbol{R}_j),$$

where the summation extends over all possible $\boldsymbol{R}_j$, i.e. over the whole crystal lattice. The function ϕ is *automatically* an eigenfunction of any translation operator and may be labelled by the index $\boldsymbol{k}$ similarly $\psi_{\boldsymbol{k}}$. Indeed, first

$$\begin{aligned}\hat{T}(\boldsymbol{R}_l)\phi(\boldsymbol{r}) &= \hat{T}(\boldsymbol{R}_l)\sum_j \exp(i\boldsymbol{k}\boldsymbol{R}_j)\chi(\boldsymbol{r}-\boldsymbol{R}_j)=\sum_j \exp(i\boldsymbol{k}\boldsymbol{R}_j)\hat{T}(\boldsymbol{R}_l)\chi(\boldsymbol{r}-\boldsymbol{R}_j)\\ &= \sum_j \exp(i\boldsymbol{k}\boldsymbol{R}_j)\chi(\boldsymbol{r}-\boldsymbol{R}_j-\boldsymbol{R}_l).\end{aligned}$$

Instead of the summation over $\boldsymbol{R}_j$ let us introduce a summation over $\boldsymbol{R}_{j'}=\boldsymbol{R}_j+\boldsymbol{R}_l$, which means an *identical* summation as before, but we begin to sum the term up from another point of the lattice. Then, we can write

$$\begin{aligned}\sum_{j'} \exp\big(i\boldsymbol{k}(\boldsymbol{R}_{j'}-\boldsymbol{R}_l)\big)\chi(\boldsymbol{r}-\boldsymbol{R}_{j'}) &= \exp(-i\boldsymbol{k}\boldsymbol{R}_l)\sum_{j'} \exp(i\boldsymbol{k}\boldsymbol{R}_{j'})\chi(\boldsymbol{r}-\boldsymbol{R}_{j'})\\ &= \exp(-i\boldsymbol{k}\boldsymbol{R}_l)\phi(\boldsymbol{r})\end{aligned}$$

which had to be proven.

Our function ϕ represents, therefore, an eigenfunction of the translation operator with the same eigenvalue as that corresponding to $\psi_{\boldsymbol{k}}$. In the following, very often $\psi_{\boldsymbol{k}}$ will be constructed as a linear combination of Bloch functions ϕ.

A Bloch function is nothing but a symmetry orbital built from the functions $\chi(\boldsymbol{r}-\boldsymbol{R}_j)$.

A symmetry orbital is a linear combination of atomic orbitals, that transforms according to an irreducible representation Γ of the symmetry group of the Hamiltonian (cf. Appendix C). In order to obtain such a function we may use the corresponding projection operator [see eq. (C.13)].

There is also another way to construct a function $\phi_{\boldsymbol{k}}(\boldsymbol{r})$ of a given $\boldsymbol{k}$ from an auxiliary function $u(\boldsymbol{r})$ satisfying an equation similar to eq. (9.2) for the potential V

$$\hat{T}(\boldsymbol{R}_i)u(\boldsymbol{r}) = u(\boldsymbol{r}-\boldsymbol{R}_i) = u(\boldsymbol{r}). \tag{9.13}$$

Then, $\phi_{\boldsymbol{k}}(\boldsymbol{r}) = \exp(i\boldsymbol{k}\boldsymbol{r})u(\boldsymbol{r})$. Indeed, let us check

$$\begin{aligned}\hat{T}(\boldsymbol{R}_j)\phi_{\boldsymbol{k}}(\boldsymbol{r}) &= \hat{T}(\boldsymbol{R}_j)\exp(i\boldsymbol{k}\boldsymbol{r})u(\boldsymbol{r}) = \exp\bigl(i\boldsymbol{k}(\boldsymbol{r}-\boldsymbol{R}_j)\bigr)u(\boldsymbol{r}-\boldsymbol{R}_j)\\ &= \exp(-i\boldsymbol{k}\boldsymbol{R}_j)\phi_{\boldsymbol{k}}(\boldsymbol{r}),\end{aligned} \tag{9.14}$$

9.3 INVERSE LATTICE

Let us now construct the so called *biorthogonal basis* $\boldsymbol{b}_1, \boldsymbol{b}_2, \boldsymbol{b}_3$ with respect to the basis vectors $\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3$ of the primitive lattice, i.e. the vectors that satisfy the *biorthogonality relations*:

biorthogonality

$$\boldsymbol{b}_i\boldsymbol{a}_j = 2\pi\delta_{ij}. \tag{9.15}$$

The vectors $\boldsymbol{b}_i$ can be expressed by the vectors $\boldsymbol{a}_i$ in the following way

$$\boldsymbol{b}_i = 2\pi\sum_j \boldsymbol{a}_j\bigl(\boldsymbol{S}^{-1}\bigr)_{ji}, \tag{9.16}$$

$$\boldsymbol{S}_{ij} = \boldsymbol{a}_i\cdot\boldsymbol{a}_j. \tag{9.17}$$

The vectors $\boldsymbol{b}_1$, $\boldsymbol{b}_2$ and $\boldsymbol{b}_3$ form the basis of a lattice in a 3D space. This lattice will be called the *inverse lattice*.

The inverse lattice vectors are, therefore,

$$\boldsymbol{K}_j = \sum_{i=1}^{i=3} g_{ji}\boldsymbol{b}_i, \tag{9.18}$$

where g_{ij} represent arbitrary *integers*. We have

$$\boldsymbol{K}_j\boldsymbol{R}_i = 2\pi M_{ij},$$

where M_{ij} are integer numbers.

Indeed,

$$\boldsymbol{K}_j\cdot\boldsymbol{R}_i = \sum_{l=1}^{3} g_{jl}\boldsymbol{b}_l\cdot\sum_{k=1}^{3} n_{ik}\boldsymbol{a}_k = \sum_{l=1}^{3}\sum_{k=1}^{3} n_{ik}g_{jl}\boldsymbol{b}_l\cdot\boldsymbol{a}_k \tag{9.19}$$

$$= \sum_{l=1}^{3}\sum_{k=1}^{3} n_{ik}g_{jl}(2\pi)\delta_{lk} = 2\pi\sum_{l=1}^{3} n_{il}g_{jl} = 2\pi M_{ij} \tag{9.20}$$

with n_{ik}, g_{jl} and, therefore also M_{ij} as *integers*. The inverse lattice is composed, therefore, from the isolated points indicated from the origin by the vectors $\boldsymbol{K}_j$. All the vectors that begin at the origin form the *inverse space*.

Examples

Let us see how we obtain the inverse lattice (1D, 2D, 3D) in practice.

1D

We have only a single biorthogonality relation: $\boldsymbol{b}_1\,\boldsymbol{a}_1 = 2\pi$, i.e. after skipping the index $\boldsymbol{b}\boldsymbol{a} = 2\pi$. Because of the single dimension, we have to have $\boldsymbol{b} = \frac{2\pi}{a}(\frac{\boldsymbol{a}}{a})$, where $|\boldsymbol{a}| \equiv a$. Therefore,

> the vector $\boldsymbol{b}$ has length $\frac{2\pi}{a}$ and the same direction as $\boldsymbol{a}$.

2D

This time we have to satisfy: $\boldsymbol{b}_1\boldsymbol{a}_1 = 2\pi$, $\boldsymbol{b}_2\boldsymbol{a}_2 = 2\pi$, $\boldsymbol{b}_1\boldsymbol{a}_2 = 0$, $\boldsymbol{b}_2\boldsymbol{a}_1 = 0$. This means that the game takes place within the plane determined by the lattice vectors $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$. The vector $\boldsymbol{b}_1$ has to be perpendicular to $\boldsymbol{a}_2$, while $\boldsymbol{b}_2$ has to be perpendicular to $\boldsymbol{a}_1$, their directions such as shown in Fig. 9.2 (each of the $\boldsymbol{b}$ vectors is a linear combination of $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ according to (9.16)).

3D

In the 3D case the biorthogonality relations are equivalent to setting

$$\boldsymbol{b}_1 = \boldsymbol{a}_2 \times \boldsymbol{a}_3 \,\frac{2\pi}{V}, \tag{9.21}$$

$$\boldsymbol{b}_2 = \boldsymbol{a}_3 \times \boldsymbol{a}_1 \,\frac{2\pi}{V}, \tag{9.22}$$

$$\boldsymbol{b}_3 = \boldsymbol{a}_1 \times \boldsymbol{a}_2 \,\frac{2\pi}{V}, \tag{9.23}$$

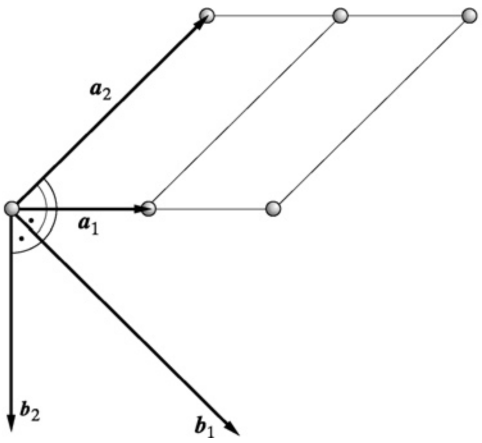


Fig. 9.2. Construction of the inverse lattice in 2D. In order to satisfy the biorthogonality relations (9.15) the vector $\boldsymbol{b}_1$ has to be orthogonal to $\boldsymbol{a}_2$, while $\boldsymbol{b}_2$ must be perpendicular to $\boldsymbol{a}_1$. The lengths of the vectors $\boldsymbol{b}_1$ and $\boldsymbol{b}_2$ also follow from the biorthogonality relations: $\boldsymbol{b}_1 \cdot \boldsymbol{a}_1 = \boldsymbol{b}_2 \cdot \boldsymbol{a}_2 = 2\pi$.

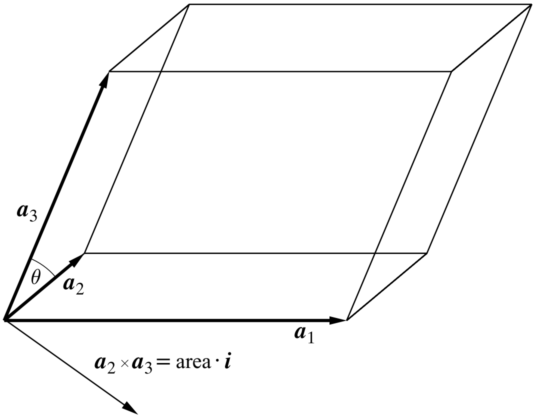


Fig. 9.3. The volume V of the unit cell is equal to $V = a_1 \cdot (\textit{area of the base})\, i = a_1 \cdot (a_2 \times a_3)$.

where

$$V = a_1 \cdot (a_2 \times a_3) \tag{9.24}$$

is the volume of the unit cell of the crystal (Fig. 9.3).

9.4 FIRST BRILLOUIN ZONE (FBZ)

Wigner–Seitz cell

As was remarked at the beginning of this chapter, the example of a jigsaw puzzle shows us that a parallelepiped unit cell does not represent the only choice. Now, we will profit from this extra freedom and will define the so called *Wigner–Seitz* unit cell. Here is the prescription of how to construct it (Fig. 9.4):

We focus on a node W, saw the crystal along the plane that dissects (symmetrically) the distance to a nearest neighbour node, throw the part that *does not* contain W into the fire-place, then repeat the procedure until we are left with a solid containing W. This solid represents the First Brillouin Zone (FBZ).

Léon Nicolas Brillouin (1889–1969), French physicist, professor at the Sorbonne and College de France in Paris, after 1941 in the USA: at the University of Madison, Columbia University, Harvard University. His contributions included quantum mechanics and solid state theory (he is one of the founders of electronic band theory).

9.5 PROPERTIES OF THE FBZ

> The vectors $\boldsymbol{k}$, which begin at the origin and end in the FBZ, label *all different* irreducible representations of the translational symmetry group.

Let us imagine two inverse space vectors $\boldsymbol{k}'$ and $\boldsymbol{k}''$ related by the equality $\boldsymbol{k}'' = \boldsymbol{k}' + \boldsymbol{K}_s$, where $\boldsymbol{K}_s$ stands for an inverse lattice vector. Taking into account the way the FBZ has been constructed, if one of them, say, $\boldsymbol{k}'$ indicates a point in the

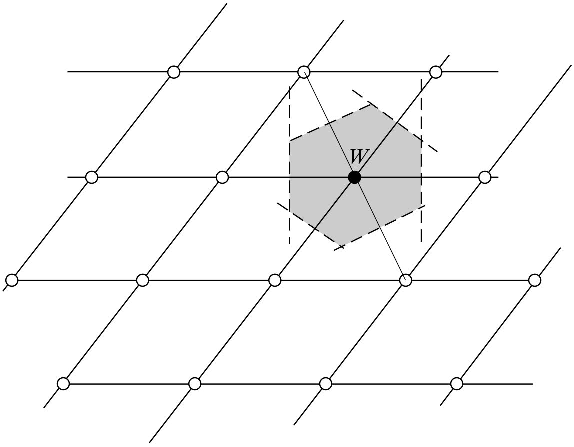


Fig. 9.4. Construction of the First Brillouin Zone (FBZ) as a Wigner–Seitz unit cell of the inverse lattice in 2D. The circles represent the nodes of the inverse lattice. We cut the lattice in the middle between the origin node W and all the other nodes (here it turns out to be sufficient to take only the nearest and the next nearest neighbours) and remove all the sawn-off parts that do not contain W. Finally we obtain the FBZ in the form of a hexagon. The Wigner–Seitz unit cells (after performing all allowed translations in the inverse lattice) reproduce the complete inverse space.

interior of the FBZ, then the second, $\boldsymbol{k}''$, "protrudes" outside the FBZ. Let us try to construct a Bloch function that corresponds to $\boldsymbol{k}''$:

$$\phi_{\boldsymbol{k}''} = \sum_j \exp(i\boldsymbol{k}''\boldsymbol{R}_j)\chi(\boldsymbol{r}-\boldsymbol{R}_j) = \sum_j \exp\big(i(\boldsymbol{k}'+\boldsymbol{K}_s)\boldsymbol{R}_j\big)\chi(\boldsymbol{r}-\boldsymbol{R}_j) \quad (9.25)$$

$$= \exp(i\boldsymbol{K}_s\boldsymbol{R}_j)\sum_j \exp(i\boldsymbol{k}'\boldsymbol{R}_j)\chi(\boldsymbol{r}-\boldsymbol{R}_j) \quad (9.26)$$

$$= \exp(i2\pi M_{sj})\sum_j \exp(i\boldsymbol{k}'\boldsymbol{R}_j)\chi(\boldsymbol{r}-\boldsymbol{R}_j) \quad (9.27)$$

$$= \sum_j \exp(i\boldsymbol{k}'\boldsymbol{R}_j)\chi(\boldsymbol{r}-\boldsymbol{R}_j) = \phi_{\boldsymbol{k}'}. \quad (9.28)$$

It turns out that our function ϕ does behave in a way identical to $\boldsymbol{k}'$. We say that the two vectors are *equivalent*.

equivalent vectors

> Vector $\boldsymbol{k}$ outside the FBZ is always equivalent to a vector from inside the FBZ, while two vectors from inside of the FBZ are never equivalent.

Therefore, if we are interested in electronic states (the irreducible representation of the translation group are labelled by $\boldsymbol{k}$ vectors) it is sufficient to limit ourselves to those $\boldsymbol{k}$ vectors that are enclosed in the FBZ.

9.6 A FEW WORDS ON BLOCH FUNCTIONS

9.6.1 WAVES IN 1D

Let us take a closer look of a Bloch function corresponding to the vector $\boldsymbol{k}$:

$$\phi_{\boldsymbol{k}}(\boldsymbol{r}) = \sum_j \exp(i\boldsymbol{k}\boldsymbol{R}_j)\chi(\boldsymbol{r}-\boldsymbol{R}_j) \quad (9.29)$$

and limit ourselves to 1D. In such a case, the wave vector $\boldsymbol{k}$ reduces to a *wave number* k, and the vectors $\boldsymbol{R}_j$ can all be written as $\boldsymbol{R}_j = aj\boldsymbol{z}$, where $\boldsymbol{z}$ stands for the unit vector along the periodicity axis, a means the lattice constant (i.e. the nearest-neighbour distance), while $j = 0, \pm 1, \pm 2, \ldots$ Let us assume that in the lattice nodes we have hydrogen atoms with orbitals $\chi = 1s$. Therefore, in 1D we have:

$$\phi_k(\boldsymbol{r}) = \sum_j \exp(ikja)\chi(\boldsymbol{r} - aj\boldsymbol{z}). \tag{9.30}$$

Let me stress that ϕ_k represents a function of position $\boldsymbol{r}$ in the 3D space and only the periodicity has a 1D character. The function is a linear combination of the hydrogen atom $1s$ orbitals. The linear combination depends exclusively on the value of k. Eq. (9.28) tells us that the allowed $k \in (0, \frac{2\pi}{a})$, or alternatively $k \in (-\frac{\pi}{a}, \frac{\pi}{a})$. If we exceed the FBZ *length* $\frac{2\pi}{a}$, then we would simply repeat the Bloch functions. For $k = 0$ we get

$$\phi_0 = \sum_j \exp(0)\chi(\boldsymbol{r} - aj\boldsymbol{z}) = \sum_j \chi(\boldsymbol{r} - aj\boldsymbol{z}), \tag{9.31}$$

i.e. simply a sum of the $1s$ orbitals. Such a sum has a large value on the nuclei, and close to a nucleus ϕ_0 will be delusively similar to its $1s$ orbital, Fig. 9.5.a.

The function looks like a chain of buoys floating on a perfect water surface. If we ask whether ϕ_0 represents a wave, the answer could be, that if it does then its wave length is ∞. What about $k = \frac{\pi}{a}$? In such a case:

$$\begin{aligned}\phi_{\frac{\pi}{a}}(\boldsymbol{r}) &= \sum_j \exp(ij\pi)\chi(\boldsymbol{r} - aj\boldsymbol{z}) = \sum_j (\cos \pi j + i \sin \pi j)\chi(\boldsymbol{r} - aj\boldsymbol{z}) \\ &= \sum_j (-1)^j \chi(\boldsymbol{r} - aj\boldsymbol{z}).\end{aligned}$$

If we decide to draw the function in space, we would obtain Fig. 9.5.b. When asked this time, we would answer that the wave length is equal to $\lambda = 2a$, which by the way is equal to[16] $\frac{2\pi}{|k|}$. There is a problem. Does the wave correspond to $k = \frac{\pi}{a}$ or $k = -\frac{\pi}{a}$? It corresponds to *both* of them. Well, does it contradict the theorem that the FBZ contains all *different* states? No, everything is OK. Both functions are from the border of the FBZ, their k values differ by $\frac{2\pi}{a}$ (one of the inverse lattice vectors) and therefore both functions represent *the same state*.

Now, let us take $k = \frac{\pi}{2a}$. We obtain

$$\begin{aligned}\phi_k(\boldsymbol{r}) &= \sum_j \exp\left(\frac{i\pi j}{2}\right)\chi(\boldsymbol{r} - aj\boldsymbol{z}) \\ &= \sum_j \left(\cos\left(\frac{\pi j}{2}\right) + i \sin\left(\frac{\pi j}{2}\right)\right)\chi(\boldsymbol{r} - aj\boldsymbol{z})\end{aligned} \tag{9.32}$$

with some coefficients being complex numbers. For $j = 0$ the coefficient is equal to 1, for $j = 1$ equals i, for $j = 2$ it takes the value -1, for $j = 3$ it attains $-i$, for $j = 4$ it is again 1, and the values repeat periodically. This is depicted in Fig. 9.5.c.

---

[16] In the preceding case the formula $\lambda = \frac{2\pi}{k}$ also worked, because it gave $\lambda = \infty$.

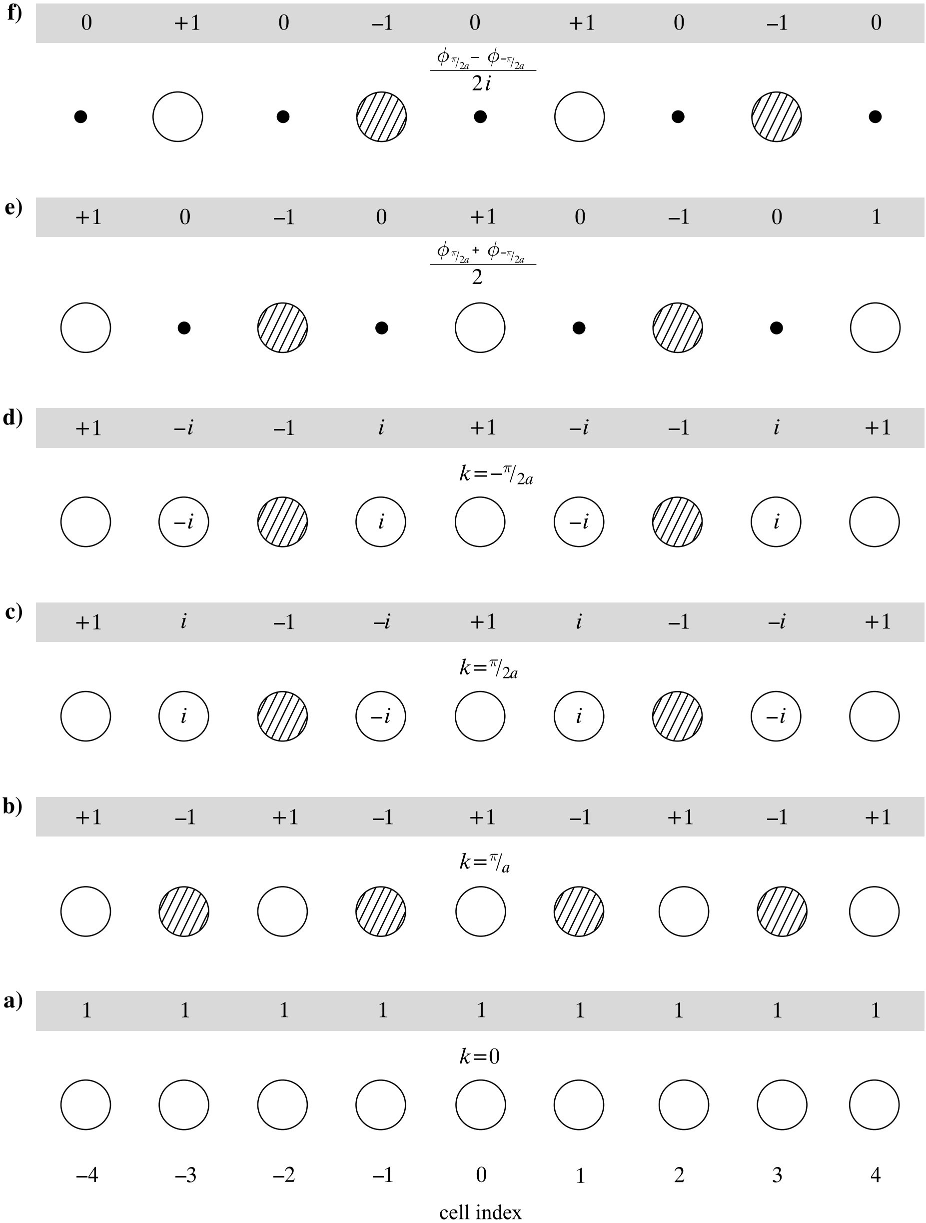


Fig. 9.5. Waves in 1D. Shadowed (white) circles mean negative (positive) value of the function. Despite the fact that some waves are complex, in each of the cases (a)–(f) we are able to determine their wave length.

If this time we ask whether we see any wave there, we have to answer that yes we do, because after the length $4a$ everything begins to repeat. Therefore, $\lambda = 4a$ and again it equals to $\frac{2\pi}{k} = \frac{2\pi}{\frac{\pi}{2a}}$. Everything is OK except that humans like pictures more than schemes. Can we help it somehow? Let us take a look of $\phi_k(\boldsymbol{r})$ which corresponds to $k = -\frac{\pi}{2a}$. We may easily convince ourselves that this situation corresponds to what we have in Fig. 9.5.d.

Let us stress that $\phi_{-k} = \phi_k^*$ represents *another* complex wave. By adding and subtracting $\phi_k(\boldsymbol{r})$ and $\phi_{-k}(\boldsymbol{r})$ we receive the real functions, which can be plotted and that is all we need. By adding $\frac{1}{2}(\phi_k + \phi_{-k})$, we obtain

$$\frac{1}{2}(\phi_k + \phi_{-k}) = \sum_j \cos\left(\frac{\pi j}{2}\right)\chi(\boldsymbol{r} - ajz), \tag{9.33}$$

while $\frac{1}{2i}(\phi_k - \phi_{-k})$ results in

$$\frac{1}{2i}(\phi_k - \phi_{-k}) = \sum_j \sin\left(\frac{\pi j}{2}\right)\chi(\boldsymbol{r} - ajz). \tag{9.34}$$

Now, there is no problem with plotting the new functions (Fig. 9.5.e,f).[17]

A similar technique may be applied to any k. Each time we will find that the wave we see exhibits the wave length $\lambda = \frac{2\pi}{k}$.

9.6.2 WAVES IN 2D

Readers confident in their understanding of the wave vector concept may skip this subsection.

This time we will consider the crystal as two-dimensional rectangular lattice, therefore, the corresponding inverse lattice is also two-dimensional as well as the wave vectors $\boldsymbol{k} = (k_x, k_y)$.

Let us take first $\boldsymbol{k} = (0, 0)$. We immediately obtain $\phi_{\boldsymbol{k}}$ shown in Fig. 9.6.a, which corresponds to infinite wave length (again $\lambda = \frac{2\pi}{k}$) or "no wave" at all.

Let us try $\boldsymbol{k} = (\frac{\pi}{a}, 0)$. The summation over j may be replaced by a double summation (indices m and n along the x and y axes, respectively), therefore, $\boldsymbol{R}_j = ma\boldsymbol{x} + nb\boldsymbol{y}$, where m and n correspond to the unit cell j, a and b denote the lattice constants along the axes shown by the unit vectors $\boldsymbol{x}$ and $\boldsymbol{y}$. We have

$$\begin{aligned}\phi_{\boldsymbol{k}} &= \sum_{mn} \exp\bigl(i(k_x ma + k_y nb)\bigr)\chi(\boldsymbol{r} - ma\boldsymbol{x} - nb\boldsymbol{y}) \\ &= \sum_{mn} \exp(i\pi m)\chi(\boldsymbol{r} - ma\boldsymbol{x} - nb\boldsymbol{y}) = \sum_{mn}(-1)^m\chi(\boldsymbol{r} - ma\boldsymbol{x} - nb\boldsymbol{y}).\end{aligned}$$

If we go through all m and n, it easily seen that moving along x we will meet the signs $+1, -1, +1, -1, \ldots,$ while moving along y we have the same sign all the time. This will correspond to Fig. 9.6.b.

This is a wave.

---

[17] And what would happen if we took $k = \frac{\pi}{a}\frac{m}{n}$, with the integer $m < n$? We would again obtain a wave with the wave length $\lambda = \frac{2\pi}{k}$, i.e. in this case $\lambda = \frac{n}{m}2a$. It would be quite difficult to recognize such a wave computed at the lattice nodes, because the closest wave maxima would be separated by $n2a$ and this length would have been covered by m wavelengths.

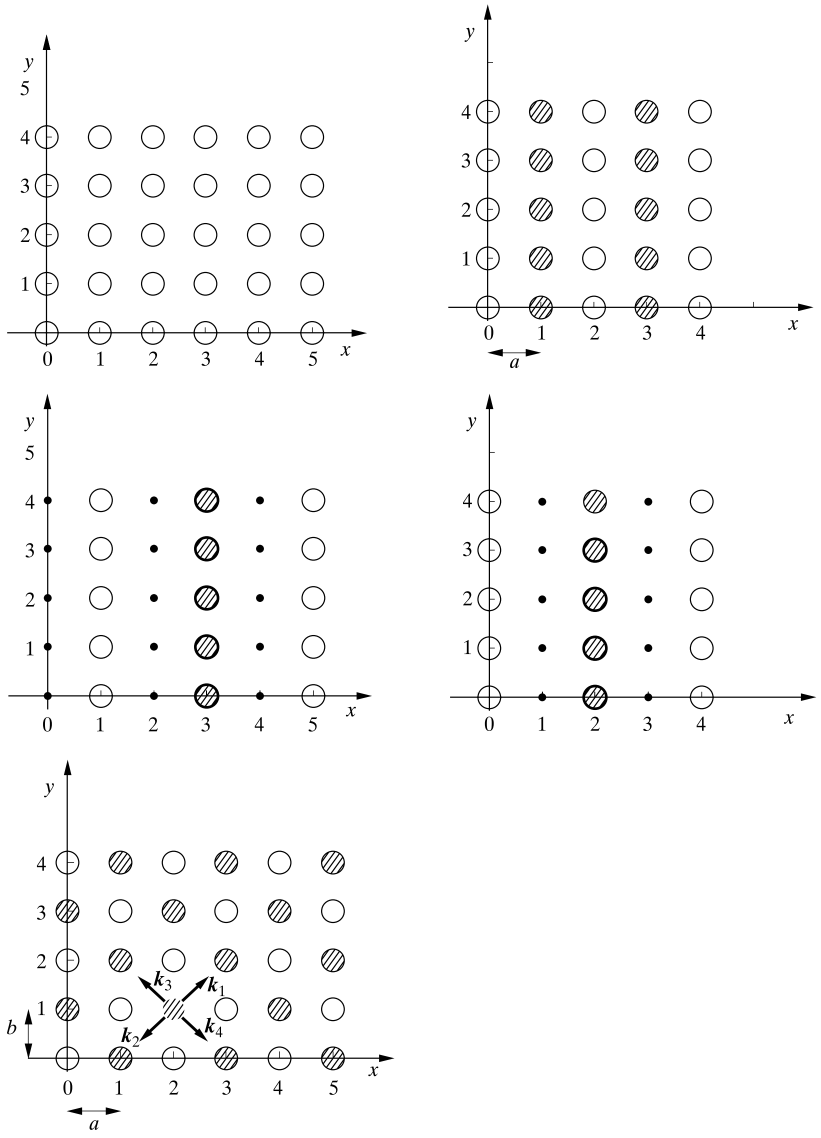


Fig. 9.6. Waves in 2D. Shadowed (white) circles mean negative (positive) value of the function. In any case $\lambda = \frac{2\pi}{k}$, while the wave vector $\boldsymbol{k}$ points to the direction of the wave propagation. a) $\boldsymbol{k} = (0,0)$; b) $\boldsymbol{k} = (\frac{\pi}{a}, 0)$; c) $\boldsymbol{k} = (\frac{\pi}{2a}, 0)$, $\frac{1}{2i}(\phi_{\boldsymbol{k}} - \phi_{-\boldsymbol{k}})$; d) $\boldsymbol{k} = (\frac{\pi}{2a}, 0)$, $\frac{1}{2}(\phi_{\boldsymbol{k}} + \phi_{-\boldsymbol{k}})$; e) $\boldsymbol{k} = (\frac{\pi}{a}, \frac{\pi}{b})$.

The wave fronts are oriented along y, i.e. the wave runs along the x axis, therefore, in the direction of the wave vector $\boldsymbol{k}$. The same happened in the 1D cases, but we did not express that explicitly: the wave moved along the (1D) vector $\boldsymbol{k}$.

Exactly as before the wave length is equal to 2π divided by the length of $\boldsymbol{k}$. Since we are at the FBZ border, a wave with $-\boldsymbol{k}$ simply means the same wave as for $\boldsymbol{k}$.

If we take $\boldsymbol{k} = [\frac{\pi}{2a}, 0]$, then

$$\begin{aligned}\phi_{\boldsymbol{k}} &= \sum_{mn} \exp\bigl(i(\boldsymbol{k}_x ma + \boldsymbol{k}_y nb)\bigr)\chi(\boldsymbol{r} - ma\boldsymbol{x} - nb\boldsymbol{y}) \\ &= \sum_{mn} \exp\left(\frac{i\pi m}{2}\right)\chi(\boldsymbol{r} - ma\boldsymbol{x} - nb\boldsymbol{y}).\end{aligned}$$

This case is very similar to that in 1D for $k = \frac{\pi}{2a}$, when we look at the index m and $k = 0$, and when we take into account the index n. We may carry out the same trick with addition and subtraction, and immediately get Figs. 9.6.c and d.

Is there any wave over there? Yes, there is. The wave length equals $4a$, i.e. exactly $\lambda = \frac{2\pi}{k}$, and the wave is directed along vector $\boldsymbol{k}$. When making the figure, we also used the wave corresponding to $-\boldsymbol{k}$, therefore, neither the sum nor the difference correspond to $\boldsymbol{k}$ or $-\boldsymbol{k}$, but rather to both of them (we have two standing waves). The reader may guess the wave length and direction of propagation for $\phi_{\boldsymbol{k}}$ corresponding to $\boldsymbol{k} = [0, \frac{\pi}{2b}]$.

Let us see what happens for $\boldsymbol{k} = [\frac{\pi}{a}, \frac{\pi}{b}]$. We obtain

$$\begin{aligned}\phi_{\boldsymbol{k}} &= \sum_{mn} \exp\bigl(i(k_x ma + k_y nb)\chi(\boldsymbol{r} - ma\boldsymbol{x} - nb\boldsymbol{y}) \\ &= \sum_{mn} \exp\bigl(i(m\pi + n\pi)\bigr)\chi(\boldsymbol{r} - ma\boldsymbol{x} - nb\boldsymbol{y}) \\ &= \sum_{mn} (-1)^{m+n}\chi(\boldsymbol{r} - ma\boldsymbol{x} - nb\boldsymbol{y}),\end{aligned}$$

which produces waves propagating along $\boldsymbol{k}$. And what about the wave length? We obtain[18]

$$\lambda = \frac{2\pi}{\sqrt{(\frac{\pi}{a})^2 + (\frac{\pi}{b})^2}} = \frac{2ab}{\sqrt{a^2 + b^2}}. \tag{9.35}$$

In the last example there is something that may worry us. As we can see, our figure corresponds not only to $\boldsymbol{k}_1 = (\frac{\pi}{a}, \frac{\pi}{b})$ and $\boldsymbol{k}_2 = (-\frac{\pi}{a}, -\frac{\pi}{b})$, which is understandable (as discussed above), but also to the wave with $\boldsymbol{k}_3 = (-\frac{\pi}{a}, \frac{\pi}{b})$ and to the wave evidently coupled to it, namely, with $\boldsymbol{k}_4 = (\frac{\pi}{a}, -\frac{\pi}{b})$! What is going on? Again, let us recall that we are on the FBZ border and this identity is natural, because the vectors $\boldsymbol{k}_2$ and $\boldsymbol{k}_3$ as well as $\boldsymbol{k}_1$ and $\boldsymbol{k}_4$ differ by the inverse lattice vector $(0, \frac{2\pi}{b})$, which makes the two vectors equivalent.

---

[18] The formula can be easily verified in two limiting cases. The first corresponds to $a = b$. Then, $\lambda = a\sqrt{2}$, and this agrees with Fig. 9.6.e. The second case is, when $b = \infty$, which gives $\lambda = 2a$, exactly as in the 1D case with $k = \frac{\pi}{a}$. This is what we expected.

9.7 THE INFINITE CRYSTAL AS A LIMIT OF A CYCLIC SYSTEM

Band structure

Let us consider the hydrogen atom in its ground state (cf. p. 178). The atom is described by the atomic orbital 1*s* and corresponds to energy −0.5 a.u. Let us now take two such atoms. We have two molecular orbitals: bonding and antibonding (cf. p. 371), which correspond, respectively, to energies a bit lower than −0.5 and a bit higher than −0.5 (this bit is larger if the overlap of the atomic orbitals gets larger). We therefore have two energy levels, which stem directly from the 1*s* levels of the two hydrogen atoms. For three atoms we would have three levels, for 10^{23} atoms we would get 10^{23} energy levels, that would be densely distributed along the energy scale, but would not cover the whole scale. There will be a bunch of energy levels stemming from 1*s*, i.e. an *energy band* of allowed electronic states. If we had an infinite chain of hydrogen atoms, there would be a band resulting from 1s levels, a band stemming from 2s, 2p, etc., the bands might be separated by *energy gaps*.

bands

energy gap

How dense would the distribution of the electronic levels be? Will the distribution be uniform? Answers to such questions are of prime importance for the electronic theory of crystals. It is always advisable to come to a conclusion by steps, starting from something as simple as possible, which we understand very well.

Fig. 9.7 shows how the energy level distribution looks for longer and longer rings (regular polygon) of hydrogen atoms. One of important features of the distribution is that

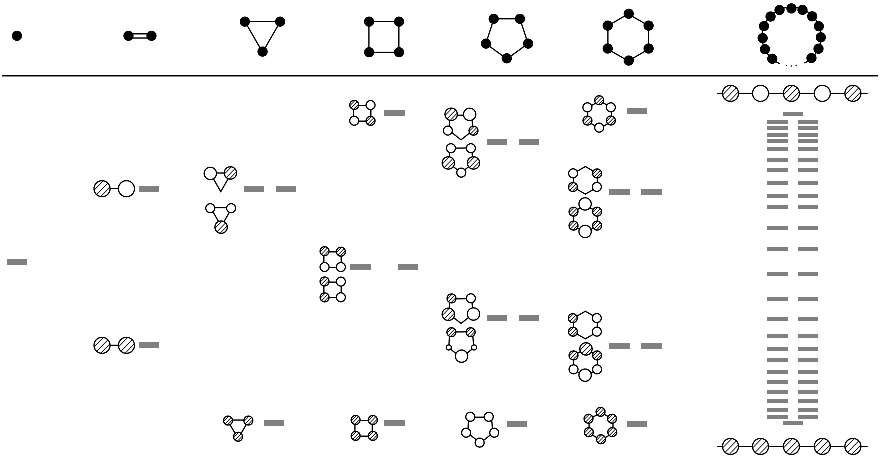


Fig. 9.7. Energy level distribution for a regular polygon built from hydrogen atoms. It is seen that the energy levels are located within an energy band, and are closer to one another at the band edges. The centre of the band is close to energy 0, taken as the binding energy in the isolated hydrogen atom (equal to −0.5 a.u.). Next to energy levels the molecular orbitals are shown schematically (the shadowed circles mean negative values). R. Hoffmann, "*Solids and Surfaces. A Chemist's View of Bonding in Extended Structures*", VCH Publishers, New York, © 1988 VCH Publishers. Reprinted with permission of John Wiley & Sons, Inc.

the levels extend over an energy interval and are more numerous at energy extremes.

How do the wave functions that correspond to higher and higher-energy levels in a band look? Let us see the situation in the ring H_n molecules. Fig. 9.7 indicates that the rule is very simple. The number of nodes of the wave function increases by one, when we go to the next level (higher in the energy scale).[19]

Born–von Kármán condition in 1D

How is it in the case of a crystal? Here we are confronted with the first difficulty. Which crystal, of what shape? Should it be an ideal crystal, i.e. with perfectly ordered atoms? There is nothing like the perfect crystal in Nature. For the sake of simplicity (as well as generality) let us assume, however, that our crystal *is* perfect indeed. Well, and now what about its surface (shape)? Even if we aimed at studying the surface of a crystal, the first step would be the infinite crystal (i.e. with no surface). This is the way theoreticians always operate.[20]

One of the ingenious ideas in this direction is known as the *Born–von Kármán boundary conditions*. The idea is that instead of considering a crystal treated as a stick (let us consider 1D case) we treat it as a circle, i.e. *the value of the wavefunction at one end of the stick has to be equal to the wavefunction value at the other end*. In this way we remove the problem of the crystal ends, and on top of that, all the unit cells become equivalent.

Theodore von Kármán (1881–1963), American physicist of Hungarian origin, director of the Guggenheim Aeronautical Laboratory at the California Institute of Technology in Pasadena. von Kármán was also a founder of the NASA Jet Propulsion Laboratory and father of the concept of the first supersonic aeroplane. On the Hungarian stamp one can see the famous "Kármán vortex street" behind an aeroplane. He was asked by the father of the young mathematical genius John von Neumann to persuade him that the job of a mathematician is far less exciting than that of a banker. Theodore von Kármán (to the benefit of science) did not accomplish this mission well.

The same may be done in 2D and 3D cases. We introduce usually the Born–von Kármán boundary conditions for a finite N and then go with N to ∞. After such a procedure is carried out, we are pretty sure that the solution we are going to

---

[19] They are bound to differ by the number of nodes, because this assures their mutual orthogonality (required for the eigenfunctions of a Hermitian operator).

[20] People say that when theoreticians attack the problem of stability of a table as a function of the number n of its legs, they do it in the following way. First, they start with $n = 0$, then they proceed with $n = 1$, then they go to $n = \infty$, and after that they have no time to consider other values of n.

obtain will not only be true for an infinite cycle but also for the mass (bulk) of the infinite crystal. This stands to reason, provided that the crystal surface does not influence the (deep) bulk properties at all.[21] In the ideal periodic case, we have to do with the cyclic translational symmetry group (Appendix C on p. 903). The group is Abelian and, therefore, all the irreducible representations have dimension 1.

Let us assume we have to do with N equidistant atoms located on a circle, the nearest-neighbour distance being a. From the Bloch theorem for the wave function ψ we have

$$\psi(aN) = \exp(-ikaN)\psi(0), \tag{9.36}$$

where we have assumed that the wave function ψ corresponds to the wave vector $\boldsymbol{k}$ (here, in 1D, wave number k).

The Born–von Kármán condition means:

$$\psi(aN) = \psi(0), \tag{9.37}$$

or

$$\exp(-ikaN) = 1. \tag{9.38}$$

From this follows that:

$$kaN = 2\pi J, \tag{9.39}$$

where $J = 0, \pm 1, \pm 2, \ldots$. This means that only *some* k are allowed, namely $k = \frac{2\pi}{a}\frac{J}{N}$.

The Bloch functions take the form [cf. eq. (9.29)]:

$$\sum_j \exp(ikja)\chi_j, \tag{9.40}$$

where χ_j denotes a given atomic orbital (e.g., 1s) centred on atom j. The summation over j in our case is finite, because we only have N atoms, $j = 0, 1, 2, \ldots,$ $N - 1$. Let us consider $J = 0, 1, 2, \ldots, N - 1$ and the corresponding values of $k = \frac{2\pi}{a}\frac{J}{N}$. For each k we have a Bloch function, altogether we have, therefore, N Bloch functions. Now, we may try to increase J and take $J = N$. The corresponding Bloch function may be written as

$$\sum_j \exp(i2\pi j)\chi_j = \sum_j \chi_j, \tag{9.41}$$

which turns out to be identical to the Bloch function with $k = 0$, i.e. with $J = 0$. We are reproducing what we already have. It is clear, therefore, that we have a set of those k, that form a *complete set of non-equivalent states*, they correspond to $J = 0, 1, 2, \ldots, N - 1$. It is also seen that if the limits of this set are shifted by the same integer, then we still have the same complete set of non-equivalent states. Staying for the time being with our primary choice of the set, we will get

---

[21] We circumvent the difficult problem of the crystal surface. The boundary (surface) problem is extremely important for obvious reasons: we usually *have to do with this, not with the bulk*. The existence of the surface leads to some specific surface-related electronic states.

N values of $k \in [0, \frac{2\pi}{a}\frac{N-1}{N}]$, i.e. $k \in \{0, \frac{2\pi}{a}\frac{1}{N}, \frac{2\pi}{a}\frac{2}{N}, \dots, \frac{2\pi}{a}\frac{N-1}{N}\}$. *Those k values are equidistant*. When $N \to \infty$ then the section to be divided attains the length $\frac{2\pi}{a}$. Hence

> the non-equivalent states (going with N to infinity) correspond to those k's that are from section $[0, \frac{2\pi}{a}]$ or shifted section $[-\frac{\pi}{a}, +\frac{\pi}{a}]$, called the FBZ. We are allowed to make any shift, because, as we have shown, we keep the same non-equivalent values of k. The allowed k values are distributed *uniformly* within the FBZ. The number of the allowed k's is equal to ∞, because $N = \infty$ (and the number of the allowed k's is always equal to N).

k-dependence of orbital energy

Let us take the example of benzene ($N = 6$, Fig. 9.7) and consider only those molecular orbitals that can be written as linear combinations of the carbon $2p_z$, where z is the axis orthogonal to the plane of the molecule. The wave vectors[22] ($k = \frac{2\pi}{a}\frac{J}{N}$) may be chosen as corresponding to $J = 0, 1, 2, \dots, 5$, or equivalently to $J = -3, -2, -1, 0, +1, +2$. It is seen that $J = 0$ gives a nodeless function,[23] $J = \pm 1$ lead to a pair of the Bloch functions with a single node, $J = \pm 2$ give a pair of the two-node functions, and finally $J = -3$ corresponds to a three-node function.

bonding and antibonding interaction

It has occasionally been remarked in this book (cf., e.g., Chapter 4), that increasing the number of nodes[24] results in higher energy. This rule becomes most transparent in the present case. A nodeless Bloch function means that all the contacts between the $2p$ orbitals are *bonding*, which results in *low energy*. A single node means introducing two nearest-neighbour *antibonding* interactions, and this is bound to cause an energy increase. Two nodes result in four antibonding interactions, and the energy goes up even more. Three nodes already give all the nearest-neighbour contacts of antibonding character and the energy is the highest possible.

9.8 A TRIPLE ROLE OF THE WAVE VECTOR

As has already been said, the wave vector (in 1D, 2D and 3D) plays several roles. Here they are:

1. The wave vector $\boldsymbol{k}$ tells us which type of plane wave arranged from certain objects (like atomic orbitals) we are concerned with. The direction of $\boldsymbol{k}$ is the propagation direction, the wave length is $\lambda = \frac{2\pi}{|\boldsymbol{k}|}$.

---

[22] In this case this is a wave number.

[23] We neglect here the node that follows from the reflection in the molecular plane as being shared by all the molecular orbitals considered.

[24] That is, considering another wavefunction that has a larger number of nodes.

2. The wave vector may also be treated as a label for the irreducible representation of the translation group.

In other words, $\boldsymbol{k}$ determines which irreducible representation we are dealing with (Appendix C on p. 903). This means that $\boldsymbol{k}$ tells us *which permitted rhythm* is exhibited by the coefficients at atomic orbitals in a particular Bloch function (permitted, i.e. assuring that the square has the symmetry of the crystal). There are a lot of such rhythms, e.g., all the coefficients equal each other ($k = 0$), or one node introduced, two nodes, etc. The FBZ represents a set of such $\boldsymbol{k}$, which corresponds to *all possible rhythms*, i.e. non-equivalent Bloch functions.[25] In other words the FBZ gives us all the possible symmetry orbitals that can be formed from an atomic orbital.

3. The longer the $\boldsymbol{k}$, the more nodes the Bloch function $\phi_{\boldsymbol{k}}$ has: $|\boldsymbol{k}| = 0$ means no nodes, at the boundary of the FBZ there is the maximum number of nodes.

9.9 BAND STRUCTURE

9.9.1 BORN–VON KÁRMÁN BOUNDARY CONDITION IN 3D

The Hamiltonian $\hat{H}$ we were talking about represents an effective one-electron Hamiltonian. From Chapter 8, we know that it may be taken as the Fock operator. A crystal represents nothing but a huge (quasi-infinite) molecule, and assuming the Born–von Kármán condition, a huge cyclic molecule. This is how we will get the Hartree–Fock solution for the crystal – by preparing the Hartree–Fock solution for a cyclic molecule and then letting the number of unit cells N go to infinity.

Hence, let us take a large piece of crystal – a parallelepiped with the number of unit cells *in each of the periodicity directions* (i.e. along the three basis vectors) equal to $2N + 1$ (the reference cell 0, N cells on the right, N cells on the left). The particular number, $2N + 1$, is not very important, we have only to be sure that such a number is large. We assume that the Born–von Kármán condition is fulfilled. This means that we treat the crystal like a snake eating its tail, and this will happen on every of the three periodicity axes. This enables us to treat the translation group as a cyclic group, which gives an enormous simplification to our task. The cyclic group of the lattice constants a, b, c implies that [cf. eq. (9.38)]

$$\exp\bigl(-ik_x a(2N+1)\bigr) = 1, \tag{9.42}$$

$$\exp\bigl(-ik_y b(2N+1)\bigr) = 1, \tag{9.43}$$

$$\exp\bigl(-ik_z c(2N+1)\bigr) = 1, \tag{9.44}$$

which can be satisfied only for some special vectors $\boldsymbol{k} = (k_x, k_y, k_z)$ satisfying:

---

[25] That is, linearly independent.

$$k_x = \frac{2\pi}{a} \frac{J_x}{2N+1}, \tag{9.45}$$

$$k_y = \frac{2\pi}{b} \frac{J_y}{2N+1}, \tag{9.46}$$

$$k_z = \frac{2\pi}{c} \frac{J_z}{2N+1}, \tag{9.47}$$

with any of J_x, J_y, J_z taking $2N+1$ consecutive integer numbers. We may, for example, assume that $J_x, J_y, J_z \in \{-N, -N+1, \ldots, 0, 1, 2, \ldots, N\}$. Whatever N is, $\boldsymbol{k}$ will always satisfy

$$-\frac{\pi}{a} < k_x < \frac{\pi}{a}, \tag{9.48}$$

$$-\frac{\pi}{b} < k_y < \frac{\pi}{b}, \tag{9.49}$$

$$-\frac{\pi}{c} < k_z < \frac{\pi}{c}, \tag{9.50}$$

which is what we call the FBZ. We may therefore say that before letting $N \to \infty$

> the FBZ is filled with the allowed vectors $\boldsymbol{k}$ in a *grain-like way*; the number being equal to the number of unit cells, i.e. $(2N+1)^3$. Note that the distribution of the vectors allowed in the FBZ is *uniform*. This is assured by the numbers J, which divide the axes k_x, k_y, k_z in the FBZ into equal pieces.

9.9.2 CRYSTAL ORBITALS FROM BLOCH FUNCTIONS (LCAO CO METHOD)

crystal orbitals

What we expect to obtain finally in the Hartree–Fock method for an infinite crystal are the molecular orbitals, which in this context will be called the *crystal orbitals* (CO). As usual we will plan to expand the CO as linear combinations of atomic orbitals (cf. p. 360). Which atomic orbitals? Well, those which we consider appropriate[26] for a satisfactory description of the crystal, e.g., the atomic orbitals of all the atoms of the crystal. We feel, however, that we have to be defeated in trying to perform this task.

> There will be a lot of atomic orbitals, and therefore also an astronomic number of integrals to compute (infinite for the infinite crystal) and that is it, we cannot help this. On the other hand, if we begin such a hopeless task, the value of any integral would repeat an infinite number of times. This indicates a chance to simplify the problem. Indeed, we have not yet used the translational symmetry of the system.

If we are going to use the symmetry, then *we may create the Bloch functions representing the building blocks that guarantee the proper symmetry in advance*. Each

---

[26] As for molecules.

Bloch function is built from an atomic orbital χ:

$$\phi_{\boldsymbol{k}} = (2N+1)^{-\frac{3}{2}} \sum_j \exp(i\boldsymbol{k}\boldsymbol{R}_j)\chi(\boldsymbol{r}-\boldsymbol{R}_j). \tag{9.51}$$

The function is identical to that of eq. (9.29), except it has a factor $(2N+1)^{-\frac{3}{2}}$, which makes the function approximately normalized.[27]

Any CO will be a linear combination of such Bloch functions, each corresponding to a given χ. This is equivalent to the LCAO expansion for molecular orbitals, the only difference is that we have cleverly preorganized the atomic orbitals (of one type) into symmetry orbitals (Bloch functions). Hence, it is indeed appropriate to call this approach as the LCAO CO method (*Linear Combination of Atomic Orbitals – Crystal Orbitals*), analogous to the LCAO MO (cf. p. 362). There is, however, a problem. Each CO should be a linear combination of the $\phi_{\boldsymbol{k}}$ for various types of χ and for *various* $\boldsymbol{k}$. Only then would we have the full analogy: a molecular orbital is a linear combination of all the atomic orbitals belonging to the atomic basis set.[28]

It will be shown below that the situation is far better:

> each CO corresponds to a single vector $\boldsymbol{k}$ from the FBZ and is a linear combination of the Bloch functions, each characterized by this $\boldsymbol{k}$.

There are, however, only a few Bloch functions – their number is equal to the number of the atomic orbitals per unit cell (denoted by ω). Our optimism pertains, of course, to taking a modest atomic basis set (small ω).

It is easy to show that, indeed, we can limit ourselves to a single vector $\boldsymbol{k}$. Imagine this is false, and our CO is a linear combination of all the Bloch functions

---

[27] The function without this factor is of class Q, i.e. normalizable for any finite N, but non-normalizable for $N=\infty$. The approximate normalization makes the function square integrable, even for $N=\infty$. Let us see:

$$\begin{aligned}\langle\phi_{\boldsymbol{k}}|\phi_{\boldsymbol{k}}\rangle &= (2N+1)^{-3}\sum_j\sum_{j'}\exp\big(i\boldsymbol{k}(\boldsymbol{R}_j-\boldsymbol{R}_{j'})\big)\int\chi(\boldsymbol{r}-\boldsymbol{R}_j)\chi(\boldsymbol{r}-\boldsymbol{R}_{j'})\,\mathrm{d}\tau\\ &= (2N+1)^{-3}\sum_j\sum_{j'}\exp\big(i\boldsymbol{k}(\boldsymbol{R}_j-\boldsymbol{R}_{j'})\big)\int\chi(\boldsymbol{r})\chi\big(\boldsymbol{r}-(\boldsymbol{R}_j-\boldsymbol{R}_{j'})\big)\,\mathrm{d}\tau,\end{aligned}$$

because the integral does depend on a *relative* separation in space of the atomic orbitals. Further,

$$\langle\phi_{\boldsymbol{k}}|\phi_{\boldsymbol{k}}\rangle = \sum_j \exp(i\boldsymbol{k}\boldsymbol{R}_j)\int\chi(\boldsymbol{r})\chi(\boldsymbol{r}-\boldsymbol{R}_j)\,\mathrm{d}\tau, \tag{9.52}$$

because we can replace a double summation over j and j' by a double summation over j and $j''=j-j'$ (both double summations exhaust all the lattice nodes), and the later summation always gives the same independent of j; the number of such terms is equal to $(2N+1)^3$. Finally, we may write $\langle\phi_{\boldsymbol{k}}|\phi_{\boldsymbol{k}}\rangle = 1+$ various integrals. The largest of these integrals is the nearest neighbour overlap integral of the functions χ. For normalized χ each of these integrals represents a fraction of 1 and additionally the contributions for further neighbours decay exponentially (cf. p. 1009). As a result, $\langle\phi_{\boldsymbol{k}}|\phi_{\boldsymbol{k}}\rangle$ is a number of the order of 1 or 2. This is what we have referred to as an approximate normalization.

[28] Indeed, for any $\boldsymbol{k}$ the number of distinct Bloch functions is equal to the number of atomic orbitals per unit cell. The number of allowed vectors, $\boldsymbol{k}$, is equal to the number of unit cells in the crystal. Hence, using the Bloch functions for all allowed $\boldsymbol{k}$ would be justified, any CO would represent a linear combination of all the atomic orbitals of the crystal.

corresponding to a given $\boldsymbol{k}$, then, of all the Bloch functions corresponding to the next $\boldsymbol{k}$, etc., up to the exhaustion of all the allowed $\boldsymbol{k}$. When, in the next step, we solve the orbital equation with the effective (i.e. Fock) Hamiltonian using the Ritz method, then we will end up computing the integrals $\langle\phi_{\boldsymbol{k}}|\hat{F}\phi_{\boldsymbol{k}'}\rangle$ and $\langle\phi_{\boldsymbol{k}}|\phi_{\boldsymbol{k}'}\rangle$. For $\boldsymbol{k} \neq \boldsymbol{k}'$ such integrals *are equal zero* according to group theory (Appendix C on p. 903), because $\hat{F}$ transforms according to the fully symmetric irreducible representation of the translation group,[29] while $\phi_{\boldsymbol{k}}$ and $\phi_{\boldsymbol{k}'}$ transform according to *different* irreducible representations.[30] Therefore the secular determinant in the Ritz method will have a *block form* (cf. Appendix C). The first block will correspond to the first $\boldsymbol{k}$, the second to the next $\boldsymbol{k}$, etc., where every block[31] would look as if in the Ritz method we used the Bloch functions corresponding uniquely to that particular $\boldsymbol{k}$. Conclusion: since a CO has to be a wave with a *given* $\boldsymbol{k}$, let us construct it with Bloch functions, which already have just this type of behaviour with respect to translation operators, i.e. have just this $\boldsymbol{k}$. This is fully analogous with the situation in molecules, if we used atomic symmetry orbitals.

> Thus each vector, $\boldsymbol{k}$, from the FBZ is associated with a crystal orbital, and therefore with a set of LCAO CO coefficients.

The number of such CO sets (each $\boldsymbol{k}$ – one set) *in principle* has to be equal to the number of unit cells, i.e. infinite.[32] The only profit we may expect could be associated with the hope that the computed quantities do not depend on $\boldsymbol{k}$ too much, but will rather change smoothly when $\boldsymbol{k}$ changes. This is indeed what will happen, *then a small number of vectors* **k** will be used, *and the quantities requiring other* **k** *will be computed by interpolation*.

Only a part of the computed COs will be occupied, and this depends on the orbital energy of a given CO, the number of electrons, and the corresponding $\boldsymbol{k}$, similar to which we had for molecules.

> The set of SCF LCAO CO equations will be very similar to the set for the molecular orbital method (SCF LCAO MO). In principle, the only difference will be that, in the crystal case, we will consequently use symmetry orbitals (Bloch functions) instead of atomic orbitals.

That's it. The rest of this section is associated with some technical details accompanying the operation $N \to \infty$.

9.9.3 SCF LCAO CO EQUATIONS

Let us write down the SCF LCAO CO equations as if they corresponded to a large molecule (Bloch functions will be used instead of atomic orbitals). Then the n-th

---

[29] Unit cells (by definition) are identical.

[30] Recall that $\boldsymbol{k}$ also has the meaning of the irreducible representation index (of the translation group).

[31] The whole problem can be split into the independent problems for individual blocks.

[32] Well, we cannot fool Mother Nature! Was there an infinite molecule (crystal) to be computed or not? Then the number of such sets of computations has to be infinite. Full stop.

CO may be written as (cf. eq. (8.49))

$$\psi_n(\boldsymbol{r},\boldsymbol{k})=\sum_q c_{qn}(\boldsymbol{k})\phi_q(\boldsymbol{r},\boldsymbol{k}), \tag{9.53}$$

where ϕ_q is the Bloch function corresponding to the atomic orbital χ_q:

$$\phi_q(\boldsymbol{r},\boldsymbol{k})=(2N+1)^{-\frac{3}{2}}\sum_j \exp(i\boldsymbol{k}\boldsymbol{R}_j)\chi_q^j, \tag{9.54}$$

with $\chi_q^j \equiv \chi_q(\boldsymbol{r}-\boldsymbol{R}_j)$ $(q=1,2,\dots,\omega)$.

The symbol χ_q^j means the q-th atomic orbital (from the set we prepared for the unit cell motif) located in the cell indicated by vector $\boldsymbol{R}_j$ (j-th cell).

In the expression for ψ_n, we have taken into account that there is no reason whatsoever that the coefficients c were $\boldsymbol{k}$-independent, since the expansion functions ϕ depend on $\boldsymbol{k}$. This situation does not differ from that, which we encountered in the Hartree–Fock–Roothaan method (cf. p. 365), with one technical exception: instead of the atomic orbitals we have symmetry orbitals, in our case Bloch functions.

The secular equations for the Fock operator will have, of course, the form of the Hartree–Fock–Roothaan equations (cf. Chapter 8, p. 365):

$$\sum_{q=1}^{\omega} c_{qn}[F_{pq}-\varepsilon_n S_{pq}]=0 \quad \text{for } p=1,2,\dots,\omega,$$

where the usual notation has been applied. For the sake of simplicity, we have not highlighted the $\boldsymbol{k}$-dependence of c, F and S. Whenever we decide to do this in future, we will put it in the form $F_{pq}(\boldsymbol{k})$, $S_{pq}(\boldsymbol{k})$, etc. Of course, ε_n will become a function of $\boldsymbol{k}$, as will be stressed by the symbol $\varepsilon_n(\boldsymbol{k})$. Theoretically, the secular equation has to be solved for every $\boldsymbol{k}$ of the FBZ.

Therefore, despite the fact that the secular determinant is of rather low rank (ω), the infinity of the crystal, forces us to solve this equation an infinite number of times. For the time being, though, do not worry too much.

9.9.4 BAND STRUCTURE AND BAND WIDTH

The number of secular equation solutions is equal to ω, and let us label them using index n. If we focus on one such solution, and check how $\varepsilon_n(\boldsymbol{k})$ and $\psi_n(\boldsymbol{r},\boldsymbol{k})$ are sensitive to a tiny change of $\boldsymbol{k}$ within the FBZ, it turns out that $\varepsilon_n(\boldsymbol{k})$ and $\psi_n(\boldsymbol{r},\boldsymbol{k})$ change smoothly. This may not be true when $\boldsymbol{k}$ passes through the border of the FBZ.

The function $\varepsilon_n(\boldsymbol{k})$ is called the n-th electronic band.

If we travelled in the FBZ, starting from the origin and continuing along a straight line, then $\varepsilon_1, \varepsilon_2, \ldots$ etc. would change as functions of $\boldsymbol{k}$ and we would be concerned with several energy bands. If $\varepsilon_n(\boldsymbol{k})$ changes very much during our travel over the FBZ, we would say that the n-th band has a large width or *dispersion*.

band width

As it was shown on p. 445 for the hydrogen atoms an energy band forms due to the bonding and antibonding effects. If instead of hydrogen atoms, we put a unit cell with a few atoms inside (motif), then the story is similar: the motif has some one-electron energy levels (orbital energies), putting together the unit cells makes changing these energy levels into energy bands, the number of levels in any band is equal to the number of unit cells, or the number of allowed $\boldsymbol{k}$ vectors in FBZ.

The band width is related to interactions among the unit cell contents, and is roughly proportional to the overlap integral between the orbitals of the interacting unit cells.

How do we plot the band structure? For the 1D crystal, e.g., a periodic polymer, there is no problem: the wave vector $\boldsymbol{k}$ means the number k and changes from $-\frac{\pi}{a}$ to $\frac{\pi}{a}$, we plot the function $\varepsilon_n(k)$. For each n we have a single plot, e.g., for the hydrogen atom the band ε_1 collects energies resulting from the 1s atomic orbital interacting with other atoms, the band ε_2, which resulted from 2s, etc. In the 3D case we usually choose a path in FBZ. We start from the point Γ defined as $\boldsymbol{k} = \boldsymbol{0}$. Then, we continue to some points located on the faces and edges of the FBZ surface. It is impossible to go through the whole FBZ. The band structure in the 3D case is usually shown by putting the described itinerary through the FBZ on the abscissa (Fig. 9.8), and $\varepsilon_n(\boldsymbol{k})$ on the ordinate. Fig. 9.8 shows an example of what we might obtain from such calculations.

9.9.5 FERMI LEVEL AND ENERGY GAP: INSULATORS, SEMICONDUCTORS AND METALS

First of all we have to know how many electrons we have in the crystal. The answer is simple: the infinite crystal contains an infinite number of electrons. But infinities are often different. The decider is the number of electrons per unit cell. Let us denote this number by n_0.

If this means a double occupation of the molecular orbitals of the unit cell, then the corresponding band in the crystal will also be fully occupied, because the number of energy levels in a band is equal to the number of unit cells, and each unit cell contributes two electrons from the above mentioned molecular orbital. Therefore,

conduction band

doubly occupied orbitals lead to fully occupied bands. Accordingly, singly occupied orbitals lead to bands that are *half-occupied*, while empty (virtual) orbitals lead to empty bands (unoccupied, or conduction bands).

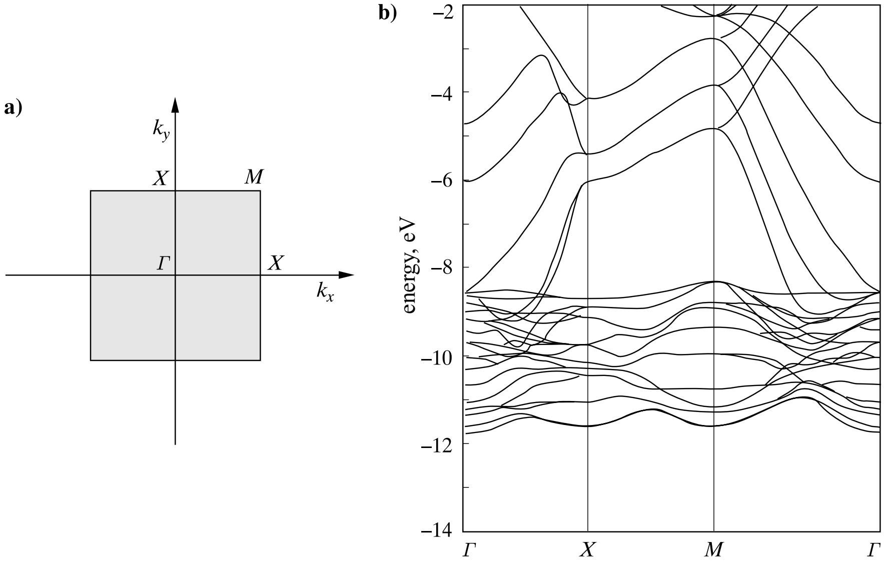


Fig. 9.8. (a) FBZ for four regular layers of nickel atoms (a crystal surface model) (b) the band structure for this system. We see that we cannot understand much: just a horrible irregular mess of lines. All the band structures look equally clumsy. Despite this, from such a plot we may determine the electrical and optical properties of the nickel slab. We will see later on why the bands have such a mysterious form. R. Hoffmann, "*Solids and Surfaces. A Chemist's View of Bonding in Extended Structures*", VCH Publishers, New York, © 1988 VCH Publishers. Reprinted with permission of John Wiley & Sons, Inc.

The Fermi level in the band structure of a crystal is equivalent to the HOMO of the crystal.[33] The two levels HOMO and LUMO, as always, decide the chemistry of the system, in our case the crystal. This concept leads to the possibilities depicted in Fig. 9.9, which we may find in ideal crystals

metal

A metal is characterized by empty levels (conduction band) immediately (zero distance) above doubly occupied valence ones (highest occupied by electrons).

Metals are conductors of electric current, and the reason for this is the zero gap.[34]

valence bands

A semiconductor exhibits a conduction band separated by a small energy gap (band gap) from the valence band.

---

[33] We sometimes find a thermodynamic definition of the Fermi level, but in this book it will always be the energy of the highest occupied crystal orbital.

[34] When an electric field is applied to a crystal, its energy levels change. If the field is weak then the changes may be computed by perturbation theory (treating the zero–field situation as the unperturbed one). This means that the perturbed states acquire some admixtures of the excited states (cf. Chapter 5). The lower the energy gap, the more mixing is taking place. For metallic systems (gap zero), such perturbation theory certainly would not be applicable, but real excitation to the conduction band may take place.

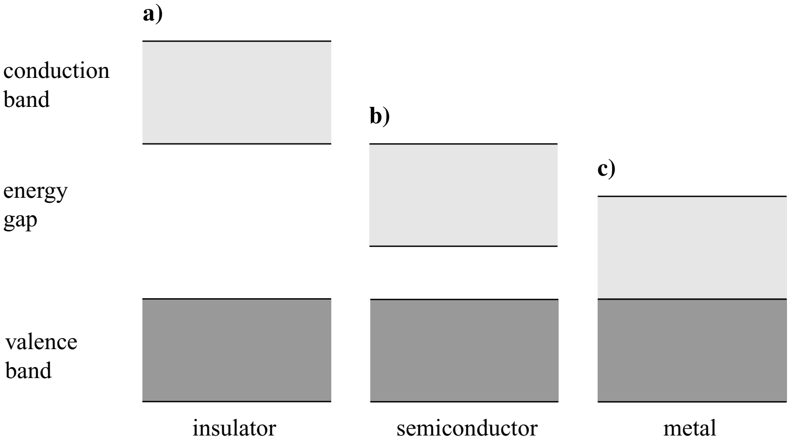


Fig. 9.9. Valence bands (highest occupied by electrons) and conduction bands (empty). The electric properties of a crystal depend on the energy gap between them (i.e. HOMO–LUMO separation). A large gap (a) is typical for an insulator, a medium gap (b) means a semiconductor, and a zero gap (c) is typical of metals.

Finally,

insulator

> an insulator has a large band gap separating the valence band from the conduction band.

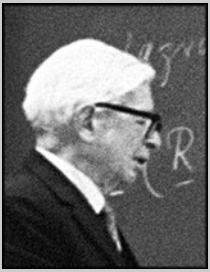
band gap

molecular wires

Peierls transition

We know metallic systems typically represent microscopically 3D objects. Recently, 2D and 1D metals have become more and more fashionable, the latter called *molecular wires*. They may have unusual properties, but are difficult to prepare for they often undergo spontaneous dimerization of the lattice (known as the *Peierls transition*).

Rudolph Peierls (1907–1995), British physicist, professor at the universities of Birmingham and Oxford. Peierls participated in the Manhattan Project (atomic bomb) as leader of the British group.



As Fig. 9.10.a shows, dimerization makes the bonding (and antibonding) effects stronger a little below (and above) the middle of the band, whereas at $k = 0$ the effect is almost zero (since dimerization makes the bonding or antibonding effects cancel within a pair of consecutive bonds). As a result, the degeneracy is removed in the middle of the band (Fig. 9.10.b), i.e. the band gap appears and the system undergoes metal–insulator or metal–semiconductor transition (Fig. 9.10.c). This is why polyacetylene, instead of having all the CC bonds equivalent (Fig. 9.10.d), which would make it a metal, exhibits alternation of bond lengths (Fig. 9.10.e) and it becomes an insulator or semiconductor.

To a chemist, the Peierls transition is natural. The hydrogen atoms will not stay equidistant in a chain, but will simply react and form hydrogen molecules, i.e. will

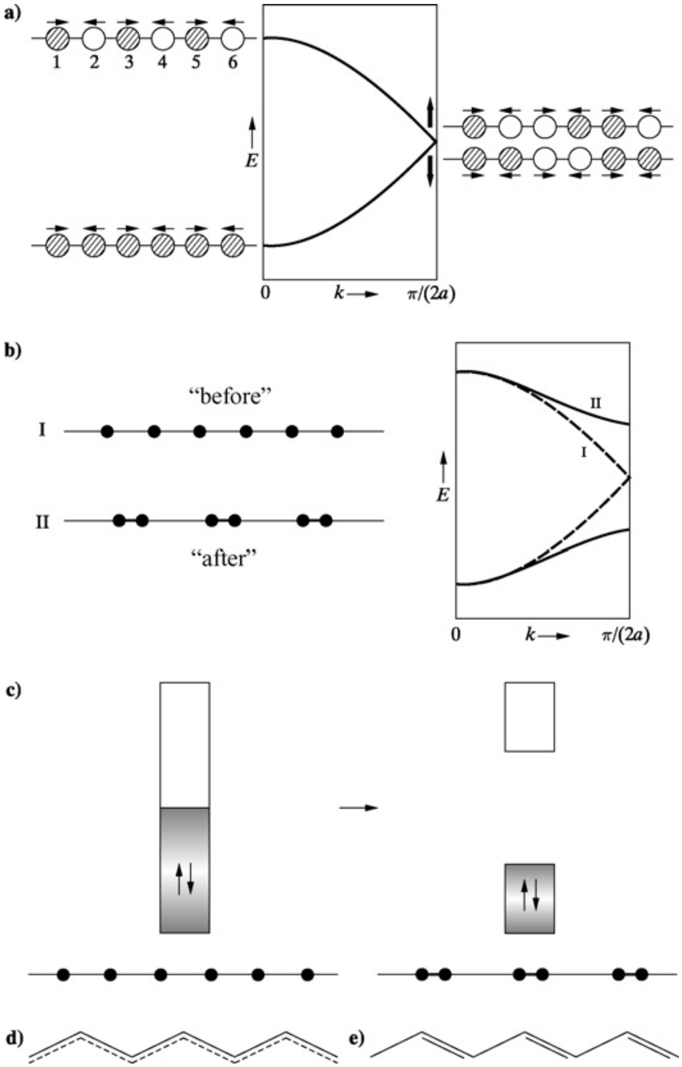


Fig. 9.10. The Peierls effect has the same origin as the Jahn–Teller effect in removing the electronic level degeneracy by distorting the system (H.A. Jahn, E. Teller, *Proc. Roy. Soc.* A161 (1937) 220). (a) The electrons occupy half the FBZ, i.e. $-\frac{\pi}{2a} \leqslant k \leqslant \frac{\pi}{2a}$, a standing for the nearest-neighbour distance. The band has been plotted assuming that the period is equal to $2a$, hence a characteristic back folding of the band (similarly as we would fold a sheet of paper with band structure drawn, the period equal a). A lattice dimerization amplifies the bonding and antibonding effects close to the middle of the FBZ, i.e. in the neighbourhood of $k = \pm\frac{\pi}{2a}$. (b) As a result, the degeneracy at $k = \frac{\pi}{2a}$ is removed and the band gap appears, which corresponds to lattice dimerization. (c) The system lowers its energy when undergoing metal–insulator or metal–semiconductor transition. (d) The polyacetylene chain, forcing equivalence of all CC bonds, represents a metal. However, due to the Peierls effect, the system undergoes dimerization (e) and becomes an insulator. R. Hoffmann, "*Solids and Surfaces. A Chemist's View of Bonding in Extended Structures*", VCH Publishers, New York, © 1988 VCH Publishers. Reprinted with permission of John Wiley & Sons, Inc.

dimerize like lightning. Also the polyacetylene will try to form π bonds by binding the carbon atoms in *pairs*. There is simply a shortage of electrons to keep *all* the CC bonds strong, *there are only enough for only every second*, which means simply dimerization through creating π bonds. On the other hand, the Peierls transition may be seen as the Jahn–Teller effect: there is a degeneracy of the occupied and empty levels at the Fermi level, and it is therefore possible to lower the energy by removing the degeneracy through a distortion of geometry (i.e. dimerization). Both pictures are correct and represent the thing.

intrinsic semiconductor

When a semiconductor is heated, this may cause a non-zero electron population in the conduction band (according to Boltzmann's law), and these electrons may contribute to electric conductance, as for metals. The higher the temperature, the larger the conductance of such a semiconductor (called an *intrinsic semiconductor*).

The electric field will not do great things there (apart from some polarization).

Small energy gaps may appear when we dope an insulator with some dopants offering their own energy levels within the energy gap (Fig. 9.11).

p- and n-type semiconductors

If the empty energy levels of the dopant are located just over the occupied band, the dopant may serve as an electron acceptor for the electrons from the occupied band (thus introducing its own conduction band), we have a *p-type semiconductor*. If the dopant energy levels are occupied and located just under the conduction band, the dopant may serve as a *n-type semiconductor*.

Polyacetylene (mentioned at the beginning of this chapter), after doping becomes ionized if the dopants are electron acceptors, or receives extra electrons if the dopant represents an electron donor (symbolized by D^+ in Fig. 9.12). The perfect polyacetylene exhibits the bond alternation discussed above, but it may

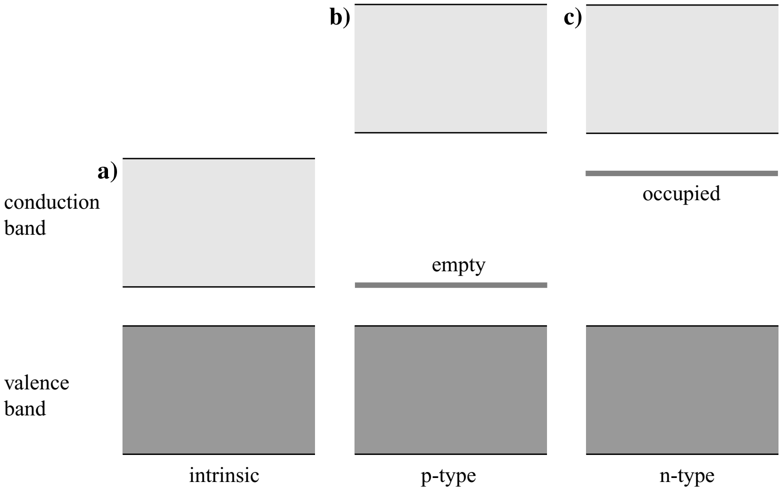


Fig. 9.11. Energy bands for semiconductors. (a) intrinsic semiconductor (small gap), (b) p type semiconductor (electron acceptor levels close to the occupied band) (c) n type semiconductor (electron donor levels close to the conduction band).

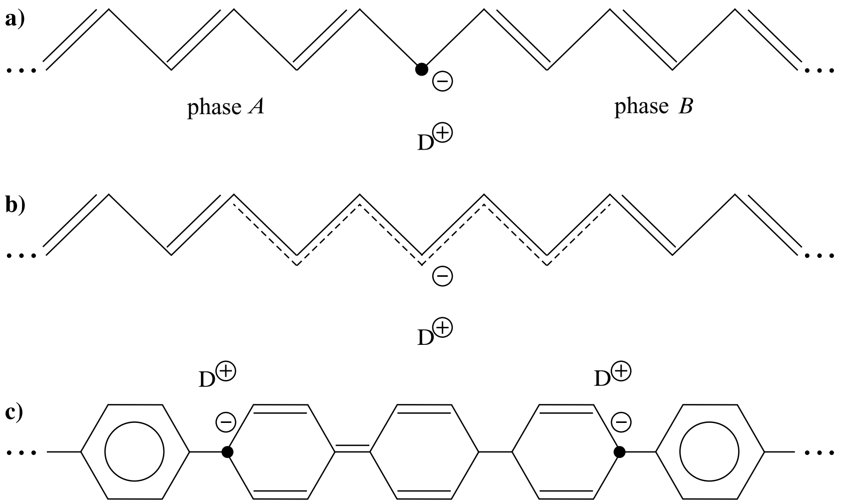


Fig. 9.12. Solitons and bipolarons as models of electric conductivity in polymers. (a) two phases of polyacetylene separated by a defect. Originally the defect was associated with an unpaired electron, but when a donor, D, gave its electron to the chain, the defect became negatively charged. The energy of such a defect is independent of its position in the chain (important for charge transportation) (b) in reality the change of phase takes place in sections of about 15 CC bonds, not two bonds as Fig.a suggests. Such a situation is sometimes modelled by a non-linear differential equation, which describes a soliton motion ("solitary wave") that represents the travelling phase boundary. (c) in the polyparaphenylene chain two phases (low-energy aromatic and high-energy quinoid) are possible as well, but in this case they are of different energies. Therefore, the energy of a single defect (aromatic structures-kink-quinoid structures) depends on its position in the chain (therefore, no charge transportation). However, a *double* defect with a (higher-energy) section of a quinoid structure has a position-independent energy, and when charged by dopants (*bipolaron*) can conduct electricity. The above mentioned polymers can be doped either by electron donors (e.g., arsenium, potassium) or electron acceptors (iodine), which results in a spectacular increase in their electric conductivity.

be that we have a defect that is associated with a region of "changing rhythm" (or "phase"): from[35] (= − = − =) to (− = − = −). Such a kink is sometimes described as a *soliton* wave (Fig. 9.12.a,b), i.e. a "solitary" wave first observed in the XIX century in England on a water channel, where it preserved its shape while moving over a distance of several kilometres. The soliton defects cause some new energy levels ("solitonic levels") to appear within the gap. These levels too form their own solitonic band.

soliton

Charged solitons may travel when subject to an electric field, and therefore the doped polyacetylene turns out to be a good conductor (organic metal).

In polyparaphenylene, soliton waves are impossible, because the two phases (aromatic and quinoid, Fig. 9.12.c) differ in energy (low-energy aromatic phase and high-energy quinoid phase). However, when the polymer is doped, a charged double defect (*bipolaron*, Fig. 9.12.c) may form, and the defect may travel when

bipolaron

---

[35] This possibility was first recognized by J.A. Pople, S.H. Walmsley, *Mol. Phys.* 5 (1962) 15, fifteen years before the experimental discovery of this effect.

an electric field is applied. Hence, the doped polyparaphenylene, similarly to the doped polyacetylene, is an "organic metal".

9.10 SOLID STATE QUANTUM CHEMISTRY

A calculated band structure, with information about the position of the Fermi level, tell us a lot about the electric properties of the material under study (insulator, semiconductor, metal). They tell us also about basic optical properties, e.g., the band gap indicates what kind of absorption spectrum we may expect. We can calculate any measurable quantity, because we have at our disposal the computed wave function.

However, despite this very precious information, which is present in the band structure, there is a little worry. When we stare at any band structure, such as that shown in Fig. 9.8, the overwhelming feeling is a kind of despair. All band structures look similar, well, just a tangle of plots. Some go up, some down, some stay unchanged, some, it seems without any reason, change their direction. Can we understand this? What is the theory behind this band behaviour?

9.10.1 WHY DO SOME BANDS GO UP?

Let us take our beloved chain of hydrogen atoms in the $1s$ state, to which we already owe so much (Fig. 9.13).

When will the state of the chain have the lowest energy possible? Of course, when all the atoms interact in a bonding, and not antibonding, way. This corresponds to Fig. 9.13.a (no nodes of the wave function). When, in this situation,

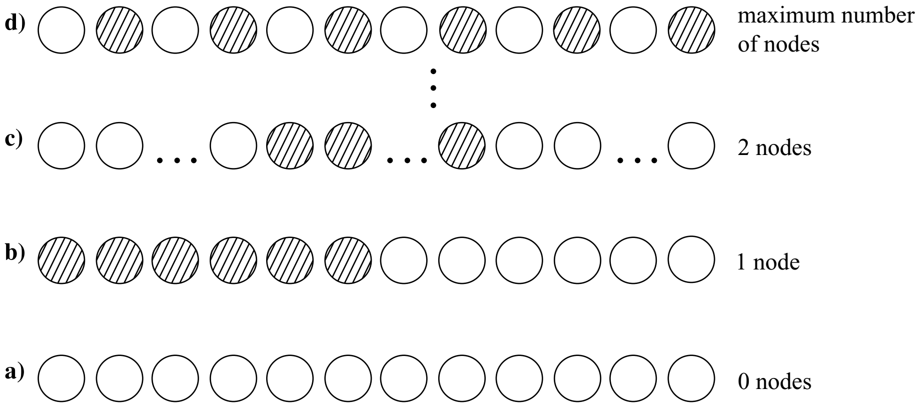


Fig. 9.13. The infinite chain of ground-state hydrogen atoms and the influence of bonding and antibonding effects, p. 371. a) all interactions are bonding; b) introduction of a single node results in an energy increase; c) two nodes increase the energy even more; d) maximum number of nodes – the energy is the highest possible.

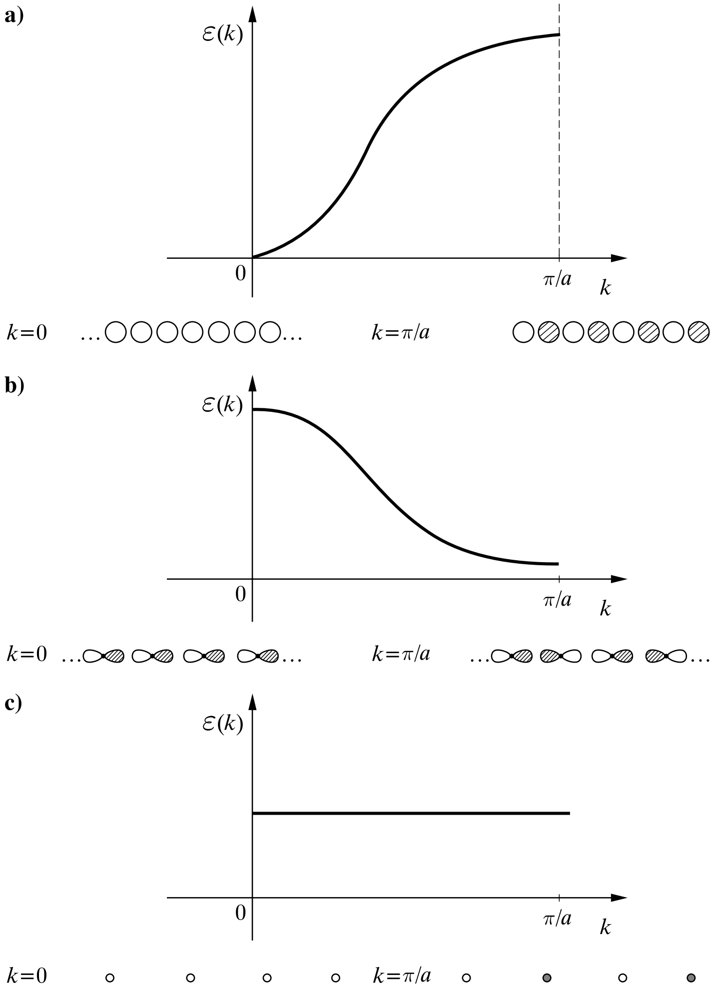


Fig. 9.14. Three typical band plots in the FBZ; a) 1*s* orbitals. Increasing *k* is accompanied by an *increase* of the antibonding interactions and this is why the energy goes up; b) $2p_z$ orbitals (z denotes the periodicity axis). Increasing *k* results in *decreasing* the number of antibonding interactions and the energy goes down; c) inner shell orbitals. The overlap is small as it is, therefore, the band width is practically zero.

we introduce a single nearest-neighbour antibonding interaction, the energy will for sure increase a bit (Fig. 9.13.b). When two such interactions are introduced (Fig. 9.13.c), the energy goes up even more, and the plot corresponds to two nodes. Finally, the highest-energy situation: all nearest-neighbour interactions are antibonding (maximum number of nodes), Fig. 9.13.d. Let us recall that the wave vector was associated with the number of nodes. Hence, if k increases from zero to $\frac{\pi}{a}$, the energy increases from the energy corresponding to the nodeless wave function to the energy characteristic for the maximum-node wave function. We understand, therefore, that some band plots are such as in Fig. 9.14.a.

9.10.2 WHY DO SOME BANDS GO DOWN?

Sometimes the bands go in the opposite direction: the lowest energy corresponds to $k = \frac{\pi}{a}$, the highest energy to $k = 0$. What happens over there? Let us once more take the hydrogen atom chain, this time, however, in the $2p_z$ state (z is the periodicity axis). This time the Bloch function corresponding to $k = 0$, i.e. a function that follows just from locating the orbitals $2p_z$ side by side, describes the highest-energy interaction – the *nearest-neighbour interactions are all antibonding*. Introduction of a node (increasing k) means a relief for the system – instead of one painful antibonding interaction we get a soothing bonding one. The energy goes down. No wonder, therefore, some bands look like those shown in Fig. 9.14.b.

9.10.3 WHY DO SOME BANDS STAY CONSTANT?

According to numerical rules (p. 362) inner shell atomic orbitals do not form effective linear combinations (crystal orbitals). Such orbitals have very large exponential coefficients and the resulting overlap integral, and therefore the band width (bonding vs antibonding effect), is negligible. This is why the nickel $1s$ orbitals (deep-energy level) result in a low-energy band of almost zero width (Fig. 9.14.c), i.e. staying flat as a pancake all the time. Since they are always of very low energy, they are doubly occupied and their plot is so boring, they are not even displayed (as in Fig. 9.8).

9.10.4 HOW CAN MORE COMPLEX BEHAVIOUR BE EXPLAINED?

We understand, therefore, at least why some bands are monotonically going down, some up, some stay constant. In explaining these cases, we have assumed that a given CO is dominated by a single Bloch function. Other behaviours can be explained as well by detecting what kind of Bloch function *combination* we have in a given crystal orbital.

2D regular lattice of the hydrogen atoms

Let us take a planar regular lattice of hydrogen atoms in their ground state.[36] Fig. 9.8 shows the FBZ of similar lattice, we (arbitrarily) choose as the itinerary through the FBZ: $\Gamma - X - M - \Gamma$. From Fig. 9.6.a we easily deduce, that the band energy for the point Γ has to be the lowest, because it corresponds to all the interaction bonding. What will happen at the point X? This situation is related to Fig. 9.6.b. If we focus on any of the hydrogen atoms, it has four nearest neighbour interactions: two bonding and two antibonding. This corresponds, to good approximation, to the non-bonding situation (hydrogen atom ground-state energy), because the two effects nearly cancel. Halfway between Γ and X, we go through the point that corresponds to Fig. 9.6.c,d. For such a point, any hydrogen atom has two bonding and two non-bonding interactions, i.e. the energy is the average of the Γ

---

[36] A chemist's first thought would be that this could never stay like this, when the system is isolated. We are bound to observe the formation of hydrogen molecules.

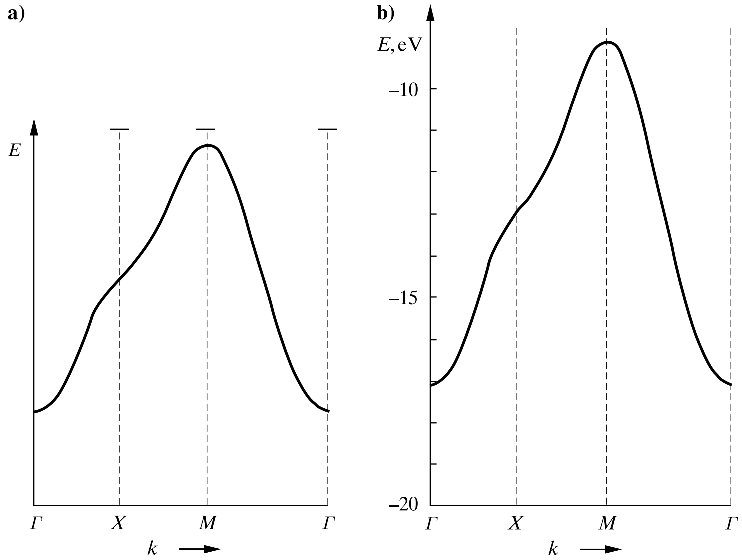


Fig. 9.15. a) A sketch of the valence band for a regular planar lattice of ground-state hydrogen atoms and b) the valence band, as computed in the laboratory of Roald Hoffmann, for nearest neighbour distance equal to 2 Å. The similarity of the two plots confirms that we are able, at least in some cases, to predict band structure. R. Hoffmann, "*Solids and Surfaces. A Chemist's View of Bonding in Extended Structures*", VCH Publishers, New York, © 1988 VCH Publishers. Reprinted with permission of John Wiley & Sons, Inc.

and X energies. The point M is located in the corner of the FBZ, and corresponds to Fig. 9.6.e. All the nearest-neighbour interactions are antibonding there, and the energy will be very high. We may, therefore, anticipate a band structure of the kind sketched in Fig. 9.15.a. The figure has been plotted to reflect the fact that the density of states for the band edges is the largest, and therefore the slope of the curves has to reflect this. Fig. 9.15 shows the results of the computations.[37] It is seen that, even very simple reasoning may rationalize the main features of band structure plots.

Trans-polyacetylene (regular 1D polymer)

Polyacetylene already has quite a complex band structure, but as usual the bands close to the Fermi level (valence bands) are the most important in chemistry and physics. All these bands are of the π type, i.e. their COs are antisymmetric with respect to the plane of the polymer. Fig. 9.16 shows how the valence bands are formed. We can see, the principle is identical to that for the chain of the hydrogen atoms: the more nodes the higher the energy. The highest energy corresponds to the band edge.

The resulting band is only *half-filled* (metallic regime), because each of the carbon atoms offers one electron, and the number of COs is equal to the number

---

[37] R. Hoffmann, "*Solids and Surfaces. A Chemist's View of Bonding in Extended Structures*", VCH Publishers, New York, 1988.

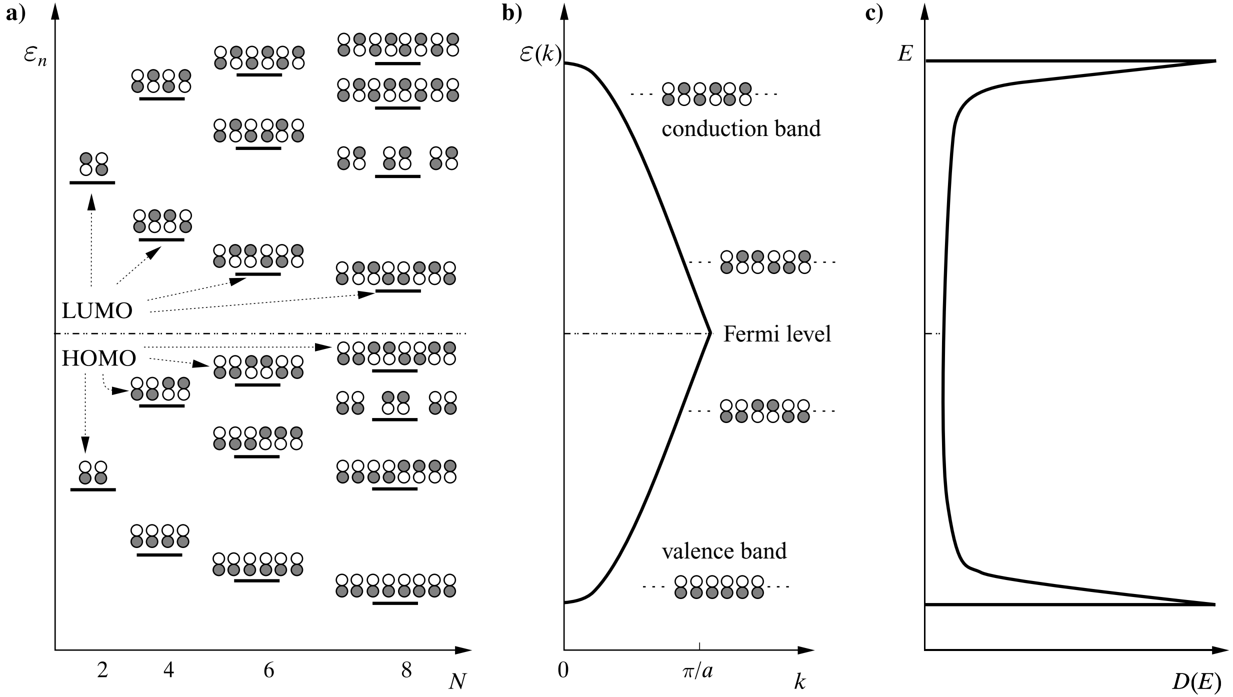


Fig. 9.16. a) π-band formation in polyenes (N stands for the number of carbon atoms) with the assumption of CC bond equivalence (each has length $a/2$). For $N = \infty$ this gives the metallic solution (no Peierls effect). As we can see, the band formation principle is identical to that, which we have seen for hydrogen atoms. b) band structure; c) density of states $D(E)$, i.e. the number of states per energy unit at a given energy E. The density has maxima at the extremal points of the band. If we allowed the Peierls transition, at $k = \pm\pi/a$ we would have a gap. J.-M. André, J. Delhalle, J.-L. Brédas, "*Quantum Chemistry Aided Design of Organic Polymers*", World Scientific, Singapore, 1991. Reprinted with permission from the World Scientific Publishing Co. Courtesy of the authors.

of carbon atoms (each CO can accommodate two electrons). Therefore, Peierls mechanism (Fig. 9.10) is bound to enter into play, and in the middle of the band a gap will open. The system is, therefore, predicted to be an insulator (or semiconductor) and indeed it is. It may change to a metal when doped. Fig. 9.16 shows a situation analogous to the case of a chain of the ground state hydrogen atoms.

Polyparaphenylene

The extent to which the COs conform to the rule of increasing number of nodes with energy (or k) will be seen in the example of a planar conformation of polyparaphenylene.[38] On the left-hand side of Fig. 9.17 we have the valence π-orbitals of benzene:

- the lowest-energy has a nodeless[39] doubly occupied molecular orbital φ_1,
- then, we have a doubly degenerate and fully occupied level with the corresponding orbitals, φ_2 and φ_3, each having a single node,
- next, a similar double degenerate empty level with orbitals φ_4 and φ_5 (each with two nodes),
- and finally, the highest-energy empty three-node orbital φ_6.

Thus, even in the single monomer we have fulfilled the rule.

Binding phenyl rings by using CC σ bonds results in polyparaphenylene. Let us see what happens when the wave number k increases (the middle and the right-hand side of Fig. 9.17). What counts now is how two *complete monomer orbitals* combine: in-phase or out-of-phase. The lowest-energy π-orbitals of benzene (φ_1) arranged in-phase ($k = 0$) give point Γ – the lowest-energy in the polymer, while out-of-phase, point $k = \frac{\pi}{a}$ – the highest-energy. At $k = \frac{\pi}{a}$ there is a degeneracy of this orbital and of φ_3 arranged out-of-phase. The degeneracy is quite interesting because, despite a superposition of the orbitals with the different number of nodes, the result, for obvious reasons, corresponds to the same number of nodes. Note the extremely small dispersion of the band which results from the arrangement of φ_2. The figure shows that it is bound to be small, because it is caused by the arrangement of two molecular orbitals that are further away in space than those so far considered (the overlap results from the overlap of the atomic orbitals separated by three bonds, and not by a single bond as it has been). We see a similar regularity in the conduction bands that correspond to the molecular orbitals φ_4, φ_5 and φ_6. The rule works here without any exception and results from the simple statement that a bonding superposition has a lower energy than the corresponding antibonding one.

Thus, when looking at the band structure for polyparaphenylene we stay cool: we understand every detail of this tangle of bands.

---

[38] J.-M. André, J. Delhalle, J.-L. Brédas, "*Quantum Chemistry Aided Design of Organic Polymers*", World Scientific, Singapore, 1991.

[39] Besides the nodal plane of the nuclear framework.

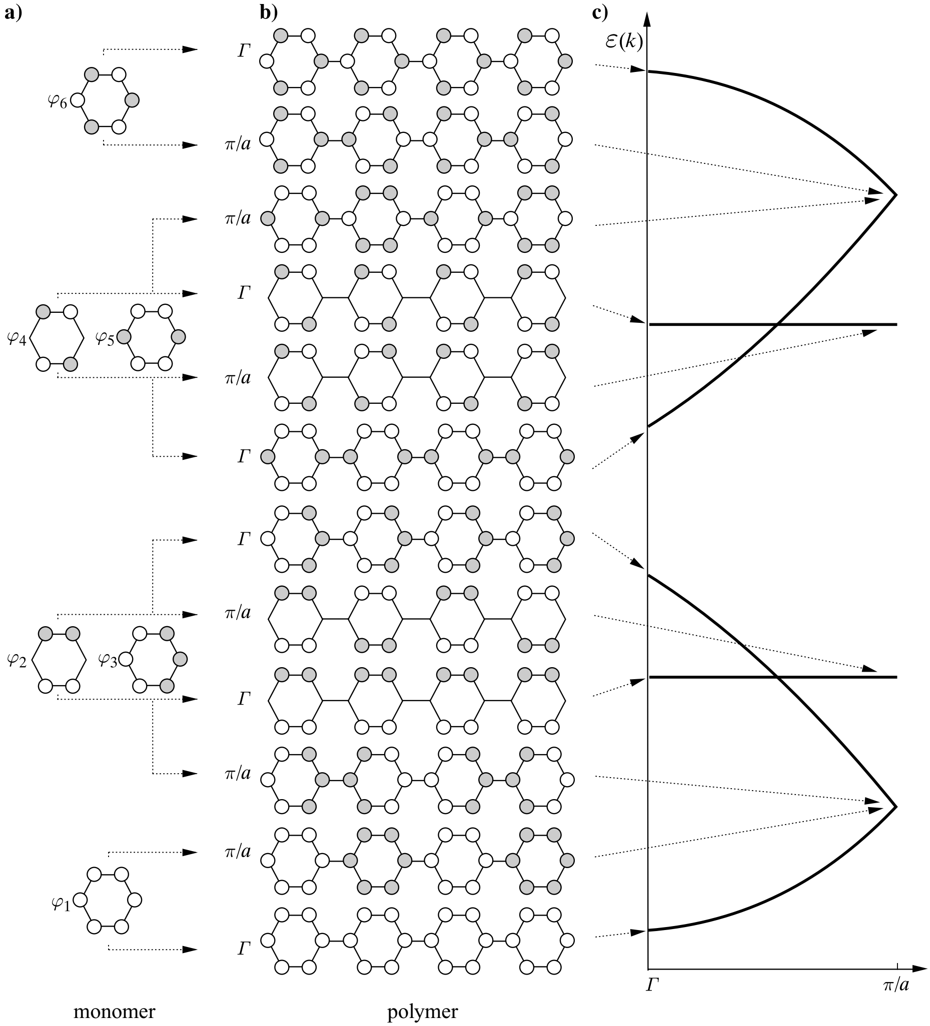


Fig. 9.17. Rationalizing the band structure of polyparaphenylene (π-bands). The COs (in centre) built as in-phase or out-of-phase combinations of the benzene π molecular orbitals (left-hand side). It is seen that energy of the COs for $k = 0$ and $k = \frac{\pi}{a}$ agree with the rule of increasing number of nodes. A small band width corresponds to small overlap integrals of the monomer orbitals. J.-M. André, J. Delhalle, J.-L. Brédas, "*Quantum Chemistry Aided Design of Organic Polymers*", World Scientific, Singapore, 1991. Reprinted with permission from the World Scientific Publishing Co. Courtesy of the authors.

A stack of Pt(II) square planar complexes

Let us try to predict[40] qualitatively (without making calculations) the band structure of a stack of platinum square planar complexes, typically $[\mathrm{Pt(CN^-)_4^{2-}}]_\infty$. Con-

---

[40] R. Hoffmann, "*Solids and Surfaces. A Chemist's View of Bonding in Extended Structures*", VCH Publishers, New York, 1988.

sider the eclipsed configuration of all the monomeric units. Let us first simplify our task. Who likes cyanides? Let us throw them away and take something theoreticians really love: H^-. This is a little less than just laziness. If needed, we are able to make calculations for cyanides too, but to demonstrate that we really understand the machinery we are always recommended to make the system as simple as possible (but not simpler). We suspect that the main role of CN^- is just to interact electrostatically, and H^- does this too (being much smaller).

The electronic dominant configuration of the platinum atom in its ground state is[41] $(Xe)(4f^{14})5d^96s^1$ (see the Mendeleev Table in the Web Annex). As we can see, we have the xenon-like closed shell and also the full closed subshell $4f$. The orbital energies corresponding to these closed shells are much lower than the orbital energy of the hydrogen anion (they are to be combined to). This is why they will not participate in the Pt–H bonds. They will of course contribute to the band structure, but this contribution will be trivial: flat bands (because of small overlap integrals) with energies very close to the energies characterizing the corresponding atomic orbitals. The Pt valence shell is therefore $5d^96s^16p^0$ for Pt^0, and $5d^86s^06p^0$ for Pt^{2+}, which we have in our stack. The corresponding orbital energies are shown on the left-hand side of Fig. 9.18.a.

Let us choose a Cartesian coordinate system with the origin on the platinum atom and the four ligands at equal distances on the x and y axes. In the Koopmans

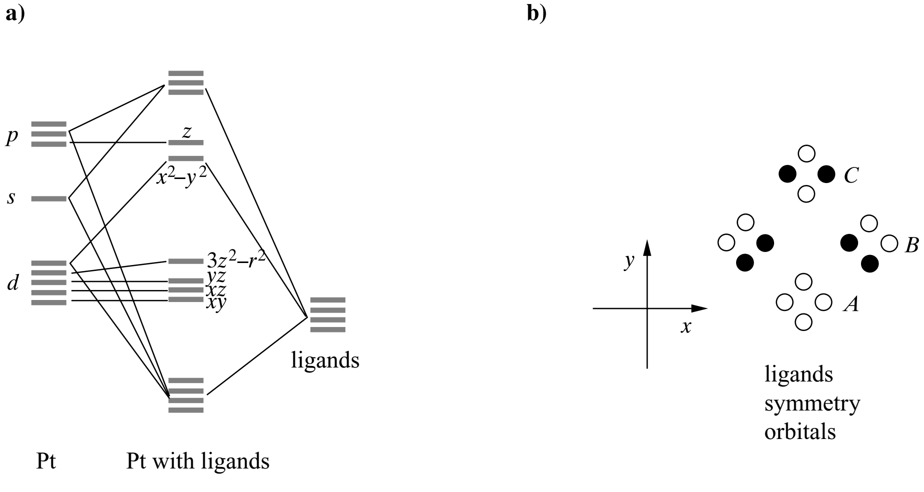


Fig. 9.18. Predicting the band structure of $(PtH_4^{2-})_\infty$. (a) monomer (PtH_4^{2-}) molecular orbitals built of the atomic orbitals of Pt^{2+} (the three p and five d Pt atomic orbitals correspond to two degenerate energy p and d levels) and four ligand (H^-) orbitals. One of the platinum orbitals $(5d_{x^2-y^2})$ corresponds to high energy, because it protrudes right across to the negatively charged ligands. The four ligand AOs, due to the long distance practically do not overlap, and are shown as a quadruply degenerate level. (b) The ligand orbitals form linear combinations with those of the metal. See the text.

---

[41] Xe denotes the xenon-like configuration of electrons.

approximation (cf. Chapter 8, p. 393) an orbital energy represents the electron energy on a given orbital. We see, that because the ligands are negatively charged, all the platinum atom orbital energies will go up (destabilization; in Fig. 9.18.a this shift is not shown, only a relative shift is given). The largest shift up will be undergone by the $5d_{x^2-y^2}$ orbital energy, because the orbital lobes protrude right across to the negative ligands. Eight electrons of Pt^{2+} will therefore occupy four other orbitals[42] $(5d_{xy}, 5d_{xz}, 5d_{yz}, 5d_{3z^2-r^2})$, while $5d_{x^2-y^2}$ will become LUMO. The four ligand atomic orbitals practically do not overlap (long distance) and this is why in Fig. 9.18.a they are depicted as a quadruply degenerate level. The ligand symmetry orbitals are shown in Fig. 9.18.b: the nodeless orbital (A), and two single-node orbitals (B) corresponding to the same energy, and the two-node orbital (C). The effective linear combinations (cf. p. 362, what counts most is symmetry) are formed by the following pairs of orbitals: $6s$ with A, $6p_x$ and $6p_y$ with B, and the orbital $5d_{x^2-y^2}$ with C (in each case we obtain the bonding and the antibonding orbital); the other platinum orbitals, $5d$ and $6p_z$ do not have partners of the appropriate symmetry (and therefore their energy does not change). Thus we obtain the energy level diagram of the monomer in Fig. 9.18.a.

Now, we form a stack of PtH_4^{2-} along the periodicity axis z. Let us form the Bloch functions (Fig. 9.19.a) for each of the valence orbitals at two points of the FBZ: $k = 0$ and $k = \frac{\pi}{a}$. The results are given in Fig. 9.19.b. Because of the large overlap of the $6p_z$ orbitals with themselves, and $3\mathrm{d}_{3z^2-r^2}$ also with themselves, these σ bands will have very large dispersions. The smallest dispersion will correspond to the $5d_{xy}$ band (as well as to the empty band $5d_{x^2-y^2}$), because the orbital lobes of $5d_{xy}$ (also of $5d_{x^2-y^2}$) are oriented perpendicularly to the periodicity axis. Two bands $5d_{xz}$ and $5d_{yz}$ have a common fate (i.e. the same plot) due to the symmetry, and a medium band width (Fig. 9.19.b). We predict therefore,[40] the band structure shown in Fig. 9.20. The prediction turns out to be correct.

9.11 THE HARTREE–FOCK METHOD FOR CRYSTALS

9.11.1 SECULAR EQUATION

What has been said previously about the Hartree–Fock method is only a sort of general theory. The time has now arrived to show how the method works in practice. We have to solve the Hartree–Fock–Roothaan equation (cf. Chapter 8, pp. 365 and 453).

---

[42] Of these four the lowest-energy will correspond to the orbitals $5d_{xz}, 5d_{yz}$, because their lobes just avoid the ligands. The last two orbitals $5d_{xy}$ and $5d_{3z^2-r^2} = 5d_{z^2-x^2} + 5d_{z^2-y^2}$ will go up somewhat in the energy scale (each to different extent), because they aim in part at the ligands. However, these splits will be smaller when compared to the fate of the orbital $5d_{x^2-y^2}$ and therefore, these levels have been shown in the figure as a single degenerate level.

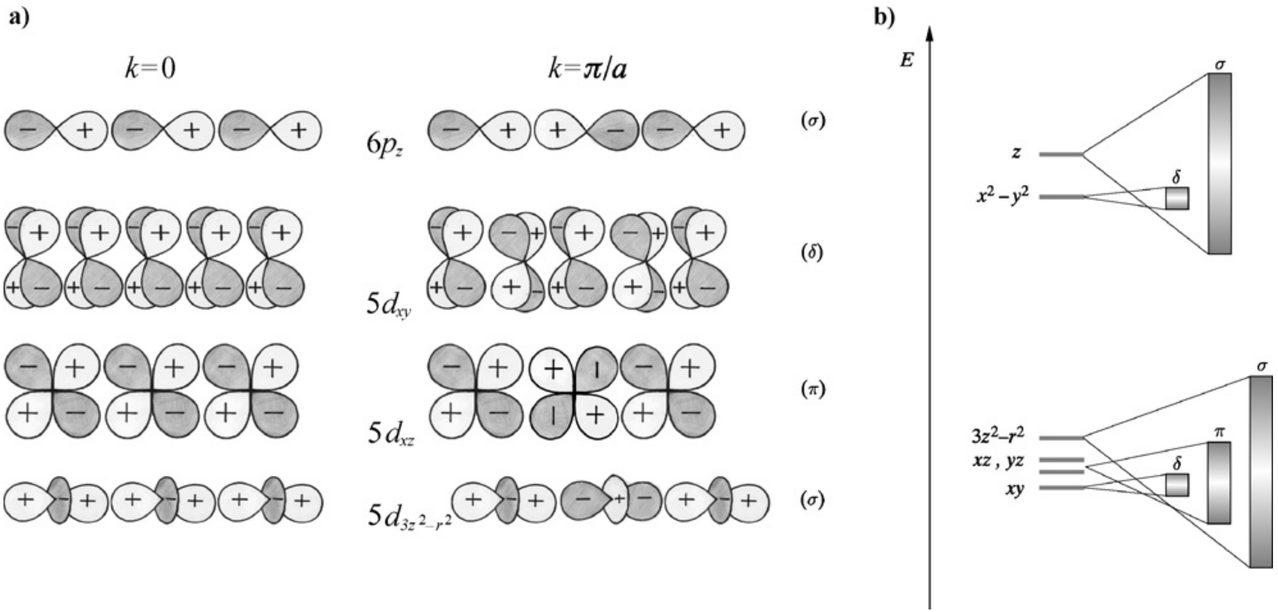


Fig. 9.19. Predicting the band structure of $(\mathrm{PtH}_4^{2-})_\infty$. (a) the Bloch functions for $k = 0$ and $k = \frac{\pi}{a}$ corresponding to the atomic orbitals $6p_z$ (σ type orbitals), $5d_{xy}$ (δ type orbitals), $5d_{xz}$ (π type orbitals, similarly for $5d_{yz}$), $5d_{3z^2-r^2}$ (σ type orbitals); (b) the band width is very sensitive to the overlap of the atomic orbitals. The band widths in $(\mathrm{PtH}_4^{2-})_\infty$ resulting from the overlap of the (PtH_4^{2-}) orbitals.

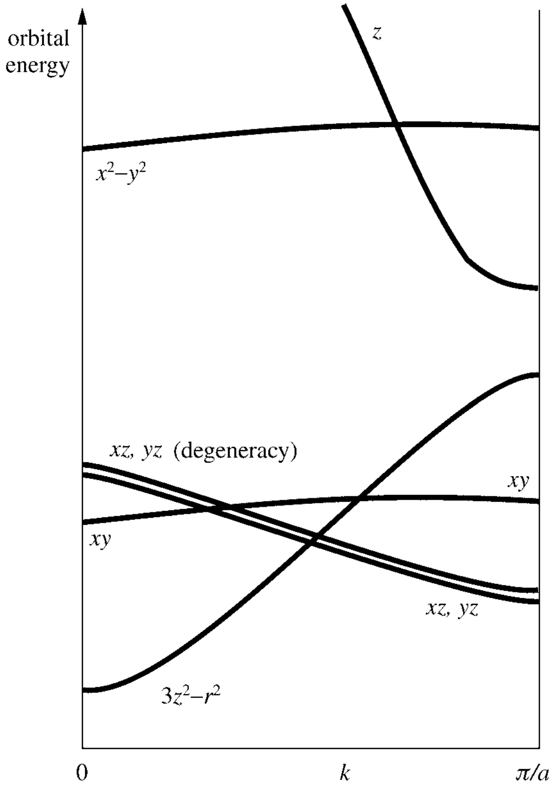


Fig. 9.20. The predicted band structure of $(\mathrm{PtH}_4^{2-})_\infty$.

The Fock matrix element is equal to (noting that $(\chi_p^j|\hat{F}\chi_q^{j'}) \equiv F_{pq}^{jj'}$ depends on the *difference*[43] between the vectors $\boldsymbol{R}_j$ and $\boldsymbol{R}_{j'}$):

$$F_{pq} = (2N+1)^{-3}\sum_{jj'} \exp\bigl(i\boldsymbol{k}(\boldsymbol{R}'_j - \boldsymbol{R}_j)\bigr)\bigl(\chi_p^j|\hat{F}\chi_q^{j'}\bigr) = \sum_j \exp(i\boldsymbol{k}\boldsymbol{R}_j)F_{pq}^{0j}.$$

The same can be done with S_{pq} and therefore the Hartree–Fock–Roothaan secular equation (see p. 453) has the form:

$$\sum_{q=1}^{\omega} c_{qn}(\boldsymbol{k})\biggl(\sum_j \exp(i\boldsymbol{k}\boldsymbol{R}_j)\bigl(F_{pq}^{0j}(\boldsymbol{k}) - \varepsilon_n(\boldsymbol{k})S_{pq}^{0j}(\boldsymbol{k})\bigr)\biggr) = 0, \tag{9.55}$$

for $p = 1, 2, \ldots, \omega$. The integral S_{pq} equals

$$S_{pq} = \sum_j \exp(i\boldsymbol{k}\boldsymbol{R}_j)S_{pq}^{0j}, \tag{9.56}$$

---

[43] As a matter of fact, all depends on how distant the unit cells j and j' are. We have used the fact that $\hat{F}$ exhibits the crystal symmetry and the sums over j all give the same result, independent of j'.

the summation $\sum_j$ goes over the lattice nodes. In order to be explicit, let us see what is inside the Fock matrix elements $F^{0j}_{pq}(\boldsymbol{k})$. *We have* to find a dependence there on the Hartree–Fock–Roothaan solutions (determined by the coefficients c_{pn}), and more precisely on the bond order matrix.[44] Any CO, according to (9.53), has the form

$$\psi_n(\boldsymbol{r},\boldsymbol{k}) = (2N+1)^{-\frac{3}{2}} \sum_q \sum_j c_{qn}(\boldsymbol{k}) \exp(i\boldsymbol{k}\boldsymbol{R}_j)\chi^j_q(\boldsymbol{r}), \tag{9.57}$$

where we promise to use such c_{qn} that ψ_n are normalized. For molecules the bond order matrix element (for the atomic orbitals χ_p and χ_q) has been defined (p. 365) as $P_{pq} = 2\sum c_{pi}c^*_{qi}$ (the summation is over the doubly occupied orbitals), where the factor 2 results from the double occupation of the closed shell. We have exactly the same for the crystal, where we define the bond order matrix element corresponding to atomic orbitals χ^j_q and χ^l_p as:

$$P^{lj}_{pq} = 2(2N+1)^{-3} \sum c_{pn}(\boldsymbol{k}) \exp(i\boldsymbol{k}\boldsymbol{R}_l) c_{qn}(\boldsymbol{k})^* \exp(-i\boldsymbol{k}\boldsymbol{R}_j), \tag{9.58}$$

where the summation goes over all the occupied COs (we assume double occupation, hence factor 2). This means that, in the summation we have to go over all the occupied bands (index n), and in each band over all allowed COs, i.e. all the allowed $\boldsymbol{k}$ vectors in the FBZ. Thus,

$$P^{lj}_{pq} = 2(2N+1)^{-3} \sum_n \sum_{\boldsymbol{k}}^{\text{FBZ}} c_{pn}(\boldsymbol{k}) c_{qn}(\boldsymbol{k})^* \exp\bigl(i\boldsymbol{k}(\boldsymbol{R}_l - \boldsymbol{R}_j)\bigr). \tag{9.59}$$

The matrix element has to have four indices (instead of the two indices in the molecular case), because we have to describe the atomic orbitals indicating that atomic orbital p is from unit cell l, and atomic orbital q from unit cell j. It is easily seen that P^{lj}_{pq} depends on the *difference* $\boldsymbol{R}_l - \boldsymbol{R}_j$, not on the $\boldsymbol{R}_l$, $\boldsymbol{R}_j$ themselves. The reason for this is that in a crystal everything is repeated and the important thing are the *relative* distances.

9.11.2 INTEGRATION IN THE FBZ

There is a problem with $\boldsymbol{P}$, because it requires a summation over $\boldsymbol{k}$. We do not like this, because the number of the permitted $\boldsymbol{k}$ is huge for large N (and N has to be large, because we are dealing with a crystal). We have to do something with it.

Let us try a small exercise. Imagine, we have to perform a summation $\sum_{\boldsymbol{k}} f(\boldsymbol{k})$, where f represents a smooth function in the FBZ. Let us denote the sum to be found by X. Let us multiply X by a small number $\Delta = \frac{V_{\text{FBZ}}}{(2N+1)^3}$, where V_{FBZ} stands

---

[44] We have met the same in the Hartree–Fock method for molecules, where the Coulomb and exchange operators depended on the solutions to the Fock equation, cf. p. 346.

for the FBZ volume:

$$X\Delta = \sum_{\boldsymbol{k}}^{\text{FBZ}} f(\boldsymbol{k})\Delta. \tag{9.60}$$

In other words we just cut the FBZ into tiny segments of volume Δ, their number equal to the number of the permitted $\boldsymbol{k}$'s. It is clear that if N is large (as it is in our case), then a very good approximation of $X\Delta$ would be

$$X\Delta = \int_{\text{FBZ}} f(\boldsymbol{k})\,\mathrm{d}^3\boldsymbol{k}. \tag{9.61}$$

Hence,

$$X = \frac{(2N+1)^3}{V_{\text{FBZ}}} \int_{\text{FBZ}} f(\boldsymbol{k})\,\mathrm{d}^3\boldsymbol{k}. \tag{9.62}$$

After applying this result to the bond order matrix we obtain

$$P_{pq}^{lj} = \frac{2}{V_{\text{FBZ}}} \int \sum_{n}^{\text{FBZ}} c_{pn}(\boldsymbol{k}) c_{qn}(\boldsymbol{k})^* \exp\bigl(i\boldsymbol{k}(\boldsymbol{R}_l - \boldsymbol{R}_j)\bigr)\,\mathrm{d}^3\boldsymbol{k}. \tag{9.63}$$

For a periodic *polymer* (in 1D: $V_{\text{FBZ}} = \frac{2\pi}{a}$, $\Delta = \frac{V}{2N+1}$) we would have:

$$P_{pq}^{lj} = \frac{a}{\pi} \int \sum_{n} c_{pn}(k) c_{qn}(k)^* \exp\bigl(ika(l-j)\bigr)\,\mathrm{d}k. \tag{9.64}$$

9.11.3 FOCK MATRIX ELEMENTS

In full analogy with the formula (8.53), we can express the Fock matrix elements by using the bond order matrix $\boldsymbol{P}$ for the crystal:

$$F_{pq}^{0j} = T_{pq}^{0j} - \sum_{h}\sum_{u} Z_u V_{pq}^{0j}\bigl(A_u^h\bigr) + \sum_{hl}\sum_{rs} P_{sr}^{lh}\Bigl[\bigl(\begin{smallmatrix}0h\\pr\end{smallmatrix}\big|\begin{smallmatrix}jl\\qs\end{smallmatrix}\bigr) - \frac{1}{2}\bigl(\begin{smallmatrix}0h\\pr\end{smallmatrix}\big|\begin{smallmatrix}lj\\sq\end{smallmatrix}\bigr)\Bigr], \tag{9.65}$$

this satisfies the normalization condition[45]

---

[45] The $\boldsymbol{P}$ matrix satisfies the normalization condition, which we obtain in the following way. As in the molecular case the normalization of CO's means:

$$\begin{aligned}
1 &= \bigl\langle \psi_n(\boldsymbol{r},\boldsymbol{k}) | \psi_n(\boldsymbol{r},\boldsymbol{k}) \bigr\rangle \\
&= (2N+1)^{-3} \sum_{pq}\sum_{jl} c_{pn}(\boldsymbol{k})^* c_{qn}(\boldsymbol{k}) \exp\bigl[i\boldsymbol{k}(\boldsymbol{R}_j - \boldsymbol{R}_l)\bigr] S_{pq}^{lj} \\
&= (2N+1)^{-3} \sum_{pq}\sum_{jl} c_{pn}(\boldsymbol{k})^* c_{qn}(\boldsymbol{k}) \exp\bigl[i\boldsymbol{k}(\boldsymbol{R}_j - \boldsymbol{R}_l)\bigr] S_{pq}^{0(j-l)} \\
&= \sum_{pq}\sum_{j} c_{pn}(\boldsymbol{k})^* c_{qn}(\boldsymbol{k}) \exp(i\boldsymbol{k}\boldsymbol{R}_j) S_{pq}^{0j}.
\end{aligned}$$

Now let us do the same for all the occupied COs and sum the results. On the left-hand side we sum just 1, therefore we obtain the number of doubly occupied COs, i.e. $n_0(2N+1)^3$, because n_0 denotes

$$\boxed{\sum_j \sum_{pq} P^{j0}_{qp} S^{0j}_{pq} = 2n_0, \qquad (9.68)}$$

where $2n_0$ means the number of electrons in the unit cell.

The first term on the right-hand side of (9.65) represents the kinetic energy matrix element

$$T^{0j}_{pq} = \left(\chi^0_p \middle| -\frac{1}{2}\Delta \middle| \chi^j_q\right), \tag{9.69}$$

the second term is a sum of matrix elements, each corresponding to the nuclear attraction of an electron and the nucleus of index u and charge Z_u in the unit cell h:

$$V^{0j}_{pq}(\boldsymbol{A}^h_u) = \left(\chi^0_p \middle| \frac{1}{|\boldsymbol{r} - \boldsymbol{A}^h_u|} \middle| \chi^j_q\right), \tag{9.70}$$

where the upper index of χ denotes the cell number, the lower index – the number of the atomic orbital in a cell, the vector $\boldsymbol{A}^h_u$ indicates nucleus u (numbering within the unit cell) in unit cell h (from the coordinate system origin). The third term is connected to the Coulombic operator (the first of two terms) and the exchange operator (the second of two terms). The summations over h and l go over the unit cells of the whole crystal, and therefore are very difficult and time consuming.

The definition of the two-electron integral

$$\left(\begin{smallmatrix}0h\\pr\end{smallmatrix}\middle|\begin{smallmatrix}jl\\qs\end{smallmatrix}\right) = \int \mathrm{d}^3\boldsymbol{r}_1\, \mathrm{d}^3\boldsymbol{r}_2\, \chi^0_p(\boldsymbol{r}_1)^* \chi^h_r(\boldsymbol{r}_2)^* \frac{1}{r_{12}} \chi^j_q(\boldsymbol{r}_1) \chi^l_s(\boldsymbol{r}_2) \tag{9.71}$$

is analogous to eq. (8.5) and Appendix M, p. 986.

---

the number of doubly occupied bands, and in each band we have in 3D $(2N+1)^3$ allowed vectors $\boldsymbol{k}$. Therefore, we have

$$\begin{aligned} n_0(2N+1)^3 &= \sum_{pq}\sum_j \left(\sum_n \sum_{\boldsymbol{k}}^{\mathrm{FBZ}} c_{pn}(\boldsymbol{k})^* c_{qn}(\boldsymbol{k}) \exp(i\boldsymbol{k}\boldsymbol{R}_j)\right) S^{0j}_{pq} \\ &= \sum_{pq}\sum_j \frac{1}{2}(2N+1)^3 P^{j0}_{qp} S^{0j}_{pq}, \end{aligned}$$

where from (9.59) after exchanging $p \leftrightarrow q$, $j \leftrightarrow l$ we had:

$$P^{jl}_{qp} = 2(2N+1)^{-3} \sum_n \sum_{\boldsymbol{k}}^{\mathrm{FBZ}} c_{qn}(\boldsymbol{k}) c_{pn}(\boldsymbol{k})^* \exp\big(i\boldsymbol{k}(\boldsymbol{R}_j - \boldsymbol{R}_l)\big) \tag{9.66}$$

and then

$$P^{j0}_{qp} = 2(2N+1)^{-3} \sum_n \sum_{\boldsymbol{k}}^{\mathrm{FBZ}} c_{qn}(\boldsymbol{k}) c_{pn}(\boldsymbol{k})^* \exp(i\boldsymbol{k}\boldsymbol{R}_j). \tag{9.67}$$

Hence, $\sum_{pq}\sum_j P^{j0}_{qp} S^{0j}_{pq} = 2n_0$.

9.11.4 ITERATIVE PROCEDURE

How does the Hartree–Fock method for periodic systems work?

- First (zeroth iteration), we start from a guess[46] for $\boldsymbol{P}$.
- Then, we calculate the elements F_{pq}^{0j} for all atomic orbitals p, q for unit cells $j = 0, 1, 2, \ldots, j_{\max}$. What is $j_{\max}$? The answer is certainly non-satisfactory: $j_{\max} = \infty$. In practice, however, we often take $j_{\max}$ as being of the order of a few cells, most often we take[47] $j_{\max} = 1$.
- For each $\boldsymbol{k}$ from the FBZ we calculate the elements F_{pq} and S_{pq} and then solve the secular equations within the Hartree–Fock–Roothaan procedure. This step requires diagonalization[48] (see Appendix K, p. 982). As a result, for each $\boldsymbol{k}$ we obtain a set of coefficients c.
- We repeat all this for the values of $\boldsymbol{k}$ covering in some optimal way (some recipes exist) the FBZ. We are then all set to carry out the numerical integration in the FBZ and we calculate an approximate matrix $\boldsymbol{P}$.
- This enables us to calculate a new approximation to the matrix $\boldsymbol{F}$ and so on, until the procedure converges in a self-consistent way, i.e. produces $\boldsymbol{P}$ very close to that matrix $\boldsymbol{P}$ which has been inserted into the Fock matrix $\boldsymbol{F}$. In this way we obtain the band structure $\varepsilon_n(\boldsymbol{k})$ and all the corresponding COs.

9.11.5 TOTAL ENERGY

How do we calculate the total energy for an infinite crystal? We know the answer without any calculation: $-\infty$. Indeed, since the energy represents an extensive quantity, for an infinite number of the unit cells we get $-\infty$, because a single cell usually represents a bound state (negative energy). Therefore, the question has to be posed in another way.

How to calculate the total energy *per unit cell*? Aha, this is a different story. Let us denote this quantity by E_T. Since a crystal only represents a *very* large molecule, we may use the expression (8.41) for the total energy of a molecule [noting that $\varepsilon_i = (i|\hat{F}|i)$]. In the 3D case we have:

$$(2N+1)^3 E_T = \frac{1}{2}\sum_{pq}\sum_{lj} P_{qp}^{jl}\left(h_{pq}^{lj} + F_{pq}^{lj}\right) + \frac{1}{2}\sum_{lj}\sum_{uv}{}' \frac{Z_u Z_v}{R_{uv}^{lj}}, \tag{9.72}$$

where the summation over p and q extends over the ω atomic orbitals that any unit cell offers, and l and j tell us in which cells these orbitals are located. The last term on the right-hand side refers to the nuclear repulsion of all the nuclei in the crystal, u, v number the nuclei in a unit cell, while l, j indicate the cells (a prime

---

[46] The result is presumed to be independent of this choice.

[47] "Nearest-neighbour approximation". We encounter a similar problem *inside* the F_{pq}^{0j}, because we have somehow to truncate the summations over h and l. These problems will be discussed later in this chapter.

[48] Unlike the molecular case, this time the matrix to diagonalize is Hermitian, and not necessarily symmetric. Methods of diagonalization exist for such matrices, and there is a guarantee that their eigenvalues are real.

means that there is no contribution from the charge interaction with itself). Since the summations over l and j extend over the whole crystal, therefore

$$\begin{aligned}(2N+1)^3 E_T &= \frac{1}{2}(2N+1)^3 \sum_{pq}\sum_{j} P^{j0}_{qp}\left[h^{0j}_{pq}+F^{0j}_{pq}\right]\\ &\quad + (2N+1)^3 \frac{1}{2}\sum_{j}\sum_{uv}{}' \frac{Z_u Z_v}{R^{0j}_{uv}}, \qquad (9.73)\end{aligned}$$

because each term has an equal contribution, and the number of such terms is equal to $(2N+1)^3$.

Therefore, the total energy per unit cell amounts to

$$\boxed{E_T = \frac{1}{2}\sum_{j}\sum_{pq} P^{j0}_{qp}\left(h^{0j}_{pq}+F^{0j}_{pq}\right) + \frac{1}{2}\sum_{j}\sum_{u}\sum_{v}{}' \frac{Z_u Z_v}{R^{0j}_{uv}}. \qquad (9.74)}$$

The formula is correct, but we can easily see that we are to be confronted with some serious problems. For example, the summation over nuclei represents a divergent series and we will get $+\infty$. *This problem appears only because we are dealing with an infinite system and we confront the long-range interactions*. We have to manage the problem somehow.

9.12 LONG-RANGE INTERACTION PROBLEM

What is left to be clarified are some problems about how to go with N to infinity.[49] It will be soon shown how dangerous this problem is.

We see from eqs. (9.65) and (9.74) that, despite using the translational symmetry to simplify the problem, we may treat each $\boldsymbol{k}$ separately. There are an infinite

---

[49] I will tell you about my adventure with this problem, because I remember very well how as a student I wanted to hear about struggles with understanding matter and ideas, instead of dry summaries.

The story began quite accidentally. In 1977, at the University of Namur (Belgium) Professor Joseph Delhalle asked the PhD student Christian Demanet to perform a numerical test. The test consisted of taking a simple infinite polymer (the infinite chain … LiH LiH LiH … had been chosen), to use the simplest atomic basis set possible and to see what we should take as N, to obtain the Fock matrix with sufficient accuracy. Demanet first took $N=1$, then $N=2$, $N=3$ – the Fock matrix changed all the time. He got impatient, took $N=10$, $N=15$ – the matrix continued to change. Only when he used $N=200$ did the Fock matrix elements stabilize within the accuracy of six significant figures. We could take $N=200$ for an extremely poor basis set and for a few such tests, but the quality of calculations will never be good and their cost would become astronomic. Even for the case in question the computations had to be done overnight. In a casual discussion at the beginning of my six-week stay at the University of Namur, Joseph Delhalle told me about the problem. He said also that in a recent paper the Austrian scientists Alfred Karpfen and Peter Schuster also noted that the results depend strongly on the chosen value of N. They made a correction *after* the calculations with a small N had been performed. They added the dipole–dipole electrostatic interaction of the cell 0 with a few hundred neighbouring cells, and for the dipole moment of a cell, they took the dipole moment of the isolated LiH molecule. As a result the Fock matrix elements changed much less with N. This information made me think about implementing the multipole expansion right from the beginning of the self-consistent Hartree–Fock–Roothaan procedure for a polymer. Below you will see what has been done. The presented theory pertains to a regular polymer (a generalization to 2D and 3D is possible).

number of them and this makes us a little nervous. In eq. (9.65) for F^{0j}_{pq} we have a summation (over the whole infinite crystal) of the interactions of an electron with all the nuclei, and in the next term a summation over the whole crystal of the electron–electron interactions. This is of course perfectly natural, because our system is infinite. The problem is, however, that both summations diverge: the first tends to $-\infty$, the second to $+\infty$. On top of this to compute the bond order matrix $\boldsymbol{P}$ we have to perform another summation in eq. (9.63) over the FBZ of the crystal. We have a similar, very unpleasant, situation in the total energy expression, where the first term tends to $-\infty$, while the nuclear repulsion term goes $+\infty$.

The routine approach was to replace the infinity by taking the first-neighbour interactions. This approach is quite understandable, because any attempt to take further neighbours ends up with an exorbitant bill to pay.[50]

9.12.1 FOCK MATRIX CORRECTIONS

A first idea we may think of is to carefully separate the long-range part of the Fock matrix elements and of the total energy from these quantities as calculated in a traditional way, i.e. by limiting the infinite-range interactions to those for the N neighbours on the left from cell 0 and N neighbours on the right of it. For the Fock matrix element we would have:

$$F^{0j}_{pq} = F^{0j}_{pq}(N) + C^{0j}_{pq}(N), \tag{9.75}$$

where $C^{0j}_{pq}(N)$ stands for the long-range correction, while $F^{0j}_{pq}(N)$ is calculated assuming interactions with the N right and N left neighbours of cell 0:

$$\begin{aligned} &F^{0j}_{pq}(N) \\ &\quad = T^{0j}_{pq} + \sum_{h=-N}^{h=+N} \left(-\sum_u Z_u V^{0j}_{pq}\left(A^h_u\right) + \sum_{l=h-N}^{l=h+N} \sum_{rs} P^{lh}_{sr} \left(\left(\begin{smallmatrix}0h\\pr\end{smallmatrix}\middle|\begin{smallmatrix}jl\\qs\end{smallmatrix}\right) - \frac{1}{2}\left(\begin{smallmatrix}0h\\pr\end{smallmatrix}\middle|\begin{smallmatrix}lj\\sq\end{smallmatrix}\right) \right)\right), \end{aligned} \tag{9.76}$$

$$C^{0j}_{pq}(N) = \sum_h{}^{\#} \left(-\sum_u Z_u V^{0j}_{pq}\left(A^h_u\right) + \sum_{l=h-N}^{l=h+N} \sum_{rs} P^{lh}_{sr} \left(\begin{smallmatrix}0h\\pr\end{smallmatrix}\middle|\begin{smallmatrix}jl\\qs\end{smallmatrix}\right)\right), \tag{9.77}$$

where the symbol $\sum_h^{\#}$ will mean a summation over all the unit cells *except* the section of unit cells with numbers $-N, -N+1, \ldots, 0, 1, \ldots, N$, i.e. the neighbourhood of cell 0 ("short-range"). The nuclear attraction integral:[51]

$$V^{0j}_{pq}\left(A^h_u\right) = \left(\chi^0_p \middle| \frac{1}{|\boldsymbol{r} - (\boldsymbol{A}_u + ha\boldsymbol{z})|} \middle| \chi^j_q \right), \tag{9.78}$$

where the vector $\boldsymbol{A}_u$ shows the position of the nucleus u in cell 0, while $\boldsymbol{A}^h_u \equiv \boldsymbol{A}_u + ha\boldsymbol{z}$ points to the position of the equivalent nucleus in cell h ($\boldsymbol{z}$ denotes the unit vector along the periodicity axis).

---

[50] The number of two-electron integrals, which quantum chemistry positively dislikes, increases with the number of neighbours taken (N) and the atomic basis set size per unit cell (ω) as $N^3\omega^4$. Besides, the nearest-neighbours *are* indeed the most important.

[51] Without the minus sign in the definition the name is not quite adequate.

The expression for $C^{0j}_{pq}(N)$ has a clear physical interpretation. The first term represents the interaction of the charge distribution $-\chi^0_p(1)^*\chi^j_q(1)$ (of electron 1, hence the sign $-$) with *all the nuclei*,[52] except those enclosed in the short-range region (i.e. extending from $-N$ to $+N$). The second term describes the interaction of the same electronic charge distribution with the *total electronic distribution* outside the short-range region. How do we see this? The integral $\left({}^{0h}_{pr}|{}^{jl}_{qs}\right)$ means the Coulombic interaction of the distribution under consideration $-\chi^0_p(1)^*\chi^j_q(1)$ with its partner-distribution $-\chi^h_r(2)^*\chi^l_s(2)$, doesn't it? This distribution is multiplied by P^{lh}_{sr} and then summed over all possible atomic orbitals r and s in cell h and its neighbourhood (the sum over cells l from the neighbourhood of cell h), which gives the total partner electronic distribution $-\sum_{l=h-N}^{l=h+N}\sum_{rs}P^{lh}_{sr}$ $\chi^h_r(2)^*\chi^l_s(2)$. This, however, simply represents the electronic charge distribution of cell h. Indeed, the distribution, when integrated gives [(just look at eq. (9.68)] $-\sum_{l=h-N}^{l=h+N}\sum_{rs}P^{lh}_{sr}S^{hl}_{rs}=2n_0$. Therefore, our electron distribution, $-\chi^0_p(1)^*\chi^j_q(1)$, interacts electrostatically with the charge distribution of all cells except those enclosed in the short-range region, because eq. (9.77) contains the summation over all cells h except the short-range region. Finally,

> the long-range correction to the Fock matrix elements $C^{0j}_{pq}(N)$ represents the Coulombic interaction of the charge distribution $-\chi^0_p(1)^*\chi^j_q(1)$ with all the unit cells (nuclei and electrons) from outside the short-range region.

In the $C^{0j}_{pq}(N)$ correction, in the summation over l, *we have neglected the exchange term* $-\frac{1}{2}\sum_{l=h-N}^{l=h+N}\sum_{rs}P^{lh}_{sr}\left({}^{0h}_{pr}|{}^{lj}_{sq}\right)$. The reason for this was that we have been *convinced*, that P^{lh}_{sr} vanishes very fast, when cell l separates from cell h. *Subsequent* reasoning would then be easy: the most important term ($l=h$) would be $-\frac{1}{2}\sum_{rs}P^{hh}_{sr}\left({}^{0h}_{pr}|{}^{hj}_{sq}\right)$. It contains the differential overlap $\chi^0_p(1)^*\chi^h_s(1)$, which decays exponentially when the cells 0 and h separate, and we have a *guarantee* [eq. (9.77)], that this separation is large.[53] We will come back to this problem.

9.12.2 TOTAL ENERGY CORRECTIONS

The total energy per unit cell could similarly be written as

$$E_T = E_T(N) + C_T(N), \tag{9.79}$$

---

[52] Cf. interpretation of the integral $-V^{0j}_{pq}(\boldsymbol{A}^h_u) = -(\chi^0_p(\boldsymbol{r})|\frac{1}{|\boldsymbol{r}-\boldsymbol{A}^h_u|}|\chi^j_q(\boldsymbol{r}))$.

[53] The exchange interactions are notorious for an exponential decay with distance when the two object separate. The matrix elements of $\boldsymbol{P}$ corresponding to distant atomic orbitals "should be" small. For the time being let us postpone the problem. We will come back to it and will see how delusive such feelings may be. We have to stress, however, that trouble will come only in some "pathological" situations. In most common cases everything will be all right.

where $E_T(N)$ means the total energy per unit cell as calculated by the traditional approach, i.e. with truncation of the infinite series on the N left and N right neighbours of the cell 0. The quantity $C_T(N)$ therefore represents the error, i.e. the long-range correction. The detailed formulae for $E_T(N)$ and $C_T(N)$ are the following

$$E_T(N) = \frac{1}{2}\sum_{j=-N}^{j=+N}\sum_{pq} P_{qp}^{j0}\left(h_{pq}^{0j} + F_{pq}^{0j}(N)\right) + \frac{1}{2}\sum_{j=-N}^{j=+N}\sum_{u}\sum_{v}{}' \frac{Z_u Z_v}{R_{uv}^{0j}}, \tag{9.80}$$

$$\begin{aligned} C_T(N) = {} & \frac{1}{2}\sum_{j}\sum_{pq} P_{qp}^{j0} C_{pq}^{0j}(N) \\ & + \frac{1}{2}\sum_{h}{}^{\#}\left(\sum_{j}\sum_{pq} P_{qp}^{j0}\sum_{u}\left[-Z_u V_{pq}^{0j}\left(A_u^h\right)\right] + \sum_{u}\sum_{v}{}' \frac{Z_u Z_v}{R_{uv}^{0h}}\right), \end{aligned} \tag{9.81}$$

where we have already separated from F_{pq}^{0j} its long-range contribution $C_{pq}^{0j}(N)$, so that $C_T(N)$ contains *all* the long-range corrections.

Eq. (9.81) for $C_T(N)$ may be obtained *by just looking* at eq. (9.80). The first term with $C_{pq}^{0j}(N)$ is evident,[54] it represents the Coulombic interaction of the *electronic distribution* (let us recall condition (9.68)) associated with cell 0 with *the whole polymer chain* except the short-range region. What, therefore, is yet to be added to $E_T(N)$? What it lacks is the Coulombic interaction of the *nuclei* of cell 0 with *the whole polymer chain*, except the short-range region. Let us see whether we have it in eq. (9.81). The last term means the Coulombic interaction of the nuclei of cell 0 with *all the nuclei of the polymer* except the short-range region (and again we know, why we have the factor $\frac{1}{2}$). What, therefore, is represented by the middle term?[55] It

---

[54] The factor $\frac{1}{2}$ may worry us a little. Why just $\frac{1}{2}$? Let us see. Imagine N *identical objects* $i = 0, 1, 2, \ldots, N-1$ *playing identical roles in a system* (like our unit cells). We will be interested in the energy per object, E_T. The total energy may be written as (let us assume here pairwise interactions only) $NE_T = \sum_j E_j + \sum_{i<j} E_{ij}$, where E_j and E_{ij} are, respectively, the isolated object energy and the pairwise interaction energy. Since the objects are identical, then

$$NE_T = NE_0 + \frac{1}{2}\sum_{i,j}{}' E_{ij} = NE_0 + \frac{1}{2}\sum_{i}\Bigl(\sum_{j}{}' E_{ij}\Bigr) = NE_0 + \frac{1}{2}N\Bigl(\sum_{j}{}' E_{0j}\Bigr),$$

where the prime means excluding self-interaction and the term in parentheses means the interaction of object 0 with all others. Finally, $E_T = E_0 + \frac{1}{2}(\sum_j' E_{0j})$, where we have the factor $\frac{1}{2}$ before the interaction of one of the objects with the rest of the system.

[55] As we can see, we have to sum (over j) to infinity the expressions h_{pq}^{0j}, which contain T_{pq}^{0j} [but these terms decay very fast with j and can all be taken into account in $E_T(N)$] and the long-range terms, the Coulombic interaction of the electronic charge distribution of cell 0 with the nuclei beyond the short-range region (the middle term in $C_T(N)$). The argument about fast decay with j of the kinetic energy matrix elements mentioned before follows from the double differentiation with respect to the coordinates of the electron. Indeed, this results in another atomic orbital, but with the same centre. This leads to the overlap integral of the atomic orbitals centred like those in $\chi_p^0 \chi_q^j$. Such an integral decays exponentially with j.

is clear, that it has to be (with the factor $\frac{1}{2}$) the Coulombic interaction of the *nuclei of cell* 0 with the total electronic distribution outside the short-range region. We look at the middle term. We have the sign "$-$". This is very good indeed, because we have to have an attraction. Further, we have the factor $\frac{1}{2}$, that is also OK, then we have $\sum_h^{\#}$, that is perfect, because we expect a summation over the long-range only, and finally we have $\sum_j \sum_{pq} P_{qp}^{j0} \sum_u [-Z_u V_{pq}^{0j}(A_u^h)]$ and we do not like this. This is the Coulombic interaction of the total *electronic distribution of cell* 0 with the *nuclei of the long-range region,* while we expected the interaction of the *nuclei* of cell 0 with the *electronic charge distribution of the long-range region*. What is going on? Everything is OK. Just count the interactions pairwise and at each of them reverse the locations of the interacting objects – the two interactions mean the same. Therefore,

> the long-range correction to the total energy per cell $C_T(N)$ represents the Coulombic interaction of cell 0 with all the cells from outside the short-range region.

We are now all set to calculate the long-range corrections $C_{pq}^{0j}(N)$ and $C_T(N)$. It is important to realize that all the interactions to be calculated pertain to objects that are *far away in space*.[56] This is what we have carefully prepared. This is the condition that enables us to apply the multipole expansion to each of the interactions (Appendix X).

9.12.3 MULTIPOLE EXPANSION APPLIED TO THE FOCK MATRIX

Let us first concentrate on $C_{pq}^{0j}(N)$. As seen from eq. (9.77) there are two type of interactions to calculate: the nuclear attraction integrals $V_{pq}^{0j}(A_u^h)$ and the electron repulsion integrals $\left({}^{0h}_{pr}|{}^{jl}_{qs}\right)$. In the second term, we may use the multipole expansion of $\frac{1}{r_{12}}$ given in the Appendix X on p. 1039. In the first term, we will do the same,

---

[56] Let us check this. What objects are we talking about? Let us begin from $C_{pq}^{0j}(N)$. As it is seen from the formula one of the interacting objects is the charge distribution of the first electron $\chi_p^0(1)^*\chi_q^j(1)$. The second object is the whole polymer except the nuclei and electrons of the neighbourhood of the cell 0. The charge distributions $\chi_p^0(1)^*\chi_q^j(1)$ with various j are always close to cell 0, because the orbital $\chi_p^0(1)$ is anchored at cell 0, and such a distribution decays exponentially when cell j goes away from cell 0. The fact that the nuclei with which the distribution $\chi_p^0(1)^*\chi_q^j(1)$ interacts are far apart is evident, but less evident is that the electrons with which the distribution interacts are also far away from cell 0. Let us have a closer look at the electron–electron interaction. The charge distribution of electron 2 is $\chi_r^h(2)^*\chi_s^l(2)$, and the summation over cells h excludes the neighbourhood of cell 0. Hence, because of the exponential decay there is a guarantee that the distribution $\chi_r^h(2)^*\chi_s^l(2)$ is bound to be close to cell h, if this distribution is to be of any significance. Therefore, the charge distribution $\chi_r^h(2)^*\chi_s^l(2)$ is certainly far away from cell 0.

Similar reasoning may be used for $C_T(N)$. The interacting objects are of the type $\chi_p^0(1)^*\chi_q^j(1)$, i.e. always close to cell 0, with the nuclei of cell h, and there is a guarantee that h is far away from cell 0. The long distance of the interacting nuclei (second term) is evident.

but this time one of the interacting particles will be the nucleus indicated by vector A_u^h. The corresponding multipole expansion reads as (in a.u.; the nucleus u of the charge Z_u interacts with the electron of charge -1, $n_k = n_l = \infty$, $S = min(k, l)$):

$$-\frac{Z_u}{r_{u1}} = \sum_{k=0}^{n_k} \sum_{l=0}^{n_l} \sum_{m=-S}^{m=+S} A_{kl|m|} R^{-(k+l+1)} \hat{M}_a^{(k,m)}(1)^* \hat{M}_b^{(l,m)}(u), \tag{9.82}$$

where R stands for the distance between the origins of the coordinate system centred in cell 0 and the coordinate system in cell h, which, of course, is equal to $R = ha$. The multipole moment operator of electron 1, $\hat{M}_a^{(k,m)}(1)$, reads as

$$\hat{M}_a^{(k,m)}(1) = -r_a^k P_k^{|m|}(\cos\theta_{a1}) \exp(im\phi_{a1}), \tag{9.83}$$

while

$$\hat{M}_b^{(l,m)}(u) = Z_u r_u^l P_l^{|m|}(\cos\theta_u) \exp(im\phi_u) = M_b^{(l,m)}(u) \tag{9.84}$$

denotes the multipole moment of nucleus u computed in the coordinate system of the cell h. When this expansion as well as the expansion for $\frac{1}{r_{12}}$, are inserted into (9.77) for $C_{pq}^{0j}(N)$, we obtain

$$\begin{aligned} C_{pq}^{0j}(N) &= \sum_h{}^{\#} \sum_{k=0}^{n_k} \sum_{l=0}^{n_l} \sum_{m=-S}^{m=+S} A_{kl|m|} R^{-(k+l+1)} \\ &\quad \times \Bigg(\big(\chi_p^0 \big| \hat{M}_a^{(k,m)}(1)^* \big| \chi_q^j\big) \Big[\sum_u M_b^{(l,m)}\big(A_u^h\big) \Big] \\ &\quad + \big(\chi_p^0 \big| \hat{M}_a^{(k,m)}(1)^* \big| \chi_q^j\big) \sum_{l'=h-N}^{l'=h+N} \sum_{rs} P_{sr}^{l'h} \big(\chi_r^h \big| \hat{M}_b^{(l,m)}(2) \big| \chi_s^{l'}\big) \Bigg) \\ &= \sum_h{}^{\#} \sum_{k=0}^{n_k} \sum_{l=0}^{n_l} \sum_{m=-S}^{m=+S} A_{kl|m|} R^{-(k+l+1)} \big(\chi_p^0 \big| \hat{M}_a^{(k,m)}(1)^* \big| \chi_q^j\big) \\ &\quad \times \Big[\sum_u M_b^{(l,m)}\big(A_u^h\big) + \sum_{l'=h-N}^{l'=h+N} \sum_{rs} P_{sr}^{l'h} \big(\chi_r^h \big| \hat{M}_b^{(l,m)}(2) \big| \chi_s^{l'}\big) \Big]. \end{aligned}$$

Let us note that in the square parentheses we have nothing but a multipole moment of *unit cell h*. Indeed, the first term represents the multipole moment of all the nuclei of cell h, while the second term is the multipole moment of electrons of unit cell h. The later can best be seen if we recall the normalization condition (9.68): $\sum_{l'=h-N}^{l'=h+N} \sum_{rs} P_{sr}^{l'h} S_{rs}^{hl'} = \sum_{l'=-N}^{l'=+N} \sum_{rs} P_{sr}^{l'0} S_{rs}^{0l'} = 2n_0$, with $2n_0$ denoting the number of electrons per cell. Hence, we can write

$$C_{pq}^{0j}(N)=\sum_{h}^{\#}\sum_{k=0}\sum_{l=0}\sum_{m=-S}^{m=+S}A_{kl|m|}R^{-(k+l+1)}\big(\chi_p^0\big|\hat{M}_a^{(k,m)}(1)^*\big|\chi_q^j\big)M^{(l,m)}(h), \tag{9.85}$$

where the dipole moment of cell h is given by:

$$M^{(l,m)}(h)=\left[\sum_u M_b^{(l,m)}\big(A_u^h\big)+\sum_{l'=h-N}^{l'=h+N}\sum_{rs}P_{sr}^{l'h}\big(\chi_r^h\big|\hat{M}_b^{(l,m)}(2)\big|\chi_s^{l'}\big)\right], \tag{9.86}$$

because the summation over u goes over the nuclei belonging to cell h, and the coordinate system b is anchored in cell h. Now it is time to say something most important.

> Despite the fact that $M^{(l,m)}(h)$ depends formally on h, in reality it is h-independent, because all the unit cells are identical.

Therefore, we may safely write that $M^{(l,m)}(h)=M^{(l,m)}$.

Now we will try to avoid a well hidden trap, and then we will be all set to prepare ourselves to pick the fruit from our orchard. The trap is that $A_{kl|m|}$ depends on h. How is this? Well, in the $A_{kl|m|}$ there is $(-1)^l$, while the corresponding $(-1)^k$ is absent, i.e. there is a thing that is associated with the 2^l-pole in the coordinate system b, and there no an analogous expression for its partner, the 2^k-pole of coordinate system a. Remember, however (Appendix X), that the axes z of both coordinate systems have been chosen in such a way that a "shoots" towards b, and *b does not* shoot towards a. Therefore, the two coordinate systems are not equivalent, and hence one may have $(-1)^l$, and not $(-1)^k$. The coordinate system a is associated with cell 0, the coordinate system b is connected to cell h. If $h>0$, then it is true that a shoots to b, but if $h<0$ their roles are exchanged. In such a case, in $A_{kl|m|}$ we should not put $(-1)^l$, but $(-1)^k$. If we do this then in the summation over h in eq. (9.85) the only dependence on h appears in a simple term $(ha)^{-(k+l+1)}$!

> It appears, therefore, to be a possibility of exactly summing the electrostatic interaction along an infinite polymer chain.

Indeed, the sum

$$\sum_{h=1}^{\infty}h^{-(k+l+1)}=\zeta(k+l+1), \tag{9.87}$$

where $\zeta(n)$ stands for the Riemann dzeta function, which is known to a high degree of accuracy and available in mathematical tables.[57]

Riemann dzeta function

---

[57] For example, M. Abramovitz, I. Stegun (eds.), "*Handbook of Mathematical Functions*", Dover, New York, 1968, p. 811.

Georg Friedrich Bernhard Riemann (1826–1866), German mathematician and physicist, professor at the University of Göttingen. Nearly all his papers gave rise to a new mathematical theory. His life was full of personal tragedies, he lived only 40 years, but despite this he made a giant contribution to mathematics, mainly in non-Euclidean geometries (his geometry plays an important role in the general theory of relativity), in the theory of integrals (Riemann integral), and in the theory of trigonometric series.

The interactions of cell 0 with all the other cells are enclosed in this number. When this is inserted into $C_{pq}^{0j}(N)$, then we obtain

$$C_{pq}^{0j}(N) = \sum_{k=0}^{\infty}\sum_{l=0}^{\infty} U_{pq}^{0j(k,l)} \frac{\Delta_N^{(k+l+1)}}{a^{(k+l+1)}}, \tag{9.88}$$

where

$$U_{pq}^{0j(k,l)} = \sum_{m=-S}^{m=+S} (-1)^m \big[(-1)^k + (-1)^l\big] \frac{(k+l)!}{(k+|m|)!(l+|m|)!} M_{pq}^{0j(k,m)*} M^{(l,m)}, \tag{9.89}$$

$$\Delta_N^{(n)} = \zeta(n) - \sum_{h=1}^{N} h^{-n}. \tag{9.90}$$

Note that the formula for $C_{pq}^{0j}(N)$ represents a sum of the multipole–multipole interactions. The formula also shows that

electrostatic interactions in a regular polymer come from a multipole–multipole interaction with the same parity of the multipoles,

which can be seen from the term[58] $[(-1)^k + (-1)^l]$.

According to the discussion in Appendix X, to preserve the invariance of the energy with respect to translation of the coordinate system, when computing $C_{pq}^{0j}(N)$

---

[58] The term appears due to the problem discussed above of "who shoots to whom" in the multipole expansion. What happens, is that the interaction of an even (odd) multipole of cell 0 with an odd (even) multipole on the right-hand side of the polymer cancels with a similar interaction with the left-hand side. It is easy to understand. Let us imagine the multipoles as non-pointlike objects built of the appropriate point charges. We look along the periodicity axis. An even multipole has the same signs at both ends, an odd one has the opposite signs. Thus, when the even multipole is located in cell 0, and the odd one on its right-hand side, this interaction will cancel exactly with the interaction of the odd one located on the left-hand side (at the same distance).

we have to add all the terms with $k+l+1=\text{const}$, i.e.:

$$C_{pq}^{0j}(N)=\sum_{n=3,5,\dots}^{\infty}\frac{\Delta_N^{(n)}}{a^n}\sum_{l=1}^{n-1}U_{pq}^{0j(n-l-1,l)}. \tag{9.91}$$

The above expression is equivalent to eq. (9.88), but automatically assures the translational invariance by taking into account all the necessary multipole–multipole interactions.[59]

What should we know, therefore, to compute the long-range correction $C_{pq}^{0j}(N)$ to the Fock matrix?[60] From eq. (9.91) it is seen that one has to know how to calculate three numbers: $U_{pq}^{0j(k,l)}$, a^{-n} and $\Delta_N^{(n)}$. The equation for the first one is given in Table 9.1, the other two are trivial, Δ is easy to calculate knowing the Riemann ζ function (from tables): in fact we have to calculate the multipole moments, and these are *one*-electron integrals (easy to calculate). Originally, before the multipole expansion method was designed we also had a large number of *two*-electron integrals (expensive to calculate). Instead of overnight calculations, the computer time was reduced to about 1 s and the results were more accurate.

9.12.4 MULTIPOLE EXPANSION APPLIED TO THE TOTAL ENERGY

As shown above, the long-range correction to the total energy means the interaction of cell 0 with all the cells from the long-range region multiplied by $\frac{1}{2}$. The reasoning pertaining to its computation may be repeated exactly in the way we have shown in the previous subsection. We have, however, to remember a few differences:

- what interacts is not the charge distribution $\chi_p^{0*}\chi_q^j$, but the complete cell 0,
- the result has to be multiplied by $\frac{1}{2}$ for reasons discussed earlier.

Finally we obtain:

$$C_T(N)=\frac{1}{2}\sum_{k=0}^{\infty}\sum_{l=0}^{\infty}U_T^{(k,l)}\frac{\Delta_N^{(k+l+1)}}{a^{k+l+1}}, \tag{9.92}$$

---

[59] Indeed, $\sum_{l=1}^{n-1}U_{pq}^{0j(n-l-1,l)}=U_{pq}^{0j(n-2,1)}+U_{pq}^{0j(n-3,2)}+\cdots+U_{pq}^{0j(0,n-1)}$, i.e. a review of all terms with $k+l+1=n$ except $U_{pq}^{0j(n-1,0)}$. This term is absent and that is OK, because it requires calculation of $M^{(0,0)}$, i.e. of the *charge of the elementary cell*, which has to stay electrically neutral (otherwise the polymer falls apart), therefore $M^{(0,0)}=0$. Why, however, does the summation over n not simply represent $n=1,2,\dots,\infty$, but contains only odd n's except $n=1$? What would happen if we took $n=1$? Look at eq. (9.88). The value $n=1$ requires $k=l=0$. This leads to the "monopole–monopole" interaction, but this is 0, since the whole unit cell (and one of the multipoles is that of the unit cell) carries no charge. The summation in (9.91) does not contain any even n's, because they would correspond to k and l of different parity, and such interactions (as we have shown before) are equal to 0. Therefore, indeed, (9.91) contains all the terms that are necessary.

[60] L. Piela, J. Delhalle, *Intern. J. Quantum Chem.* 13 (1978) 605.

Table 9.1. The quantities $U^{(k,l)}$ for $k+l<7$ are necessary for calculating the long-range corrections to the Fock matrix elements $U_{pq}^{0j(k,l)}$ and to the total energy per unit cell $U_T^{(k,l)}$. The parentheses [ ] mean the corresponding multipole moment. When computing the Fock matrix correction the first multipole moment [ ] stands for the multipole moment of the charge distribution $\chi_p^0\chi_q^j$, the second, of the unit cell. For example, $U^{(0,2)}$ for the correction $C_{pq}^{0j}(N)$ is equal to $\left({}^{0}_{p}\middle|{}^{j}_{q}\right)\left(\sum_u Z_u(3z_u^2-r_u^2)-\sum_{l'=-N}^{l'=+N}\sum_{rs}P_{sr}^{l'0}(\chi_r^0|3z^2-r^2|\chi_s^{l'})\right)$, while $U^{(0,2)}$ for $C_T(N)$ is equal 0, because [1] means the *charge* of the unit cell, which is equal to zero. In the table only U's for $k\leqslant l$ are given. If $l<k$, then the formula is the same, but the order of the moments is reversed

n	$U^{(k,l)},\ k+l+1=n$
3	$U^{(0,2)}=[1][3z^2-r^2]$
	$U^{(1,1)}=2[x][x]+2[y][y]-4[z][z]$
5	$U^{(0,4)}=\frac{1}{4}[1][35z^4-30z^2r^2+3r^4]$
	$U^{(1,3)}=4[z][3r^2z-5z^3]+3[x][5xz^2-r^2x]+3[y][5yz^2-r^2y]$
	$U^{(2,2)}=3[3z^2-r^2][3z^2-r^2]-24[xz][xz]-24[yz][yz]$ $+\frac{3}{2}[x^2-y^2][x^2-y^2]+6[xy][xy]$
7	$U^{(0,6)}=\frac{1}{8}[1][231z^6-315z^4r^2+105z^2r^4-5r^6]$
	$U^{(1,5)}=-\frac{3}{2}[z][63z^5-70z^3r^2+15zr^4]+\frac{15}{4}[x][21z^4x-14z^2xr^2+xr^4]$ $+\frac{15}{4}[y][21z^4y-14z^2yr^2+yr^4]$
	$U^{(2,4)}=\frac{15}{8}[3z^2-r^2][35z^4-30z^2r^2+3r^4]-30[xz][7z^3x-3xzr^2]$ $-30[yz][7z^3y-3yzr^2]+\frac{15}{4}[x^2-y^2][7z^2(x^2-y^2)-r^2(x^2-y^2)]$ $+15[xy][7z^2xy-xyr^2]$
	$U^{(3,3)}=-10[5z^3-3zr^2][5z^3-3zr^2]+\frac{45}{4}[5z^2x-xr^2][5z^2x-xr^2]$ $+\frac{45}{4}[5z^2y-yr^2][5z^2y-yr^2]-45[zx^2-zy^2][zx^2-zy^2]$ $-180[xyz][xyz]+\frac{5}{4}[x^3-3xy^2][x^3-3xy^2]+\frac{5}{4}[y^3-3x^2y][y^3-3x^2y]$

where

$$U_T^{(k,l)}=\sum_{m=-S}^{m=+S}\left((-1)^k+(-1)^l\right)\frac{(k+l)!(-1)^m}{(k+|m|)!(l+|m|)!}M^{(k,m)*}M^{(l,m)}. \tag{9.93}$$

Let us note that (for the same reasons as before)

> interaction of multipoles of different parity gives zero

and this time we have to do with the interaction of the multipoles of complete cells. The quantities $U_T^{(k,l)}$ are given in Table 9.1.

Do the Fock matrix elements and the total energy per cell represent finite values?

If the Fock matrix elements were infinite, then we could not manage to carry out the Hartree–Fock–Roothaan self-consistent procedure. If E_T were infinite, the periodic system could not exist at all. It is, therefore, important to know when we can safely *model* an infinite system.

For any finite system there is no problem: the results are always finite. The only danger, therefore, is the summation to infinity ("lattice sums"), which always ends with the interaction of a part or whole unit cell with an infinite number of distant cells. Let us take such an example in the simplest case of a single atom per cell. Let us assume that the atoms interact by the Lennard-Jones pairwise potential (p. 284):

$$E = \varepsilon\left[\left(\frac{r_0}{r}\right)^{12} - 2\left(\frac{r_0}{r}\right)^{6}\right],$$

where r means the interatomic distance, r_0 means the equilibrium distance and ε the depth of the potential well. Let us try to compute the lattice sum $\sum_j E_{0j}$, where E_{0j} means the interaction energy of the cells 0 and j. We see that, due to the form of the potential, for long distances what counts is the uniquely attractive term $-2\varepsilon(\frac{r_0}{r})^6$. When we take such interactions which pertain to a sphere of the radius R (with the origin located on atom 0), each individual term (i.e. its absolute value) decreases with increasing R. This is very important, because when we have a 3D lattice, the number of such interactions within the sphere *increases* as R^3. We see that the decay rate of the interactions will finally win and the lattice sum will converge. We can, however, easily see that if the decay of the pairwise interaction energy were slower, then we might have had trouble calculating the lattice sum. For example, if, instead of the neutral Lennard-Jones atoms, we took ions of the same charge, the interaction energy would explode to ∞. It is evident, therefore, that for each periodic system there are some conditions to be fulfilled if we want to have finite lattice sums.

These conditions are more severe for the Fock matrix elements because each of the terms represent the interaction of a *charge* with complete distant unit cells. The convergence depends on the asymptotic interaction energy of the potential. In the case of the multipole–multipole interaction, we know what the asymptotic behaviour looks like, it is $R^{-(0+l+1)} = R^{-(l+1)}$, where R stands for the intercell distance. The lattice summation in a nD lattice ($n = 1, 2, 3$) gives the partial sum dependence on R as $\frac{R^n}{R^{l+1}} = R^{n-l-1}$. This means that[61]

> in 1D the unit cell cannot have any non-zero net charge ($l = 0$), in 2D it cannot have a non-zero charge and dipole moment ($l = 1$), in 3D it cannot have a non-zero charge, dipole moment and quadrupole moment ($l = 2$).

9.13 BACK TO THE EXCHANGE TERM

The long-range effects discussed so far result from the Coulomb interaction in the Fock equation for a regular polymer. There is, however, also an exchange contri-

---

[61] L.Z. Stolarczyk, L. Piela, *Intern. J. Quantum Chem.* 22 (1982) 911.

bution, which has been postponed in the long-range region (p. 477). It is time now to consider this again. The exchange term in the Fock matrix element F_{pq}^{0j} had the form (see (9.65))

$$-\frac{1}{2}\sum_{h,l}\sum_{rs} P_{sr}^{lh}\left({}^{0h}_{pr}\middle|{}^{lj}_{sq}\right) \tag{9.94}$$

and gave the following contribution to the total energy per unit cell

$$E_{\text{exch}} = \sum_{j} E_{\text{exch}}(j), \tag{9.95}$$

where the cell 0-cell j interaction has the form (see (9.81)):

$$E_{\text{exch}}(j) = -\frac{1}{4}\sum_{h,l}\sum_{pqrs} P_{qp}^{j0} P_{sr}^{lh}\left({}^{0h}_{pr}\middle|{}^{lj}_{sq}\right). \tag{9.96}$$

It would be very nice to have the exchange contribution $E_{\text{exch}}(j)$ decaying fast, when j increases, because it could be enclosed in the short-range contribution. Do we have good prospects for this? The above formula shows (the integral) that the summation over l is *safe*: the contribution of those cells l that are far from cell 0 is negligible due to differential overlaps of type $\chi_p^0(1)^*\chi_s^l(1)$. The summation over cells h is safe as well (for the same reasons), because it is bound to be limited to the neighbourhood of cell j (see the integral).

In contrast, the only guarantee of a satisfactory convergence of the sum over j is that *we hope* the matrix element P_{qp}^{j0} decays fast if j increases.

So far, exchange contributions have been neglected, and there has been an indication suggesting that this was the right procedure. This was the magic word "exchange". All the experience of myself and my colleagues in intermolecular interactions whispers "this is surely a short-range type". In a manuscript by Sandor Suhai, I read that the exchange contribution is of a long-range type. To our astonishment this turned out to be right (just a few numerical experiments). We have a long-range exchange. After an analysis was performed it turned out that

the long-range exchange interaction appears if and only if the system is metallic.

A metallic system is notorious for its HOMO–LUMO quasidegeneracy, therefore, we began to suspect that when the HOMO–LUMO gap decreases, the P_{qp}^{j0} coefficients do not decay with j.

Such things are most clearly seen when the simplest example is taken, and the hydrogen molecule at long internuclear distance is the simplest prototype of a metal. Indeed, this is a system with half-filled orbital energy levels when the LCAO MO method is applied (in the simplest case: two atomic orbitals). Note that, after subsequently adding two extra electrons, the resulting system (let us not worry too much that such a molecule could not exist!) would model an insulator, i.e. all the levels are doubly occupied.[62]

Analysis of these two cases convinces us that indeed our suspicions were justified. Here are the bond order matrices we obtain in both cases (see Appendix S, p. 1015, S denotes the overlap integral of the 1s atomic orbitals of atoms a and b):

$$\boldsymbol{P} = (1+S)^{-1}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \quad \text{for } \mathrm{H}_2, \tag{9.97}$$

$$\boldsymbol{P} = \left(1-S^2\right)^{-1}\begin{pmatrix} 1 & -S \\ -S & 1 \end{pmatrix} \quad \text{for } \mathrm{H}_2^{2-}. \tag{9.98}$$

We see[63] how profoundly these two cases differ in the off-diagonal elements (they are analogues of P_{qp}^{j0} for $j \neq 0$).

> In the second case the proportionality of P_{qp}^{j0} and S ensures an *exponential, therefore very fast, decay* if j tends to ∞. In the first case there is no decay of P_{qp}^{j0} at all.

A detailed analysis for an infinite chain of hydrogen atoms ($\omega = 1$) leads to the following formula[64] for P_{qp}^{j0}:

$$P_{11}^{j0} = \frac{2}{\pi j}\sin\left(\frac{\pi j}{2}\right). \tag{9.99}$$

This means an extraordinarily slow decay of these elements (and therefore of the exchange contribution) with j. When the metallic regime is even slightly removed, the decay gets much, much faster.

> This result shows that the long-range character of the exchange interactions does not exist in reality. It represents an indication that the Hartree–Fock method fails in such a case.

---

[62] Of course, we could take two helium atoms. This would also be good. However, the first principle in research is “*in a single step only change a single parameter, analyze the result, draw the conclusions, and make the second step*”.

Just *en passant*, a second principle, also applies here. If we do not understand an effect, what should we do? Just divide the system in two parts and look where the effect persists. Keep dividing until the effect disappears. Take the simplest system in which the effect still exists, analyze the problem, understand it and go back slowly to the original system (this is why we have H_2 and H_2^{2-} here).

[63] L. Piela, J.-M. André, J.G. Fripiat, J. Delhalle, *Chem. Phys. Lett.* 77 (1981) 143.

[64] I.I. Ukrainski, *Theor. Chim. Acta* 38 (1975) 139, $q = p = 1$ means that we have a single 1s hydrogen orbital per unit cell.

9.14 CHOICE OF UNIT CELL

The concept of the unit cell has been important throughout the present chapter. The unit cell represents an object that, when repeated by translations, gives an infinite crystal. In this simple definition almost every word can be a trap.

Is it feasible? Is the choice unique? If not, then what are the differences among them? How is the motif connected to the unit cell choice? Is the motif unique? Which motifs may we think about?

As we have already noted, the choice of unit cell as well as of motif is not unique. This is easy to see. Indeed Fig. 9.21 shows that the unit cell and the motif can be chosen in many different and equivalent ways.

Moreover, there is no chance of telling, in a responsible way, which of the choices are reasonable and which are not. And it happens that in this particular case we really have a plethora of choices. Putting no limits to our fantasy, we may choose a unit cell in a particularly capricious way, Figs. 9.21.b and 9.22.

Fig. 9.22 shows six different, fully legitimate, choices of motifs associated with a unit cell in a 1D "polymer" $(\mathrm{LiH})_\infty$. Each motif consists of the lithium nucleus, a proton and an electronic charge distribution in the form of two Gaussian $1s$ orbitals that accommodate four electrons altogether. By repeating any of these motifs we reconstitute the same original chain.

We may say there may be many legal choices of motif, but this is without any theoretical meaning, because all the choices lead to the same infinite system. Well, this is true with respect to theory, but in practical applications the choice of motif may be of prime importance. We can see this from Table 9.2, which corresponds to Fig. 9.22.

The results without the long-range interactions, depend very strongly on the choice of unit cell motif.

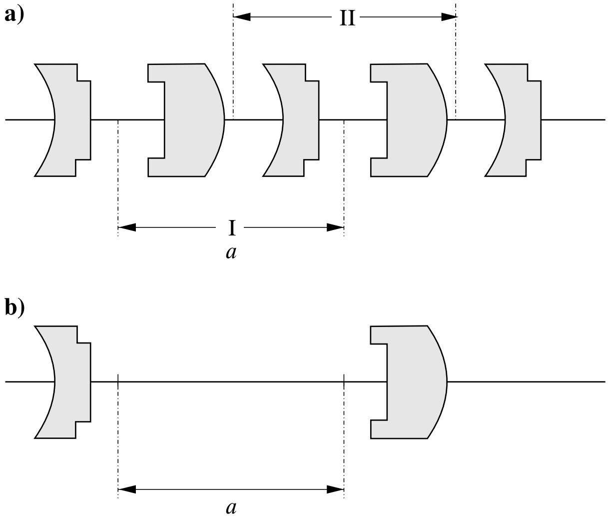


Fig. 9.21. Three of many possible choices of the unit cell motif. a) choices I and II differ, but both look "reasonable"; b) choice III might be called strange. Despite this strangeness, choice III is as legal (from the point of view of mathematics) as I or II.

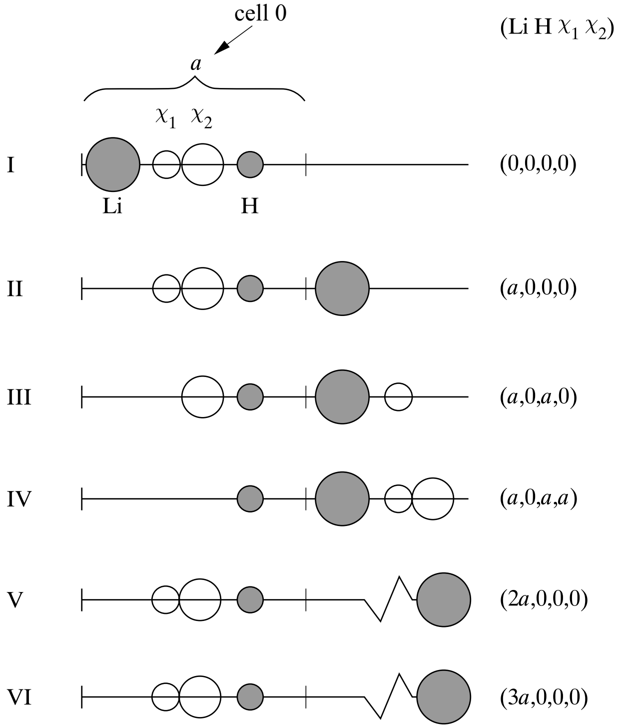


Fig. 9.22. Six different choices (I–VI) of unit cell content (motifs) for a linear chain $(\mathrm{LiH})_\infty$. Each cell has the same length $a = 6.3676$ a.u. There are two nuclei: Li^{3+} and H^+ and two Gaussian doubly occupied 1*s* atomic orbitals (denoted by χ_1 and χ_2 with exponents 1.9815 and 0.1677, respectively) *per cell*. Motif I corresponds to a "common sense" situation: both nuclei and electron distribution determined by χ_1 and χ_2 are within the section $(0,a)$. The other motifs (II–VI) all correspond to the same unit cell $(0,a)$ of length a, but are very strange. Each motif is characterized by the symbol (ka, la, ma, na). This means that the Li nucleus, H nucleus, χ_1 and χ_2 are shifted to the right by ka, la, ma, na, respectively. *All the unit cells with their contents (motifs) are fully justified, equivalent from the mathematical point of view, and, therefore, "legal" from the point of view of physics*. Note that the nuclear framework and the electronic density corresponding to a cell are very different for all the choices.

Use of the multipole expansion greatly improves the results and, to very good accuracy, makes them *independent of the choice of unit cell motif*.

Note that the larger the dipole moment of the unit cell the worse the results. This is understandable, because the first non-vanishing contribution in the multipole expansion is the dipole–dipole term (cf. Appendix X). Note how considerably the unit cell dependence drops after this term is switched on (a^{-3}).

The conclusion is that in the standard (i.e. short-range) calculations we should always choose the unit cell motif that corresponds to the smallest dipole moment. It seems however that such a motif is what everybody would choose using their "common sense".

Table 9.2. Total energy per unit cell E_T in the "polymer" LiH as a function of unit cell definition (Fig. 9.22, I–V). For each choice of unit cell this energy is computed in four ways: (1) without long-range forces (long range = 0), i.e. unit cell 0 interacts with $N = 6$ unit cells on its right-hand side and N unit cells on its left-hand side; (2), (3), (4) with the long range computed with multipole interactions up to the a^{-3}, a^{-5} and a^{-7} terms. The bold figures are exact. The corresponding dipole moment μ of the unit cell (in Debyes) is also given.

Unit cell	Long range	μ	$-E_T$
I	0	6.6432	**6.61**0869
	a^{-3}	6.6432	**6.61279469**2
	a^{-5}	6.6432	**6.61279468**7
	a^{-7}	6.6432	**6.612794674**
II	0	−41.878	**6**.524802885
	a^{-3}	−41.878	**6.612**519674
	a^{-5}	−41.878	**6.61279**0564
	a^{-7}	−41.878	**6.6127946**04
III	0	−9.5305	**6.60**7730984
	a^{-3}	−9.5305	**6.61278**8446
	a^{-5}	−9.5305	**6.612794**633
	a^{-7}	−9.5305	**6.61279467**3
IV	0	22.82	**6.5**7395630
	a^{-3}	22.82	**6.612**726254
	a^{-5}	22.82	**6.612793**807
	a^{-7}	22.82	**6.6127946**62
V	0	−90.399	6.148843431
	a^{-3}	−90.399	**6.60**7530384
	a^{-5}	−90.399	**6.61**2487745
	a^{-7}	−90.399	**6.6127**74317

9.14.1 FIELD COMPENSATION METHOD

In a moment we will unexpectedly find a quite different conclusion. The logical chain of steps that led to it has, in my opinion, a didactic value, and contains a considerable amount of optimism. When this result was obtained by Leszek Stolarczyk and myself, we were stunned by its simplicity.

> Is it possible to design a unit cell motif with a dipole moment of zero? This would be a great unit cell, because its interaction with other cells would be weak and it would decay fast with intercellular distance. We could therefore compute the interaction of a few cells like this and the job would be over: we would have an accurate result at very low cost.

There is such a unit cell motif.

Imagine we start from the concept of the unit cell with its motif (with lattice constant a). This motif is, of course, electrically neutral (otherwise the total energy

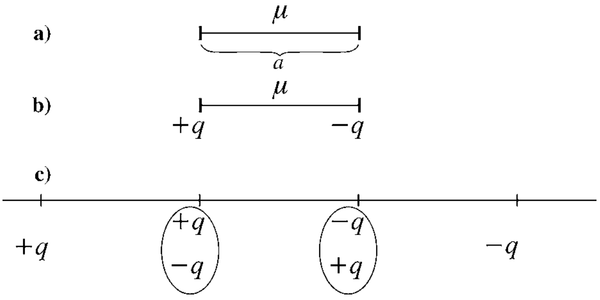


Fig. 9.23. Field compensation method. (a) the unit cell with length a and dipole moment $\mu > 0$. (b) the modified unit cell with additional fictitious charges ($|q| = \frac{\mu}{a}$) which cancel the dipole moment. (c) the modified unit cells (with $\mu' = 0$) give the original polymer, when added together.

would be $+\infty$), and its dipole moment component along the periodicity axis is equal to μ. Let us put its symbol in the unit cell, Fig. 9.23.a.

Now let us add to the motif two extra pointlike opposite charges ($+q$ and $-q$), located on the periodicity axis and separated by a. The charges are chosen in such a way ($q = \frac{\mu}{a}$) that they alone give the dipole moment component along the periodicity axis equal to $-\mu$, Fig. 9.23.b.

In this way the new unit cell dipole moment (with the additional fictitious charges) is equal to zero. Is this an acceptable choice of motif? Well, what does acceptable mean? The only requirement is that by repeating the new motif with period a, we have to reconstruct the whole crystal. What will we get when repeating the new motif? Let us see (Fig. 9.23.c).

We get the original periodic structure, because the charges all along the polymer, *except the boundaries*, have cancelled each other. Simply, the pairs of charges $+q$ and $-q$, when located at a point result in nothing.

In practice we would like to repeat just a few neighbouring unit cell motifs (a cluster) and then compute their interaction. In such case, we will observe the charge cancellation inside the cluster, but no cancellation on its boundaries ("surface").

Therefore we get a sort of point charge distribution at the boundaries.

If the boundary charges did not exist, it would correspond to the traditional calculations of the original unit cells without taking any long-range forces into account. The boundary charges therefore play the important role of replacing the electrostatic interaction with the rest of the *infinite* crystal, by the boundary charge interactions with the cluster ("field compensation method").

This is all. The consequences are simple.

Let us not only kill the dipole moment, but also other multipole moments of the unit cell content (up to a maximum moment), and the resulting cell will be unable to interact electrostatically with anything. Therefore, interaction within a small cluster of such cells will give us an accurate energy per cell result.

This multipole killing (field compensation) may be carried out in several ways.[65]

Application of the method is extremely simple. Imagine unit cell 0 and its neighbour unit cells (a cluster). Such a cluster is sometimes treated as a molecule and its role is to represent a bulk crystal. This is a very expensive way to describe the bulk crystal properties, for the cluster surface atom ratio to the bulk atom is much higher than we would wish (the surface still playing an important role). What is lacking is the crystal field that will change the cluster properties. In the field compensation method we do the same, but there are some fictitious charges at the cluster boundaries that take care of the crystal field. This enables us to use a smaller cluster than before (low cost) and still get the influence of the infinite crystal. The fictitious charges are treated in computations the same way as are the nuclei (even if some of them are negatively charged). However artificial it may seem, the results are far better when using the field compensation method than without it.[66]

9.14.2 THE SYMMETRY OF SUBSYSTEM CHOICE

The example described above raises an intriguing question, pertaining to our understanding of the relation between a part and the whole.

There are an infinite number of ways to reconstruct the same system from parts. *These ways are not equivalent in practical calculations*, if for any reason we are unable to compute all the interactions in the system. However, if we have a theory (in our case the multipole method) that is able to compute the interactions,[67] including the long-range forces, then it turns out the final result is virtually independent of the choice of unit cell motif. This arbitrariness of choice of subsystems looks analogous to the arbitrariness of the choice of coordinate system. The final results do not depend on the coordinate system used, but still the numerical results (as well as the effort to get the solution) do.

Bogolyubov transformation

The separation of the whole system into subsystems is of key importance to many physical approaches, but we rarely think of the freedom associated with the choice. For example, an atomic nucleus does not in general represent an elementary particle, and yet in quantum mechanical calculations we treat it as a point particle, without an internal structure and we are successful.[68] Further, in the *Bogolyubov*[69] *transformation*, the Hamiltonian is represented by creation and annihilation operators, each being a linear combination of the creation and annihilation

---

[65] L. Piela, L.Z. Stolarczyk, *Chem. Phys. Letters* 86 (1982) 195.

[66] Using "negative" nuclei looked so strange that some colleagues doubted receiving anything reasonable from such a procedure.

[67] With controlled accuracy, i.e. we still neglect the interactions of higher multipoles.

[68] This represents only a fragment of the story-like structure of science (cf. p. 60), one of its most intriguing features. It makes science operate, otherwise when considering the genetics of peas in biology we have had to struggle with the quark theory of matter.

[69] Nicolai Nicolaevitch Bogolyubov (1909–1992), Russian physicist, director of the Dubna Nuclear Institute, outstanding theoretician.

operators for electrons (described in Appendix U, p. 1023). The new operators also fulfil the anticommutation rules – only the Hamiltonian contains more additional terms than before (Appendix U). A particular Bogolyubov transformation may describe the creation and annihilation of quasi-particles, such as the electron hole (and others). We are dealing with the same physical system as before, but we look at it from a completely different point of view, by considering it is composed of something else. Is there any theoretical (i.e. *serious*) reason for preferring one division into subsystems over another? Such a reason may only be of practical importance.[70]

SYMMETRY WITH RESPECT TO DIVISION INTO SUBSYSTEMS

The symmetry of *objects* is important for the description of *them*, and therefore may be viewed as of limited interest. The symmetry of the laws of Nature, i.e. of the theory that describes all objects (whether symmetric or not) is much more important. This has been discussed in detail in Chapter 2 (cf. p. 61), but it seems that we did not list there a fundamental symmetry of any correct theory: the *symmetry with respect to the choice of subsystems. A correct theory has to describe the total system independently of what we decide to treat as subsystems*.

We will meet this problem once more in intermolecular interactions (Chapter 13). However, in the periodic system it has been possible to use, in computational practice, the symmetry described above.

Our problem resembles an excerpt which I found in "*Dreams of a Final Theory*" by Steven Weinberg[71] pertaining to *gauge symmetry*: "*The symmetry underlying it has to do with changes in our point of view about the identity of the different types of elementary particle. Thus it is possible to have a particle wave function that is neither definitely an electron nor definitely a neutrino, until we look at it*". Here also we have freedom in the choice of subsystems and a correct theory has to reconstitute the description of the whole system.

An intriguing problem.

Summary

- A crystal is often approximated by an infinite crystal (*primitive) lattice*, which leads to the concept of the *unit cell*. By repeating a chosen atomic *motif* associated with a unit cell, we reconstruct the whole infinite crystal.
- The Hamiltonian is invariant with respect to translations by any lattice vector. Therefore its eigenfunctions are simultaneously eigenfunctions of the translation operators (Bloch theorem): $\phi_{\boldsymbol{k}}(\boldsymbol{r}-\boldsymbol{R}_j)=\exp(-i\boldsymbol{k}\boldsymbol{R}_j)\phi_{\boldsymbol{k}}(\boldsymbol{r})$ and transform according to the irreducible representation of the translation group labelled by the *wave vector* $\boldsymbol{k}$.

---

[70] For example, at temperature $t<0\,^\circ\mathrm{C}$ we may solve the equations of motion for N frozen water drops and we may obtain reasonable dynamics of the system. At $t>0\,^\circ\mathrm{C}$, obtaining dynamics of the same drops will be virtually impossible.

[71] Pantheon Books, New York (1992), Chapter 6.

- *Bloch functions* may be treated as atomic symmetry orbitals $\phi = \sum_j \exp(i\boldsymbol{k}\boldsymbol{R}_j)\chi(\boldsymbol{r} - \boldsymbol{R}_j)$ formed from the atomic orbital $\chi(\boldsymbol{r})$.
- The crystal lattice basis vectors allow the formation of the basis vectors of the *inverse lattice.*
- Linear combinations of them (with integer coefficients) determine the *inverse lattice*.
- The *Wigner–Seitz* unit cell of the inverse lattice is called the First Brillouin Zone (FBZ).
- The vectors $\boldsymbol{k}$ inside the FBZ label all possible irreducible representations of the translation group.
- The wave vector plays a triple role:
  - it indicates the *direction of the wave*, which is an eigenfunction of $\hat{T}(\boldsymbol{R}_j)$ with eigenvalue $\exp(-i\boldsymbol{k}\boldsymbol{R}_j)$,
  - it labels the *irreducible representations* of the translation group,
  - the longer the wave vector $\boldsymbol{k}$, the more nodes the wave has.
- In order to neglect the crystal surface, we apply the *Born–von Kármán boundary condition*: “instead of a stick-like system we take a circle”.
- In full analogy with molecules, we can formulate the SCF LCAO CO Hartree–Fock–Roothaan method (CO instead MO). Each CO is characterized by a vector $\boldsymbol{k} \in$ FBZ and is a linear combination of the Bloch functions (with the same $\boldsymbol{k}$).
- The orbital energy dependence on $\boldsymbol{k} \in$ FBZ is called the *energy band*. The stronger the intercell interaction, the wider the bandwidth (dispersion).
- Electrons occupy (besides the inner shells) the *valence bands*, the *conduction bands* are empty. The *Fermi level* is the HOMO energy of the crystal. If the HOMO–LUMO energy difference (*energy gap* between the valence and conduction bands) is zero, we have a *metal*; if it is large, we have an *insulator*; if, it is medium, we have a *semiconductor*.
- Semiconductors may be intrinsic, or n-type (if the donor dopant levels are slightly below the conduction band), or p-type (if the acceptor dopant levels are slightly above the occupied band).
- Metals when cooled may undergo what is known as the Peierls transition, which denotes lattice dimerization and band gap formation. The system changes from a metal to a semiconductor or insulator. This transition corresponds to the Jahn–Teller effect in molecules.
- Polyacetylene is an example of a Peierls transition (“dimerization”), which results in shorter bonds (a little “less–multiple” than double ones) and longer bonds (a little “more multiple” than single ones). Such a dimerization introduces the possibility of a defect separating two rhythms (“phases”) of the bonds: from “double–single” to “single–double”. This defect can move within the chain, which may be described as a solitonic wave. The soliton may become charged and in this case, participates in electric conduction (increasing it by many orders of magnitude).
- In polyparaphenylene, a soliton wave is not possible, because the two phases, quinoid and aromatic, are not of the same energy. A double defect is possible though, a bipolaron. Such a defect represents a section of the quinoid structure (in the aromatic-like chain) at the end of which we have two unpaired electrons. The electrons, when paired with extra electrons from donor dopants, or when removed by acceptor dopants, form a double ion (bipolaron), which may contribute to electric conductance.
- The band structure may be foreseen in simple cases and logically connected to the subsystem orbitals.
- To compute the Fock matrix elements or the total energy per cell, we have to calculate the interaction of cell 0 with all other cells.

- The interaction with neighbouring cells is calculated without approximations, while that with distant cells uses multipole expansion. Multipole expansion applied to the electrostatic interaction gives accurate results, while the numerical effort is dramatically reduced.
- In some cases (metals), we meet long-range exchange interaction, which disappears as soon as the energy gap emerges. This indicates that the Hartree–Fock method is not applicable in this case.
- The choice of unit cell motif is irrelevant from the theoretical point of view, but leads to different numerical results when the long-range interactions are omitted. By including the interactions the theory becomes independent of the division of the whole system into arbitrary motifs.

Main concepts, new terms

lattice constant (p. 431)
primitive lattice (p. 432)
translational symmetry (p. 432)
unit cell (p. 432)
motif (p. 432)
wave vector (p. 434)
Bloch theorem (p. 434)
Bloch function (p. 435)
symmetry orbital (p. 435)
biorthogonal basis (p. 436)
inverse lattice (p. 436)
Wigner–Seitz cell (p. 438)
First Brillouin Zone (p. 438)
Born–von Kármán boundary condition (p. 446)
crystal orbitals (p. 450)
band structure (p. 453)
band width (p. 454)
Fermi level (p. 454)
valence band (p. 455)
band gap (p. 455)
conduction band (p. 455)
insulators (p. 455)
metals (p. 455)
semi-conductor (p. 455)
Peierls transition (p. 456)
n-type semiconductor (p. 458)
p-type semiconductor (p. 458)
Jahn–Teller effect (p. 458)
soliton (p. 459)
bipolaron (p. 459)
long-range interactions (p. 475)
multipole expansion (p. 479)
exchange interaction (p. 485)
field compensation method (p. 490)
symmetry of division into subsystems (p. 492)

From the research front

The Hartree–Fock method for periodic systems nowadays represents a routine approach coded in several *ab initio* computer packages. We may analyze the total energy, its dependence on molecular conformation, the density of states, the atomic charges, etc. Also calculations of first-order responses to the electric field (polymers are of interest for optoelectronics) have been successful in the past. However, non-linear problems (like the second harmonic generation, see Chapter 12) still represent a challenge. On the one hand, the experimental results exhibit wide dispersion, which partly comes from market pressure. On the other hand, the theory itself has not yet elaborated reliable techniques.

Ad futurum…

Probably there will soon be no problem in carrying out the Hartree–Fock or DFT (see Chapter 11) calculations, even for complex polymers and crystals. What will remain for a few decades is the very important problem of lowest-energy crystal packing and of solid state reactions and phase transitions. Post-Hartree–Fock calculations (taking into account electronic correlation effects) will be more and more important. The real challenge will start in designing non-periodic materials, where the polymer backbone will serve as a molecular rack for installing some functions (transport, binding, releasing, signal transmitting). The functions will be expected to cooperate ("smart materials", cf. Chapter 15).

Additional literature

A.A. Levin, "Vviedienije w kvantovuyou khimiyou tverdovo tiela. Khimicheskaya sviaz i struktura energeticheskikh zon w tietraedricheskikh poluprovodnikakh", Khimija, Moscow, 1974.

This is the first textbook of solid state chemistry. The theory of periodic systems (especially semiconductors) is presented in about 230 pages.

R. Hoffmann, "Solids and Surfaces. A Chemist's View of Bonding in Extended Structures", VCH publishers, New York, 1988.

A masterpiece written by a Nobel Prize winner, one of the founders of solid state quantum chemistry. More oriented towards chemistry than Levin's book. Solid state theory was traditionally the domain of physicists, some concepts typical of chemistry as, e.g., atomic orbitals, bonding and antibonding effects, chemical bonds and localization of orbitals were usually absent in such descriptions.

J.-M. André, J. Delhalle, J.-L. Brédas, "Quantum Chemistry Aided Design of Organic Polymers", World Scientific, Singapore, 1991.

A well written book oriented mainly towards the response of polymers to the electric field.

Questions

1. Bloch theorem says that:
   a) $\hat{T}(\boldsymbol{R}_j)\phi_{\boldsymbol{k}}(\boldsymbol{r}) = \exp(i\boldsymbol{k}\boldsymbol{r})\phi_{\boldsymbol{k}}(\boldsymbol{r})$; b) $\phi_{\boldsymbol{k}}(\boldsymbol{r}) = \phi_{-\boldsymbol{k}}(\boldsymbol{r})$; c) $\phi_{\boldsymbol{k}}(\boldsymbol{r}) + \phi_{-\boldsymbol{k}}(\boldsymbol{r}) = 0$; d) $\phi_{\boldsymbol{k}}(\boldsymbol{r} - \boldsymbol{R}_j) = \exp(-i\boldsymbol{k}\boldsymbol{R}_j)\phi_{\boldsymbol{k}}(\boldsymbol{r})$.
2. The First Brillouin Zone ($\boldsymbol{k}$ stands for the wave vector, CO – for a crystal orbital):
   a) represents the smallest unit cell of the primitive lattice; b) represents the smallest motif in the crystal; c) its interior contains only non-equivalent vectors; d) represents a basis in the inverse cell.
3. Function $\phi_{\boldsymbol{k}}$ corresponding to the wave vector $\boldsymbol{k}$:
   a) has to satisfy the Schrödinger equation; b) represents a wave with direction $\boldsymbol{k}$; c) always has $|\boldsymbol{k}|$ nodes; d) represents the CO.
4. Crystal orbital ($\boldsymbol{k}$ is a wave vector, CO means a crystal orbital):
   a) represents an arbitrary linear combination of the atomic orbitals of cell 0; b) represents an arbitrary linear combination of the atomic orbitals of cells 0, ±1; c) with $\boldsymbol{k} = \boldsymbol{0}$ corresponds to the lowest energy in the band; d) always corresponds to a given $\boldsymbol{k}$.
5. The infinite polyacetylene chain:
   a) is an electrical conductor; b) exhibits all CC bonds of equal length; c) if doped becomes a conductor due to soliton defects; d) has a zero band gap.
6. Band width ($\boldsymbol{k}$ is a wave vector, CO means a crystal orbital) gets larger if:
   a) the orbital overlap is larger; b) the band energy is lower; c) the CO has a larger number of nodes; d) $|\boldsymbol{k}|$ is smaller.

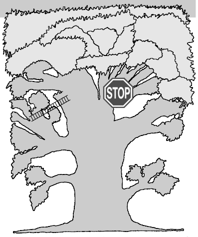
7. A semiconductor:
   a) has a large band gap; b) means a resistor; c) has a small band gap; d) conducts electric current, but only in one direction.
8. Fermi level means:
   a) the electric affinity of an insulator; b) HOMO energy for the crystal; c) the mean value of the occupied band; d) the lowest energy of a band.
9. In regular polymer the dipole-quadrupole interaction is:
   a) 0; b) the difference of the quadrupole–quadrupole and dipole–dipole interactions; c) $\frac{1}{3!}$; d) the mean value of the dipole–dipole and quadrupole–quadrupole interactions.
10. The dipole moment of a unit cell in a polymer:
    a) is uniquely determined if the polymer is electrically neutral; b) depends on the position of the cell with respect to cell 0; c) depends on the choice of the motif; d) is equal to 0.

Answers

1d, 2c, 3b, 4d, 5c, 6a, 7c, 8b, 9a, 10c

Chapter 10

CORRELATION OF THE ELECTRONIC MOTIONS



WHERE ARE WE?

Where are we?

The main road on the trunk lead us to the middle of the crown of the tree.

An example

As usual let us consider the simplest example: the hydrogen molecule. The normalized Restricted Hartree–Fock determinant

$$\psi_{\mathrm{RHF}}(1,2)=\frac{1}{\sqrt{2}}\begin{vmatrix}\phi_1(1) & \phi_1(2)\\ \phi_2(1) & \phi_2(2)\end{vmatrix}$$

with double occupancy of the normalized molecular orbital φ ($\phi_1=\varphi\alpha$, $\phi_2=\varphi\beta$), after expansion, gives immediately:

$$\psi_{\mathrm{RHF}}(1,2)=\varphi(1)\varphi(2)\frac{1}{\sqrt{2}}\big\{\alpha(1)\beta(2)-\beta(1)\alpha(2)\big\}.$$

The key quantity here is the square of the complex modulus of $\psi_{\mathrm{RHF}}(1,2)$, since the $|\psi_{\mathrm{RHF}}(1,2)|^2$ tells us about the probability density of the occurrence of certain coordinates of the electrons. We will study the fundamental problem for the motion of electrons: whether the electrons react to their presence.

Let us ask some very important questions. Firstly what is the (conditional) probability density of the occurrence of the situation when

- electron 1 occupies *different positions* in space *on the contour line* $\varphi=\mathrm{const}$ and has spin coordinate $\sigma_1=1/2$ while electron 2 has spin coordinate $\sigma_2=-1/2$, and its space coordinates are x_2, y_2, z_2.

We calculate

$$\begin{aligned}\big|\psi_{\mathrm{RHF}}(1,2)\big|^2&=\left[\varphi(1)\varphi(2)\frac{1}{\sqrt{2}}\big\{\alpha(\sigma_1)\beta(\sigma_2)-\beta(\sigma_1)\alpha(\sigma_2)\big\}\right]^2\\&=\left[\mathrm{const}\times\varphi(x_2,y_2,z_2)\frac{1}{\sqrt{2}}\big\{\alpha(1/2)\beta(-1/2)-\beta(1/2)\alpha(-1/2)\big\}\right]^2\\&=\left[\mathrm{const}\times\varphi(x_2,y_2,z_2)\frac{1}{\sqrt{2}}\{1\times1-0\times0\}\right]^2\\&=\frac{1}{2}(\mathrm{const})^2\times\varphi^2(x_2,y_2,z_2).\end{aligned}$$

Electron 1 *changes its position on the contour line, but the distribution of the probability density of electron* 2 *(of the opposite spin) does not change a bit, while electron* 2 *should move away from its partner, since the electrons repel each other.* Electron 2 is not afraid to approach electron 1. The latter can even touch electron 2 and it does not react at all. For such a deficiency we have to pay through the high mean value of the Hamiltonian (since there is a high average energy of the electron repulsion). The Hartree–Fock method therefore has an obvious shortcoming.

- We now ask about the probability density of finding a situation in which we leave everything the same as before, but now electron 2 has spin coordinate $\sigma_2 = 1/2$ (so this is the situation where both electrons have identical projections of spin angular momentum[1]). What will the response to this change be of $|\psi_{\mathrm{RHF}}(1,2)|^2$ as a function of the position of electron 2?
- Again we calculate

$$\begin{aligned}|\psi_{\mathrm{RHF}}(1,2)|^2 &= \left[\varphi(1)\varphi(2)\frac{1}{\sqrt{2}}\{\alpha(\sigma_1)\beta(\sigma_2)-\beta(\sigma_1)\alpha(\sigma_2)\}\right]^2\\ &= \left[\text{const}\times\varphi(x_2,y_2,z_2)\frac{1}{\sqrt{2}}\left\{\alpha\left(\frac{1}{2}\right)\beta\left(\frac{1}{2}\right)-\beta\left(\frac{1}{2}\right)\alpha\left(\frac{1}{2}\right)\right\}\right]^2\\ &= \left[\text{const}\times\varphi(x_2,y_2,z_2)\frac{1}{\sqrt{2}}\{1\times 0-0\times 1\}\right]^2=0.\end{aligned}$$

We ask about the distribution of the electron of the same spin. The answer is that this distribution is *everywhere equal to zero*, i.e. we do not find electron 2 with spin coordinate $\frac{1}{2}$ independently of the position of the electron 1 with spin coordinate $\frac{1}{2}$ (in whatever point on the contour line or beyond it).

The second conclusion can be accepted, since it follows from the pairing of the spins,[2] but the first conclusion is just absurd. Such nonsense is admitted by the Hartree–Fock method. In this chapter we will ponder how we can introduce a correlation of electronic motions.

correlation energy

The correlation of the motions of the electrons results in an energy gain called *a correlation energy* which is defined as

$$E_{\mathrm{corel}} = E - E_{\mathrm{RHF}},$$

where E is the energy from the Schrödinger equation,[3] and E_{RHF} is the Restricted Hartree–Fock energy.[4]

---

[1] We may ask: "How come?" After all, we consider a singlet state, hence the spin projections are opposite. We will not find the situation with parallel spin projections. Take it easy. If, in fact, we are right then we will get 0 as the density of the respective conditional probability. Let us see whether it really will be so.

[2] And this is ensured by the singlet form of the spin part of the function.

[3] This is the rigorous nonrelativistic energy of the system in its ground state. This quantity is not available experimentally, we can *evaluate* it by subtraction of the calculated relativistic corrections from the energy of the total ionization of the system.

[4] Usually we define the correlation energy for the case of double occupancy of the molecular orbitals (the RHF method, see p. 330). In the case of open shells, especially when the multideterminantal description is required, the notion of correlation energy still remains to be defined. These problems will not be discussed in this book.

What is it all about

The outline of the chapter is as follows:

- First we will discuss the methods which explicitly (*via* the form of the suggested wave function) allow the electrons to control their mutual distance ("a correlation of motions").
- In the second part of the chapter the correlation will be less visible, since it will be accounted for by the application of linear combinations of the Slater determinants. First we will discuss the variational methods (VB, CI, MC SCF), and then the non-variational ones (CC, EOM-CC, MBPT).

NON-VARIATIONAL METHODS WITH SLATER DETERMINANTS

- Derivation of the EOM-CC equations

Many body perturbation theory (MBPT) (♦) **p. 551**

- Unperturbed Hamiltonian
- Perturbation theory – slightly different approach
- Reduced resolvent or the "almost" inverse of $(E_0^{(0)} - \hat{H}^{(0)})$
- MBPT machinery
- Brillouin–Wigner perturbation theory
- Rayleigh–Schrödinger perturbation theory

Møller–Plesset version of Rayleigh–Schrödinger perturbation theory (△) **p. 558**

- Expression for MP2 energy
- Convergence of the Møller–Plesset perturbational series
- Special status of double excitations

In the previous chapter we dealt with the description of electronic motion in the mean field approximation. Now *we use this approximation as a starting point towards methods accounting for electron correlation*. Each of the methods considered in this chapter, when rigorously applied, should give an exact solution of the Schrödinger equation. Thus this chapter will give us access to methods providing accurate solutions of the Schrödinger equation.

Why is this important?

Perhaps, in our theories, the electrons do not need to correlate their motion and the results will be still all right?

Unfortunately, this is not so. The mean field method provides, to be sure, ca. 99% of the total energy of the system. This is certainly a lot, in many cases the mean field method gives quite good results, but still falls short of treating several crucial problems correctly. For example:

- *Only through electron correlation* do the noble gas atoms attract each other in accordance with experiment (liquefaction of gases).
- According to the Hartree–Fock method, the F_2 molecule *does not exist* at all, whereas the fact is that it exists, and is doing quite well (bonding energy equal to 38 kcal/mol).[5]
- About *half* the interaction energy of large molecules (often of biological importance) calculated at the equilibrium distance originates purely from the correlation effect.
- The RHF method used to describe the dissociation of the chemical bond gives simply tragic results (cf. Chapter 8, p. 371), *qualitatively wrong* (the UHF method gives a qualitatively correct description).

We see that in many cases electronic correlation must be taken into account.

What is needed?

- Operator algebra (Appendix B, necessary).
- Hartree–Fock method (Chapter 8, necessary).
- Eigenvalue problem (Appendix L, p. 984, necessary).
- Variational method (Chapter 5, necessary).
- Perturbation theory (Chapter 5, recommended).
- Matrix diagonalization (Appendix K, p. 982, recommended).
- Second quantization (Appendix U, p. 1023, necessary).

---

[5] Yet this is not a strong bond. For example, the bonding energy of the H_2 molecule equals 104 kcal/mol, that of HF equals 135 kcal/mol.

Classic papers

The first calculations incorporating electron correlation in an atom (helium) were published by Egil Andersen Hylleraas in an article "*Neue Berechnung der Energie des Heliums im Grundzustande, sowie des tiefsten Terms von Ortho-Helium*", *Zeitschrift für Physik*, 54 (1929) 347. ★ The first calculations with electron correlation for molecules were performed by Walter Heitler and Fritz Wolfgang London in a paper "*Wechselwirkung neutraler Atome und homöopolare Bindung nach der Quantenmechanik*" published in *Zeitschrift für Physik*, 44 (1927) 455. Formation of the covalent bond (in H_2) could be correctly described only after the electron correlation has been included. June 30, 1927, when Heitler and London submitted the paper, is the birth date of quantum chemistry. ★ Later, significantly more accurate results were obtained for the hydrogen molecule by Hubert M. James and Albert S. Coolidge in an article "*The Ground State of the Hydrogen Molecule*", *Journal of the Chemical Physics*, 1 (1933) 825, and a contemporary reference point for that molecule are papers by Włodzimierz Kołos and Lutosław Wolniewicz, among others an article entitled "*Potential Energy Curves for the* $X^1\Sigma_g^+$, $B^3\Sigma_u^+$, $C^1\Pi_u$ *States of the Hydrogen Molecule*" published in *Journal of Chemical Physics*, 43 (1965) 2429. ★ Christian Møller and Milton S. Plesset in *Physical Review*, 46 (1934) 618 published a paper "*Note on an Approximation Treatment for Many-Electron Systems*", where they presented a perturbational approach to electron correlation. ★ The first calculations with the Multi-Configurational Self-Consistent Field (MC SCF) method for atoms was published by Douglas R. Hartree, William Hartree and Bertha Swirles in a paper "*Self-Consistent Field, Including Exchange and Superposition of Configurations, with some Results for Oxygen*", *Philosophical Transactions of the Royal Society (London)*, A 238 (1939) 229, and the general MC SCF theory was presented by Roy McWeeny in a work "*On the Basis of Orbital Theories*", *Proceedings of the Royal Society (London)*, A 232 (1955) 114. ★ As a classic paper in electronic correlation we also consider an article by Per-Olov Löwdin "*Correlation Problem in Many-Electron Quantum Mechanics*" in *Advances in Chemical Physics*, 2 (1959) 207. ★ The idea of the Coupled Cluster (CC) method was introduced by Fritz Coester in a paper in *Nuclear Physics*, 7 (1958) 421 entitled "*Bound States of a Many-Particle System*". Jiří Čížek introduced the (diagrammatic) CC method into electron correlation theory in the paper "*On the Correlation Problem in Atomic and Molecular Systems. Calculation of Wavefunction Components in Ursell-type Expansion Using Quantum-Field Theoretical Methods*" published in the *Journal of Chemical Physics*, 45 (1966) 4256. ★ The book edited by Oktay Sinanoğlu and Keith A. Brueckner "*Three Approaches to Electron Correlation in Atoms*", Yale Univ. Press, New Haven and London, 1970, contains several reprints of the papers which cleared the path towards the coupled-cluster method. ★ A derivation of the coupled cluster equations (for interacting nucleons) was presented by Herman Kümmel and Karl-Heinz Lührmann, *Nuclear Physics*, A191 (1972) 525 in a paper entitled "*Equations for Linked Clusters and the Energy Variational Principle*".

VARIATIONAL METHODS USING EXPLICITLY CORRELATED WAVE FUNCTION

We have learnt, from the example given at the beginning of this chapter, that the "genetic defect" of mean field methods is, that they describe electrons and com-

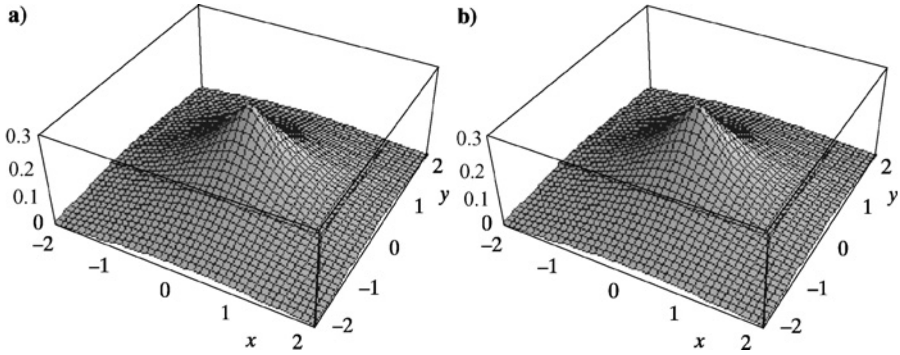


Fig. 10.1. Absence of electronic correlation in the helium atom as seen by the Hartree–Fock method. Visualization of the cross-section of the square of the wave function (probability density distribution) describing electron 2 within the plane xy provided electron 1 is located in a certain point in space: a) at $(-1, 0, 0)$; b) at $(1, 0, 0)$. *Note, that in both cases the conditional probability density distributions of electron 2 are identical.* This means electron 2 does not react to the motion of electron 1, i.e. there is no correlation whatsoever of the electronic motions (when the total wave function is the Hartree–Fock one).

pletely ignore the fact that they are close or far away from each other. For example, in the two-electron case previously considered when we established the coordinates of electron 1, electron 2 has a certain distribution of the probability density. *This distribution does not change when the electron 1 moves to a different position.* This means that the electrons "are not afraid" to get close to each other, although they should, since when electrons are close the energy increases (Fig. 10.1.a,b).

The explicitly correlated wave function (we will get to it in a moment) *has the interelectronic distance built in its mathematical form*. We may compare this to making the electrons wear spectacles.[6] Now they avoid each other. One of my students said that it would be the best if the electrons moved apart to infinity. Well, they cannot. They are attracted by the nucleus (energy gain), and being close to it, are necessarily close to each other too (energy loss). There is a compromise to achieve.

10.1 CORRELATION CUSP CONDITION

Short distances are certainly most important for the Coulombic interaction of two charges, although obviously, the regions of configurational space connected with the long interelectronic distances are much larger. Thus the region is not large, but *very* important, within it "collisions" take place. It turns out that the wave function calculated *in the region of collision* must satisfy some very simple mathematical condition (called *correlation cusp* condition). This is what we want to demonstrate. The derived formulae[7] are universal, they apply to any pair of charged particles.

---

[6] Of course, the methods described further also provide their own "spectacles" (otherwise they would not give the solution of the Schrödinger equation), but the spectacles in the explicitly correlated functions are easier to construct with a small number of parameters.

[7] T. Kato, *Commun. Pure Appl. Math.* 10 (1957) 151.

Let us consider *two* particles with charges q_i and q_j and masses m_i and m_j *separated from other particles.* This, of course, makes sense since simultaneous collisions of three or more particles occur very rarely in comparison with two-particle collisions. Let us introduce a Cartesian system of coordinates (say, in the middle of the beautiful market square in Brussels), so that the system of two particles is described with six coordinates. Then (atomic units are used) the sum of the kinetic energy operators of the particles is

$$\hat{T} = -\frac{1}{2m_i}\Delta_i - \frac{1}{2m_j}\Delta_j. \tag{10.1}$$

Now we separate the motion of the centre of mass of the two particles with position vectors $\boldsymbol{r}_i$ and $\boldsymbol{r}_j$. The centre of mass in our coordinate system is indicated by the vector $\boldsymbol{R}_{\mathrm{CM}} = (X_{\mathrm{CM}}, Y_{\mathrm{CM}}, Z_{\mathrm{CM}})$

$$\boldsymbol{R}_{\mathrm{CM}} = \frac{m_i\boldsymbol{r}_i + m_j\boldsymbol{r}_j}{m_i + m_j}. \tag{10.2}$$

Let us also introduce the total mass of the system $M = m_i + m_j$, the reduced mass of the two particles $\mu = \frac{m_i m_j}{m_i + m_j}$ and the vector of their relative positions $\boldsymbol{r} = \boldsymbol{r}_i - \boldsymbol{r}_j$. Introducing the three coordinates of the centre of mass measured with respect to the market square in Brussels and the three coordinates x, y, z which are components of the vector $\boldsymbol{r}$, we get (Appendix I, Example 1)

$$\hat{T} = -\frac{1}{2M}\Delta_{\mathrm{CM}} - \frac{1}{2\mu}\Delta, \tag{10.3}$$

$$\Delta_{\mathrm{CM}} = \frac{\partial^2}{\partial X_{\mathrm{CM}}^2} + \frac{\partial^2}{\partial Y_{\mathrm{CM}}^2} + \frac{\partial^2}{\partial Z_{\mathrm{CM}}^2}, \tag{10.4}$$

$$\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}. \tag{10.5}$$

After this operation, the Schrödinger equation for the system is separated (as always in the case of two particles, see Appendix I, p. 971) into two equations: the first describing the motion of the centre of mass (seen from Brussels) and the second describing the *relative* motion of the two particles (with Laplacian of x, y, z and reduced mass μ). We are not interested in the first equation, but the second one is what we are after. Let us write down the Hamiltonian corresponding to the second equation

$$\hat{H} = -\frac{1}{2\mu}\Delta + \frac{q_i q_j}{r}. \tag{10.6}$$

We are interested in how the wave function looks when the distance between the two particles r gets very small. If r is small, it makes sense to expand the wave function in a power series[8] of r: $\psi = C_0 + C_1 r + C_2 r^2 + \cdots$. Let us calculate $\hat{H}\psi$

---

[8] Assuming such a form we exclude the possibility that the wave function goes to ±∞ for $r \to 0$. This must be so, since otherwise either the respective probability would go to infinity or the operators

in the vicinity of $r = 0$. The Laplacian expressed in the spherical coordinates is the sum of three terms (Appendix I, p. 971): the first, which contains the differentiation with respect to r and the remaining two, which contain the differentiation with respect to the angles θ and ϕ: $\Delta = \frac{1}{r^2}\frac{\partial}{\partial r}r^2\frac{\partial}{\partial r}+$ terms depending on θ and ϕ. Since we have assumed the function to be dependent on r, upon the action of the Laplacian only the first term gives a nonzero contribution.

We obtain

$$\begin{aligned}\hat{H}\psi &= \left(-\frac{1}{2\mu}\Delta + \frac{q_i q_j}{r}\right)\psi \\ &= 0 - \frac{C_1}{2\mu}\frac{2}{r} + 6C_2 + 12C_3 r + \cdots \\ &\quad + C_0\frac{q_i q_j}{r} + C_1 q_i q_j + C_2 q_i q_j r + \cdots \qquad (10.7)\end{aligned}$$

The wave function cannot go to infinity when r tends to zero, while in the above expression we have two terms which would then "explode" to infinity.

These terms must cancel.

Hence, we obtain

$$C_0 q_i q_j = \frac{C_1}{\mu}. \qquad (10.8)$$

This condition is usually expressed in a different way. We use the fact that $\psi(r = 0) = C_0$ and $(\frac{\partial\psi}{\partial r})_{r=0} = C_1$. We then obtain the cusp condition as

$$\left(\frac{\partial\psi}{\partial r}\right)_{r=0} = \mu q_i q_j \psi(r=0).$$

- The case of two electrons:

  Then $m_i = m_j = 1$, hence $\mu = \frac{1}{2}$ and $q_i = q_j = -1$. We get the cusp condition for the collision of two electrons as

  $$\left(\frac{\partial\psi}{\partial r}\right)_{r=0} = \frac{1}{2}\psi(r=0)$$

  or

  the wave function should be of the form

  $$\psi = \phi(\boldsymbol{r}_1, \boldsymbol{r}_2)\left[1 + \frac{1}{2}r_{12} + \cdots\right],$$

  where $+\cdots$ means higher powers of r_{12}.

---

would become non-Hermitian, cf. p. 73. Both possibilities are unacceptable. We covertly assumed also (to simplify our considerations) that the wave function does not depend on the angles θ and ϕ. This dependence can be accounted for by making the constants C_0, C_1, C_2 the functions of θ and ϕ. Then the final results still holds, but for the coefficients C_0 and C_1 averaged over θ and ϕ.

- The nucleus–electron case:

  When one of the particles is a nucleus of charge Z then $\mu \simeq 1$ and we get

$$\left(\frac{\partial\psi}{\partial r}\right)_{r=0} = -Z\psi(r=0).$$

Thus

> the correct wave function for the electron in the vicinity of a nucleus should have an expansion $\psi = \text{const}(1 - Zr_{a1} + \cdots)$, where r_{a1} is the distance from the nucleus

Let us see how it is with the $1s$ function for the hydrogen-like atom (the nucleus has charge Z) expanded in a Taylor series in the neighbourhood of $r = 0$. We have $1s = N\exp(-Zr) = N(1 - Zr + \cdots)$. It works.

The correlation cusp condition shows that the wave function is not differentiable at $r = 0$.

10.2 THE HYLLERAAS FUNCTION

In 1929, two years after the birth of quantum chemistry, a paper by Hylleraas[9] appeared, where, for the ground state of the helium atom, a trial variational function, containing the interelectronic distance explicitly, was applied. This was a brilliant idea, since it showed that already a small number of terms provide very good results. Even though no fundamental difficulties were encountered for larger atoms, the enormous numerical problems were prohibitive for atoms with larger numbers of electrons. In this case, the progress made from the nineteen twenties to the end of the twentieth century is exemplified by transition from two- to ten-electron systems.

10.3 THE HYLLERAAS CI METHOD

In this method,[10] we exploit the Hylleraas idea in such a way that the electronic wave function is expressed as a linear combinations of Slater determinants, and in front of each determinant $\Phi_i\ (1, 2, 3, \ldots, N)$ we insert, next to the variational coefficient c_i, correlational factors with some powers $(v, u, \ldots)$ of the interelectronic

---

[9] E.A. Hylleraas, *Zeit. Phys.* 54 (1929) 347. Egil Andersen Hylleraas arrived in 1926 in Göttingen, to collaborate with Max Born. His professional experience was related to crystallography and to the optical properties of quartz. When one of the employees fell ill, Born told Hylleraas to continue his work on the helium atom in the context of the newly developed quantum mechanics. The helium atom problem had already been attacked by Albrecht Unsöld in 1927 using first order perturbation theory, but Unsöld obtained the ionization potential equal to 20.41 eV, while the experimental value was equal to 24.59 eV. In the reported calculations (done on a recently installed calculator) Hylleraas obtained a value of 24.47 eV (cf. contemporary accuracy, p. 134).

[10] CI, *Configuration Interaction*.

distances (r_{mn} between electron m and electron n, etc.):

$$\psi = \sum_i c_i \hat{A}\left[r_{mn}^{v_i} r_{kl}^{u_i} \dots \Phi_i(1,2,3,\dots,N)\right], \tag{10.9}$$

where $\hat{A}$ denotes an antisymmetrization operator (see Appendix U, p. 1023). If $v_i = u_i = 0$, we have the CI expansion: $\psi = \sum_i c_i \Phi_i$ (we will discuss it on p. 525). If $v_i \neq 0$, we include a variationally proper treatment of the appropriate distance r_{mn}, i.e. correlation of the motions of the electrons m and n, etc. The antisymmetrization operator ensures fulfilment of the requirement for symmetry of the wave function with respect to the exchange of the arbitrary two electrons. The method described was independently proposed in 1971 by Wiesław Woźnicki[11] and by Sims and Hagstrom.[12] The method of correlational factors has a nice feature, in that even a short expansion should give a very good total energy for the system, since we combine the power of the CI method with the great success of the explicitly correlated approaches. Unfortunately, the method has also a serious drawback. To make practical calculations, it is necessary to evaluate the integrals occurring in the variational method, and they are very difficult to calculate. It is enough to realize that, in the matrix element of the Hamiltonian containing two terms of the above expansion, we may find, e.g., a term $1/r_{12}$ (from the Hamiltonian) and r_{13} (from the factor in front of the determinant), as well as the product of 6 spinorbitals describing the electrons 1, 2, 3. Such integrals have to be computed and the existing algorithms are inefficient.

10.4 THE HARMONIC HELIUM ATOM

An unpleasant feature of the electron correlation is that we deal either with intuitive concepts or, if our colleagues want to help us, they bring wave functions with formulae as long as the distance from Cracow to Warsaw (or longer[13]) and say: look, this is what *really* happens. It would be good to analyze such formulae term by term, but this does not make sense, because there are too many terms. Even the helium atom, when we write down the formula for its ground-state wave function, becomes a mysterious object. Correlation of motion of whatever seems to be so difficult to grasp mathematically that we easily give up. A group of scientists published a paper in 1993 which aroused enthusiasm.[14] They obtained a rigorous solution of the Schrödinger equation (described in Chapter 4, p. 188), the only exact solution which has been obtained so far for correlational problems.

---

[11] W. Woźnicki, in "*Theory of Electronic Shells in Atoms and Molecules*" (ed. A. Yutsis), Mintis, Vilnius, 1971, p. 103.

[12] J.S. Sims, S.A. Hagstrom, *Phys. Rev. A*4 (1971) 908. This method is known as a Hylleraas–CI.

[13] This is a very conservative estimate. Let us calculate – half jokingly. Writing down a single Slater determinant would easily take 10 cm. The current world record amounts to several billion such determinants in the CI expansion. Say, three billion. Now let us calculate: $10 \text{ cm} \times 3 \times 10^9 = 3 \times 10^{10} \text{ cm} = 3 \times 10^8 \text{ m} = 3 \times 10^5 \text{ km} = 300000 \text{ km}$. So, this not Warsaw to Cracow, but Earth to Moon.

[14] S. Kais, D.R. Herschbach, N.C. Handy, C.W. Murray, G.J. Laming, *J. Chem. Phys.* 99 (1993) 417.

Note that the exact wave function (its spatial part[15]) is a *geminal* (i.e. two-electron function).

$$\psi(\boldsymbol{r}_1,\boldsymbol{r}_2)=N\left(1+\frac{1}{2}r_{12}\right)e^{-\frac{1}{4}(r_1^2+r_2^2)}. \tag{10.10}$$

Let me be naive. Do we have two harmonic springs here? Yes, we do. Then, let us treat them first as independent oscillators and take the *product* of the ground-state functions of both oscillators: $\exp[-\frac{1}{4}(r_1^2+r_2^2)]$. Well, it would be good to account for the cusp condition $\psi=\phi(\boldsymbol{r}_1,\boldsymbol{r}_2)[1+\frac{1}{2}r_{12}+\cdots]$ and take care of it even in a naive way. Let us just implement the crucial correlation factor $(1+\frac{1}{2}r_{12})$, *the simplest* that satisfies the cusp condition (see p. 505). It turns out, that such a recipe leads to a *rigorous* wave function.[16]

From (10.10) we see that for $r_1=r_2=\text{const}$ (in such a case both electrons move on the surface of the sphere), the larger value of the function (and *eo ipso* of the probability) is obtained for *larger* r_{12}. This means that, it is most probable that the electrons prefer to occupy opposite sides of a nucleus. This is a practical manifestation of the existence of the Coulomb hole around electrons, i.e. the region of the reduced probability of finding a second electron: the electrons simply repel each other. They cannot move apart to infinity since both are held by the nucleus. The only thing they can do is to be close to the nucleus and to avoid each other – this is what we observe in (10.10).

10.5 JAMES–COOLIDGE AND KOŁOS–WOLNIEWICZ FUNCTIONS

One-electron problems are the simplest. For systems with *two* electrons[17] we can apply certain mathematical tricks which allow very accurate results. We are going to talk about such calculations in a moment.

Kołos and Wolniewicz applied the Ritz variational method (see Chapter 5) to the hydrogen molecule with the following trial function:

$$\Psi=\frac{1}{\sqrt{2}}\big[\alpha(1)\beta(2)-\alpha(2)\beta(1)\big]\sum_i^M c_i\big(\Phi_i(1,2)+\Phi_i(2,1)\big), \tag{10.11}$$

$$\Phi_i(1,2)=\exp\big(-A\xi_1-\bar{A}\xi_2\big)\xi_1^{n_i}\eta_1^{k_i}\xi_2^{m_i}\eta_2^{l_i}\left(\frac{2r_{12}}{R}\right)^{\mu_i}$$
$$\cdot\big(\exp\big(B\eta_1+\bar{B}\eta_2\big)+(-1)^{k_i+l_i}\exp\big(-B\eta_1-\bar{B}\eta_2\big)\big),$$

---

[15] For one- and two-electron systems the wave function is *a product* of the spatial and spin factors. A normalized spin factor for two-electron systems, $\frac{1}{\sqrt{2}}\{\alpha(1)\beta(2)-\beta(1)\alpha(2)\}$, guarantees that the state in question is a singlet (see Appendix Q, p. 1006). Since we will only manipulate the spatial part of the wave function, the spin is the default. Since the total wave function has to be antisymmetric, and the spin function is antisymmetric, the spatial function should be symmetric and it is.

[16] As a matter of fact, only for a single force constant. Nevertheless, the unusual simplicity of that analytic formula is most surprising.

[17] For a larger number of electrons it is much more difficult.

where the elliptic coordinates of the electrons with index $j = 1, 2$ are given by:

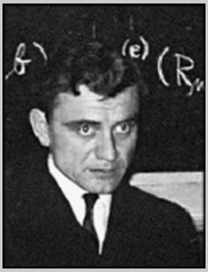
$$\xi_j = \frac{r_{aj} + r_{bj}}{R}, \tag{10.12}$$

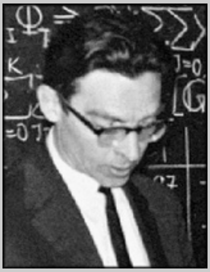
$$\eta_j = \frac{r_{aj} - r_{bj}}{R}, \tag{10.13}$$

R denotes the internuclear distance, r_{aj} and r_{bj} are nucleus–electron distances (the nuclei are labelled by a, b), r_{12} is the (crucial to the method) interelectronic distance, c_i, A, $\bar{A}$, B, $\bar{B}$ are variational parameters, and n, k, l, m are integers.

The simplified form of this function with $A = \bar{A}$ and $B = \bar{B} = 0$ is the James–Coolidge[18] function, thanks to which the later authors enjoyed the most accurate result for the hydrogen molecule in 27 years.

Włodzimierz Kołos (1928–1996), Polish chemist, professor at the Warsaw University. His calculations on small molecules (with Roothaan, Wolniewicz, Rychlewski) took into account all known effects and were of unprecedented accuracy in quantum chemistry. The Department of Chemistry of Warsaw University and the Polish Chemical Society established the Włodzimierz Kołos Medal accompanying a Lecture (the first lecturers were: Roald Hoffmann, Richard Bader and Paul von Ragué Schleyer). In the Ochota quarter in Warsaw there is a Włodzimierz Kołos Street. Lutosław Wolniewicz (born 1927), Polish physicist, professor at the Nicolaus Copernicus University in Toruń.





Kołos and Roothaan,[19] and later on, Kołos and Wolniewicz[20] as well as Kołos and Rychlewski and others[21] applied longer and longer expansions (computer technology was improving fast) up to M of the order of thousands. The results obtained exceeded the accuracy of experiments, although the latter represented one of the most accurate spectroscopic measurements ever done. Owing to the great precision of these calculations it was proved that quantum mechanics, and in particular the Schrödinger equation, describe the reality with remarkable accuracy, Tables 10.1 and 10.2.

As can be seen from Tables 10.1 and 10.2, there was a competition between theoreticians and the experimental laboratory of Herzberg. When, in 1964, Kołos and Wolniewicz obtained 36117.3 cm^{-1} (Table 10.1, bold face) for the dissociation energy of the hydrogen molecule, quantum chemists held their breath. The experimental result of Herzberg and Monfils, obtained four years earlier (Table 10.1, bold face), was

---

[18] H.M. James, A.S. Coolidge, *J. Chem. Phys.* 1 (1933) 825. Hubert M. James in the sixties was professor at Purdue University (USA).

[19] W. Kołos, C.C.J. Roothaan, *Rev. Modern Phys.* 32 (1960) 205.

[20] For the first time in quantum chemical calculations relativistic corrections and corrections resulting from quantum electrodynamics were included. This accuracy is equivalent to hitting, from Earth, an object on the Moon the size of a car. These results are cited in nearly all textbooks on quantum chemistry to demonstrate that the theoretical calculations have a solid background.

[21] The description of these calculations is given in the review article by Kołos cited in Table 10.1.

Table 10.1. Dissociation energy of H_2 in the ground state (in cm^{-1}). Comparison of the results of theoretical calculations and experimental measurements. The references to the cited works can be found in the paper by W. Kołos, *Pol. J. Chem.* 67 (1993) 553. Bold numbers are used to indicate the values connected with the Herzberg–Kołos–Wolniewicz controversy

Year	Author	Experiment	Theory
1926	Witmer	35000	
1927	Heitler–London		23100[a)]
1933	James–Coolidge		36104[a)]
1935	Beutler	36116 ± 6	
1960	Kołos–Roothaan		36113.5[a)]
1960	Herzberg–Monfils	**36113.6 ± 0.6**	
1964	Kołos–Wolniewicz		**36117.3**[a)]
1968	Kołos–Wolniewicz		**36117.4**[a)]
1970	Herzberg	**36118.3**[c)]	
1970	Stwalley	36118.6 ± 0.5	
1975	Kołos–Wolniewicz		36118.0
1978	Kołos–Rychlewski		36118.12[b)]
1978	Bishop–Cheung		36117.92
1983	Wolniewicz		36118.01
1986	Kołos–Szalewicz–Monkhorst		36118.088
1991	McCormack–Eyler	36118.26 ± 0.20	
1992	Balakrishnan–Smith–Stoicheff	36118.11 ± 0.08	
1992	Kołos–Rychlewski		36118.049

[a)] Obtained from calculated binding energy by subtracting the energy of zero vibrations.

[b)] Obtained by treating the improvement of the binding energy as an additive correction to the dissociation energy.

[c)] Upper bound.

Table 10.2. Ionization energy of H_2 (in cm^{-1}). See the caption for Table 10.1

Year	Author	Experiment	Theory
1934	Richardson	124569.2	
1933	James–Coolidge		124438
1938	Beutler–Jünger	124429±13	
1969	Jeziorski–Kołos		124417.3
1969	Takezawa	124417±2	
1970	Takezawa	124417.4 ± 0.6	
1972	Herzberg–Jungen	124417.2 ± 0.4	
1978	Kołos–Rychlewski		124417.44
1986	Jungen–Herzberg	124417.5 ± 0.1	
1986/7	Eyler–Short–Pipkin	124417.42 ± 0.15	
1987	Glab–Hessler	124417.61 ± 0.07	
1989	McCormack–Gilligan–Comaggia–Eyler	124417.524±0.015	
1990	Jungen–Dabrowski–Herzberg–Vervloet	124417.501 ± 0.015	
1992	Gilligan–Eyler	124417.507 ± 0.018	
1992	Jungen–Dabrowski–Herzberg–Vervloct	124417.484 ± 0.017	
1992	Eyler et al.	124417.507 ± 0.012	
1992	Kołos–Rychlewski		124417.471

higher and this seemed to contradict the variational principle (Chapter 5) a foundation of quantum mechanics. There were only three possibilities:

- the theoretical result is wrong,
- the experimental result is wrong,
- quantum mechanics has internal inconsistency.

Kołos and Wolniewicz increased the accuracy of their calculations in 1968 and excluded the first possibility. It soon turned out that the problem lay in the accuracy of the experiment.[22] When Herzberg increased the accuracy, he obtained 36118.3 cm^{-1} as the dissociation energy (Table 10.1, bold face), which was then consistent with the variational principle.

Nowadays, these results are recognized in the world as the most reliable source of information on small molecules. For example, Kołos and Wolniewicz's results for the H_2 molecule were used to estimate the hydrogen concentration on Jupiter.

Gerhard Herzberg (1904–1999), Canadian chemist of German origin professor at the National Research Council and at the University of Saskatchewan in Saskatoon and the University of Ottawa. The greatest spectroscopist of the XX century. Herzberg laid the foundations of molecular spectroscopy, is author of the fundamental monograph on this subject, received a Nobel prize in 1971 "*for his contribution to knowledge of the electronic structure and geometry of molecules, particularly free radicals*".

10.5.1 NEUTRINO MASS

Calculations like those above required unique software, especially in the context of the non-adiabatic effects included. Additional gains appeared unexpectedly, when Kołos and others[23] initiated work aiming at explaining whether the electronic neutrino has a non-zero mass or not.[24] In order to interpret the expensive experiments,

---

[22] At that time Herzberg was hosting them in Canada and treated them to a home made fruit liquor, the latter event was considered by his coworkers to be absolutely exceptional. This is probably the best time to give the recipe for this exquisite drink which is known in the circles of quantum chemists as "kolosovka".

Pour a pint of pure spirits into a beaker. Hang an orange on a piece of gauze directly over the meniscus. Cover tightly and wait for two weeks. Then throw the orange away – there is nothing of value left in it, and turn your attention to the spirits. It should contain now all the flavours from orange. Next, slowly pour some spring water until the liquid becomes cloudy and some spirits to make it clear again. Propose a toast to the future of quantum chemistry!

[23] W. Kołos, B. Jeziorski, H.J. Monkhorst, K. Szalewicz, *Int. J. Quantum Chem.* S19 (1986) 421.

[24] Neutrinos are stable fermions of spin $\frac{1}{2}$. Three types of neutrinos exist (each has its own antiparticle): electronic, muonic and taonic. The neutrinos are created in the weak interactions (e.g., in β-decay) and do not participate either in the strong, or in electromagnetic interactions. The latter feature expresses itself in an incredible ability to penetrate matter (e.g., crossing the Earth almost as through a vacuum). The existence of the electronic neutrino was postulated in 1930 by Wolfgang Pauli and discovered in 1956 by F. Reines and C.L. Cowan; the muonic neutrino was discovered in 1962 by L. Lederman, M. Schwartz and J. Steinberger.

Alexandr Alexandrovich Friedmann (1888–1925), Russian mathematician and physicist, in his article in *Zeit. Phys.* 10 (1922) 377 proved on the basis of Einstein's general theory of relativity, that the curvature of the Universe must change, which became the basis of cosmological models of the expanding Universe. During World War I, Friedman was a pilot in the Russian army and made bombing raids over my beloved Przemyśl. In one of his letters he asked his friend cheerfully, the eminent Russian mathematician Steklov, for advice about the integration of equations he derived to describe the trajectories of his bombs. Later, in a letter to Steklov of February 28, 1915 he wrote: "*Recently I had an opportunity to verify my theory during a flight over Przemyśl, the bombs fell exactly in the places predicted by the theory. To get the final proof of my theory I intend to test it in flights during next few days.*"

More information in: http://www-groups.dcs.st-and.ac.uk/~history/Mathematicians/Friedmann.html

Edwin Powell Hubble (1889–1953), American astronomer, explorer of galaxies, found in 1929, that the distance between galaxies is proportional to the infrared shift in their spectrum caused by the Doppler effect, which is consequently interpreted as expansion of the Universe. A surprise from recent astronomical studies is that the expansion is faster and faster (for reasons unknown).

precise calculations were required for the β-decay of the tritium molecule as a function of the neutrino mass. The emission of the antineutrino (ν) in the process of β-decay:

$$T_2 \rightarrow HeT^+ + e + \nu$$

should have consequences for the final quantum states of the HeT^+ molecule. To enable evaluation of the neutrino mass by the experimentalists Kołos et al. performed precise calculations of all possible final states of HeT^+ and presented them as a function of the hypothetical mass of the neutrino. There is such a large number of neutrinos in the Universe that, if its mass exceeded a certain value, even a very small threshold value of the order of 1 eV,[25] the mass of the Universe would exceed the critical value predicted by Alexandr Friedmann in his cosmological theory (based on the general theory of relativity of Einstein). This would mean that the currently occurring expansion of the Universe (discovered by Hubble) would finally stop and its collapse would follow. If the neutrino mass would turn out to be too small, then the Universe would continue its expansion. Thus, quantum chemical calculations for the HeT^+ molecule may turn out to be helpful in predicting our fate (unfortunately, being crushed or frozen). So far, the estimate of neutrino mass gives a value smaller than 1 eV, which indicates the Universe expansion.[26]

---

[25] The mass of the elementary particle is given in the form of its energetic equivalent mc^2.

[26] At this moment there are other candidates for contributing significantly to the mass of the Universe, mainly the mysterious "dark matter". This constitutes the major part of the mass of the Universe. We know veeeery little.

Recently it turned out that neutrinos undergo what are called oscillations, e.g., an electronic neutrino travels from the Sun and on its way spontaneously changes to a muonic neutrino. The oscillations indicate that the mass of the neutrino is nonzero. According to current estimations, it is much smaller, however, than the accuracy of the tritium experiments.

10.6 METHOD OF EXPONENTIALLY CORRELATED GAUSSIAN FUNCTIONS

In 1960, Boys[27] and Singer[28] noticed that the functions which are products of Gaussian orbitals and correlational factors of Gaussian type, $\exp(-br_{ij}^2)$, where r_{ij} is the distance between electron i and electron j, generate relatively simple integrals in the quantum chemical calculations. A product of two Gaussian orbitals (with positions shown by the vectors $\boldsymbol{A}, \boldsymbol{B}$) and of an exponential correlation factor is called *an exponentially correlated Gaussian geminal*:[29]

geminal

$$g(\boldsymbol{r}_i, \boldsymbol{r}_j; \boldsymbol{A}, \boldsymbol{B}, a_1, a_2, b) = Ne^{-a_1(\boldsymbol{r}_i-\boldsymbol{A})^2}e^{-a_2(\boldsymbol{r}_j-\boldsymbol{B})^2}e^{-br_{ij}^2}.$$

A geminal is an analogue of an orbital, which is a one-electron function. Here is a two-electron one. A single geminal is very rarely used in computations,[30] we apply hundreds or even thousands of Gaussian geminals. When we want to find out what are the optimal positions $\boldsymbol{A}, \boldsymbol{B}$ and the optimal exponents a and b in these thousands of geminals, it turns out that nothing sure is known about them, the $\boldsymbol{A}, \boldsymbol{B}$ positions are scattered chaotically,[31] and in the $a > 0$ and $b > 0$ exponents, there is no regularity either. Nevertheless, the above formula for a single Gaussian geminal looks like if it suggested $b > 0$.

10.7 COULOMB HOLE ("CORRELATION HOLE")

It is always good to count "on fingers" to make sure that everything is all right. Let us see how a *single* Gaussian geminal describes the correlation of the electronic motion. Let us begin with the helium atom with the nucleus in the position $\boldsymbol{A} = \boldsymbol{B} = \boldsymbol{0}$. The geminal takes the form:

$$g_{\mathrm{He}} = Ne^{-a_1r_1^2}e^{-a_1r_2^2}e^{-br_{12}^2}, \tag{10.14}$$

where N is a normalization factor. Let us assume[32] that electron 1 is at $(x_1, y_1, z_1) = (1, 0, 0)$. Where in such situation does electron 2 prefer to be? We will find out (Fig. 10.2) from the position of electron 2 for which g_{He} assumes the largest value.

Just to get an idea, let us try to restrict the motion of electron 2. For instance, let us demand that it moves only on the sphere of radius equal to 1 centred at the nucleus. So we insert $r_1 = r_2 = 1$. Then, $g_{\mathrm{He}} = \mathrm{const}\exp[-br_{12}^2]$ and we will easily find out what electron 2 likes most. With $b > 0$ the latter factor tells us that

---

[27] S.F. Boys, *Proc. Royal Soc. A* 258 (1960) 402.

[28] K. Singer, *Proc. Royal Soc. A* 258 (1960) 412.

[29] This is an attempt to go beyond the two-electron systems with the characteristic (for these systems) approach of James, Coolidge, Hylleraas, Kołos, Wolniewicz and others.

[30] Ludwik Adamowicz introduced an idea of the minimal basis of the Gaussian geminals (equal to the number of the electron pairs) and applied to the LiH and HF molecules, L. Adamowicz, A.J. Sadlej, *J. Chem. Phys.* 69 (1978) 3992.

[31] The methods in which those positions are selected at random scored a great success.

[32] We use atomic units.

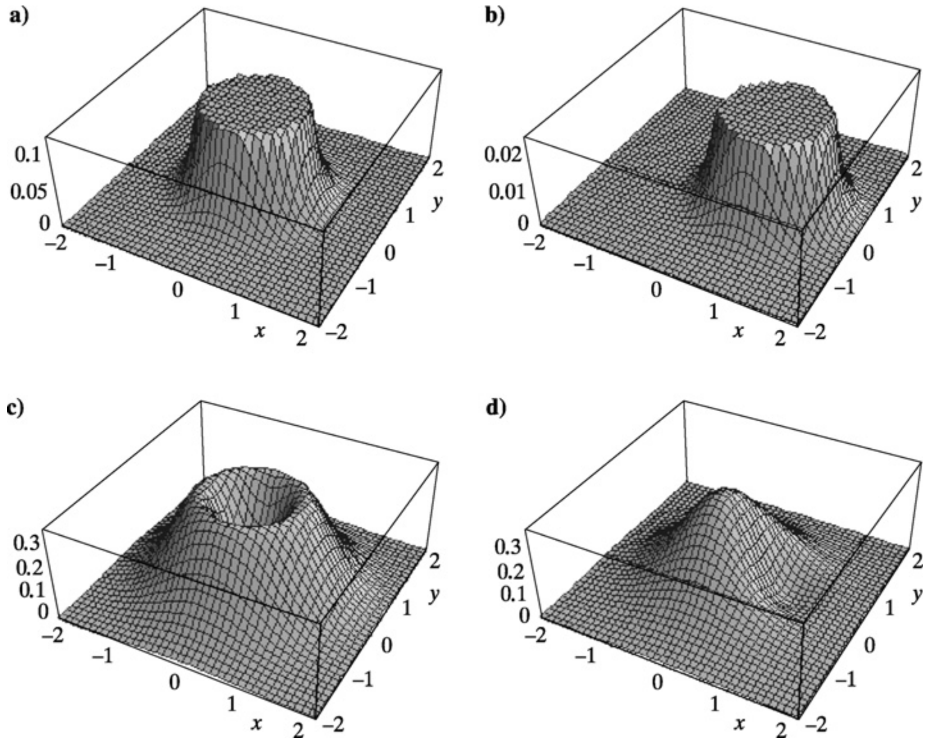


Fig. 10.2. Illustration of the correlation and … anticorrelation of the electrons in the helium atom. Figs. (a) and (b) present the machinery of the "anticorrelation" connected with the geminal $g_{\mathrm{He}} = N\exp[-r_1^2]\exp[-r_2^2]\exp[-2r_{12}^2]$. In Fig. (a) electron 1 has a position $(0, 0, 0)$, while Fig. (b) corresponds to electron 1 being at point $(1, 0, 0)$ (cutting off the top parts of the plots is caused by graphical limitations, not by the physics of the problem). It can be seen that electron 2 *holds on to electron 1*, i.e. it behaves in a completely unphysical manner (since electrons repel each other). Figs. (c) and (d) show how electron 2 will respond to such two positions of electron 1, if the wave function is described by the geminal $g_{\mathrm{He}} = N\exp[-r_1^2]\exp[-r_2^2][1 - \exp[-2r_{12}^2]]$. In Fig. (c) we see that electron 2 runs away "with all its strength" (the hollow in the middle) from electron 1 placed at $(0, 0, 0)$. We have correlation. Similarly, Fig. (d), if electron 1 is in point $(1, 0, 0)$, then it causes a slight depression for electron 2 in this position. Again we do have correlation. However, the graphs (c) and (d) differ widely. This is understandable since the nucleus is all the time at the point $(0, 0, 0)$. Figs. (e), (f) correspond to the same displacements of electron 1, but this time the correlation function is equal to $\psi(\boldsymbol{r}_1, \boldsymbol{r}_2) = (1 + \frac{1}{2}r_{12})\exp[-(r_1^2 + r_2^2)]$, i.e. is similar to the wave function of the harmonic helium atom. It can be seen (particularly in Fig. (e)) that there *is* a correlation, although much less visible than in the previous examples. To amplify (artificially) the correlation effect Figs. (g), (h) show the same as Figs. (e), (f) but for the function $\psi(\boldsymbol{r}_1, \boldsymbol{r}_2) = (1 + 25r_{12})\exp[-(r_1^2 + r_2^2)]$, which (unlike Figs. (e), (f)) does not satisfy the correlation cusp condition.

what electron 2 likes best is just to sit on electron 1! Is it what the correlation is supposed to mean that one electron sits on the other? Here we have rather an anticorrelation. Something is going wrong. According to this analysis we should rather take the geminal of the form, e.g.:

$$g_{\mathrm{He}} = Ne^{-a_1 r_1^2} e^{-a_1 r_2^2}\left[1 - e^{-b r_{12}^2}\right].$$

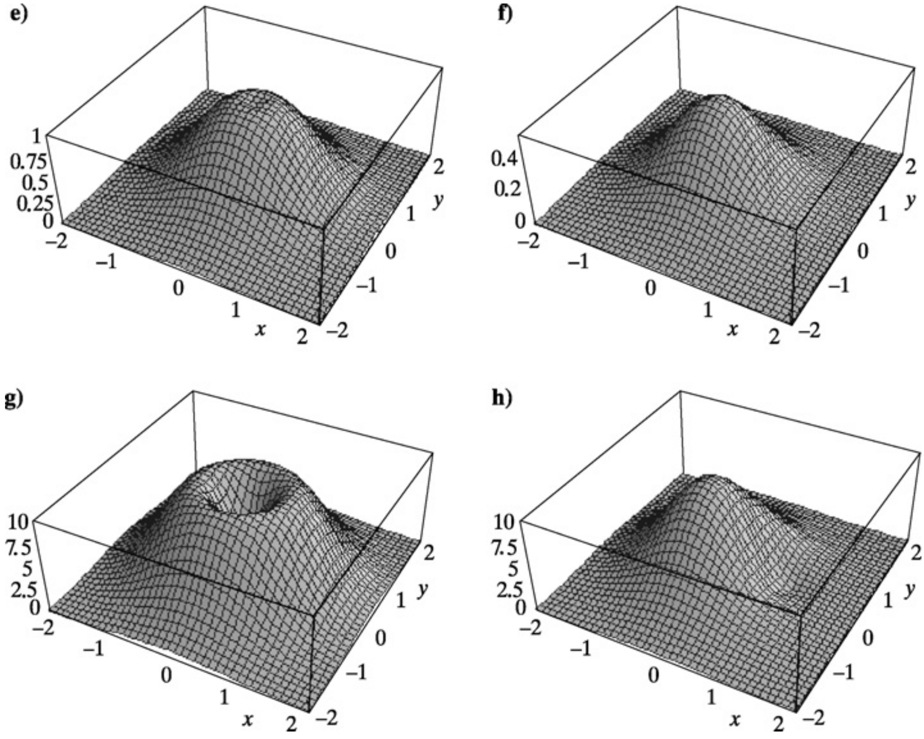


Fig. 10.2. *Continued.*

Now everything is qualitatively in order. When the interelectronic distance increases, the value of the g_{He} function also increases, which means that such a situation is *more* probable than that corresponding to a short distance. If the electrons become too agitated and begin to think that it would be better when their distance gets very long, they would be called to order by the factors $\exp[-a_1 r_1^2]\exp[-a_1 r_2^2]$. Indeed, in such a case, the distance between the nucleus and at least one of the electrons is long and the probability of such a situation is quenched by one or both exponential factors. For large r_{12} distances, the factor $[1-\exp[-br_{12}^2]]$ is practically equal to 1. This means that for large interelectronic distances g_{He} is practically equal to $N\exp[-a_1 r_1^2]\exp[-a_1 r_2^2]$, i.e. to the product of the orbitals (no correlation of motions at long interelectronic distances, and rightly so).

> Around electron 1 there is a region of low probability of finding electron 2. This region is called the Coulomb hole.

The Gaussian geminals do not satisfy the correlation cusp condition (p. 505), because of factor $\exp(-br_{ij}^2)$. It is required (for simplicity we write $r_{ij}=r$) that $(\frac{\partial g}{\partial r})_{r=0} = \frac{1}{2} g(r=0)$, whereas the left-hand side is equal to 0, while the right-hand side $\frac{1}{2} N\exp[-a_1(\boldsymbol{r}_i - \boldsymbol{A})^2]\exp[-a_2(\boldsymbol{r}_j - \boldsymbol{B})^2]$ is not equal to zero. This is not a disqualifying feature, since the region of space in which this condition should be fulfilled, is very small.

The area of application of this method is – for practical (computational) reasons – relatively small. The method of Gaussian geminals has been applied in unusually accurate calculations for three- and four-electron systems.[33]

10.8 EXCHANGE HOLE ("FERMI HOLE")

The mutual avoidance of electrons in helium atom or in hydrogen molecule is caused by Coulombic repulsion of electrons ("Coulomb hole", see above). As we have shown in this Chapter, in the Hartree–Fock method the Coulomb hole is absent, whereas methods which account for electron correlation generate such a hole. However, electrons avoid each other not only because of their charge. The Pauli principle is an additional reason. One of the consequences is the fact that electrons with the same spin coordinate cannot reside in the same place, see p. 33. The continuity of the wave function implies that the probability density of them staying *in the vicinity* of each other is small, i.e.

> around the electron there is a NO PARKING area for other electrons with the same spin coordinate ("exchange, or Fermi hole").

Let us see how such exchange holes arise. We will try to make the calculations as simple as possible.

We have shown above that the Hartree–Fock function does not include any electron correlation. We must admit, however, that we have come to this conclusion on the basis of the two-electron closed shell case. This is a special situation, since both electron have *different* spin coordinates ($\sigma = \frac{1}{2}$ and $\sigma = -\frac{1}{2}$). Is it really true that the Hartree–Fock function does not include any correlation of electronic motion?

We take the H_2^- molecule in the simplest formulation of the LCAO MO method (two atomic orbitals only: $1s_a = \chi_a$ and $1s_b = \chi_b$, two molecular orbitals: bonding $\varphi_1 = \frac{1}{\sqrt{2(1+S)}}(\chi_a + \chi_b)$ and antibonding $\varphi_2 = \frac{1}{\sqrt{2(1-S)}}(\chi_a - \chi_b)$, cf. p. 371; the overlap integral $S \equiv (\chi_a|\chi_b)$). We have three electrons. As a wave function we will take the single (normalized) Hartree–Fock determinant (UHF) with the following orthonormal spinorbitals occupied: $\phi_1 = \varphi_1\alpha$, $\phi_2 = \varphi_1\beta$, $\phi_3 = \varphi_2\alpha$:

$$\psi_{\mathrm{UHF}}(1,2,3) = \frac{1}{\sqrt{3!}} \begin{vmatrix} \phi_1(1) & \phi_1(2) & \phi_1(3) \\ \phi_2(1) & \phi_2(2) & \phi_2(3) \\ \phi_3(1) & \phi_3(2) & \phi_3(3) \end{vmatrix}.$$

Example 1

We are interested in electron 3 with electron 1 residing at nucleus *a* with space coordinates (0, 0, 0) and with spin coordinate $\sigma_1 = \frac{1}{2}$ and with electron 2 located at

---

[33] W. Cencek, Ph.D. Thesis, Adam Mickiewicz University, Poznań, 1993, also J. Rychlewski, W. Cencek, J. Komasa, *Chem. Phys. Letters* 229 (1994) 657; W. Cencek, J. Rychlewski, *Chem. Phys. Letters* 320 (2000) 549. All these results were world records.

nucleus b with coordinates $(R,0,0)$ and $\sigma_2=-\frac{1}{2}$, whereas the electron 3 itself has spin coordinate $\sigma_3=\frac{1}{2}$. The square of the absolute value of the function calculated for these values depends on x_3, y_3, z_3 and represents the *conditional probability* density distribution for finding electron 3 (provided electrons 1 and 2 have the fixed coordinates given above and denoted by $1_0, 2_0$). So let us calculate individual elements of the determinant $\psi_{\mathrm{UHF}}(1_0,2_0,3)$, taking into account the properties of spin functions α and β (cf. p. 28):

$$\psi_{\mathrm{UHF}}(1_0,2_0,3)=\frac{1}{\sqrt{3!}}\begin{vmatrix}\varphi_1(0,0,0) & 0 & \varphi_1(x_3,y_3,z_3)\\ 0 & \varphi_1(R,0,0) & 0\\ \varphi_2(0,0,0) & 0 & \varphi_2(x_3,y_3,z_3)\end{vmatrix}.$$

Using the Laplace expansion (Appendix A on p. 889) we get

$$\psi_{\mathrm{UHF}}(1_0,2_0,3)=\frac{1}{\sqrt{3!}}\big[\varphi_1(0,0,0)\varphi_1(R,0,0)\varphi_2(x_3,y_3,z_3)$$
$$-\varphi_1(x_3,y_3,z_3)\varphi_1(R,0,0)\varphi_2(0,0,0)\big].$$

The plot of this function (the overlap integral S is included in normalization factors of the molecular orbitals) is given in Fig. 10.3.

Qualitatively, however, everything is clear even without the calculations. Due to the forms of the molecular orbitals (S is small) $\varphi_1(0,0,0)=\varphi_1(R,0,0)\approx\varphi_2(0,0,0)=\mathrm{const}$ we get:

$$\psi_{\mathrm{UHF}}(1_0,2_0,3)\approx-\mathrm{const}^2\frac{1}{\sqrt{3}}\chi_b(3)$$

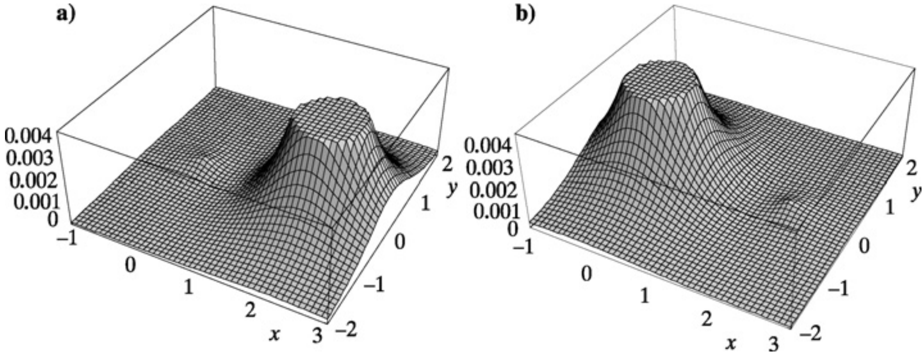


Fig. 10.3. Demonstration of the exchange ("Fermi") hole in the H_2^- molecular ion (truncation of the hills is artificial, without this it would be more difficult to see the details of the figure). (a) $|\psi_{\mathrm{UHF}}(1_0,2_0,3)|^2$ is the probability density of finding the spatial coordinates of electron 3 (having $\sigma_3=\frac{1}{2}$) provided that electron 1 resides on the nucleus a at $(0,0,0)$ having $\sigma_1=\frac{1}{2}$ and electron 2 sits on nucleus b at $(R=2,0,0)$ and has $\sigma_2=-\frac{1}{2}$; (b) the same as above, but this time electron 1 has moved to nucleus b (i.e. it shares b with electron 2).

so the conditional probability density of finding electron 3 is

$$\rho(3) \approx \frac{1}{3}\mathrm{const}^4\big[\chi_b(3)\big]^2. \tag{10.15}$$

We can see that *for some reason electron 3 has moved in the vicinity of nucleus* *b*. What scared it so much, when we placed one of the two electrons at *each* nucleus? Electron 3 ran to be as far away as possible from electron 1 residing on *a*. It hates electron 1 so much that it has just ignored the Coulomb repulsion with electron 2 sitting on *b* and jumped on it! What the hell has happened? Well, we have some suspicions. Electron 3 could have been scared only by the spin coordinate of electron 1, *the same as its own*.

This is just an indication of the exchange hole around each electron.

Example 2

Maybe electron 3 does not run away from anything, but simply always resides at nucleus b? Let us make sure of that. Let us move electron 1 to nucleus b (there is already electron 2 sitting over there, but it does not matter). What then will electron 3 do? Let us see. We have electrons 1 and 2 at nucleus b with space coordinates $(R,0,0)$ and spin coordinates $\sigma_1=\frac{1}{2}$, $\sigma_2=-\frac{1}{2}$, whereas electron 3 has spin coordinate $\sigma_3=\frac{1}{2}$. To calculate the conditional probability we have to calculate the value of the wave function.

This time

$$\psi_{\mathrm{UHF}}(1_0,2_0,3)=\frac{1}{\sqrt{3!}}\begin{vmatrix}\varphi_1(R,0,0) & 0 & \varphi_1(x_3,y_3,z_3)\\ 0 & \varphi_1(R,0,0) & 0\\ \varphi_2(R,0,0) & 0 & \varphi_2(x_3,y_3,z_3)\end{vmatrix}$$

$$\approx \mathrm{const}^2\frac{1}{\sqrt{3}}\chi_a(3)$$

or

$$\rho(3) \approx \frac{1}{3}\mathrm{const}^4\big[\chi_a(3)\big]^2. \tag{10.16}$$

We see that electron 3 with spin coordinate $\sigma_3=\frac{1}{2}$ runs in panic to nucleus a, because it is as scared of electron 1 with spin $\sigma_1=\frac{1}{2}$ as the devil is of holy water.

Example 3

And what would happen if we made the decision for electron 3 more difficult? Let us put electron 1 ($\sigma_1=\frac{1}{2}$) *in the centre* of the molecule and electron 2 ($\sigma_2=-\frac{1}{2}$) as before, at nucleus b. According to what we think about the whole machinery, electron 3 (with $\sigma_3=\frac{1}{2}$) should run away from electron 1, because both electrons have the same spin coordinates, and this is what they hate most. But *where* should it run? Will electron 3 select nucleus a or nucleus b? The nuclei do not look equivalent. There is an electron sitting at b, while the a centre is empty. Maybe electron 3 will jump to a then? Well, the function analyzed is Hartree–Fock – electron 3 ignores the Coulomb hole (it does not see electron 2 sitting on b) and therefore will not prefer the empty nucleus a to sit at. It looks like electron 3 will treat both nuclei on the same basis. In the case of two atomic orbitals, electron 3 has only the choice: either bonding orbital φ_1 or antibonding orbital φ_2 (in both situations the electron densities on a and on b are equal, no nucleus is distinguished). Out of the two molecular orbitals, φ_2 looks much more attractive to electron 3, because it has a node[34] exactly, where electron 1 with its nasty spin is. This means that there is a chance for electron 3 to take care of the Fermi hole of electron 1: we predict that electron 3 will “select” only φ_2. Let us check this step by step:

$$\begin{aligned}
\psi_{\mathrm{UHF}}(1_0, 2_0, 3) &= \frac{1}{\sqrt{3!}} \begin{vmatrix} \varphi_1\left(\frac{R}{2}, 0, 0\right) & 0 & \varphi_1(x_3, y_3, z_3) \\ 0 & \varphi_1(R, 0, 0) & 0 \\ \varphi_2\left(\frac{R}{2}, 0, 0\right) & 0 & \varphi_2(x_3, y_3, z_3) \end{vmatrix} \\
&= \frac{1}{\sqrt{3!}} \begin{vmatrix} \varphi_1\left(\frac{R}{2}, 0, 0\right) & 0 & \varphi_1(x_3, y_3, z_3) \\ 0 & \varphi_1(R, 0, 0) & 0 \\ 0 & 0 & \varphi_2(x_3, y_3, z_3) \end{vmatrix} \\
&= \frac{1}{\sqrt{3!}} \varphi_1\left(\frac{R}{2}, 0, 0\right) \varphi_1(R, 0, 0) \varphi_2(x_3, y_3, z_3) \\
&= \mathrm{const}_1\, \varphi_2(x_3, y_3, z_3).
\end{aligned}$$

And it does exactly so.

Example 4

Why is the hole called the exchange hole? Perhaps it would be enough to take the product function[35] and then we would also see that electron 3 runs away in panic from the other electron with the same spin? Let us see how it is in the first case (Example 1):

$$\begin{aligned}
\psi_{\mathrm{Hartree}}(1, 2, 3) &= \phi_1(1)\phi_2(2)\phi_3(3) = \varphi_1(1)\alpha(1)\varphi_1(1)\beta(1)\varphi_2(3)\alpha(3), \\
\psi_{\mathrm{Hartree}}(1_0, 2_0, 3) &= \varphi_1(0, 0, 0)\varphi_1(R, 0, 0)\varphi_2(x_3, y_3, z_3) = \mathrm{const}^2\, \varphi_2(x_3, y_3, z_3).
\end{aligned}$$

We get the distribution

$$\rho_{\mathrm{Hartree}} = \mathrm{const}^4 \left|\varphi_2(x_3, y_3, z_3)\right|^2.$$

And what do we get in the second case (Example 2)?

$$\begin{aligned}
\psi_{\mathrm{Hartree}}(1, 2, 3) &= \phi_1(1)\phi_2(2)\phi_3(3) = \varphi_1(1)\alpha(1)\varphi_1(1)\beta(1)\varphi_2(3)\alpha(3), \\
\psi_{\mathrm{Hartree}}(1_0, 2_0, 3) &= \varphi_1(R, 0, 0)\varphi_1(R, 0, 0)\varphi_2(x_3, y_3, z_3) = \mathrm{const}^2\, \varphi_2(x_3, y_3, z_3).
\end{aligned}$$

---

[34] That is, low probability of finding electron 3 over there.

[35] “Illegal” (Hartree approximation), since it does not obey the Pauli principle.

Hence, electron 3 occupies the antibonding orbital φ_2 and does not even think of running away from anything. Its distribution is entirely insensitive to the position of electron 1.

Thus, *this hole results from the Pauli principle, i.e. from the exchange of electron numbering, hence the name "exchange hole".*

Summing up, the wave function of the electronic system:
- should account for the existence of the Coulomb hole around each electron, i.e. for the reduced probability for finding any other electron there;
- should also account for the exchange hole, i.e., in the vicinity of an electron with a definite spin coordinate there should be reduced probability for finding any other electron with the same spin coordinate;
- as we saw, the Hartree–Fock function does not account at all for the Coulomb hole, however, it takes into account the existence of the exchange hole.

Which hole is more important: Coulomb or exchange? This question will be answered in Chapter 11.

VARIATIONAL METHODS WITH SLATER DETERMINANTS

10.9 VALENCE BOND (VB) METHOD

10.9.1 RESONANCE THEORY – HYDROGEN MOLECULE

Slater determinants are usually constructed from *molecular* spinorbitals. If, instead, we use *atomic* spinorbitals and the Ritz variational method (Slater determinants as the expansion functions) we would get the most general formulation of the valence bond (VB) method. The beginning of VB theory goes back to papers by Heisenberg. The first application was made by Heitler and London, and later theory was generalized by Hurley, Lennard-Jones and Pople.[36]

The essence of the VB method can be explained by an example. Let us take the hydrogen molecule with atomic spinorbitals of type $1s_a\ \alpha$ and $1s_b\ \beta$ denoted shortly as $a\alpha$ and $b\beta$ centred at two nuclei. Let us construct from them several (non-normalized) Slater determinants, for instance:

$$\psi_1 = \begin{vmatrix} a(1)\alpha(1) & a(2)\alpha(2) \\ b(1)\beta(1) & b(2)\beta(2) \end{vmatrix} = \left[a(1)\alpha(1)b(2)\beta(2) - a(2)\alpha(2)b(1)\beta(1)\right],$$

$$\psi_2 = \begin{vmatrix} a(1)\beta(1) & a(2)\beta(2) \\ b(1)\alpha(1) & b(2)\alpha(2) \end{vmatrix} = \left[a(1)\beta(1)b(2)\alpha(2) - a(2)\beta(2)b(1)\alpha(1)\right],$$

---

[36] W. Heisenberg, *Zeit. Phys*. 38 (1926) 411, *ibid*. 39 (1926) 499, *ibid*. 41 (1927) 239; W. Heitler, F. London, *Zeit. Phys.* 44 (1927) 455; A.C. Hurley, J.E. Lennard-Jones, J.A. Pople, *Proc. Roy. Soc. London* A220 (1953) 446.

$$\psi_3 = \begin{vmatrix} a(1)\alpha(1) & a(2)\alpha(2) \\ a(1)\beta(1) & a(2)\beta(2) \end{vmatrix} = \left[a(1)\alpha(1)a(2)\beta(2) - a(2)\alpha(2)a(1)\beta(1)\right]$$

$$= \left[a(1)a(2)\right]\left[\alpha(1)\beta(2) - \alpha(2)\beta(1)\right] \equiv \psi_{H^-H^+},$$

$$\psi_4 = \begin{vmatrix} b(1)\alpha(1) & b(2)\alpha(2) \\ b(1)\beta(1) & b(2)\beta(2) \end{vmatrix} = \left[b(1)b(2)\right]\left[\alpha(1)\beta(2) - \alpha(2)\beta(1)\right] \equiv \psi_{H^+H^-}.$$

Heitler–London function

The functions ψ_3, ψ_4 and the normalized difference $\psi_1 - \psi_2$ (N_{HL} is a normalization factor)

$$\boxed{\begin{aligned} \psi_{\mathrm{HL}} &= N_{\mathrm{HL}}(\psi_1 - \psi_2) \\ &= N_{\mathrm{HL}}\left[a(1)b(2) + a(2)b(1)\right]\left[\alpha(1)\beta(2) - \alpha(2)\beta(1)\right] \quad (10.17) \end{aligned}}$$

are eigenfunctions of the operators $\hat{S}^2$ and $\hat{S}_z$ (cf. Appendix Q, p. 1006) corresponding to the singlet state. The functions ψ_3, ψ_4 for obvious reasons are called *ionic structures* (H^-H^+ and H^+H^-),[37] whereas the function ψ_{HL} is called a Heitler–London function or *a covalent structure*.[38]

ionic structure

The VB method relies on optimization of the expansion coefficients c in front of these structures in the Ritz procedure (p. 202)

covalent structure

$$\boxed{\psi = c_{\mathrm{cov}}\psi_{\mathrm{HL}} + c_{\mathrm{ion1}}\psi_{H^-H^+} + c_{\mathrm{ion2}}\psi_{H^+H^-}. \qquad (10.18)}$$

The covalent structure itself, ψ_{HL}, was one great success of Walter Heitler[39] and Fritz London. For the first time the correct description of the chemical bond was

Fritz Wolfgang London (1900–1954) was born in Breslau (now Wrocław) and studied in Bonn, Frankfurt, Göttingen, Munich (Ph.D. at 21) and in Paris. Later worked in Zurich, Rome and Berlin. Escaped from nazism to UK, where he worked at Oxford University (1933–1936). In 1939 London emigrated to the USA, where he became professor of theoretical chemistry at Duke University in Durham.

Fritz London rendered great services to quantum chemistry. He laid the foundations of the theory of the *chemical* (covalent) bond and also, in addition, introduced dispersion interactions, one of the most important *intermolecular* interactions. This is nearly all of what chemistry is about. He also worked in the field of superconductivity.

---

[37] Since both electrons reside at the same nucleus.

[38] Since both electrons belong to the same extent to each of the nuclei.

[39] Walter Heitler (1904–1981), German chemist, professor at the University in Göttingen, later in Bristol and Zürich.

obtained. The crucial point turned out to be an inclusion – in addition to the product function $a(1)b(2)$ – its counterpart *with exchanged electron numbers* $a(2)b(1)$, since the electrons *are* indistinguishable. If we expand the Hartree–Fock determinant with doubly occupied bonding orbital $a + b$, we would also obtain a certain linear combination of the three structures mentioned, but with *the constant coefficients independent of the interatomic distance*:

Hartree–Fock function in AO

$$\boxed{\psi_{\text{RHF}} = N\left(\frac{1}{N_{\text{HL}}}\psi_{\text{HL}} + \psi_{H^-H^+} + \psi_{H^+H^-}\right). \qquad (10.19)}$$

This leads to a very bad description of the H_2 molecule at long internuclear distances with the Hartree–Fock method. The true wave function should contain, among other things, both the covalent structure (i.e. the Heitler–London function) and the ionic structures. However, for long internuclear distances the Heitler–London function should dominate, because it corresponds to the (exact) dissociation limit (two ground-state hydrogen atoms). The trouble is that, with fixed coefficients, *the Hartree–Fock function overestimates the role of the ionic structure* for long interatomic distances. Fig. 10.4 shows that the Heitler–London function describes the electron correlation (Coulomb hole), whereas the Hartree–Fock function does not.

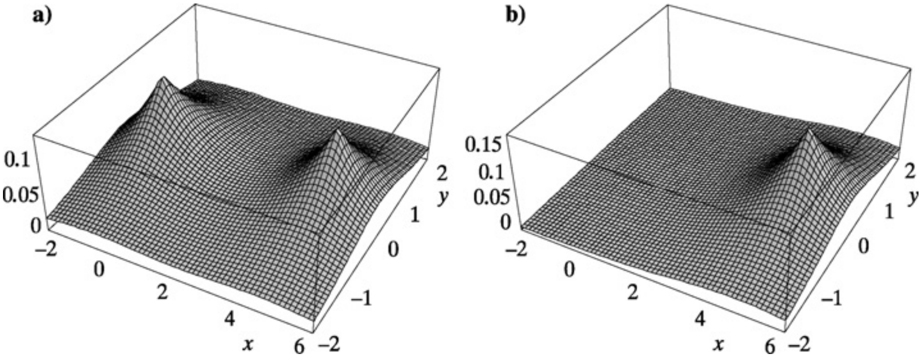
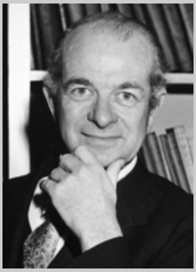


Fig. 10.4. Illustration of electron correlation in the hydrogen molecule. The nuclear positions are $(0, 0, 0)$ and $(4, 0, 0)$ in a.u. Slater orbitals of $1s$ type have orbital exponent equal to 1. (a) Visualization of the xy cross-section of the wave function of electron 2, assuming that electron 1 resides on the nucleus (either the first or the second one), has spin coordinate $\sigma_1 = \frac{1}{2}$, whereas electron 2 has spin coordinate $\sigma_2 = -\frac{1}{2}$ and the total wave function is equal $\psi = N\{ab + ba + aa + bb\}\{\alpha\beta - \beta\alpha\}$, i.e. it is a Hartree–Fock function. The plot is the same independently of which nucleus electron 1 resides, i.e., we observe the *lack of any correlation* of the motions of electrons 1 and 2. If we assume the spins to be parallel ($\sigma_2 = \frac{1}{2}$), the wave function vanishes. (b) A similar plot, but for the Heitler–London function $\psi_{\text{HL}} = N_{\text{HL}}[a(1)b(2) + a(2)b(1)][\alpha(1)\beta(2) - \alpha(2)\beta(1)]$ and with electron 1 residing at nucleus $(0, 0, 0)$. Electron 2 runs to the nucleus in position $(4, 0, 0)$. We have the correlation of the electronic motion. If we assume parallel spins ($\sigma_2 = \frac{1}{2}$), the wave function vanishes.

10.9.2 RESONANCE THEORY – POLYATOMIC CASE

The VB method was developed by Linus Pauling under the name of *theory of resonance*.

resonance theory

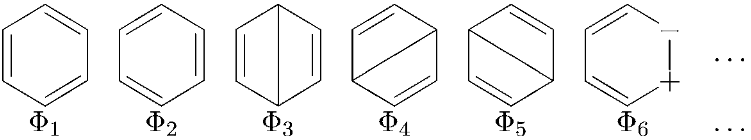


Linus Carl Pauling (1901–1994), American physicist and chemist, in the years 1931–1964 professor at the California Institute of Technology in Pasadena, in 1967–1969 professor at the University of California, San Diego, from 1969–1974 professor at the Stanford University. He received the 1954 Nobel prize: "*for his research into the nature of the chemical bond and its application to the elucidation of the structure of complex substances*". In 1962 he received the Nobel peace prize. His major achievements are the development of the theory of chemical bond, i.a., the VB method (also called resonance theory), and determining the structure of one of the fundamental structural elements of proteins, the α-helix.

The method can be applied to all molecules, although a particularly useful field of applications of resonance theory can be found in the organic chemistry of aromatic systems. For example, the total electronic wave function of the benzene molecule is presented as a linear combination of resonance structures[40]

$$\psi = \sum_I c_I \Phi_I, \tag{10.20}$$

to each (in addition to the mathematical form), a graph is assigned. For example, six π electrons can participate in the following "adventures" (forming covalent and ionic bonds)



The first two structures are famous Kekulé structures, the next three are Dewar structures, the sixth is an example of the possible mixed covalent-ionic structures. From these graphs, we may deduce which atomic orbitals (out of the $2p_z$ orbital of carbon atoms, z is perpendicular to the plane of the benzene ring) takes part in the covalent bond (of the π type). As far as the mathematical form of the Φ_1 structure is concerned, we can write it as the antisymmetrized (cf. antisymmetrization operator, p. 986) product of three Heitler–London functions (involving the proper pairs of $2p_z$ carbon atomic orbitals), the first for electrons $1, 2$, the second for electrons $3, 4$, and the third for $5, 6$. Within the functions Φ_I, the ionic structures can

---

[40] Similar to the original applications, we restrict ourselves to the π electrons, the σ electrons are treated as inactive in each structure, forming, among other things, the six C–C bonds presented below.

also occur. The rules for writing the structures were not quite clear, and the electrons were located to some extent in an arbitrary manner, making the impression that it is up to theoretical chemists to use their imaginations and draw imaginary pictures and – next – to translate them into mathematical form to obtain – after applying the variational method – an approximation to the wave function (and to the energy).

In fact, the problem is connected to the Ritz method and to expansion into the complete set of functions,[41] i.e. a purely mathematical problem. Although it may seem very strange to students (fortunately), many people were threatened for supporting the theory of resonance. Scientists serving the totalitarian regime decided to attack eq. (10.20). How, was this possible?[42] The Stalinists did not like the idea that "*the sum of fictitious structures can describe reality*". Wait a second! If some artificial functions could interfere with reality then socialist realism may lose to abstraction, a kolkhoz (collective farm) member to an intellectual, Lysenkoism to Mendelism,[43] goulags to the idea of freedom, and you are on the brink of disaster.

---

[41] In principle, *they should* form the complete set, but even so, in practical calculations, we never deal with true complete sets.

[42] Of course, *the true* reason was not a convergence of a series in the Hilbert space, but their personal careers *at any price*. Totalitarian systems never have problems finding such "scientists". In chemistry, there was the danger of losing a job, in biology, of losing a life.

It is rather difficult to think about Joseph Stalin as a quantum chemist. He was, however, kept informed about the current situation of a group of people involved in carrying out the summations in eq. (10.20), i.e. working in the resonance theory. To encourage young people to value and protect the freedom they have, and to reflect on human nature, some excerpts from the resolution adopted by the All-Soviet Congress of Chemists of the Soviet Union are reported. The resolution pertains, i.a., to the theory of resonance (after the disturbing and reflective book by S.E. Schnoll, "*Gieroi i zlodiei rossijskoj nauki*", Kron-Press, Moscow, 1997, p. 297):

"*Dear Joseph Vissarionovich (Stalin),*

the participants of the All-Soviet Congress send to you, the Great Leader and Teacher of all progressive mankind, our warm and cordial greetings. We Soviet chemists gathered together to decide, by means of broad and free discussion, the fundamental problems of the contemporary theory of the structure of molecules, want to express our deepest gratitude to you for the everyday attention you pay to Soviet science, particularly to chemistry. Our Soviet chemistry is developing in the Stalin era, which offers unlimited possibilities for the progress of science and industry. Your brilliant work in the field of linguistics put the tasks for still swifter progress in front of all scientists of our fatherland (...). Motivated by the resolutions of the Central Committee of the Bolshevik Communist Party concerning ideological matters and by your instructions, Comrade Stalin, the Soviet chemists wage war against the ideological concepts of bourgeois science. The lie of the so called "resonance theory" has been disclosed, and the remains of this idea will be thrown away from the Soviet chemistry. We wish you, our dear Leader and Teacher, good health and many, many years of famous life to the joy and happiness of the whole of progressive mankind(...)."

The events connected with the theory of resonance started in the autumn of 1950 at Moscow University. Quantum chemistry lecturers, Yakov Kivovitch Syrkin and Mirra Yefimovna Diatkina, were attacked. The accusation was about dissemination of the theory of resonance and was launched by former assistants of Syrkin. Since everything was in the hands of the professionals, Syrkin and Diatkina pleaded guilty with respect to each of the charges.

[43] Trofim Lysenko (1898–1976), Soviet scientist of enormous political influence, rejected the genetic laws of Mendel. In my 7th grade school biology textbook virtually only his "theory" was mentioned. As a pupil, I recall wanting to learn this theory. It was impossible to find any information. With difficulties I finally found something: acorns should be placed in a hole in the ground in large numbers to permit something like the class struggle. The winner will be the strongest oak-tree and this is what we all want.

Gregor Johann Mendel (1822–1884), modest Moravian monk, from 1843 a member of the Augustinian order in Brno (abbot from 1868). His unusually precise and patient experiments with sweet peas of two colours and seeds of two degrees of smoothness, allowed him to formulate the principal laws of genetics. Only in 1900 were his fundamental results remembered, and since then the rapid progress of contemporary genetics began.

10.10 CONFIGURATION INTERACTION (CI) METHOD

In this method[44]

> the variational wave function is a linear combination of Slater determinants constructed from *molecular* spinorbitals, an expansion analogous to eq. (10.20).

In most cases we are interested in the function ψ for *the electronic ground state of the system* (in addition when solving the CI equations we also get approximations to the excited states with different values of the c_I coefficients).

Generally we construct the Slater determinants Φ_I by placing electrons on the molecular spinorbitals obtained with the Hartree–Fock method,[45] in most cases the set of determinants is additionally limited by imposing an upper bound for the orbital energy. In that case, the expansion in (10.20) is finite. The Slater determinants Φ_I are obtained by the replacement of occupied spinorbitals with virtual ones in the single Slater determinant, which is – in most cases – the Hartree–Fock

CI method

---

[44] Also called the method of superposition of configurations.

[45] In this method we obtain M molecular orbitals, i.e. $2M$ molecular spinorbitals, where M is the number of atomic orbitals employed. The Hartree–Fock determinant Φ_0 is the best form of wave function as long as the electronic correlation is not important. The criterion of this "goodness" is the mean value of the Hamiltonian. If we want to include the electron correlation, we may think of another form of the one-determinantal function, more suitable the starting point. Of course, we do not change our definition of correlation energy, i.e. we consider the RHF energy as that which does not contain any correlation effects. For instance, we may ask which of the normalized single-determinant functions Φ is closest to the normalized exact function ψ. As a measure of this we might use:

$$\left|\langle\psi|\Phi\rangle\right| = \text{maximum}. \tag{10.21}$$

The single determinantal function $\Phi = \Phi_B$, which fulfils the above condition, is called a Bruckner function (O. Sinanoğlu, K.A. Brueckner, "*Three Approaches to Electron Correlation in Atoms*", Yale Univ. Press, New Haven and London, 1970).

function (Φ_0, i.e. ψ_{RHF}). When one spinorbital is replaced, the resulting determinant is called singly excited, when two – doubly excited, etc.[46,47]

The virtual spinorbitals form an orthonormal basis in *the virtual space* (Appendix B, p. 895). If we carry out any non-singular linear transformation (cf. p. 396) of virtual spinorbitals, each "new" n-tuply excited Slater determinant becomes a linear combination of all "old" n-tuply excited determinants and only n-tuply excited ones.[48] In particular, the unitary transformation preserves the mutual orthogonality of the n-tuply excited determinantal functions.

Thus, the total wave function (10.20) is a linear combination of the *known* Slater determinants (we assume that the spinorbitals are always known) with *unknown c* coefficients.

The name of the CI methods refers to the linear combination of the configurations rather than to the Slater determinants.

CSF (configuration)

A configuration (CSF, i.e. Configuration State Function) is a linear combination of determinants which is an eigenfunction of the operators: $\hat{S}^2$ and $\hat{S}_z$, and belongs to the proper irreducible representation of the symmetry group of the Hamiltonian. We say that this is a linear combination of the (spatial and spin) symmetry adapted determinants. Sometimes we refer to the spin-adapted configurations which are eigenfunctions only of the $\hat{S}^2$ and $\hat{S}_z$ operators.

The particular terms in the CI expansion may refer to the respective CSFs or to the Slater determinants. Both versions lead to the same results, but using CSFs

---

[46] In the language of the second quantization (see Appendix U, p. 1023) the wave function in the CI method has the form (the Φ_0 function is a Slater determinant which does not necessarily need to be a Hartree–Fock determinant)

$$\psi = c_0\Phi_0 + \sum_{a,p} c_p^a \hat{p}^\dagger \hat{a}\Phi_0 + \sum_{a<b,\ p<q} c_{pq}^{ab} \hat{q}^\dagger \hat{p}^\dagger \hat{a}\hat{b}\Phi_0 + \text{higher excitations}, \tag{10.22}$$

where c are the expansion coefficients, the creation operators $\hat{q}^\dagger, \hat{p}^\dagger, \ldots$ refer to the virtual spinorbitals $\phi_p, \phi_q, \ldots$ and the annihilation operators $\hat{a}, \hat{b}, \ldots$ refer to occupied spinorbitals $\phi_a, \phi_b, \ldots$ (the operators are denoted with the same indices as spinorbitals but the former are equipped with hat symbols), and the inequalities satisfied by the summation indices ensure that the given Slater determinant occurs only once in the expansion.

[47] The Hilbert space corresponding to N electrons is the sum of the orthogonal subspaces $\Omega_n, n = 0, 1, 2, \ldots, N$, which are spanned by the n-tuply excited (orthonormal) Slater determinants. Elements of the space Ω_n are all linear combinations of n-tuply excited Slater determinants. It does not mean, of course, that each element of this space is an n-tuply excited Slater determinant. For example, the sum of two doubly excited Slater determinants is a doubly excited Slater determinant only when one of the excitations is common to both determinants.

[48] Indeed, the Laplace expansion (Appendix A) along the row corresponding to the first new virtual spinorbital leads to the linear combination of the determinants containing new (*virtual, which means that the rank of excitation is not changed by this*) orbitals in this row. Continuing this procedure with the Slater determinants obtained, we finally get a linear combination of n-tuply excited Slater determinants expressed in old spinorbitals.

may be more efficient if we are looking for a wave function which transforms itself according to a single irreducible representation.

Next this problem is reduced to the Ritz method (see Appendices L, p. 984, and K, p. 982), and subsequently to the secular equations $(\boldsymbol{H} - \varepsilon \boldsymbol{S})\boldsymbol{c} = \boldsymbol{0}$. It is worth noting here that, e.g., the CI wave function for the ground state of the helium atom would be linear combinations of the determinants where the largest c coefficient occurs in front of the Φ_0 determinant constructed from the spinorbitals $1s\alpha$ and $1s\beta$, but a nonzero contribution would also come from the other determinants, e.g., constructed from the $2s\alpha$ and $2s\beta$ spinorbitals (one of the doubly excited determinants). The CI wave functions for all states (ground and excited) are linear combinations of *the same Slater determinants* – they differ only in the c coefficients.

The state energies obtained from the solution of the secular equations always approach the exact values from above.

10.10.1 BRILLOUIN THEOREM

In the CI method we have to calculate matrix elements H_{IJ} of the Hamiltonian. The Brillouin theorem says that:

$$\langle \Phi_0 | \hat{H} \Phi_1 \rangle = 0 \tag{10.23}$$

if Φ_0 is a solution of the Hartree–Fock problem ($\Phi_0 \equiv \psi_{\mathrm{RHF}}$), and Φ_1 is a singly excited Slater determinant in which the spinorbital $\phi_{i'}$ is orthogonal to all spinorbitals used in Φ_0.

Proof: From the II Slater–Condon rule (Appendix M, p. 986) we have:

$$\langle \Phi_0 | \hat{H} \Phi_1 \rangle = \langle i | \hat{h} i' \rangle + \sum_j \big[\langle ij | i'j \rangle - \langle ij | ji' \rangle \big]. \tag{10.24}$$

On the other hand, considering the integral $\langle i | \hat{F} i' \rangle$, where $\hat{F}$ is a Fock operator, we obtain from (8.27) (using the definition of the Coulomb and exchange operators from p. 337):

$$\begin{aligned} \langle i | \hat{F} i' \rangle &= \langle i | \hat{h} i' \rangle + \sum_j \big[\langle i | \hat{J}_j i' \rangle - \langle i | \hat{K}_j i' \rangle \big] = \langle i | \hat{h} i' \rangle + \sum_j \big[\langle ij | i'j \rangle - \langle ij | ji' \rangle \big] \\ &= \langle \Phi_0 | \hat{H} \Phi_1 \rangle. \end{aligned}$$

From the Hermitian character of $\hat{F}$ it follows that

$$\langle i | \hat{F} i' \rangle = \langle \hat{F} i | i' \rangle = \varepsilon_i \delta_{ii'} = 0. \tag{10.25}$$

We have proved the theorem.

The Brillouin theorem is sometimes useful in discussion of the importance of particular terms in the CI expansion for the ground state.

10.10.2 CONVERGENCE OF THE CI EXPANSION

Increasing the number of expansion functions by adding a new function lowers or keeps unchanged the energy (due to the variational principle). It often happens that the inclusion of only two determinants gives qualitative improvement with respect to the Hartree–Fock method, however when going further, the situation becomes more difficult. The convergence of the CI expansion is very slow, i.e. to achieve a good approximation to the wave function, the number of determinants in the expansion must usually be large. Theoretically, the shape of the wave function ensures solution of the Schrödinger equation $H\psi = E\psi$, but in practice we are *always* limited by the basis of the atomic orbitals employed. To obtain satisfactory results, we need to increase the number M of atomic orbitals in the basis. The number of molecular orbitals produced by the Hartree–Fock method is also equal to M, while the number of spinorbitals is equal to $2M$. In this case, the number of all determinants is equal to $\binom{2M}{N}$, where N refers to the number of electrons.

10.10.3 EXAMPLE OF H_2O

We are interested in the ground state of the water molecule. This is a singlet state ($S = 0$, $M_S = 0$).

double dzeta

The minimal basis set, composed of 7 atomic orbitals (two $1s$ orbitals of the hydrogen atoms, $1s, 2s$ and three $2p$ orbitals of the oxygen atom), is considered too poor, therefore we prefer what is called double dzeta basis, which provides two functions with different exponents for each orbital of the minimal basis. This creates a basis of $M = 14$ atomic orbitals. There are 10 electrons, hence $\binom{28}{10}$ gives 13 123 110 determinants. For a matrix of that size to be diagonalized is certainly impressive. Even more impressive is that we achieve only *an approximation* to the correlation energy which amounts to about 50% of the exact correlation energy,[49] since M is only equal to 14, but in principle it should be equal to ∞. Nevertheless, for *comparative* purposes we assume the correlation energy obtained is 100%.

block form

The simplest remedy is to get rid of some determinants in such a way that the correlation energy is not damaged. Which ones? Well, many of them correspond to the incorrect projection S_z of the total spin. For instance, we are interested in the singlet state (i.e. $S = 0$ and $S_z = 0$), but some determinants are built of spinorbitals containing exclusively α spin functions. This is a pure waste of resources, since the non-singlet functions do not make any contributions to the singlet state. When we remove these and other *incorrect* determinants, we obtain a smaller matrix to be diagonalized. The number of Slater determinants with $S_z = 0$ is equal $\binom{M}{N/2}^2$. In our case, this makes slightly over 4 million determinants (instead of about 13 million). What would happen if we diagonalized the huge original matrix anyway? Well, nothing would happen. There would be more work, but the computer would create *the block form*[50] (see Appendix C) from our enormous matrix, and each block would correspond to the particular S^2 and S_z, *while the whole contribution to the correlation energy of the ground state comes from the block corresponding to* $S = 0$ and $S_z = 0$.

Let us continue throwing away determinants. This time, however, we have to make a compromise, i.e. some of the Slater determinants are arbitrarily consid-

---

[49] We see here how vicious the dragon of electron correlation is.

[50] These square blocks would be easily noticed after proper ordering of the expansion functions.

ered not to be important (which will worsen the results, if they are rejected). Which of the determinants should be considered as not important? The general opinion in quantum chemistry is that the multiple excitations are less and less important (when the multiplicity increases). If we take only the singly, doubly, triply and quadruply excited determinants, the number of determinants will reduce to about 25000 and we will obtain 99% of the approximate correlation energy defined above. If we take the singly and doubly excited determinants only, there are only 360 of them, and 94% of the correlation effect is obtained. This is why this CISD (*CI Singles and Doubles*) method is used so often.

CI SD

For larger molecules this selection of determinants becomes too demanding, therefore we have to decide individually for each configuration: to include or reject it? The decision is made either on the basis of the perturbational estimate of the importance of the determinant[51] or by a test calculation with inclusion of the determinant in question, Fig. 10.5.

To obtain very good results, we need to include a large number of determinants, e.g., of the order of thousands, millions or even billions.[52] This means that contemporary quantum chemistry has made enormous technical progress.[53] This, however, is a sign, not of the strength of quantum chemistry, but of its weakness. What are we going to do with such a function? We may load it back into the computer and calculate all the properties of the system with high accuracy (although this cannot be guaranteed). To answer a student's question about why we obtained some particular numbers, we have to say that we do not know, it is the computer which knows. This is a trap. It would be better to get, say, two Slater determinants, which describe the system to a reasonable approximation and we can understand what is going on in the molecule.

10.10.4 WHICH EXCITATIONS ARE MOST IMPORTANT?

The convergence can be particularly bad if we use the virtual spinorbitals obtained by the Hartree–Fock method. Not all excitations are equally important. It turns out that usually, although this is not a rule, low excitations dominate the ground state wave function.[54] The single excitations *themselves* do not contribute anything to the ground state *energy* (if the spinorbitals are generated with the Hartree–Fock

---

[51] The perturbational estimate mentioned relies on the calculation of the weight of the determinant based on the first order correction to the wave function in perturbation theory, see Chapter 5. In such an estimate the denominator contains the excitation energy evaluated as the difference in orbital energies between the Hartree–Fock determinant and the one in question. In the numerator there is a respective matrix element of the Hamiltonian calculated with the help of the known Slater–Condon rules (Appendix M, p. 986).

[52] Recently calculations with 3.6 billion Slater determinants were reported.

[53] To meet such needs, quantum chemists have had to develop entirely new techniques of applied mathematics.

[54] That is, requiring the lowest excitation energies. Later, a psychological mechanism began to work supported by economics: the *high energy* excitations are numerous and, because of that, very expensive and they correspond to high *excitations rank* (the number of electrons excited). Due to this, a reasonable restriction for the number of configurations in the CI expansion is excitation rank. We will come back to this problem later.

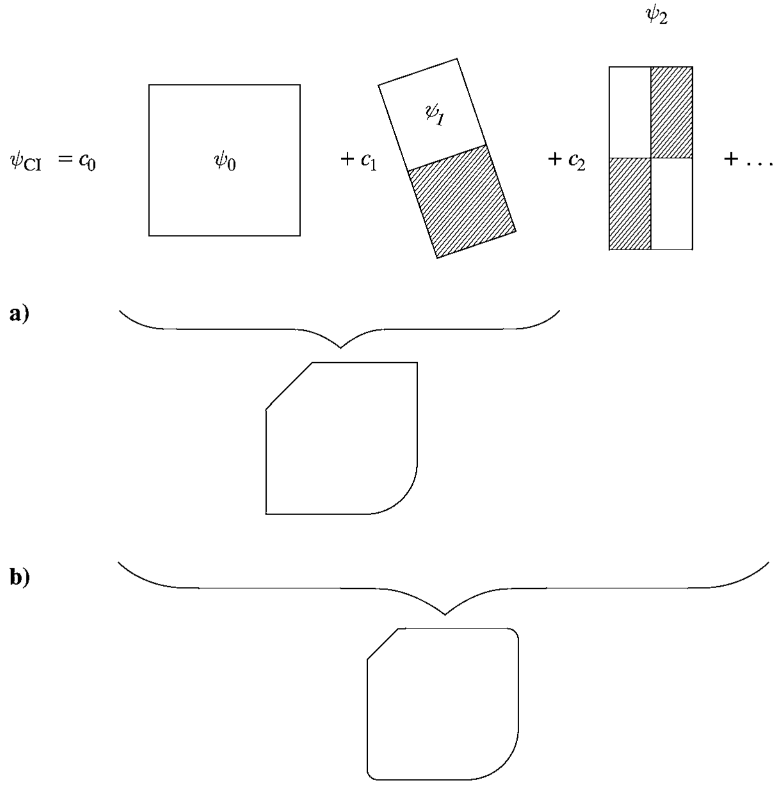


Fig. 10.5. *Symbolic illustration of the principle* of the CI method with one Slater determinant ψ_0 dominant in the ground state (this is a problem of the many electron wave function so the picture cannot be understood literally). The purpose of this diagram is to emphasize a relatively small role of electronic correlation (more exactly, of what is known as the dynamical correlation, i.e. correlation of electronic motion). The function ψ_{CI} is a linear combination (the c coefficients) of the determinantal functions of different shapes in the many electron Hilbert space. The shaded regions correspond to the negative sign of the function; the nodal surfaces of the added functions allow for the effective deformation of ψ_0 to have lower and lower average energy. (a) Since c_1 is small in comparison to c_0, the result of the addition of the first two terms is a slightly deformed ψ_0. (b) Similarly the additional excitations just make cosmetic changes to the function (although they may substantially affect a quantity calculated with it).

method, then the Brillouin theorem mentioned above applies). *They are crucial, however, for excited states or in dipole moment calculations.* Only when coupled to other types of excitation do they assume non-zero (although small) values. Indeed, if in the CI expansion we only use the Hartree–Fock determinant and the determinants corresponding to single excitations, then, due to the Brillouin theorem, the secular determinant would be factorized.[55] This factorization (Fig. 10.6) pertains to the single-element determinant corresponding to the Hartree–Fock function

---

[55] That is, could be written out in the block form, which would separate the problem into two subproblems of smaller size.

$$\boldsymbol{H} = \begin{array}{c|c|c|c|c|c|c|}
 & \text{HF} & \text{S} & \text{D} & \text{T} & \text{Q} & \cdots \\
\hline
\text{HF} & E_{\text{HF}} & 0^{\text{a}} & \text{III} & 0^{\text{b}} & 0^{\text{b}} & 0^{\text{b}} \\
\hline
\text{S} & 0^{\text{a}} & \text{block S} & \text{II} & \text{III} & 0^{\text{b}} & 0^{\text{b}} \\
\hline
\text{D} & \text{III} & \text{II} & \text{block D} & \text{II} & \text{III} & 0^{\text{b}} \\
\hline
\text{T} & 0^{\text{b}} & \text{III} & \text{II} & \text{block T} & \text{II} & \text{III} \\
\hline
\text{Q} & 0^{\text{b}} & 0^{\text{b}} & \text{III} & \text{II} & \text{block Q} & \text{II} \\
\hline
\vdots & 0^{\text{b}} & 0^{\text{b}} & 0^{\text{b}} & \text{III} & \text{II} & \text{block} \ldots \\
\hline
\end{array}$$

Fig. 10.6. The block structure of the Hamiltonian matrix ($\boldsymbol{H}$) is the result of the Slater–Condon rules (Appendix M, p. 986). S – single excitations, D – double excitations, T – triple excitations, Q – quadruple excitations. (a) Block of zero values due to the Brillouin theorem. (b) The block of zero values due to the IV Slater–Condon rule, (II) the non-zero block obtained according to II and III Slater–Condon rules, (III) the non-zero block obtained according to III Slater–Condon rule. All the non-zero blocks are sparse matrices dominated by zero values, which is important in the diagonalization process.

and to the determinants corresponding exclusively to single excitations. Since we are interested in the ground state, only the first determinant (Hartree–Fock) is of importance to us, and it does not change whether we include or not, a contribution coming from single excitations into the wave function.

Performing CI calculations with the inclusion of all excitations (for the assumed value of M), i.e. the *full CI*, is not possible in practical calculations due to the too long expansion. We are forced to truncate the CI basis somewhere. It would be good to terminate it in such a way that all *essential* (the problem is what we mean by essential) terms are retained. *The most significant terms for the correlation energy come from the double excitations since these are the first excitations coupled to the Hartree–Fock function*, Fig. 10.6. Smaller, although important, contributions come from other excitations (usually of low excitation rank). We certainly wish that it would be like this for large molecules. Nobody knows what the truth is.

full CI

10.10.5 NATURAL ORBITALS (NO)

The fastest convergence is achieved in the basis set of *natural* orbitals (NO), i.e. when we construct spinorbitals with *these* orbitals and from them the Slater determinants. The NO is defined *a posteriori* in the following way. After carrying out the CI calculations, we construct the density matrix ρ (see Appendix S, p. 1015)

$$\begin{aligned}
\rho(1,1') &= \int \psi^*(1',2,3,\ldots,N)\psi(1,2,3,\ldots,N)\,\mathrm{d}\tau_2\,\mathrm{d}\tau_3\ldots\mathrm{d}\tau_N \\
&= \sum_{ij} D_{ji}\phi_i^*(1')\phi_j(1), \quad D_{ij} = D_{ji}^*,
\end{aligned} \tag{10.26}$$

where the summation runs over all the spinorbitals. By diagonalization of matrix $\boldsymbol{D}$ (a rotation in the Hilbert space spanned by the spinorbitals) we obtain the density matrix expressed in the natural spinorbitals (NO) transformed by the unitary transformation

$$\rho(1,1') = \sum_i (D_{\text{diag}})_{ii} \phi_i'^*(1)\phi_i'(1'). \tag{10.27}$$

NO occupancies

pseudonatural orbitals (PNO)

The most important ϕ_i' from the viewpoint of the correlation are the NOs with large *occupancies*, i.e. $(D_{\text{diag}})_{ii}$ values. Inclusion of only the most important ϕ_i' in the CI expansion creates a short and quite satisfactory wave function.[56] Meyer introduced the PNO CI, i.e. *pseudonatural orbitals*.[57] In the first step, we perform the CI calculations for excitations obtained by replacement of two selected spinorbitals. The process is repeated for all spinorbital pairs and at the end we carry out a "large" CI, which includes all important determinants engaged in the partial calculations (i.e. those with large weights).

10.10.6 SIZE CONSISTENCY

A truncated CI expansion has one unpleasant feature which affects the applicability of the method.

Let us imagine we want to calculate the interaction energy of two beryllium atoms. Let us suppose that we decide that to describe the beryllium atom we have to include, not only the $1s^2 2s^2$ configuration, but also the doubly excited $1s^2 2p^2$. In the case of beryllium, this is a very reasonable step, since both configurations have similar energies. Let us assume now that we calculate the wave function for *two* beryllium atoms. If we want this function to describe the system correctly, also at large interatomic distances, we have to make sure that the departing atoms have appropriate excitations at their disposition, i.e. in our case $1s^2 2p^2$. To achieve this we *must incorporate quadruple excitations into the method*.[58]

> If we include quadruples, we have a chance to achieve (an approximate) size consistency, i.e., the energy will be proportional to the number of atoms, otherwise our results will not be size consistent.

Let us imagine 10 beryllium atoms. In order to have size consistency we need to include 20-fold excitations. This would be very expensive. We clearly see that, for many systems, the size consistency requires inclusion of multiple excitations. If we carried out CI calculations for all possible (for a given number of spinorbitals) excitations, such a CI method (i.e. *full CI*) would be size consistent.

---

[56] Approximate natural orbitals can also be obtained directly without performing the CI calculations.

[57] R. Ahlrichs, W. Kutzelnigg, *J. Chem. Phys*. 48 (1968) 1819; W. Meyer, *Intern. J. Quantum Chem.* S5 (1971) 341.

[58] See J.A. Pople, R. Seeger, R. Krishnan, *Intern. J. Quantum Chem.* S11 (1977) 149, also p. 47 of the book by P. Jørgensen and J. Simons, "*Second Quantization-Based Methods in Quantum Chemistry*", Academic Press, 1981.

10.11 DIRECT CI METHOD

We have already mentioned that the CI method converges slowly. Due to this, the Hamiltonian matrices and overlap integral matrices are sometimes so large that they cannot fit into the computer memory. In practice, such a situation occurs in all good quality calculations for small systems and in all calculations for medium and large systems. Even for quite large atomic orbital basis, the number of integrals is much smaller than the number of Slater determinants in the CI expansion.

Björn Roos[59] first noticed that to find the lowest eigenvalues and their eigenvectors we do not need to store a huge $\boldsymbol{H}$ matrix in computer memory. Instead, we need to calculate the *residual vector* $\boldsymbol{\sigma} = (\boldsymbol{H} - E\mathbf{1})\boldsymbol{c}$, where $\boldsymbol{c}$ is a trial vector (defining the trial function in the variational method, p. 196). If $\boldsymbol{\sigma} = \mathbf{0}$, it means that the solution is found. Knowing $\boldsymbol{\sigma}$, we may find (on the basis of first order perturbation theory) slightly improved $\boldsymbol{c}$, etc. The product $\boldsymbol{Hc}$ can be obtained by going through the set of integrals and assigning to each a coefficient resulting from $\boldsymbol{H}$ and $\boldsymbol{c}$, and adding the results to the new $\boldsymbol{c}$ vector. Then the procedure is repeated. Until 1971, CI calculations with 5000 configurations were considered a significant achievement. After Roos's paper, there was a leap of several orders of magnitude, bringing the number of configurations to the range of billions. For the computational method this was a revolution.

residual vector

10.12 MULTIREFERENCE CI METHOD

Usually in the CI expansion, the dominant determinant is Hartree–Fock. We construct the CI expansion, replacing the spinorbitals in this determinant (*single reference method*). We can easily imagine a situation in which taking one determinant is not justified, since the shell is not well closed (e.g., four hydrogen atoms). We already know that certain determinants (or, in other words: configurations) absolutely need to be present ("*static correlation*") in the correct wave function. To be sure, we are the judges, deciding which is good or bad. This set of determinants is a basis in the *model space*.

single reference method

static correlation

model space

> In the single reference CI method, the model space (Fig. 10.7) is formed by one Slater determinant. In the multireference CI method, the set of determinants constitute the model space. This time, the CI expansion is obtained by replacement of the spinorbitals participating in the model space by other virtual orbitals. We proceed further as in CI.

multireference method

There is no end to the problems yet, since, again we have billions of possible excitations.[60] We do other tricks to survive in this situation. We may, for instance,

---

[59] B.O. Roos, *Chem. Phys. Letters* 15 (1972) 153.

[60] There is another trouble known as *intruder states*, i.e. states which are of unexpectedly low energy. How can these states appear? Firstly, the CI states known as "front door intruders" appear, if some im-

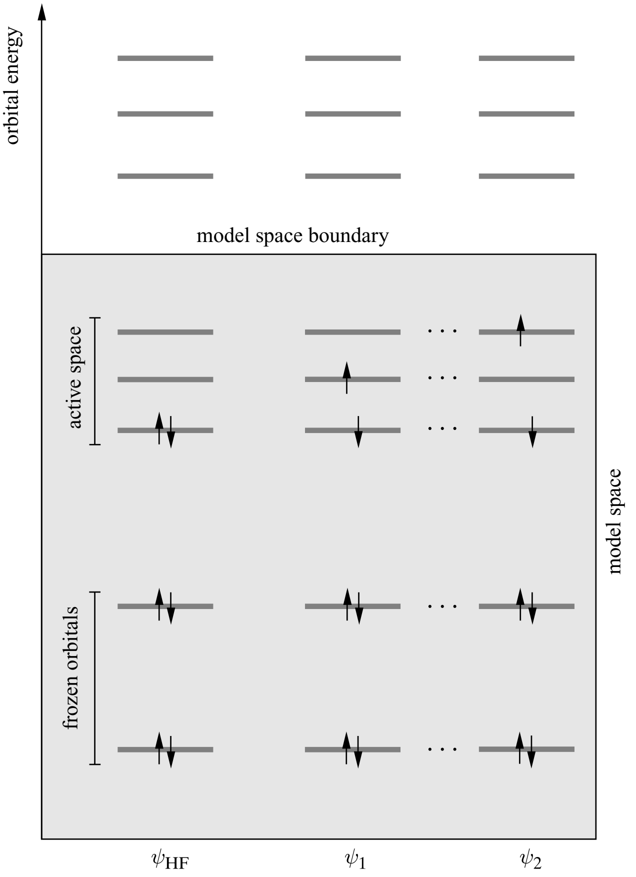


Fig. 10.7. Illustration of the model space in the multireference CI method used mainly in the situation when no single Slater determinant dominates the CI expansion. In the figure the orbital levels of the system are presented. Part of them are occupied in all Slater determinants considered ("frozen spinorbitals"). Above them is a region of closely spaced orbital levels called *active space*. In the optimal case, an energy gap occurs between the latter and unoccupied levels lying higher. The model space is spanned by all or some of the Slater determinants obtained by various occupancies of the active space levels.

frozen orbitals

inactive orbitals

get the idea not to excite the inner shell orbitals, since the numerical effort is serious, the lowering of the total energy can also be large, but the effect on the energy *differences* (this is what chemists are usually interested in) is negligible. We say that such orbitals are *frozen*. Some of the orbitals are kept doubly occupied in all Slater determinants but we optimize their shape. Such orbitals are called *inactive*. Finally, the orbitals of varied occupancy in different Slater determinants are called

---

portant (low-energy) configurations were for some reason not included into the model space. Secondly, we may have the "back door intruder" states. When the energy gaps between the model space and the other configurations are too small (quasi-degeneracy), some CI states became low energy states (enter the model space energy zone) even if they are not composed of the model space configurations.

active. The frozen orbitals are, in our method, important spectators of the drama, the inactive orbitals contribute a little towards lowering the energy, but the most efficient work is done by the active orbitals.

active orbitals

10.13 MULTICONFIGURATIONAL SELF-CONSISTENT FIELD METHOD (MC SCF)

In the configuration interaction method, it is sometimes obvious that certain determinants of the CI expansion *must* contribute to the wave function, if the latter is to correctly describe the system. For example, if we want to describe the system in which a bond is being broken (or is being formed), for its description we need several determinants for sure (cf. description of the dissociation of the hydrogen molecule on p. 371).

Why is this? In the case of dissociation, that we are dealing with here, there is a quasidegeneracy of the bonding and antibonding orbital of the bond in question, i.e. the approximate equality of their energies (the bond energy is of the order of the overlap integral and the latter goes to zero when the bond is being broken). The determinants, which can be constructed by various occupancies of these orbitals, have very similar energies and, consequently, their contributions to the total wave function are of similar magnitude and *should be included* in the wave function.

In the MC SCF method, as in CI, it is *up to us* to decide which set of determinants we consider sufficient for the description of the system.

> Each of the determinants is constructed from molecular spinorbitals which are not fixed (as in the CI method) but are modified in such a way as to have the total energy as low as possible.

The MC SCF method is the most general scheme of the methods that use a linear combination of Slater determinants as an approximation to the wave function. In the limiting case of the MC SCF, when the number of determinants is equal to 1, we have, of course, the Hartree–Fock method.

10.13.1 CLASSICAL MC SCF APPROACH

We will describe first the classical MC SCF approach. This is a variational method. As was mentioned, the wave function in this method has the form of a finite linear combination of Slater determinants Φ_I

$$\psi = \sum_I d_I \Phi_I, \tag{10.28}$$

where d are variational coefficients.

In the classical MC SCF method we:

1. take a finite CI expansion (the Slater determinants and the orbitals for their construction are fixed)
2. calculate the coefficients for the determinants by the Ritz method (the orbitals do not change)
3. vary the LCAO coefficients in the orbitals at the fixed CI coefficients to obtain the best MOs
4. return to point 1 until self-consistency is achieved

10.13.2 UNITARY MC SCF METHOD

Another version of the MC SCF problem, *a unitary method* suggested by Lévy and Berthier[61] and later developed by Dalgaard and Jørgensen[62] is gaining increasing importance. The eigenproblem does not appear in this method.

We need two mathematical facts to present the unitary MC SCF method. The first is a theorem:

If $\hat{A}$ is a Hermitian operator, i.e. $\hat{A}^\dagger = \hat{A}$, then $\hat{U} = \exp(i\hat{A})$ is a unitary operator satisfying $\hat{U}^\dagger\hat{U} = 1$.

Let us see how $\hat{U}^\dagger$ looks:

$$\begin{aligned}\hat{U}^\dagger = \left(\exp\left(i\hat{A}\right)\right)^\dagger &= \left(1 + i\hat{A} + \frac{1}{2!}(i\hat{A})^2 + \frac{1}{3!}(i\hat{A})^3 + \cdots\right)^\dagger \\ &= \left(1 + (-i)\hat{A}^\dagger + \frac{1}{2!}(-i\hat{A}^\dagger)^2 + \frac{1}{3!}(-i\hat{A}^\dagger)^3 + \cdots\right) \\ &= \left(1 + (-i)\hat{A} + \frac{1}{2!}(-i\hat{A})^2 + \frac{1}{3!}(-i\hat{A})^3 + \cdots\right) = \exp(-i\hat{A}).\end{aligned}$$

Hence, $\hat{U}\hat{U}^\dagger = 1$, i.e. $\hat{U}$ is a unitary operator.[63]

---

[61] B. Lévy, G. Berthier, *Intern. J. Quantum Chem.* 2 (1968) 307.

[62] E. Dalgaard, P. Jørgensen, *J. Chem. Phys.* 69 (1978) 3833.

[63] Is an operator ($\hat{C}$) of multiplication by a constant c Hermitian?

$$\begin{aligned}\langle\varphi|\hat{C}\psi\rangle &\overset{?}{=} \langle\hat{C}\varphi|\psi\rangle,\\ \text{l.h.s.} = \langle\varphi|c\psi\rangle &= c\langle\varphi|\psi\rangle,\\ \text{r.h.s.} = \langle c\varphi|\psi\rangle &= c^*\langle\varphi|\psi\rangle.\end{aligned}$$

Both sides are equal, if $c = c^*$. An operator conjugate to c is c^*.

Further: $\hat{B} = i\hat{A}$, what is a form of $\hat{B}^\dagger$?

$$\begin{aligned}\langle\hat{B}^\dagger\varphi|\psi\rangle &= \langle\varphi\hat{B}|\psi\rangle,\\ \langle\varphi|i\hat{A}|\psi\rangle &= \langle -i\hat{A}^\dagger\varphi|\psi\rangle,\\ \hat{B}^\dagger &= -i\hat{A}^\dagger.\end{aligned}$$

Now the second mathematical fact. This is a commutator expansion:

$$e^{-\hat{A}}\hat{H}e^{\hat{A}} = \hat{H} + [\hat{H},\hat{A}] + \frac{1}{2!}[[\hat{H},\hat{A}],\hat{A}] + \frac{1}{3!}[[[\hat{H},\hat{A}],\hat{A}],\hat{A}] + \cdots \quad (10.29)$$

This theorem can be proved by induction, expanding the exponential functions.

Now we are all set to describe the unitary method. We introduce two new operators:

$$\hat{\lambda} = \sum_{ij} \lambda_{ij} \hat{i}^{\dagger} \hat{j}, \quad (10.30)$$

where $\hat{i}^{\dagger}$ and $\hat{j}$ are the creation and annihilation operators, respectively, associated to spinorbitals i, j, see Appendix U. Further,

$$\hat{S} = \sum_{IJ} S_{IJ} |\Phi_I\rangle\langle\Phi_J|. \quad (10.31)$$

We assume that λ_{ij} and S_{IJ} are elements of the Hermitian matrices $\boldsymbol{\lambda}$ and $\boldsymbol{S}$ (their determination is the goal of the method), Φ_I are determinants from the MC SCF expansion (10.28).

> It can be seen that the $\hat{\lambda}$ operator replaces a single spinorbital in a Slater determinant and forms a linear combination of such modified determinantal functions; the $\hat{S}$ operator replaces such a combination with another. The "knobs" which control these changes are coefficients λ_{ij} and S_{IJ}.

We will need transformations $\exp(i\hat{\lambda})$ and $\exp(i\hat{S})$. They are unitary, because $\hat{\lambda}^{\dagger} = \hat{\lambda}$ and $\hat{S}^{\dagger} = \hat{S}$, i.e. $\hat{\lambda}$ and $\hat{S}$ are Hermitian.[64]

We suggest the form of our variational function:

$$|\tilde{0}\rangle = \exp(i\hat{\lambda})\exp(i\hat{S})|0\rangle \quad (10.32)$$

where $|0\rangle$ denotes a starting combination of determinants with specific spinorbitals and the matrices $\boldsymbol{\lambda}$ and $\boldsymbol{S}$ contain the variational parameters as the matrix elements. So, we modify the spinorbitals and change the coefficients in front of the determinants to obtain a new combination of the modified determinants, $|\tilde{0}\rangle$. The mean energy value for that function is[65]

$$E = \langle\tilde{0}|\hat{H}|\tilde{0}\rangle = \langle 0|\exp(-i\hat{S})\exp(-i\hat{\lambda})\hat{H}\exp(i\hat{\lambda})\exp(i\hat{S})|0\rangle. \quad (10.33)$$

Taking advantage of the commutator expansion (10.29), we have

$$E = \langle 0|\hat{H}|0\rangle - i\langle 0|[\hat{S}+\hat{\lambda},\hat{H}]|0\rangle + \frac{1}{2}\langle 0|[\hat{S},[\hat{H},\hat{S}]]|0\rangle + \frac{1}{2}\langle 0|[\hat{\lambda},[\hat{H},\hat{\lambda}]]|0\rangle$$
$$+ \langle 0|[\hat{S},[\hat{H},\hat{\lambda}]]|0\rangle + \cdots$$

---

[64] Considering the matrix elements of the operators $\hat{\lambda}$ and $\hat{S}$, we would easily be convinced that both operators are also Hermitian.

[65] Here we use the equality $[\exp(i\hat{A})]^{\dagger} = \exp(-i\hat{A})$.

It follows from the last equation, that in order to calculate E, we have to know the result of the operation of $\hat{\lambda}$ on $|0\rangle$, i.e. on the linear combination of determinants, which comes down to the operation of the creation and annihilation operators on the determinants, which is simple. It can also be seen that we need to apply the operator $\hat{S}$ to $|0\rangle$, but its definition shows that this is trivial. This expression[66] can now be optimized, i.e. the best Hermitian matrices $\boldsymbol{\lambda}$ and $\boldsymbol{S}$ can be selected. It is done in the same step (this distinguishes the current method from the classical one). Usually the calculations are carried out in a matrix form neglecting the higher terms and retaining only the quadratic ones in $\hat{S}$ and $\hat{\lambda}$. Neglecting the higher terms is equivalent to allowing for very small rotations in the transformation (10.32), but instead we have a large number of rotations (iterative solution).[67]

The success of the method depends on the starting point. The latter strongly affects the energy and its hypersurface (in the space of the parameters of the matrices $\boldsymbol{\lambda}$ and $\boldsymbol{S}$) is very complicated, it has many local minima. This problem is not yet solved, but various procedures accelerating the convergence are applied, e.g., the new starting point is obtained by averaging the starting points of previous iterations. The method also has other problems, since the orbital rotations partially replace the rotation in the space of the Slater determinants (the rotations do not commute and are not independent). In consequence, linear dependencies may appear.

10.13.3 COMPLETE ACTIVE SPACE METHOD (CAS SCF)

An important special case of the MC SCF method is the CAS SCF (Complete Active Space Self-Consistent Field, Fig. 10.8) of Roos, Taylor and Siegbahn.[68] Let us assume that we are dealing with a closed-shell molecule. The RHF method (p. 342) provides the molecular orbitals and the orbital energies. From them we select the low energy orbitals. Part of them are *inactive*, i.e. are doubly occupied in all determinants, but they are varied, which results in lowering the mean value of the Hamiltonian (some of the orbitals may be frozen, i.e. kept unchanged). These are the spinorbitals corresponding to the inner shells. The remaining spinorbitals belong to the *active space*. Now we consider all possible occupancies and excitations of the active spinorbitals (this is where the adjective "*complete*" comes from) to obtain the set of determinants in the expansion of the MC SCF. By taking all possible

---

[66] The term with i gives a real number

$$i \cdot \langle 0|[\hat{S}+\hat{\lambda},\hat{H}]|0\rangle = i \cdot \big(\langle(\hat{S}+\hat{\lambda})0|\hat{H}0\rangle - \langle\hat{H}0|(\hat{S}+\hat{\lambda})0\rangle\big) \to i \cdot (z - z^*) = i(2i\,\mathrm{Im}\,z) \in R.$$

R is a set of real numbers.

[67] In the classical MC SCF method when minimizing the energy with respect to the parameters, we use only linear terms in the expansion of the energy with respect to these parameters. In the unitary formulation, on the other hand, we use both linear and quadratic terms. This implies much better convergence of the unitary method.

[68] B.O. Roos, P.E.M. Siegbahn, in "*Modern Theoretical Chemistry*", vol. III, ed. H.F. Schaefer, Plenum Press, New York, 1977; P.E.M. Siegbahn, *J. Chem. Phys.* 70 (1979) 5391; B.O. Roos, P.R. Taylor, P.E.M. Siegbahn, *Chem. Phys.* 48 (1980) 157.

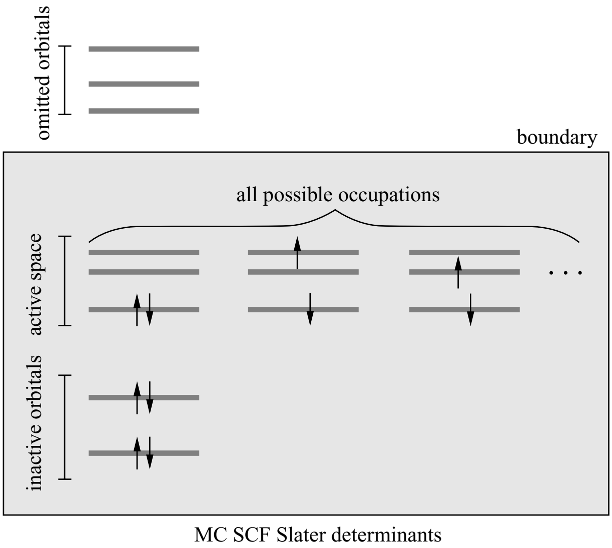


Fig. 10.8. CAS SCF, a method of construction of the Slater determinants in the MC SCF expansion. The inner shell orbitals are usually inactive, i.e. are doubly occupied in each Slater determinant. Within the active space + inactive spinorbitals we create the complete set of possible Slater determinants to be used in the MC SCF calculations. The spinorbitals of energy higher than a certain selected threshold are entirely ignored in the calculations.

excitations within the active space, we achieve *a size consistency*, i.e. when dividing the system into subsystems and separating them (infinite distances) we obtain the sum of the energies calculated for each subsystem separately. By taking the complete set of excitations we also show that the results do not depend on any (non-singular) linear transformation of the molecular spinorbitals within the given subgroup of orbitals, i.e. within the inactive or active spinorbitals. This makes the result invariant with respect to the localization of the molecular orbitals.

NON-VARIATIONAL METHODS WITH SLATER DETERMINANTS

10.14 COUPLED CLUSTER (CC) METHOD

The problem of a many-body correlation of motion of anything is extremely difficult and so far unresolved (e.g., weather forecasting). The problem of electron correlation also seemed to be hopelessly difficult. It still remains so, however, it turns out that we can exploit a certain observation made by Sinanoğlu.[69] This author noticed that the major portion of the correlation is taken into account through the in-

---

[69] O. Sinanoğlu and K.A. Brueckner, "*Three Approaches to Electron Correlation in Atoms*", Yale Univ. Press, New Haven and London, 1970.

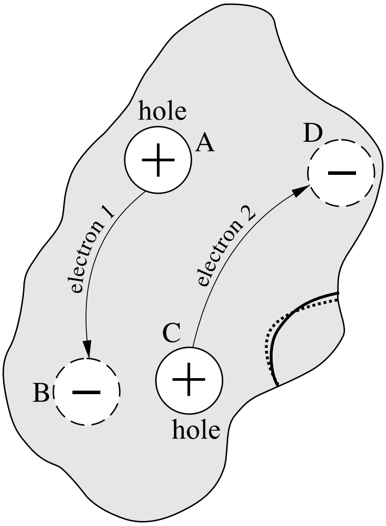


Fig. 10.9. In order to include the electron correlation, the wave function should somehow reflect the fact that electrons avoid each other. Electron 1 jumping from A (an orbital) to B (another orbital) should make electron 2 escape from C (close to B) to D (close to A). This is the very essence of electron correlation. The other orbitals play a role of spectators. However, the spectators change upon the excitations described above. These changes are performed by allowing their own excitations. This is how triple, quadruple and higher excitations emerge and contribute to electronic correlation.

troducing of correlation within electron pairs, next through pair–pair interactions, then pair–pair–pair interactions, etc. The canonical molecular spinorbitals, which we can use, are in principle delocalized over the whole molecule, but in practice the delocalization is not so large. Even in the case of canonical spinorbitals, and certainly when using localized molecular spinorbitals, we can think about electron excitation as a transfer of an electron from one place in the molecule to another. Inclusion of the correlation of electronic motion represents, in the language of electron excitations, the following philosophy: when electron 1 jumps from an orbital localized in place A to an orbital localized in place B, it would be good – from the point of view of the variational principle – if electron 2 jumped from the orbital localized at C to the orbital localized at D (strong electrostatic stabilization), Fig. 10.9.

The importance of a given double excitation depends on the energy connected with the electron relocation and the arrangement of points A,B,C,D. Yet this simplistic reasoning suggests single excitations do not carry any correlation (this is confirmed by the Brillouin theorem) and this is why their role is very small. Moreover, it also suggests that double excitations should be very important.

The general idea of the coupled cluster method relies on the more and more accurate description of the many-electron system, beginning with the picture of the independent electrons, next of independent pairs, next of independent pair–pair sets, etc.

10.14.1 WAVE AND CLUSTER OPERATORS

vacuum state

At the beginning we introduce a special Slater determinant, the *reference determinant* (called the *vacuum state*, it can be the Hartree–Fock determinant) Φ_0 and we write that the exact wave function for the ground state is

$$\psi = \exp(\hat{T})\Phi_0 \tag{10.34}$$

wave and cluster operators

where $\exp(\hat{T})$ is *a wave operator*, and $\hat{T}$ itself is *a cluster operator*. In the CC method *an intermediate normalization*[70] of the function ψ is assumed, i.e.

$$\langle\psi|\Phi_0\rangle = 1.$$

intermediate normalization

Eq. (10.34) represents a very ambitious task. It assumes that we will find an operator $\hat{T}$ such that the wave operator ($e^{\hat{T}}$), as with the touch of a wizard's wand, will make an ideal solution of the Schrödinger equation from the Hartree–Fock function. The formula with $\exp(\hat{T})$ is *an Ansatz*. The charming sounding word Ansatz[71] can be translated as arrangement or order, but in mathematics it refers to the construction assumed.

In literature we use the argument that the wave operator ensures the size consistency of the CC. According to this reasoning, for an infinite distance between molecules A and B, both ψ and Φ_0 functions can be expressed in the form of the product of the wave functions for A and B. When the cluster operator is assumed to be of the form (obvious for infinitely separated systems) $\hat{T} = \hat{T}_A + \hat{T}_B$, then the exponential form of the wave operator $\exp(\hat{T}_A + \hat{T}_B)$ ensures a desired form of the product of the wave function $[\exp(\hat{T}_A + \hat{T}_B)]\Phi_0 = \exp\hat{T}_A \exp\hat{T}_B\Phi_0$. *If we took a finite CI expansion*: $(\hat{T}_A + \hat{T}_B)\Phi_0$, *then we would not get the product but the sum which is incorrect*. With this reasoning there is a problem, since due to the Pauli principle (antisymmetry of the wave function with respect to the electron exchange) for long distance neither the function ψ nor the function Φ_0 are the product of the functions for the subsystems.[72] Although the reasoning is not quite correct, the conclusion is correct, as will be shown at the end of the description of the CC method (p. 547).

> The CC method is automatically size consistent.

As a cluster operator $\hat{T}$ we assume a sum of the excitation operators (see Appendix U)

$$\hat{T} = \hat{T}_1 + \hat{T}_2 + \hat{T}_3 + \cdots + \hat{T}_{l_{\max}} \tag{10.35}$$

where

$$\hat{T}_1 = \sum_{a,r} t_a^r \hat{r}^\dagger \hat{a} \tag{10.36}$$

---

[70] It contributes significantly to the numerical efficiency of the method.

[71] This word has survived in the literature in its original German form.

[72] For instance, the RHF function for the hydrogen molecule is not a product function for long distances, see p. 520.

is an operator for single excitations,

$$\hat{T}_2 = \frac{1}{4}\sum_{\substack{ab\\rs}} t_{ab}^{rs}\hat{s}^\dagger\hat{r}^\dagger\hat{a}\hat{b}, \tag{10.37}$$

is an operator for double excitations, etc. The subscript l in $\hat{T}_l$ indicates the rank of the excitations involved (with respect to the vacuum state). The symbols $a, b, \dots$ refer to the spinorbitals occupied in Φ_0, and $p, q, r, s, \dots$ refer to the unoccupied ones, and

> t represents *amplitudes*, i.e. the numbers whose determination is the goal of the CC method. The rest of this chapter will be devoted to the problem of how we can obtain these miraculous amplitudes.

In the CC method we want to obtain correct results with the assumption that $l_{\max}$ *of eq.* (10.35) *is relatively small* (usually $2 \div 5$). If $l_{\max}$ were equal to N, i.e. to the number of electrons, then the CC method would be identical to the full (usually unfeasible) CI method.

10.14.2 RELATIONSHIP BETWEEN CI AND CC METHODS

Obviously, there is a relation between the CI and CC methods. For instance, if we write $\exp(\hat{T})\Phi_0$ in such a way as to resemble the CI expansion

$$\begin{aligned}\exp(\hat{T})\Phi_0 &= \left[1 + (\hat{T}_1 + \hat{T}_2 + \hat{T}_3 + \cdots) + \frac{1}{2}(\hat{T}_1 + \hat{T}_2 + \hat{T}_3 + \cdots)^2 + \cdots\right]\Phi_0\\ &= (1 + \hat{C}_1 + \hat{C}_2 + \hat{C}_3 + \cdots)\Phi_0,\end{aligned} \tag{10.38}$$

the operators $\hat{C}_i$, pertaining to the CI method, have the following structure

$$\begin{aligned}\hat{C}_1 &= \hat{T}_1,\\ \hat{C}_2 &= \hat{T}_2 + \frac{1}{2!}\hat{T}_1^2,\\ \hat{C}_3 &= \hat{T}_3 + \frac{1}{3!}\hat{T}_1^3 + \hat{T}_1\hat{T}_2,\\ \hat{C}_4 &= \hat{T}_4 + \frac{1}{4!}\hat{T}_1^4 + \frac{1}{2!}\hat{T}_2^2 + \hat{T}_3\hat{T}_1 + \frac{1}{2!}\hat{T}_1^2\hat{T}_2,\\ &\dots\end{aligned} \tag{10.39}$$

factorizable part of CI coefficient

We see that the multiple excitations $\hat{C}_l$ result from mathematically distinct terms, e.g., $\hat{C}_3$ is composed of $\hat{T}_3$, $\hat{T}_1^3$ and $\hat{T}_1\hat{T}_2$. Sometimes we speak about the *factorizable part* of the CI coefficient (like $\hat{T}_1^3$ and $\hat{T}_1\hat{T}_2$) multiplying the particular Slater deter-

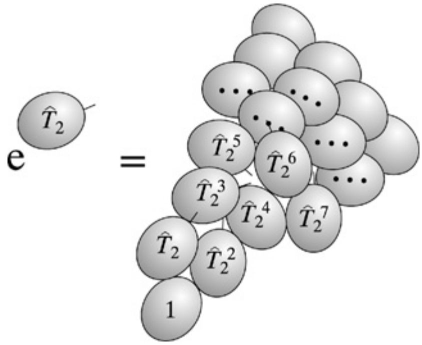


Fig. 10.10. Why such a name? An artistic impression on *coupled clusters*.

minant (corresponding to an n-tuple excitation) as the part which can be expressed in terms of the *lower* rank amplitudes.

On the basis of current numerical experience,[73] we believe that, within the excitation of a given rank, the contributions coming from the correlational interactions of the electron pairs are the most important, e.g., within C_4 the $\frac{1}{2!}\hat{T}_2^2$ excitations containing the product of amplitudes for two electron pairs are the most important, $\hat{T}_4$ (which contains the amplitudes of quadruple excitations) is of little importance, since they correspond to the coupling of the motions of four electrons, the terms $\hat{T}_1$, $\hat{T}_3\hat{T}_1$ and $\hat{T}_1^2\hat{T}_2$ can be made small by using the MC SCF orbitals. Contemporary quantum chemists use diagrammatic language following Richard Feynman. The point is that the mathematical terms (the energy contributions) appearing in CC theory can be translated – one by one – into the figures according to certain rules. It turns out that it is much easier (at least at lower orders) to think in terms of diagrams than to speak about the mathematical formulae or to write them out.

The coupled cluster method, terminated at $\hat{T}_2$ in the cluster operator automatically includes $\hat{T}_2^2$, etc. We may see in it some resemblance to a group of something (excitations), or in other words to a cluster, Fig. 10.10.

10.14.3 SOLUTION OF THE CC EQUATIONS

The strategy of the CC method is the following: first, we make a decision with respect to $l_{\max}$ in the cluster expansion (10.35) ($l_{\max}$ should be small[74]).

The exact wave function $\exp(\hat{T})\Phi_0$ satisfies the Schrödinger equation, i.e.

$$\hat{H}\exp(\hat{T})\Phi_0 = E\exp(\hat{T})\Phi_0, \tag{10.40}$$

which, after operating from the left with $\exp(-\hat{T})$ gives:

$$\exp(-\hat{T})\hat{H}\exp(\hat{T})\Phi_0 = E\Phi_0. \tag{10.41}$$

---

[73] This is a contribution by Oktay Sinanoğlu; O. Sinanoğlu, K.A. Brueckner (eds.), "*Three Approaches to Electron Correlation in Atoms*", Yale Univ. Press, New Haven and London, 1970.

[74] Only then is the method cost-effective.

The $\exp(-\hat{T})\hat{H}\exp(\hat{T})$ operator can be expressed in terms of the commutators (see (10.29))[75]

$$e^{-\hat{T}}\hat{H}e^{\hat{T}} = \hat{H} + [\hat{H},\hat{T}] + \frac{1}{2!}[[\hat{H},\hat{T}],\hat{T}] + \frac{1}{3!}[[[\hat{H},\hat{T}],\hat{T}],\hat{T}] + \frac{1}{4!}[[[[\hat{H},\hat{T}],\hat{T}],\hat{T}],\hat{T}]. \tag{10.42}$$

The expansion (10.42) is *finite* (justification can be only diagrammatic, and is not given here) since in the Hamiltonian $\hat{H}$ we have only two-particle interactions. Substituting this into the Schrödinger equation we have:

$$\left\{\hat{H} + [\hat{H},\hat{T}] + \frac{1}{2!}[[\hat{H},\hat{T}],\hat{T}] + \frac{1}{3!}[[[\hat{H},\hat{T}],\hat{T}],\hat{T}] + \frac{1}{4!}[[[[\hat{H},\hat{T}],\hat{T}],\hat{T}],\hat{T}]\right\}\Phi_0 = E\Phi_0. \tag{10.43}$$

Multiplying from the left with the function $\langle^{mn\ldots}_{ab\ldots}|$ representing the determinant obtained from the vacuum state by the action of the annihilators $\hat{a}, \hat{b}, \ldots$ and creators $\hat{n}^\dagger, \hat{m}^\dagger, \ldots$ and integrating, we obtain one equation for each function used:

$$\langle^{mn\ldots}_{ab\ldots}|\hat{H} + [\hat{H},\hat{T}] + \frac{1}{2!}[[\hat{H},\hat{T}],\hat{T}] + \frac{1}{3!}[[[\hat{H},\hat{T}],\hat{T}],\hat{T}] + \frac{1}{4!}[[[[\hat{H},\hat{T}],\hat{T}],\hat{T}],\hat{T}]|\Phi_0\rangle = 0, \tag{10.44}$$

where we have zero on the right-hand side due to the orthogonality. The Slater determinants $|^{mn}_{ab}\rangle$ represent all excitations from Φ_0 resulting from the given cluster expansion $\hat{T} = \hat{T}_1 + \hat{T}_2 + \cdots + \hat{T}_{l_{\max}}$. This is the fundamental equation of the CC method. For such a set of excited configurations the number of CC equations is equal to the number of the amplitudes sought.

$t^{mn\ldots}_{ab\ldots}$ are unknown quantities, i.e. amplitudes determining the $\hat{T}_l$, and, consequently, the wave operator (10.34) and wave function for the ground state $\Psi = \Psi_0$. The equations we get in the CC method are nonlinear

since the t's occur at higher powers than the first (it can be seen from eq. (10.44) that the highest power of t is 4), which, on one hand, requires much more demanding and capricious (than linear ones) numerical procedures, and, on the other, contributes to the greater efficiency of the method. The number of such equations very

nonlinearity

---

[75] It is straightforward to demonstrate the correctness of the first few terms by expanding the wave operator in the Taylor series.

often exceeds 100 000 or a million.[76] These equations are solved iteratively assuming certain starting amplitudes t and iterating the equations until self-consistency is achieved.

We hope that in such a procedure an approximation to the ground state wave function is obtained, although sometimes an unfortunate starting point may lead to some excited state.[77]

We usually use as a starting point that which is obtained from the linear version (reduced to obtain a linearity) of the CC method. We will write down these equations as $t_{ab}^{mn} = \ldots$ various powers of all t for all amplitudes. First we neglect the non-linear terms, this represents the initial approximation. The amplitudes are substituted into the right-hand side and we iterate until self-consistency is achieved. When all the amplitudes are found, then we obtain the energy E by projecting eq. (10.44) against Φ_0 function instead of $|{}^{mn}_{ab}\rangle$:

$$E = \langle \Phi_0 | e^{-\hat{T}} \hat{H} e^{\hat{T}} \Phi_0 \rangle. \tag{10.45}$$

The operator $(e^{-T})^\dagger$, conjugate to e^{-T}, is $e^{-T^\dagger}$, i.e.

$$E = \langle e^{-\hat{T}^\dagger} \Phi_0 | \hat{H} e^{\hat{T}} \Phi_0 \rangle, \tag{10.46}$$

which *is not* the mean value of the Hamiltonian. Hence, the CC method is not variational. If we multiplied eq. (10.40) from the left by $e^{\hat{T}^\dagger}$ we would obtain the variational character of E

$$E = \frac{\langle \Phi_0 | e^{\hat{T}^\dagger} \hat{H} e^{\hat{T}} \Phi_0 \rangle}{\langle \Phi_0 | e^{\hat{T}^\dagger} e^{\hat{T}} \Phi_0 \rangle} = \frac{\langle e^{\hat{T}} \Phi_0 | \hat{H} | e^{\hat{T}} \Phi_0 \rangle}{\langle e^{\hat{T}} \Phi_0 | e^{\hat{T}} \Phi_0 \rangle}. \tag{10.47}$$

However, it would not be possible to apply the commutator expansion and instead of the four terms in eq. (10.42) we would have an infinite number (due to the full normalization of the final function[78]). For this reason, we prefer the non-variational approach.

10.14.4 EXAMPLE: CC WITH DOUBLE EXCITATIONS

How does the CC machinery work? Let us show it for a relatively simple case $\hat{T} = \hat{T}_2$. Eq. (10.44), written without the commutator expansion, has the form

$$\langle {}^{mn}_{ab} | e^{-\hat{T}_2} \hat{H} e^{\hat{T}_2} \Phi_0 \rangle = 0. \tag{10.48}$$

---

[76] This refers to calculations with $\hat{T} = \hat{T}_2$ for ca. 10 occupied orbitals (for instance, two water molecules) and 150 virtual orbitals. These are not calculations for large systems.

[77] The first complete analysis of all CC solutions was performed by K. Jankowski and K. Kowalski, *Phys. Rev. Letters* 81 (1998) 1195; *J. Chem. Phys.* 110 (1999) 37, 93; *ibid.* 111 (1999) 2940, 2952. Recapitulation can be found in K. Jankowski, K. Kowalski, I. Grabowski, H.J. Monkhorst, *Intern. J. Quantum Chem.* 95 (1999) 483.

[78] The (non-variational) CC method benefits from the very economical condition of the intermediate normalization.

Taking advantage of the commutator expansion we have

$$\begin{aligned}\left\langle {}^{mn}_{ab}\middle| e^{-\hat{T}_2}\hat{H}e^{\hat{T}_2}\Phi_0\right\rangle &= \left\langle {}^{mn}_{ab}\middle|\left(1-\hat{T}_2+\frac{1}{2}\hat{T}_2^2+\cdots\right)\hat{H}\left(1+\hat{T}_2+\frac{1}{2}\hat{T}_2^2+\cdots\right)\Phi_0\right\rangle\\ &= \left\langle {}^{mn}_{ab}\middle|\hat{H}\Phi_0\right\rangle+\left\langle {}^{mn}_{ab}\middle|\hat{H}\hat{T}_2\Phi_0\right\rangle+\frac{1}{2}\left\langle {}^{mn}_{ab}\middle|\hat{H}\hat{T}_2^2\Phi_0\right\rangle-\left\langle {}^{mn}_{ab}\middle|\hat{T}_2\hat{H}\Phi_0\right\rangle\\ &\quad-\left\langle {}^{mn}_{ab}\middle|\hat{T}_2\hat{H}\hat{T}_2\Phi_0\right\rangle+A=0.\end{aligned}$$

However,

$$\begin{aligned}A &= -\frac{1}{2}\left\langle {}^{mn}_{ab}\middle|\hat{T}_2\hat{H}\hat{T}_2^2\Phi_0\right\rangle+\frac{1}{2}\left\langle {}^{mn}_{ab}\middle|\hat{T}_2^2\hat{H}\Phi_0\right\rangle+\frac{1}{2}\left\langle {}^{mn}_{ab}\middle|\hat{T}_2^2\hat{H}\hat{T}_2\Phi_0\right\rangle+\frac{1}{4}\left\langle {}^{mn}_{ab}\middle|\hat{T}_2^2\hat{H}\hat{T}_2^2\Phi_0\right\rangle\\ &=0.\end{aligned}$$

excitations and deexcitations

The last equality follows from the fact that each term is equal to zero. The first vanishes since both determinants differ by four excitations. Indeed, $\langle(\hat{T}_2^\dagger)^{mn}_{ab}|$ denotes a double *deexcitation*[79] of the doubly excited function, i.e. something proportional to $\langle\Phi_0|$. For similar reasons (too strong deexcitations give zero) the remaining terms in A also vanish. As a result we need to solve the equation:

$$\left\langle {}^{mn}_{ab}\middle|\hat{H}\Phi_0\right\rangle+\left\langle {}^{mn}_{ab}\middle|\hat{H}\hat{T}_2\Phi_0\right\rangle+\frac{1}{2}\left\langle {}^{mn}_{ab}\middle|\hat{H}\hat{T}_2^2\Phi_0\right\rangle-\left\langle {}^{mn}_{ab}\middle|\hat{T}_2\hat{H}\Phi_0\right\rangle-\left\langle {}^{mn}_{ab}\middle|\hat{T}_2\hat{H}\hat{T}_2\Phi_0\right\rangle=0.$$

After *several days*[80] of algebraic manipulations, we get the equations for the t amplitudes (for each t^{mn}_{ab} amplitude one equation):

$$\begin{aligned}&(\varepsilon_m+\varepsilon_n-\varepsilon_a-\varepsilon_b)t^{mn}_{ab}\\ &\quad=\langle mn|ab\rangle-\sum_{p>q}\langle mn|pq\rangle t^{pq}_{ab}-\sum_{\gamma>\delta}\langle cd|ab\rangle t^{mn}_{cd}\\ &\qquad+\sum_{c,p}\left[\langle cn|bp\rangle t^{mp}_{ac}-\langle cm|bp\rangle t^{np}_{ac}-\langle cn|ap\rangle t^{mp}_{bc}+\langle cm|ap\rangle t^{np}_{bc}\right]\\ &\qquad+\sum_{c>d,\ p>q}\langle cd|pq\rangle\left[t^{pq}_{ab}t^{mn}_{cd}-2\left(t^{mp}_{ab}t^{nq}_{cd}+t^{nq}_{ab}t^{mp}_{cd}\right)\right.\\ &\qquad\left.-2\left(t^{mn}_{ac}t^{pq}_{bd}+t^{pq}_{ac}t^{mn}_{bd}\right)+4\left(t^{mp}_{ac}t^{nq}_{bd}+t^{nq}_{ac}t^{mp}_{bd}\right)\right].\end{aligned}\tag{10.49}$$

It can be seen that the last expression includes: the term independent of t, the linear terms and the quadratic terms.

How can we find the t's? We do it with the help of the iterative method.[81] First, we substitute zeros for all t's on the right-hand side of the equation. Thus, from

---

[79] Opposite to excitation.

[80] Students – more courage!

[81] We organize things in such a way that a given unknown parameter will occur in the simple form on one side of the equation, whereas the more complicated terms, also containing the parameter sought,

the left-hand side the first approximation to t_{ab}^{mn} is[82] $t_{ab}^{mn} \cong \frac{\langle mn|ab\rangle}{(\varepsilon_m+\varepsilon_n-\varepsilon_a-\varepsilon_b)}$. We have now an estimate of each amplitude – we are making progress. The approximation to t obtained in this way is substituted into the right-hand side to evaluate the left-hand side and so forth. Finally, we achieve a self-consistency of the iterative process and obtain the CC wave function for the ground state of our system. With the amplitudes we calculate the energy of the system with eq. (10.45).

This is how the CCD (the CC with double excitations in the cluster operator) works from the practical viewpoint. It is more efficient when the initial amplitudes are taken from a short CI expansion,[83] with subsequent linearization (as above) of terms containing the initial (known) amplitudes.

The computational cost of the CCD and CCSD (singles and doubles) methods scales as N^6 where N is a number of molecular orbitals (occupied and virtual[84]), whereas the analogous cost of the CCSDT (singles, doubles, triples) method requires N^8 scaling. This means that, if we increase the orbital basis twice, the increase in the computational cost of the CCSDT method will be four times larger than that of the CCSD scheme. This is a lot, and because of this, wide-spread popularity has been gained for the CCSD(T) method, which only partly uses the triple excitations.

10.14.5 SIZE CONSISTENCY OF THE CC METHOD

The size consistency of the CC method can be proved on the basis of eq. (10.43) and (10.44). Let us assume that the system dissociates into two[85] non-interacting subsystems A and B (i.e. at infinite distance). Then the orbitals can be also divided into two separable (mutually orthogonal) subsets. We will show[86] that the cluster amplitudes, having *mixed* indices (from the first and second groups of orbitals), are equal to 0.

Let us note first that, for an infinite distance, the Hamiltonian $\hat{H} = \hat{H}_A + \hat{H}_B$. In such a situation the wave operator can be expressed as

$$\hat{T} = \hat{T}_A + \hat{T}_B + \hat{T}_{AB}, \tag{10.50}$$

---

are kept on the other side of the equation. Then we take a certain random value of the unknown and calculate the complicated side of the equation. On the left-hand side we will then have its *new* approximation to the true value. We repeat the whole procedure so many times until we start getting the same value (if the procedure converges). Then the equation is solved.

[82] As we see we would have trouble if $(\varepsilon_m + \varepsilon_n - \varepsilon_a - \varepsilon_b)$ is close to 0 (quasidegeneracy of the vacuum state with some other state), because then $t_{ab}^{mn} \to \infty$.

[83] The configuration interaction method with inclusion of single and double excitations only:

CCD: J.A. Pople, R. Krishnan, H.B. Schlegel, J.S. Binkley, *Intern. J. Quantum Chem.* S14 (1978) 545; R.J. Bartlett, G.D. Purvis III, *Intern. J. Quantum Chem.* S14 (1978) 561.

CCSD: G.D. Purvis III, *J. Chem. Phys*. 76 (1982) 1910.

[84] These estimations are valid for the same relative increase of the number of occupied and virtual orbitals, as it is, e.g., for going from a molecule to its dimer. In the case of calculations for the same molecule, but two atomic basis sets (that differ in size) the cost increases only as N^4.

[85] This can be generalized to many non-interacting subsystems.

[86] B. Jeziorski, J. Paldus, P. Jankowski, *Intern. J. Quantum Chem*. 56 (1995) 129.

where $\hat{T}_A$,$\hat{T}_B$, $\hat{T}_{AB}$ include the operators corresponding to spinorbitals from the subsystems A, B and from the system AB, respectively. Of course, in this situation we have the following commutation condition:

$$[\hat{H}_A, \hat{T}_B] = [\hat{H}_B, \hat{T}_A] = 0. \tag{10.51}$$

Then, owing to the commutator expansion in eq. (10.42), we obtain:

$$e^{-\hat{T}}(\hat{H}_A + \hat{H}_B)e^{\hat{T}} = e^{-\hat{T}_A}\hat{H}_A e^{\hat{T}_A} + e^{-\hat{T}_B}\hat{H}_B e^{\hat{T}_B} + O(\hat{T}_{AB}), \tag{10.52}$$

where $O(\hat{T}_{AB})$ denotes the linear and higher terms in $\hat{T}_{AB}$. Substituting this into eq. 10.44 with *bra* $\langle\text{mixed}|$ vector representing mixed excitation, we observe that the first two terms on the right-hand side of the last equation give zero. It means that we get the equation

$$\langle\text{mixed}|O(\hat{T}_{AB})\Phi_0\rangle = 0, \tag{10.53}$$

which, due to the linear term in $O(\hat{T}_{AB})$, is fulfilled by $\hat{T}_{AB} = 0$. Conclusion: for the infinite distance between the subsystems we do not have mixed amplitudes and the energy of the AB system is bound to be the sum of the energies of subsystem A and subsystem B (size consistency).

10.15 EQUATION-OF-MOTION METHOD (EOM-CC)

The CC method is used to calculate the *ground* state energy and wave function. What about the excited states? This is a task for the equation-of-motion CC method, the primary goal being not the excited states themselves, but the excitation energies with respect to the ground state.

10.15.1 SIMILARITY TRANSFORMATION

Let us note that for the Schrödinger equation $\hat{H}\psi = E\psi$, we can perform an interesting sequence of transformations based on the wave operator $e^{\hat{T}}$:

$$e^{-\hat{T}}\hat{H}\psi = Ee^{-\hat{T}}\psi,$$
$$e^{-\hat{T}}\hat{H}e^{\hat{T}}e^{-\hat{T}}\psi = Ee^{-\hat{T}}\psi.$$

We obtain the eigenvalue equation again, but for the *similarity transformed Hamiltonian*

$$\hat{\mathcal{H}}\bar{\psi} = E\bar{\psi},$$

where $\hat{\mathcal{H}} = e^{-\hat{T}}\hat{H}e^{\hat{T}}$, $\bar{\psi} = e^{-\hat{T}}\psi$, and the energy E does not change at all after this transformation. This result will be very useful in a moment.

10.15.2 DERIVATION OF THE EOM-CC EQUATIONS

As the reference function in the EOM-CC method, we take the coupled-cluster wave function for the ground state:

$$\psi_0 = \exp(\hat{T})\Phi_0, \tag{10.54}$$

where Φ_0 is usually a Hartree–Fock determinant. Now, we define the operator $\hat{U}_k$ ("EOM-CC Ansatz"), which performs a miracle: from the wave function of the ground state ψ_0 it creates the wave function ψ_k for the k-th excited state of the system:

$$\psi_k = \hat{U}_k \psi_0.$$

The operators $\hat{U}_k$ change the coefficients in front of the configurations (see p. 526). The operators $\hat{U}_k$ are (unlike the wave operator $\exp(\hat{T})$) linear with respect to the excitations, i.e. the excitation amplitudes occur there in the first powers. For the case of the single and double excitations (EOM-CCSD) we have $\hat{T}$ in the form of the sum of single and double excitations:

$$\hat{T} = \hat{T}_1 + \hat{T}_2$$

and

$$\hat{U}_k = \hat{U}_{k,0} + \hat{U}_{k,1} + \hat{U}_{k,2},$$

where the task for the $\hat{U}_{k,0}$ operator is to change the coefficient in front of the function Φ_0 to that appropriate to the $|k\rangle$ function. The role of the operators $\hat{U}_{k,1}$, $\hat{U}_{k,2}$ is an appropriate modification of the coefficients in front of the singly and doubly excited configurations. These tasks are done by the excitation operators with τ amplitudes (they have to be distinguished from the t amplitudes of the CC method):

$$\begin{aligned}
\hat{U}_{k,0} &= \tau_0(k),\\
\hat{U}_{k,1} &= \sum_{a,p} \tau_a^p(k)\hat{p}^\dagger \hat{a},\\
\hat{U}_{k,2} &= \sum_{a,b,p,q} \tau_{ab}^{pq}(k)\hat{q}^\dagger \hat{p}^\dagger \hat{a}\hat{b},
\end{aligned}$$

where the amplitudes $\tau(k)$ are numbers, which are the targets of the EOM-CC method. The amplitudes give the wave function ψ_k and the energy E_k.

We write down the Schrödinger equation for the excited state:

$$\hat{H}\psi_k = E_k \psi_k.$$

Now we substitute the EOM-CC *Ansatz*:

$$\hat{H}\hat{U}_k \psi_0 = E_k \hat{U}_k \psi_0,$$

and from the definition of the CC wave operator we get[87]

$$\hat{H}\hat{U}_k \exp(\hat{T})\Phi_0 = E_k \hat{U}_k \exp(\hat{T})\Phi_0.$$

---

[87] By neglecting higher than single and double excitations the equation represents an approximation.

Due to the missing *deexcitation* part (i.e. that which lowers the excitation rank, e.g., from doubles to singles) the operators $\hat{U}_k$ and $\hat{T}$ commute,[88] hence the operators $\hat{U}_k$ and $\exp(\hat{T})$ also commute:

$$\hat{U}_k \exp(\hat{T}) = \exp(\hat{T})\hat{U}_k.$$

Substituting this we have:

$$\hat{H}\exp(\hat{T})\hat{U}_k\Phi_0 = E_k \exp(\hat{T})\hat{U}_k\Phi_0$$

and multiplying from the left with $\exp(-\hat{T})$ we get:

$$\bigl[\exp(-\hat{T})\hat{H}\exp(\hat{T})\bigr]\hat{U}_k\Phi_0 = E_k\hat{U}_k\Phi_0$$

or introducing the similarity transformed Hamiltonian

$$\hat{\mathcal{H}} = e^{-\hat{T}}\hat{H}e^{\hat{T}}$$

we obtain

$$\hat{\mathcal{H}}\hat{U}_k\Phi_0 = E_k\hat{U}_k\Phi_0.$$

From the last equation we subtract the CC equation for the ground state

$$\bigl[\exp(-\hat{T})\hat{H}\exp(\hat{T})\bigr]\Phi_0 = E_0\Phi_0$$

multiplied from the left with $\hat{U}_k$, i.e. $\hat{U}_k\hat{\mathcal{H}}\Phi_0 = E_0\hat{U}_k\Phi_0$ and we get

$$\hat{\mathcal{H}}\hat{U}_k\Phi_0 - \hat{U}_k\hat{\mathcal{H}}\Phi_0 = E_k\hat{U}_k\Phi_0 - E_0\hat{U}_k\Phi_0.$$

Finally, we obtain an important result:

$$\bigl[\hat{\mathcal{H}}, \hat{U}_k\bigr]\Phi_0 = (E_k - E_0)\hat{U}_k\Phi_0.$$

The operator $\hat{U}_k$ contains the sought amplitudes $\tau(k)$.

We find them in a similar manner as in the CC method. For that purpose we make a scalar product of the left- and right-hand side of that equation with each excitation $|^{mn\ldots}_{ab\ldots}\rangle$ used in $\hat{U}_k$, including[89] that of no excitation, i.e. the function Φ_0. We get the set of the EOM-CC equations whose number is equal to the number of sought amplitudes plus one more equation due to normalization condition of ψ_k.

---

[88] If $\hat{U}_k$ contains true excitations, then it does not matter whether excitations are performed by $\hat{U}_k\hat{T}$ or $\hat{T}\hat{U}_k$ (commutation), because both $\hat{U}_k$ and $\hat{T}$ mean going up in the energy scale. If, however, $\hat{U}_k$ contains deexcitations, then it may happen that there is an attempt in $\hat{T}\hat{U}_k$ to deexcite the ground state wave function – that makes immediately 0, whereas $\hat{U}_k\hat{T}$ may be still OK, because the excitations in $\hat{T}$ may be more important than the deexcitations in $\hat{U}_k$.

[89] More precisely: to get only the excitation energy we do not need the coefficient next to Φ_0.

The unknown parameters are amplitudes and the excitation energies $E_k - E_0$:

$$\left\langle^{mn\ldots}_{ab\ldots}\middle|[\hat{\mathcal{H}}, \hat{U}_k]\middle|\Phi_0\right\rangle = (E_k - E_0)\left\langle^{mn\ldots}_{ab\ldots}\middle|\hat{U}_k\middle|\Phi_0\right\rangle.$$

We solve these equations and the problem is over.

10.16 MANY BODY PERTURBATION THEORY (MBPT)

The majority of routine calculations in quantum chemistry are done with variational methods (mainly the Hartree–Fock scheme). If we consider post-Hartree–Fock calculations then non-variational [CCSD, CCSD(T)] as well as perturbational (among them MBPT) approaches take the lead. The perturbational methods are based on the simple idea that the system, in a slightly modified condition, is similar to that before the perturbation is applied (cf. p. 203).

In the formalism of perturbation theory, knowing the unperturbed system and the perturbation we are able to provide successive corrections to obtain the solution of the perturbed system. Thus, for instance, the energy of the perturbed system is the energy of the unperturbed system plus the first-order correction plus the second-order correction plus…, etc.[90] If the perturbation is small then we *hope*[91] the series is convergent, even then however, there is no guarantee that the series converges fast.

10.16.1 UNPERTURBED HAMILTONIAN

In the perturbational approach (cf. p. 204) to the electron correlation the Hartree–Fock function, Φ_0, is treated as the zero-order approximation to the true ground state wave function, i.e. $\Phi_0 = \psi_0^{(0)}$. Thus, the Hartree–Fock wave function stands at the starting point, while the goal is the exact ground-state electronic wave function ψ_0.

In majority of cases this is a reasonable approximation, since the Hartree–Fock method usually provides as much as 98–99% of the total energy.[92] A Slater determinant Φ_I is constructed from the spinorbitals satisfying the Fock equation. How to construct the operator for which the Slater determinant is an eigenfunction? We will find out in a moment that this operator is the sum of the Fock operators (cf. Appendix U)

$$\hat{H}^{(0)} = \sum_i \hat{F}(i) = \sum_i \varepsilon_i \hat{\imath}^\dagger \hat{\imath}. \tag{10.55}$$

---

[90] This is an old trick of perturbation theory, equivalent to saying that the shape of a bridge loaded with a car is the shape of the bridge without the car plus the deformation proportional to the mass of the car plus the deformation proportional to the square of the mass of the car, etc. This works, if the bridge is solid and the car is light (the perturbation is small).

[91] There is not much known concerning the convergence of series occurring in quantum chemistry. Commonly, only a few perturbational corrections are computed.

[92] Sometimes, as we know, the method fails and then the perturbation theory based on the Hartree–Fock starting point is a risky business, since the perturbation is very large.

Indeed,

$$\hat{H}^{(0)}\Phi_I = \sum_i \varepsilon_i \hat{\imath}^\dagger \hat{\imath} \cdot \Phi_I = \left(\sum_i \varepsilon_i\right) \cdot \Phi_I, \tag{10.56}$$

since the annihilation of one spinorbital in the determinant and the creation of the same spinorbital leaves the determinant unchanged. This is so on condition that the spinorbital ϕ_i *is present* in $\psi_0^{(0)}$.

> The eigenvalue of $\hat{H}_0 = \sum_i \varepsilon_i \hat{\imath}^\dagger \hat{\imath}$ is always the sum of the orbital energies corresponding to all spinorbitals in the Slater determinant Φ_I.

This means that the sum of several determinants, each built from a different (in the sense of the orbital energies) set of spinorbitals, is not an eigenfunction of $\hat{H}^{(0)}$.

10.16.2 PERTURBATION THEORY – SLIGHTLY DIFFERENT APPROACH

We have to solve the Schrödinger equation for the ground state[93] $\hat{H}\psi_0 = E\psi_0$, with $\hat{H} = \hat{H}^{(0)} + \hat{H}^{(1)}$, where $\hat{H}^{(0)}$ denotes the unperturbed Hamiltonian, and $\hat{H}^{(1)}$ is a perturbation operator. We assumed that $\hat{H}^{(0)}$ has eigenfunctions and corresponding energy eigenvalues

$$\hat{H}^{(0)}\psi_k^{(0)} = E_k^{(0)}\psi_k^{(0)}. \tag{10.57}$$

The ground state $\psi_0^{(0)}$ is non-degenerate (assumption).

The Schrödinger equation does not force the normalization of the function. It is convenient to use the *intermediate normalization* (Fig. 10.11.a), i.e. to require that $\langle\psi_0|\psi_0^{(0)}\rangle = 1$. This means that the (non-normalized) ψ_0 must include the normalized function of zeroth order $\psi_0^{(0)}$ *and, possibly, something orthogonal to it*.

Let us write $\hat{H}\psi_0$ as $\hat{H}\psi_0 = (\hat{H}^{(0)} + \hat{H}^{(1)})\psi_0$, or, in another way, as $\hat{H}^{(1)}\psi_0 = (\hat{H} - \hat{H}^{(0)})\psi_0$. Multiplying this equation by $\psi_0^{(0)}$ and integrating, we get (taking advantage of the intermediate normalization)

$$\begin{aligned}\langle\psi_0^{(0)}|\hat{H}^{(1)}\psi_0\rangle &= \langle\psi_0^{(0)}|(\hat{H} - \hat{H}^{(0)})\psi_0\rangle = E_0\langle\psi_0^{(0)}|\psi_0\rangle - \langle\psi_0^{(0)}|\hat{H}^{(0)}\psi_0\rangle \\ &= E_0 - E_0^{(0)} = \Delta E_0.\end{aligned} \tag{10.58}$$

Thus,

> $$\Delta E_0 = \langle\psi_0^{(0)}|\hat{H}^{(1)}\psi_0\rangle. \tag{10.59}$$

---

[93] We use the notation from Chapter 5.

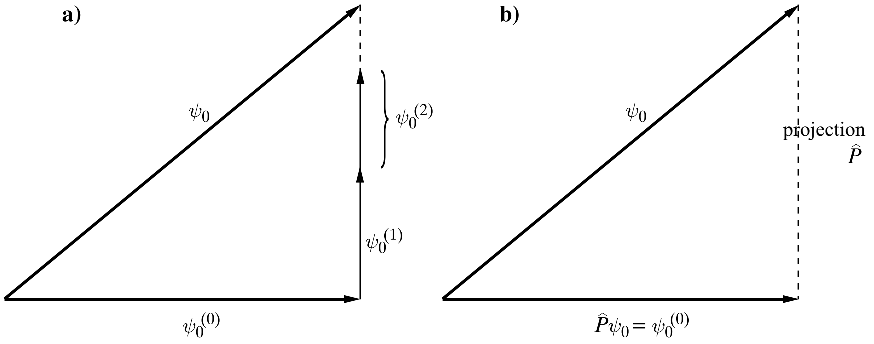


Fig. 10.11. Pictorial presentation of (a) the intermediate normalization $\langle\psi|\psi_0^{(0)}\rangle = 1$, $\psi_0^{(n)}$ is the n-th correction, and (b) the projection onto the axis $\psi_0^{(0)}$ in the Hilbert space using the operator $\hat{P} = |\psi_0^{(0)}\rangle\langle\psi_0^{(0)}|$.

10.16.3 REDUCED RESOLVENT OR THE "ALMOST" INVERSE OF $(E_0^{(0)} - \hat{H}^{(0)})$

Let us define several useful quantities – we need to get familiar with them now – which will introduce a certain elegance into our final equations.

Let the first be *a projection operator* on the ground-state zeroth order function

projection operator

$$\hat{P} = |\psi_0^{(0)}\rangle\langle\psi_0^{(0)}|. \tag{10.60}$$

This means that $\hat{P}\chi$ is, within accuracy to a constant, equal to $\psi_0^{(0)}$ for an arbitrary function χ. Indeed, if χ is expressed as a linear combination of the eigenfunctions $\psi_n^{(0)}$ (these functions form an orthonormal complete set as eigenfunctions of the Hermitian operator)

$$\chi = \sum_n c_n \psi_n^{(0)}, \tag{10.61}$$

then (Fig. 10.11.b)

$$\hat{P}\chi = \sum_n c_n \hat{P}\psi_n^{(0)} = \sum_n c_n |\psi_0^{(0)}\rangle\langle\psi_0^{(0)}|\psi_n^{(0)}\rangle = \sum_n c_n \delta_{0n}\psi_0^{(0)} = c_0\psi_0^{(0)}. \tag{10.62}$$

Let us now introduce a projection operator

$$\hat{Q} = 1 - \hat{P} = \sum_{n=1}^{\infty} |\psi_n^{(0)}\rangle\langle\psi_n^{(0)}| \tag{10.63}$$

in the space orthogonal to $\psi_0^{(0)}$. Obviously, $\hat{P}\hat{Q} = 0$, $\hat{P}^2 = \hat{P}$ and $\hat{Q}^2 = \hat{Q}$. The latter holds since $\hat{Q}^2 = (1-\hat{P})^2 = 1 - 2\hat{P} + \hat{P}^2 = 1 - \hat{P} = \hat{Q}$.

reduced resolvent

Now we define *a reduced resolvent*

$$\hat{R}_0 = \sum_{n=1}^{\infty} \frac{|\psi_n^{(0)}\rangle\langle\psi_n^{(0)}|}{E_0^{(0)} - E_n^{(0)}}. \tag{10.64}$$

For functions orthogonal to $\psi_0^{(0)}$, the action of the operator $\hat{R}_0$ is identical to that of the operator $(E_0^{(0)} - \hat{H}^{(0)})^{-1}$. Let us make sure of this. Let us operate first on the function ϕ orthogonal to $\psi_0^{(0)}$ with the operator $\hat{R}_0(E_0^{(0)} - \hat{H}^{(0)})$. The result should be equal to ϕ. Let us see:

$$\hat{R}_0\big(E_0^{(0)} - \hat{H}^{(0)}\big)\phi = \sum_{n=1}^{\infty}\big(E_0^{(0)} - E_n^{(0)}\big)^{-1}\big|\psi_n^{(0)}\big\rangle\big\langle\psi_n^{(0)}\big|\big(E_0^{(0)} - \hat{H}^{(0)}\big)\big|\phi\big\rangle \tag{10.65}$$

$$= \sum_{n=1}^{\infty}\big(E_0^{(0)} - E_n^{(0)}\big)^{-1}\big(E_0^{(0)} - E_n^{(0)}\big)\big|\psi_n^{(0)}\big\rangle\big\langle\psi_n^{(0)}\big|\phi\big\rangle \tag{10.66}$$

$$= \sum_{n=1}^{\infty}\big|\psi_n^{(0)}\big\rangle\big\langle\psi_n^{(0)}\big|\phi\big\rangle = \hat{Q}\phi = \phi, \tag{10.67}$$

since for ϕ orthogonal to $\psi_0^{(0)}$ the projection $\hat{Q}\phi$ equals ϕ. Let us now operate on the same function with the operator $(E_0^{(0)} - \hat{H}^{(0)})\hat{R}_0$ (i.e. the operators are in reverse order):

$$\big(E_0^{(0)} - \hat{H}^{(0)}\big)\hat{R}_0\phi = \big(E_0^{(0)} - \hat{H}^{(0)}\big)\sum_{n=1}^{\infty}\big(E_0^{(0)} - E_n^{(0)}\big)^{-1}\big|\psi_n^{(0)}\big\rangle\big\langle\psi_n^{(0)}\big|\phi\big\rangle$$

$$= \sum_{n=1}^{\infty}\big(E_0^{(0)} - E_n^{(0)}\big)^{-1}\big(E_0^{(0)} - \hat{H}^{(0)}\big)\big|\psi_n^{(0)}\big\rangle\big\langle\psi_n^{(0)}\big|\phi\big\rangle$$

$$= \sum_{n=1}^{\infty}\big|\psi_n^{(0)}\big\rangle\big\langle\psi_n^{(0)}\big|\phi\big\rangle = \hat{Q}\phi = \phi. \tag{10.68}$$

It really looks as if the $\hat{R}_0$ is the inverse of $(E_0^{(0)} - \hat{H}^{(0)})$. This is not so, since when acting on the function $\psi_0^{(0)}$ we get

$$\hat{R}_0\big(E_0^{(0)} - \hat{H}^{(0)}\big)\psi_0^{(0)} = \hat{R}_0 \cdot 0 = 0, \tag{10.69}$$

and not $\psi_0^{(0)}$. In other words $\hat{R}_0(E_0^{(0)} - \hat{H}^{(0)})\psi_0^{(0)} = 0 \neq \psi_0^{(0)}$.

Similarly,

$$\big(E_0^{(0)} - \hat{H}^{(0)}\big)\hat{R}_0\psi_0^{(0)} = \big(E_0^{(0)} - \hat{H}^{(0)}\big)\sum_{n=1}^{\infty}\big(E_0^{(0)} - E_n^{(0)}\big)^{-1}\big|\psi_n^{(0)}\big\rangle\big\langle\psi_n^{(0)}\big|\psi_0^{(0)}\big\rangle$$

$$= \big(E_0^{(0)} - \hat{H}^{(0)}\big)\cdot 0 = 0 \neq \psi_0^{(0)}.$$

Thus, the reduced resolvent is "almost" the inverse of $(E_0^{(0)} - \hat{H}^{(0)})$, almost, because it happens only when acting on the functions from the space orthogonal to $\psi_0^{(0)}$. When the reduced resolvent operates on an arbitrary function, the result belongs to the Q space, but it does not represent a projection on the Q space. Indeed, let us operate with $\hat{R}_0$ on function ϕ:

$$\begin{aligned}\hat{R}_0\phi &= \sum_{n=1}^{\infty} \big(E_0^{(0)} - E_n^{(0)}\big)^{-1} \big|\psi_n^{(0)}\big\rangle\big\langle\psi_n^{(0)}\big|\phi\big\rangle \\ &= \text{linear combination of functions orthogonal to } \psi_0^{(0)}. \qquad (10.70)\end{aligned}$$

Such a linear combination always belongs to the Q space, but we have not obtained ϕ, hence $\hat{R}_0$ is not a projection operator.

10.16.4 MBPT MACHINERY

Our goal now will be to present the Schrödinger equation in a different form. Let us first write it down as follows

$$\big(E_0 - \hat{H}^{(0)}\big)\psi_0 = \hat{H}^{(1)}\psi_0. \qquad (10.71)$$

We aim at having $(E_0^{(0)} - \hat{H}^{(0)})\psi_0$ on the left-hand side. Let us add $(E_0^{(0)} - E_0)\psi_0$ to both sides of that equation to obtain

$$\big(E_0 - \hat{H}^{(0)}\big)\psi_0 + \big(E_0^{(0)} - E_0\big)\psi_0 = \hat{H}^{(1)}\psi_0 + \big(E_0^{(0)} - E_0\big)\psi_0 \qquad (10.72)$$

or

$$\big(E_0^{(0)} - \hat{H}^{(0)}\big)\psi_0 = \big(E_0^{(0)} - E_0 + \hat{H}^{(1)}\big)\psi_0. \qquad (10.73)$$

Let us now operate on both sides of this equation with the reduced resolvent $\hat{R}_0$

$$\hat{R}_0\big(E_0^{(0)} - \hat{H}^{(0)}\big)\psi_0 = \hat{R}_0\big(E_0^{(0)} - E_0 + \hat{H}^{(1)}\big)\psi_0. \qquad (10.74)$$

On the left-hand side we have $\hat{Q}\psi_0$ (as follows from eq. (10.67)), but $\hat{Q}\psi_0 = (1 - \hat{P})\psi_0 = \psi_0 - |\psi_0^{(0)}\rangle\langle\psi_0^{(0)}|\psi_0\rangle = \psi_0 - \psi_0^{(0)}$, due to the intermediate normalization. As a result, the equation takes the form

$$\psi_0 - \psi_0^{(0)} = \hat{R}_0\big(E_0^{(0)} - E_0 + \hat{H}^{(1)}\big)\psi_0. \qquad (10.75)$$

Thus, we obtain

$$\psi_0 = \psi_0^{(0)} + \hat{R}_0\big(E_0^{(0)} - E_0 + \hat{H}^{(1)}\big)\psi_0. \qquad (10.76)$$

At the same time, based on the expression for ΔE in perturbation theory (eq. (10.59)), we have:

$$E_0 = E_0^{(0)} + \big\langle\psi_0^{(0)}\big|\hat{H}^{(1)}\psi_0\big\rangle. \qquad (10.77)$$

These are the equations of the many body perturbation theory, in which the exact wave function and energy are expressed in terms of the unperturbed functions and energies plus certain corrections. The problem is that, as can be seen, these corrections involve the unknown function and unknown energy.

Let us not despair in this situation, but try to apply an iterative technique. First substitute for ψ_0 in the right-hand side of (10.76) that, which most resembles ψ_0, i.e. $\psi_0^{(0)}$. We obtain

$$\psi_0 \cong \psi_0^{(0)} + \hat{R}_0\left(E_0^{(0)} - E_0 + \hat{H}^{(1)}\right)\psi_0^{(0)}, \tag{10.78}$$

and then the new approximation to ψ_0 should again be plugged into the right-hand side and this procedure is continued *ad infinitum*. It can be seen that the successive terms form a series (let us hope that it is convergent).

$$\psi_0 = \sum_{n=0}^{\infty}\left[\hat{R}_0\left(E_0^{(0)} - E_0 + \hat{H}^{(1)}\right)\right]^n \psi_0^{(0)}. \tag{10.79}$$

Now only known quantities occur on the right-hand side except for E_0, the exact energy. Let us pretend that its value is known and insert into the energy expression (10.77) the function ψ_0

$$\begin{aligned} E_0 &= E_0^{(0)} + \langle\psi_0^{(0)}|\hat{H}^{(1)}\psi_0\rangle \\ &= E_0^{(0)} + \langle\psi_0^{(0)}|\hat{H}^{(1)}\sum_{n=0}^{M}\left[\hat{R}_0\left(E_0^{(0)} - E_0 + \hat{H}^{(1)}\right)\right]^n|\psi_0^{(0)}\rangle. \end{aligned} \tag{10.80}$$

Let us go back to our problem: we want to have E_0 on the left-hand side of the last equation, while – for the time being – E_0 occurs on the right-hand sides of both equations. To exit the situation we will treat E_0 occurring on the right-hand side as a parameter manipulated in such a way as to obtain equality in both above equations. We may do it in two ways. One leads to Brillouin–Wigner perturbation theory, the other to Rayleigh–Schrödinger perturbation theory.

10.16.5 BRILLOUIN–WIGNER PERTURBATION THEORY

Let us decide first at what $n = M$ we terminate the series, i.e. to what order of perturbation theory the calculations will be carried out. Say, $M = 3$. Let us now take any reasonable value[94] as a parameter of E_0. We insert this value into the right-hand side of eq. (10.80) for E_0 and calculate the left-hand side, i.e. E_0. Then let us again insert the new E_0 into the right-hand side and continue in this way until self-consistency, i.e. until (10.80) is satisfied. After E_0 is known we go to eq. (10.79) and compute ψ_0 (through a certain order, e.g., M).

---

[94] A "unreasonable" value will lead to numerical instabilities. *Then* we will learn that it was unreasonable to take it.

Brillouin–Wigner perturbation theory has, as seen, the somewhat unpleasant feature that successive corrections to the wave function depend on the M assumed at the beginning.

We may suspect[95] – and this is true – that the Brillouin–Wigner perturbation theory is not size consistent.

10.16.6 RAYLEIGH–SCHRÖDINGER PERTURBATION THEORY

As an alternative to Brillouin–Wigner perturbation theory, we may consider Rayleigh–Schrödinger perturbation theory, which is size consistent. In this method the total energy is computed in a stepwise manner

$$E_0 = \sum_{k=0}^{\infty} E_0^{(k)} \tag{10.81}$$

in such a way that first we calculate the first order correction $E_0^{(1)}$, i.e. of the order of $\hat{H}^{(1)}$, then the second order correction, $E_0^{(2)}$, i.e. of the order of $(\hat{H}^{(1)})^2$, etc. If we insert into the right-hand side of (10.79) and (10.80) the expansion $E_0 = \sum_{k=0}^{\infty} E_0^{(k)}$ and then, by using the usual perturbation theory argument, we equalize the terms of the same order, we get for $n = 0$:

$$E_0^{(1)} = \langle \psi_0^{(0)} | \hat{H}^{(1)} \psi_0^{(0)} \rangle, \tag{10.82}$$

for $n = 1$:

$$E^{(2)} = \langle \psi_0^{(0)} | \hat{H}^{(1)} \hat{R}_0 \big(E_0^{(0)} - E_0 + \hat{H}^{(1)}\big) \psi_0^{(0)} \rangle = \langle \psi_0^{(0)} | \hat{H}^{(1)} \hat{R}_0 \hat{H}^{(1)} \psi_0^{(0)} \rangle, \tag{10.83}$$

since $\hat{R}_0 \psi_0^{(0)} = 0$; for $n = 2$:

$E^{(3)} =$ the third order terms from the expression:

$$\begin{aligned} &\langle \psi_0^{(0)} | \hat{H}^{(1)} \big[\hat{R}_0 \big(E_0^{(0)} - E_0^{(0)} - E_0^{(1)} - E_0^{(2)} - \cdots + \hat{H}^{(1)}\big)\big]^2 \psi_0^{(0)} \rangle \\ &= \langle \psi_0^{(0)} | \hat{H}^{(1)} \hat{R}_0 \big(-E_0^{(1)} - E_0^{(2)} - \cdots + \hat{H}^{(1)}\big) \hat{R}_0 \big(-E_0^{(1)} - E_0^{(2)} - \cdots + \hat{H}^{(1)}\big) \psi_0^{(0)} \rangle \end{aligned}$$

and the only terms of the third order are:

$$E^{(3)} = \langle \psi_0^{(0)} | \hat{H}^{(1)} \hat{R}_0 \hat{H}^{(1)} \hat{R}_0 \hat{H}^{(1)} \psi_0^{(0)} \rangle - E_0^{(1)} \langle \psi_0^{(0)} | \hat{H}^{(1)} R_0^2 \hat{H}^{(1)} \psi_0^{(0)} \rangle, \tag{10.84}$$

etc.

Unfortunately, we cannot give a general expression for the k-th correction to the energy although we can give an algorithm for the construction of such an expression.[96] Rayleigh–Schrödinger perturbation theory (unlike the Brillouin–Wigner approach) has the nice feature that the corrections of the particular orders are independent of the maximum order chosen.

---

[95] Due to the iterative procedure.

[96] J. Paldus and J. Čížek, *Adv. Quantum Chem.* 105 (1975).

10.17 MØLLER–PLESSET VERSION OF RAYLEIGH–SCHRÖDINGER PERTURBATION THEORY

Let us consider the case of a closed shell.[97] In the Møller–Plesset perturbation theory we assume as $\hat{H}^{(0)}$ the sum of the Hartree–Fock operators (from the RHF method), and $\psi_0^{(0)} = \psi_{\mathrm{RHF}}$, i.e.:

$$\hat{H}^{(0)} = \sum_i^{\infty} \varepsilon_i i^{\dagger} i,$$

$$\hat{H}^{(0)} \psi_{\mathrm{RHF}} = E_0^{(0)} \psi_{\mathrm{RHF}}, \tag{10.85}$$

$$E_0^{(0)} = \sum_i \varepsilon_i, \tag{10.86}$$

(the last summation is over spinorbitals occupied in the RHF function) hence the perturbation, known in the literature as *a fluctuation potential*, is equal

fluctuation potential

$$\hat{H}^{(1)} = \hat{H} - \hat{H}^{(0)}. \tag{10.87}$$

We may carry out calculations through a given order for such a perturbation. A very popular method relies on the inclusion of the perturbational corrections to the energy through the second order (known as MP2 method) and through the fourth order (MP4).

10.17.1 EXPRESSION FOR MP2 ENERGY

What is the expression for the total energy in the MP2 method?

Let us note first that, when calculating the mean value of the Hamiltonian in the standard Hartree–Fock method, we automatically obtain the sum of the zeroth order energies $\sum_i \varepsilon_i$ and the first-order correction to the energy $\langle \psi_{\mathrm{RHF}} | \hat{H}^{(1)} \psi_{\mathrm{RHF}} \rangle$:

$$E_{\mathrm{RHF}} = \langle \psi_{\mathrm{RHF}} | \hat{H} \psi_{\mathrm{RHF}} \rangle = \big\langle \psi_{\mathrm{RHF}} \big| \big(\hat{H}^{(0)} + \hat{H}^{(1)}\big) \psi_{\mathrm{RHF}} \big\rangle$$
$$= \Big(\sum_i \varepsilon_i \Big) + \big\langle \psi_{\mathrm{RHF}} \big| \hat{H}^{(1)} \psi_{\mathrm{RHF}} \big\rangle.$$

So what is left to be done (in the MP2 approach) is the addition of the second order correction to the energy (p. 208, the prime in the summation symbol indicates that the term making the denominator equal to zero is omitted), where, as the complete set of functions, we assume the Slater determinants $\psi_k^{(0)}$ corresponding to the energy $E_k^{(0)}$ (they are generated by various spinorbital occupancies):

$$E_{\mathrm{MP2}} = E_{\mathrm{RHF}} + \sum_k{}' \frac{|\langle \psi_k^{(0)} | \hat{H}^{(1)} \psi_{\mathrm{RHF}} \rangle|^2}{E_0^{(0)} - E_0^{(k)}}$$
$$= E_{\mathrm{RHF}} + \sum_k{}' \frac{|\langle \psi_k^{(0)} | \hat{H} \psi_{\mathrm{RHF}} \rangle|^2}{E_0^{(0)} - E_0^{(k)}}, \tag{10.88}$$

---

[97] Møller–Plesset perturbation theory also has its multireference formulation when the function Φ_0 is a linear combination of determinants (K. Woliński, P. Pulay, *J. Chem. Phys.* 90 (1989) 3647).

since ψ_{RHF} is an eigenfunction of $\hat{H}^{(0)}$, and $\psi_k^{(0)}$ and ψ_{RHF} are orthogonal. It can be seen that among possible functions $\psi_k^{(0)}$, we may ignore all but doubly excited ones. Why? This is because

- the single excitations give $\langle\psi_k^{(0)}|\hat{H}\psi_{\mathrm{RHF}}\rangle = 0$ due to the Brillouin theorem,
- the triple and higher excitations differ by more-than-two excitations from the functions ψ_{RHF} and, due to the IV Slater–Condon rule (see Appendix M, p. 986), give a contribution equal to 0.

In such a case, we take as the functions $\psi_k^{(0)}$ only doubly excited Slater determinants ψ_{ab}^{pq}, which means that we replace the occupied spinorbitals: $a \to p$, $b \to q$, and, to avoid repetitions $a < b$, $p < q$. These functions are eigenfunctions of $\hat{H}^{(0)}$ with the eigenvalues being the sum of the respective orbital energies (eq. (10.56)). Thus, using the III Slater–Condon rule, we obtain the energy accurate through the second order

$$E_{\mathrm{MP2}} = E_{\mathrm{RHF}} + \sum_{a<b,\ p<q}{}' \frac{|\langle ab|pq\rangle - \langle ab|qp\rangle|^2}{\varepsilon_a + \varepsilon_b - \varepsilon_p - \varepsilon_q}, \tag{10.89}$$

hence, the MP2 scheme viewed as an approximation to the correlation energy gives[98]

$$E_{\mathrm{corel}} \approx E_{\mathrm{MP2}} - E_{\mathrm{RHF}} = \sum_{a<b,\ p<q}{}' \frac{|\langle ab|pq\rangle - \langle ab|qp\rangle|^2}{\varepsilon_i + \varepsilon_j - \varepsilon_m - \varepsilon_n}. \tag{10.90}$$

10.17.2 CONVERGENCE OF THE MØLLER–PLESSET PERTURBATION SERIES

Does the Møller–Plesset perturbation procedure converge? Very often this question can be considered surrealist, since *most frequently* we carry out calculations through the second, third, and – at most – fourth order of perturbation theory. Such calculations usually give quite a satisfactory description of the physical quantities considered and we do not think about going to high orders requiring major computational effort. There were, however, scientists interested to see how fast the convergence is if very high orders are included (MPn) for $n < 45$. And there was a surprise (see Fig. 10.12).

It is true that the first few orders of the MP perturbation theory give reasonably good results, but later, the accuracy of the MP calculations gets worse. A lot depends on the atomic orbital basis set adopted and the wealthy people (using the augmented basis sets – which is much more rare) encounter some difficulties whereas poor ones (modest basis sets) do not. Moreover, for long bond lengths

---

[98] The MP2 method usually gives satisfactory results, e.g., the frequencies of the normal modes. There are indications, however, that the deformations of the molecule connected with some vibrations strongly affecting the electron correlation (vibronic coupling) create too severe a test for the method – the error may amount to 30–40% for frequencies of the order of hundreds of cm^{-1} as has been shown by D. Michalska, W. Zierkiewicz, D.C. Bieńko, W. Wojciechowski, T. Zeegers-Huyskens, *J. Phys. Chem.* A 105 (2001) 8734.

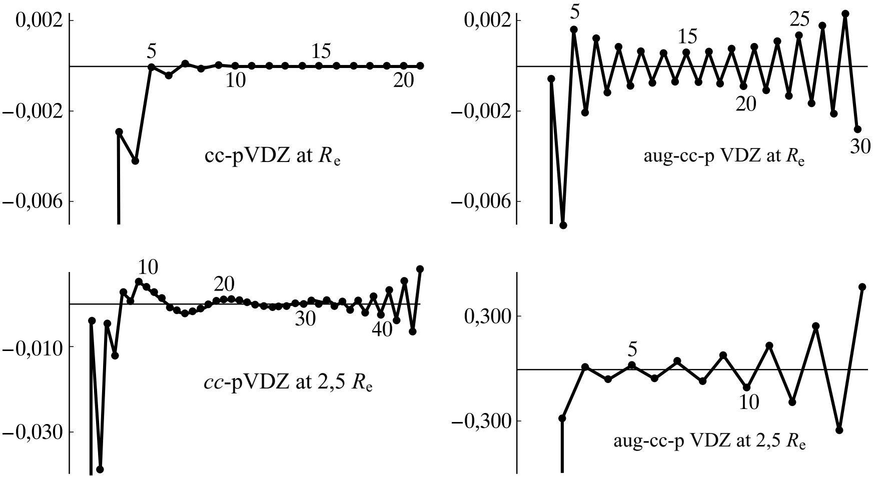


Fig. 10.12. Convergence of the Møller–Plesset perturbation theory (deviation from the exact value, a.u.) for the HF molecule as a function of the basis set used (cc-pVDZ and augmented cc-pVDZ) and assumed bond length, R_e denotes the HF equilibrium distance (T. Helgaker, P. Jørgensen, J. Olsen, "*Molecular Electronic-Structure Theory*", Wiley, Chichester, 2000, p. 780, Fig. 14.6. © 2000, John Wiley and Sons. Reproduced with permission of John Wiley and Sons Ltd.).

(2.5 of the equilibrium distance R_e) the MPn performance is worse. For high orders, the procedure is heading for the catastrophe[99] already described on p. 210. The reason for this is the highly excited and diffuse states used as the expansion functions.[100]

10.17.3 SPECIAL STATUS OF DOUBLE EXCITATIONS

In Møller–Plesset perturbation theory $\Delta E = E_0 - E_0^{(0)} = E_0 - E_{\text{RHF}} - E_0^{(0)} + E_{\text{RHF}} = E_{\text{corel}} + (E_{\text{RHF}} - E_0^{(0)})$. On the other hand $\Delta E = \langle \psi_0^{(0)} | \hat{H}^{(1)} \psi \rangle$. Substituting[101] the operator $\hat{H} - \hat{H}^{(0)}$ instead of $\hat{H}^{(1)}$ gives

$$\begin{aligned}
\Delta E &= \langle \psi_0^{(0)} | (\hat{H} - \hat{H}^{(0)}) \psi_0 \rangle = \langle \psi_0^{(0)} | \hat{H} \psi_0 \rangle - \langle \psi_0^{(0)} | \hat{H}^{(0)} \psi_0 \rangle \\
&= \langle \psi_0^{(0)} | \hat{H} \psi_0 \rangle - \langle \hat{H}^{(0)} \psi_0^{(0)} | \psi_0 \rangle = \langle \psi_0^{(0)} | \hat{H} \psi_0 \rangle - E_0^{(0)} \langle \psi_0^{(0)} | \psi_0 \rangle \\
&= \langle \psi_0^{(0)} | \hat{H} \psi_0 \rangle - E_0^{(0)}.
\end{aligned}$$

The function ψ_0 can be expanded in Slater determinants of various excitation rank (we use intermediate normalization): $\psi_0 = \psi_0^{(0)} +$ excitations. Then, by equalizing the two expressions for ΔE obtained above, we have

---

[99] Except for the smaller basis set and the equilibrium bond length, but the problem has been studied up to $n = 21$.

[100] An analysis of this problem is given in the book cited in the caption to Fig. 10.12, p. 769.

[101] Also taking advantage of the intermediate normalization and the fact that $\psi_0^{(0)}$ is an eigenfunction of $\hat{H}^{(0)}$.

$$E_{\mathrm{corel}} + E_{\mathrm{RHF}} = \langle \psi_0^{(0)} | \hat{H} \psi_0 \rangle = \langle \psi_0^{(0)} | \hat{H} (\psi_0^{(0)} + \text{excitations}) \rangle$$
$$= E_{\mathrm{RHF}} + \langle \psi_0^{(0)} | \hat{H} (\text{excitations}) \rangle,$$

hence

$$E_{\mathrm{corel}} = \langle \psi_0^{(0)} | \hat{H} (\text{excitations}) \rangle. \tag{10.91}$$

The Slater–Condon rules (Appendix M, p. 986) show immediately that the only excitations which give non-zero contributions are the single and double excitations. Moreover, taking advantage of the Brillouin theorem, we obtain single excitation contributions exactly equal to zero. So we get the result that

> the exact correlation energy can be obtained from a formula containing exclusively double excitations.

The problem, however, lies in the fact that these doubly excited determinants are equipped with coefficients obtained in the *full* CI method, i.e. with all possible excitations. How is this? We should draw attention to the fact that, in deriving the formula for ΔE, intermediate normalization is used. If someone gave us the normalized FCI (Full CI) wave functions as a Christmas gift,[102] then the coefficients occurring in the formula for ΔE would not be the double excitation coefficient in the FCI function. We would have to denormalize this function to have the coefficient for the Hartree–Fock determinant equal to 1. We cannot do this without knowledge of the coefficients for higher excitations, cf. Fig. 10.9.

It is as if somebody said: the treasure is hidden in our room, but to find it we have to solve a very difficult problem in the Kingdom of Far Far Away. Imagine a compass which leads us unerringly to that place in our room where the treasure is hidden. Perhaps a functional exists whose minimization would provide us directly with the solution, but we do not know it yet.[103]

Summary

- In the Hartree–Fock method *electrons of opposite spins do not correlate their motion*[104] which is an absurd situation (electrons of the same spins avoid each other – which is reasonable). In many cases (the F_2 molecule, incorrect description of dissociation of chemical bonds, interaction of atoms and non-polar molecules) this leads to wrong results. In this chapter we have learnt about the methods which do take into account a correlation of electronic motions.

VARIATIONAL METHODS USING EXPLICITLY CORRELATED WAVE FUNCTION

- Such methods rely on employing in the variational method a trial function which contains the explicit distance between the electrons. This improves the results significantly, but requires an evaluation of *very* complex integrals.

---

[102] Dreams…

[103] It looks like the work by H. Nakatsuji, *Phys. Rev*. A 14 (1976) 41 and M. Nooijen, *Phys. Rev. Letters* 84 (2000) 2108 go in this direction.

[104] Although they repel each other (mean field) as if they were electron clouds.

- An exact wave function satisfies the correlation cusp condition, $(\frac{\partial\psi}{\partial r})_{r=0} = \mu q_i q_j \psi(r = 0)$, where r is the distance of two particles with charges q_i and q_j, and μ is the reduced mass of the particles. This condition helps to determine the correct form of the wave function ψ. For example, for the two electrons the correct wave function has to satisfy (in a.u.): $(\frac{\partial\psi}{\partial r})_{r=0} = \frac{1}{2}\psi(r = 0)$.
- The family of variational methods with explicitly correlated functions includes: the Hylleraas method, the Hylleraas CI method, the James–Coolidge and the Kołos–Wolniewicz approaches, and the method with exponentially correlated Gaussians. The method of explicitly correlated functions is very successful for 2-, 3- and 4-electron systems. For larger systems, due to the excessive number of complicated integrals, variational calculations are not yet feasible.

VARIATIONAL METHODS WITH SLATER DETERMINANTS

- The CI (Configuration Interaction) approach is *a Ritz method* (Chapter 5) which uses the expansion in terms of *known* Slater determinants. These determinants are constructed from the molecular spinorbitals, usually occupied and virtual ones, produced by the Hartree–Fock method.
- Full CI expansion usually contains an enormous number of terms and is not feasible. Therefore, the CI expansion must be somewhere truncated. Usually we truncate it at a certain maximum rank of excitations with respect to the Hartree–Fock determinant (i.e. the Slater determinants corresponding to single, double, … up to some maximal excitations are included).
- Truncated (limited) CI expansion *is not size consistent*, i.e. the energy of the system of non-interacting objects is not equal to the sum of the energies of the individual objects (calculated separately with the same truncation pattern).
- The MC SCF (Multiconfiguration Self Consistent Field) method is similar to the CI scheme, but we vary not only *the coefficients in front of the Slater determinants, but also the Slater determinants themselves* (changing the analytical form of the orbitals in them). We have learnt about two versions: the classic one (we optimize alternatively coefficients of Slater determinants and the orbitals) and a unitary one (we optimize *simultaneously* the determinantal coefficients and orbitals).
- The CAS SCF (Complete Active Space) method is a special case of the MC SCF approach and relies on *selection* of a set of spinorbitals (usually separated energetically from others) and on construction from them of all possible Slater determinants within the MC SCF scheme. Usually low energy spinorbitals are "*inactive*" during this procedure, i.e. they are doubly occupied in *each* Slater determinant (and are either frozen or allowed to vary). Most important active spinorbitals correspond to HOMO and LUMO.

NON-VARIATIONAL METHODS WITH SLATER DETERMINANTS

- The CC (Coupled-Cluster) method is an attempt to find such an expansion of the wave function in terms of the Slater determinants, which would preserve size consistency. In this method the wave function for the electronic ground state is obtained as a result of the operation of the wave operator $\exp(\hat{T})$ on the Hartree–Fock function (this ensures size consistency). The wave operator $\exp(\hat{T})$ contains the cluster operator $\hat{T}$, which is defined as the sum of the operators for the l-tuple excitations, $\hat{T}_l$ up to a certain maximum $l = l_{\max}$. Each $\hat{T}_l$ operator is the sum of the operators each responsible for *a particular l-tuple excitation* multiplied by its *amplitude* t. The aim of the CC method is to find the t values, since they determine the wave function and energy. The method generates non-linear (with respect to unknown t amplitudes) equations. The CC method usually provides very good results.

- The EOM-CC ("Equation-of-Motion" CC) method is based on the CC wave function obtained for the ground state and is designed to provide the electronic excitation energies and the corresponding excited-state wave functions.
- The MBPT (Many Body Perturbation Theory) method is a perturbation theory in which the unperturbed system is usually described by a single Slater determinant. We obtain two basic equations of the MBPT approach: $\psi_0 = \psi_0^{(0)} + \hat{R}_0(E_0^{(0)} - E_0 + \hat{H}^{(1)})\psi_0$ and $E_0 = E_0^{(0)} + \langle\psi_0^{(0)}|\hat{H}^{(1)}\psi_0\rangle$, where $\psi_0^{(0)}$ is usually the Hartree–Fock function, $E_0^{(0)}$ the sum of the orbital energies, $\hat{H}^{(1)} = \hat{H} - \hat{H}^{(0)}$ is the fluctuation potential, and $\hat{R}_0$ the reduced resolvent (i.e. "almost" inverse of the operator $E_0^{(0)} - \hat{H}^{(0)}$). These equations are solved in an iterative manner. Depending on the iterative procedure chosen, we obtain either the Brillouin–Wigner or the Rayleigh–Schrödinger perturbation theory. The latter is applied in the Møller–Plesset (MP) method. *One of the basic computational methods for the correlation energy is the MP2 method*, which gives the result correct through the second order of the Rayleigh–Schrödinger perturbation theory (with respect to the energy).

Main concepts, new terms

correlation energy (p. 499)
explicit correlation (p. 502)
cusp condition (p. 503)
Hylleraas function (p. 506)
harmonic helium atom (p. 507)
James–Coolidge function (p. 508)
Kołos–Wolniewicz function (p. 508)
geminal (p. 513)
exponentially correlated function (p. 513)
Coulomb hole (p. 513)
exchange hole (p. 516)
Valence bond (VB) method (p. 520)
covalent structure (p. 521)
resonance theory (p. 520)
Heitler–London function (p. 521)
ionic structure (p. 521)
Brueckner function (p. 525)
configuration mixing (p. 525)
configuration interaction (p. 526)
configuration (p. 526)
Brillouin theorem (p. 527)
density matrix (p. 531)
natural orbitals (p. 531)
full CI method (p. 531)
direct CI method (p. 533)
size consistency (p. 532)
multireference methods (p. 533)
active space (p. 535)
frozen orbitals (p. 534)
multiconfigurational SCF methods (p. 535)
unitary MC SCF method (p. 536)
commutator expansion (p. 537)
cluster operator (p. 540)
wave operator (p. 540)
CC amplitudes (p. 542)
EOM-CC method (p. 548)
deexcitations (p. 550)
many body perturbation theory (MBPT) (p. 551)
reduced resolvent (p. 554)
Brillouin–Wigner perturbation theory (p. 556)
Rayleigh–Schrödinger perturbation theory (p. 557)
Møller–Plesset perturbation theory (p. 558)

From the research front

The computational cost in the Hartree–Fock method scales with the size N of the atomic orbital basis set as N^4 and, while using devices similar to *direct* CI, even[105] as N^3. How-

---

[105] This reduction is caused mainly by a preselection of the two-electron integrals. The preselection allows us to estimate the value of the integral without its computation and to reject the large number of integrals of values close to zero.

ever, after doing the Hartree–Fock computations for small (say, up to 10 electrons) systems, we perform more and more frequently calculations of the electronic correlation. The main approaches used to this end are: the MP2 method, the CC method with single and double excitations in $\hat{T}$ and partial inclusion of triple ones (the so called CCSD(T) approach). The-state-of-the art in CC theory currently includes the full CCSDTQP model, which incorporates into the cluster expansion all the operators through pentuple excitations.[106] The computational cost of the CCSD scheme scales as N^6. The computational strategy often adopted relies on obtaining the optimum geometry of the system with a less sophisticated method (e.g., Hartree–Fock) and, subsequently, calculating the wave function for that geometry with a more sophisticated approach (e.g., the MP2 that scales as N^5, MP4 or CCSD(T) scaling as N^7). In the next chapter we will learn about the density functional theory (DFT) which joins the above mentioned methods and is used for large systems.

Ad futurum…

Experimental chemistry is focused, in most cases, on molecules of *larger* size than those for which fair calculations with correlation are possible. However, after thorough analysis of the situation, it turns out that the cost of the calculations does not necessarily increase very fast with the size of a molecule. Employing localized molecular orbitals and using the multipole expansion of the integrals involving the orbital separated in space causes, for elongated molecules, the cost of the post-Hartree–Fock calculations to scale linearly with the size of a molecule.[107] It can be expected that, if the methods described in this Chapter are to survive in practical applications, such a step has to be made.

There is one more problem which will probably be faced by quantum chemistry when moving to larger molecules containing heteroatoms. Nearly all the methods including electron correlation described so far (with the exception of the explicitly correlated functions) are based on the silent and pretty "obvious" assumption, that the higher the excitation we consider the higher the configuration energy we get. This assumption seems to be satisfied so far, but the molecules considered were always small, and the method has usually been limited to a small number of excited electrons. This assumption can be challenged in certain cases. The multiple excitations in large molecules containing easily polarizable fragments can result in electron transfers which cause energetically favourable strong electrostatic interactions ("mnemonic effect"[108]) which lower the energy of the configuration. The reduction can be large enough to make the energy of the *formally* multiply excited determinant close to that of the Hartree–Fock determinant. Therefore, it should be taken into account on the same footing as the Hartree–Fock. This is rather unfeasible for the methods discussed above.

The explicitly correlated functions have a built-in adjustable and efficient basic mechanism accounting for the correlation within the interacting electronic pair. The mechanism is based on the obvious thing: the electrons should avoid each other.[109]

Let us imagine the CH_4 molecule. Let us look at it from the viewpoint of localized orbitals. With the method of explicitly correlated geminal functions for bonds we would succeed in making the electrons avoid each other within the same bond. And what should

---

[106] M. Musiał, S.A. Kucharski, R.J. Bartlett, *J. Chem. Phys.* 116 (2002) 4382.

[107] H.-J. Werner, *J. Chem. Phys.* 104 (1996) 6286.

[108] L.Z. Stolarczyk, L. Piela, *Chem. Phys. Letters* 85 (1984) 451, see also A. Jagielska, L. Piela, *J. Chem. Phys.* 112 (2000) 2579.

[109] In special conditions one electron can follow the other together forming a Cooper pair. The Cooper pairs are responsible for the mechanism of superconductivity. This will be a fascinating field of research for chemist-engineered materials in the future.

happen if the centre of gravity of the electron pair of one of the bonds shifts towards the carbon atom? The centres of gravity of the electron pairs of the remaining three bonds should move away along the CH bonds. The wave function must be designed in such a way that it accounts for this. In current theories, this effect is either deeply hidden or entirely neglected. A similar effect may happen in a polymer chain. One of the natural correlations of electronic motions should be a shift of electron pairs of all bonds in the same phase. As a highly many-electron effect the latter is neglected in current theories. However, the purely correlational Axilrod–Teller effect in the case of linear configuration, discussed in Chapter 13 (three-body dispersion interaction in the third order of perturbation theory), suggests clearly that the correlated motion of many electrons should occur.

It seems that the explicitly correlated functions, in spite of serious problems at the integral level, can be generalized in future towards the collective motions of electrons, perhaps on the basis of the renormalization theory of Kenneth Wilson (introduced into chemistry for the first time by Martin Head-Gordon).[110]

Kenneth Geddes Wilson (born 1936), American theoretical physicist. The authorities of Cornell University worried by Wilson's low number of published papers. Pressed by his supervisors, he finally started to publish, and won in 1982 the Nobel prize for the renormalization theory. It is a theory of the mathematical transformations describing a system viewed at various scales (with variable resolution). The renormalization theory, as applied by Head-Gordon to the hydrocarbon molecule, first uses the LCAO (the usual atomic orbitals), then, in subsequent approximations, some linear combinations of functions that are more and more diffused in space.

Additional literature

A. Szabo, N.S. Ostlund, "Modern Quantum Chemistry", McGraw-Hill, New York, 1989, p. 231–378.

Excellent book.

T. Helgaker, P. Jørgensen, J. Olsen, "Molecular Electronic-Structure Theory", Wiley, Chichester, 2000, p. 514.

Practical information on the various methods accounting for electron correlation presented in a clear and competent manner.

Questions

1. The Hartree–Fock method for the helium atom in its ground state. If electron 1 resides on the one side of the nucleus then electron 2 can be found most likely:
   a) on the other side of the nucleus; b) at the nucleus; c) on the same side of a nucleus; d) at infinite distance from the nucleus.

2. The Gaussian geminal for the helium atom $\psi(\boldsymbol{r}_1, \boldsymbol{r}_2) = N(1 + \kappa r_{12}) \exp[-\frac{1}{4}(r_1^2 + r_2^2)]$, N is the normalization constant:
   a) to satisfy the cusp condition should have $\kappa = \frac{1}{2}$;
   b) represents the exact wave function for $\kappa = \frac{1}{2}$;
   c) for $\kappa < 0$ takes care of electron repulsion;
   d) to satisfy the cusp condition has to have $\exp[-\frac{1}{2}r_{12}^2]$ instead of $(1 + \kappa r_{12})$.

---

[110] M. Head-Gordon, "*Proc. 5th Intern. Conf. Computers in Chemistry*", Szklarska Poręba, Poland, 1999, p. L33.

3. The wave function for the H_2^- molecule [positions of nuclei a and b: $(0,0,0)$ and $(R,0,0)$, respectively] in the form of a single Slater determinant, built of three spinorbitals $\phi_1(\boldsymbol{r},\sigma)=\varphi_1(\boldsymbol{r})\alpha(\sigma)$, $\phi_2=\varphi_1(\boldsymbol{r})\beta(\sigma)$, $\phi_3=\varphi_2(\boldsymbol{r})\alpha(\sigma)$ (φ_1 is the doubly occupied bonding, and φ_2 is the singly occupied antibonding one). If $\boldsymbol{r}_1=(\frac{R}{2},0,0)$, $\sigma_1=\frac{1}{2}, \boldsymbol{r}_2=(0,0,0)$, $\sigma_2=-\frac{1}{2}$, $\sigma_3=\frac{1}{2}$ then the probability density of finding electron 3 is:
   a) almost zero on nucleus a; b) almost zero on nucleus b; c) equal to 0 everywhere; d) proportional to $|\varphi_2|^2$.

4. A Hartree–Fock function:
   a) correlates the positions of all electrons;
   b) correlates the positions of electrons with the same spin coordinates;
   c) correlates the positions of electrons with opposite spin coordinates;
   d) does not correlate the positions of electrons, since in the Hartree–Fock method electron correlation is not accounted for.

5. The Brillouin theorem says that ($\hat{H}$ is the Hamiltonian, Φ_0 is the Hartree–Fock function, Φ_1 is a singly and Φ_2 a doubly excited Slater determinant):
   a) $\langle\Phi_0|\hat{H}\Phi_1\rangle=0$ if all the spinorbitals are orthogonal; b) $\langle\Phi_1|\hat{H}\Phi_1\rangle=0$; c) $\langle\Phi_2|\hat{H}\Phi_1\rangle=0$; d) $\langle\Phi_0|\Phi_1\rangle=0$.

6. In the Coupled Cluster method ($\hat{T}$ is the cluster operator, Φ_0 is the Hartree–Fock wave function) the wave function:
   a) is $\psi=\exp(i\hat{T})\Phi_0$; b) does not vanish in infinity; c) contains only single and double excitations; d) is $\psi=\exp(\hat{T})\Phi_0$ and ensures size consistency.

7. MBPT: If the projector $\hat{P}=|\psi_0^{(0)}\rangle\langle\psi_0^{(0)}|$ and $\hat{Q}=\sum_{n=1}^{\infty}|\psi_n^{(0)}\rangle\langle\psi_n^{(0)}|$ ($\psi_n^{(0)}$ form the complete orthonormal set) then:
   a) $\hat{P}\hat{Q}=1$; b) $(\hat{P}+\hat{Q})^2=1$; c) $[\hat{P},\hat{Q}]=i\hbar$; d) $\hat{Q}=\exp(\hat{P})$.

8. The Møller–Plesset method (MP2) is:
   a) a variational method with two variational parameters; b) a perturbation theory with unperturbed wave function in the form of a Gaussian geminal; c) a perturbation theory with the energy computed through the second order; d) a Ritz method limited to double excitations.

9. To calculate the exact correlation energy:
   a) it is enough to have the expansion in singly excited Slater determinants;
   b) it is enough to know the Hartree–Fock function;
   c) we must use explicitly correlated functions;
   d) it is enough to have a certain wave function containing double excitations only.

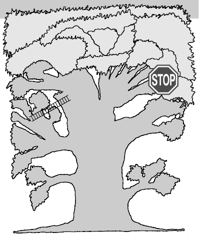
10. We have the following order of mean values of the Hamiltonian calculated for the functions: *I*: $\psi_1=$ the Hartree–Fock function, *II*: $\psi_2=$ the Hartree–Fock function + doubly excited Slater determinant, *III*: $\psi_3=$ the Hartree–Fock function $+\lambda\cdot$ doubly excited Slater determinant (the same as in ψ_2), where λ is an optimal variational coefficient:
    a) $I>III>II$; b) $I>III$ and $II>III$; c) $III>II>I$; d) $I>II>III$.

Answers

1b, 2a, 3d, 4b, 5a, 6d, 7b, 8c, 9d, 10b

Chapter 11

ELECTRONIC MOTION: DENSITY FUNCTIONAL THEORY (DFT)



WHERE ARE WE?

Where are we?

We are on an upper right-hand side branch of the TREE.

An example

A metal represents a system that is very difficult to describe using the quantum chemistry methods given so far. The Restricted Hartree–Fock method here offers a very bad, if not pathological, approximation (cf. Chapter 8, p. 371), because the HOMO-LUMO gap is equal to zero in metals. The methods based on the Slater determinants (CI, MC SCF, CC, etc., Chapter 10) are ruled out as involving a giant number of excited configurations to be taken into account, because of the continuum of the occupied and virtual energy levels (see Chapter 9). Meanwhile, in the past some properties of metals could be obtained, from simple theories that assumed that the electrons in a metal behave similarly to a homogeneous electron gas (also known as jellium), and the nuclear charge (to make the whole system neutral) has been treated as smeared out uniformly in the metal volume. There has to be something physically important captured in such theories.

electron gas

What is it all about

The preceding chapter has shown how difficult it is to calculate the correlation energy. Basically there are two approaches: either to follow configuration interaction type methods (CI, MC SCF, CC, etc.), or to go in the direction of explicitly correlated functions. The first means a barrier of more and more numerous excited configurations to be taken into account, the second, very tedious and time-consuming integrals. In both cases we know the Hamiltonian and fight for a satisfactory wave function (often using the variational principle, Chapter 5). It turns out that there is also a third direction (presented in this chapter) that does not regard configurations (except a single special one) and does not have the bottleneck of difficult integrals. Instead, we have the kind of wave function in the form of a single Slater determinant, but we have a serious problem in defining the proper Hamiltonian.

The ultimate goal of the DFT method is the calculation of the total energy of the system and the ground-state electron density distribution without using the wave function of the system.

Why is this important?

The DFT calculations (despite taking electronic correlation into account) are not expensive, their cost is comparable with that of the Hartree–Fock method. Therefore, the same computer power allows us to explore much larger molecules than with other post-Hartree–Fock (correlation) methods.

What is needed?

- The Hartree–Fock method (Chapter 8, necessary).
- The perturbational method (Chapter 5, advised).
- Lagrange multipliers (Appendix N, p. 997, advised).

Classic works

The idea of treating electrons in metal as an electron gas was conceived in 1900, independently by Lord Kelvin[1] and by Paul Drude.[2] ★ The concept explained the electrical conductivity of metals, and was then used by Llewellyn Hilleth Thomas in "*The Calculation of Atomic Fields*" published in *Proceedings of the Cambridge Philosophical Society*, 23 (1926)

---

[1] Or William Thomson (1824–1907), British physicist and mathematician, professor at the University of Glasgow. His main contributions are in thermodynamics (the second law, internal energy), theory of electric oscillations, theory of potentials, elasticity, hydrodynamics, etc. His great achievements were honoured by the title of Lord Kelvin (1892).

[2] Paul Drude (1863–1906), German physicist, professor at the universities in Leipzig, Giessen and Berlin.

542 as well as by Enrico Fermi in "*A Statistical Method for the Determination of Some Atomic Properties and the Application of this Method to the Theory of the Periodic System of Elements*" in *Zeitschrift für Physik*, 48 (1928) 73. They (independently) calculated the electronic kinetic energy per unit volume (this is therefore the *kinetic energy density*) as a function of the local electron density ρ. ★ In 1930 Paul Adrien Maurice Dirac presented a similar result in "*Note on the Exchange Phenomena in the Thomas Atom*", *Proceedings of the Cambridge Philosophical Society*, 26 (1930) 376 for the *exchange energy* as a function of ρ. ★ In a classic paper "*A Simplification of the Hartree–Fock Method*" published in *Physical Review*, 81 (1951) 385, John Slater showed that the Hartree–Fock method applied to metals gives the exchange energy density proportional to $\rho^{\frac{1}{3}}$. ★ For classical positions specialists often use a book by Pál Gombas "*Die statistische Theorie des Atoms und ihre Anwendungen*", Springer Verlag, Wien, 1948. ★ The contemporary theory was born in 1964–1965, when two fundamental works appeared: Pierre Hohenberg and Walter Kohn in *Physical Review*, 136 (1964) B864 entitled "*Inhomogeneous Electron Gas*" and Walter Kohn and Lu J. Sham in *Physical Review*, A140 (1965) 1133 under the title "*Self-Consistent Equations including Exchange and Correlation Effects*". ★ Mel Levy in "*Electron Densities in Search of Hamiltonians*" published in *Physical Review*, A26 (1982) 1200 proved that the variational principle in quantum chemistry can be equivalently presented as a minimization of the Hohenberg–Kohn functional that depends on the electron density ρ. ★ Richard F.W. Bader in 1994 wrote a book on mathematical analysis of the electronic density "*Atoms in Molecules. A Quantum Theory*", Clarendon Press, Oxford, that enabled chemists to look at molecules in a synthetic way, independently of the level of theory that has been used to describe it.

11.1 ELECTRONIC DENSITY – THE SUPERSTAR

In the DFT method we begin from the Born–Oppenheimer approximation, that allows us to obtain the electronic wave function corresponding to fixed positions of the nuclei. We will be interested in the ground-state of the system.

As it will turn out later on,

> to describe this state instead of the N electron wave function $\Psi(1,2,\dots,N)$, we need only the *electron density distribution* defined as:
>
> $$\rho(\boldsymbol{r}) = N \sum_{\sigma_1=-\frac{1}{2},\frac{1}{2}} \int \mathrm{d}\tau_2\,\mathrm{d}\tau_3 \dots \mathrm{d}\tau_N \left|\Psi(\boldsymbol{r},\sigma_1,\boldsymbol{r}_2,\sigma_2,\dots,\boldsymbol{r}_N,\sigma_N)\right|^2. \tag{11.1}$$

It is seen that we obtain ρ by carrying out the integration of $|\Psi|^2$ over the coordinates (*space and spin*) of all the electrons except one (in our case electron 1 with coordinates $\boldsymbol{r}, \sigma_1$) and in addition the summation over its spin coordinate (σ_1). Thus we obtain a function of the position of electron 1 in space: $\rho(\boldsymbol{r})$. The wave function Ψ is antisymmetric with respect to the exchange of the coordinates of any two electrons, and, therefore, $|\Psi|^2$ is symmetric with respect to such an exchange.

Hence, the definition of ρ is independent of the label of the electron we do not integrate over. According to this definition, ρ represents nothing else but the density of the electron cloud carrying N electrons, hence (integration over the whole 3D space):

$$\int \rho(\boldsymbol{r})\,\mathrm{d}^3\boldsymbol{r} = N. \tag{11.2}$$

Therefore the electron density distribution $\rho(\boldsymbol{r})$ is given for a point $\boldsymbol{r}$ in the units: the number of electrons per volume unit (e.g., 0.37 Å^{-3}). Since $\rho(\boldsymbol{r})$ represents an integral of a non-negative integrand, $\rho(\boldsymbol{r})$ is always non-negative. Let us check that ρ may also be defined as the mean value of the density operator $\hat{\rho}(\boldsymbol{r})$, or sum of the Dirac delta operators (cf. Appendix E on p. 951) for individual electrons at position $\boldsymbol{r}$:

$$\rho(\boldsymbol{r}) = \langle\Psi|\hat{\rho}\Psi\rangle \equiv \left\langle\Psi\middle|\left(\sum_{i=1}^{N}\delta(\boldsymbol{r}_i-\boldsymbol{r})\right)\Psi\right\rangle = \sum_{i=1}^{N}\big\langle\Psi\big|\big(\delta(\boldsymbol{r}_i-\boldsymbol{r})\big)\Psi\big\rangle. \tag{11.3}$$

Indeed, each of the integrals in the summation is equal to $\rho(\boldsymbol{r})/N$, the summation over i gives N, therefore, we obtain $\rho(\boldsymbol{r})$.

If the function Ψ is taken as a normalized Slater determinant built of N spinorbitals ϕ_i, from the I rule of Slater–Condon (Appendix M) for $\langle\Psi|(\sum_{i=1}^{N}\delta(\boldsymbol{r}_i-\boldsymbol{r}))\Psi\rangle$ we obtain (after renaming the electron coordinates in the integrals, the integration is over the spatial and spin coordinates of electron 1)[3]

$$\begin{aligned}\rho(\boldsymbol{r}) &= \big\langle\phi_1(1)\big|\delta(\boldsymbol{r}_1-\boldsymbol{r})\phi_1(1)\big\rangle_1 + \big\langle\phi_2(1)\big|\delta(\boldsymbol{r}_1-\boldsymbol{r})\phi_2(1)\big\rangle_1 + \cdots\\ &\quad + \big\langle\phi_N(1)\big|\delta(\boldsymbol{r}_1-\boldsymbol{r})\phi_N(1)\big\rangle_1\\ &= \sum_{i=1}^{N}\sum_{\sigma_1=-\frac{1}{2},+\frac{1}{2}}\big|\phi_i(\boldsymbol{r},\sigma_1)\big|^2.\end{aligned} \tag{11.4}$$

If we assume the double occupancy of the molecular orbitals, we have

$$\begin{aligned}\rho(\boldsymbol{r}) &= \sum_{i=1}^{N}\sum_{\sigma_1}\big|\phi_i(\boldsymbol{r},\sigma_1)\big|^2 = \sum_{i=1}^{N/2}\sum_{\sigma_1}\big|\varphi_i(\boldsymbol{r})\alpha(\sigma_1)\big|^2 + \sum_{i=1}^{N/2}\sum_{\sigma_1}\big|\varphi_i(\boldsymbol{r})\beta(\sigma_1)\big|^2\\ &= \sum_{i=1}^{N/2}2\big|\varphi_i(\boldsymbol{r})\big|^2,\end{aligned}$$

where φ_i stand for the molecular orbitals. We see that admitting the open shells we have

$$\rho(\boldsymbol{r}) = \sum_i n_i\big|\varphi_i(\boldsymbol{r})\big|^2 \tag{11.5}$$

with $n_i = 0, 1, 2$ denoting orbital occupancy in the Slater determinant.

---

[3] This expression is invariant with respect to any unitary transformation of the molecular orbitals, cf. Chapter 8.

11.2 BADER ANALYSIS

11.2.1 OVERALL SHAPE OF ρ

Imagine an electron cloud with a charge distribution[4] that carries the charge of N electrons. Unlike a storm cloud, the electron cloud does not change in time (stationary state), but has density $\rho(\boldsymbol{r})$ that changes in space (similar to the storm cloud). Inside the cloud the nuclei are located. The function $\rho(\boldsymbol{r})$ exhibits non-analytical behaviour (discontinuity of its gradient) at the positions of the nuclei, which results from the poles ($-\infty$) of the potential energy at these positions. Recall the shape of the 1s wave function for the hydrogen-like atom (see Fig. 4.17), it has a spike at $r = 0$. In Chapter 10 it was shown that the correct electronic wave function has to satisfy the cusp condition in the neighbourhood of each of the nuclei, where ρ changes as $\exp(-2Zr)$. This condition results in spikes of $\rho(\boldsymbol{r})$ exactly at the positions of the nuclei, Fig. 11.1.a. How sharp such a spike is[5] depends on the charge of the nucleus Z: an infinitesimal deviation from the position of the nucleus (p. 505) has to be accompanied by such a decreasing of the density[6] that $\frac{\partial\rho}{\partial r}/\rho = -2Z$.

Thus, because of the Coulombic interactions, the electrons will concentrate close to the nuclei, and therefore we will have maxima of ρ right on them. It is evident also, that at long distances from the nuclei the density ρ will decay to practically zero. Further details will be of great interest, e.g., are there any concentrations of ρ in the regions *between* nuclei? If yes, will it happen for every pair of nuclei or for some pairs only? This is of obvious importance for chemistry, which deals with the concept of chemical bonds and a model of the molecule as the nuclei kept together by a chemical bond pattern.

11.2.2 CRITICAL POINTS

For analysis of any function, including the electronic density as a function of the position in space, the *critical (or stationary) points* are defined as those that have

critical points

---

[4] Similar to a storm cloud in the sky.

[5] If non-zero size nuclei were considered, the cusps would be rounded (within the size of the nuclei), the discontinuity of the gradient would be removed and regular maxima would be observed.

[6] It has been shown (P.D. Walker, P.G. Mezey, *J. Am. Chem. Soc.* 116 (1994) 12022) that despite the non-analytical character of ρ (because of the spikes) the function ρ has the following remarkable property: *if we know* ρ even in the smallest volume, this determines ρ in the whole space. A by-product of this theorem is of interest for chemists. Namely, this means that a functional group in two different molecules or in two conformations of the same molecule cannot have an identical ρ characteristic for it. If it had, from ρ in its neighbourhood we would be able to reproduce the whole density distribution $\rho(\boldsymbol{r})$, but for which of the molecules or conformers? Therefore, by *reductio ad absurdum* we have the result: it is impossible to define (with all details) the notion of a functional group in chemistry. This is analogous to the conclusion drawn in Chapter 8 about the impossibility of a rigorous definition of a chemical bond (p. 397). This also shows that chemistry and physics (relying on mathematical approaches) profit very much, and further, are heavily based on, some ideas that mathematics destroys in a second. Nevertheless, without these ideas natural sciences would lose their generality, efficiency and …beauty.

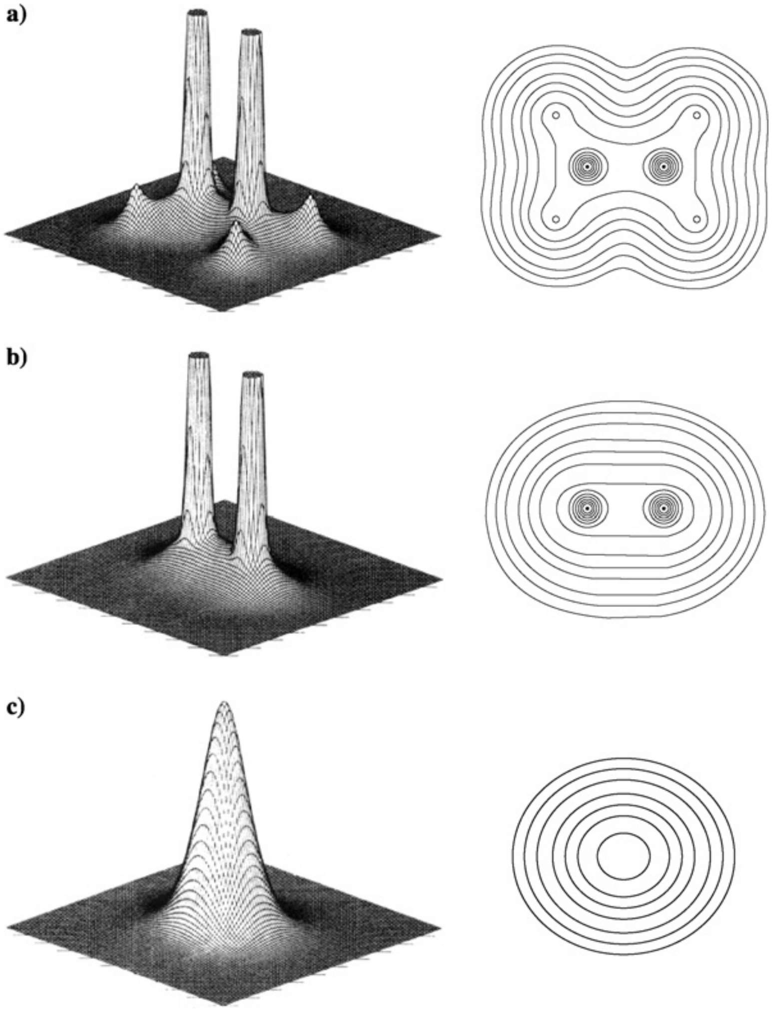


Fig. 11.1. Electron density ρ for the planar ethylene molecule shown in three cross sections. $\int \rho(\boldsymbol{r})\,\mathrm{d}^3\boldsymbol{r} = 16$, the number of electrons in the molecule. Fig. (a) shows the cross section *within the molecular plane*. The positions of the nuclei can be easily recognized by the "spikes" of ρ (obviously much more pronounced for the carbon atoms than for the hydrogens atoms), their charges can be computed from the slope of ρ. Fig. (b) shows the cross section *along the CC bond perpendicular to the molecular plane*, therefore, only the maxima at the positions of the carbon nuclei are visible. Fig. (c) is the cross section *perpendicular to the molecular plane and intersecting the CC bond* (through its centre). It is seen that ρ decays monotonically with the distance from the bond centre. Most interesting, however, is that the cross section resembles an ellipse rather than a circle. Note that we do not see any separate σ or π densities. This is what the concept of π bond is all about, just to reflect the bond cross section ellipticity. R.F.W. Bader, T.T. Nguyen-Dang, Y. Tal, *Rep. Progr. Phys.* 44 (1981) 893, courtesy of Institute of Physics Publishing, Bristol, UK.

vanishing gradient

$$\nabla \rho = \mathbf{0}.$$

These are maxima, minima and saddle points. If we start from an arbitrary point and follow the direction of $\nabla \rho$, we end up at a maximum of ρ. The compact set of starting points which converge in this way to the same maximum is called the *basin of attraction of this maximum*, and the position of the maximum is known as *attractor*. The position may correspond to any of the nuclei or to a non-nuclear electronic distribution (*non-nuclear attractors*,[7] Fig. 11.2.a). The largest maxima correspond to the positions of the nuclei. Formally, positions of the nuclei are not the stationary points, because $\nabla \rho$ has a discontinuity here connected to the cusp condition (see Chapter 10, p. 504). A basin has its neighbour-basins and the border between the basins (a surface) satisfies $\nabla \rho \cdot \boldsymbol{n} = 0$, where $\boldsymbol{n}$ is a unit vector perpendicular to the surface (Fig. 11.2.b,c).

non-nuclear attractors

In order to tell whether a particular critical point represents a maximum (non-nuclear attractor), a minimum or a saddle point we have to calculate at this point the Hessian, i.e. the matrix of the second derivatives: $\{\frac{\partial^2 \rho}{\partial \xi_i \partial \xi_j}\}$, where $\xi_1 = x$, $\xi_2 = y$, $\xi_3 = z$. Now, the stationary point is used as the origin of a local Cartesian coordinate system, which will be rotated in such a way as to obtain the Hessian matrix (computed in the rotated coordinate system) diagonal. This means that the rotation has been performed in such a way that the axes of the new local coordinate system are collinear with the principal axes of a quadratic function that approximates ρ in the neighbourhood of the stationary point (this rotation is achieved simply by diagonalization of the Hessian $\{\frac{\partial^2 \rho}{\partial \xi_i \partial \xi_j}\}$, cf. Appendix K). The diagonalization gives three eigenvalues. We have the following possibilities:

- *All three eigenvalues are negative – we have a maximum of* ρ (non-nuclear attractor, Fig. 11.2.a).
- *All three eigenvalues are positive – we have a minimum of* ρ. The minimum appears when we have a cavity, e.g., in the centre of fullerene. When we leave this point, independently of the direction of this motion, the electron density increases.
- *Two eigenvalues are positive, while one is negative – we have a first-order saddle point of* ρ. The centre of the benzene ring may serve as an example (Fig. 11.2.d). If we leave this point in the molecular plane in any of the two independent directions, ρ increases (thus, a minimum of ρ within the plane, the two eigenvalues positive), but when leaving perpendicularly to the plane the electronic density decreases (thus a maximum of ρ along the axis, the negative eigenvalue).

cavity

ring

---

[7] For example, imagine a few dipoles with their positive poles oriented towards a point in space. If the dipole moments exceed some value, it may turn out that around this point there will be a concentration of electron density having a maximum there. This is what happens in certain dipoles, in which an electron is far away from the nuclear framework (sometimes as far as 50 Å) and keeps following the positive pole of the dipole ("a dipole-bound electron") when the dipole rotates in space, see, e.g., J. Smets, D.M.A. Smith, Y. Elkadi, L. Adamowicz, *Pol. J. Chem.* 72 (1998) 1615.

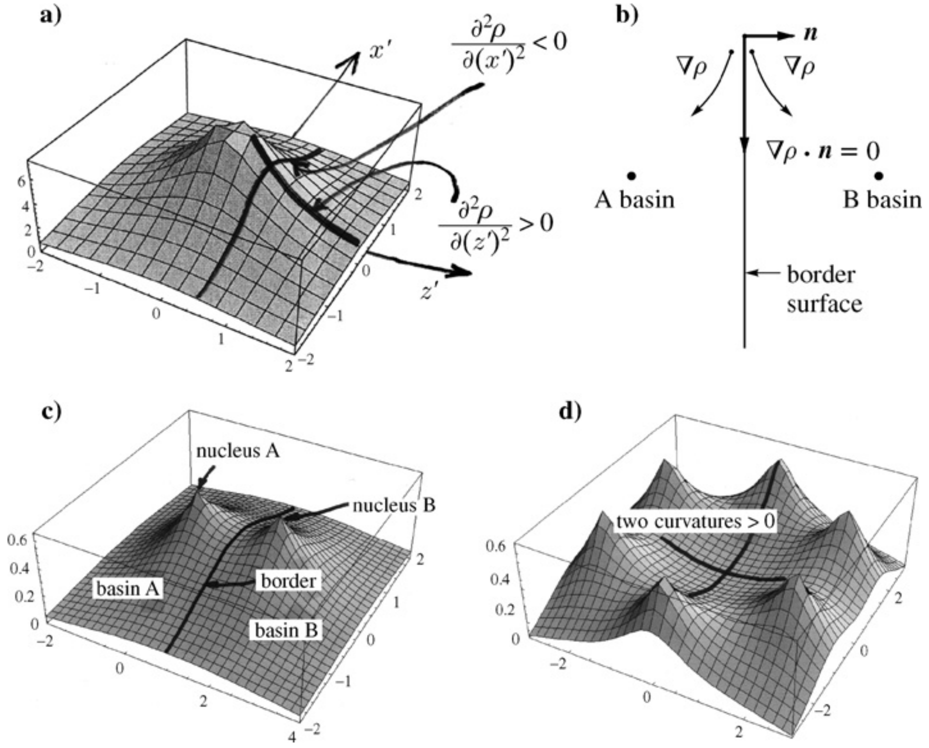


Fig. 11.2. How does the electronic density change when we leave a critical point? Fig. (a) illustrates a non-nuclear attractor (maximum of ρ, no cusp). Note that we can tell the signs of some second derivatives (curvatures) computed at the intersection of black lines (slope), the radial curvature $\frac{\partial^2\rho}{\partial(z')^2}$ is positive, while the two lateral ones (only one of them: $\frac{\partial^2\rho}{\partial(x')^2}$ is shown) are negative. If for the function shown the curvatures were computed at the maximum, all three curvatures would be negative, Fig. (b) shows the idea of the border surface separating two basins of ρ corresponding to two nuclei: A and B. Right at the border between the two basins the force lines of $\nabla\rho$ diverge: if you make a step left from the border, you end up in the basin of nucleus A, if you make a step right, you get into the basin of B. Just at the border you have to have $\nabla\rho \cdot \boldsymbol{n} = 0$, because the two vectors: $\nabla\rho$ and $\boldsymbol{n}$ are perpendicular. Fig. (c): The same showing additionally the density function for chemical bond AB. The border surface is shown as a black line. Two of three curvatures are negative (one of them shown), the third one (along AB line) is positive. Fig. (d) illustrates the electronic density distribution in benzene. In the middle of the ring two curvatures are positive (shown), the third curvature is negative (not shown). If the curvatures were computed in the centre of the fullerene (not shown), all three curvatures would be positive (because the electron density increases when going out of the centre).

chemical bond

- *One eigenvalue is positive, while two are negative – we have a second-order saddle point of ρ.* It is a very important case, because this is what happens at any covalent chemical bond (Figs. 11.1, 11.2.c). In the region between *some*[8] nuclei of a polyatomic molecule we may have such a critical point. When we go perpendicu-

---

[8] Only *some* pairs of atoms correspond to chemical bonds.

larly to the bond in any of the two possible directions, ρ decreases (a maximum within the plane, two eigenvalues negative), while going towards any of the two nuclei ρ increases (to achieve maxima at the nuclei; a minimum along one direction, i.e. one eigenvalue positive). The critical point needs not be located along the straight line going through the nuclei ("banana" bonds are possible), also its location may be closer to one of the nuclei (polarization). Thus the nuclei are connected by a kind of electronic density "rope" (most dense at its core and decaying outside) extending from one nucleus to the other along a curved line, having a single critical point on it, its cross section for some bonds circular, for others elliptic-like.[9] Calculations have shown that when the two nuclei separate, the rope elongates and *suddenly*, at a certain internuclear distance it breaks down (this corresponds to zeroing one of the eigenvalues). The set of parameters (like the internuclear distance) at which $\det\{\frac{\partial^2\rho}{\partial\xi_i\partial\xi_j}\} = 0$ (corresponding to an eigenvalue equal to 0) is called the catastrophe set. Thus the catastrophe theory of René Thom turns out to be instrumental in chemistry.

Richard Bader (born 1931), Canadian chemist, professor emeritus at McMaster University in Canada. After his PhD at the Massachusetts Institute of Technology he won an international fellowship to study at Cambridge University in UK under Christopher Longuet-Higgins. At their first meeting Bader was given the titles of two books together with: "*When you have read these books, maybe we can talk again*". From these books Bader found out about theories of electron density. From that time on he became convinced that electron density was the quantity of prime importance for the theory. Photo reproduced thanks to courtesy of Richard Bader.

René Thom (1923–2002), French mathematician, professor at the Université de Strasbourg and founder of catastrophe theory (1966). The theory analyzes abrupt changes of functions (change of the number and character of stationary points) upon changing some parameters. In 1958 René Thom received the Fields Medal, the highest distinction for a mathematician.

11.2.3 LAPLACIAN OF THE ELECTRONIC DENSITY AS A "MAGNIFYING GLASS"

Now we will focus on the Hessian of ρ *beyond* the critical points.

Fig. 11.3.a shows a decreasing function $f(x)$, i.e. $f' \equiv \frac{df}{dx} < 0$, with a single well developed maximum at $x = 0$ and a small hump close to x_2. The function somewhat resembles the electron density decay, say, for the neon atom, when we go out of the nucleus. Note, that the function $-f'' \equiv -\frac{d^2f}{dx^2}$ exhibits an easily visible

---

[9] All the details may be computed nowadays by using quantum mechanical methods, often most demanding ones (with the electronic correlation included). Contemporary crystallography is able to measure the same quantities in some fine X-ray experiments. Therefore, the physicochemical methods are able to indicate precisely which atoms are involved in a chemical bond: is it strong or not, is it straight or curved ("rope-like" banana bond), what is the thickness of the "rope", has it a cylindrical or oval cross-section (connected to its σ or π character), etc. A good review is available: T.S. Koritsanszky, P. Coppens, *Chem. Rev.* 101 (2001) 1583.

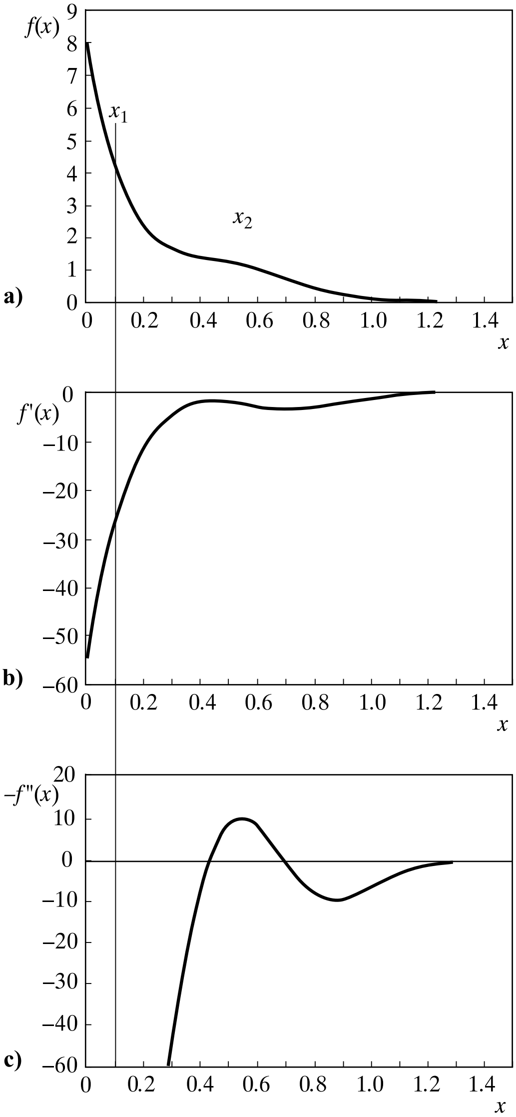


Fig. 11.3. The principle of the Bader analysis of electronic density distribution. The quantity $-f'' = -\frac{d^2f}{dx^2}$ works as a "magnifying glass" for almost invisible humps of the function $f(x)$. Fig. (a) shows a decreasing function $f(x)$, that is similar to the dependence of the electron density of an atom as a function of the distance from the nucleus, with a trace of a hump near x_2. Figs. (b,c) show the derivatives f' and $-f''$, respectively. The almost invisible hump on f turns out to be very visible when the function $-f''$ has been plotted. This is why, in the Bader approach, $-\Delta$ plays the role of a "magnifying glass". R.F.W. Bader, "*Atoms in Molecules. A Quantum Theory*", Clarendon Press, Oxford 1994, by permission from the Oxford University Press.

maximum close to x_2. If the cusp at $x = 0$ were absent,[10] $-f''$ would also exhibit a maximum at $x = 0$.

> We may say that $-\frac{d^2f}{dx^2}$ can detect some subtle features of the $f(x)$ plot and gives maxima where the original function $f(x)$ has only almost invisible "humps".

---

[10] Non-zero size of the nucleus or/and Gaussian type orbitals.

There is a similar story with the function $-\Delta\rho(x, y, z) = -(\frac{\partial^2\rho}{\partial x^2} + \frac{\partial^2\rho}{\partial y^2} + \frac{\partial^2\rho}{\partial z^2})$, except that here we have three Cartesian coordinates. The way we choose the directions of the Cartesian axes is irrelevant, because at any point of space $-\Delta\rho(x, y, z)$ does not depend on such a choice. Indeed, the coordinate systems, which we may choose, differ by an orthogonal transformation, which is peculiar for it does leave the *trace* of the Hessian (i.e. $\Delta\rho$) invariant.

Imagine now ρ of an atom decaying with the distance to the nucleus as $f(x)$, similar to the decay of a smoke cloud (Fig. 11.4.a), dense in the centre and vanishing outward. Let us calculate the Hessian at every point along the radius. It is easy to calculate $\Delta\rho(x, y, z)$ simply by summing up the diagonal terms of the Hessian. If we diagonalized the Hessian (i.e. rotated the axes in a particular way) its eigenvalues would correspond to the curvatures of the sections of ρ along the new coordinate axes (x', y', z'):

- *The section along the radius* (say z'). This curvature (see Fig. 11.2.a) is expected to be large and positive, since this is the direction ρ exponentially decays.
- *Two other sections: along x' and along y'* (only the first of them is shown in Fig. 11.2.a). These sections at a given radius mean cutting perpendicularly to the radius, and whether looking along x' or along y' we see the same: a larger value at the radius and a decay outside, i.e. both eigenvalues are negative.

> Fig. 11.4 displays ρ and $-\Delta\rho$ for the argon atom. Despite an apparent lack of any internal structure of the function ρ (left figure), the function $-\Delta\rho$ detected three concentrations of charge similar to the hump of the function $f(x)$ at x_2, Fig. 11.3. Thus, $-\Delta\rho(x, y, z)$ plays the role of a "magnifying glass": these are the K, L, M shells of the argon atom, seen very clearly.

Fig. 11.5 shows $-\Delta\rho$ for the systems N_2, Ar_2 and F_2. The figure highlights the shell character of the electronic structure of each of the atoms.[11] Fig. 11.2.c shows

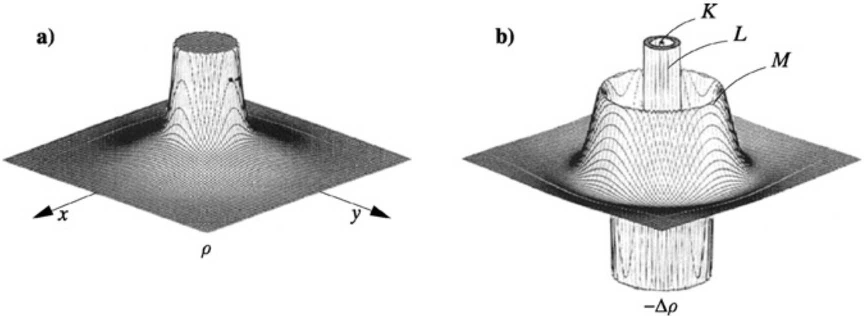


Fig. 11.4. A cross section of ρ (Fig. (a)) as well as a cross section of $-\Delta\rho$ (Fig. (b)) for the argon atom. The three humps (b) correspond to the K, L, M electron shells, cf. p. 380. R.F.W. Bader, "*Atoms in Molecules. A Quantum Theory*", Clarendon Press, Oxford, 1994, by permission from the Oxford University Press.

---

[11] Note, that the nitrogen as well as the fluorine have two shells (K and L), while the argon atom has three shells (K, L, M), cf. Chapter 8.

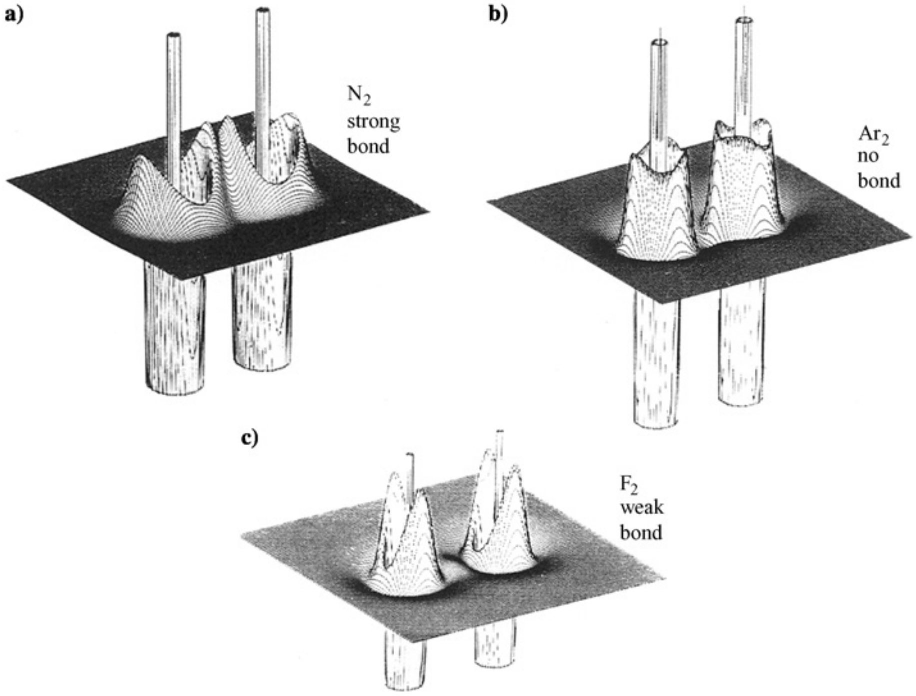


Fig. 11.5. A cross section of the quantity $-\Delta\rho$ for N_2, Ar_2 and F_2. We will focus now on the $-\Delta\rho$ value, computed in the middle of the internuclear distance. We can see that (a) for N_2 the value of $-\Delta\rho > 0$ (chemical bond), (b) for Ar–Ar, $-\Delta\rho < 0$ (no chemical bond) and (c) a very small positive $-\Delta\rho$ for F_2 (weak chemical bond). J.D. Dunitz, P. Seiler, *J. Am. Chem. Soc.* 105 (1983) 7056 even talk about the *decreasing* of ρ in the middle of F_2, when compared to the sum of densities of the non-interacting atoms. R.F.W. Bader, "*Atoms in Molecules. A Quantum Theory*", Clarendon Press, Oxford, 1994, by permission from the Oxford University Press.

that the electronic density is the greatest along the bond and drops outside in *each of the two orthogonal* directions. If, however, we went along the bond approaching any of the nuclei, the *density would increase*. This means that there is a saddle point of the second order, because one eigenvalue of the Hessian is positive and two negative.

If there were no covalent bond at all (non-bonded atoms or ionic bond: no electron density "rope" connecting the nuclei), the last two values would be zero, and this means that $-\Delta\rho < 0$. Thus, if it happens that for a bond $-\Delta\rho > 0$, this means a large perpendicular contribution, i.e. a "rope-like" covalent bond.

For the N_2 molecule we have a large value of $-\Delta\rho > 0$ between the nuclei, which means an electronic charge concentrated in a strong bond. The nuclei have, therefore, a dilemma, whether to run off, because they repel each other, or to run only a little, because there in the middle of the bond is such a beautiful negative charge (the nuclei choose the second possibility). This dilemma is absent in the Ar_2 system (Fig. 11.5.b): the electronic charge runs off the middle of the bond, the nuclei get uncovered and they run off. The molecule F_2 sticks together but not very strongly, just look at the internuclear region, $-\Delta\rho$ is quite low over there.[12]

11.3 TWO IMPORTANT HOHENBERG–KOHN THEOREMS

11.3.1 EQUIVALENCE OF THE ELECTRONIC WAVE FUNCTION AND ELECTRON DENSITY

Hohenberg and Kohn proved in 1964 an interesting theorem.[13]

> The ground-state electronic density $\rho_0(\boldsymbol{r})$ and the ground-state wave function Ψ_0 can be used alternatively as full descriptions of the ground state of the system.

This theorem is sometimes proved in a quite special way. Imagine somebody gave us $\rho_0(\boldsymbol{r})$ without a single word of explanation. We have no idea which system it corresponds to. First, we calculate $\int \rho_0(\boldsymbol{r})\,d^3\boldsymbol{r}$, where the integration goes over the whole space. This gives a natural number N, which is the number of electrons in the system. We did not know it, now we do. Next, we investigate the function $\rho_0(\boldsymbol{r})$ looking at its values, point by point. We are searching for the "spikes" (cusps), because every cusp tells us, where a nucleus is.[14] After this is done, we know all the positions of the nuclei. Now, we concentrate on each of the nuclei and look how fast the density drops, when leaving the nucleus. The calculated slope has to be equal to a negative even number: $-2Z$ (see p. 571), and Z gives us the charge of the nucleus. Thus, we have deduced the composition of our system. Now we are in a position to write down the Hamiltonian for the system and *solve the Schrödinger equation*. After that we know the ground-state wave function.[15]

Walter Kohn (b. 1923), American physicist of the Austrian origin, professor at the University of California – Santa Barbara. His conviction about the primary role the electronic density plays, led him to fundamental theoretical discoveries. Kohn shared the Nobel Prize with John A. Pople in 1998, receiving it "*for his development of the density-functional theory*".

---

[12] We see now why the F_2 molecule does not represent an easy task for the Hartree–Fock method (Chapter 8, the method indicated that the molecule ... does not exist).

[13] P. Hohenberg, W. Kohn, *Phys. Rev. B* 136 (1964) 864.

[14] $\rho(\boldsymbol{r})$ represents a cloud similar to those that float in the sky. This "spike", therefore, means simply a high density in the cloud.

[15] And all the excited states wave functions as well!

> We started, therefore, from $\rho_0(\boldsymbol{r})$, and we got the ground-state wave function Ψ_0. According to eq. (11.1), from the wave function by integration we obtain the density distribution $\rho_0(\boldsymbol{r})$. Hence, $\rho_0(\boldsymbol{r})$ contains the same precise information about the system as does Ψ_0.

Thus, if we know ρ_0, we also know Ψ_0, and, if we know Ψ_0, we also know ρ_0.[16]

v-representability

The proof we carried out pertains only to the case when the *external potential* (everything except the interelectronic interaction) acting on the electrons stems, from the nuclei. The Hohenberg–Kohn theorem can be proved for an arbitrary external potential – this property of the density is called the *v-representability*. The arbitrariness mentioned above is necessary in order to define in future the functionals for more general densities (than for isolated molecules). We will need that generality when introducing the functional derivatives (p. 587) in which $\rho(\boldsymbol{r})$ has to correspond to any external potential (or, to be a *v-representable density*). Also we will be interested in a *non-Coulombic* potential corresponding to the harmonic helium atom (cf. harmonium, p. 507) to see how exact the DFT method is. We may imagine a ρ, which is *not* v-representable, e.g., discontinuous (in one, two or even in every point like the Dirichlet function). The density distributions that are not v-representable are out of our field of interest.

Johann Lejeune-Dirichlet (1805–1859), German mathematician, professor at the universities in Berlin and Göttingen. His achievements are in number theory, infinite series, variational calculus and theory of potential. He is also notorious for designing the “most strange function ever” (discontinuous and nondifferentiable in every point) defined as $f(x)$ taking the value 1 for commensurate x and 0 for non-commensurate x.

11.3.2 EXISTENCE OF AN ENERGY FUNCTIONAL MINIMIZED BY ρ_0

Hohenberg and Kohn also proved an analogue of the variational principle (p. 196):

> For a given number of electrons (the integral over ρ equals N) and external potential v, there *exists* a functional of ρ, denoted by $E_v^{\mathrm{HK}}[\rho]$, for which the following variational principle is satisfied:
>
> $$E_v^{\mathrm{HK}}[\rho] \geqslant E_v^{\mathrm{HK}}[\rho_0] = E_0,$$
>
> where ρ_0 stands for the (ideal) electronic density distribution for the ground state of the energy E_0.

---

[16] The theorem just proved shines in its simplicity. People thought that the wave function, usually a very complicated mathematical object (that depends on $3N$ space and N spin coordinates) is indispensable for computing the properties of the system. Moreover, the larger the system the greater the difficulties in calculating it (please recall Chapter 10 with billions of excitations, non-linear parameters, etc.). Besides, how to interpret such a complex object? Horror. And it turns out that everything about the system just sits in $\rho(\boldsymbol{r})$, a function of position in our well known 3D space. It turns out that information about nuclei is hidden in such a simple object. This seems trivial (cusps), but we also have in $\rho(\boldsymbol{r})$ much

It appears that the energy E_0 represents the minimum value of a certain, unfortunately unknown, functional, and, this minimum value is obtained when inserting the density distribution ρ equal to the perfect ground-state density distribution ρ_0 into the functional.

We will prove this theorem using the variational principle in a way first given by Levy.[17] The variational principle states that $E_0 = \min\langle\Psi|\hat{H}|\Psi\rangle$, where we search among the wave functions Ψ normalized to 1 and describing N electrons. This minimization may be carried out in two steps:

$$E_0 = \min_{\rho,\int \rho\,\mathrm{d}V=N} \ \min_{\Psi\to\rho} \left\langle\Psi\middle|\hat{T}+U+V\middle|\Psi\right\rangle \tag{11.6}$$

where $\hat{T}, U, V$ represent the kinetic energy, the electron repulsion and the electron-nuclei attraction operators, respectively, for all the N electrons of our system (the hat in operators will be omitted, if the operator has a multiplicative character).

The minimization goes in two stages:

- The internal minimization is performed at the condition labelled as "$\Psi \to \rho$", which means minimization among the N-electron functions that are normalized to 1, and any of them give a *fixed* density distribution ρ "carrying" N electrons. As a result of this minimization we obtain a functional of ρ given as $\min_{\Psi\to\rho}\langle\Psi|\hat{T}+U+V|\Psi\rangle$.
- In the external minimization symbolized by "$\rho, \int\rho\,\mathrm{d}V = N$" we go over all the density distributions ρ that integrate to N (i.e. describe N electrons), and, we choose $\rho = \rho_0$ which minimizes $\min_{\Psi\to\rho}\langle\Psi|\hat{T}+U+V|\Psi\rangle$. According to the variational principle (p. 197), this minimum is bound to be the exact ground-state energy E_0, while ρ_0 is the exact ground-state density distribution.

Therefore, both minimizations do the same as the variational principle.

It is easy to show that $\langle\Psi|V\Psi\rangle$ may be expressed as an integral involving the density distribution ρ instead of Ψ. Indeed, since

$$V = \sum_{i=1}^{N} v(i), \quad \text{where } v(i) \equiv v(\boldsymbol{r}_i) = \sum_A -\frac{Z_A}{|\boldsymbol{r}_i - \boldsymbol{r}_A|} \tag{11.7}$$

(i.e. V is a sum of one-electron operators, $\langle\Psi|V\Psi\rangle = \sum_{i=1}^{N}\langle\Psi|v_i\Psi\rangle$), then in each of the resulting integrals $\langle\Psi|v(i)\Psi\rangle$ we may carry out the integration over all the electrons except one, and for this single one we sum over its spin coordinate. It is easy to see that every such term (their number is N) gives the same result $\langle\Psi|v(i)\Psi\rangle = \frac{1}{N}\int v(\boldsymbol{r})\rho(\boldsymbol{r})\,\mathrm{d}^3\boldsymbol{r}$, because the electrons are indistinguishable. Because of this we will get

$$\langle\Psi|V\Psi\rangle = \int v(\boldsymbol{r})\rho(\boldsymbol{r})\,\mathrm{d}^3\boldsymbol{r}. \tag{11.8}$$

---

more subtle information about how electrons avoid each other due to Coulombic repulsion and the Pauli exclusion principle.

[17] M. Levy, *Phys. Rev. A* 26 (1982) 1200.

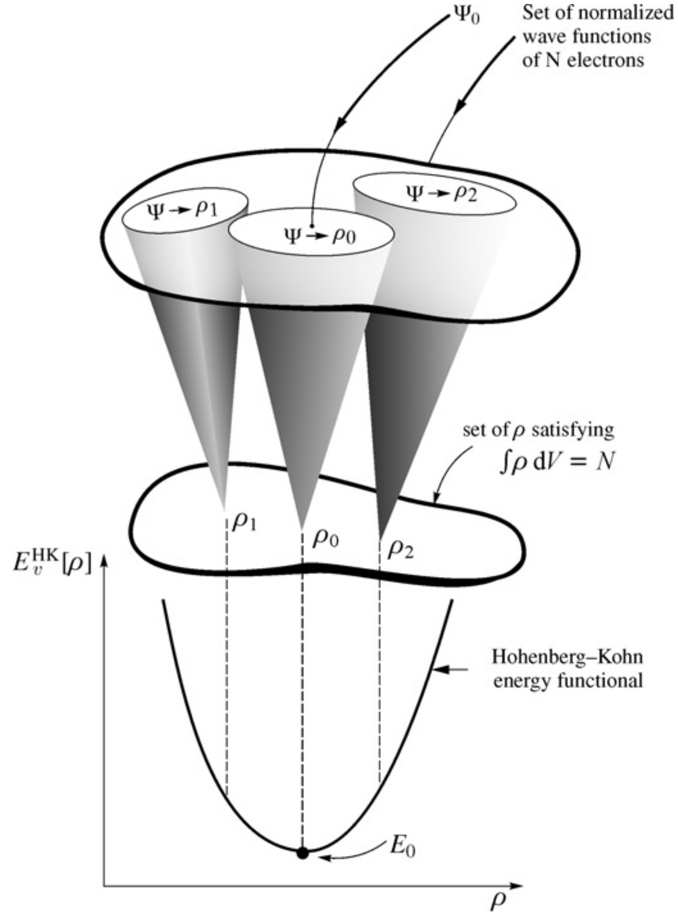


Fig. 11.6. Levy variational principle. The task of the internal minimization is: at a *given fixed* density distribution ρ carrying N electrons, choose among those normalized functions Ψ, *that all produce* ρ (we will denote this by the symbol "$\Psi \to \rho$"), *such a function that minimizes* $\langle \Psi | \hat{T} + U | \Psi \rangle$ of eq. (11.10). In the upper part of the figure three sets of such functions Ψ are shown: one set gives ρ_1, the second ρ_0, the third ρ_2. The external minimization symbolized by "$\rho, \int \rho \, dV = N$" chooses among all possible electron distributions ρ (that correspond to N electrons, the centre part of the figure) such a distribution $\rho = \rho_0$, that gives the lowest value (the ground state energy E_0, see the bottom part of the figure) of the Hohenberg–Kohn functional E_v^{HK}, i.e. $E_0 = \min_{\rho, \int \rho \, dV = N} E_v^{HK}$. Note that among the functions Ψ that give ρ_0 there is the exact ground-state wave function Ψ_0.

Therefore, the Levy minimization may be written as

$$E_0 = \min_{\rho, \int \rho \, dV = N} \left\{ \int v(\boldsymbol{r}) \rho(\boldsymbol{r}) \, d^3\boldsymbol{r} + \min_{\Psi \to \rho} \langle \Psi | (\hat{T} + U) \Psi \rangle \right\}. \qquad (11.9)$$

At this point we define the auxiliary functional[18] F^{HK}:

---

[18] A functional is always defined in a domain, in our case a domain of the allowed ρ's. How do the allowed ρ's look? Here are the conditions to fulfil: a) $\rho \geqslant 0$; b) $\int \rho \, dV = N$; c) $\nabla \rho^{1/2}$ square-integrable.

$$F^{\mathrm{HK}}[\rho]=\min_{\Psi\to\rho}\langle\Psi|(\hat{T}+U)\Psi\rangle\equiv\langle\Psi_{\min}|(\hat{T}+U)\Psi_{\min}\rangle, \tag{11.10}$$

where $\Psi_{\min}$ stands for a normalized function which has been chosen among those that produce a given ρ, and makes the smallest $\langle\Psi|\hat{T}+U|\Psi\rangle+\mathrm{const}$, where $\mathrm{const}=\int v\rho\,\mathrm{d}^3\boldsymbol{r}$. In the DFT we define the crucial *Hohenberg–Kohn functional* $E_v^{\mathrm{HK}}[\rho]$ as

Hohenberg–Kohn functional

$$E_v^{\mathrm{HK}}[\rho]=\int v(\boldsymbol{r})\rho(\boldsymbol{r})\,\mathrm{d}^3\boldsymbol{r}+F^{\mathrm{HK}}[\rho], \tag{11.11}$$

and the minimum of this functional is the ground-state energy

$$E_0=\min_{\rho,\int\rho\,\mathrm{d}V=N}E_v^{\mathrm{HK}}[\rho], \tag{11.12}$$

while ρ that minimizes $E_v^{\mathrm{HK}}[\rho]$ represents the exact ground-state density distribution ρ_0, Fig. 11.6. Each ρ corresponds to at least one antisymmetric electronic wave function (the "N-representability" mentioned above), and there is no better wave function than the ground-state, which, of course, corresponds to the density distribution ρ_0. This is why we have:

HOHENBERG–KOHN FUNCTIONAL:
The Hohenberg–Kohn functional $E_v^{\mathrm{HK}}[\rho]$ attains minimum $E_0=E_v^{\mathrm{HK}}[\rho_0]$ for the ideal density distribution.
Now our job will be to find out what mathematical form the functional could have. And here we meet the basic problem of the DFT method: nobody has so far been able to give such a formula. The best which has been achieved are some approximations. These approximations, however, are so good that they begin to supply results that satisfy chemists.

Therefore, when the question is posed: "*Is it possible to construct a quantum theory, in which the basic concept would be electronic density?*", we have to answer: "*yes, it is*". This answer, however, has only an existential value ("*yes, there exists*"). We have no information about how such a theory could be constructed.

An indication may come from the concept of the wave function. In order to proceed towards the above mentioned unknown functional, we will focus on the ingenious idea of a *fictitious Kohn–Sham system of non-interacting electrons*.

Among these conditions we do not find any that would require the existence of such an antisymmetric Ψ of N electrons that would correspond [in the sense of eq. (11.1)] to the density ρ under consideration (this is known as N-representability). It turns out that such a requirement is not needed, since it was proved by Thomas Gilbert (the proof may be found in the book by R.G. Parr and W. Yang "*Density-Functional Theory of Atoms and Molecules*", Oxford University Press, New York, 1989), that every ρ, that satisfies the above conditions, is N-representable, because it corresponds to at least one antisymmetric N-electron Ψ.

11.4 THE KOHN–SHAM EQUATIONS

11.4.1 THE KOHN–SHAM SYSTEM OF NON-INTERACTING ELECTRONS

Let us consider an electron subject to some "external" potential[19] $v(\boldsymbol{r})$, *for example* coming from the Coulombic interaction with the nuclei (with charges Z_α in a.u. and positions $\boldsymbol{r}_A$)

$$v(\boldsymbol{r}) = \sum_A -\frac{Z_A}{|\boldsymbol{r}-\boldsymbol{r}_A|}. \tag{11.13}$$

In this system we have N electrons, which also interact by Coulombic forces between themselves. All these interactions produce the ground-state electronic density distribution ρ_0 (ideal, i.e. that we obtain from the exact, 100% correlated wave function). Now let us consider

> FICTITIOUS KOHN–SHAM SYSTEM:
> the fictitious Kohn–Sham system of N model electrons (fermions), that *do not interact* at all (as if their charge were equal zero), but, instead of the interaction with the nuclei, they are subject to an external potential $v_0(\boldsymbol{r})$ so ingeniously tailored that ρ does not change, i.e. we still have the ideal ground-state electronic density $\rho = \rho_0$.

Let us assume for a while that we have found such a wonder potential $v_0(\boldsymbol{r})$. We will worry later about how to find it in reality. Now we assume the problem has been solved. Can we find ρ? Of course, we can. Since the Kohn–Sham electrons do not interact between themselves, we have only to solve the one-electron equation (with the wonder v_0)

$$\left(-\frac{1}{2}\Delta + v_0\right)\phi_i = \varepsilon_i\phi_i \tag{11.14}$$

where ϕ_i are some spinorbitals, of course, called the Kohn–Sham spinorbitals.[20]

---

[19] In the DFT the word "external potential" means any potential that is external with respect to a system of N electrons.

[20] If the electrons do not interact, the corresponding wave function can be taken as a *product* of the spinorbitals for individual electrons. Such a function for electrons is not antisymmetric, and, therefore, is "illegal". Taking the *Kohn–Sham determinant* (instead of the product) helps, because it is antisymmetric and represents an eigenfunction of the total Hamiltonian of the fictitious system [i.e. the sum of the one-electron operators given in parenthesis in eq. (11.14)]. This is easy to show, because a determinant represents a sum of products of the spinorbitals, the products differing only by permutation of electrons. If the total Hamiltonian of the fictitious system acts on such a sum, each term (product) is its eigenfunction, and each eigenvalue amounts to $\sum_{i=1}^{N}\varepsilon_i$, i.e. is the same for each product. Hence, the Kohn–Sham determinant represents an eigenfunction of the fictitious system. Scientists compared the Kohn–Sham orbitals with the canonical Hartree–Fock orbitals with great interest. It turns out that the differences are small.

The total wave function is a Slater determinant, which in our situation should rather be called the Kohn–Sham determinant. The electronic density distribution of such a system is given by eq. (11.5) and the density distribution ρ means *exact*, i.e. correlated 100% (thanks to the "wonder" and unknown operator v_0).

11.4.2 TOTAL ENERGY EXPRESSION

Let us try to write down a general expression for the electronic ground-state energy of the system under consideration. Obviously, we have to have in it the kinetic energy of the electrons, their interaction with the nuclei and their repulsion among themselves. However, in the DFT approach we write the following

$$E = T_0 + \int v(\boldsymbol{r})\rho(\boldsymbol{r})\,\mathrm{d}^3\boldsymbol{r} + J[\rho] + E_{\mathrm{xc}}[\rho], \tag{11.15}$$

where

- instead of the electronic kinetic energy of the system we write down (in cold blood) the electronic kinetic energy *of the fictitious Kohn–Sham system of (non-interacting) electrons* T_0 (please, recall the Slater–Condon rules, p. 986),

$$T_0 = -\frac{1}{2}\sum_{i=1}^{N}\langle\phi_i|\Delta\phi_i\rangle; \tag{11.16}$$

- next, there is the correct electron-nuclei interaction (or other external potential) term: $\int v(\boldsymbol{r})\rho(\boldsymbol{r})\,\mathrm{d}^3\boldsymbol{r}$;
- then, there is a *self-interaction* of the electron cloud with itself:[21]

self-interaction

$$J[\rho] = \frac{1}{2}\iint \frac{\rho(\boldsymbol{r}_1)\rho(\boldsymbol{r}_2)}{|\boldsymbol{r}_1 - \boldsymbol{r}_2|}\,\mathrm{d}^3\boldsymbol{r}_1\,\mathrm{d}^3\boldsymbol{r}_2. \tag{11.17}$$

No doubt, the energy expression might contain such a self-interaction, but this is certainly not all that should be included in the electron–electron interaction. Two electrons repel each other electrostatically and therefore around each of them there has to exist a kind of no-parking zone for the other one ("Coulomb hole",

---

[21] How to compute the Coulombic interaction within a storm cloud exhibiting certain charge distribution ρ? At first sight it looks like a difficult problem, but remember we know how to calculate the Coulombic interaction of two *point* charges. Let us divide the whole cloud into tiny cubes, each of volume $\mathrm{d}V$. The cube that is pointed by the vector $\boldsymbol{r}_1$ contains a tiny charge $\rho(\boldsymbol{r}_1)\,\mathrm{d}V \equiv \rho(\boldsymbol{r}_1)\,\mathrm{d}^3\boldsymbol{r}_1$. We know that when calculating the Coulombic interaction of two such cubes we have to write: $\frac{\rho(\boldsymbol{r}_1)\rho(\boldsymbol{r}_2)}{|\boldsymbol{r}_1-\boldsymbol{r}_2|}\,\mathrm{d}^3\boldsymbol{r}_1\,\mathrm{d}^3\boldsymbol{r}_2$. This has to be summed over all possible positions of the first and the second cube:$\iint \frac{\rho(\boldsymbol{r}_1)\rho(\boldsymbol{r}_2)}{|\boldsymbol{r}_1-\boldsymbol{r}_2|}\,\mathrm{d}^3\boldsymbol{r}_1\,\mathrm{d}^3\boldsymbol{r}_2$, but in such a way each interaction is computed twice, because they represent parts of the same cloud. Hence, the final self-interaction of the storm cloud is $\frac{1}{2}\iint \frac{\rho(\boldsymbol{r}_1)\rho(\boldsymbol{r}_2)}{|\boldsymbol{r}_1-\boldsymbol{r}_2|}\,\mathrm{d}^3\boldsymbol{r}_1\,\mathrm{d}^3\boldsymbol{r}_2$. The expression for the self-interaction of the *electron* cloud is the same.

cf. p. 515). Also a no-parking zone results, because electrons of the same spin coordinate hate one another[22] ("exchange", or "Fermi hole", cf. p. 516). The integral J does not take such a correlation of motions into account.

Thus, we have written a few terms and we do not know what to write down next. Well,

> in the DFT in the expression for E we write in (11.15) the lacking remainder as E_{xc}, and we call it the *exchange–correlation energy* (label x stands for "exchange", c is for "correlation") and declare, courageously, that we will manage somehow to get it.

exchange–correlation energy

The above formula represents a definition of the *exchange–correlation energy*, although it is rather a strange definition – it requires us to know E. We should not forget that in E_{xc} a correction to the kinetic energy has also to be included (besides the exchange and correlation effects) that takes into account that kinetic energy has to be calculated for the true (i.e. interacting) electrons, not for the non-interacting Kohn–Sham ones. All this stands to reason if E_{xc} is small as compared to E. The next question is connected to what kind of mathematical form E_{xc} might have. Let us assume, for the time being we have no problem with this mathematical form. For now we will establish a relation between our wonder external potential v_0 and our mysterious E_{xc}, both quantities performing miracles, but not known.

11.4.3 DERIVATION OF THE KOHN–SHAM EQUATIONS

Now we will make a variation of E, i.e. we will find the linear effect of changing E due to a variation of the spinorbitals (and therefore also of the density). We make a spinorbital variation denoted by $\delta\phi_i$ (as before, p. 336, it is justified to vary either ϕ_i or ϕ_i^*, the result is the same, we choose, therefore, $\delta\phi_i^*$) and see what effect it will have on E keeping only the linear term. We have (see eq. (11.4)),

$$\phi_i^* \to \phi_i^* + \delta\phi_i^*, \tag{11.18}$$

$$\rho \to \rho + \delta\rho, \tag{11.19}$$

$$\delta\rho(\boldsymbol{r}) = \sum_\sigma \sum_{i=1}^{N} \delta\phi_i^*(\boldsymbol{r},\sigma)\phi_i(\boldsymbol{r},\sigma). \tag{11.20}$$

We insert the right-hand sides of the above expressions into E, and identify the variation, i.e. the linear part of the change of E. The variations of the individual terms of E look like (note, see p. 334, that the symbol $\langle|\rangle$ stands for an integral over space coordinates and a summation over the spin coordinates):

$$\delta T_0 = -\frac{1}{2}\sum_{i=1}^{N}\langle\delta\phi_i|\Delta\phi_i\rangle, \tag{11.21}$$

---

[22] A correlated density and a non-correlated density differ in that in the correlated one we have smaller values in the high-density regions, because the holes make the overcrowding of space by electrons less probable.

$$\delta \int v\rho \, \mathrm{d}^3\boldsymbol{r} = \int v\delta\rho \, \mathrm{d}^3\boldsymbol{r} = \sum_{i=1}^{N} \langle \delta\phi_i | v\phi_i \rangle, \tag{11.22}$$

$$\begin{aligned}
\delta J &= \frac{1}{2}\left[\int \frac{\rho(\boldsymbol{r}_1)\delta\rho(\boldsymbol{r}_2)}{|\boldsymbol{r}_1 - \boldsymbol{r}_2|} \, \mathrm{d}^3\boldsymbol{r}_1 \, \mathrm{d}^3\boldsymbol{r}_2 + \int \frac{\delta\rho(\boldsymbol{r}_1)\rho(\boldsymbol{r}_2)}{|\boldsymbol{r}_1 - \boldsymbol{r}_2|} \, \mathrm{d}^3\boldsymbol{r}_1 \, \mathrm{d}^3\boldsymbol{r}_2 \right] \\
&= \int \frac{\rho(\boldsymbol{r}_1)\delta\rho(\boldsymbol{r}_2)}{|\boldsymbol{r}_1 - \boldsymbol{r}_2|} \, \mathrm{d}^3\boldsymbol{r}_1 \, \mathrm{d}^3\boldsymbol{r}_2 \\
&= \sum_{i,j=1}^{N} \big\langle \delta\phi_i(\boldsymbol{r}_2, \sigma_2) \big| \hat{J}_j(\boldsymbol{r}_2)\phi_i(\boldsymbol{r}_2, \sigma_2) \big\rangle_2,
\end{aligned} \tag{11.23}$$

where $\langle \ldots | \ldots \rangle_2$ means integration over spatial coordinates and the summation over the spin coordinate of electron 2, with the Coulomb operator $\hat{J}_j$ associated with the spinorbital ϕ_j

$$\hat{J}_j(\boldsymbol{r}_2) = \sum_{\sigma_1} \int \frac{\phi_j(\boldsymbol{r}_1, \sigma_1)^* \phi_j(\boldsymbol{r}_1, \sigma_1)}{|\boldsymbol{r}_1 - \boldsymbol{r}_2|} \, \mathrm{d}^3\boldsymbol{r}_1. \tag{11.24}$$

Finally, we come to the variation of E_{xc}, i.e. δE_{xc}. We are in a quite difficult situation, because we do not know the mathematical dependence of the functional E_{xc} on ρ, and therefore also on $\delta\phi_i^*$. Nevertheless, we somehow have to get the linear part of E_{xc}, i.e. the variation.

A change of functional $F[f]$ (due to $f \to f + \delta f$) contains a part linear in δf denoted by δF plus some higher powers[23] of δf denoted by $\mathrm{O}((\delta f)^2)$

$$F[f + \delta f] - F[f] = \delta F + \mathrm{O}\big((\delta f)^2\big). \tag{11.25}$$

functional derivative

The δF is defined through the *functional derivative* (Fig. 11.7) of F with respect to the function f (denoted by $\frac{\delta F[f]}{\delta f(x)}$), for a single variable[24] x

$$\delta F = \int_a^b \mathrm{d}x \, \frac{\delta F[f]}{\delta f(x)} \delta f(x). \tag{11.26}$$

Indeed, in our case we obtain as δE_{xc}:

$$\delta E_{\mathrm{xc}} = \int \mathrm{d}^3\boldsymbol{r} \, \frac{\delta E_{\mathrm{xc}}}{\delta\rho(\boldsymbol{r})} \delta\rho(\boldsymbol{r}) = \sum_{i=1}^{N} \left\langle \delta\phi_i \middle| \frac{\delta E_{\mathrm{xc}}}{\delta\rho} \phi_i \right\rangle. \tag{11.27}$$

---

[23] If δf is very small, the higher terms are negligible.

[24] Just for the sake of simplicity. The functional derivative itself is a functional of f and a function of x. An example of a functional derivative may be found in eq. (11.23), when looking at $\delta J = \int \frac{\rho(\boldsymbol{r}_1)\delta\rho(\boldsymbol{r}_2)}{|\boldsymbol{r}_1 - \boldsymbol{r}_2|} \, \mathrm{d}\boldsymbol{r}_1^3 \, \mathrm{d}\boldsymbol{r}_2^3 = \int \mathrm{d}\boldsymbol{r}_2^3 \{\int \mathrm{d}\boldsymbol{r}_1^3 \frac{\rho(\boldsymbol{r}_1)}{|\boldsymbol{r}_1 - \boldsymbol{r}_2|}\} \delta\rho(\boldsymbol{r}_2)$. Indeed, as we can see from eq. (11.26) $\int \mathrm{d}\boldsymbol{r}_1^3 \frac{\rho(\boldsymbol{r}_1)}{|\boldsymbol{r}_1 - \boldsymbol{r}_2|} \equiv \frac{\delta J[\rho]}{\delta\rho(\boldsymbol{r}_2)}$, which is a 3D equivalent of $\frac{\delta F[f]}{\delta f(x)}$. Note, that $\int \mathrm{d}\boldsymbol{r}_1^3 \frac{\rho(\boldsymbol{r}_1)}{|\boldsymbol{r}_1 - \boldsymbol{r}_2|}$ is a functional of ρ and a function of $\boldsymbol{r}_2$.

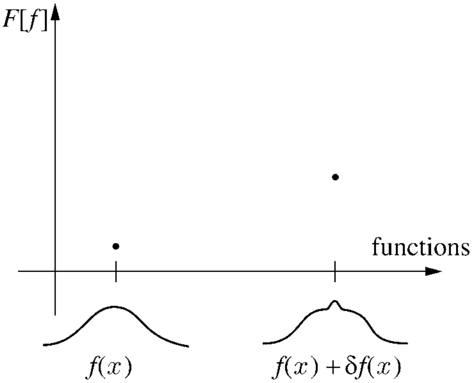


Fig. 11.7. A scheme showing what a functional derivative is about. The ordinate represents the values of a functional $F[f]$, while each point of the horizontal axis represents a function $f(x)$. The functional $F[f]$ depends, of course, on details of the function $f(x)$. If we consider a *small local change* of $f(x)$, this change may result in a large change of F – then the derivative $\frac{\delta F}{\delta f}$ is large, or in a small change of F – then the derivative $\frac{\delta F}{\delta f}$ is small (this depends on the particular functional).

Therefore, the unknown quantity E_{xc} is replaced by the unknown quantity $\frac{\delta E_{\text{xc}}}{\delta \rho}$, but there is profit from this: the functional derivative enables us to write an equation for spinorbitals. The variations of the spinorbitals are not arbitrary in this formula – they have to satisfy the orthonormality conditions (because our formulae, e.g., (11.4), are valid only for such spinorbitals) for $i, j = 1, \dots, N$, which gives

$$\langle \delta\phi_i | \phi_j \rangle = 0 \quad \text{for } i, j = 1, 2, \dots, N. \tag{11.28}$$

Let us multiply each of eqs. (11.28) by a Lagrange multiplier[25] ε_{ij}, add them together, then subtract from the variation δE and write the result as equal to zero (in the minimum we have $\delta E = 0$). We obtain

$$\delta E - \sum_{i,j}^{N} \varepsilon_{ij} \langle \delta\phi_i | \phi_j \rangle = 0 \tag{11.29}$$

or

$$\sum_{i=1}^{N} \left\langle \delta\phi_i \middle| \left\{ \left[-\frac{1}{2}\Delta + v + \sum_{j=1}^{N} \hat{J}_j + \frac{\delta E_{\text{xc}}}{\delta \rho} \right] \phi_i - \sum_{i,j}^{N} \varepsilon_{ij} \phi_j \right\} \right\rangle = 0. \tag{11.30}$$

After inserting the Lagrange multipliers, the variations of ϕ_i^* are already *independent* and the only way to have zero on the right-hand side is that every individual ket $|\rangle$ is zero (Euler equation, cf. p. 998):

$$\left\{ -\frac{1}{2}\Delta + v + v_{\text{coul}} + v_{\text{xc}} \right\} \phi_i = \sum_{i,j}^{N} \varepsilon_{ij} \phi_j, \tag{11.31}$$

$$v_{\text{coul}}(\boldsymbol{r}) \equiv \sum_{j=1}^{N} \hat{J}_j(\boldsymbol{r}), \tag{11.32}$$

$$v_{\text{xc}}(\boldsymbol{r}) \equiv \frac{\delta E_{\text{xc}}}{\delta \rho(\boldsymbol{r})}. \tag{11.33}$$

It would be good now to get rid of the non-diagonal Lagrange multipliers in order to obtain a beautiful one-electron equation analogous to the Fock equation. To

---

[25] Appendix N, p. 997.

this end we need the operator in the curly brackets in (11.31) to be invariant with respect to an arbitrary unitary transformation of the spinorbitals. The sum of the Coulomb operators (v_{coul}) is invariant, as has been demonstrated on p. 340. As to the unknown functional derivative $\delta E_{\text{xc}}/\delta\rho$, i.e. potential v_{xc}, its invariance follows from the fact that it is a functional of ρ (and ρ of eq. (11.4) is invariant). Finally, we obtain the Kohn–Sham equation ($\varepsilon_{ii} = \varepsilon_i$).

KOHN–SHAM EQUATION

$$\left\{-\frac{1}{2}\Delta + v + v_{\text{coul}} + v_{\text{xc}}\right\}\phi_i = \varepsilon_i\phi_i. \tag{11.34}$$

The equation is analogous to the Fock equation (p. 341). We solve the Kohn–Sham equation by an iterative method. We start from any zero-iteration orbitals. This enables us to calculate a zero approximation to ρ, and then the zero approximations to the operators v_{coul} and v_{xc} (in a moment we will see how to compute E_{xc}, and then using (11.33), we obtain v_{xc}). The solution to the Kohn–Sham equation gives new orbitals and new ρ. The procedure is then repeated until consistency is achieved.

iterative method

Hence, finally we "know" what the wonder operator v_0 looks like:

$$v_0 = v + v_{\text{coul}} + v_{\text{xc}}. \tag{11.35}$$

There is no problem with v_{coul}, a serious difficulty arises with the exchange–correlation operator v_{xc}, or (equivalent) with the energy E_{xc}. The second Hohenberg–Kohn theorem says that the functional $E_v^{\text{HK}}[\rho]$ exists, but it does not guarantee that it is simple. For now we worry about this potential, but we courageously go ahead.

Kohn–Sham equations with spin polarization

Before searching for v_{xc} let us generalize the Kohn–Sham formalism and split ρ into that part which comes from electrons with the α spin function and those with the β spin function. If these contributions are not equal (even for some $\boldsymbol{r}$), we will have a *spin polarization*). In order to do this, we consider two non-interacting fictitious electron systems: one described by the spin functions α, and the other – by functions β, with the corresponding density distributions $\rho_\alpha(\boldsymbol{r})$ and $\rho_\beta(\boldsymbol{r})$ exactly equal to ρ_α and ρ_β, respectively, in the (original) interacting system. Of course, for any system we have

spin polarization

$$\rho = \rho_\alpha + \rho_\beta, \tag{11.36}$$

which follows from the summation over σ in eq. (11.1). Then, we obtain two coupled[26] Kohn–Sham equations with potential v_0 that depends on the spin coordi-

---

[26] Through the common operator v_{coul}, a functional of $\rho_\alpha + \rho_\beta$ and through v_{xc}, because the later is in general a functional of both, ρ_α and ρ_β.

nate σ

$$v_0^{\sigma} = v + v_{\text{coul}} + v_{\text{xc}}^{\sigma}. \tag{11.37}$$

The situation is analogous to the unrestricted Hartree–Fock method (UHF), cf. p. 342.

This extension of the DFT is known as spin density functional theory (SDFT).

11.5 WHAT TO TAKE AS THE DFT EXCHANGE–CORRELATION ENERGY E_{xc}?

We approach the point where we promised to write down the mysterious exchange–correlation energy. Well, how to tell you the truth? Let me put it straightforwardly: we do not know the analytical form of this quantity. Nobody knows what the exchange–correlation is, there are only guesses. The number of formulae will be, as usual with guesses, almost unlimited.[27] Let us take the simplest ones.

11.5.1 LOCAL DENSITY APPROXIMATION (LDA)

The electrons in a molecule are in quite a complex situation, because they not only interact between themselves, but also with the nuclei. However, a simpler system has been elaborated theoretically for years: a homogeneous gas model in a box,[28] an electrically neutral system (the nuclear charge is smeared out uniformly). It does not represent the simplest system to study, but it turns out that theory is able to determine (exactly) some of its properties. For example, it has been deduced how E_{xc} depends on ρ, and even how it depends on ρ_α and ρ_β. Since the gas is homogeneous and the volume of the box is known, then we could easily work out how the E_{xc} per unit volume depends on these quantities.

Then, the reasoning is the following.[29]

> The electronic density distribution in a molecule is certainly inhomogeneous, but locally (within a small volume) we may assume its homogeneity. Then, if someone asks about the exchange–correlation energy contribution from this small volume, we would say that in principle we do not know, but to a good approximation the contribution could be calculated as a product of the small volume and the exchange–correlation energy density from the homogeneous gas theory (calculated inside the small volume).

Thus, everything is decided locally: we have a sum of contributions from each infinitesimally small element of the electron cloud with the corresponding density.

---

[27] Some detailed formulae are reported in the book by J.B. Foresman and A. Frisch, "*Exploring Chemistry with Electronic Structure Methods*", Gaussian, Pittsburgh, PA, USA, p. 272.

[28] With periodic boundary conditions. This is a common trick to avoid the surface problem. We consider a box having such a property, that if something goes out through one wall it enters through the opposite wall (cf. p. 446).

[29] W. Kohn, L.J. Sham, *Phys. Rev*. 140 (1965) A1133.

This is why it is called the local density approximation (LDA, when the ρ dependence is used) or the local spin density approximation (LSDA, when the ρ_α and ρ_β dependencies are exploited).

LSDA or LDA

11.5.2 NON-LOCAL APPROXIMATIONS (NLDA)

Gradient Expansion Approximation

There are approximations that go beyond the LDA. They consider that the dependence $E_{\text{xc}}[\rho]$ may be *non-local*, i.e. E_{xc} may depend on ρ at a given point (locality), but also on ρ nearby (non-locality). When we are at a point, what happens further off depends not only on ρ at that point, but also the gradient of ρ at the point, etc.[30] This is how the idea of the gradient expansion approximation (GEA) appeared

non-local functionals

$$E_{\text{xc}}^{\text{GEA}} = E_{\text{xc}}^{\text{LSDA}} + \int B_{\text{xc}}(\rho_\alpha, \rho_\beta, \nabla\rho_\alpha, \nabla\rho_\beta)\, \mathrm{d}^3\boldsymbol{r}, \tag{11.38}$$

where the exchange–correlation function B_{xc} is carefully selected as a function of ρ_α, ρ_β and their gradients, in order to maximize the successes of the theory/experiment comparison. However, this recipe was not so simple and some strange unexplained discrepancies were still taking place.

Perdew–Wang functional (PW91)

A breakthrough in the quality of results is represented by the following proposition of Perdew and Wang:

$$E_{\text{xc}}^{\text{PW91}} = \int f(\rho_\alpha, \rho_\beta, \nabla\rho_\alpha, \nabla\rho_\beta)\, \mathrm{d}^3\boldsymbol{r} \tag{11.39}$$

where the function f of ρ_α, ρ_β and their gradients has been tailored in an ingenious way. It sounds unclear, but it will be shown below that their approximation used some fundamental properties and this enabled them *without introducing any parameters* to achieve a much better agreement between the theory and experiment.

The famous B3LYP hybrid functional

The B3LYP approach belongs to the *hybrid approximations* for the exchange–correlation functional. The approximation is famous, because it gives very good results and, therefore, is extremely popular. So far so good, but there is a danger of Babylon type science.[31] It smells like a witch's brew for the B3LYP exchange–correlation potential E_{xc}: *take the exchange–correlation energy from the LSDA method, add a pinch (20%) of the difference between the Hartree–Fock exchange energy*[32] E_{x}^{KS} *and the LSDA* $E_{\text{x}}^{\text{LSDA}}$. *Then, mix well 72% of Becke exchange potential*

hybrid approximation

---

[30] As in a Taylor series, then we may need not only the gradient, but also the Laplacian, etc.

[31] The Chaldean priests working out "Babylonian science" paid attention to making their small formulae efficient. The ancient Greeks (contemporary science owes them so much) were in favour of crystal clear reasoning.

[32] In fact, this is Kohn–Sham exchange energy, see eq. (11.69), because the Slater determinant wave function, used to calculate it, is the Kohn–Sham determinant, not the Hartree–Fock one.

E_x^{B88} *which includes the 1988 correction, then strew in 81% of the Lee–Yang–Parr correlation potential* E_c^{LYP}. You will like this homeopathic magic potion most (a "hybrid") if you conclude by putting in 19% of the Vosko–Wilk–Nusair potential[33] E_c^{VWN}:

$$E_{\mathrm{xc}} = E_{\mathrm{xc}}^{\mathrm{LSDA}} + 0.20\left(E_{\mathrm{x}}^{\mathrm{HF}} - E_{\mathrm{x}}^{\mathrm{LSDA}}\right) + 0.72E_{\mathrm{x}}^{B88} + 0.81E_{\mathrm{c}}^{\mathrm{LYP}} + 0.19E_{\mathrm{c}}^{\mathrm{VWN}}. \quad (11.40)$$

If you do it this way – satisfaction is (almost) guaranteed, your results will agree very well with experiment.

11.5.3 THE APPROXIMATE CHARACTER OF THE DFT VS APPARENT RIGOUR OF *ab initio* COMPUTATIONS

There are lots of exchange–correlation potentials in the literature. There is an impression that their authors worried most about theory/experiment agreement. We can hardly admire this kind of science, but the alternative, i.e. the practice of *ab initio* methods with the intact and "holy" Hamiltonian operator, has its own dark sides and smells a bit of witch's brew too. Yes, because finally we have to choose a given atomic basis set, and this influences the results. It is true that we have the variational principle at our disposal, and it is possible to tell which result is more accurate. But more and more often in quantum chemistry we use some non-variational methods (cf. Chapter 10). Besides, the Hamiltonian holiness disappears when the theory becomes relativistic (cf. Chapter 3).

Everybody would like to have agreement with experiment and no wonder people tinker at the exchange–correlation enigma. This tinkering, however, is by no means arbitrary. There are some serious physical restraints behind it, which will be shown in a while.

11.6 ON THE PHYSICAL JUSTIFICATION FOR THE EXCHANGE CORRELATION ENERGY

We have to introduce several useful concepts such as the "electron pair distribution function", and the "electron hole" (in a more formal way than in Chapter 10, p. 515), etc.

11.6.1 THE ELECTRON PAIR DISTRIBUTION FUNCTION

pair correlation function

From the N-electron wave function we may compute what is called the *electron pair correlation function* $\Pi(\boldsymbol{r}_1, \boldsymbol{r}_2)$, in short, a pair function defined as[34]

$$\Pi(\boldsymbol{r}_1, \boldsymbol{r}_2) = N(N-1) \sum_{\sigma_1,\sigma_2} \int |\Psi|^2 \, d\tau_3 \, d\tau_4 \ldots d\tau_N \quad (11.41)$$

---

[33] S.H. Vosko, L. Wilk, M. Nusair, *Can. J. Phys.* 58 (1980) 1200.

[34] The function represents the diagonal element of the *two-particle electron density matrix*:

$$\Gamma(\boldsymbol{r}_1, \boldsymbol{r}_2; \boldsymbol{r}_1', \boldsymbol{r}_2') = N(N-1) \sum_{\text{all } \sigma} \int \Psi^*(\boldsymbol{r}_1'\sigma_1, \boldsymbol{r}_2', \sigma_2, \boldsymbol{r}_3, \sigma_3, \ldots, \boldsymbol{r}_N, \sigma_N) \times \Psi(\boldsymbol{r}_1, \sigma_1, \boldsymbol{r}_2, \sigma_2, \boldsymbol{r}_3, \sigma_3, \ldots, \boldsymbol{r}_N, \sigma_N) \, d^3\boldsymbol{r}_3 \, d^3\boldsymbol{r}_4 \ldots d^3\boldsymbol{r}_N,$$

$$\Pi(\boldsymbol{r}_1, \boldsymbol{r}_2) \equiv \Gamma(\boldsymbol{r}_1, \boldsymbol{r}_2; \boldsymbol{r}_1, \boldsymbol{r}_2).$$

where the summation over spin coordinates pertains to all electrons (for the electrons $3, 4, \ldots, N$ the summation is hidden in the integrals over $d\tau$), while the integration is over the space coordinates of the electrons $3, 4, \ldots, N$.

The function $\Pi(\boldsymbol{r}_1, \boldsymbol{r}_2)$ measures the probability density of finding one electron at the point indicated by $\boldsymbol{r}_1$ and another at $\boldsymbol{r}_2$, and tells us how the motions of two electrons are correlated. If Π were a *product* of two functions $\rho_1(\boldsymbol{r}_1) > 0$ and $\rho_2(\boldsymbol{r}_2) > 0$, then this motion is not correlated (because the probability of two events represents a product of the probabilities for each of the events only for *independent, i.e. uncorrelated events*).

Function Π appears in a natural way, when we compute the mean value of the total electronic repulsion $\langle\Psi|U|\Psi\rangle$ with the Coulomb operator $U = \sum_{i<j}^{N} \frac{1}{r_{ij}}$ and a normalized N-electron wave function Ψ. Indeed, we have ("prime" in the summation corresponds to omitting the diagonal term)

$$\begin{aligned}
\langle\Psi|U\Psi\rangle &= \frac{1}{2}\sum_{i,j=1}^{N}{}' \left\langle \Psi \middle| \frac{1}{r_{ij}}\Psi \right\rangle \\
&= \frac{1}{2}\sum_{i,j=1}^{N}{}' \left\{ \sum_{\sigma_i,\sigma_j} \int \mathrm{d}^3\boldsymbol{r}_i\, \mathrm{d}^3\boldsymbol{r}_j \frac{1}{r_{ij}} \int |\Psi|^2 \frac{\mathrm{d}\tau_1\, \mathrm{d}\tau_2 \ldots \mathrm{d}\tau_N}{\mathrm{d}\tau_i\, \mathrm{d}\tau_j} \right\} \\
&= \frac{1}{2}\sum_{i,j=1}^{N}{}' \int \mathrm{d}^3\boldsymbol{r}_i\, \mathrm{d}^3\boldsymbol{r}_j \frac{1}{r_{ij}} \frac{1}{N(N-1)} \Pi(\boldsymbol{r}_i, \boldsymbol{r}_j) \\
&= \frac{1}{2}\frac{1}{N(N-1)} \sum_{i,j=1}^{N}{}' \int \mathrm{d}^3\boldsymbol{r}_1\, \mathrm{d}^3\boldsymbol{r}_2 \frac{1}{r_{12}} \Pi(\boldsymbol{r}_1, \boldsymbol{r}_2) \\
&= \frac{1}{2}\frac{1}{N(N-1)} \int \mathrm{d}^3\boldsymbol{r}_1\, \mathrm{d}^3\boldsymbol{r}_2 \frac{\Pi(\boldsymbol{r}_1, \boldsymbol{r}_2)}{r_{12}} \sum_{i,j=1}^{N}{}' 1 \\
&= \frac{1}{2}\int \mathrm{d}^3\boldsymbol{r}_1\, \mathrm{d}^3\boldsymbol{r}_2 \frac{\Pi(\boldsymbol{r}_1, \boldsymbol{r}_2)}{r_{12}}.
\end{aligned} \tag{11.42}$$

We will need this result in a moment. We see, that to determine the contribution of the electron repulsions to the total energy we need the two-electron function Π. The first Hohenberg–Kohn theorem tells us that it is sufficient to know something simpler, namely, the electronic density ρ. How to reconcile these two demands?

The further DFT story will pertain to the question: how to change the potential in order to replace Π by ρ?

11.6.2 THE QUASI-STATIC CONNECTION OF TWO IMPORTANT SYSTEMS

To begin let us write two Hamiltonians that are certainly very important for our goal: the first is the total Hamiltonian of our system (of course, with the Coulombic electron–electron interactions). Let us denote the operator for some reasons as $H(\lambda = 1)$, cf. eqs. (11.6) and (11.7):

$$\hat{H}(\lambda = 1) = \sum_{i=1}^{N}\left[-\frac{1}{2}\Delta_i + v(i)\right] + U. \tag{11.43}$$

The second Hamiltonian $H(\lambda = 0)$ pertains to the Kohn–Sham fictitious system of the *non-interacting* electrons (it contains our wonder v_0, which we solemnly promise to search for, and the kinetic energy operator and …nothing else, cf. eq. (11.14)):

$$\hat{H}(\lambda = 0) = \sum_{i=1}^{N}\left[-\frac{1}{2}\Delta_i + v_0(i)\right]. \tag{11.44}$$

We will try to connect these two important systems by generating some intermediate Hamiltonians $\hat{H}(\lambda)$ for λ *intermediate* between 0, and 1:

$$\hat{H}(\lambda) = \sum_{i=1}^{N}\left[-\frac{1}{2}\Delta_i + v_\lambda(i)\right] + U(\lambda), \tag{11.45}$$

where

$$U(\lambda) = \lambda \sum_{i<j}^{N} \frac{1}{r_{ij}}.$$

Note, that our electrons are not real for intermediate values of λ (each electron carries the electric charge $\sqrt{\lambda}$).

> The intermediate Hamiltonian $\hat{H}(\lambda)$ contains a mysterious v_λ, which generates the exact density distribution ρ that corresponds to the Hamiltonian $\hat{H}(\lambda = 1)$, i.e. with all interactions in place. The same exact ρ corresponds to $\hat{H}(\lambda = 0)$.

We have, therefore, the ambition to go from the $\lambda = 0$ situation to the $\lambda = 1$ situation, all the time guaranteeing that the antisymmetric ground-state eigenfunction of $\hat{H}(\lambda)$ for any λ gives *the same electron density distribution* ρ, *the ideal (exact)*. We decide to follow the path of the exact electron density distribution and measure our way by the value of λ. The way chosen represents a kind of "path of life" for us, because by sticking to it we do not lose the most precious of our treasures: the ideal density distribution ρ. We will call this path the *quasi-static transition*, because all the time we will adjust the correction computed to our actual position on the path.

quasi-static transition

Our goal will be the total energy $E(\lambda = 1)$. The quasi-static transition will be carried out by tiny steps. We will start with $E(\lambda = 0)$, and end up with $E(\lambda = 1)$:

$$E(\lambda = 1) = E(\lambda = 0) + \int_0^1 E'(\lambda)\,\mathrm{d}\lambda, \tag{11.46}$$

where the increments $\mathrm{d}E(\lambda) = E'(\lambda)\,\mathrm{d}\lambda$ will be calculated as the first-order perturbation energy correction, eq. (5.22). The first-order correction is sufficient, because we are going to apply only infinitesimally small λ increments.[35] Each time, when λ changes from λ to $\lambda + \mathrm{d}\lambda$, the situation at λ [i.e. the Hamiltonian $\hat{H}(\lambda)$ and the wave function $\Psi(\lambda)$] will be treated as unperturbed. What, therefore, does the perturbation operator look like? Well, when we go from λ to $\lambda + \mathrm{d}\lambda$, the Hamiltonian changes by perturbation $\hat{H}^{(1)}(\lambda) = \mathrm{d}\hat{H}(\lambda)$. Then, the first-order perturbation correction to the energy given by (5.22), represents the mean value of $\mathrm{d}\hat{H}(\lambda)$ with the unperturbed function $\Psi(\lambda)$:

$$\mathrm{d}E(\lambda) = \big\langle \Psi(\lambda) \big| \mathrm{d}\hat{H}(\lambda)\Psi(\lambda)\big\rangle, \tag{11.47}$$

where in $\mathrm{d}\hat{H}$ we only have a change of v_λ and of $U(\lambda)$ due to the change of λ:

$$\mathrm{d}\hat{H}(\lambda) = \sum_{i=1}^{N} \mathrm{d}v_\lambda(i) + \mathrm{d}\lambda \sum_{i<j}^{N} \frac{1}{r_{ij}}. \tag{11.48}$$

Note that we have succeeded in writing such a simple formula, *because the kinetic energy operator stays unchanged all the time* (*it does not depend on* λ). Let us insert this into the first-order correction to the energy in order to get $\mathrm{d}E(\lambda)$:

$$\begin{aligned}\mathrm{d}E(\lambda) &= \big\langle \Psi(\lambda) \big| \mathrm{d}\hat{H}(\lambda)\Psi(\lambda)\big\rangle \\ &= \int \rho(\boldsymbol{r})\,\mathrm{d}v_\lambda(\boldsymbol{r})\,\mathrm{d}^3\boldsymbol{r} + \frac{1}{2}\mathrm{d}\lambda \iint \mathrm{d}^3\boldsymbol{r}_1\,\mathrm{d}^3\boldsymbol{r}_2\,\frac{\Pi_\lambda(\boldsymbol{r}_1,\boldsymbol{r}_2)}{r_{12}}.\end{aligned} \tag{11.49}$$

In the last formula we introduced a function Π_λ that is an analogue of the pair function Π, but pertains to the electrons carrying the charge $\sqrt{\lambda}$ (we have used the formula (11.42), noting that we have a λ-dependent wave function $\Psi(\lambda)$).

In order to go from $E(\lambda = 0)$ to $E(\lambda = 1)$, it is sufficient just to integrate this expression from 0 to 1 over λ (this corresponds to the infinitesimally small increments of λ as mentioned before). Note that (by definition) *ρ does not depend on* λ, which is of fundamental importance in the success of the integration $\int \rho(\boldsymbol{r})\,\mathrm{d}v_\lambda(\boldsymbol{r})\,\mathrm{d}^3\boldsymbol{r}$ and gives the result

$$\begin{aligned}E(\lambda = 1) - E(\lambda = 0) = {}& \int \rho(\boldsymbol{r})\{v - v_0\}(\boldsymbol{r})\,\mathrm{d}^3\boldsymbol{r} \\ & + \frac{1}{2}\int_0^1 \mathrm{d}\lambda \iint \mathrm{d}^3\boldsymbol{r}_1\,\mathrm{d}^3\boldsymbol{r}_2\,\frac{\Pi_\lambda(\boldsymbol{r}_1,\boldsymbol{r}_2)}{r_{12}}.\end{aligned} \tag{11.50}$$

The energy for $\lambda = 0$, i.e. for the non-interacting electrons in an unknown external

---

[35] λ plays a different role here than the perturbational parameter λ on p. 205.

potential v_0 will be written as (cf. the formulas (11.14) and (11.16)):

$$E(\lambda=0)=\sum_i \varepsilon_i = T_0+\int \rho(\boldsymbol{r})v_0(\boldsymbol{r})\,\mathrm{d}^3\boldsymbol{r}. \tag{11.51}$$

Inserting this into (11.50) we obtain $E(\lambda=1)$, i.e. the energy of our original system:

$$E(\lambda=1)=T_0+\int \rho(\boldsymbol{r})v(\boldsymbol{r})\,\mathrm{d}^3\boldsymbol{r}+\frac{1}{2}\int_0^1 \mathrm{d}\lambda \iint \mathrm{d}^3\boldsymbol{r}_1\,\mathrm{d}^3\boldsymbol{r}_2\,\frac{\Pi_\lambda(\boldsymbol{r}_1,\boldsymbol{r}_2)}{r_{12}}. \tag{11.52}$$

This expression may be simplified by introducing the pair distribution function Π_{aver} which is the $\Pi_\lambda(\boldsymbol{r}_1,\boldsymbol{r}_2)$ averaged over $\lambda=[0,1]$:

$$\Pi_{\text{aver}}(\boldsymbol{r}_1,\boldsymbol{r}_2)\equiv\int_0^1 \Pi_\lambda(\boldsymbol{r}_1,\boldsymbol{r}_2)\,\mathrm{d}\lambda. \tag{11.53}$$

Finally we obtain the following expression for the total energy E:

$$\boxed{E(\lambda=1)=T_0+\int \rho(\boldsymbol{r})v(\boldsymbol{r})\,\mathrm{d}^3\boldsymbol{r}+\frac{1}{2}\iint \mathrm{d}^3\boldsymbol{r}_1\,\mathrm{d}^3\boldsymbol{r}_2\,\frac{\Pi_{\text{aver}}(\boldsymbol{r}_1,\boldsymbol{r}_2)}{r_{12}}.} \tag{11.54}$$

Note that this equation is similar to the total energy expression appearing in traditional quantum chemistry[36] (without repulsion of the nuclei),

$$E=T+\int \rho(\boldsymbol{r})v(\boldsymbol{r})\,\mathrm{d}^3\boldsymbol{r}+\frac{1}{2}\iint \mathrm{d}^3\boldsymbol{r}_1\,\mathrm{d}^3\boldsymbol{r}_2\,\frac{\Pi(\boldsymbol{r}_1,\boldsymbol{r}_2)}{r_{12}}, \tag{11.55}$$

where in the last term we recognize the mean repulsion energy of electrons (obtained a while before). As we can see, the DFT total energy expression, instead of the mean kinetic energy of the fully interacting electrons T, contains T_0, i.e. the mean kinetic energy of the non-interacting (Kohn–Sham) electrons.[37] We pay, however, a price, which is that we need to compute the function Π_{aver} somehow. *Note, however, that the correlation energy dragon has been driven into the problem of finding a two-electron function* Π_{aver}.

11.6.3 EXCHANGE–CORRELATION ENERGY vs Π_{aver}

What is the relation between Π_{aver} and the exchange–correlation energy E_{xc} introduced earlier? We find that immediately, comparing the total energy given in eqs. (11.15) and (11.17) and now in (11.54). It is seen that the exchange–correlation energy

$$\boxed{E_{\text{xc}}=\frac{1}{2}\iint \mathrm{d}^3\boldsymbol{r}_1\,\mathrm{d}^3\boldsymbol{r}_2\,\frac{1}{r_{12}}\bigl\{\Pi_{\text{aver}}(\boldsymbol{r}_1,\boldsymbol{r}_2)-\rho(\boldsymbol{r}_1)\rho(\boldsymbol{r}_2)\bigr\}.} \tag{11.56}$$

---

[36] It is evident from the mean value of the total Hamiltonian [taking into account the mean value of the electron–electron repulsion, eq. (11.8)] and (11.42).

[37] As a matter of fact the whole Kohn–Sham formalism with the fictitious system of the non-interacting electrons has been designed precisely because of this.

The energy looks as if it were a potential energy, but it implicitly incorporates (in Π_{aver}) the kinetic energy correction for changing the electron non-interacting to the electron-interacting system.

Now let us try to get some information about the integrand, i.e. Π_{aver}, by introducing the notion of the electron hole.

11.6.4 ELECTRON HOLES

Electrons do not like each other, which manifests itself by Coulombic repulsion. On top of that, two electrons having the same spin coordinates hate each other (Pauli exclusion principle) and also try to get out of the other electron way. This has been analyzed in Chapter 10, p. 516. We should somehow highlight these features, because both concepts are basic and simple.

First, we will decompose the function Π_{aver} into the components related to the spin functions[38] of electrons 1 and 2 $\alpha\alpha$, $\alpha\beta$, $\beta\alpha$, $\beta\beta$,

$$\Pi_{\text{aver}} = \Pi_{\text{aver}}^{\alpha\alpha} + \Pi_{\text{aver}}^{\alpha\beta} + \Pi_{\text{aver}}^{\beta\alpha} + \Pi_{\text{aver}}^{\beta\beta} \equiv \sum_{\sigma\sigma'} \Pi_{\text{aver}}^{\sigma\sigma'}, \tag{11.57}$$

where $\Pi_{\text{aver}}^{\alpha\beta}\,\mathrm{d}V_1\,\mathrm{d}V_2$ represents a measure of the probability density[39] that two electrons are in their small boxes indicated by the vectors $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$, have the volumes $\mathrm{d}V_1$ and $\mathrm{d}V_2$, and are described by the spin functions α and β (the other components of Π_{aver} are defined in a similar way). Since $\rho = \rho_\alpha + \rho_\beta$, the exchange–correlation energy can be written as[40]

$$E_{\text{xc}} = \frac{1}{2}\sum_{\sigma\sigma'} \iint \mathrm{d}^3\boldsymbol{r}_1\,\mathrm{d}^3\boldsymbol{r}_2\,\frac{\Pi_{\text{aver}}^{\sigma\sigma'}(\boldsymbol{r}_1,\boldsymbol{r}_2) - \rho_\sigma(\boldsymbol{r}_1)\rho_{\sigma'}(\boldsymbol{r}_2)}{r_{12}}, \tag{11.58}$$

where the summation goes over the spin coordinates. It is seen that

> E_{xc} tells us how the behaviour of electrons deviates from their *independence* (the later is described by the product of the probability densities, i.e. the second term in the nominator). This means that E_{xc} has to contain the electron–electron correlation resulting from Coulombic interaction and their avoidance from the Pauli exclusion principle.

We wish to represent the integral as a Coulombic interaction of $\rho_\sigma(\boldsymbol{r}_1)$ with the density distribution of electron 2 to see how electron 2 "flees in panic" when seeing electron 1. We will try do this by inserting $\rho_\sigma(\boldsymbol{r}_1)$ into the nominator of (11.58) and the correctness of the formula will be assured by the unknown hole function h:

$$E_{\text{xc}} = \frac{1}{2}\sum_{\sigma\sigma'} \int \mathrm{d}^3\boldsymbol{r}_1 \int \mathrm{d}^3\boldsymbol{r}_2\,\frac{\rho_\sigma(\boldsymbol{r}_1)}{r_{12}} h_{\text{xc}}^{\sigma\sigma'}(\boldsymbol{r}_1,\boldsymbol{r}_2). \tag{11.59}$$

---

[38] Such a decomposition follows from eq. (11.41). We average all the contributions $\Pi^{\sigma\sigma'}$ separately and obtain the formula.

[39] λ-averaged.

[40] Simply, each Π_{aver} "in the spin resolution" will find its product of the spin density distributions – this is what we have as the nominator in the integrand.

It is precisely the function h that describes the distribution of electron 2, when electron 1 with spin coordinate σ is characterized by its density distribution $\rho_\sigma(\boldsymbol{r}_1)$

EXCHANGE–CORRELATION HOLE

$$h_{\mathrm{xc}}^{\sigma\sigma'}(\boldsymbol{r}_1,\boldsymbol{r}_2)=\frac{\Pi_{\mathrm{aver}}^{\sigma\sigma'}(\boldsymbol{r}_1,\boldsymbol{r}_2)}{\rho_\sigma(\boldsymbol{r}_1)}-\rho_{\sigma'}(\boldsymbol{r}_2). \tag{11.60}$$

This means that the hole represents that part of the pair distribution function that is inexplicable by a product-like dependence. Since a product function describes independent electrons, the hole function grasps the "intentional" avoidance of the two electrons.

We have, therefore, four exchange–correlation holes: $h_{\mathrm{xc}}^{\alpha\alpha}$, $h_{\mathrm{xc}}^{\alpha\beta}$, $h_{\mathrm{xc}}^{\beta\alpha}$, $h_{\mathrm{xc}}^{\beta\beta}$.

11.6.5 PHYSICAL BOUNDARY CONDITIONS FOR HOLES

Note that if a hole h were not spherically symmetric (i.e. contained a spherically symmetric component plus some non-spherical components) with respect to the position of electron 1, the contribution of such a hole to the integral in (11.59) over $d^3\boldsymbol{r}_2$ would come only from its spherically symmetric component, because the operator $1/r_{12}$ is spherically symmetric.

The hole functions are of fundamental importance in the DFT, because they have to fulfil some boundary requirements. The requirements are unable to fix the precise mathematical form of the hole functions (and therefore of E_{xc}), but the form is heavily restricted by the boundary conditions.

What boundaries we are talking about? These boundaries are different for electrons with the same spin coordinates[41] to those corresponding to the opposite spins. For a pair of electrons with the same spins, we have to have the following result of integration over $\boldsymbol{r}_2$ (for any $\boldsymbol{r}_1$):[42]

$$\int h_{\mathrm{xc}}^{\alpha\alpha}(\boldsymbol{r}_1,\boldsymbol{r}_2)\,d^3\boldsymbol{r}_2=\int h_{\mathrm{xc}}^{\beta\beta}(\boldsymbol{r}_1,\boldsymbol{r}_2)\,d^3\boldsymbol{r}_2=-1, \tag{11.61}$$

which means that

---

[41] These boundaries come from the symmetry properties of the wave function forced by the Pauli exclusion principle. For example, from the antisymmetry of the wave function (with respect to the exchange of labels of two electrons), it follows that two electrons of the same spin coordinate cannot meet in a point in space. Indeed, if they did, *all* their coordinates would be the same, their exchange meant, therefore, nothing, whereas it has to change the sign of the wave function. The only possibility is to make the value of the function equal to zero. This reasoning cannot be repeated with two electrons of opposite spins, and such a meeting is, therefore, possible. Of course, around any electron there should be a Coulombic hole because of Coulomb repulsion. However, the degree of taking such a hole into account depends on the quality of the wave function (e.g., a Hartree–Fock function will not give any Coulomb hole, cf. p. 515).

[42] We do not prove that here.

just one electron (of the same spin) has escaped from the space around electron 1 as compared to the independent particle model.

Since for $\boldsymbol{r}_1 = \boldsymbol{r}_2 = \boldsymbol{r}$ we have to have (Pauli exclusion principle) $\Pi^{\alpha\alpha}(\boldsymbol{r},\boldsymbol{r}) = \Pi^{\beta\beta}(\boldsymbol{r},\boldsymbol{r}) = 0$, therefore, we extend this property on the corresponding Π_{aver} noting that the Pauli exclusion principle operates at any λ (the model electrons are fermions). Then, after inserting $\Pi^{\alpha\alpha}_{\mathrm{aver}}(\boldsymbol{r},\boldsymbol{r}) = \Pi^{\beta\beta}_{\mathrm{aver}}(\boldsymbol{r},\boldsymbol{r}) = 0$ into (11.60) we obtain the following conditions

$$h^{\alpha\alpha}_{\mathrm{xc}}(\boldsymbol{r},\boldsymbol{r}) = -\rho_{\alpha}(\boldsymbol{r}), \tag{11.62}$$

$$h^{\beta\beta}_{\mathrm{xc}}(\boldsymbol{r},\boldsymbol{r}) = -\rho_{\beta}(\boldsymbol{r}). \tag{11.63}$$

If similar calculations were done for the electrons with opposite spins, we would obtain

$$\int h^{\alpha\beta}_{\mathrm{xc}}(\boldsymbol{r}_1,\boldsymbol{r}_2)\,\mathrm{d}^3\boldsymbol{r}_2 = \int h^{\beta\alpha}_{\mathrm{xc}}(\boldsymbol{r}_1,\boldsymbol{r}_2)\,\mathrm{d}^3\boldsymbol{r}_2 = 0, \tag{11.64}$$

which means that the *Pauli exclusion principle* alone does not give any restriction to the residence of electron 2 in the neighbourhood of electron 1 of *opposite* spin, compared to what happens when the electrons are independent.

11.6.6 EXCHANGE AND CORRELATION HOLES

The restrictions introduced come from the Pauli exclusion principle and hence have been related to the exchange energy. So far no restriction has appeared that would stem from the Coulombic interactions of electrons.[43] This made people think of differentiating the holes into two contributions: exchange hole h_{x} and correlation hole h_{c} (so far called the Coulombic hole). Let us begin with a formal division of the exchange–correlation energy into the exchange and the correlation parts:

EXCHANGE–CORRELATION ENERGY

$$E_{\mathrm{xc}} = E_{\mathrm{x}} + E_{\mathrm{c}} \tag{11.65}$$

and *we will say that we know, what the exchange part is*.

The DFT exchange energy (E_x) is calculated in the same way as in the Hartree–Fock method, but with the Kohn–Sham determinant. The correlation energy E_{c} represents just a rest.

This is again the same strategy of chasing the electronic correlation dragon into a hole, this time into the correlation hole. When we do not know a quantity, we

---

[43] This is the role of the Hamiltonian.

write down what we know plus a remainder. And the dragon with a hundred heads sits in it. Because of this division, the Kohn–Sham equation will contain the sum of the exchange and correlation potentials instead of v_{xc}:

$$\boxed{v_{xc} = v_x + v_c \qquad (11.66)}$$

with

$$v_x \equiv \frac{\delta E_x}{\delta\rho}, \qquad (11.67)$$

$$v_c \equiv \frac{\delta E_c}{\delta\rho}. \qquad (11.68)$$

Let us recall what the Hartree–Fock exchange energy[44] looks like [Chapter 8, eq. (8.35)]. The Kohn–Sham exchange energy looks, of course, the same, except that the spinorbitals are now Kohn–Sham, not Hartree–Fock. Therefore, we have the exchange energy E_x as (the sum is over the molecular spinorbitals[45])

$$\begin{aligned} E_x &= -\frac{1}{2}\sum_{i,j=1}^{\mathrm{SMO}} K_{ij} = -\frac{1}{2}\sum_{i,j=1}^{\mathrm{SMO}} \langle ij|ji\rangle \\ &= -\frac{1}{2}\sum_{\sigma}\int \frac{\{\sum_{i=1}^{N}\phi_i^*(1)\phi_i(2)\}\{\sum_{j=1}^{N}\phi_j^*(2)\phi_j(1)\}}{r_{12}}\,\mathrm{d}^3\boldsymbol{r}_1\,\mathrm{d}^3\boldsymbol{r}_2 \\ &= -\frac{1}{2}\sum_{\sigma}\int \frac{|\rho_\sigma(\boldsymbol{r}_1,\boldsymbol{r}_2)|^2}{r_{12}}\,\mathrm{d}^3\boldsymbol{r}_1\,\mathrm{d}^3\boldsymbol{r}_2, \end{aligned} \qquad (11.69)$$

where (cf. p. 531)

$$\rho_\sigma(\boldsymbol{r}_1,\boldsymbol{r}_2) \equiv \sum_{i=1}^{N}\phi_i(\boldsymbol{r}_1,\sigma)\phi_i^*(\boldsymbol{r}_2,\sigma) \qquad (11.70)$$

represents the *one-particle* density matrix for the σ subsystem, and ρ_σ is obtained from the Kohn–Sham determinant. Note that density $\rho_\sigma(\boldsymbol{r})$ is its diagonal, i.e. $\rho_\sigma(\boldsymbol{r}) \equiv \rho_\sigma(\boldsymbol{r},\boldsymbol{r})$.

The above may be incorporated into the exchange energy E_x equal to

$$\boxed{E_x = \frac{1}{2}\sum_{\sigma\sigma'}\iint \mathrm{d}^3\boldsymbol{r}_1\,\mathrm{d}^3\boldsymbol{r}_2\,\frac{\rho_\sigma(\boldsymbol{r}_1)}{r_{12}}h_x^{\sigma\sigma'}(\boldsymbol{r}_1,\boldsymbol{r}_2), \qquad (11.71)}$$

if the exchange hole h is proposed as

$$h_x^{\sigma\sigma'}(\boldsymbol{r}_1,\boldsymbol{r}_2) = \delta_{\sigma\sigma'}\left\{-\frac{|\rho_\sigma(\boldsymbol{r}_1,\boldsymbol{r}_2)|^2}{\rho_\sigma(\boldsymbol{r}_1)}\right\}. \qquad (11.72)$$

---

[44] The one which appeared from the exchange operator, i.e. containing the exchange integrals.

[45] Note that spinorbital i has to have the same spin function as spinorbital j (otherwise $K_{ij} = 0$).

It is seen that the exchange hole is negative everywhere[46] and diagonal in the spin index. What, therefore, does the correlation hole look like? According to the philosophy of dragon hunting it is the rest

$$h_{\mathrm{xc}}^{\sigma\sigma'} = h_{\mathrm{x}}^{\sigma\sigma'} + h_{\mathrm{c}}^{\sigma\sigma'}. \tag{11.73}$$

The correlation energy from eq. (11.65) has, therefore, the form:

$$\boxed{E_{\mathrm{c}} = \frac{1}{2}\sum_{\sigma\sigma'}\iint \mathrm{d}^3\boldsymbol{r}_1\,\mathrm{d}^3\boldsymbol{r}_2\,\frac{\rho_\sigma(\boldsymbol{r}_1)}{r_{12}}h_{\mathrm{c}}^{\sigma\sigma'}(\boldsymbol{r}_1,\boldsymbol{r}_2).} \tag{11.74}$$

Since the exchange hole has already fulfilled the boundary conditions (11.62)–(11.64), forced by the Pauli exclusion principle, the correlation hole satisfies a simple boundary condition

$$\int h_{\mathrm{c}}^{\sigma\sigma'}(\boldsymbol{r}_1,\boldsymbol{r}_2)\,\mathrm{d}^3\boldsymbol{r}_2 = 0. \tag{11.75}$$

The dragon of electronic correlation has been chased into the correlation hole. Numerical experience turns out to conclude later on[47] that

> the exchange energy E_{x} is more important than the correlation energy E_{c}

and, therefore, the dragon in the hole has been considerably weakened by scientists.

11.6.7 PHYSICAL GROUNDS FOR THE DFT APPROXIMATIONS

LDA

The LDA is not as primitive as it looks. The electron density distribution for the homogeneous gas model satisfies the Pauli exclusion principle and, therefore, this approximation gives the Fermi holes that fulfil the boundary conditions with eqs. (11.62), (11.63) and (11.64). The LDA is often used because it is rather inexpensive, while still giving reasonable geometry of molecules and vibrational frequencies.[48] The quantities that the LDA fails to reproduce are the binding energies,[49] ionization potentials and the intermolecular dispersion interaction.

The Perdew–Wang functional (PW91)

Perdew noted a really dangerous feature in an innocent and reasonable looking GEA potential. It turned out that in contrast to the LDA the boundary conditions for the electron holes were not satisfied. For example, the exchange hole was not

---

[46] Which has its origin in the minus sign before the exchange integrals in the total energy expression.

[47] Below we give an example.

[48] Some colleagues of mine sometimes add a malicious remark that the frequencies are so good that they even take into account the anharmonicity of the potential.

[49] The average error in a series of molecules may even be of the order of 40 kcal/mol; this is a lot, since the chemical bond energy is of the order of about 100 kcal/mol.

negative everywhere as eq. (11.72) requires. Perdew and Wang corrected this deficiency in a way similar to that of Alexander the Great, when he cut in 333 B.C. the Gordian knot. They tailored the formula for E_{xc} in such a way as to change the positive values of the function just to zero, while the peripheral parts of the exchange holes were cut to force the boundary conditions to be satisfied anyway. The authors noted an important improvement in the results.

The functional B3LYP

It was noted that the LDA and even GEA models systematically give too large chemical bond energies. On the other hand it was known that the Hartree–Fock method is notorious for making the bonds too weak. What are we to do? Well, just mix the two types of potential and hope to have an improvement with respect to any of the models. Recall the formula (11.53) for Π_{aver}, where the averaging extended from $\lambda = 0$ to $\lambda = 1$. The contribution to the integral for λ close to 0 comes from the situations similar to the fictitious model of non-interacting particles, where the wave function has the form of the Kohn–Sham determinant. Therefore, those contributions contain the exchange energy E_x corresponding to such a determinant. We may conclude that a contribution from the Kohn–Sham exchange energy E_x^{HF} might look quite natural.[50] This is what the B3LYP method does, eq. (11.40). Of course, it is not possible to justify the particular proportions of the B3LYP ingredients. Such things are justified only by their success.[51]

11.7 REFLECTIONS ON THE DFT SUCCESS

The DFT method has a long history behind it, which began with Thomas, Dirac, Fermi, etc. At the beginning the successes were quite modest (the electron gas theory, known as the Xα method). Real success came after a publication by Jan Andzelm and Erich Wimmer.[52] The DFT method, offering results at a correlated level for a wide spectrum of physical quantities, turned out to be roughly as inexpensive as the Hartree–Fock procedure – this is the most sensational feature of the method.

We have a beacon – exact electron density distribution of harmonium

Hohenberg and Kohn proved their famous theorem on the existence of the energy functional, but nobody was able to give the functional for any system. All the

---

[50] The symbol HF pertains rather to Kohn–Sham than to Hartree–Fock.

[51] As in homeopathy.

[52] J. Andzelm, E. Wimmer, *J. Chem. Phys*. 96 (1992) 1280. Jan was my PhD student in the old days. In the paper by A. Scheiner, J. Baker, J. Andzelm, *J. Comp. Chem*. 18 (1997) 775 the reader will find an interesting comparison of the methods used. One of the advantages (or deficiencies) of the DFT methods is that they offer a wide variety of basis functions (in contrast to the *ab initio* methods, where Gaussian basis sets rule), recommended for some particular problems to be solved. For example, in electronics (Si, Ge) the plane wave $\exp(i\boldsymbol{kr})$ expansion is a preferred choice. On the other hand these functions are not advised for catalysis phenomena with rare earth atoms. The Gaussian basis sets in the DFT had a temporary advantage (in the nineties of the twentieth century) over others, because the standard Gaussian programs offered analytically computed gradients (for optimization of the geometry). Now this is also offered by many DFT methodologies.

DFT efforts are directed towards elaborating such a potential, and the only criterion of whether a model is any good, is comparison with experiment. However, it turned out that there is a system for which every detail of the DFT can be verified. Uniquely, the dragon may be driven out the hole and we may fearlessly and with impunity analyze all the details of its anatomy. The system is a bit artificial, it is the harmonic helium atom (harmonium) discussed on p. 185, in which the two electrons attract the nucleus by a harmonic force, while repelling each other by Coulombic interaction. For some selected force constants k, e.g., for $k = \frac{1}{4}$, the Schrödinger equation *can be solved analytically*. The wave function is extremely simple, see p. 507. The electron density (normalized to 2) is computed as

$$\rho_0(\boldsymbol{r}) = 2N_0^2 \mathrm{e}^{-\frac{1}{2}r^2}\left\{\left(\frac{\pi}{2}\right)^{\frac{1}{2}}\left[\frac{7}{4}+\frac{1}{4}r^2+\left(r+\frac{1}{r}\right)\mathrm{erf}\left(\frac{r}{\sqrt{2}}\right)\right]+\mathrm{e}^{-\frac{1}{2}r^2}\right\}, \tag{11.76}$$

where erf is the error function, $\mathrm{erf}(z) = \frac{2}{\sqrt{\pi}}\int_0^z \exp(-u^2)\,\mathrm{d}u$, and

$$N_0^2 = \frac{\pi^{\frac{3}{2}}}{(8+5\sqrt{\pi})}. \tag{11.77}$$

We should look at the $\rho_0(\boldsymbol{r})$ with a great interest – it is a unique occasion, it is probable you will never again see an *exact* result. The formula is not only exact, but on top of this it is simple. Kais et al. compare the exact results with two DFT methods: the BLYP (a version of B3LYP) and the Becke–Perdew (BP) method.[53]

Because of the factor $\exp(-0.5r^2)$ the density distribution ρ is concentrated on the nucleus.[54] The authors compare this density distribution with the corresponding Hartree–Fock density (appropriate for the potential used), and even with the density distribution related to the hydrogen-like atom (after neglecting $1/r_{12}$ in the Hamiltonian the wave function becomes an antisymmetrized product of the two hydrogen-like orbitals). In the later case the electrons do not see each other[55] and the corresponding density distribution is too concentrated on the nucleus. As soon as the term $1/r_{12}$ is restored, the electrons immediately move apart and ρ on the nucleus decreases by about 30%. The second result is also interesting: the Hartree–Fock density is very close to ideal – it is almost the same curve.[56]

Total energy components

It turns out that in the case analyzed (and so far only in this case) we can calculate the *exact* total energy E [eq. (11.15)], "wonder" potential v_0 that in the Kohn–Sham model gives the exact density distribution ρ [eq. (11.76)], exchange potential v_x and correlation v_c [eqs. (11.67) and (11.68)].[57] Let us begin from the total energy.

---

[53] The detailed references to these methods are given in S. Kais, D.R. Herschbach, N.C. Handy, C.W. Murray, G.J. Laming, *J. Chem. Phys.* 99 (1993) 417.

[54] As it should be.

[55] Even in the sense of the mean field (as it is in the Hartree–Fock method).

[56] This is why the HF method is able to give 99.6% of the total energy. Nevertheless, in some cases this may not be a sufficient accuracy.

[57] These potentials as functions of ρ or r.

Table 11.1. Harmonium (harmonic helium atom). Comparison of the components (a.u.) of the total energy $E[\rho_0]$ calculated by the HF, BLYP and BP methods with the exact values (row KS)

	$E[\rho_0]$	$T_0[\rho_0]$	$\int v\rho_0\,\mathrm{d}\boldsymbol{r}$	$J[\rho_0]$	$E_{\mathrm{x}}[\rho_0]$	$E_{\mathrm{c}}[\rho_0]$
KS	**2.0000**	**0.6352**	**0.8881**	**1.032**	**−0.5160**	**−0.0393**
HF	2.0392	0.6318	0.8925	1.030	−0.5150	0
BLYP	2.0172	0.6313	0.8933	1.029	−0.5016	−0.0351
BP	1.9985	0.6327	0.8926	1.028	−0.5012	−0.0538

In the second row of Table 11.1 labelled KS, the exact total energy is reported ($E[\rho_0] = 2.0000$ a.u.) and its components (bold figures) calculated according to eqs. (11.15), (11.16), (11.8), (11.17), (11.65) and (11.69). The exact correlation energy E_{c} is calculated as the difference between the exact total energy and the listed components. Thus, $T_0[\rho_0]$ stands for the kinetic energy of the non-interacting electrons, $\int v\rho_0\,\mathrm{d}^3\boldsymbol{r}$ means the electron–nucleus attraction (positive, because the harmonic potential is positive), and $J[\rho_0]$ represents the self-interaction energy of ρ_0. According to eq. (11.17) and taking into account ρ_0, i.e. twice a square of the orbital, we obtain $J[\rho_0] = 2\mathcal{J}_{11}$ with the Coulombic integral $\mathcal{J}_{11}$. On the other hand the exchange energy is given by eq. (11.69): $E_x = -\frac{1}{2}\sum_{i,j=1}^{\mathrm{SMO}} K_{ij}$, and after summing over the spin coordinates we obtain the exchange energy $E_{\mathrm{x}} = -\mathcal{K}_{11} = -\mathcal{J}_{11}$. We see such a relation between J and E_{x} in the second row (KS[58]). The other rows report already various approximations computed by: HF, BLYP, BP, each of them giving its own Kohn–Sham spinorbitals and its own approximation of the density distribution ρ_0. This density distribution was used for the calculation of the components of the total energy within each approximate method. Of course, the Hartree–Fock method gave 0 for the correlation energy (third row), because there is no correlation in it except that which follows from the Pauli exclusion principle fully taken into account in the exchange energy (cf. Chapter 10, p. 516).

We see that all the methods are doing quite well. The BLYP gives the total energy with an error of 0.87% – twice as small as the Hartree–Fock method, while the BP functional missed by as little as 0.08%. The total energy components are a bit worse, which proves that a certain cancellation of errors occurs among the energy components. The KS kinetic energy T_0 amounts to 0.6352, while that calculated as the mean value of the kinetic energy operator (of two electrons) is 0.6644, a bit larger – the rest is in the exchange–correlation energy.[59]

Exact "wonder" v_0 potential

Fig. 11.8 shows our "wonder" long awaited potential v_0 as a function of r, and alternatively as a function of $\rho^{\frac{1}{3}}$. The exact $v_0(r)$ represents a monotonic function increasing with r and represents a modification (influence of the second electron) of the external potential v, we see that v_0 is shifted *upwards* with respect to v, because

---

[58] Only for spin-compensated two-electron systems we have $E_x[\rho] = -\frac{1}{2}J[\rho_0]$ and, therefore $v_x = \frac{\delta E_x}{\delta\rho}$ can be calculated analytically. In all other cases, although E_x can be easily evaluated (knowing orbitals), the calculation of v_x is very difficult and costly (it can only be done numerically). In the present two-electron case v_x^{HF} is a multiplicative operator rather than integral operator.

[59] As we have described before.

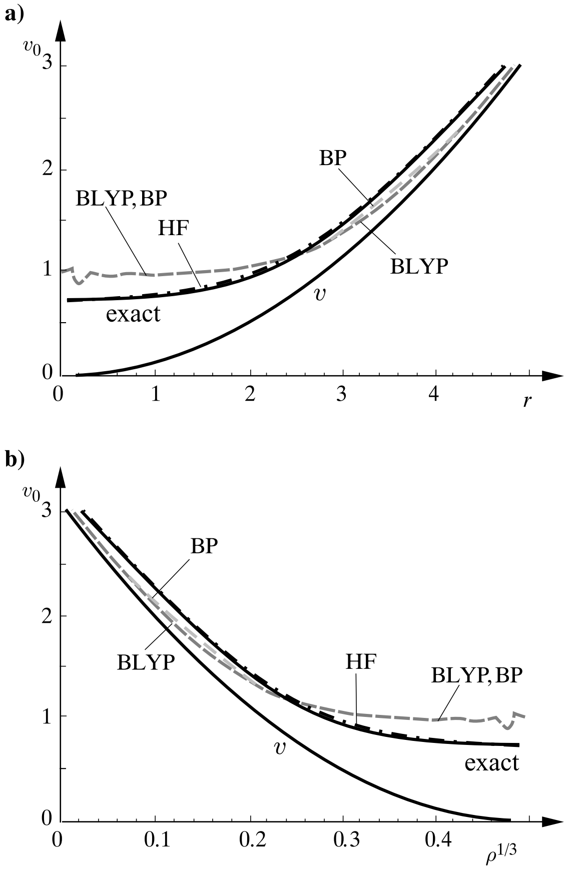


Fig. 11.8. Efficiency analysis of various DFT methods and comparison with the exact theory for the harmonium (with force constant $k = \frac{1}{4}$) according to Kais et al. Fig. (a) shows one-electron effective potential $v_0 = v + v_{\text{coul}} + v_{\text{xc}}$, with external potential $v = \frac{1}{2}kr^2$. Fig. (b) presents the same quantities as functions of $\rho^{\frac{1}{3}}$. The solid line corresponds to the exact results. The symbol HF pertains to the Fock potential (for the harmonic helium atom), the symbols BLYP and BP stand for two popular DFT methods. Reused with permission from S. Kais, D.R. Herschbach, N.C. Handy, C.W. Murray, and G.J. Laming, *J. Chem. Phys.* 99 (1993) 417, © 1993, American Institute of Physics.

the electron repulsion is effectively included. As we can see, the best approximate potential is the Hartree–Fock.

Exchange potential

> As for the exchange potential v_x (Fig. 11.9, it has to be negative and indeed it is), it turns out to correspond to forces 10–20 *times larger* than those typical for correlation potential v_c (just look at the corresponding slopes). This is an important message, because, as the reader may remember, at the very end we tried to push the dragon into the correlation hole and, as we see now, we have succeeded, the dragon turned out to be quite a small beast.

How are the BLYP and BP exchange potential doing? Their plots are very close to each other and go almost parallel to the exact exchange potential for most values of r, i.e. they are very good (any additive constant does not count). For small r both DFT potentials undergo some strange vibration. This region (high density) is

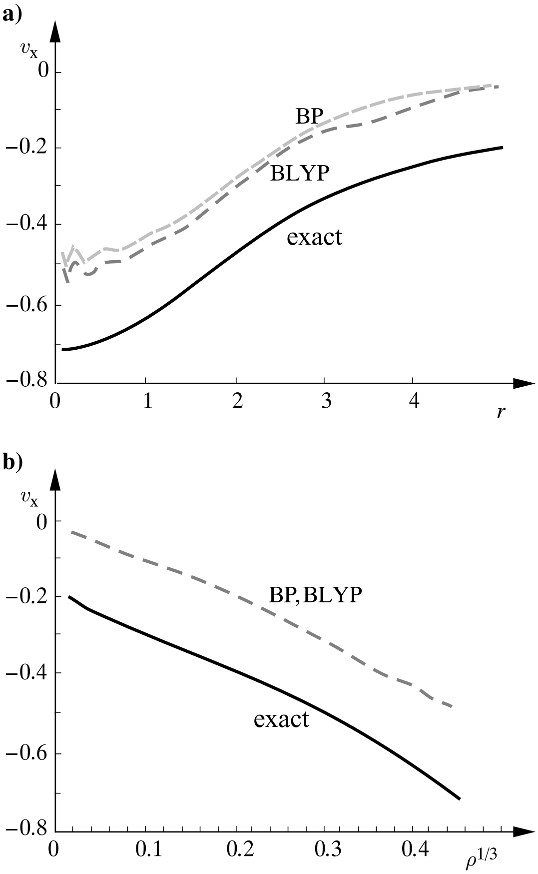


Fig. 11.9. Exchange potential. Efficiency analysis of various DFT methods and comparison with the exact theory for the harmonium (with the force constant $k = \frac{1}{4}$) according to Kais et al. Fig. (a) shows exchange potential v_x as a function of the radius r and Fig. (b) as a function of the density distribution ρ. The notation of Fig. 11.8 is used. It is seen that both DFT potentials produce plots that differ by nearly a constant from the exact potential (it is, therefore, an almost exact potential). The two DFT methods exhibit some nonphysical oscillations for small r. Reused with permission from S. Kais, D.R. Herschbach, N.C. Handy, C.W. Murray, and G.J. Laming, *J. Chem. Phys.* 99 (1993) 417, © 1993, American Institute of Physics.

surely the most difficult to describe, and no wonder that simple formulae cannot accurately describe the exact electronic density distribution.

Correlation potential

The correlation potential E_c is more intriguing (Fig. 11.10). The exact potential represents a smooth "hook-like" curve. The BLYP and BP correlation plots twine loosely like eels round about the exact curve,[60] and for small r exhibit some vibration similar to that for v_x. It is most impressive that the BLYP and BP curves twine as if they were in counter-phase, which suggests that, if added, they might produce good[61] results.

Conclusion

The harmonic helium atom represents an instructive example that pertains to medium electronic densities. It seems that the dragon of the correlation energy

---

[60] The deviations are very large.

[61] Such temptations give birth to Babylon-type science.

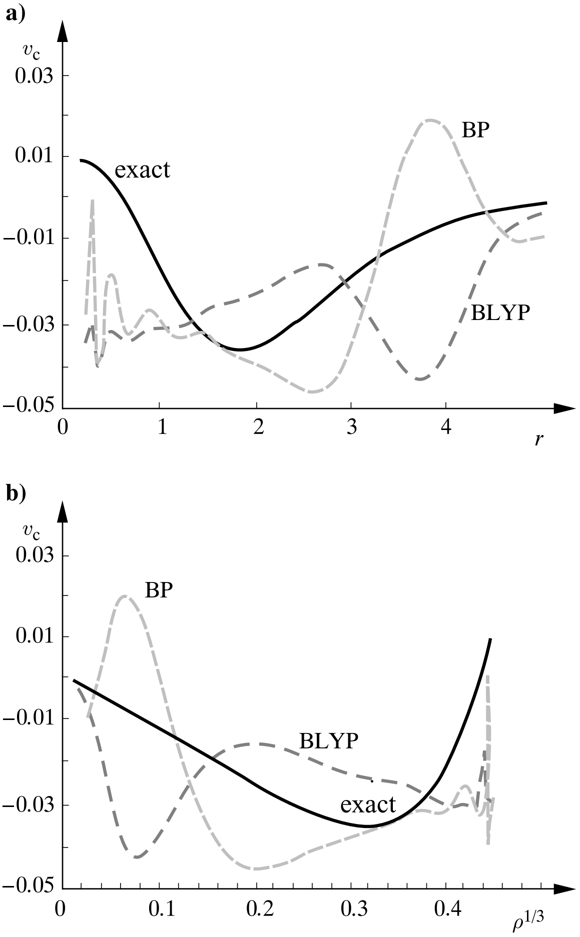


Fig. 11.10. Correlation potential – efficiency analysis of various DFT methods and comparison with the exact theory for the harmonic helium atom (with the force constant $k = \frac{1}{4}$) according to Kais et al. Fig. (a) shows correlation potential v_c (less important than the exchange potential) as a function of the radius r (a) and of density distribution ρ (b). Notation as in Fig. 11.8. The DFT potentials produce plots that differ widely from the exact correlation potential. Reused with permission from S. Kais, D.R. Herschbach, N.C. Handy, C.W. Murray, and G.J. Laming, *J. Chem. Phys.* 99 (1993) 417, © 1993, American Institute of Physics.

does not have a hundred heads and is quite mild (which is good), though a little bit unpredictable.

The results of various DFT versions are generally quite good, although this comes from a cancellation of errors. Nevertheless, great progress has been made. At present many chemists prefer the DFT method (economy and accuracy) than to getting stuck at the barrier of the configuration interaction excitations. And yet the method can hardly be called *ab initio*, since the exchange–correlation potential is tailored in a somewhat blind manner.

Summary

- The main theoretical concept of the DFT method is the electronic density distribution

$$\rho(\boldsymbol{r}) = N \sum_{\sigma_1=\frac{1}{2}}^{-\frac{1}{2}} \int \mathrm{d}\tau_2\,\mathrm{d}\tau_3 \ldots \mathrm{d}\tau_N \left|\Psi(\boldsymbol{r}, \sigma_1, \boldsymbol{r}_2, \sigma_2, \ldots, \boldsymbol{r}_n, \sigma_N)\right|^2,$$

where $\boldsymbol{r}$ indicates a point in 3D space, and the sum is over all the spin coordinates of N electrons, while the integration is over the space coordinates of $N-1$ electrons. For example, within the molecular orbital (RHF) approximation $\rho=\sum_i 2|\varphi_i(\boldsymbol{r})|^2$ is the sum of the squares of all the molecular orbitals multiplied by their occupation number. The electronic density distribution ρ is a function of position in the 3D space.

- ρ carries a lot of information. The density ρ exhibits maxima at nuclei (with a discontinuity of the gradient, because of the cusp condition, p. 504). The Bader analysis is based on identification of the critical (stationary) points of ρ (i.e. those for which $\nabla\rho=\mathbf{0}$), for each of them the Hessian is computed (the second derivatives matrix). Diagonalization of the Hessian tells us whether the critical point corresponds to a maximum of ρ (non-nuclear attractor[62]), a minimum (e.g., cavities), a first-order saddle point (e.g., a ring centre), or a second-order saddle point (chemical bond).
- The DFT relies on the two Hohenberg–Kohn theorems:
  - The *ground-state electronic density distribution* (ρ_0) *contains the same information as the ground-state wave function* (Ψ_0). Therefore, instead of a complex mathematical object (the ground-state wave function Ψ_0) depending on $4N$-variables we have a much simpler object (ρ_0) that depends on three variables (Cartesian coordinates) only.
  - A total energy functional of ρ *exists* that attains its minimum at $\rho=\rho_0$. This mysterious functional is not yet known.
- Kohn and Sham presented the concept of a system with *non-interacting electrons*, subject however to some "wonder" external field $v_0(\boldsymbol{r})$ (instead of that of the nuclei), such that the resulting *density ρ remains identical to the exact ground-state density distribution* ρ_0. This *fictitious* system of electrons plays a very important role in the DFT.
- Since the Kohn–Sham electrons do not interact, its wave function represents a single Slater determinant (known as the Kohn–Sham determinant).
- We write the total energy expression $E=T_0+\int v(\boldsymbol{r})\rho(\boldsymbol{r})\,d^3\boldsymbol{r}+J[\rho]+E_{\text{xc}}[\rho]$ that contains:
  - the kinetic energy of the *non-interacting* electrons (T_0),
  - the potential energy of the electron–nuclei interaction ($\int v(\boldsymbol{r})\rho(\boldsymbol{r})\,d^3\boldsymbol{r}$),
  - the Coulombic electron–electron self-interaction energy ($J[\rho]$),
  - the remainder E_{xc}, i.e. the unknown exchange–correlation energy.
- Using the single-determinant Kohn–Sham wave function (which gives the exact ρ_0) we vary the Kohn–Sham spinorbitals in order to find the minimum of the energy E.
- We are immediately confronted with the problem of how to find the unknown exchange–correlation energy E_{xc}, which is replaced also by an unknown exchange–correlation potential in the form of a functional derivative $v_{\text{xc}}\equiv\frac{\delta E_{\text{xc}}}{\delta\rho}$. We obtain the Kohn–Sham equation (resembling the Fock equation) $\{-\frac{1}{2}\Delta+v_0\}\phi_i=\varepsilon_i\phi_i$, where "wonder-potential" $v_0=v+v_{\text{coul}}+v_{\text{xc}}$, v_{coul} stands for the sum of the usual Coulombic operators (as in the Hartree–Fock method)[63] (built from the Kohn–Sham spinorbitals) and v_{xc} is the potential to be found.
- The main problem now resides in the nature of E_{xc} (and v_{xc}). We are forced to make a variety of practical guesses here.
- The simplest guess is the local density approximation (LDA). We assume that E_{xc} can be summed up from the contributions of all the points in space, and that the individual

---

[62] The maxima on the nuclei are excluded from the analysis, because of the discontinuity of $\nabla\rho$ mentioned above.

[63] It is, in fact, $\frac{\delta J[\rho]}{\delta\rho}$.

contribution depends only on ρ computed at this point. Now, the key question is *what does this dependence* $E_{\mathrm{xc}}[\rho]$ *look like*? The LDA answers this question by using the following approximation: each point $\boldsymbol{r}$ in the 3D space contributes to E_{xc} depending on the computed value of $\rho(\boldsymbol{r})$ *as if it were a homogeneous gas of uniform density* ρ, where the dependence $E_{\mathrm{xc}}[\rho]$ is *exactly* known.

- There are also more complex $E_{\mathrm{xc}}[\rho]$ functionals that go beyond the local approximation. They not only use the local value of ρ but sometimes also $\nabla\rho$ (gradient approximation).
- In each of these choices there is a lot of ambiguity. This, however, is restricted by some physical requirements.
- The requirements are related to the *electron pair distribution function*

$$\Pi(\boldsymbol{r}_1, \boldsymbol{r}_2) = N(N-1) \sum_{\text{all } \sigma_i} \int |\Psi|^2 \, \mathrm{d}^3\boldsymbol{r}_3 \, \mathrm{d}^3\boldsymbol{r}_4 \ldots \mathrm{d}^3\boldsymbol{r}_N,$$

which takes account of the fact that the two electrons, shown by $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$, avoid each other.

- First-order perturbation theory leads to the exact expression for the total energy E as

$$E = T_0 + \int \rho(\boldsymbol{r}) v(\boldsymbol{r}) \, \mathrm{d}^3\boldsymbol{r} + \frac{1}{2} \iint \mathrm{d}^3\boldsymbol{r}_1 \, \mathrm{d}^3\boldsymbol{r}_2 \, \frac{\Pi_{\mathrm{aver}}(\boldsymbol{r}_1, \boldsymbol{r}_2)}{r_{12}},$$

where

$$\Pi_{\mathrm{aver}}(\boldsymbol{r}_1, \boldsymbol{r}_2) = \int_0^1 \Pi_\lambda(\boldsymbol{r}_1, \boldsymbol{r}_2) \, \mathrm{d}\lambda,$$

with the parameter $0 \leqslant \lambda \leqslant 1$ instrumental when transforming the system of *non-interacting* electrons ($\lambda = 0$, Kohn–Sham model) into the system of *fully interacting* ones ($\lambda = 1$) *all the time preserving the exact density distribution* ρ. Unfortunately, the function $\Pi_\lambda(\boldsymbol{r}_1, \boldsymbol{r}_2)$ remains unknown.

- The function $\Pi_\lambda(\boldsymbol{r}_1, \boldsymbol{r}_2)$ serves to define the electron hole functions, which will tell us where electron 2 prefers to be, if electron 1 occupies the position $\boldsymbol{r}_1$. The exchange–correlation energy is related to the $\Pi_{\mathrm{aver}}^{\sigma\sigma'}$ function by:

$$E_{\mathrm{xc}} = \frac{1}{2} \sum_{\sigma\sigma'} \iint \mathrm{d}^3\boldsymbol{r}_1 \, \mathrm{d}^3\boldsymbol{r}_2 \, \frac{\Pi_{\mathrm{aver}}^{\sigma\sigma'}(\boldsymbol{r}_1, \boldsymbol{r}_2) - \rho_\sigma(\boldsymbol{r}_1)\rho_{\sigma'}(\boldsymbol{r}_2)}{r_{12}},$$

where the sum is over the spin coordinate σ of electron 1 and spin coordinate σ' of electron 2, with the decomposition $\Pi_{\mathrm{aver}} = \Pi_{\mathrm{aver}}^{\alpha\alpha} + \Pi_{\mathrm{aver}}^{\alpha\beta} + \Pi_{\mathrm{aver}}^{\beta\alpha} + \Pi_{\mathrm{aver}}^{\beta\beta}$. For example, the number $\Pi_{\mathrm{aver}}^{\alpha\beta} \, \mathrm{d}V_1 \, \mathrm{d}V_2$ stands for the probability of finding simultaneously an electron with the spin function α in the volume $\mathrm{d}V_1$ located at $\boldsymbol{r}_1$ and another electron with the spin function β in the volume $\mathrm{d}V_2$ located at $\boldsymbol{r}_2$, etc.

- The definition of the exchange–correlation hole function $h_{\mathrm{xc}}^{\sigma\sigma'}(\boldsymbol{r}_1, \boldsymbol{r}_2)$ is as follows:

$$E_{\mathrm{xc}} = \frac{1}{2} \sum_{\sigma\sigma'} \int \mathrm{d}^3\boldsymbol{r}_1 \int \mathrm{d}^3\boldsymbol{r}_2 \, \frac{\rho_\sigma(\boldsymbol{r}_1)}{r_{12}} h_{\mathrm{xc}}^{\sigma\sigma'}(\boldsymbol{r}_1, \boldsymbol{r}_2),$$

which is equivalent to setting

$$h_{\mathrm{xc}}^{\sigma\sigma'}(\boldsymbol{r}_1, \boldsymbol{r}_2) = \frac{\Pi_{\mathrm{aver}}^{\sigma\sigma'}(\boldsymbol{r}_1, \boldsymbol{r}_2)}{\rho_\sigma(\boldsymbol{r}_1)} - \rho_{\sigma'}(\boldsymbol{r}_2).$$

This means that the hole function is related to that part of the pair distribution function that indicates the *avoidance of the two electrons* (i.e. beyond their independent motion described by the *product* of the densities $\rho_\sigma(\boldsymbol{r}_1)\rho_{\sigma'}(\boldsymbol{r}_2)$).

- Due to the antisymmetry requirement for the wave function (Chapter 1) the holes have to satisfy some general (integral) conditions. The electrons with parallel spins have to avoid each other:
$$\int h_{\mathrm{xc}}^{\alpha\alpha}(\boldsymbol{r}_1,\boldsymbol{r}_2)\,\mathrm{d}^3\boldsymbol{r}_2=\int h_{\mathrm{xc}}^{\beta\beta}(\boldsymbol{r}_1,\boldsymbol{r}_2)\,\mathrm{d}^3\boldsymbol{r}_2=-1$$
(one electron disappears from the neighbourhood of the other), while the electrons with opposite spins are not influenced by the Pauli exclusion principle:
$$\int h_{\mathrm{xc}}^{\alpha\beta}(\boldsymbol{r}_1,\boldsymbol{r}_2)\,\mathrm{d}^3\boldsymbol{r}_2=\int h_{\mathrm{xc}}^{\beta\alpha}(\boldsymbol{r}_1,\boldsymbol{r}_2)\,\mathrm{d}^3\boldsymbol{r}_2=0.$$
- The exchange correlation hole is a sum of the exchange hole and the correlation hole: $h_{\mathrm{xc}}^{\sigma\sigma'}=h_{\mathrm{x}}^{\sigma\sigma'}+h_{\mathrm{c}}^{\sigma\sigma'}$, where the exchange hole follows in a simple way from the Kohn–Sham determinant (and is therefore supposed to be known). Then, we have to guess the correlation holes. *All the correlation* holes have to satisfy the condition $\int h_{\mathrm{c}}^{\sigma\sigma'}(\boldsymbol{r}_1,\boldsymbol{r}_2)\,\mathrm{d}^3\boldsymbol{r}_2=0$, which only means that the average has to be zero, but says nothing about the particular form of $h_{\mathrm{c}}^{\sigma\sigma'}(\boldsymbol{r}_1,\boldsymbol{r}_2)$. The only sure thing is, e.g., that close to the origin the function $h_{\mathrm{c}}^{\sigma\sigma'}$ has to be negative, and, therefore, for longer distances it has to be positive.
- The popular approximations, e.g., LDA, PW91, in general, satisfy the integral conditions for the holes.
- The hybrid approximations (e.g., B3LYP), i.e. such a linear combination of the potentials that it will ensure good agreement with experiment, become more and more popular.
- The DFT models can be tested when applied to exactly solvable problems with electronic correlation (like the harmonium, Chapter 4). It turns out that despite the exchange, and especially correlation and DFT potentials deviating from the exact ones, the total energy is quite accurate.

Main concepts, new terms

electron gas (p. 567)
electronic density distribution (p. 569)
Bader analysis (p. 571)
critical points (p. 571)
non-nuclear attractor (p. 573)
catastrophe set (p. 575)
Hohenberg–Kohn functional (p. 580)
v-representability (p. 580)
Kohn–Sham system (p. 584)
self-interaction energy (p. 585)
exchange–correlation energy (p. 586)
exchange–correlation potential (p. 588)
spin polarization (p. 589)
local density approximation, LDA (p. 590)
gradient approximation, NLDA (GEA) (p. 591)
hybrid approximations, NLDA (p. 591)
electron pair distribution (p. 592)
quasi-static transition (p. 594)
exchange–correlation hole (p. 598)
exchange hole (p. 599)
correlation hole (p. 599)

From the research front

Computer technology has been revolutionary, not only because computers are fast. Much more important is that each programmer uses the full experience of his predecessors and easily “stands on the shoulders of giants”. The computer era has made an unprecedented transfer of the most advanced theoretical tools from the finest scientists to practically everybody. Experimentalists often investigate large molecules. If there is a method like DFT, which gives answers to their vital questions in a shorter time than the *ab initio* methods, they will not hesitate and choose the DFT, even if the method is notorious for failing to reproduce the intermolecular interactions correctly (especially the dispersion energy, see Chapter 13). Something like this has now happened. Nowadays the DFT procedure is applicable to systems with hundreds of atoms.

The DFT method is developing fast also in the conceptual sense,[64] e.g., the theory of reactivity ("charge sensitivity analysis"[65]) has been derived, which established a link between the intermolecular electron transfer and the charge density changes in atomic resolution. For systems in magnetic fields, current DFT was developed.[66] Relativistic effects[67] and time dependent phenomena[68] are included in some versions of the theory.

Ad futurum...

The DFT will of course be further elaborated. There are already investigations under way, which will allow us to calculate the dispersion energy.[69] The impetus will probably be directed towards such methods as the Density Matrix Functional Theory (DMFT) proposed by Levy,[70] and currently being developed by Jerzy Ciosłowski.[71] The idea is to abandon $\rho(\boldsymbol{r})$ as the central quantity, and instead use the one-particle density matrix $\Gamma(\boldsymbol{r}, \boldsymbol{r}')$

$$\Gamma(\boldsymbol{r}, \boldsymbol{r}') = N \sum_{\sigma_1=\frac{1}{2}}^{\frac{1}{2}} \int d\tau_2\, d\tau_3 \ldots d\tau_N\, \Psi(\boldsymbol{r}, \sigma_1, \boldsymbol{r}_2, \sigma_2, \ldots, \boldsymbol{r}_N, \sigma_N) \times \Psi^*(\boldsymbol{r}', \sigma_1, \boldsymbol{r}_2, \sigma_2, \ldots, \boldsymbol{r}_N, \sigma_N), \tag{11.78}$$

in which the coordinates for electron 1 (integration pertains to electrons $2, 3, \ldots, N$) are different in Ψ^* and Ψ. We see that the diagonal element $\Gamma(\boldsymbol{r}, \boldsymbol{r})$ of $\Gamma(\boldsymbol{r}, \boldsymbol{r}')$ is simply $\rho(\boldsymbol{r})$. The method has the advantage that we are not forced to introduce the non-interacting Kohn–Sham electrons, because the mean value of the electron kinetic energy may be expressed

---

[64] See, e.g., P. Geerlings, F. De Proft, W. Langenaeker, *Chem. Rev.* 103 (2003) 1793.

[65] R.F. Nalewajski, J. Korchowiec, "*Charge Sensitivity Approach to Electronic Structure and Chemical Reactivity*", World Scientific, Singapore, 1997; R.F. Nalewajski, J. Korchowiec, A. Michalak, "*Reactivity Criteria in Charge Sensitivity Analysis*", *Topics in Current Chemistry* 183 (1996) 25; R.F. Nalewajski, "*Charge Sensitivities of Molecules and Their Fragments*", *Rev. Mod. Quant. Chem.*, ed. K.D. Sen, World Scientific, Singapore (2002) 1071; R.F. Nalewajski, R.G. Parr, *Proc. Natl. Acad. Sci. USA* 97 (2000) 8879.

[66] G. Vignale, M. Rasolt, *Phys. Rev. Letters* 59 (1987) 2360, *Phys. Rev. B* 37 (1988) 10685.

[67] A.K. Rajagopal, J. Callaway, *Phys. Rev. B* 7 (1973) 1912; A.H. Mac Donald, S.H. Vosko, *J. Phys. C* 12 (1979) 2977.

[68] E. Runge, E.K.U. Gross, *Phys. Rev. Letters* 52 (1984) 997, R. van Leeuwen, *Phys. Rev. Letters* 82 (1999) 3863.

[69] W. Kohn, Y. Meir, D. Makarov, *Phys. Rev. Letters* 80 (1998) 4153; E. Hult, H. Rydberg, B.I. Lundqvist, D.C. Langreth, *Phys. Rev. B* 59 (1999) 4708; J. Ciosłowski, K. Pernal, *J. Chem. Phys.* 116 (2002) 4802.

[70] M. Levy, *Proc. Nat. Acad. Sci.(USA)* 76 (1979) 6062.

[71] J. Ciosłowski, K. Pernal, *J. Chem. Phys.* 111 (1999) 3396; J. Ciosłowski, K. Pernal, *Phys. Rev. A* 61 (2000) 34503; J. Ciosłowski, P. Ziesche, K. Pernal, *Phys. Rev. B* 63 (2001) 205105; J. Ciosłowski, K. Pernal, *J. Chem. Phys*. 115 (2001) 5784; J. Ciosłowski, P. Ziesche, K. Pernal, *J. Chem. Phys.* 115 (2001) 8725.

directly by the new quantity (this follows from the definition):

$$T = -\frac{1}{2}\int d^3\boldsymbol{r}\,[\Delta_{\boldsymbol{r}}\Gamma(\boldsymbol{r},\boldsymbol{r}')]|_{\boldsymbol{r}'=\boldsymbol{r}},$$

where the symbol $|_{\boldsymbol{r}'=\boldsymbol{r}}$ means replacing $\boldsymbol{r}'$ by $\boldsymbol{r}$ after the result $\Delta_{\boldsymbol{r}}\Gamma(\boldsymbol{r},\boldsymbol{r}')$ is ready. Thus, in the DMFT exchange–correlation we have no kinetic energy left.

The success of the DFT approach will probably make the traditional *ab initio* procedures faster, up to the development of methods with linear scaling (with the number of electrons for long molecules). The massively parallel "computer farms" with 2000 processors currently (and millions expected soon), will saturate most demands of experimental chemistry. The results will be calculated fast and it will be much more difficult to define an interesting target to compute. We will be efficient.

We will have an efficient hybrid potential, say, of the B3LYP5PW2001/2002-type. There remains, however, a problem that already appears in laboratories. A colleague delivers a lecture and proposes a hybrid B3LYP6PW2003update,[72] which is more effective for aromatic molecules. What will these two scientists talk about? It is very good that the computer understands all this, but what about the scientists? In my opinion science will move into such areas as planning new materials and new molecular phenomena (cf. Chapter 15) with the programs mentioned above as tools.

Additional literature

A.D. Becke, in "Modern Electronic Structure Theory. Part II" , D.R. Yarkony, ed., World Scientific, p. 1022.

An excellent and comprehensible introduction into DFT written by a renowned expert in the field.

J. Andzelm, E. Wimmer, *J. Chem. Phys.* 96 (1992) 1280.

A competent presentation of DFT technique introduced by the authors.

Richard F.W. Bader, "Atoms in Molecules. A Quantum Theory", Clarendon Press, Oxford, 1994.

An excellent book.

E.J. Baerends, O.V. Gritsenko, "A Quantum Chemical View of Density Functional Theory", *J. Phys. Chem.* A101 (1997) 5383.

A very well written article.

Other sources:

R.G. Parr, W. Yang, "Density Functional Theory of Atoms and Molecules", Oxford Univ. Press, Oxford, 1989.

"Density Functional Theory of Many Fermion Systems", ed. S.B. Trickey, Academic Press, New York, 1990.

R.H. Dreizler, E.K.U. Gross, "Density Functional Theory", Springer, Berlin, 1990.

"Density Functional Theory", ed. E.K.U. Gross, R.H. Dreizler, Plenum, New York, 1994.

---

[72] The same pertains to the traditional methods. Somebody operating billions of the expansion functions meets a colleague using even more functions. It would be very pity if we changed into experts ("*this is what we are paid for...*") knowing, which particular BLYP is good for calculating interatomic distances, which for charge distribution, etc.

N.H. March, “Electron Density Theory of Atoms and Molecules”, Academic Press, London, 1992.

“DFT I,II,III,IV” Topics in Current Chemistry, vols. 180–183, ed. R. Nalewajski, Springer, Berlin, 1996.

“Density Functionals: Theory and Applications”, ed. D. Joubert, *Lecture Notes in Physics*, vol. 500, Springer, Berlin, 1998.

A. Freeman, E. Wimmer, “DFT as a major tool in computational materials science”, *Ann. Rev. Mater. Sci*. 25 (1995) 7.

W. Kohn, A.D. Becke, R.G. Parr, “DFT of electronic structure”, *J. Chem. Phys.* 100 (1996) 12974.

A. Nagy, “Density Functional Theory and Applications to Atoms and Molecules”, *Phys. Reports* 298 (1998) 1.

Questions

1. The Hessian of the electronic density distribution computed for the critical point within a covalent chemical bond has:
   a) exactly one negative eigenvalue; b) the number of eigenvalues equal to the number of electrons in the bond; c) exactly one positive eigenvalue; d) has two positive eigenvalues.

2. Hohenberg and Kohn (ρ, ρ_0, E, E_0 stand for the density distribution, the ground-state density distribution, the mean value of the Hamiltonian, and the ground-state energy, respectively):
   a) have proposed a functional $E[\rho]$ that exhibits minimum $E[\rho_0] = E_0$;
   b) have proved that a functional $E[\rho]$ exists that satisfies $E[\rho] \geqslant E[\rho_0] = E_0$;
   c) have proved that an energy functional $E[\rho] \geqslant 0$;
   d) have proved that a total energy functional $E[\rho] > E_0$.

3. The Kohn–Sham system represents:
   a) any set of non-interacting N electrons;
   b) a set of N non-interacting electrons subject to an external potential that preserves the exact density distribution ρ of the system;
   c) a set of N electrons interacting among themselves in such a way that preserves the exact density distribution ρ of the system under consideration;
   d) a set of N paired electrons that satisfies $\rho_\alpha = \rho/2$.

4. In the LDA (E_{xc} stands for the exchange–correlation energy):
   a) the $E_{\mathrm{xc}}[\rho]$ for molecules is computed as a sum of local contributions as if they came from a homogeneous electronic gas of density ρ;
   b) $E_{\mathrm{xc}} = \int \frac{1}{r_{12}} \rho(\boldsymbol{r}_1)\rho(\boldsymbol{r}_2)$, where ρ corresponds to the electronic homogeneous gas density distribution;
   c) E_{xc} is neglected;
   d) $E_{\mathrm{xc}}[\rho(\boldsymbol{r})]$ is calculated by multiplying ρ by a constant.

5. In the DFT hybrid approximations (E_{xc} stands for the exchange–correlation energy):
   a) the Kohn–Sham orbitals represent the hybrid atomic orbitals described in Chapter 8;
   b) $E_{\mathrm{xc}} = \frac{1}{2} \int d^3\boldsymbol{r}_1\, d^3\boldsymbol{r}_2 \frac{\Pi(\boldsymbol{r}_1, \boldsymbol{r}_2)}{r_{12}}$;

c) E_{xc} is identical to the exchange energy corresponding to the Kohn–Sham determinant built from hybrid orbitals;
d) as E_{xc} we use a linear combination of the expressions for E_{xc} from several different DFT approximations.

6. The electron pair distribution function $\Pi_\lambda(\mathbf{r}_1, \mathbf{r}_2)$:
   a) pertains to two electrons each of charge $\sqrt{\lambda}$ and with ρ that equals the exact electron density distribution; b) satisfies $\int \Pi_\lambda(\mathbf{r}_1, \mathbf{r}_2)\, d\lambda = 1$; c) is the correlation energy per electron pair for $\lambda \in [0, 1]$; d) is the pair distribution function for the electrons with charge λ.

7. In the DFT the exchange–correlation energy E_{xc}: a) contains a part of the electronic kinetic energy; b) contains the total electronic kinetic energy; c) contains only that part of the electronic kinetic energy that corresponds to non-interacting electrons; d) does not contain any electronic kinetic energy.

8. The exchange–correlation holes $h_{\mathrm{xc}}^{\alpha\beta}(\mathbf{r}_1, \mathbf{r}_2)$ and $h_{\mathrm{xc}}^{\beta\beta}(\mathbf{r}_1, \mathbf{r}_2)$ satisfy:
   a) $h_{\mathrm{xc}}^{\alpha\beta}(\mathbf{r}_1, \mathbf{r}_2)\, \mathrm{d}^3\mathbf{r}_2 = 0$ and $h_{\mathrm{xc}}^{\beta\beta}(\mathbf{r}_1, \mathbf{r}_2)\, \mathrm{d}^3\mathbf{r}_2 = -2$;
   b) $\int h_{\mathrm{xc}}^{\alpha\beta}(\mathbf{r}_1, \mathbf{r}_2)\, \mathrm{d}^3\mathbf{r}_2 = -1$ and $\int h_{\mathrm{xc}}^{\beta\beta}(\mathbf{r}_1, \mathbf{r}_2)\, \mathrm{d}^3\mathbf{r}_2 = 0$;
   c) $\int h_{\mathrm{xc}}^{\alpha\beta}(\mathbf{r}_1, \mathbf{r}_2)\, \mathrm{d}^3\mathbf{r}_2 = 0$ and $\int h_{\mathrm{xc}}^{\beta\beta}(\mathbf{r}_1, \mathbf{r}_2)\, \mathrm{d}^3\mathbf{r}_2 = -1$;
   d) $\int h_{\mathrm{xc}}^{\alpha\beta}(\mathbf{r}_1, \mathbf{r}_2)\, \mathrm{d}^3\mathbf{r}_2 = \int h_{\mathrm{xc}}^{\beta\beta}(\mathbf{r}_1, \mathbf{r}_2)\, \mathrm{d}^3\mathbf{r}_2 = 0$.

9. The DFT exchange energy E_x:
   a) $E_x > 0$; b) turns out to be more important than the correlation energy; c) is identical to the Hartree–Fock energy; d) represents a repulsion.

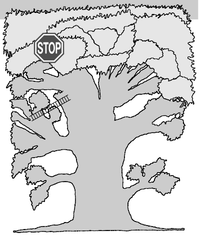
10. The DFT:
    a) describes the argon–argon equilibrium distance correctly;
    b) is roughly as time-consuming as the CI procedure;
    c) cannot take into account any electronic correlation since it uses a single Kohn–Sham determinant;
    d) is incorrect when describing the dispersion interaction of two water molecules.

Answers

1c, 2b, 3b, 4a, 5d, 6a, 7a, 8c, 9b, 10d

Chapter 12

THE MOLECULE IN AN ELECTRIC OR MAGNETIC FIELD


WHERE ARE WE?

Where are we?

We are already in the crown of the TREE (left-hand side)

An example

How does a molecule react to an applied electric field? How do you calculate the changes it undergoes? In some materials there is a strange phenomenon: a monochromatic *red* laser light beam enters a transparent substance, and leaves the specimen as a *blue* beam. Why?

Another example, this time with a magnetic field. We apply a long wavelength electromagnetic radiation to a specimen. We do not see any absorption whatsoever. However, if, in addition, we apply a static magnetic field gradually increasing in intensity, at some intensities we observe absorption. If we analyze the magnetic field values corresponding to the absorption then they cluster into mysterious groups that depend on the chemical composition of the specimen. Why?

What is it all about

The properties of a substance with and without an external electric field *differ*. The problem is how to compute the molecular properties in the electric field from the properties of the isolated molecule and the characteristics of the applied field. Molecules react also upon application of a magnetic field, which changes the internal electric currents and modifies the local magnetic field. A nucleus may be treated as a small magnet, which reacts to the local magnetic field it encounters. This local field depends not only on the external magnetic field, but also on those from other nuclei, and on the electronic structure in the vicinity. This produces some energy levels in the spin system, with transitions leading to the nuclear magnetic resonance (NMR) phenomenon which has wide applications in chemistry, physics and medicine.

The following topics will be described in the present chapter.

Why is this important?

There is no such a thing as an isolated molecule, since any molecule interacts with its neighbourhood. In most cases this is the electric field of another molecule or an external electric field and represents the only information about the external world the molecule has. The source of the electric field (another molecule or a technical equipment) is of no importance. *Any molecule will respond to the electric field, but some will respond dramatically, while others may respond quite weakly*. This is of importance in designing new materials.

The molecular electronic structure does not respond to a change in orientation of the nuclear magnetic moments, because the corresponding perturbation is too small. On the other hand, the molecular electronic structure influences the subtle energetics of interaction of the nuclear spin magnetic moments and these effects may be recorded in the NMR spectrum. This is of great practical importance, because it means *we have in the molecule un-*

der study a system of sounds (nuclear spins) which characterize the electronic structure almost without perturbing it.

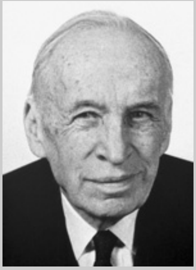
What is needed?

- Perturbation theory (Chapter 5, necessary).
- Variational method (Chapter 5, advised).
- Harmonic oscillator and rigid rotator (Chapter 4, advised).
- Breit Hamiltonian (Chapter 3, advised).
- Appendix S, p. 1015 (advised).
- Appendix G, p. 962 (necessary for magnetic properties).
- Appendix M, p. 986 (advised).
- Appendix W, p. 1032 (advised).

Classical works

Peter Debye, as early as 1921, predicted in "*Molekularkräfte und ihre Elektrische Deutung*", *Physikalische Zeitschrift*, 22 (1921) 302 that a non-polar gas or liquid of molecules with a non-zero quadrupole moment, when subject to an inhomogeneous electric field, will exhibit the birefringence phenomenon due to the orientation of the quadrupoles in the electric field gradient. ★ The book by John Hasbrouck Van Vleck "*Electric and Magnetic Susceptibilities*", Oxford University Press, 1932 represented enormous progress. ★ The theorem that forces acting on nuclei result from classical interactions with electron density (computed by a quantum mechanical method) was first proved by Hans Gustav Adolf Hellmann in the world's first textbook of quantum chemistry "*Einführung in die Quantenchemie*", Deuticke, Leipzig und Wien,[1] (1937), p. 285, and then, independently, by Richard Philips Feynman in "*Forces in Molecules*" published in *Physical Review*, 56 (1939) 340. ★ The first idea of nuclear magnetic resonance (NMR) came from a Dutch scholar, Cornelis Jacobus Gorter, in "*Negative Result in an Attempt to Detect Nuclear Spins*" in *Physica*, 3 (1936) 995. ★ The first *electron* paramagnetic resonance (EPR) measurement was carried out by Evgenii Zavoiski from Kazan University (USSR) and reported in "*Spin-Magnetic Resonance in Paramagnetics*" published in *Journal of Physics (USSR)*, 9 (1945) 245, 447. ★ The first NMR absorption experiment was performed by Edward M. Purcell, Henry C. Torrey and Robert V. Pound and published in "*Resonance Absorption by Nuclear Magnetic Moments in a Solid*", which appeared in *Physical Review*, 69 (1946) 37, while the first correct explanation of nuclear spin–spin coupling (through the chemical bond) was given by Norman F. Ramsey and Edward M. Purcell in "*Interactions between Nuclear Spins in Molecules*" published in *Physical Review*, 85 (1952) 143. ★ The first successful experiment in non-linear optics with frequency doubling was reported by Peter A. Franken, Alan E. Hill, Wilbur C. Peters and Gabriel Weinreich in "*Generation of Optical Harmonics*" published in *Physical Review Letters*, 7 (1961) 118. ★ Hendrik F. Hameka's book "*Advanced Quantum Chemistry. Theory of Interactions between Molecules and Electromagnetic Fields*" (1965) is also considered a classic work. ★ Although virtually unknown outside Poland, the book "*Mole-*

John Hasbrouck Van Vleck (1899–1980), American physicist, professor at the University of Minnesota, received the Nobel Prize in 1977 for "*fundamental theoretical investigations of the electronic structure of magnetic and disordered systems*".



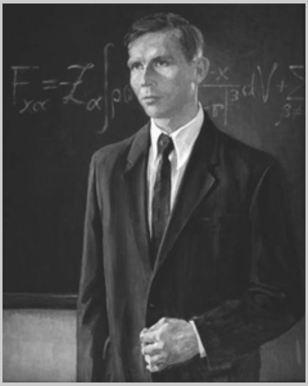
---

[1] A Russian edition had appeared a few months earlier, but it does not contain the theorem.

cular Non-Linear Optics", Warsaw–Poznań, PWN (1977) (in Polish) by Stanisław Kielich, deserves to be included in the list of classic works.

12.1 HELLMANN–FEYNMAN THEOREM

Let us assume that a system with Hamiltonian $\hat{H}$ is in a *stationary state* described by the (normalized) function ψ. Now let us begin to do a little "tinkering" with the Hamiltonian by introducing a parameter P. So we have $\hat{H}(P)$, and assume we may change the parameter smoothly. For example, as the parameter P we may take the electric field intensity, or, if we assume the Born–Oppenheimer approximation, then as P we may take a nuclear coordinate.[2] If we change P in the Hamiltonian $\hat{H}(P)$, then we have a response in the eigenvalue $E(P)$. The eigenfunctions and eigenvalues of $\hat{H}$ become functions of P.



Hans Gustav Adolf Hellmann (1903–1938), German physicist, one of the pioneers of quantum chemistry. He contributed to the theory of dielectric susceptibility, theory of spin, chemical bond theory (semiempirical calculations, also virial theorem and the role of kinetic energy), intermolecular interactions theory, electronic affinity, etc. Hellmann wrote the world's first textbook of quantum chemistry "*Vviedieniye v kvantovuyu khimiyu*", a few months later edited in Leipzig as "*Einführung in die Quantenchemie*". In 1933 Hellmann presented his habilitation thesis at the Veterinary College of Hannover. As part of the paper work he filled out a form, in which according to the recent Nazi requirement he wrote that his wife was of Jewish origin. The Nazi ministry rejected the habilitation. The situation grew more and more dangerous (many students of the School were active Nazis) and the Hellmanns decided to emigrate. Since his wife originated from the Ukraine they chose the Eastern route. Hellmann obtained a position at the Karpov Institute of Physical Chemistry in Moscow as a theoretical group leader. A leader of another group, the Communist Party First Secretary of the Institute (Hellmann's colleague and a coauthor of one of his papers) A.A. Zukhovitskyi as well as the former First Secretary, leader of the Heterogenic Catalysis Group Mikhail Tiomkin, denounced Hellmann to the institution later called the KGB, which soon arrested him. Years later an investigation protocol was found in the KGB archives, with a text about Hellmann's spying written by somebody else but with Hellmann's signature. This was a common result of such "investigations". On May 16, 1938 Albert Einstein, and on May 18 three other Nobel prize recipients: Irene Joliot-Curie, Frederick Joliot-Curie and Jean-Baptiste Perrin, asked Stalin for mercy for Hellmann. Stalin ignored the eminent scholars' supplication, and on May 29, 1938 Hans Hellmann faced the firing squad and was executed.

After W.H.E. Schwarz et al., *Bunsen-Magazin* (1999) 10, 60. Portrait reproduced from a painting by Tatiana Livschitz, courtesy of Professor Eugen Schwarz.

---

[2] Recall please that in the adiabatic approximation, the electronic Hamiltonian depends parametrically on the nuclear coordinates (Chapter 6). Then $E(P)$ corresponds to $E_k^0(R)$ from eq. (6.8).

Richard Philips Feynman (1919–1988), American physicist, for many years professor at the California Institute of Technology. His father was his first informal teacher of physics, who taught him the extremely important skill of independent thinking. Feynman studied at Massachusetts Institute of Technology, then in Princeton University, where he defended his Ph.D. thesis under the supervision of John Archibald Wheeler.

In 1945–1950 Feynman served as a professor at Cornell University. A paper plate thrown in the air by a student in the Cornell cafe was the first impulse for Feynman to think about creating a new version of quantum electrodynamics. For this achievement Feynman received the Nobel prize in 1965, cf. p. 14.

Feynman was a genius, who contributed to several branches of physics (superfluidity, weak interactions, quantum computers, nanotechnology). His textbook "*The Feynman Lectures on Physics*" is considered an unchallenged achievement in academic literature. Several of his books became best-sellers. Feynman was famous for his unconventional, straightforward and crystal-clear thinking, as well as for his courage and humour. Curiosity and courage made possible his investigations of the ancient Maya calendar, ant habits, as well as his activity in painting and music.

From John Slater's autobiography "*Solid State and Molecular Theory*", London, Wiley (1975):

"... The theorem known as the Hellmann–Feynman theorem, stating that the force on a nucleus can be rigorously calculated by electrostatics (...), remained, as far as I was concerned, only a surmise for several years. Somehow, I missed the fact that Hellmann, in Germany, proved it rigorously in 1936, and when a very bright undergraduate turned up in 1938–1939 wanting a topic for a bachelor's thesis, I suggested to him that he see if it could be proved. He come back very promptly with a proof. Since he was Richard Feynman (...), it is not surprizing that he produced his proof without trouble."

The Hellmann–Feynman theorem pertains to the rate of the change[3] of $E(P)$:

HELLMANN–FEYNMAN THEOREM:

$$\frac{\partial E}{\partial P} = \left\langle \psi \middle| \frac{\partial \hat{H}}{\partial P} \middle| \psi \right\rangle. \tag{12.1}$$

The proof is simple. The differentiation with respect to P of the integrand in $E = \langle \psi | H | \psi \rangle$ gives

$$\begin{aligned} \frac{\partial E}{\partial P} &= \left\langle \frac{\partial \psi}{\partial P} \middle| \hat{H}\psi \right\rangle + \left\langle \psi \middle| \frac{\partial \hat{H}}{\partial P}\psi \right\rangle + \left\langle \psi \middle| \hat{H}\frac{\partial \psi}{\partial P} \right\rangle \\ &= E\left(\left\langle \frac{\partial \psi}{\partial P} \middle| \psi \right\rangle + \left\langle \psi \middle| \frac{\partial \psi}{\partial P} \right\rangle \right) + \left\langle \psi \middle| \frac{\partial \hat{H}}{\partial P}\psi \right\rangle = \left\langle \psi \middle| \frac{\partial \hat{H}}{\partial P}\psi \right\rangle, \end{aligned} \tag{12.2}$$

because the expression in parentheses is equal to zero (we have profited from the

---

[3] We may define $(\frac{\partial \hat{H}}{\partial P})_{P=P_0}$ as an operator, being a limit when $P \to P_0$ of the operator sequence $\frac{\hat{H}(P)-\hat{H}(P_0)}{P-P_0}$.

facts that the $\hat{H}$ is Hermitian, and that ψ represents its eigenfunction[4]). Indeed, differentiating $\langle\psi|\psi\rangle = 1$ we have:

$$0 = \left\langle \frac{\partial\psi}{\partial P} \middle| \psi \right\rangle + \left\langle \psi \middle| \frac{\partial\psi}{\partial P} \right\rangle, \tag{12.3}$$

which completes the proof.

Soon we will use the Hellmann–Feynman theorem to compute the molecular response to an electric field.[5]

ELECTRIC PHENOMENA

12.2 THE MOLECULE IMMOBILIZED IN AN ELECTRIC FIELD

The electric field intensity $\boldsymbol{\mathcal{E}}$ at a point represents the force acting on a unit positive point charge (probe charge): $\boldsymbol{\mathcal{E}} = -\nabla V$, where V stands for the electric field potential energy at this point.[6] When the potential changes linearly in space

---

[4] If, instead of the exact eigenfunction, we use an approximate function ψ, then the theorem would have to be modified. In such a case we have to take into account the terms $\langle \frac{\partial\psi}{\partial P}|\hat{H}|\psi\rangle + \langle\psi|\hat{H}|\frac{\partial\psi}{\partial P}\rangle$.

[5] In case P is a nuclear coordinate (say, x coordinate of the nucleus C, denoted by X_C), and E stands for the potential energy for the motion of the nuclei (cf. Chapter 6, the quantity corresponds to E_0^0 of eq. (6.8)), the quantity $-\frac{\partial E}{\partial P} = F_{X_C}$ represents the x component of the force acting on the nucleus. The Helmann–Feynman theorem says that this component can be computed as the mean value of the derivative of the Hamiltonian with respect to the parameter P. Since the electronic Hamiltonian reads

$$\hat{H}_0 = -\frac{1}{2}\sum_i \Delta_i + V,$$

$$V = -\sum_A\sum_i \frac{Z_A}{r_{Ai}} + \sum_{i<j}\frac{1}{r_{ij}} + \sum_{A<B}\frac{Z_A Z_B}{R_{AB}},$$

then, after differentiating, we have as $\frac{\partial\hat{H}}{\partial P}$

$$\frac{\partial\hat{H}_0}{\partial X_C} = \sum_i \frac{Z_C}{(r_{Ci})^3}(X_C - x_i) - \sum_{B(\neq C)}\frac{Z_C Z_B}{(R_{BC})^3}(X_C - X_B).$$

Therefore,

$$F_{X_C} = -\left\langle \psi \middle| \frac{\partial\hat{H}}{\partial P} \middle| \psi \right\rangle = Z_C\left[\int \mathrm{d}V_1\, \rho(1)\frac{x_1 - X_C}{(r_{C1})^3} - \sum_{B(\neq C)}\frac{Z_B}{(R_{BC})^3}(X_B - X_C)\right],$$

where $\rho(1)$ stands for the electronic density defined in Chapter 11, eq. (11.1).

The last term can be easily calculated from the positions of the nuclei. The first term requires the calculation of the one-electron integrals. Note, that the resulting formula states that the forces acting on the nuclei follow from the *classical* Coulomb interaction involving the electronic density ρ, even if the electronic density has been (and has to be) computed from quantum mechanics.

[6] We see that two potential functions that differ by a constant will give the same forces, i.e. will describe identical physical phenomena (this is why this constant is arbitrary).

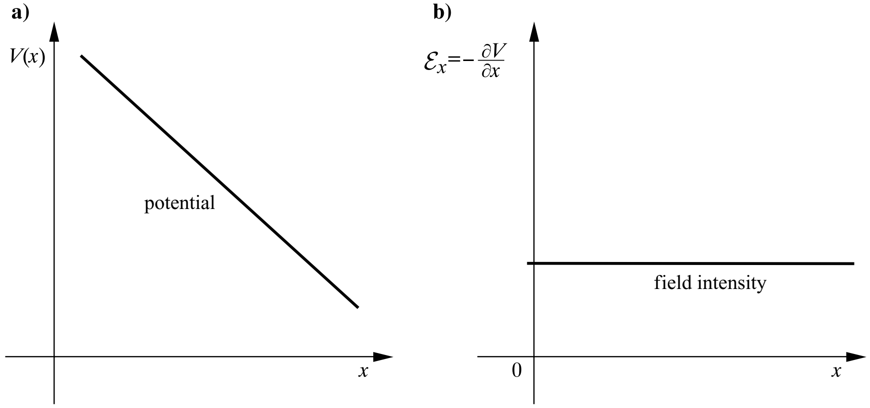


Fig. 12.1. Recalling the electric field properties. (a) 1D: the potential V decreases with x. This means that the electric field intensity $\boldsymbol{\mathcal{E}}$ is constant, i.e. the field is homogeneous (b) 3D; (c) homogeneous electric field $\boldsymbol{\mathcal{E}} = (\mathcal{E}, 0, 0)$; (d) inhomogeneous electric field $\boldsymbol{\mathcal{E}} = (\mathcal{E}(x), 0, 0)$; (e) inhomogeneous electric field $\boldsymbol{\mathcal{E}} = (\mathcal{E}_x(x, y), \mathcal{E}_y(x, y), 0)$.

(Fig. 12.1.a), the electric field intensity is constant (Fig. 12.1.b,c). If at such a potential we shift the probe charge from a to $x + a$, $x > 0$, then the energy will *lower* by $V(x + a) - V(a) = -\mathcal{E}x < 0$. This is similar to the lowering of the of potential energy of a stone as it slides downhill.

If, instead of a unit charge, we shift the charge Q, then the energy will change by $-\mathcal{E}Qx$.

It is seen that if we change the *direction* of the shift or the *sign* of the probe charge, then the energy will go *up* (in case of the stone we may change only the direction).

12.2.1 THE ELECTRIC FIELD AS A PERTURBATION

The inhomogeneous field at a slightly shifted point

Imagine a Cartesian coordinate system in 3D space and an inhomogeneous electric field (Fig. 12.1.d,e) in it $\boldsymbol{\mathcal{E}} = [\mathcal{E}_x(x, y, z), \mathcal{E}_y(x, y, z), \mathcal{E}_z(x, y, z)]$.

Assume the electric field vector $\boldsymbol{\mathcal{E}}(\boldsymbol{r}_0)$ is measured at a point indicated by the vector $\boldsymbol{r}_0$. What will we measure at a point shifted by a small vector $\boldsymbol{r} = (x, y, z)$ with respect to $\boldsymbol{r}_0$? The components of the electric field intensity represent smooth functions in space and this is why we may compute the electric field from the Taylor expansion (for each of the components $\mathcal{E}_x$, $\mathcal{E}_y$, $\mathcal{E}_z$ separately, all the derivatives are computed at point $\boldsymbol{r}_0$):

$$\mathcal{E}_x = \mathcal{E}_{x,0} + \left(\frac{\partial \mathcal{E}_x}{\partial x}\right)_0 x + \left(\frac{\partial \mathcal{E}_x}{\partial y}\right)_0 y + \left(\frac{\partial \mathcal{E}_x}{\partial z}\right)_0 z$$
$$+ \frac{1}{2}\left(\frac{\partial^2 \mathcal{E}_x}{\partial x^2}\right)_0 x^2 + \frac{1}{2}\left(\frac{\partial^2 \mathcal{E}_x}{\partial x \partial y}\right)_0 xy + \frac{1}{2}\left(\frac{\partial^2 \mathcal{E}_x}{\partial x \partial z}\right)_0 xz$$

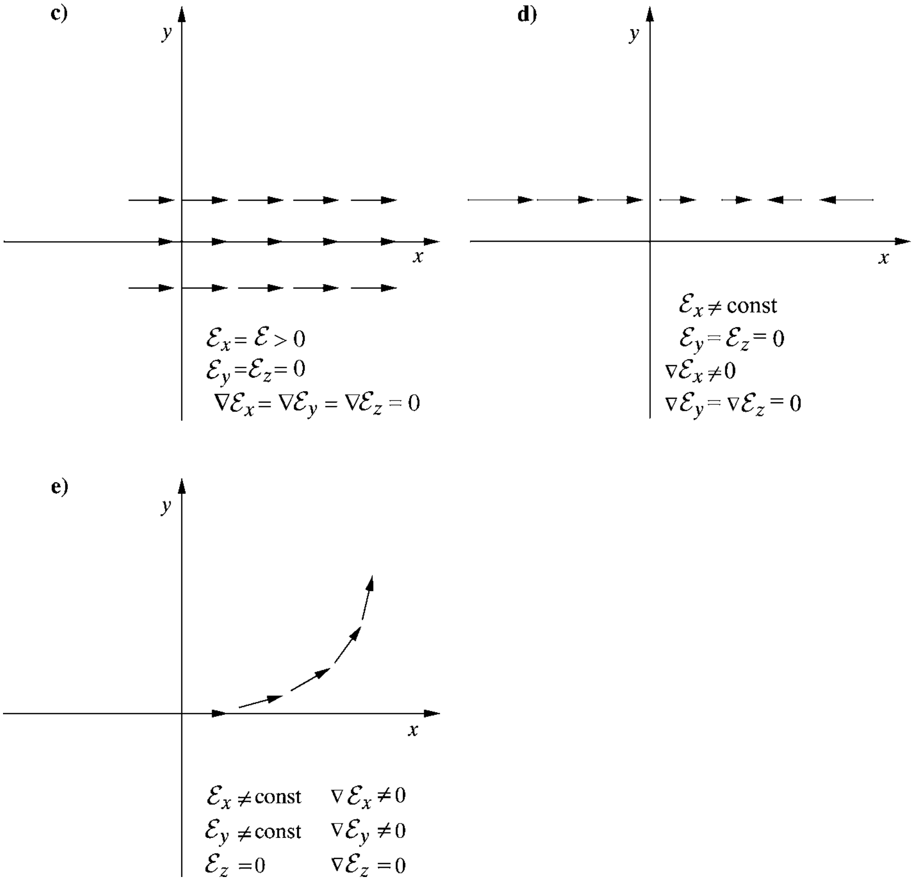


Fig. 12.1. *Continued.*

$$
\begin{aligned}
&+\frac{1}{2}\left(\frac{\partial^2\mathcal{E}_x}{\partial y\partial x}\right)_0 yx+\frac{1}{2}\left(\frac{\partial^2\mathcal{E}_x}{\partial y^2}\right)_0 y^2+\frac{1}{2}\left(\frac{\partial^2\mathcal{E}_x}{\partial y\partial z}\right)_0 yz\\
&+\frac{1}{2}\left(\frac{\partial^2\mathcal{E}_x}{\partial z\partial x}\right)_0 zx+\frac{1}{2}\left(\frac{\partial^2\mathcal{E}_x}{\partial z\partial y}\right)_0 zy+\frac{1}{2}\left(\frac{\partial^2\mathcal{E}_x}{\partial z^2}\right)_0 z^2+\cdots
\end{aligned}
$$

$\mathcal{E}_y = \ldots$ similarly
$\mathcal{E}_z = \ldots$ similarly (Fig. 12.2).

Energy gain due to a shift of the electric charge Q

These two electric field intensities (at points $\boldsymbol{r}_0$ and $\boldsymbol{r}_0 + \boldsymbol{r}$) have been calculated in order to consider the energy gain associated with the shift $\boldsymbol{r}$ of the electric point charge Q. Similar to the 1D case just considered, we have the energy gain $\Delta E = -Q\boldsymbol{\mathcal{E}} \cdot \boldsymbol{r}$. There is only one problem: *which* of the two electric field intensities is

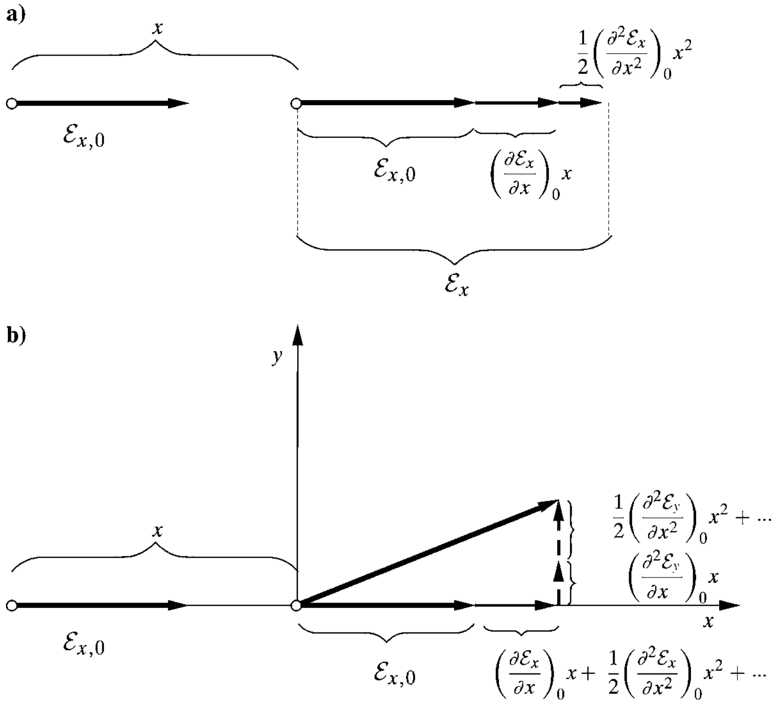


Fig. 12.2. The electric field computed at point $x \ll 1$ from its value (and the values of its derivatives) at point 0. (a) 1D case; (b) 2D case.

to be inserted into the formula? Since the vector $\boldsymbol{r} = \boldsymbol{i}x + \boldsymbol{j}y + \boldsymbol{k}z$ is small ($\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k}$ stand for unit vectors corresponding to axes x, y, z, respectively), we may insert, e.g., the mean value of $\boldsymbol{\mathcal{E}}(\boldsymbol{r}_0)$ and $\boldsymbol{\mathcal{E}}(\boldsymbol{r}_0 + \boldsymbol{r})$. We quickly get the following (indices $q, q', q'' \in \{x, y, z\}$):

$$\begin{aligned}
\Delta E &= -Q\boldsymbol{\mathcal{E}} \cdot \boldsymbol{r} = -Q\frac{1}{2}\big[\boldsymbol{\mathcal{E}}(\boldsymbol{r}_0) + \boldsymbol{\mathcal{E}}(\boldsymbol{r}_0 + \boldsymbol{r})\big]\boldsymbol{r} \\
&= -\frac{1}{2}Q\big[\boldsymbol{i}(\mathcal{E}_{x,0} + \mathcal{E}_x) + \boldsymbol{j}(\mathcal{E}_{y,0} + \mathcal{E}_y) + \boldsymbol{k}(\mathcal{E}_{z,0} + \mathcal{E}_z)\big](\boldsymbol{i}x + \boldsymbol{j}y + \boldsymbol{k}z) \\
&= -\mathcal{E}_{x,0}Qx - \mathcal{E}_{y,0}Qy - \mathcal{E}_{z,0}Qz \\
&\quad - Q\frac{1}{2}\sum_q \left(\frac{\partial \mathcal{E}_x}{\partial q}\right)_0 qx - Q\frac{1}{4}\sum_{q,q'} \left(\frac{\partial^2 \mathcal{E}_x}{\partial q \partial q'}\right)_0 qq'x \\
&\quad - Q\frac{1}{2}\sum_q \left(\frac{\partial \mathcal{E}_y}{\partial q}\right)_0 qy - Q\frac{1}{4}\sum_{q,q'} \left(\frac{\partial^2 \mathcal{E}_y}{\partial q \partial q'}\right)_0 qq'y \\
&\quad - Q\frac{1}{2}\sum_q \left(\frac{\partial \mathcal{E}_z}{\partial q}\right)_0 qz - Q\frac{1}{4}\sum_{q,q'} \left(\frac{\partial^2 \mathcal{E}_z}{\partial q \partial q'}\right)_0 qq'z + \cdots
\end{aligned}$$

$$= -\sum_q \mathcal{E}_{q,0}\tilde{\mu}_q - \frac{1}{2}\sum_{q,q'}\left(\frac{\partial \mathcal{E}_q}{\partial q'}\right)_0 \tilde{\Theta}_{qq'}$$
$$-\frac{1}{4}\sum_{q,q',q''}\left(\frac{\partial^2 \mathcal{E}_q}{\partial q'\partial q''}\right)_0 \tilde{\Omega}_{qq'q''} + \cdots, \tag{12.4}$$

electric moments

where "$+\cdots$" denotes higher order terms, while $\tilde{\mu}_q = Qq$, $\tilde{\Theta}_{qq'} = Qqq'$, $\tilde{\Omega}_{qq'q''} = Qqq'q''$, ... represent the components of the successive *moments* of a particle with electric charge Q pointed by the vector $\boldsymbol{r}_0 + \boldsymbol{r}$ and calculated within the coordinate system located at $\boldsymbol{r}_0$. For example, $\tilde{\mu}_x = Qx$, $\tilde{\Theta}_{xy} = Qxy$, $\tilde{\Omega}_{xzz} = Qxz^2$, etc.

Traceless multipole moments

The components of such moments in general are not independent. The three components of the dipole moment are indeed independent, but among the quadrupole components we have the obvious relations $\tilde{\Theta}_{qq'} = \tilde{\Theta}_{q'q}$ from their definition, which reduces the number of independent components from 9 to 6. This however is not all. From the Maxwell equations (see Appendix G, p. 962), we obtain the *Laplace equation*, $\Delta V = 0$ (Δ means the Laplacian), valid for points without electric charges. Since $\boldsymbol{\mathcal{E}} = -\nabla V$ and therefore $-\nabla\boldsymbol{\mathcal{E}} = \Delta V$ we obtain

$$\nabla\boldsymbol{\mathcal{E}} = \sum_q \frac{\partial \mathcal{E}_q}{\partial q} = 0. \tag{12.5}$$

Thus, in the energy expression

$$-\frac{1}{2}\sum_{q,q'}\left(\frac{\partial \mathcal{E}_q}{\partial q'}\right)_0 \tilde{\Theta}_{qq'}$$

of eq. (12.4), the quantities $\tilde{\Theta}_{qq'}$ are not independent, since we have to satisfy the condition (12.5).

We have therefore only five independent moments that are quadratic in coordinates. For the same reasons we have only seven (among 27) independent moments with the third power of coordinates. Indeed, ten original components $\Omega_{q,q',q''}$ with $(q, q', q'') = xxx, yxx, yyx, yyy, zxx, zxy, zzx, zyy, zzy, zzz$ correspond to all permutationally non-equivalent moments. We have, however, three relations these components have to satisfy. They correspond to the three equations, each obtained from the differentiation of eq. (12.5) over x, y, z, respectively. This results in only seven *independent* components[7] $\Omega_{q,q',q''}$.

These relations between moments can be taken into account (adding to the energy expression the zeros resulting from the Laplace equation (12.5)) and we

---

[7] In Appendix X on p. 1038 the definition of the multipole moments based on polar coordinates is reported. The number of independent components of such moments is equal to the number of independent Cartesian components and equals $(2l+1)$ for $l = 0, 1, 2, \ldots$ with the consecutive l pertaining, respectively, to the monopole (or charge) $(2l+1=1)$, dipole (3), quadrupole (5), octupole (7), etc. (in agreement with what we find now for the particular moments).

may introduce what are known as the *traceless Cartesian multipole moments*[8] (the symbol without tilde), which may be chosen in the following way

traceless moments

$$\mu_q \equiv \tilde{\mu}_q, \tag{12.6}$$

$$\Theta_{qq'} \equiv \frac{1}{2}\left[3\tilde{\Theta}_{qq'} - \delta_{qq'}\sum_q \tilde{\Theta}_{qq}\right]. \tag{12.7}$$

The adjective "traceless" results from relations of the type $\mathrm{Tr}\,\Theta = \sum_q \Theta_{qq} = 0$, etc.

Then, the expression for the energy contribution changes to (please check that both expressions are identical after using the Laplace formula)

$$\Delta E = -\sum_q \mathcal{E}_{q,0}\mu_q - \frac{1}{3}\sum_{q,q'}\left(\frac{\partial \mathcal{E}_q}{\partial q'}\right)_0 \Theta_{qq'} - \cdots. \tag{12.8}$$

Most often we compute first the moments and then use them to calculate the traceless multipole moments (cf. Table 9.1 on p. 484).

System of charges in an inhomogeneous electric field

Since we are interested in constructing the *perturbation operator* that is to be added to the Hamiltonian, from now on, according to the postulates of quantum mechanics (Chapter 1), we will treat the coordinates x, y, z in eq. (12.8) as *operators* of multiplication (by just x, y, z). In addition we would like to treat many charged particles, not just one, because we want to consider molecules. To this end we will sum up all the above expressions, computed for each charged particle, separately. As a result the Hamiltonian for the total system (nuclei and electrons) in the electric field $\mathcal{E}$ represents the Hamiltonian of the system without field ($\hat{H}^{(0)}$) and the perturbation ($\hat{H}^{(1)}$):

$$\hat{H} = \hat{H}^{(0)} + \hat{H}^{(1)}, \tag{12.9}$$

where

$$\hat{H}^{(1)} = -\sum_q \hat{\mu}_q \mathcal{E}_q - \frac{1}{3}\sum_{qq'} \hat{\Theta}_{qq'}\mathcal{E}_{qq'} \cdots \tag{12.10}$$

with the convention

$$\mathcal{E}_{qq'} \equiv \frac{\partial \mathcal{E}_q}{\partial q'},$$

where the field component and its derivatives are computed at a given point ($\boldsymbol{r}_0$), e.g., at the centre of mass of the molecule, while $\hat{\mu}_q, \hat{\Theta}_{qq'}, \ldots$ denote the operators of the components of the traceless Cartesian multipole moments of the total system, i.e. of the molecule.[9] How can we imagine multipole moments? We may

---

[8] The reader will find the corresponding formulae in the article by A.D. Buckingham, *Advan. Chem. Phys.* 12 (1967) 107 or by A.J. Sadlej, "*Introduction to the Theory of Intermolecular Interactions*", Lund's Theoretical Chemistry Lecture Notes, Lund, 1990.

[9] Also calculated with respect to this point. This means that if the molecule is large, then $\boldsymbol{r}$ may become dangerously large. In such a case, as a consequence, the series (12.8) may converge slowly.

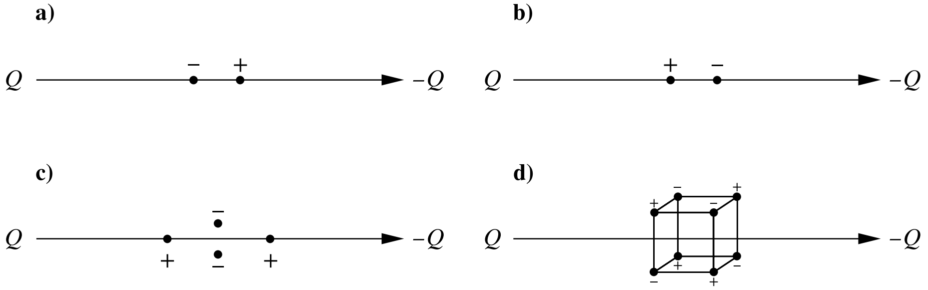


Fig. 12.3. Explanation of why a dipole moment interacts with the electric field intensity, a quadrupole moment with its gradient, while the octupole moment does not interact either with the first or with the second. The external electric field is produced by two *distant* electric charges $Q > 0$ and $-Q$ (for long distances between them the field in the central region between the charges resembles a homogeneous field) and interacts with an object (a dipole, a quadrupole, etc.) located in the central region. A favourable orientation of the object corresponds to the lowest interaction energy with Q and $-Q$. Fig. (a) shows such a low-energy situation for a dipole: the charge "+" protrudes towards $-Q$, while the charge "−" protrudes towards Q. Fig. (b) corresponds to the opposite situation, energetically non-favourable. As we can see, the interaction energy of the dipole with the electric field *differentiates* these two situations. Now, let us locate a quadrupole in the middle (c). Let us imagine that a neutral point object has just split into four point charges (of the same absolute value). The system lowers its energy by the "−" charges going off the axis, because they have increased their distance from the charge $-Q$, but at the same time the system energy has increased by the same amount, since the charges went off the symmetrically located charge $+Q$. What about the "+" charges? The splitting of the "++" charges leads to an energy gain for the right-hand side "+" charge, because it approached $-Q$, and went off the charge $+Q$, but the left-hand side "+" charge gives the opposite energy effect. Altogether the net result is zero. Conclusion: *the quadrupole does not interact with the homogeneous electric field.* Now, let us imagine an inhomogeneous field having a non-zero gradient along the axis (e.g., both Q charges differ by their absolute values). There will be no energy difference for the minus charges, but one of the plus charges will be attracted more strongly than the other. Therefore, the *quadrupole interacts with the field gradient.* We may foresee that the quadrupole will align with its longer axis along the field. Fig. (d) shows an octupole (all charges have the same absolute value). Indeed, the total charge, all the components of the dipole as well as of quadrupole moment are equal to zero, but the octupole (eight charges in the vertices of a cube) is non-zero. Such an octupole does not interact with a homogeneous electric field (because the right and left sides of the cube do not gain anything when interacting). It also does not interact with the field gradient (because each of the above mentioned sides of the cube is composed of two plus and two minus charges; what the first ones gain the second ones lose).

associate a given multipole moment with a simple object that exhibits a non-zero value for this particular moment, but all lower multipole moments equal zero.[10] Some of such objects are shown in Fig. 12.3, located between two charges Q and $-Q$ producing an "external field". Note that the multipole moment names (*di*pole, *quadru*pole, *octu*pole) indicate the *number* of the point charges from which these objects are built.

Eq. (12.10) means that if the system exhibits non-zero multipole moments (before any interaction or due to the interaction), they will interact with the external electric field: the dipole with the electric field intensity, the quadrupole with its gradient, etc. Fig. 12.3 shows why this happens.

---

[10] *Higher* moments in general will be non-zero.

12.2.2 THE HOMOGENEOUS ELECTRIC FIELD

In case of a *homogeneous* external electric field, the contribution to $\hat{H}^{(1)}$ comes from the first term in eq. (12.10):

$$\hat{H} = \hat{H}^{(0)} + \hat{H}^{(1)} = \hat{H}^{(0)} - \hat{\mu}_x \mathcal{E}_x - \hat{\mu}_y \mathcal{E}_y - \hat{\mu}_z \mathcal{E}_z = \hat{H}^{(0)} - \hat{\boldsymbol{\mu}} \cdot \boldsymbol{\mathcal{E}}, \tag{12.11}$$

where the dipole moment operator $\hat{\boldsymbol{\mu}}$ has the form:

$$\hat{\boldsymbol{\mu}} = \sum_i \boldsymbol{r}_i Q_i, \tag{12.12}$$

with the vector $\boldsymbol{r}_i$ indicating the particle i of charge Q_i.

Hence,

$$\frac{\partial \hat{H}}{\partial \mathcal{E}_q} = -\hat{\mu}_q. \tag{12.13}$$

From this it follows

$$\left\langle \psi \middle| \frac{\partial \hat{H}}{\partial \mathcal{E}_q} \psi \right\rangle = -\langle \psi | \hat{\mu}_q \psi \rangle = -\mu_q, \tag{12.14}$$

where μ_q is the expected value of the q-th component of the dipole moment.

From the Hellmann–Feynman theorem we have:

$$\left\langle \psi \middle| \frac{\partial \hat{H}}{\partial \mathcal{E}_q} \psi \right\rangle = \frac{\partial E}{\partial \mathcal{E}_q}, \tag{12.15}$$

therefore

$$\frac{\partial E}{\partial \mathcal{E}_q} = -\mu_q. \tag{12.16}$$

On the other hand, in the case of a *weak electric field* $\boldsymbol{\mathcal{E}}$ we certainly may write the Taylor expansion:

$$E(\boldsymbol{\mathcal{E}}) = E^{(0)} + \sum_q \left(\frac{\partial E}{\partial \mathcal{E}_q} \right)_{\boldsymbol{\mathcal{E}}=\mathbf{0}} \mathcal{E}_q + \frac{1}{2!} \sum_{q,q'} \left(\frac{\partial^2 E}{\partial \mathcal{E}_q \partial \mathcal{E}_{q'}} \right)_{\boldsymbol{\mathcal{E}}=\mathbf{0}} \mathcal{E}_q \mathcal{E}_{q'} + \frac{1}{3!} \sum_{q,q',q''} \left(\frac{\partial^3 E}{\partial \mathcal{E}_q \partial \mathcal{E}_{q'} \partial \mathcal{E}_{q''}} \right)_{\boldsymbol{\mathcal{E}}=\mathbf{0}} \mathcal{E}_q \mathcal{E}_{q'} \mathcal{E}_{q''} + \cdots, \tag{12.17}$$

where $E^{(0)}$ stands for the energy of the unperturbed molecule.

Linear and non-linear responses to a homogeneous electric field

Comparing (12.16) and (12.17) we get,

$$\frac{\partial E}{\partial \mathcal{E}_q} = -\mu_q = \left(\frac{\partial E}{\partial \mathcal{E}_q} \right)_{\boldsymbol{\mathcal{E}}=\mathbf{0}} + \sum_{q'} \left(\frac{\partial^2 E}{\partial \mathcal{E}_q \partial \mathcal{E}_{q'}} \right)_{\boldsymbol{\mathcal{E}}=\mathbf{0}} \mathcal{E}_{q'} + \frac{1}{2} \sum_{q',q''} \left(\frac{\partial^3 E}{\partial \mathcal{E}_q \partial \mathcal{E}_{q'} \partial \mathcal{E}_{q''}} \right)_{\boldsymbol{\mathcal{E}}=\mathbf{0}} \mathcal{E}_{q'} \mathcal{E}_{q''} \cdots, \tag{12.18}$$

induced dipole moment

or replacing the derivatives by their equivalents (permanent dipole moment, molecular polarizability and hyperpolarizabilities)

$$\mu_q = \mu_{0q} + \sum_{q'} \alpha_{qq'} \mathcal{E}_{q'} + \frac{1}{2} \sum_{q'q''} \beta_{qq'q''} \mathcal{E}_{q'} \mathcal{E}_{q''} + \cdots. \tag{12.19}$$

The meaning of the formula for μ_q is clear: in addition to the permanent dipole moment $\boldsymbol{\mu}_0$ of the isolated molecule, we have its modification, i.e. an induced dipole moment, which consists of the *linear* part in the field ($\sum_{q'} \alpha_{qq'} \mathcal{E}_{q'}$) and of the *nonlinear* part ($\frac{1}{2} \sum_{q'q''} \beta_{qq'q''} \mathcal{E}_{q'} \mathcal{E}_{q''} + \cdots$). The quantities that characterize the molecule: vector $\boldsymbol{\mu}_0$ and tensors $\boldsymbol{\alpha}, \boldsymbol{\beta}, \ldots$ are of key importance. By comparing (12.18) with (12.19) we have the following relations:

dipole polarizability

the permanent (field-independent) dipole moment of the molecule (component q):

$$\mu_{0q} = -\left(\frac{\partial E}{\partial \mathcal{E}_q}\right)_{\boldsymbol{\mathcal{E}}=\mathbf{0}}, \tag{12.20}$$

the total dipole moment (field-dependent):

$$\mu_q = -\left(\frac{\partial E}{\partial \mathcal{E}_q}\right), \tag{12.21}$$

the component qq' of the *dipole polarizability* tensor:

$$\alpha_{qq'} = -\left(\frac{\partial^2 E}{\partial \mathcal{E}_q \partial \mathcal{E}_{q'}}\right)_{\boldsymbol{\mathcal{E}}=\mathbf{0}} = \left(\frac{\partial \mu_q}{\partial \mathcal{E}_{q'}}\right)_{\boldsymbol{\mathcal{E}}=\mathbf{0}}, \tag{12.22}$$

the component $qq'q''$ of the *dipole hyperpolarizability* tensor:

$$\beta_{qq'q''} = -\left(\frac{\partial^3 E}{\partial \mathcal{E}_q \partial \mathcal{E}_{q'} \partial \mathcal{E}_{q''}}\right)_{\boldsymbol{\mathcal{E}}=\mathbf{0}}. \tag{12.23}$$

dipole hyperpolarizabilities

Next, we obtain higher-order dipole hyperpolarizabilities ($\gamma, \ldots$), which will contribute to the characteristics of the way the molecule is polarized when subject to a weak electric field.

The homogeneous field: dipole polarizability and dipole hyperpolarizabilities

From eq. (12.17) we have the following expression for the energy of the molecule in the electric field

$$E(\boldsymbol{\mathcal{E}}) = E^{(0)} - \sum_q \mu_{0q} \mathcal{E}_q - \frac{1}{2} \sum_{qq'} \alpha_{qq'} \mathcal{E}_q \mathcal{E}_{q'} - \frac{1}{3!} \sum_{qq'q''} \beta_{qq'q''} \mathcal{E}_q \mathcal{E}_{q'} \mathcal{E}_{q''} - \frac{1}{4!} \sum_{qq'q''q'''} \gamma_{qq'q''q'''} \mathcal{E}_q \mathcal{E}_{q'} \mathcal{E}_{q''} \mathcal{E}_{q'''} \cdots. \tag{12.24}$$

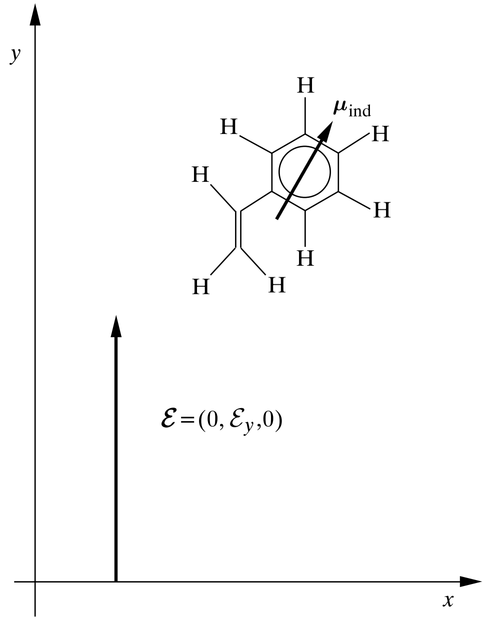


Fig. 12.4. The direction of the induced dipole moment may differ from the direction of the electric field applied (due to the tensor character of the polarizability and hyperpolarizabilities). Example: the vinyl molecule in a planar conformation. Assume the following Cartesian coordinate system: x (horizontal in the figure plane), y (vertical in the figure plane) and z (perpendicular to the figure plane), and the external electric field: $\mathcal{E}=(0,\mathcal{E}_y,0)$. The component x of the induced dipole moment is equal to (within the accuracy of linear terms, eq. (12.19)) $\mu_{\text{ind},x}=\mu_x-\mu_{0x}\approx\alpha_{xy}\mathcal{E}_y$, $\mu_{\text{ind},y}\approx\alpha_{yy}\mathcal{E}_y$, $\mu_{\text{ind},z}\approx\alpha_{zy}\mathcal{E}_y$. Due to the symmetry plane $z=0$ of the molecule (cf. p. 630) $\alpha_{zy}=\alpha_{zx}=0$, and similarly for the hyperpolarizabilities, we have $\mu_{\text{ind},z}=0$. As we can see, despite the field having its x component equal to zero, the induced dipole moment x component does not ($\mu_{\text{ind},x}\neq 0$).

This formula pertains exclusively to the interaction of the molecular *dipole* (the permanent dipole plus the induced linear and non-linear response) with the electric field. As seen from (12.19), the induced dipole moment with the components $\mu_q-\mu_{0q}$ may have a different direction from the applied electric field (due to the tensor character of the polarizability and hyperpolarizabilities). This is quite understandable, because the electrons will move in a direction which will represent a compromise between the direction of the electric field which forces them to move, and the direction where the polarization of the molecule is easiest (Fig. 12.4).

It is seen from eqs. (12.19) and (12.22) that:

- As a second derivative of a continuous function E the polarizability represents a symmetric tensor ($\alpha_{qq'}=\alpha_{q'q}$).
- The polarizability characterizes this part of the *induced dipole moment, which is proportional to the field*.
- If non-diagonal components of the polarizability tensor are non-zero, then the flow direction of the charge within the molecule will differ from the direction of the field. This would happen when the electric field forced the electrons to flow into empty space, while they had a “highway” to travel along some chemical bonds (cf. Fig. 12.4).
- If a molecule is symmetric with respect to the plane $q=0$, say, $z=0$, then all the (hyper)polarizabilities with odd numbers of the indices z, are equal to zero (cf. Fig. 12.4). It has to be like this, because otherwise a change of the electric field component from $\mathcal{E}_z$ to $-\mathcal{E}_z$ would cause a change in energy (see eq. (12.24)), which is impossible, because the molecule is symmetric with respect to the plane $z=0$.

- The dipole *hyper*polarizabilities (β and higher-order) are very important, because, if we limited ourselves to the first two terms of (12.19) containing only μ_{0q} and $\alpha_{qq'}$ (i.e. neglecting β and higher hyperpolarizabilities), the molecule would be equally easy to polarize in two opposite directions.[11] This is why, for a molecule with a centre of inversion, all *odd* dipole hyperpolarizabilities (i.e. with an odd number of indices q) have to equal zero, because the invariance of the energy with respect to the inversion will be preserved that way. If the molecule does not exhibit an inversion centre, the non-zero odd dipole hyperpolarizabilities ensure that polarization of the molecule depends, in general, on whether we change the electric field vector to the opposite direction. This is how it should be. Why were the electrons able to move to the same extent towards an electron donor (on one end of the molecule) as to an electron acceptor (on the other end)?

Does the dipole moment really exist?

Now, let us complicate things. What is $\boldsymbol{\mu}_0$? We used to say that it is the dipole moment of the molecule in its ground state. Unfortunately, no molecule has a non-zero dipole moment. This follows from the invariance of the Hamiltonian with respect to the inversion operation and was described on p. 65. The mean value of the dipole moment operator is bound to be zero since the square of the wave function is symmetric, while the dipole moment operator itself is antisymmetric with respect to the inversion. Thus for *any* molecule[12] $\mu_{0q} = 0$ for $q = x, y, z$. The reason is the rotational part of the wave function (cf. p. 230). This is quite natural. Dear reader, did you ever think why the hydrogen atom does not exhibit a dipole moment despite having two poles: the proton and the electron? The reason is the same. The electron in its ground state is described by the $1s$ orbital, which does not prefer any direction and the dipole moment integral for the hydrogen atom gives zero. Evidently, we have got into trouble.

The trouble disappears after the Born–Oppenheimer approximation (the clamped nuclei approximation, cf. p. 227) is used, i.e. if we hold the *molecule fixed in space*. In such a case, the molecule *has* the dipole moment and this dipole moment is to be inserted into formulae as $\boldsymbol{\mu}_0$, and then we may calculate the polarizability, hyperpolarizabilities, etc. But what do we do, when we do not apply the Born–Oppenheimer approximation? Yet, in experiments we do not use the Born–Oppenheimer approximation (or any other). We have to allow the molecule to rotate and then the dipole moment $\boldsymbol{\mu}_0$ disappears.

It is always good to see things working in a simple model, and simple models resulting in exact solutions of the Schrödinger equation have been described in

---

[11] According to eq. (12.19) the absolute value of the q component of the *induced dipole moment* $\boldsymbol{\mu}_{\text{ind}} = \boldsymbol{\mu} - \boldsymbol{\mu}_0$ would be identical for $\mathcal{E}_q$ as well as for $-\mathcal{E}_q$.

[12] "Everybody knows" that the HF molecule has a non-zero dipole moment. Common knowledge says that when an electric field is applied, the HF dipole aligns itself along the electric field vector. At any field, no matter how small? This would be an incredible scenario. No, the picture has to be more complex.

Chapter 4. A good model for our rotating molecule may be the rigid rotator with a dipole moment (a charge Q on one mass and $-Q$ on the other).[13] The Hamiltonian remains, in principle, the same as for the rigid rotator, because we have to add a *constant* $-\frac{Q^2}{R}$ to the potential energy, which does not change anything. Thus the ground state wave function is Y_0^0 as before, which tells us that *every* orientation of the rigid dipolar rotor in space is *equally probable*.

Well, what if the rotating molecule is located in a *very weak* electric field? After the field is switched on the molecule will of course continue to rotate, but a tiny preference of those orientations which orient the dipole at least partly *along* the electric field, will appear. We may say that the system will have a certain *polarizability*, which can be computed as a negative second derivative of the energy with respect to the electric field. This effect will be described by our perturbation theory, eq. (12.22). If the electric field were of medium intensity, instead of the orientational preferences, the rotator would already pay great attention to it, and would "oscillate" about the direction of the electric field $\boldsymbol{\mathcal{E}}$. This would already be beyond the capabilities of perturbation theory (too large perturbation). Finally, if the electric field were very strong (e.g., along the x axis), the rotator would orient *exactly along the field*, the energy gain would be equal to[14] $-\boldsymbol{\mu}\cdot\boldsymbol{\mathcal{E}} = -QR\mathcal{E}_x$ and its second derivative would be zero (as well as the polarizability).[15] Therefore,[16]

---

[13] This moment therefore has a constant length.

[14] This is what we often assume in phenomenological theories, forgetting that at weak field intensities the situation is different.

[15] The case we have been talking about pertains to the ground state of the system. What if the electric field were applied to the system in its excited state? For a medium electric field, the subsequent energy levels as functions of the field will be nearly equidistant. Why? The reason is quite simple. For medium electric fields the eigenstates of the rotator will be related to its *oscillations* about the direction of the field. In the harmonic approximation this means equidistant energy eigenvalues. The corresponding vibrational wave functions (that depend on the deviation angle θ from the direction of the field) will have large amplitudes for small θ values and an increasing number of nodes when the vibrational quantum number increases.

[16] A detailed analysis of this problem was carried out by Grzegorz Łach (these results prior to publication are acknowledged). Two asymptotic dependencies of energy as a function of electric field intensity have been obtained: $E(\boldsymbol{\mathcal{E}}) = \frac{1}{I}f(I\mu\mathcal{E})$, where I stands for the moment of the inertia of the rotator, and the function $f(x)$ for small field intensities (this results from a perturbation theory with the unperturbed system corresponding to the absence of an electric field)

$$f(x) = -\frac{1}{3}x^2 + \frac{11}{135}x^4 - \frac{376}{8505}x^6 + \cdots$$

for very large field intensities

$$f(x) = -x + \sqrt{x} - \frac{1}{4} - \frac{1}{64}\frac{1}{\sqrt{x}} + \cdots.$$

It has been shown, that the first two terms in the last formula also follow from perturbation theory. However, in this perturbation theory the unperturbed operator does not correspond to the free molecule, but in addition contains a harmonic oscillator potential (with the angle θ as the corresponding coordinate). The anharmonicity is treated as a perturbation.

at weak electric fields we expect quadratic dependence of the energy on the field and only at stronger fields may we expect linear dependence.

12.2.3 THE INHOMOGENEOUS ELECTRIC FIELD: MULTIPOLE POLARIZABILITIES AND HYPERPOLARIZABILITIES

dipole polarizabilities and hyper-polarizabilities

The formula $\mu_q = \mu_{0q} + \sum_{q'} \alpha_{qq'}\mathcal{E}_{q'} + \frac{1}{2}\sum_{q'q''} \beta_{qq'q''}\mathcal{E}_{q'}\mathcal{E}_{q''} + \cdots$ pertains to the polarizabilities and hyperpolarizabilities in a *homogeneous* electric field. The polarizability $\alpha_{qq'}$ characterizes a *linear* response of the molecular *dipole* moment to the electric field, the hyperpolarizability $\beta_{qq'q''}$ and the higher ones characterize the corresponding *non-linear* response of the molecular *dipole* moment. However, a change of the charge distribution contains more information than just that offered by the induced dipole moment. For a *non-homogeneous* electric field the energy expression changes, because besides the dipole moment, higher multipole moments (permanent as well as induced) come into play (see Fig. 12.3). Using the Hamiltonian (12.9) with the perturbation (12.10) (which corresponds to a molecule immersed in a non-homogeneous electric field) we obtain the following energy expression from the Hellmann–Feynman theorem (formula (12.15)) and eq. (12.17):

$$E(\boldsymbol{\mathcal{E}}) = E^{(0)} + E_\mu + E_\Theta + E_{\mu-\Theta} + \cdots, \tag{12.25}$$

where besides the unperturbed energy $E^{(0)}$, we have:

- the dipole–field interaction energy E_μ (including the permanent and induced dipole – these terms appeared earlier for the homogeneous field):

$$E_\mu = -\left[\sum_q \mu_{0q}\mathcal{E}_q + \frac{1}{2}\sum_{qq'} \alpha_{qq'}\mathcal{E}_q\mathcal{E}_{q'} + \frac{1}{6}\sum_{q,q',q''} \beta_{q,q',q''}\mathcal{E}_q\mathcal{E}_{q'}\mathcal{E}_{q''}\cdots\right], \tag{12.26}$$

quadrupole polarizability

- next, the terms that pertain to the *inhomogeneity* of the electric field: the energy E_Θ of the interaction of the field gradient with the quadrupole moment (the permanent one Θ – the first term, and of the induced one; C stands for the *quadrupole polarizability*, and then, in the terms denoted by "$+\cdots$" there are the non-linear responses with *quadrupole hyperpolarizabilities*):

$$E_\Theta = -\left[\frac{1}{3}\sum_{qq'} \Theta_{qq'}\mathcal{E}_{qq'} + \frac{1}{6}\sum_{qq'q''q'''} C_{qq'q''q'''}\mathcal{E}_{qq'}\mathcal{E}_{q''q'''} + \cdots\right], \tag{12.27}$$

- the dipole–quadrupole cross term $E_{\mu-\Theta}$:

$$E_{\mu-\Theta} = -\left[\frac{1}{3}\sum_{q,q',q''} A_{q,q'q''}\mathcal{E}_q\mathcal{E}_{q'q''} + \frac{1}{6}\sum_{q,q',q'',q'''} B_{qq',q''q'''}\mathcal{E}_q\mathcal{E}_{q'}\mathcal{E}_{q''q'''}\right] \tag{12.28}$$

and

- the interaction of higher multipoles (permanent as well as induced: first, the octupole Ω with the corresponding *octupole polarizabilities and hyperpolarizabilities*, etc.) with the higher derivatives of electric field together with the corresponding cross terms denoted as: $+\cdots$.

octupole polarizability

12.3 HOW TO CALCULATE THE DIPOLE MOMENT

The dipole moment in the normalized state $|n\rangle$ is calculated (according to the postulates of quantum mechanics, Chapter 1) as the mean value $\boldsymbol{\mu} = \langle n|\hat{\boldsymbol{\mu}}|n\rangle$ of the dipole moment operator[17]

$$\hat{\boldsymbol{\mu}} = -\sum_i \boldsymbol{r}_i + \sum_A Z_A \boldsymbol{R}_A, \tag{12.29}$$

where $\boldsymbol{r}_i$ are the vectors indicating the electrons and $\boldsymbol{R}_A$ shows nucleus A with the charge Z_A (in a.u.).

For a neutral molecule only, the dipole moment operator and the dipole moment itself do not depend on the choice of the origin of the coordinate system. When two coordinate systems differ by translation $\boldsymbol{R}$, then, in general, we may obtain two different results ($\boldsymbol{r}_i$ and Q_i stand for the position vector and charge of particle i):

$$\hat{\boldsymbol{\mu}} = \sum_i Q_i \boldsymbol{r}_i,$$

$$\hat{\boldsymbol{\mu}}' = \sum_i Q_i \boldsymbol{r}'_i = \sum_i Q_i(\boldsymbol{r}_i + \boldsymbol{R}) = \hat{\boldsymbol{\mu}} + \sum_i Q_i \boldsymbol{R} = \hat{\boldsymbol{\mu}} + \boldsymbol{R}\sum_i Q_i. \tag{12.30}$$

It is seen that $\hat{\boldsymbol{\mu}}' = \hat{\boldsymbol{\mu}}$, only if $\sum_i Q_i = 0$, i.e. for a neutral system.[18]

This represents a special case of the theorem, saying that the lowest nonvanishing multipole moment does not depend on the choice of the coordinate system; all others may depend on that choice.

12.3.1 HARTREE–FOCK APPROXIMATION

In order to show the reader how we calculate the dipole moment in practice, let us use the Hartree–Fock approximation. Using the normalized Slater determinant $|\Phi_0\rangle$ we have as the Hartree–Fock approximation to the dipole moment:

$$\begin{aligned}\boldsymbol{\mu} &= \langle\Phi_0| - \sum_i \boldsymbol{r}_i + \sum_A Z_A \boldsymbol{R}_A|\Phi_0\rangle = \langle\Phi_0| - \sum_i \boldsymbol{r}_i|\Phi_0\rangle + \langle\Phi_0|\sum_A Z_A \boldsymbol{R}_A|\Phi_0\rangle \\ &= \boldsymbol{\mu}_{\mathrm{el}} + \boldsymbol{\mu}_{\mathrm{nucl}},\end{aligned} \tag{12.31}$$

where the integration goes over the electronic coordinates. The dipole moment of the nuclei $\boldsymbol{\mu}_{\mathrm{nucl}} = \sum_A Z_A \boldsymbol{R}_A$ is very easy to compute, because, in the Born–

---

[17] As is seen, this is an operator having x, y and z components in a chosen coordinate system and each of its components means a *multiplication by the corresponding coordinates and electric charges*.

[18] If you ever have to debug a computer program that calculates the dipole moment, then please remember there is a simple and elegant test at your disposal that is based on the above theorem. You just make two runs of the program for a neutral system each time using a different coordinate system (the two systems differing by a translation). The two results have to be identical.

Oppenheimer approximation, the nuclei occupy some fixed positions in space. The electronic component of the dipole moment $\boldsymbol{\mu}_{\text{el}} = \langle\Phi_0| - \sum_i \boldsymbol{r}_i|\Phi_0\rangle$, according to the Slater–Condon rules (Appendix M on p. 986), amounts to: $\boldsymbol{\mu}_{\text{el}} = -\sum_i n_i(\varphi_i|\boldsymbol{r}_i\varphi_i)$, where n_i stands for the occupation number of the orbital φ_i (let us assume double occupation, i.e. $n_i = 2$). After the LCAO expansion is applied $\varphi_i = \sum_j c_{ji}\chi_j$ and combining the coefficients c_{ji} into the bond order matrix (see p. 365) $\boldsymbol{P}$, we have

$$\boldsymbol{\mu}_{\text{el}} = -\sum_{kl} P_{lk}(\chi_k|\boldsymbol{r}|\chi_l). \tag{12.32}$$

This is in principle all we can say about calculation of the dipole moment in the Hartree–Fock approximation. The rest belongs to the technical side. We choose a coordinate system and calculate all the integrals of type $(\chi_k|\boldsymbol{r}\chi_l)$, i.e. $(\chi_k|x\chi_l)$, $(\chi_k|y\chi_l)$, $(\chi_k|z\chi_l)$. The bond order matrix $\boldsymbol{P}$ is just a by-product of the Hartree–Fock procedure.

12.3.2 ATOMIC AND BOND DIPOLES

It is interesting that the total dipole moment can be decomposed into atomic and pairwise contributions:

$$\boldsymbol{\mu}_{\text{el}} = -\sum_A \sum_{k\in A}\sum_{l\in A} P_{lk}(\chi_k|\boldsymbol{r}|\chi_l) - \sum_A \sum_{k\in A}\sum_{B\neq A}\sum_{l\in B} P_{lk}(\chi_k|\boldsymbol{r}|\chi_l), \tag{12.33}$$

where we assume that the atomic orbital centres (A, B) correspond to the nuclei. If the two atomic orbitals k and l belong to *the same* atom, then we insert $\boldsymbol{r} = \boldsymbol{R}_A + \boldsymbol{r}_A$, where $\boldsymbol{R}_A$ indicates the atom (nucleus) A from the origin, and $\boldsymbol{r}_A$ indicates the separation of the electron from the local origin centred on A. If k and l belong to *different* atoms, then $\boldsymbol{r} = \boldsymbol{R}_{AB} + \boldsymbol{r}_{AB}$, where $\boldsymbol{R}_{AB}$ indicates the centre of the AB section, and $\boldsymbol{r}_{AB}$ represents the position of the electron with respect to this centre. Then,

$$\begin{aligned}\boldsymbol{\mu}_{\text{el}} = &-\sum_A \boldsymbol{R}_A \sum_{k\in A}\sum_{l\in A} S_{kl}P_{lk} - \sum_A\sum_{k\in A}\sum_{l\in A} P_{lk}(\chi_k|\boldsymbol{r}_A|\chi_l)\\ &-\sum_A\sum_{B\neq A}\boldsymbol{R}_{AB}\sum_{k\in A}\sum_{l\in B} S_{kl}P_{lk} - \sum_A\sum_{k\in A}\sum_{B\neq A}\sum_{l\in B} P_{lk}(\chi_k|\boldsymbol{r}_{AB}|\chi_l),\end{aligned} \tag{12.34}$$

where S_{kl} are the overlap integrals. After adding the dipole moment of the nuclei we obtain

$$\boldsymbol{\mu} = \sum_A \boldsymbol{\mu}_A + \sum_A\sum_{B\neq A}\boldsymbol{\mu}_{AB}, \tag{12.35}$$

where

$$\boldsymbol{\mu}_A = \boldsymbol{R}_A\Big(Z_A - \sum_{k\in A}\sum_{l\in A} S_{kl}P_{lk}\Big) - \sum_A\sum_{k\in A}\sum_{l\in A} P_{lk}(\chi_k|\boldsymbol{r}_A|\chi_l)$$

$$\boldsymbol{\mu}_{AB} = -\sum_A\sum_{B\neq A}\boldsymbol{R}_{AB}\sum_{k\in A}\sum_{l\in B} S_{kl}P_{lk} - \sum_A\sum_{k\in A}\sum_{B\neq A}\sum_{l\in B} P_{lk}(\chi_k|\boldsymbol{r}_{AB}|\chi_l).$$

We therefore have a quite interesting result:[19]

> The molecular dipole moment can be represented as the sum of the individual atomic dipole moments and the pairwise atomic dipole contributions.

The P_{lk} is large, when k and l belong to the atoms forming the *chemical bonds* (if compared to two non-bonded atoms, see Appendix S, p. 1015), therefore the dipole moments related to *pairs* of atoms come practically uniquely from *chemical bonds*. The contribution of the lone pairs of the atom A is hidden in the second term of $\boldsymbol{\mu}_A$ and may be quite large (cf. Appendix T on p. 1020).

12.3.3 WITHIN THE ZDO APPROXIMATION

In several semiempirical methods of quantum chemistry (e.g., in the Hückel method) we assume the Zero Differential Overlap (ZDO) approximation, i.e. that $\chi_k \chi_l \approx (\chi_k)^2 \delta_{kl}$ and hence the second terms in $\boldsymbol{\mu}_A$ as well as in $\boldsymbol{\mu}_{AB}$ are equal to zero,[20] and therefore

$$\boldsymbol{\mu} = \sum_A \boldsymbol{R}_A \left(Z_A - \sum_{k \in A} P_{kk} \right) = \sum_A \boldsymbol{R}_A Q_A, \tag{12.36}$$

where $Q_A = (Z_A - \sum_{k \in A} P_{kk})$ represents the net electric charge of the atom A. This result is extremely simple: the dipole moment comes from the atomic charges only.

12.4 HOW TO CALCULATE THE DIPOLE POLARIZABILITY

12.4.1 SUM OVER STATES METHOD (SOS)

Perturbation theory gives the energy of the ground state $|0\rangle$ in a weak electric field as (the sum of the zeroth, first and second-order energies,[21] eqs. (5.22) and (5.26)):

$$E(\boldsymbol{\mathcal{E}}) = E^{(0)} + \langle 0|\hat{H}^{(1)}|0\rangle + \sum_n{}' \frac{|\langle 0|\hat{H}^{(1)}|n\rangle|^2}{E_0^{(0)} - E_n^{(0)}} + \cdots. \tag{12.37}$$

If we assume a homogeneous electric field (see eq. (12.11)), the perturbation is equal to $\hat{H}^{(1)} = -\hat{\boldsymbol{\mu}} \cdot \boldsymbol{\mathcal{E}}$, and we obtain

$$E = E^{(0)} - \langle 0|\hat{\boldsymbol{\mu}}|0\rangle \cdot \boldsymbol{\mathcal{E}} + \sum_n{}' \frac{[\langle 0|\hat{\boldsymbol{\mu}}|n\rangle \cdot \boldsymbol{\mathcal{E}}][\langle n|\hat{\boldsymbol{\mu}}|0\rangle \cdot \boldsymbol{\mathcal{E}}]}{E^{(0)} - E_n^{(0)}} + \cdots. \tag{12.38}$$

The first term represents the energy of the unperturbed molecule, the second term is a correction for the interaction of the permanent dipole moment with the field.

---

[19] This does not represent a unique partitioning, only the total dipole moment should remain the same. For example, the individual atomic contributions include the lone pairs, which otherwise could be counted as a separate lone pair contribution.

[20] The second term in $\boldsymbol{\mu}_A$ is equal to zero, because the integrands $\chi_k^2 x$, $\chi_k^2 y$, $\chi_k^2 z$ are all antisymmetric with respect to transformation of the coordinate system $x \to -x$, $y \to -y$, $z \to -z$.

[21] Prime in the summation means that the m-th state is excluded.

The next term already takes into account that the *induced moment* also interacts with the electric field (eq. (12.19)):

$$\sum_n{}' \frac{[\langle 0|\hat{\boldsymbol{\mu}}|n\rangle \cdot \boldsymbol{\mathcal{E}}][\langle n|\hat{\boldsymbol{\mu}}|0\rangle \cdot \boldsymbol{\mathcal{E}}]}{E^{(0)} - E_n^{(0)}} = -\frac{1}{2}\sum_{qq'} \alpha_{qq'}\mathcal{E}_q\mathcal{E}_{q'}, \tag{12.39}$$

where the component qq' of the polarizability is equal to

$$\boxed{\alpha_{qq'} = 2\sum_n{}' \frac{\langle 0|\hat{\mu}_q|n\rangle\langle n|\hat{\mu}_{q'}|0\rangle}{\Delta_n},} \tag{12.40}$$

where $\Delta_n = E_n^{(0)} - E^{(0)}$. The polarizability has the dimension of volume.[22]

Similarly, we may obtain the perturbational expressions for the dipole, quadrupole, octupole hyperpolarizabilities, etc. For example, the ground-state dipole hyperpolarizability β_0 has the form (the $qq'q''$ component, where the prime means that the ground state is omitted – we skip the derivation):

$$\beta_{qq'q''} = \sum_{n,m}{}' \frac{\langle 0|\hat{\mu}_q|n\rangle\langle n|\hat{\mu}_{q'}|m\rangle\langle m|\mu_{q''}|0\rangle}{\Delta_n\Delta_m} - \langle 0|\mu_q|0\rangle\sum_n{}' \frac{\langle 0|\hat{\mu}_{q'}|n\rangle\langle n|\hat{\mu}_{q''}|0\rangle}{(\Delta_n)^2}. \tag{12.41}$$

A problem with the SOS method is its slow convergence and the fact that, whenever the expansion functions do not cover the energy continuum, the result is incomplete.

Example 1. *The hydrogen atom in an electric field – perturbational approach*

An atom or molecule, when located in electric field undergoes a *deformation*. We will show this in detail, taking the example of the hydrogen atom.

First, let us introduce a Cartesian coordinate system, within which the whole event will be described. Let the electric field be directed towards your right, i.e. has the form $\boldsymbol{\mathcal{E}} = (\mathcal{E}, 0, 0)$, with a constant $\mathcal{E} > 0$. The positive value of $\mathcal{E}$ means, according to the definition of electric field intensity, that a positive unit charge would move along $\boldsymbol{\mathcal{E}}$, i.e. from left to right. Thus, the anode is on your left and the cathode on your right.

We will consider a weak electric field, therefore the perturbation theory is applicable; this means just small corrections to the unperturbed situation. In our case the first-order correction to the wave function (see eq. (5.25)) will be expanded in the series of hydrogen atoms orbitals (they form the complete set,[23] cf. Chapter 5)

$$\psi_0^{(1)} = \sum_{k\,(\neq 0)} \frac{\langle k|\hat{H}^{(1)}|0\rangle}{-\frac{1}{2} - E_k^{(0)}}\psi_k^{(0)}, \tag{12.42}$$

---

[22] Because μ^2 has the dimension of charge$^2\times$ length^2, and Δ_n has the dimension of energy as for example in Coulombic energy: charge^2/length.

[23] Still they do not span the continuum.

where $|k\rangle \equiv \psi_k^{(0)}$ with energy $E_k^{(0)} = -\frac{1}{2n^2}$ (n is the principal quantum number of the state k) denotes the corresponding wave function of the hydrogen atom, $\hat{H}^{(1)}$ is the perturbation, which for a homogeneous electric field has the form $\hat{H}^{(1)} = -\hat{\boldsymbol{\mu}} \cdot \boldsymbol{\mathcal{E}} = -\hat{\mu}_x \mathcal{E}$, and $\hat{\mu}_x$ is the dipole moment operator (its x component). The operator, according to eq. (12.29), represents the sum of products: charge (in our case of the electron or proton) times the x coordinate of the corresponding particle (let us denote them x and X, respectively): $\hat{\mu}_x = -x + X$, where the atomic units have been assumed. To keep the expression as simple as possible, let us locate the proton at the origin of the coordinate system, i.e. $X = 0$. Finally, $\hat{H}^{(1)} = x\mathcal{E}$, because the electron charge is equal[24] to -1. Thus the perturbation $\hat{H}^{(1)}$ is simply proportional to the x coordinate of the electron.[25]

In order not to work in vain, let us first check which unperturbed states k will contribute to the summation on the right-hand side of (12.42). The ground state ($k = 0$), i.e. the $1s$ orbital is excluded (by the perturbation theory), next, $k = 1, 2, 3, 4$ denote the orbitals $2s, 2p_x, 2p_y, 2p_z$. The contribution of the $2s$ is equal to zero, because $\langle 2s|\hat{H}^{(1)}|1s\rangle = 0$ due to the antisymmetry of the integrand with respect to reflection $x \to -x$ ($\hat{H}^{(1)}$ changes its sign, while the orbitals $1s$ and $2s$ do not). A similar argument excludes the $2p_y$ and $2p_z$ orbitals. Hence, for the time being we have only a single candidate[26] $2p_x$. This time the integral is not zero and we will calculate it in a minute. If the candidates from the next shell ($n = 3$) are considered, similarly, the only non-zero contribution comes from $3p_x$. We will however stop our calculation at $n = 2$, because our goal is only to show how the whole machinery works. Thus, we need to calculate

$$\frac{\langle 2p_x|\hat{H}^{(1)}|1s\rangle}{E_0^{(0)} - E_1^{(0)}} = \frac{\langle 2p_x|x|1s\rangle}{E_0^{(0)} - E_1^{(0)}}\mathcal{E}.$$

The denominator is equal to $-1/2 + 1/8 = -3/8$ a.u. Calculation of the integral (a fast exercise for students[27]) gives 0.7449 a.u. At $\mathcal{E} = 0.001$ a.u. we obtain the coefficient -0.001986 at the normalized orbital $2p_x$ in the first-order correction to the wave function. The negative value of the coefficient means that the orbital

---

[24] It is, therefore, the operator of multiplication by x times a constant $\mathcal{E}$.

[25] The proton might be located anywhere. The result does not depend on this choice, because the perturbation operators will differ by a constant. This, however, means that the nominator $\langle k|\hat{H}^{(1)}|1\mathrm{s}\rangle$ in the formula will remain unchanged, because $\langle k|1\mathrm{s}\rangle = 0$.

[26] Note how fast our computation of the integrals proceeds. The main job (zero or not zero – that is the question) is done by the group theory.

[27] From p. 181 we have

$$\begin{aligned}\langle 2p_x|x|1s\rangle &= \frac{1}{4\pi\sqrt{2}}\int_0^{\infty} \mathrm{d}r\, r^4 \exp\left(-\frac{3}{2}r\right)\int_0^{\pi} \mathrm{d}\theta \sin^3\theta \int_0^{2\pi} \mathrm{d}\phi \cos^2\phi \\ &= \frac{1}{4\pi\sqrt{2}} 4!\left(\frac{3}{2}\right)^{-5}\frac{4}{3}\pi = 0.7449,\end{aligned}$$

where we have used the formula $\int_0^{\infty} x^n \exp(-\alpha x)\,\mathrm{d}x = n!\alpha^{-(n+1)}$ to calculate the integral over r.

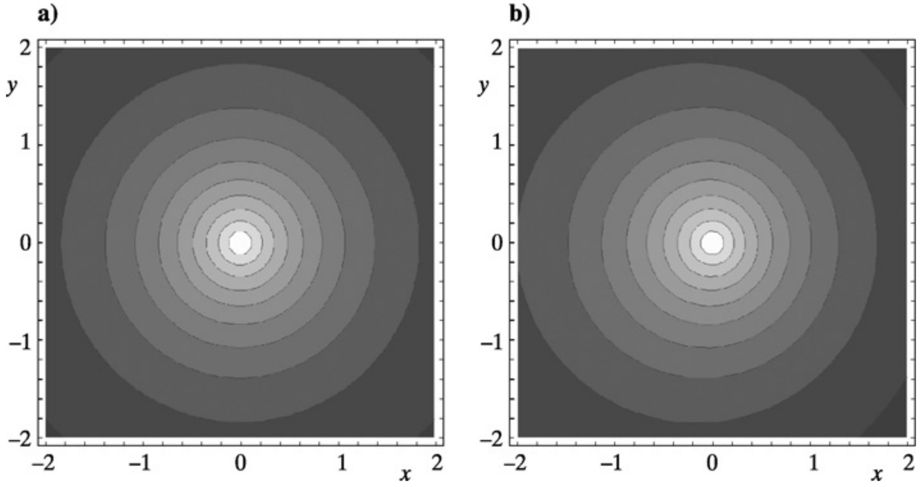


Fig. 12.5. Polarization of the hydrogen atom in an electric field. The wave function for (a) the unperturbed atom (b) the atom in the electric field (a.u.) $\mathcal{E} = (0.1, 0, 0)$. As we can see, there are differences in the corresponding electronic density distributions: in the second case the wave function is deformed towards the anode (i.e. leftwards). Please note that the wave function is less deformed in the region close to the nucleus, than in its left or right neighbourhood. This is a consequence of the fact that the deformation is made by the $-0.1986(2p_x)$ function. Its main role is to subtract on the right and add on the left, and the smallest changes are at the nucleus, because $2p_x$ has its node there.

$-0.001986(2p_x)$ has its positive lobe oriented leftward.[28] The small absolute value of the coefficient results in such a tiny modification of the $1s$ orbital after the electric field is applied, that it will be practically invisible in Fig. 12.5. In order to make the deformation visible, let us use $\mathcal{E} = 0.1$ a.u. Then, the admixture of $2p_x$ is equal to $-0.1986(2p_x)$, i.e. an approximate wave function of the hydrogen atom has the form: $1s - 0.19862p_x$. Fig. 12.5 shows the unperturbed and perturbed $1s$ orbital. As seen, the deformation makes an egg-like shape of the wave function (from a spherical one) – the electron is pulled towards the anode.[29] This is what we expected. Higher expansion functions $(3p_x, 4p_x, \ldots)$ would change the shape of the wave function by only a small amount.

Just *en passant* we may calculate an approximation to the dipole polarizability α_{xx}. From (12.40) we have

$$\alpha_{xx} \cong \frac{16}{3}\langle 2p_x|x(1s)\rangle^2 = \frac{16}{3}(0.7449)^2 = 2.96 \text{ a.u.}$$

The exact (non-relativistic) result is $\alpha_{xx} = 4.5$ a.u. This shows that the number we have received is somewhat off, but after recalling that only a single expansion

---

[28] $2p_x \equiv x\times$ the positive spherically symmetric factor, means the positive lobe of the $2p_x$ orbital is on your right (i.e. on the positive part of the x axis).

[29] This "pulling" results from adding together $1s$ and (with a negative coefficient) $2p_x$, i.e. we decrease the probability amplitude on the right-hand side of the nucleus, and increase it on the left-hand side.

function has been used (instead of infinity of them) we should be quite happy with our result.[30]

12.4.2 FINITE FIELD METHOD

The above calculation represents an example of the application to an atom of what is called the finite field method. In this method we solve the Schrödinger equation for the system in a given homogeneous (weak) electric field. Say, we are interested in the approximate values of $\alpha_{qq'}$ for a molecule. First, we choose a coordinate system, fix the positions of the nuclei in space (the Born–Oppenheimer approximation) and calculate the number of electrons in the molecule. These are the data needed for the input into the reliable method we choose to calculate $E(\boldsymbol{\mathcal{E}})$. Then, using eqs. (12.38) and (12.24) we calculate the permanent dipole moment, the dipole polarizability, the dipole hyperpolarizabilities, etc. by approximating $E(\boldsymbol{\mathcal{E}})$ by a power series of $\mathcal{E}_q$'s.

How do we put the molecule in an electric field? For example, at a long distance from the molecule we locate two point-like electric charges q_x and q_y on x and y axes, respectively. Hence, the total external field at the origin (where the "centre" of the molecule is located) has the components $\mathcal{E}_x = \frac{q_x}{R_x^2}$ and $\mathcal{E}_y = \frac{q_y}{R_y^2}$, with R_x and R_y denoting the distances of both charges from the origin. The field on the molecule will be almost homogeneous, because of the long R_x and R_y distances. In our case the $E(\boldsymbol{\mathcal{E}})$ of eq. (12.24) reads as:

$$
\begin{aligned}
E(\boldsymbol{\mathcal{E}}) = E^{(0)} - \mu_{0,x}\mathcal{E}_x + \mu_{0,y}\mathcal{E}_y - \tfrac{1}{2}\alpha_{xx}\mathcal{E}_x^2 - \tfrac{1}{2}\alpha_{xy}\mathcal{E}_x\mathcal{E}_y - \tfrac{1}{2}\alpha_{yx}\mathcal{E}_x\mathcal{E}_y \\
- \tfrac{1}{2}\alpha_{yy}\mathcal{E}_y^2 + \cdots. \qquad (12.43)
\end{aligned}
$$

Neglecting the cubic and higher terms for a very small field $\boldsymbol{\mathcal{E}}$ (approximation) we obtain an equation for α_{xx}, α_{xy} and α_{yy}, because the polarizability tensor is symmetric. Note that

- $E(\boldsymbol{\mathcal{E}}) - E^{(0)}$, as well as the components $\mu_{0,x}, \mu_{0,y}$ can be calculated;
- there is only one equation, while we have three unknowns to calculate. However, we may apply two other electric fields, which gives us two other equations.[31]

The results of the above procedure depend very much on the configuration of the point-charges. This is why the additional term $-\hat{\boldsymbol{\mu}} \cdot \boldsymbol{\mathcal{E}}$ in the Hamiltonian of the system is much more popular, which, according to eq. (12.11), is equivalent to immersing the system in a homogeneous electric field $\boldsymbol{\mathcal{E}}$.

Example 2. *Hydrogen atom in electric field – variational approach*

The polarizability of the hydrogen atom may also be computed by using the variational method (Chapter 5), in which the variational wave function $\psi = \chi_1 + c\chi_2$

---

[30] Such a situation is quite typical in the practice of quantum chemistry: the first terms of expansions give a lot, while next ones give less and less, the total result approaching, with more and more pain, its limit. Note that in the present case all terms are of the same sign, and we obtain better and better approximations when the expansion becomes longer and longer.

[31] We try to apply small fields, because then the hyperpolarizabilities play a negligible role (the cubic terms in the field intensity will be negligibly small).

where $\chi_1 \equiv 1s$ plus an admixture (as variational parameter) of the p type orbital χ_2 with a certain exponential coefficient ζ (Ritz method of Chapter 5), see Appendix V, eq. (V.1). From eq. (V.4), it can be seen that if χ_2 is taken as the $2p_x$ orbital (i.e. $\zeta = \frac{1}{2}$) we obtain $\alpha_{xx} = 2.96$ a.u., the same number we have already obtained by the perturbational method. If we take $\zeta = 1$ i.e. the same as in hydrogenic orbital $1s$, we will obtain $\alpha_{xx} = 4$ a.u. Well, a substantial improvement.

Is it possible to obtain an even better result with the variational function ψ? Yes, it is. If we use the finite field method (with the electric field equal $\mathcal{E} = 0.01$ a.u.), we will obtain[32] the minimum of E of eq. (V.3) as corresponding to $\zeta_{\text{opt}} = 0.797224$. If we insert $\zeta = \zeta_{\text{opt}}$ into eq. (V.4), we will obtain 4.475 a.u., only 0.5% off the exact result. This nearly perfect result is computed with a single correction function![33]

Sadlej relation

In order to compute accurate values of $E(\boldsymbol{\mathcal{E}})$ extended LCAO expansions have to be used. Andrzej Sadlej[34] noticed that this huge numerical task in fact only takes into account a very simple effect: just a kind of *shift*[35] of the electronic charge distribution towards the anode. Since the atomic orbitals are usually centred on the nuclei and the electronic charge distribution shifts, compensation (still using the on-nuclei atomic orbitals) requires monstrous and expensive LCAO expansions.

In LCAO calculations, nowadays, we most often use Gaussian-type orbitals (GTO, see Chapter 8). They are rarely thought of as representing wave functions of the harmonic oscillator[36] (cf. Chapter 4), but they really do. Sadlej became interested in what would happen if an electron described by a GTO were subject to the electric field $\boldsymbol{\mathcal{E}}$.

> Sadlej noticed that the Gaussian type orbital will change in a similar way to the wave functions of a charged harmonic oscillator in electric field. These however simply shift.

Indeed, this can be shown as follows. The Schrödinger equation for the harmonic oscillator (here: an electron with $m = 1$ in a.u., its position is x) without any electric field is given on p. 166. The Schrödinger equation for an electron oscillating in homogeneous electric field $\mathcal{E} > 0$ has the form:

$$\left(-\frac{1}{2}\frac{d^2}{dx^2} + \frac{1}{2}kx^2 + \mathcal{E}x\right)\psi(x,\mathcal{E}) = E(\mathcal{E})\psi(x,\mathcal{E}). \tag{12.44}$$

---

[32] You may use *Mathematica* and the command FindMinimum[E,{ζ, 1}].

[33] This means that sometimes long expansions in the Ritz method may result from an unfortunate choice of expansion functions.

[34] A.J. Sadlej, *Chem. Phys. Letters* 47 (1977) 50; A.J. Sadlej, *Acta Phys. Polon. A* 53 (1978) 297.

[35] With a deformation.

[36] At least if they represent the $1s$ GTOs.

Now, let us find such constants a and b, that:

$$\frac{1}{2}kx^2 + \mathcal{E}x = \frac{1}{2}k(x-a)^2 + b. \tag{12.45}$$

We immediately get $a = -\mathcal{E}/k$, $b = -\frac{1}{2}ka^2$. The constant b is completely irrelevant, since it only shifts the zero on the energy scale. Thus,

> the solution to a charged harmonic oscillator (oscillating electron) in a homogeneous electric field represents the same function as for the harmonic oscillator without the field, but shifted by $-\frac{\mathcal{E}}{k}$.

Indeed, inserting $x' = x + \frac{\mathcal{E}}{k}$ leads to $\mathrm{d}/\mathrm{d}x = \mathrm{d}/\mathrm{d}x'$ and $\mathrm{d}^2/\mathrm{d}x^2 = \mathrm{d}^2/\mathrm{d}x'^2$ which gives a similar Schrödinger equation except that the harmonic potential is shifted. Therefore, the solution to the equation can be written as simply a shifted zero-field solution $\psi(x') = \psi(x + \frac{\mathcal{E}}{k})$. This is quite understandable, because the operation means nothing more than adding to the parabolic potential energy $kx^2/2$ a term proportional to x, i.e. again a parabola potential (a displaced one though, Fig. 12.6.b).

To see how this displacement depends on the GTO *exponent*, let us recall its relation to the harmonic oscillator force constant k (cf. Chapter 4). The harmonic oscillator eigenfunction corresponds to a Gaussian orbital with an exponent equal to $\alpha/2$, where $\alpha^2 = k$ (in a.u.). Therefore, if we have a GTO with exponent equal to A,

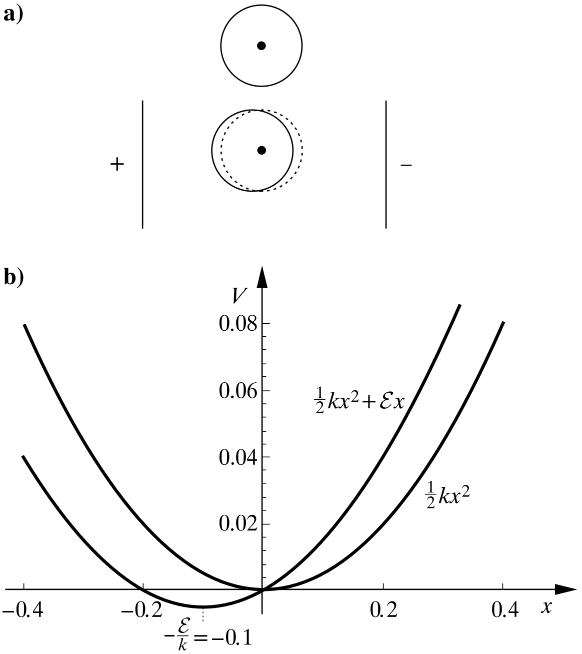


Fig. 12.6. Sadlej relation. The electric field mainly causes a shift of the electronic charge distribution towards the anode (a). A Gaussian type orbital represents the eigenfunction of a harmonic oscillator. Suppose an electron oscillates in a parabolic potential energy well (with the force constant k). In this situation a homogeneous electric field $\mathcal{E}$ corresponds to the perturbation $\mathcal{E}x$, that *conserves the harmonicity with unchanged force constant k* (Fig. b).

this means the corresponding harmonic oscillator has the force constant $k = 4A^2$. Now, if the homogeneous electric field $\mathcal{E}$ is switched on, the centre of this atomic orbital has to move by $\Delta(A) = -\mathcal{E}/k = -\frac{1}{4}\mathcal{E}/A^2$. This means that all the atomic orbitals have to move opposite to the applied electric field (as expected) and the displacement of the orbital is small, if its exponent is large, and *vice versa*. Also, if the atomic electron charge distribution results from several GTOs (as in the LCAO expansion) it deforms in the electric field in such a way that the diffuse orbitals shift a certain amount, while the compact ones (with large exponents) shift by only a small amount. Altogether this does not mean just a simple shift of the electronic charge density, it means that instead its shift is accompanied by a deformation. On the other hand, we may simply optimize the GTO positions within the finite field Hartree–Fock method, and check whether the corresponding shifts $\Delta_{\text{opt}}(A)$ indeed follow the Sadlej relation.[37] It turns out that the relation $\Delta_{\text{opt}}(A) \sim -\mathcal{E}/A^2$ is satisfied to a good accuracy,[38] despite the fact that the potential energy in an atom does not represent a parabola.

The electrostatic catastrophe of the theory

There is a serious problem in finite field theory. If even the *weakest* homogeneous electric field is applied and a very good basis set is used, we are *bound* to have some kind of catastrophe. A nasty word, but unfortunately reflecting quite adequately a mathematical horror we are going to be exposed to after adding to the Hamiltonian operator $\hat{H}^{(1)} = x\mathcal{E}$, here x symbolizes the component of the dipole moment.[39] The problem is that this operator is *unbound*, i.e. for a normalized trial function ϕ the integral $\langle\phi|\hat{H}^{(1)}\phi\rangle$ may attain ∞ or $-\infty$. Indeed, by gradually shifting the function ϕ towards the negative values of the x axis, we obtain more and more negative values of the integral, and for $x = -\infty$ we get $\langle\phi|\hat{H}^{(1)}\phi\rangle = -\infty$. In other words,

> when using atomic orbitals centred far from the nuclei in the region of the negative x (or allowing optimization of the orbital centres with the field switched on), we will lower the energy to $-\infty$, i.e. we obtain a catastrophe. This is quite understandable, because such a system (electrons separated from the nuclei and shifted far away along the x axis) has a huge dipole moment, and therefore *very low energy*.

---

[37] We have tacitly assumed that in the unperturbed molecule the atomic orbitals occupy optimal positions. This assumption may sometimes cause trouble. If the centres of the atomic orbitals in an isolated molecule are non-optimized, we may end up with a kind of antipolarizability: we apply the electric field and, when the atomic orbital centres are optimized, the electron cloud moves opposite to that which we expect. This is possible only because in such a case the orbital centres mainly follow the strong intramolecular electric field, rather than the much weaker external field $\mathcal{E}$ (J.M. André, J. Delhalle, J.G. Fripiat, G. Hennico, L. Piela, *Intern. J. Quantum Chem.* 22S (1988) 665).

[38] This is how the *electric-field–variant orbitals* (EFVO) were born: Andrzej's colleagues did not believe in this simple recipe for calculating polarizabilities, but they lost the bet (a bar of chocolate).

[39] The most dramatic form of the problem would appear if the finite field method was combined with the numerical solution of the Schrödinger or Fock equation.

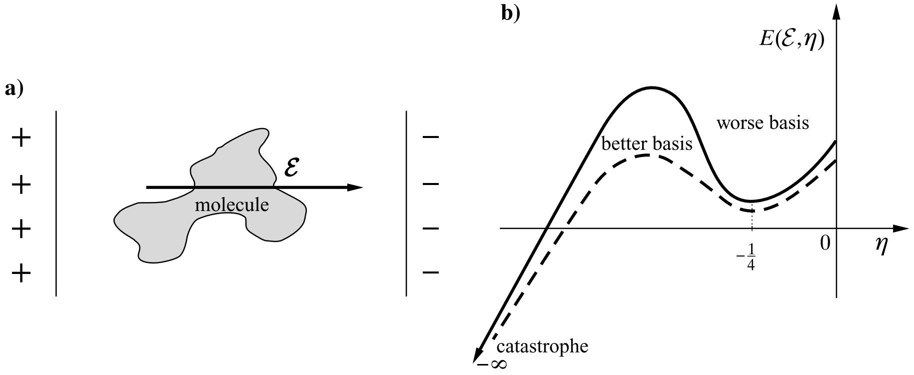


Fig. 12.7. A molecule in a homogeneous electric field (a). In Fig. (b) η is a parameter describing the shift $\eta\mathcal{E}/A^2$ of the Gaussian atomic orbitals along the electric field $\mathcal{E}$, with $\eta = 0$ showing the centring on the nuclei. The total energy $E(\mathcal{E}, \eta)$ is a function of the electric field intensity $\mathcal{E}$ and the basis set shift parameter η. Optimization of η gives a result close to the Sadlej value $\eta = -\frac{1}{4}$, larger $|\eta|$ values first lead to an increase of E, but then to a decrease towards a catastrophe: $\lim_{\eta\to-\infty} E(\mathcal{E}, \eta) = -\infty$.

Suppose the calculations for a molecule in an electric field $\mathcal{E}$ are carried out. According to the Sadlej relation, we shift the corresponding atomic orbitals proportionally to $\eta\mathcal{E}/A^2$, with $\eta \leqslant 0$, and the energy goes down. Around $\eta = -\frac{1}{4}$, which according to Sadlej corresponds to optimal shifts,[40] we may expect the lowest energy, then, for larger $|\eta|$, the energy has to go up. What if we continue to increase (Fig. 12.7) the shift parameter $|\eta|$?

The energy increase will continue only up to some critical value of η. Then, according to the discussion above, the energy will fall to $-\infty$, i.e. to a catastrophe. Thus the energy curve exhibits a barrier (Fig. 12.7), that is related to the basis set quality (its "saturation"). A poor basis means a high barrier and the ideal basis (i.e. the complete basis set), gives no barrier at all, just falling into the abyss with the polarizability going to infinity, etc. Therefore, rather paradoxically, reliable values of polarizability are obtained using a medium quality basis set. An improvement of the basis will lead to worse results.[41]

The above relate to variational calculations. What about the perturbational method? In the first- and second-order corrections to the energy, the formulae contain the zero-order approximation to the wave function $\psi_0^{(0)}$, e.g., $E^{(2)} = \langle\psi_0^{(0)}|\hat{H}^{(1)}\psi_0^{(1)}\rangle$. If the origin of the coordinate system is located on the molecule, then the exponential decay of $\psi_0^{(0)}$ forces the first-order correction to the wave function $\psi_0^{(1)}$ to be also localized close to the origin, otherwise it would tend to zero through the shifting towards the negative values of x (this prevents the integral diverging to $-\infty$). However, the third-order correction to the energy contains the term $\langle\psi_0^{(1)}|\hat{H}^{(1)}\psi_0^{(1)}\rangle$, which may already go to $-\infty$. Hence, the perturbation theory also carries the seed of future electrostatic catastrophe.

---

[40] They are optimal for a parabolic potential.

[41] Once more in this book: excessive wealth does not improve life.

12.4.3 WHAT IS GOING ON AT HIGHER ELECTRIC FIELDS

Polarization

The theory described so far is applicable only when the electric field intensity is small. Such a field can polarize (a small deformation) the electronic charge distribution. More fascinating phenomena begin when the electric field gets stronger.

Deformation

Of course, the equilibrium configurations of the molecule with and without an electric field differ. In a simple case, say the HCl molecule, the HCl distance increases in an electric field. It has to increase, since the cathode pulls the hydrogen atom and repels the chlorine atom, while the anode does the opposite. In more complex cases, like a flexible molecule, the field may change its conformation. This means that the polarizability results both from the electron cloud deformation and the displacement of the nuclei. It turns out that the later effect (called vibrational polarization) is of great importance.[42]

Dissociation

When the electric field gets stronger the molecule may dissociate into ions. To this end, the external electric field intensity has to become comparable to the electric field produced by the molecule itself in its neighbourhood. The intramolecular electric fields are huge, the intermolecular ones are weaker but also very large, of the order of 10^8 V/m, much larger than those offered by current technical installations. No wonder then, that the molecules may interact to such an extent that they may even undergo chemical reactions. When the interaction is weaker, the electric fields produced by molecules may lead to intermolecular complexes, many beautiful examples may be found in biochemistry (see Chapters 13 and 15). A strong external electric field applied to a crystal may cause a cascade of processes, e.g., the so called *displacive phase transitions*, when sudden displacements of atoms occur, and a new crystal structure appears.

displacive phase transition

Destruction

A sufficiently strong electric field will destroy the molecules through their ionization. The resulting ions accelerate in the field, collide with the molecules and ionize them even more (these phenomena are accompanied by light emission as in vacuum tubes). Such processes may lead to the final decomposition of the system (plasma) with the electrons and the nuclei finally reaching the anode and cathode. We will have a vacuum.

Creation!

Let us keep increasing the electric field applied to the vacuum. Will anything interesting happen? We know, from Chapter 3, that when huge electric field intensities

---

[42] J.-M. André, B. Champagne, in "*Conjugated Oligomers, Polymers, and Dendrimers: From Polyacetylene to DNA*", J.L. Brédas (ed.), Bibliothéque Scientifique Francqui, De Boeck Université, p. 349.

are applied (of the order of the electric field intensity in the vicinity of a proton – not feasible for the time being), then *the particles and antiparticles will leap out of the vacuum*! The vacuum is not just nothing. Formula (3.71) gives the probability of such a process.

12.5 A MOLECULE IN AN OSCILLATING ELECTRIC FIELD

Constant and oscillating components

A non-zero hyperpolarizability indicates a *non-linear* response (contributions to the dipole moment proportional to the second and higher powers of the field intensity). This may mean an "inflated" reaction to the applied field, a highly desired feature for contemporary optoelectronic materials. One such reaction is the second- and third-harmonic generation (SHG and THG, respectively), where light of frequency ω generates in a material light with frequencies 2ω and 3ω, respectively. A simple statement about why this may happen is shown below.[43]

Let us imagine a molecule immobilized in a laboratory coordinate system (like in an oriented crystal). Let us switch on a homogeneous electric field $\boldsymbol{\mathcal{E}}$, which has two components, a static component $\boldsymbol{\mathcal{E}}^0$ and an oscillating one $\boldsymbol{\mathcal{E}}^\omega$ with frequency ω:

$$\boldsymbol{\mathcal{E}} = \boldsymbol{\mathcal{E}}^0 + \boldsymbol{\mathcal{E}}^\omega \cos(\omega t). \tag{12.46}$$

We may imagine various experiments here: the steady field along x, y or z and a light beam polarized along x, y or z, we may also vary ω for each beam, etc. Such choices lead to a rich set of non-linear optical phenomena.[44] What will the reaction of the molecule be in such an experiment? Let us see.[45]

Induced dipole moment

The total dipole moment of the molecule (i.e. the permanent moment $\boldsymbol{\mu}_0$ plus the induced moment $\boldsymbol{\mu}_{\text{ind}}$) will depend on time, because $\boldsymbol{\mu}_{\text{ind}}$ does:

$$\mu_q(t) = \mu_{0,q} + \mu_{\text{ind},q}, \tag{12.47}$$

$$\mu_{\text{ind},q}(t) = \sum_{q'} \alpha_{qq'}\mathcal{E}_{q'} + \frac{1}{2}\sum_{q'q''} \beta_{qq'q''}\mathcal{E}_{q'}\mathcal{E}_{q''} + \frac{1}{6}\sum_{q',q'',q'''} \gamma_{qq'q''q'''}\mathcal{E}_{q'}\mathcal{E}_{q''}\mathcal{E}_{q'''} + \cdots. \tag{12.48}$$

Therefore, if we insert $\mathcal{E}_q = \mathcal{E}_q^0 + \mathcal{E}_q^\omega \cos(\omega t)$ as the electric field component for

---

[43] The problem of how the polarizability changes as a function of inducing wave frequency is described in detail in J. Olsen, P. Jørgensen, *J. Chem. Phys.* 82 (1985) 3235.

[44] S. Kielich, "*Molecular non-linear optics*", Warszawa–Poznań, PWN (1977).

[45] For the sake of simplicity we have used the same frequency and the same phases for the light polarized along x, y and z.

$q = x, y, z$, we obtain

$$\begin{aligned}\mu_q(t) = \mu_{0,q} &+ \sum_{q'} \alpha_{qq'}\big[\mathcal{E}^0_{q'} + \mathcal{E}^\omega_{q'}\cos(\omega t)\big] \\ &+ \frac{1}{2}\sum_{q'q''} \beta_{qq'q''}\big[\mathcal{E}^0_{q'} + \mathcal{E}^\omega_{q'}\cos(\omega t)\big] \times \big[\mathcal{E}^0_{q''} + \mathcal{E}^\omega_{q''}\cos(\omega t)\big] \\ &+ \frac{1}{6}\sum_{q',q'',q'''} \gamma_{qq'q''q'''}\big[\mathcal{E}^0_{q'} + \mathcal{E}^\omega_{q'}\cos(\omega t)\big]\big[\mathcal{E}^0_{q''} + \mathcal{E}^\omega_{q''}\cos(\omega t)\big] \\ &\times \big[\mathcal{E}^0_{q'''} + \mathcal{E}^\omega_{q'''}\cos(\omega t)\big] + \cdots. \end{aligned} \tag{12.49}$$

Second (SHG) and Third (THG) Harmonic Generation

After multiplication and simple trigonometry we have

$$\boxed{\mu_q(t) = \mu_{\omega=0,q} + \mu_{\omega,q}\cos\omega t + \mu_{2\omega,q}\cos(2\omega t) + \mu_{3\omega,q}\cos(3\omega t),} \tag{12.50}$$

where the amplitudes μ corresponding to the coordinate $q \in x, y, z$ and to the particular resulting frequencies $0, \omega, 2\omega, 3\omega$ have the following form[46]

$$\begin{aligned}\mu_{\omega=0,q} = \mu_{0,q} &+ \sum_{q'} \alpha_{qq'}(0;0)\mathcal{E}^0_{q'} + \frac{1}{2}\sum_{q',q''} \beta_{qq'q''}(0;0,0)\mathcal{E}^0_{q'}\mathcal{E}^0_{q''} \\ &+ \frac{1}{6}\sum_{q',q'',q'''} \gamma_{qq'q''q'''}(0;0,0,0)\mathcal{E}^0_{q'}\mathcal{E}^0_{q''}\mathcal{E}^0_{q'''} \\ &+ \frac{1}{4}\sum_{q'q''} \beta_{q,q',q''}(0;-\omega,\omega)\mathcal{E}^\omega_{q'}\mathcal{E}^\omega_{q''} \\ &+ \frac{1}{4}\sum_{q',q'',q'''} \gamma_{qq'q''q'''}(0;0,-\omega,\omega)\mathcal{E}^0_{q'}\mathcal{E}^\omega_{q''}\mathcal{E}^\omega_{q'''}, \\ \mu_{\omega,q} = &\sum_{q'} \alpha_{qq'}(-\omega;\omega)\mathcal{E}^\omega_{q'} + \sum_{q',q''} \beta_{qq'q''}(-\omega;\omega,0)\mathcal{E}^\omega_{q'}\mathcal{E}^0_{q''} \\ &+ \frac{1}{2}\sum_{q',q'',q'''} \gamma_{qq'q''q'''}(-\omega;\omega,0,0)\mathcal{E}^\omega_{q'}\mathcal{E}^0_{q''}\mathcal{E}^0_{q'''}\end{aligned}$$

---

[46] According to convention, a given (hyper)polarizability, e.g., $\gamma_{qq'q''q'''}(-3\omega;\omega,\omega,\omega)$, is accompanied (in parenthesis) by the frequencies ω corresponding to the three directions x, y, z of the incident light polarization (here: q', q'' and q''', preceded by minus the Fourier frequency of the term, -3ω, which symbolizes the photon energy conservation law). Some of the symbols, e.g., $\gamma_{qq'q''q'''}(-\omega;\omega,-\omega,\omega)$, after a semicolon have negative values, which means a partial (as in $\gamma_{qq'q''q'''}(-\omega;\omega,-\omega,\omega)$) or complete (as in $\beta_{q,q',q''}(0;-\omega,\omega)$) cancellation of the intensity of the oscillating electric field.

$$+\frac{1}{8}\sum\sum_{q',q'',q'''}\gamma_{qq'q''q'''}(-\omega;\,\omega,-\omega,\omega)\mathcal{E}^{\omega}_{q'}\mathcal{E}^{\omega}_{q''}\mathcal{E}^{\omega}_{q'''},$$

$$\mu_{2\omega,q}=\frac{1}{4}\sum_{q'q''}\beta_{q,q',q''}(-2\omega;\,\omega,\omega)\mathcal{E}^{\omega}_{q'}\mathcal{E}^{\omega}_{q''}$$

$$+\frac{1}{4}\sum\sum_{q',q'',q'''}\gamma_{qq'q''q'''}(-2\omega;\,\omega,\omega,0)\mathcal{E}^{\omega}_{q'}\mathcal{E}^{\omega}_{q''}\mathcal{E}^{0}_{q'''}, \qquad (12.51)$$

$$\mu_{3\omega,q}=\frac{1}{24}\sum_{q',q'',q'''}\gamma_{qq'q''q'''}(-3\omega;\,\omega,\omega,\omega)\mathcal{E}^{\omega}_{q'}\mathcal{E}^{\omega}_{q''}\mathcal{E}^{\omega}_{q'''}. \qquad (12.52)$$

We see that:

- An oscillating electric field may result in a non-oscillating dipole moment related to the hyperpolarizabilities $\beta_{q,q',q''}(0;-\omega,\omega)$ and $\gamma_{qq'q''q'''}(0;0,-\omega,\omega)$, which manifests as an electric potential difference on two opposite crystal faces.
- The dipole moment oscillates with the basic frequency ω of the incident light and in addition, with two other frequencies: the second (2ω) and third (3ω) harmonics (SHG and THG, respectively). This is supported by experiment (mentioned in the example at the beginning of the chapter), applying incident light of frequency ω we obtain emitted light with frequencies[47] 2ω and 3ω.

Note that to generate a large second harmonic the material has to have large values of the hyperpolarizabilities β and γ. The THG needs a large γ. In both cases a strong laser electric field is necessary. The SHG and THG therefore require support from the theoretical side: we are looking for high hyperpolarizability materials and quantum mechanical calculations may predict such materials *before* an expensive organic synthesis is done.[48]

MAGNETIC PHENOMENA

The electric and magnetic fields (both of them are related by the Maxwell equations, Appendix G) interact differently with matter, which is highlighted in Fig. 12.8, where the electron trajectories in both fields are shown. They are totally different, the trajectory in the magnetic field has a cycloid character, while in the electric field it is a parabola. This is why the description of magnetic properties differs so much from that of electric properties.

---

[47] This experiment was first carried out by P.A. Franken, A.E. Hill, C.W. Peters, G. Weinreich, *Phys. Rev. Letters* 7 (1961) 118.

[48] In molecular crystals it is not sufficient that particular molecules have high values of hyperpolarizability. What counts is the hyperpolarizability of the crystal unit cell.

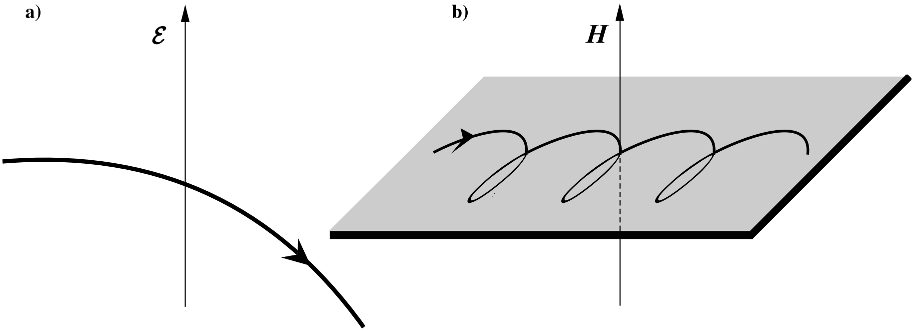


Fig. 12.8. The trajectories of an electron in the (a) electric field – the trajectory is a parabola (b) magnetic field, perpendicular to the incident velocity – the trajectory is a cycloid in a plane perpendicular to the figure.

12.6 MAGNETIC DIPOLE MOMENTS OF ELEMENTARY PARTICLES

12.6.1 ELECTRON

magnetic dipole moment

An elementary particle, besides its orbital angular momentum, may also have internal angular momentum, or spin, cf. p. 25. In Chapter 3, the Dirac theory led to a relation between the *spin* angular momentum $\boldsymbol{s}$ of the electron and its *dipole magnetic moment* $\boldsymbol{M}_{\text{spin,el}}$ (eq. (3.62), p. 122):

$$\boldsymbol{M}_{\text{spin,el}} = \gamma_{\text{el}}\boldsymbol{s},$$

with the *gyromagnetic factor*[49]

$$\gamma_{\text{el}} = -2\frac{\mu_B}{\hbar},$$

where the Bohr magneton (m_0 is the electronic rest mass)

$$\mu_B = \frac{e\hbar}{2m_0c}.$$

The relation is quite a surprise, because the gyromagnetic factor is twice as large as that appearing in the relation between the electron *orbital* angular momentum $\boldsymbol{L}$ and the associated magnetic dipole moment

$$\boldsymbol{M}_{\text{orb,el}} = -\frac{\mu_B}{\hbar}\boldsymbol{L}. \tag{12.53}$$

gyro-magnetic factor

Quantum electrodynamics explains this effect qualitatively – predicting a factor very close to the experimental value[50] 2.0023193043737, known with the breathtaking accuracy of ±0.0000000000082.

---

[49] From the Greek word *gyros*, or circle; it is believed that a circular motion of a charged particle is related to the resulting magnetic moment.

[50] R.S. Van Dyck Jr., P.B. Schwinberg, H.G. Dehmelt, *Phys. Rev. Letters* 59 (1990) 26.

12.6.2 NUCLEUS

Let us stay within the Dirac theory. If, instead of an electron, we take a *nucleus* of charge $+Ze$ and atomic mass[51] M, then we *would presume* (after insertion into the above formulae) the gyromagnetic factor should be $\gamma = 2\frac{Z}{M}\frac{\mu_N}{\hbar}$, where $\mu_N = \frac{e\hbar}{2m_H c}$ (m_H denoting the proton mass) is known as the *nuclear magneton*.[52] For a proton ($Z = 1$, $M = 1$), we would have $\gamma_p = 2\mu_N/\hbar$, whereas the experimental value[53] is $\gamma_p = 5.59\mu_N/\hbar$. What is going on? In both cases we have a single elementary particle (electron or proton), both have the spin quantum number equal to $\frac{1}{2}$, we might expect that nothing special will happen to the proton, and only the mass ratio and charge will make a difference. Instead we see that Dirac theory does relate to the electron, but not to the nuclei. Visibly, the proton is more complex than the electron. We see that even the simplest nucleus has internal machinery, which results in the observed strange deviation. There are lots of quarks in the proton (three valence quarks and a sea of virtual quarks together with the gluons, etc.). The proton and electron polarize the vacuum differently and this results in different gyromagnetic factors. Other nuclei exhibit even stranger properties. Sometimes we even have negative gyromagnetic coefficients. In such a case their magnetic moment is opposite to the spin angular momentum. The complex nature of the internal machinery of the nuclei and vacuum polarization lead to the observed gyromagnetic coefficients.[54] Science has had some success here, e.g., for leptons,[55] but for nuclei the situation is worse. This is why we are simply forced to take this into account in the present book[56] and treat the *spin magnetic moments* of the nuclei as the experimental data:

nuclear magneton

$$\boldsymbol{M}_A = \gamma_A \boldsymbol{I}_A, \tag{12.54}$$

where $\boldsymbol{I}_A$ represents the spin angular momentum of the nucleus A.

---

[51] Unitless quantity.

[52] Ca. 1840 times smaller than the Bohr magneton (for the electron).

[53] Also the gyromagnetic factor for an electron is expected to be ca. 1840 times larger than that for a proton. This means that a proton is expected to create a magnetic field ca. 1840 times weaker than the field created by an electron.

[54] The relation between spin and magnetic moment is as mysterious as that between the magnetic moment and charge of a particle (the spin is associated with a rotation, while the magnetic moment is associated with a rotation of a charged object) or its mass. A neutron has spin equal to $\frac{1}{2}$ and magnetic moment similar to that of a proton despite the zero electric charge. The neutrino has no charge, nearly zero mass and magnetic moment, and still has a spin equal to $\frac{1}{2}$.

[55] And what about the "heavier brothers" of the electron, the muon and taon (cf. p. 268)? For the muon, the coefficient in the gyromagnetic factor (2.0023318920) is similar to that of the electron (2.0023193043737), just a bit larger and agrees equally well with experiment. For the taon we have only a theoretical result, a little larger than for the two other "brothers". Thus, each of the lepton family behaves in a similar way.

[56] With a suitable respect of the Nature's complexity.

12.6.3 DIPOLE MOMENT IN THE FIELD

Electric field

The problem of an electric dipole $\boldsymbol{\mu}$ rotating in an electric field was described on p. 631. There we were interested in the ground state. When the field is switched off (cf. p. 176), the ground state is non-degenerate ($J = M = 0$). After a weak electric field ($\boldsymbol{\mathcal{E}}$) is switched on, the ground-state wave function deforms in such a way as to prefer the alignment of the rotating dipole moment along the field. Since we may always use a complete set of rigid rotator wave functions (at zero field), this means the deformed wave functions *have to be linear combinations of the wave functions corresponding to different* J.

Magnetic field

Imagine a spinning top which is a magnet. If you make it spin (with angular momentum $\boldsymbol{I}$) and leave it in space without any external torque $\boldsymbol{\tau}$, then due to the fact that space is isotropic, its angular momentum will stay constant, because $\frac{d\boldsymbol{I}}{dt} = \boldsymbol{\tau} = \mathbf{0}$ (τ is time), i.e. the top will rotate about its axis with a constant speed and the axis will not move with respect to distant stars, Fig. 12.9.a.

The situation changes if a magnetic field is switched on. Now, the space is no longer isotropic and the vector of the angular momentum is no longer conserved. However, the conservation law *for the projection of the angular momentum on the direction of the field* is still valid. This means that the top makes a precession about the

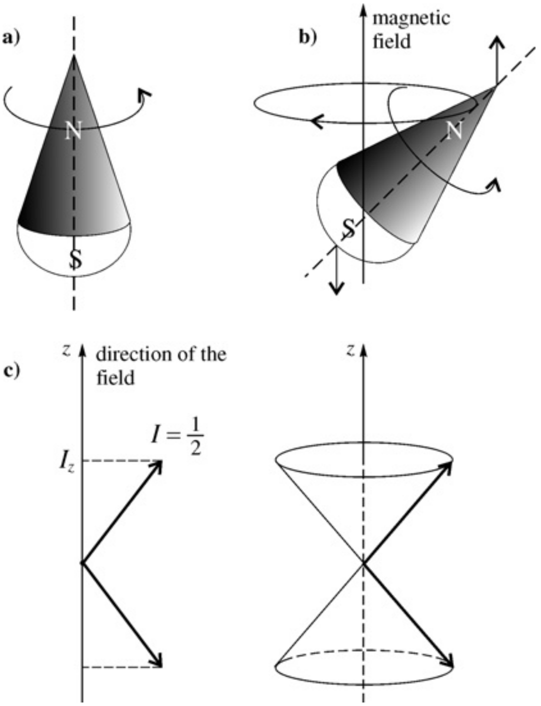


Fig. 12.9. Classical and quantum tops (magnets) in space. (a) The space is isotropic and therefore the classical top preserves its angular momentum $\boldsymbol{I}$, i.e. its axis does not move with respect to distant stars and the top rotates about its axis with a constant speed. (b) The same top in a magnetic field. The space is no longer isotropic, and therefore the total angular momentum is no longer preserved. The projection of the total momentum on the field direction is still preserved. The magnetic field causes a torque $\boldsymbol{\tau}$ (orthogonal to the picture) and $\frac{d\boldsymbol{I}}{dt} = \boldsymbol{\tau}$. This means precession of the top axis about the direction of the field. (c) A quantum top, i.e. an elementary particle with spin quantum number $I = \frac{1}{2}$ in the magnetic field. The projection I_z of its spin angular momentum $\boldsymbol{I}$ is quantized: $I_z = m_I \hbar$ with $m_I = -\frac{1}{2}, +\frac{1}{2}$ and, therefore, we have two energy eigenstates that correspond to two precession cones, directed up and down.

field axis, because $\frac{d\boldsymbol{I}}{dt} = \boldsymbol{\tau} \neq \mathbf{0}$, and $\boldsymbol{\tau}$ is orthogonal to $\boldsymbol{I}$ and to the field, Fig. 12.9.b. In quantum mechanics the magnetic dipole moment $\boldsymbol{M} = \gamma \boldsymbol{I}$ in the magnetic field $\boldsymbol{H} = (0, 0, H)$, $H > 0$, has as many stationary states as is the number of possible projections of the spin angular momentum on the field direction. From Chapter 1, we know that this number is $2I + 1$, where I is the spin quantum number of the particle (e.g., for a proton: $I = \frac{1}{2}$). The projections are equal (Fig. 12.9.c) $m_I \hbar$ with $m_I = -I, -I+1, \ldots, 0, \ldots, +I$. Therefore,

the energy levels in the magnetic field

$$E_{m_I} = -\gamma m_I \hbar H. \tag{12.55}$$

Note, that the energy level splitting is proportional to the magnetic field intensity, Fig. 12.10.

If a nucleus has $I = \frac{1}{2}$, then the energy difference ΔE between the two states in a magnetic field H: one with $m_I = -\frac{1}{2}$ and the other one with $m_I = \frac{1}{2}$, equals $\Delta E = 2 \times \frac{1}{2} \gamma \hbar H = \gamma \hbar H$, and

$$\Delta E = h\nu_L, \tag{12.56}$$

where the Larmor[57] frequency is defined as

$$\nu_L = \frac{\gamma H}{2\pi}. \tag{12.57}$$

We see (Fig. 12.10) that for nuclei with $\gamma > 0$, lower energy corresponds to $m_I = \frac{1}{2}$, i.e. to the spin moment along the field (forming an angle $\theta = 54°44'$ with the magnetic field vector, see p. 28).

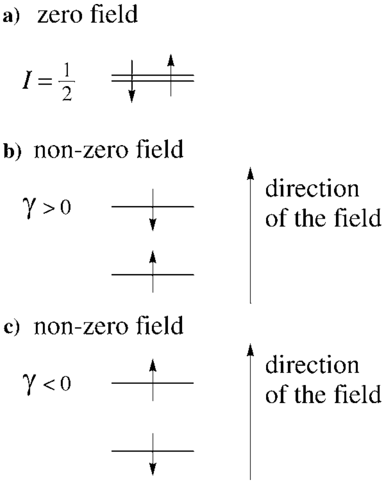


Fig. 12.10. Energy levels in magnetic field $\boldsymbol{H} = (0, 0, H)$ for a nucleus with spin angular momentum $\boldsymbol{I}$ corresponding to spin quantum number $I = \frac{1}{2}$. The magnetic dipole moment equals to $\boldsymbol{M} = \gamma \boldsymbol{I}$ (a) at the zero field the level is doubly degenerate. (b) For $\gamma > 0$ (e.g., a proton) $\boldsymbol{I}$ and $\boldsymbol{M}$ have the same direction. In a non-zero magnetic field the energy equals to $E = -\boldsymbol{M} \cdot \boldsymbol{H} = -M_Z H = -\gamma m_I \hbar H$, where $m_I = \pm\frac{1}{2}$. Thus, the degeneracy is lifted: the state with $m_I = \frac{1}{2}$, i.e. with the positive projection of $\boldsymbol{I}$ on direction of the magnetic field has lower energy. (c) For $\gamma < 0$ $\boldsymbol{I}$ and $\boldsymbol{M}$ have the opposite direction. The state with $m_I = \frac{1}{2}$, i.e. has higher energy.

---

[57] Joseph Larmor (1857–1942), Irish physicist, professor at Cambridge University.

Note that

> there is a difference between the energy levels of the electric dipole moment in an electric field and the levels of the magnetic dipole in a magnetic field. The difference is that, for the magnetic dipole of an elementary particle the states do not have admixtures from the other I values (which is given by nature), while for the electric dipole there are admixtures from states with other values of J.

This suggests that we may also expect such admixtures in a magnetic field. In fact this is true if the particle is complex. For example, the singlet state ($S = 0$) of the hydrogen molecule gets an admixture of the triplet state ($S = 1$) in the magnetic field, because the spin magnetic moments of both electrons tend to align parallel to the field.

12.7 TRANSITIONS BETWEEN THE NUCLEAR SPIN QUANTUM STATES – NMR TECHNIQUE

Is there any possibility of making the nuclear spin flip from one quantum state to another? Yes. Evidently, we have to create distinct energy levels corresponding to different spin projections, i.e. to switch the magnetic field on, Figs. 12.10 and 12.11.a. After the electromagnetic field is applied and its frequency matches the energy level difference, the system absorbs the energy. It looks as if a nucleus absorbs the energy and changes its quantum state. In a genuine NMR experiment, the electromagnetic frequency is fixed (radio wave lengths) and the specimen is scanned by a variable magnetic field. At some particular field values the energy difference matches the electromagnetic frequency and the transition (Nuclear Magnetic Resonance) is observed.

The magnetic field that a particular nucleus feels differs from the external magnetic field applied, because the electronic structure in which the nucleus is immersed in, makes its own contribution (see Fig. 12.11.b,c). Also the nuclear spins interact by creating their own magnetic fields.

We have not yet considered these effects in the non-relativistic Hamiltonian (2.1) (e.g., no spin–spin or spin–field interactions). The effects which we are now dealing with are so small, of the order of 10^{-11} kcal/mole, that they are of no importance for most applications, including UV-VIS, IR, Raman spectra, electronic structure, chemical reactions, intermolecular interactions, etc. This time, however, the situation changes: we are going to study very subtle interactions using the NMR technique which aims precisely at the energy levels that result from spin–spin and spin–magnetic field interactions. Even if these effects are very small, they can be observed. Therefore, we have to consider more exact Hamiltonians. First, we have to introduce

- the interaction of our system with the electromagnetic field,
- then we will consider the influence of the electronic structure on the magnetic field acting on the nuclei

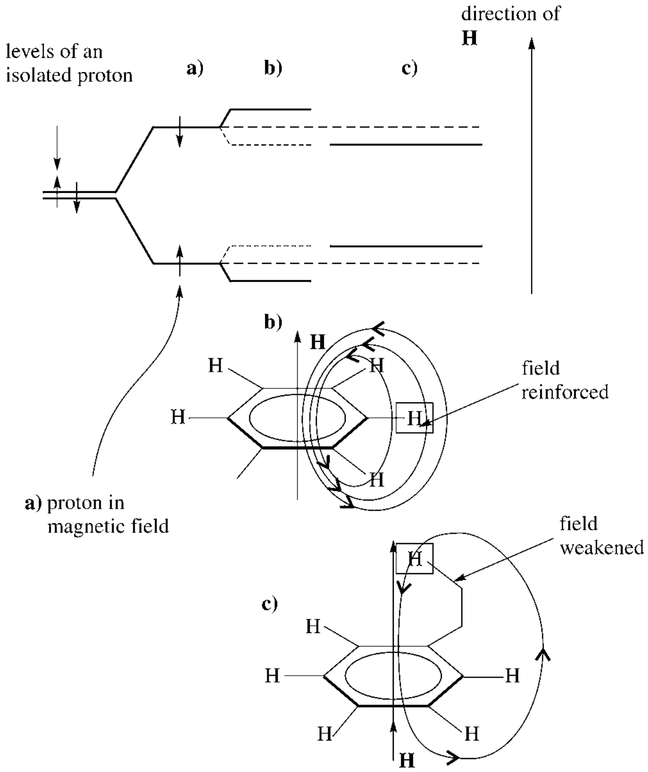


Fig. 12.11. Proton's shielding by the electronic structure. (a) The energy levels of an isolated proton in a magnetic field. (b) The energy levels of the proton of the benzene ring (no nuclear spin interaction is assumed). The most mobile π electrons of benzene (which may be treated as a conducting circular wire) move around the benzene ring in response to the external magnetic field (perpendicular to the ring) thus producing an induced magnetic field. The latter one (when considered along the ring six-fold axis) opposes the external magnetic field, but *at the position of the proton actually leads to an additional increasing of the magnetic field felt by the proton*. This is why the figure shows energy level difference increases due to the *electron shielding effect*. (c) The energy levels of another proton (located along the ring axis) in a similar molecule. This proton feels a local magnetic field that is decreased with respect to the external one (due to the induction effect).

- and finally, the nuclear magnetic moment interaction ("coupling") will be considered.

12.8 HAMILTONIAN OF THE SYSTEM IN THE ELECTROMAGNETIC FIELD

The non-relativistic Hamiltonian[58] $\hat{H}$ of the system of N particles (the j-th particle having mass m_j and charge q_j) moving in an external electromagnetic field with

---

[58] To describe the interactions of the spin magnetic moments, this Hamiltonian will soon be supplemented by the relativistic terms from the Breit Hamiltonian (p. 131).

vector potential A and scalar potential ϕ may be written as[59]

$$\hat{H} = \sum_{j=1} \left[\frac{1}{2m_j} \left(\hat{\boldsymbol{p}}_j - \frac{q_j}{c} \boldsymbol{A}_j \right)^2 + q_j \phi_j \right] + \hat{V}, \tag{12.58}$$

where $\hat{V}$ stands for the "internal" potential coming from the mutual interactions of the particles, and $\boldsymbol{A}_j$ and ϕ_j denote the external vector[60] and scalar potentials $\boldsymbol{A}$ and ϕ, respectively, calculated at the position of particle j.

12.8.1 CHOICE OF THE VECTOR AND SCALAR POTENTIALS

In Appendix G on p. 962 it is shown that there is a certain arbitrariness in the choice of both potentials, which leaves the physics of the system unchanged. If for a homogeneous magnetic field $\boldsymbol{H}$ we choose the vector potential at the point indicated by $\boldsymbol{r} = (x, y, z)$ as (eq. (G.13)) $\boldsymbol{A}(\boldsymbol{r}) = \frac{1}{2}[\boldsymbol{H} \times \boldsymbol{r}]$, then, as shown in Appendix G, we will satisfy the Maxwell equations, and in addition obtain the commonly used relation (eq. (G.12)) $\operatorname{div} \boldsymbol{A} \equiv \nabla \boldsymbol{A} = 0$, known as the *Coulombic gauge*. In this way the origin of the coordinate system ($\boldsymbol{r} = \boldsymbol{0}$) was chosen as the origin of the vector potential (which need not be a rule).

Coulombic gauge

Because $\boldsymbol{\mathcal{E}} = \boldsymbol{0}$ and $\boldsymbol{A}$ is time-independent, $\phi = \text{const}$ (p. 962), which of course means also that $\phi_j = \text{const}$, and as an additive constant, it may simply be eliminated from the Hamiltonian (12.58).

12.8.2 REFINEMENT OF THE HAMILTONIAN

Let us assume the Born–Oppenheimer approximation (p. 229). Thus, the nuclei occupy some fixed positions in space, and in the electronic Hamiltonian (12.58) we have the electronic charges $q_j = -e$ and masses $m_j = m_0 = m$ (we skip the subscript 0 for the rest mass of the electron). Now, let us refine the Hamiltonian by adding the interaction of the particle magnetic moments (of the electrons and nuclei; the moments result from the orbital motion of the electrons as well as from the spin of each particle) with themselves and with the external magnetic field. We have, therefore, a refined Hamiltonian of the system [the particular terms of

---

[59] To obtain this equation we may use eq. (3.33) as the starting point, which together with $E = mc^2$ gives with the accuracy of the first two terms in the expression $E = m_0c^2 + \frac{p^2}{2m_0}$. In the electromagnetic field, after introducing the vector and scalar potentials for particle of charge q we have to replace E by $E - q\phi$, and $\boldsymbol{p}$ by $(\boldsymbol{p} - \frac{q}{c}\boldsymbol{A})$. Then, after shifting the zero of the energy by m_0c^2, the energy operator for a single particle reads as $\frac{1}{2m}(\hat{\boldsymbol{p}} - \frac{q}{c}\boldsymbol{A})^2 + q\phi$, where $\boldsymbol{A}$ and ϕ are the values of the corresponding potentials at the position of the particle. For many particles we sum these contributions up and add the interparticle interaction potential (V). This is what we wanted to obtain (H. Hameka, "*Advanced Quantum Chemistry*", Addison-Wesley Publishing Co., Reading, Massachusetts (1965), p. 40).

[60] Note that the presence of the magnetic field (and therefore of $\boldsymbol{A}$) makes it to appear as if the charged particle moves faster on one side of the vector potential origin and slower on the opposite side.

the Hamiltonian correspond[61] to the relevant terms of the Breit Hamiltonian[62] (p. 131)]

$$\hat{\mathcal{H}} = \hat{H}_1 + \hat{\mathcal{H}}_2 + \hat{\mathcal{H}}_3 + \hat{\mathcal{H}}_4, \tag{12.59}$$

where (δ stands for the Dirac delta function, Appendix E, N is the number of electrons, and the spins have been replaced by the corresponding operators)

$$\hat{\mathcal{H}}_1 = \sum_{j=1}^{N} \frac{1}{2m}\left(\hat{\boldsymbol{p}}_j + \frac{e}{c}\boldsymbol{A}_j\right)^2 + \hat{V} + \hat{H}_{\mathrm{SH}} + \hat{H}_{\mathrm{IH}} + \hat{H}_{\mathrm{LS}} + \hat{H}_{\mathrm{SS}} + \hat{H}_{\mathrm{LL}}, \tag{12.60}$$

$$\hat{\mathcal{H}}_2 = \gamma_{\mathrm{el}} \sum_{j=1}^{N} \sum_{A} \gamma_A \left[\frac{\hat{\boldsymbol{s}}_j \cdot \hat{\boldsymbol{I}}_A}{r_{Aj}^3} - 3\frac{(\hat{\boldsymbol{s}}_j \cdot \boldsymbol{r}_{Aj})(\hat{\boldsymbol{I}}_A \cdot \boldsymbol{r}_{Aj})}{r_{Aj}^5} \right], \tag{12.61}$$

$$\hat{\mathcal{H}}_3 = -\gamma_{\mathrm{el}} \frac{8\pi}{3} \sum_{j=1}^{N} \sum_{A} \gamma_A \delta(\boldsymbol{r}_{Aj}) \hat{\boldsymbol{s}}_j \cdot \hat{\boldsymbol{I}}_A, \tag{12.62}$$

$$\hat{\mathcal{H}}_4 = \sum_{A<B} \gamma_A \gamma_B \left[\frac{\hat{\boldsymbol{I}}_A \cdot \hat{\boldsymbol{I}}_B}{R_{AB}^3} - 3\frac{(\hat{\boldsymbol{I}}_A \cdot \boldsymbol{R}_{AB})(\hat{\boldsymbol{I}}_B \cdot \boldsymbol{R}_{AB})}{R_{AB}^5} \right], \tag{12.63}$$

where in the global coordinate system the internuclear distance means the length of the vector $\boldsymbol{R}_{AB} = \boldsymbol{R}_B - \boldsymbol{R}_A$, while the electron–nucleus distance (of the electron j with nucleus A) will be the length of $\boldsymbol{r}_{Aj} = \boldsymbol{r}_j - \boldsymbol{R}_A$. We have:

- In the term $\hat{\mathcal{H}}_1$, besides the kinetic energy operator in the external magnetic field [with vector potential $\boldsymbol{A}$, and the convention $\boldsymbol{A}_j \equiv \boldsymbol{A}(\boldsymbol{r}_j)$] given by $\sum_{j=1}^{N} \frac{1}{2m}(\hat{\boldsymbol{p}}_j + \frac{e}{c}\boldsymbol{A}_j)^2$, we have the Coulomb potential $\hat{V}$ of the interaction of all the charged particles. Next, we have:
  - The interaction of the spin magnetic moments of the electrons ($\hat{H}_{\mathrm{SH}}$) and of the nuclei ($\hat{H}_{\mathrm{IH}}$) with the field $\boldsymbol{H}$. These terms come from the first part of the term $\hat{H}_6$ of the Breit Hamiltonian, and represent the simple Zeeman terms:

Pieter Zeeman (1865–1943), Dutch physicist, professor at the University of Amsterdam. He became interested in the influence of a magnetic field on molecular spectra and discovered a field-induced splitting of the absorption lines in 1896. He shared the Nobel Prize with Hendrik Lorentz "*for their researches into the influence of magnetism upon radiation phenomena*" in 1902. The Zeeman splitting of star spectra allows us to determine the value of the magnetic field that *was* on the star at the moment the light was emitted!

---

[61] All the terms used in the theory of magnetic susceptibilities and the Fermi contact term can be derived from classical electrodynamics.

[62] Not all of them. As we will see later, the NMR experimental spectra are described by using, for each nucleus, what is known as the shielding constant (related to the shielding of the nucleus by the electron cloud) and the internuclear coupling constants. The shielding and coupling constants enter in a specific way into the energy expression. Only those terms are included in the Hamiltonian that give non-zero contributions to these quantities.

- $\hat{\boldsymbol{\mu}} \cdot \boldsymbol{H}$, where $\hat{\boldsymbol{\mu}}$ is the magnetic moment operator of the corresponding particle. Why do we not have together with $\hat{H}_{\mathrm{SH}} + \hat{H}_{\mathrm{IH}}$ in $\hat{\mathcal{H}}_1$ the term $\hat{H}_{\mathrm{LH}}$, i.e. the interaction of the electron orbital magnetic moment with the field? It would be so nice to have the full set of terms: the spin and the orbital magnetic moments interacting with the field. Everything is fine though, such a term is hidden in the mixed term resulting from $\frac{1}{2m}(\hat{\boldsymbol{p}}_j + \frac{e}{c}\boldsymbol{A}_j)^2$. Indeed, we get the corresponding Zeeman term from the transformation

Zeeman term

$$\frac{e}{mc}\hat{\boldsymbol{p}}_j \cdot \boldsymbol{A}_j = \frac{e}{mc}\boldsymbol{A}_j \cdot \hat{\boldsymbol{p}}_j = \frac{e}{2mc}(\boldsymbol{H} \times \boldsymbol{r}_j) \cdot \hat{\boldsymbol{p}}_j = \frac{e}{2mc}\boldsymbol{H} \cdot (\boldsymbol{r}_j \times \hat{\boldsymbol{p}}_j)$$
$$= \frac{e}{2mc}\boldsymbol{H} \cdot \hat{\boldsymbol{L}}_j = -\boldsymbol{H} \cdot \left(-\frac{e}{2mc}\hat{\boldsymbol{L}}_j\right) = -\boldsymbol{H} \cdot \boldsymbol{M}_{\mathrm{orb,el}}(j),$$

  where $\boldsymbol{M}_{\mathrm{orb,el}}(j)$ is, according to the definition of eq. (12.53), the orbital magnetic moment of the electron j. Next, we have the terms
- the electronic spin–orbit terms ($\hat{H}_{\mathrm{LS}}$), i.e. the corresponding magnetic dipole moment interactions; related to the term $\hat{H}_3$ in the Breit Hamiltonian.
- the electronic spin–spin terms ($\hat{H}_{\mathrm{SS}}$), i.e. the corresponding spin magnetic moment interactions, related to the term $\hat{H}_5$ in the Breit Hamiltonian.
- the electronic orbit–orbit terms ($\hat{H}_{\mathrm{LL}}$), i.e. the electronic orbital magnetic dipole interactions (corresponding to the term $\hat{H}_2$ in the Breit Hamiltonian).

- The terms $\hat{\mathcal{H}}_2, \hat{\mathcal{H}}_3, \hat{\mathcal{H}}_4$ (crucial for the NMR experiment) correspond to the magnetic "dipole–dipole" interaction involving *nuclear* spins (the term $\hat{H}_5$ of the Breit Hamiltonian). In more details these are the classical electronic spin–nuclear spin interaction ($\hat{\mathcal{H}}_2$), the corresponding Fermi contact term[63] ($\hat{\mathcal{H}}_3$) and the classical interaction of the nuclear spin magnetic dipoles ($\hat{\mathcal{H}}_4$), this time without the contact term, because the nuclei are kept at long distances by the chemical bond framework.[64]

The magnetic dipole moment (of a nucleus or electron) "feels" the magnetic field acting on it through the vector potential $\boldsymbol{A}_j$ at the particle's position $\boldsymbol{r}_j$. This $\boldsymbol{A}_j$ is composed of the external field vector potential $\frac{1}{2}[\boldsymbol{H} \times (\boldsymbol{r}_j - \boldsymbol{R})]$ (i.e. associated with the external magnetic field[65] $\boldsymbol{H}$), the individual vector potentials coming from the magnetic dipoles of the nuclei[66] $\sum_A \gamma_A \frac{\boldsymbol{I}_A \times \boldsymbol{r}_{Aj}}{r_{Aj}^3}$ (and having their origins on the individual nuclei) and the vector potential $\boldsymbol{A}_{\mathrm{el}}(\boldsymbol{r}_j)$ coming from the orbital and spin

---

[63] Let us take the example of the hydrogen atom in its ground state. Just note that the highest probability of finding the electron described by the orbital 1s is on the proton. The electron and the proton have spin magnetic moments that necessarily interact after they coincide. This effect is certainly something other than just the dipole–dipole interaction, which as usual describes the magnetic interaction for long distances. We have to have a correction for very short distances – this is the Fermi contact term.

[64] And atomic electronic shell structure.

[65] The vector $\boldsymbol{R}$ indicates the origin of the external magnetic field $\boldsymbol{H}$ vector potential from the global coordinate system (cf. Appendix G and the commentary there related to the choice of origin).

[66] Recalling the force lines of a magnet, we see that the magnetic field vector $\boldsymbol{H}$ produced by the nuclear magnetic moment $\gamma_A \boldsymbol{I}_A$ should reside within the plane of $\boldsymbol{r}_{Aj}$ and $\gamma_A \boldsymbol{I}_A$. This means that $\boldsymbol{A}$ has to be orthogonal to the plane. This is assured by $\boldsymbol{A}_j$ proportional to $\gamma_A \boldsymbol{I}_A \times \boldsymbol{r}_{Aj}$.

magnetic moments of all the electrons

$$\boldsymbol{A}_j \equiv \boldsymbol{A}(\boldsymbol{r}_j) = \frac{1}{2}[\boldsymbol{H} \times \boldsymbol{r}_{0j}] + \sum_A \gamma_A \frac{\boldsymbol{I}_A \times \boldsymbol{r}_{Aj}}{r_{Aj}^3} + \boldsymbol{A}_{\mathrm{el}}(\boldsymbol{r}_j), \tag{12.64}$$

where

$$\boldsymbol{r}_{0j} = \boldsymbol{r}_j - \boldsymbol{R}. \tag{12.65}$$

For closed-shell systems (the majority of molecules) the vector potential $\boldsymbol{A}_{\mathrm{el}}$ may be neglected, i.e. $\boldsymbol{A}_{\mathrm{el}}(\boldsymbol{r}_j) \cong \boldsymbol{0}$, because the magnetic fields of the electrons cancel out for a closed-shell molecule (singlet state).

Rearranging terms

When such a vector potential $\boldsymbol{A}$ is inserted into $\hat{\mathcal{H}}_1$ (just patiently make the square of the content of the parentheses) we immediately get

$$\hat{\mathcal{H}} = \hat{H}_0 + \hat{\mathcal{H}}^{(1)}, \tag{12.66}$$

where $\hat{H}_0$ is the usual non-relativistic Hamiltonian for the isolated system

$$\hat{H}_0 = -\sum_j \frac{\hbar^2}{2m}\Delta_j + \hat{V}, \tag{12.67}$$

$$\hat{\mathcal{H}}^{(1)} = \sum_k^{11} \hat{B}_k, \tag{12.68}$$

while *a few* minutes of a careful calligraphy leads to the result[67]

$$\hat{B}_1 = \frac{e^2}{2mc^2} \sum_{A,B} \sum_j \gamma_A \gamma_B \frac{\hat{\boldsymbol{I}}_A \times \boldsymbol{r}_{Aj}}{r_{Aj}^3} \frac{\hat{\boldsymbol{I}}_B \times \boldsymbol{r}_{Bj}}{r_{Bj}^3}, \tag{12.69}$$

$$\hat{B}_2 = \frac{e^2}{8mc^2} \sum_j (\boldsymbol{H} \times \boldsymbol{r}_{0j}) \cdot (\boldsymbol{H} \times \boldsymbol{r}_{0j}), \tag{12.70}$$

$$\hat{B}_3 = -\frac{i\hbar e}{mc} \sum_A \sum_j \gamma_A \nabla_j \cdot \frac{\hat{\boldsymbol{I}}_A \times \boldsymbol{r}_{Aj}}{r_{Aj}^3}, \tag{12.71}$$

$$\hat{B}_4 = -\frac{i\hbar e}{2mc} \sum_j \nabla_j \cdot (\boldsymbol{H} \times \boldsymbol{r}_{0j}), \tag{12.72}$$

$$\hat{B}_5 = \frac{e^2}{2mc^2} \sum_A \sum_j \gamma_A (\boldsymbol{H} \times \boldsymbol{r}_{0j}) \cdot \frac{\hat{\boldsymbol{I}}_A \times \boldsymbol{r}_{Aj}}{r_{Aj}^3}, \tag{12.73}$$

$$\hat{B}_6 = \hat{\mathcal{H}}_2 = \gamma_{\mathrm{el}} \sum_{j=1}^{N} \sum_A \gamma_A \left[\frac{\hat{\boldsymbol{s}}_j \cdot \hat{\boldsymbol{I}}_A}{r_{Aj}^3} - 3\frac{(\hat{\boldsymbol{s}}_j \cdot \boldsymbol{r}_{Aj})(\hat{\boldsymbol{I}}_A \cdot \boldsymbol{r}_{Aj})}{r_{Aj}^5} \right], \tag{12.74}$$

---

[67] The operators $\hat{B}_3$ and $\hat{B}_4$ contain the nabla (differentiation) operators. It is worth noting that this differentiation pertains to *everything on the right hand side of the nabla, including any function on which* $\hat{B}_3$ and $\hat{B}_4$ *operators will act*.

$$\hat{B}_7 = \hat{\mathcal{H}}_3 = -\gamma_{\text{el}} \frac{8\pi}{3} \sum_{j=1} \sum_{A} \gamma_A \delta(\boldsymbol{r}_{Aj}) \hat{\boldsymbol{s}}_j \cdot \hat{\boldsymbol{I}}_A, \tag{12.75}$$

$$\hat{B}_8 = \hat{H}_{\text{SH}} = -\gamma_{\text{el}} \sum_{j} \hat{\boldsymbol{s}}_j \cdot \boldsymbol{H}, \tag{12.76}$$

$$\hat{B}_9 = \hat{\mathcal{H}}_4 = \sum_{A<B} \gamma_A \gamma_B \left[\frac{\hat{\boldsymbol{I}}_A \cdot \hat{\boldsymbol{I}}_B}{R_{AB}^3} - 3 \frac{(\hat{\boldsymbol{I}}_A \cdot \boldsymbol{R}_{AB})(\hat{\boldsymbol{I}}_B \cdot \boldsymbol{R}_{AB})}{R_{AB}^5} \right], \tag{12.77}$$

$$\hat{B}_{10} = \hat{H}_{\text{IH}} = -\sum_{A} \gamma_A \hat{\boldsymbol{I}}_A \cdot \boldsymbol{H}, \tag{12.78}$$

$$\hat{B}_{11} = \hat{H}_{\text{LS}} + \hat{H}_{\text{SS}} + \hat{H}_{\text{LL}}. \tag{12.79}$$

We are just approaching the coupling of our theory with the NMR experiment. To this end, let us first define an empirical Hamiltonian, which serves in the NMR experiment to find what are known as the nuclear shielding constants and the spin–spin coupling constants. Then we will come back to the perturbation $\hat{\mathcal{H}}^{(1)}$.

12.9 EFFECTIVE NMR HAMILTONIAN

NMR spectroscopy[68] means recording the electromagnetic wave absorption by a system of interacting nuclear magnetic dipole moments.[69] It is important to note that the energy differences detectable by contemporary NMR equipment are of the order of 10^{-13} a.u., while the breaking of a chemical bond corresponds to about 10^{-1} a.u. This is why

> all possible changes of the spin state of a system of nuclei do not change the chemical properties of the molecule. This is really what we could only dream of: we have something like observatory stations (the nuclear spins) that are able to detect tiny chemical bond details.

As will be seen in a moment, to reproduce NMR spectra we need an effective and rotation-averaged Hamiltonian that describes the interaction of the nuclear magnetic moments with the magnetic field and with themselves.

12.9.1 SIGNAL AVERAGING

NMR experiments usually pertain to long (many hours) recording of the radio-wave radiation coming from a liquid specimen. Therefore, we obtain a static (time-averaged) record, which involves various kinds of averaging:

---

[68] The first successful experiment of this kind was described by E.M. Purcell, H.C. Torrey, R.V. Pound, *Phys. Rev.* 69 (1946) 37.

[69] The wave lengths used in the NMR technique are of the order of meters (radio frequencies).

- over the rotations of any single molecule that contributes to the signal (we assume that each dipole keeps the same orientation in space when the molecule is rotating). These rotations can be free or restrained;
- over all the molecules present in the specimen;
- over the vibrations of the molecule (including internal rotations).

12.9.2 EMPIRICAL HAMILTONIAN

The effective NMR Hamiltonian contains some parameters that take into account the electronic cloud structure in which the nuclei are immersed. *These NMR parameters will represent our target.*

Now, let us proceed in this direction.

To interpret the NMR data, it is sufficient to consider an *effective* Hamiltonian (containing explicitly only the nuclear magnetic moments, the electron coordinates are absent and the electronic structure enters only implicitly through some interaction parameters). In the matrix notation we have

shielding constants

$$\hat{\mathcal{H}} = -\sum_A \gamma_A \boldsymbol{H}^T(\boldsymbol{1} - \boldsymbol{\sigma}_A)\boldsymbol{I}_A + \sum_{A<B} \gamma_A\gamma_B\{\boldsymbol{I}_A^T(\boldsymbol{D}_{AB} + \boldsymbol{K}_{AB})\boldsymbol{I}_B\}, \qquad (12.80)$$

where $\boldsymbol{I}_C \equiv (I_{C,x}, I_{C,y}, I_{C,z})^T$ stands for the spin angular momentum of the nucleus C, while $\boldsymbol{\sigma}_A, \boldsymbol{D}_{AB}, \boldsymbol{K}_{AB}$ denote the symmetric square matrices of dimension three (*tensors*):

- $\boldsymbol{\sigma}_A$ is a *shielding constant* tensor of the nucleus A. Due to this shielding, nucleus A feels a *local field* $\boldsymbol{H}_{\text{loc}} = (\boldsymbol{1} - \boldsymbol{\sigma}_A)\boldsymbol{H} = \boldsymbol{H} - \boldsymbol{\sigma}_A\boldsymbol{H}$ instead of the external field $\boldsymbol{H}$ applied (due to the tensor character of $\boldsymbol{\sigma}_A$ the vectors $\boldsymbol{H}_{\text{loc}}$ and $\boldsymbol{H}$ may differ in their length and direction). The formula assumes that the shielding is proportional to the external magnetic field intensity that causes the shielding. Thus, the first term in the Hamiltonian $\hat{\mathcal{H}}$ may also be written as $-\sum_A \gamma_A \boldsymbol{H}_{\text{loc}}^T \boldsymbol{I}_A$.
- $\boldsymbol{D}_{AB}$ is the 3 × 3 matrix describing the (direct) *dipole–dipole interaction through space* defined above.
- $\boldsymbol{K}_{AB}$ is also a 3 × 3 matrix that takes into account that two magnetic dipoles interact additionally through the framework of the chemical bonds or hydrogen bonds that separate them. This is known as the *reduced spin–spin intermediate coupling tensor*.

local field

Without electrons...

Let us imagine, just for fun, removing all the electrons from the molecule (and keep them safely in a drawer), while the nuclei still reside in their fixed positions in space. The Hamiltonian would consist of two types of term:

- *the Zeeman term*: interaction of the nuclear magnetic moments with the external electric field (the nuclear analogue of the first term in $\hat{H}_6$ of the Breit Hamiltonian, p. 131) $-\sum_A \boldsymbol{H} \cdot \hat{\boldsymbol{M}}_A = -\sum_A \gamma_A \boldsymbol{H} \cdot \hat{\boldsymbol{I}}_A$;
- the "through space" dipole–dipole nuclear magnetic moment interaction (the nuclear analogue of the $\hat{H}_5$ term in the Breit Hamiltonian) $\sum_{A<B} \gamma_A\gamma_B\{\hat{\boldsymbol{I}}_A \cdot$

$\boldsymbol{D}_{AB}\hat{\boldsymbol{I}}_B)\}$:

$$\boldsymbol{D}_{AB} = \frac{\boldsymbol{i}\cdot\boldsymbol{j}}{R_{AB}^3} - 3\frac{(\boldsymbol{i}\cdot\boldsymbol{R}_{AB})(\boldsymbol{j}\cdot\boldsymbol{R}_{AB})}{R_{AB}^5},$$

where $\boldsymbol{i}, \boldsymbol{j}$ denote the unit vectors along the x, y, z axes, e.g.,

$$(\boldsymbol{D}_{AB})_{xx} = \frac{1}{R_{AB}^3} - 3\frac{(R_{AB,x})^2}{R_{AB}^5}, \quad (\boldsymbol{D}_{AB})_{xy} = -3\frac{R_{AB,x}R_{AB,y}}{R_{AB}^5}, \quad \text{etc.}$$

with $\boldsymbol{R}_{AB}$ denoting the vector separating nucleus B from nucleus A (of length R_{AB}).

Rotations average out the dipole–dipole interaction

What would happen if we rotated the molecule? In the theory of NMR, there are a lot of notions stemming from classical electrodynamics. In the isolated molecule the total angular momentum has to be conserved (this follows from the isotropic properties of space). The total angular momentum comes, not only from the particles' orbital motion, but also from their spin contributions. The empirical (non-fundamental) conservation law pertains to the total spin angular momentum alone (cf. p. 68), as well as the individual spins separately. The spin magnetic moments are oriented in space and this orientation results from the history of the molecule and may be different in each molecule of the substance. These spin states are non-stationary. The *stationary states correspond to some definite values of the square of the total spin of the nuclei and of the spin projection on a chosen axis*. According to quantum mechanics (Chapter 1), only these values are to be measured. For example, in the hydrogen molecule there are two stationary nuclear spin states: one with parallel spins (ortho-hydrogen) and the other with antiparallel (para-hydrogen). Then we may assume that the hydrogen molecule has two "nuclear gyroscopes" that keep pointing them in the same direction in space when the molecule rotates (Fig. 12.12).

Let us see what will happen if we average the interaction of two magnetic dipole moments (the formula for the interaction of two dipoles will be derived in Chapter 13, p. 701):

$$E_{\text{dip–dip}} = \frac{\boldsymbol{M}_A\cdot\boldsymbol{M}_B}{R_{AB}^3} - 3\frac{(\boldsymbol{M}_A\cdot\boldsymbol{R}_{AB})(\boldsymbol{M}_B\cdot\boldsymbol{R}_{AB})}{R_{AB}^5}.$$

Assume (without losing the generality of the problem) that $\boldsymbol{M}_A$ resides at the origin of a polar coordinate system and has a constant direction along the z axis, while the dipole $\boldsymbol{M}_B$ just moves on the sphere of the radius R_{AB} around $\boldsymbol{M}_A$ (all orientations are equally probable), the $\boldsymbol{M}_B$ vector preserving the same direction in space $(\theta, \phi) = (u, 0)$ all the time. Now, let us calculate the average value of $E_{\text{dip–dip}}$ with respect to all possible positions of $\boldsymbol{M}_B$ on the sphere:

$$\begin{aligned}\bar{E}_{\text{dip–dip}} &= \frac{1}{4\pi}\int_0^{\pi} \mathrm{d}\theta\,\sin\theta \int_0^{2\pi} \mathrm{d}\phi\, E_{\text{dip–dip}} \\ &= \frac{1}{4\pi}\int_0^{\pi} \mathrm{d}\theta\,\sin\theta\end{aligned}$$

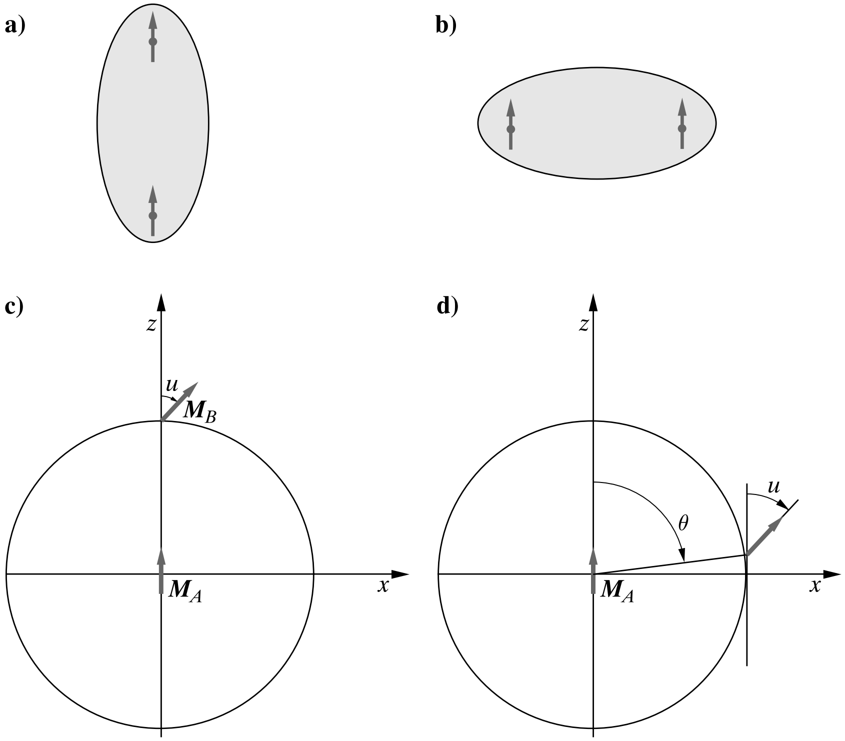


Fig. 12.12. Rotation of a molecule and the nuclear magnetic moments. Fig. (a) shows the orientation of the nuclear magnetic moments in the orthohydrogen at the vertical configuration of the nuclei. Fig. (b) shows the same, but the molecule is oriented horizontally. In the theory of NMR, we assume (in a classical way), that the motion of the molecule does not influence the orientation of both nuclear magnetic moments (c) averaging the dipole–dipole interaction over all possible orientations. Let us immobilize the magnetic moment $\boldsymbol{M}_A$ along the z axis, the magnetic moment $\boldsymbol{M}_B$ will move on the sphere of radius 1 both moments still keeping the same direction in space $(\theta,\phi)=(u,0)$. Fig. (d) shows one of such configurations. Averaging over all possible orientations gives zero (see the text).

$$
\begin{aligned}
&\times \int_0^{2\pi} \mathrm{d}\phi \left[\frac{1}{R_{AB}^3}\boldsymbol{M}_A\cdot\boldsymbol{M}_B - \frac{3}{R_{AB}^5}(\boldsymbol{M}_A\cdot\boldsymbol{R}_{AB})(\boldsymbol{M}_B\cdot\boldsymbol{R}_{AB})\right]\\
&= \frac{M_A M_B}{4\pi R_{AB}^3}\int_0^{\pi}\mathrm{d}\theta\,\sin\theta\int_0^{2\pi}\mathrm{d}\phi\,\big[\cos u - 3\cos\theta\cos(\theta-u)\big]\\
&= \frac{M_A M_B}{2R_{AB}^3}\int_0^{\pi}\mathrm{d}\theta\,\sin\theta\big[\cos u - 3\cos\theta\cos(\theta-u)\big]\\
&= \frac{M_A M_B}{R_{AB}^3}\left\{\cos u - \frac{3}{2}\int_0^{\pi}\mathrm{d}\theta\,\sin\theta\cos\theta[\cos\theta\cos u + \sin\theta\sin u]\right\}\\
&= \frac{M_A M_B}{R_{AB}^3}\left\{\cos u - \frac{3}{2}\left[\cos u\cdot\frac{2}{3} + \sin u\cdot 0\right]\right\} = 0. \qquad (12.81)
\end{aligned}
$$

Thus, the averaging gave 0 *irrespective of the radius* R_{AB} *and of the angle* u *between the two dipoles. This result was obtained when assuming the orientations of both dipoles do not change (the above mentioned "gyroscopes") and that all angles* θ and ϕ are equally probable.

Averaging over rotations

An NMR experiment requires long recording times. This means that each molecule, when rotating freely (gas or liquid[70]) with respect to the NMR apparatus, acquires all possible orientations with equal probability. The equipment will detect an averaged signal. This is why the proposed effective Hamiltonian has to be averaged over the rotations. As we have shown, such an averaging causes the mean dipole–dipole interaction (containing $\boldsymbol{D}_{AB}$) to be equal to zero. If we assume that the external magnetic field is along the z axis, then the averaged Hamiltonian reads as

$$\hat{\mathcal{H}}_{\mathrm{av}} = -\sum_A \gamma_A (1-\sigma_A) H_z \hat{I}_{A,z} + \sum_{A<B} \gamma_A \gamma_B K_{AB} \big(\hat{\boldsymbol{I}}_A \cdot \hat{\boldsymbol{I}}_B\big), \tag{12.82}$$

where $\sigma_A = \frac{1}{3}(\sigma_{A,xx} + \sigma_{A,yy} + \sigma_{A,zz}) = \frac{1}{3}\mathrm{Tr}\, \boldsymbol{\sigma}_A$, with $K_{AB} = \frac{1}{3}\mathrm{Tr}\, \boldsymbol{K}_{AB}$.

This Hamiltonian is at the basis of NMR spectra interpretation. An experimentalist adjusts σ_A for all the magnetic nuclei and K_{AB} for all their interactions, in order to reproduce the observed spectrum. Any theory of NMR spectra should explain the values of these parameters.

Adding the electrons – why the nuclear spin interaction does not average out to zero

We know already why $\boldsymbol{D}_{AB}$ averages out to zero, but why is this not true for $\boldsymbol{K}_{AB}$?

Ramsey and Purcell[71] explained this by what is known as the spin induction mechanism described in Fig. 12.13. Spin induction results in the averaging of $\boldsymbol{K}_{AB}$ and the spin–spin configurations have different weights than in the averaging of $\boldsymbol{D}_{AB}$. This effect is due to the chemical bonds, because it makes a difference if the correlating electrons have their spins oriented parallel or perpendicular to the bond line.

Where does such an effect appear in quantum chemistry? One of the main candidates may be the term $\hat{H}_3$ (the Fermi contact term in the Breit Hamiltonian, p. 131) which couples the orbital motion of the electrons with their spin magnetic

---

[70] This is not the case in the solid state.

[71] N.F. Ramsey, E.M. Purcell, *Phys. Rev.* 85 (1952) 143.

Norman F. Ramsey (born in 1915), American physicist, professor at the University of Illinois and Columbia University, then from 1947 at the Harvard University. He is first of all an outstanding experimentalist in the domain of NMR measurements in molecular jets, but his "hobby" is theoretical physics. Ramsey carried out the first accurate measurement of the neutron magnetic moment and gave a lower bound theoretical estimation to its dipole moment. In 1989 he received the Nobel prize "*for the invention of the separated oscillatory fields method and its use in the hydrogen maser and other atomic clocks*."

Edwards Mills Purcell (1912–1997), American physicist, professor at the Massachusetts Institute of Technology and Harvard University. His main domains were relaxation phenomena and magnetic properties in low temperatures. He received the Nobel prize together with Felix Bloch "*for their development of new methods for nuclear magnetic precision measurements and discoveries in connection therewith*" In 1952.

moments. This is a relativistic effect, hence it is very small and therefore the rotational averaging results in only a small value for K_{AB}.

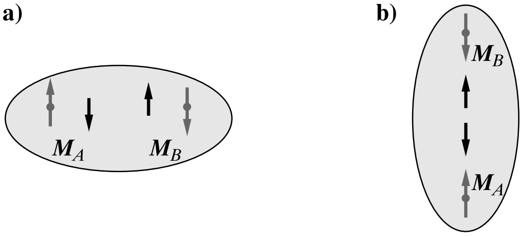


Fig. 12.13. The nuclear spin–spin coupling (Fermi contact) mechanism through chemical bond AB. The electrons repel each other and therefore correlate their motion (cf. p. 515). This is why, when one of them is close to nucleus A, the second prefers to run off to nucleus B. For some nuclei the electron–nucleus interaction of the magnetic dipole moments of A, and of the first electron near the last, will exhibit a tendency (i.e. the corresponding energy will be lower than in the opposite case) to have a spin antiparallel to the spin of A – this is what happens for protons and electrons. The second electron, close to B, must have an opposite spin to its partner, and therefore will exhibit a tendency to have its spin *the same as that of nucleus* A. We may say that the electron exposes the spin of nucleus A right at the position of the nucleus B. Such a mechanism gives a much stronger magnetic dipole interaction than that through empty space. Fig. (a) shows a favourable configuration of nuclear and electron spins, all perpendicular to the bond, Fig. (b) shows the same situation after the molecule is rotated by 90°. The electronic correlation energy will obviously differ in these two orientations of the molecule, and this results in different averaging than in the case of the interaction through space.

12.9.3 NUCLEAR SPIN ENERGY LEVELS

From calculating the mean value of the Hamiltonian (12.82), we obtain the energy of the nuclear spins in the magnetic field

$$E = -\sum_{A}(1-\sigma_A)\gamma_A H m_{I,A}\hbar + \sum_{A<B}\gamma_A\gamma_B K_{AB}\langle \hat{\boldsymbol{I}}_A \cdot \hat{\boldsymbol{I}}_B\rangle,$$

where $\langle \hat{\boldsymbol{I}}_A \cdot \hat{\boldsymbol{I}}_B\rangle$ is the mean value of the scalar product of the two spins calculated by using their spin functions. This expression can be simplified by the following transformation

$$E = -\sum_{A}(1-\sigma_A)\gamma_A H m_{I,A}\hbar + \sum_{A<B}\gamma_A\gamma_B K_{AB}\langle \hat{I}_{A,x}\hat{I}_{B,x} + \hat{I}_{A,y}\hat{I}_{B,y} + \hat{I}_{A,z}\hat{I}_{B,z}\rangle$$
$$= -\sum_{A}(1-\sigma_A)\gamma_A H m_{I,A}\hbar + \sum_{A<B}\gamma_A\gamma_B K_{AB}\left(0\cdot 0 + 0\cdot 0 + \hbar^2 m_{I,A} m_{I,B}\right),$$

because the mean values of $\hat{I}_{C,x}$ and $\hat{I}_{C,y}$ calculated for the spin functions of nucleus C both equal 0 (for the α or β functions describing a nucleus with $I_C = \frac{1}{2}$, see Chapter 1, p. 30). Therefore, the energy becomes a function of the magnetic spin quantum numbers $m_{I,C}$ for all the nuclei with a non-zero spin $\boldsymbol{I}_C$

$$E(m_{I,A}, m_{I,B}, \ldots) = -\hbar H \sum_{A}(1-\sigma_A)\gamma_A m_{I,A} + \sum_{A<B} hJ_{AB} m_{I,A} m_{I,B}, \quad (12.83)$$

coupling constant

where the commonly used nuclear spin–spin *coupling constant* is defined as

$$J_{AB} \equiv \frac{\hbar}{2\pi}\gamma_A\gamma_B K_{AB}. \quad (12.84)$$

Note that since hJ_{AB} has the dimension of the energy, then J_{AB} itself is a frequency and may be expressed in Hz.

Due to the presence of the rest of the molecule (electron shielding) the Larmor frequency $\nu_A = \frac{H\gamma_A}{2\pi}(1-\sigma_A)$ is changed by $-\sigma_A\frac{H\gamma_A}{2\pi}$ with respect to the Larmor frequency $\frac{H\gamma_A}{2\pi}$ for an isolated proton. Such changes are usually expressed (as "ppm", i.e. "parts per million"[72]) by the *chemical shift* δ_A

chemical shift

$$\delta_A = \frac{\nu_A - \nu_{\mathrm{ref}}}{\nu_{\mathrm{ref}}}\cdot 10^6 = \frac{\sigma_{\mathrm{ref}} - \sigma_A}{\sigma_{\mathrm{ref}}}\cdot 10^6, \quad (12.85)$$

where ν_{ref} is the Larmor frequency for a reference nucleus [for protons this means by convention the proton Larmor frequency in tetramethylsilane, $\mathrm{Si(CH_3)_4}$]. The chemical shifts (unlike the Larmor frequencies) are independent of the magnetic field applied.

Example 3. *The carbon nucleus in an external magnetic field*

Let us consider a single carbon ^{13}C nucleus (spin quantum number $I_C = \frac{1}{2}$) in a molecule.

---

[72] This means the chemical shift (unitless quantity) has to be multiplied by 10^{-6} to obtain $\frac{\nu_A - \nu_{\mathrm{ref}}}{\nu_{\mathrm{ref}}}$.

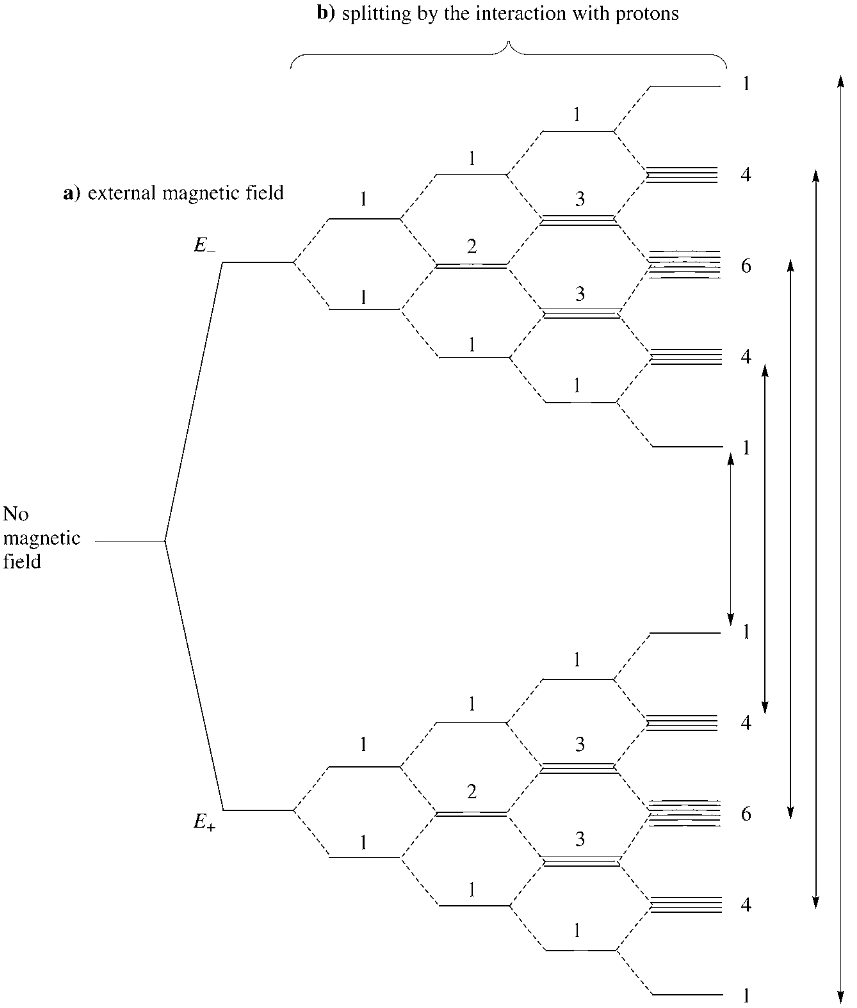


Fig. 12.14. The energy levels of the ^{13}C magnetic moment in the methane molecule and in an external magnetic field. (a) The spin energy levels of the ^{13}C atom in an external magnetic field; (b) additional interaction of the ^{13}C spin with the four equivalent proton magnetic moments switched on. As we can see the energy levels in each branch follow the Pascal triangle rule. The splits within the branch come from the coupling of the nuclei and are field-independent. The E_+ and E_- energies are field-dependent: increasing field means a tuning of the separation between the energy levels. The resonance takes place when the field-dependent energy difference matches the energy of the electromagnetic field quanta. The NMR selection rule means that only the transitions indicated take place. Since the energy split due to the coupling of the nuclei is very small, the levels E_+ are equally occupied and therefore the NMR intensities satisfy the ratio: 1 : 4 : 6 : 4 : 1.

As seen from eq. (12.83) such a nucleus in magnetic field H has two energy levels (for $m_{I,C} = \pm\frac{1}{2}$, Fig. 12.14.a)

$$E(m_{I,C}) = -\hbar H(1 - \sigma_C)\gamma_C m_{I,C},$$

where the shielding constant σ_C characterizes the vicinity of the nucleus. For the isolated nucleus $\sigma_C = 0$.

Example 4. *The methane molecule $^{13}CH_4$ in magnetic field H*

This time there is an additional magnetic field coming from four equivalent protons, each having $I_H = \frac{1}{2}$. The energy levels of the carbon magnetic spin result from the magnetic field and from the $m_{I,H}$'s of the protons according to eq. (12.83), Fig. 12.14. The resonance of the ^{13}C nucleus means a transition between energy levels that correspond to $m_{I,C} = \pm\frac{1}{2}$ with all the m_{I,H_i} being constant.[73] Thus, the lower level corresponds to

$$\begin{aligned} &E_+(m_{I,H_1}, m_{I,H_2}, m_{I,H_3}, m_{I,H_4}) \\ &\quad = -\frac{\hbar}{2}H(1-\sigma_C)\gamma_C + \frac{h}{2}\,{}^{1}J_{CH}(m_{I,H_1} + m_{I,H_2} + m_{I,H_3} + m_{I,H_4}), \end{aligned}$$

and at the higher level we have the energy

$$\begin{aligned} &E_-(m_{I,H_1}, m_{I,H_2}, m_{I,H_3}, m_{I,H_4}) \\ &\quad = \frac{\hbar}{2}H(1-\sigma_C)\gamma_C - \frac{h}{2}\,{}^{1}J_{CH}(m_{I,H_1} + m_{I,H_2} + m_{I,H_3} + m_{I,H_4}). \end{aligned}$$

Since $m_{I,H_i} = \pm\frac{1}{2}$ then each of the levels $E_\pm$ will be split into 5 levels, Fig. 12.14:

- a non-degenerate level arising from all $m_{I,H_i} = \frac{1}{2}$,
- a quadruply degenerate level that comes from all $m_{I,H_i} = \frac{1}{2}$, except one equal to $-\frac{1}{2}$ (there are four positions of this one),
- a sextuply degenerate level that results from two $m_{I,H_i} = \frac{1}{2}$ and two $m_{I,H_i} = -\frac{1}{2}$ (six ways of achieving this),
- a quadruply degenerate level that comes from all $m_{I,H_i} = -\frac{1}{2}$, except one that equals $\frac{1}{2}$ (there are four positions of this one),
- a non-degenerate level arising from all $m_{I,H_i} = -\frac{1}{2}$.

12.10 THE RAMSEY THEORY OF THE NMR CHEMICAL SHIFT

An external magnetic field $\boldsymbol{H}$ or/and the magnetic field produced by the nuclei magnetic dipole moments $\boldsymbol{M}_1, \boldsymbol{M}_2, \boldsymbol{M}_3, \ldots$ certainly represent an extremely weak perturbation to the energy of a molecule, and therefore, the changes they induce in the molecule are suitable for perturbational methods.

We decide to apply the theory through the second order. Such an effect is composed of two parts:

- the first-order correction (*the diamagnetic contribution*),
- the second-order correction (*the paramagnetic contribution*).

---

[73] The NMR selection rule for a given nucleus says that the single nucleus undergoes a flip.

The corresponding energy change due to the perturbation $\hat{\mathcal{H}}^{(1)}$ from eq. (12.68) (prime means that $k = 0$, i.e. the ground state is excluded from the summation)

$$\Delta E = E_0^{(1)} + E_0^{(2)} = \langle\psi_0^{(0)}|\hat{\mathcal{H}}^{(1)}\psi_0^{(0)}\rangle + \sum_k{}' \frac{\langle\psi_0^{(0)}|\hat{\mathcal{H}}^{(1)}\psi_k^{(0)}\rangle\langle\psi_k^{(0)}|\hat{\mathcal{H}}^{(1)}\psi_0^{(0)}\rangle}{E_0^{(0)} - E_k^{(0)}}. \tag{12.86}$$

12.10.1 SHIELDING CONSTANTS

In the Hamiltonian (12.82) the shielding constants occur in the term $\boldsymbol{I}_A \cdot \boldsymbol{H}$. The perturbation operator $\hat{H}^{(1)}$ contains a lot of terms, but most of them, when inserted into the above formula, are unable to produce terms that behave like $\boldsymbol{I}_A \cdot \boldsymbol{H}$. Only some very particular terms could produce such a dot product dependence. A minute of reflection leads directly to $\hat{B}_3$, $\hat{B}_4$, $\hat{B}_5$ and $\hat{B}_{10}$ as the only terms of the Hamiltonian that have any chance of producing the dot product form.[74] Therefore, using the definition of the reduced resolvent $\hat{R}_0$ of eq. (10.64) we have[75]

$$\begin{aligned}\Delta E = E_0^{(1)} + E_0^{(2)} &= \langle\psi_0^{(0)}|(\hat{B}_{10} + \hat{B}_5)\psi_0^{(0)}\rangle \\ &+ \langle\psi_0^{(0)}|(\hat{B}_3\hat{R}_0\hat{B}_4 + \hat{B}_4\hat{R}_0\hat{B}_3)\psi_0^{(0)}\rangle.\end{aligned} \tag{12.87}$$

After averaging the formula over rotations and extracting the proper term (details given in Appendix W, p. 1032) we obtain as the shielding constant of the nucleus A

$$\begin{aligned}\sigma_A = {}& \frac{e^2}{3mc^2}\left\langle\psi_0^{(0)}\middle|\sum_j(\boldsymbol{r}_{0j}\cdot\boldsymbol{r}_{Aj})\frac{1}{r_{Aj}^3}\psi_0^{(0)}\right\rangle - \frac{e^2}{6m^2c^2}\left\langle\psi_0^{(0)}\middle|\left[\left(\sum_j\frac{\hat{\boldsymbol{L}}_{Aj}}{r_{Aj}^3}\right)\hat{R}_0\left(\sum_j\hat{\boldsymbol{L}}_{0j}\right)\right.\right. \\ & \left.\left. + \left(\sum_j\hat{\boldsymbol{L}}_{0j}\right)\hat{R}_0\left(\sum_j\frac{\hat{\boldsymbol{L}}_{Aj}}{r_{Aj}^3}\right)\right]\psi_0^{(0)}\right\rangle,\end{aligned} \tag{12.88}$$

---

[74] There is an elegant way to single out the only necessary B_i's that give a contribution to the energy proportional to the product $x_i x_j$ (no higher terms included), where x_i and x_j stand for some components of the magnetic field intensity $\boldsymbol{H}$ or of the nuclear spin $\boldsymbol{I}_A$'s (that cause perturbation of the molecule). As to the first-order correction ("diamagnetic") we calculate the second derivative $(\frac{\partial^2\hat{\mathcal{H}}^{(1)}}{\partial x_i \partial x_i})_{\boldsymbol{H}=\boldsymbol{0},\boldsymbol{I}_i=\boldsymbol{0}}$ of the Hamiltonian $\hat{\mathcal{H}}^{(1)}$ with respect to the components of $\boldsymbol{H}$ or $\boldsymbol{I}_A$, afterwards inserting $\boldsymbol{H} = \boldsymbol{0}$ and $\boldsymbol{I}_A = \boldsymbol{0}$ (i.e. calculating the derivative at zero perturbation). Then the diamagnetic correction to the energy is $\langle\psi_0^{(0)}|(\frac{\partial^2\hat{\mathcal{H}}^{(1)}}{\partial x_i \partial x_i})_{\boldsymbol{H}=\boldsymbol{0},\boldsymbol{I}_i=\boldsymbol{0}}\psi_0^{(0)}\rangle$. As to the second-order correction ("paramagnetic"), we calculate the first derivatives: $(\frac{\partial\hat{\mathcal{H}}^{(1)}}{\partial x_i})_{\boldsymbol{H}=\boldsymbol{0},\boldsymbol{I}_i=\boldsymbol{0}}$ and $(\frac{\partial\hat{\mathcal{H}}^{(1)}}{\partial x_j})_{\boldsymbol{H}=\boldsymbol{0},\boldsymbol{I}_i=\boldsymbol{0}}$ and, therefore, the contribution to the energy is

$$\sum_k{}' \frac{\langle\psi_0^{(0)}|(\frac{\partial\hat{\mathcal{H}}^{(1)}}{\partial x_i})_{\boldsymbol{H}=\boldsymbol{0},\boldsymbol{I}_i=\boldsymbol{0}}\psi_k^{(0)}\rangle\langle\psi_k^{(0)}|(\frac{\partial\hat{\mathcal{H}}^{(1)}}{\partial x_j})_{\boldsymbol{H}=\boldsymbol{0},\boldsymbol{I}_i=\boldsymbol{0}}\psi_0^{(0)}\rangle}{E_0^{(0)} - E_k^{(0)}}.$$

[75] Note that whenever the reduced resolvent appears in a formula, infinite summation over unperturbed states is involved.

where

$$\hat{\boldsymbol{L}}_{Aj} = -i\hbar(\boldsymbol{r}_{Aj} \times \nabla_j) \tag{12.89}$$

and

$$\hat{\boldsymbol{L}}_{0j} = -i\hbar(\boldsymbol{r}_{0j} \times \nabla_j) \tag{12.90}$$

stand for the angular momenta operators for the electron j calculated with respect to the position of nucleus A and with respect the origin of vector potential $\boldsymbol{A}$, respectively.

12.10.2 DIAMAGNETIC AND PARAMAGNETIC CONTRIBUTIONS

The result (12.88) has been obtained (apparently) in two parts:

$$\sigma_A = \sigma_A^{\text{dia}} + \sigma_A^{\text{para}}, \tag{12.91}$$

called the *diamagnetic contribution*

$$\sigma_A^{\text{dia}} = \frac{e^2}{3mc^2}\left\langle \psi_0^{(0)} \middle| \sum_j (\boldsymbol{r}_{0j} \cdot \boldsymbol{r}_{Aj}) \frac{1}{r_{Aj}^3} \psi_0^{(0)} \right\rangle$$

and the *paramagnetic contribution*

$$\begin{aligned}\sigma_A^{\text{para}} = -\frac{e^2}{6m^2c^2}\Bigg\langle \psi_0^{(0)} \Bigg| \Bigg[\bigg(\sum_j \frac{\hat{\boldsymbol{L}}_{Aj}}{r_{Aj}^3} \bigg) \hat{R}_0 \bigg(\sum_j \hat{\boldsymbol{L}}_{0j} \bigg) \\ + \bigg(\sum_j \hat{\boldsymbol{L}}_{0j} \bigg) \hat{R}_0 \bigg(\sum_j \frac{\hat{\boldsymbol{L}}_{Aj}}{r_{Aj}^3} \bigg) \Bigg] \psi_0^{(0)} \Bigg\rangle.\end{aligned}$$

Each of these contributions looks suspicious. Indeed, the diamagnetic contribution explicitly depends on the choice of origin $\boldsymbol{R}$ of vector potential $\boldsymbol{A}$ through $\boldsymbol{r}_{0j} = \boldsymbol{r}_j - \boldsymbol{R}$, see (12.65). Similarly, the paramagnetic contribution also depends on choice through $\hat{\boldsymbol{L}}_{0j}$ and (12.65). We have already stressed the *practical* importance of the choice of $\boldsymbol{R}$ in Appendix G. Since both contributions depend on the choice, they separately cannot have any physical significance.

Is it possible that the *sum* of the two contributions is invariant with respect to choice of $\boldsymbol{R}$? Yes, it is! The invariance has fortunately been proved.[76] This is good, because any measurable quantity cannot depend on an arbitrary choice of the origin of the coordinate system.

12.11 THE RAMSEY THEORY OF NMR SPIN–SPIN COUPLING CONSTANTS

We will apply the same philosophy as that used for shielding constants (or, equivalently, of the NMR chemical shift) to calculate the nuclear coupling constant.

---

[76] A. Abragam, "*The Principles of Nuclear Magnetism*", Clarendon Press, Oxford (1961).

Taking into account the Hamiltonian $\hat{\mathcal{H}}^{(1)}$ from eq. (12.68), we note that the only terms in $\hat{\mathcal{H}}^{(1)}$ that have the chance to contribute to the NMR coupling constants (see eq. (12.83)) are

$$\Delta E = E_0^{(1)} + E_0^{(2)} = \langle \psi_0^{(0)} | (\hat{B}_1 + \hat{B}_9)\psi_0^{(0)} \rangle \tag{12.92}$$

$$+ \sum_k{}' \frac{\langle \psi_0^{(0)} | (\hat{B}_3 + \hat{B}_6 + \hat{B}_7)\psi_k^{(0)} \rangle \langle \psi_k^{(0)} | (\hat{B}_3 + \hat{B}_6 + \hat{B}_7)\psi_0^{(0)} \rangle}{E_0^{(0)} - E_k^{(0)}}$$

$$= E_{\text{dia}} + E_{\text{para}}, \tag{12.93}$$

because we are looking for terms that could result in the scalar product of the nuclear magnetic moments. The first term is the diamagnetic contribution (E_{dia}), the sum is the paramagnetic contribution (E_{para}).

12.11.1 DIAMAGNETIC CONTRIBUTIONS

There are two diamagnetic contributions in the total diamagnetic effect $\langle \psi_0^{(0)} | (\hat{B}_1 + \hat{B}_9)\psi_0^{(0)} \rangle$:

- The $\langle \psi_0^{(0)} | \hat{B}_9 \psi_0^{(0)} \rangle$ term simply represents the $\sum_{A<B} \gamma_A \gamma_B \boldsymbol{I}_A^T \boldsymbol{D}_{AB} \boldsymbol{I}_B$ contribution of eq. (12.80), i.e. the *direct ("through space") nuclear spin–spin interaction.* This calculation does not require anything except summation over spin–spin terms. However, *as has been shown, averaging over free rotations of the molecule in the specimen renders this term equal to zero.*
- The $\langle \psi_0^{(0)} | \hat{B}_1 \psi_0^{(0)} \rangle$ term can be transformed in the following way:

direct spin–spin contribution

$$\langle \psi_0^{(0)} | \hat{B}_1 \psi_0^{(0)} \rangle = \left\langle \psi_0^{(0)} \middle| \left(\frac{e^2}{2mc^2} \sum_{A,B} \sum_j \gamma_A \gamma_B \frac{\boldsymbol{I}_A \times \boldsymbol{r}_{Aj}}{r_{Aj}^3} \frac{\boldsymbol{I}_B \times \boldsymbol{r}_{Bj}}{r_{Bj}^3} \right) \psi_0^{(0)} \right\rangle$$

$$= \frac{e^2}{2mc^2} \sum_{A,B} \sum_j \gamma_A \gamma_B \left\langle \psi_0^{(0)} \middle| \frac{(\boldsymbol{I}_A \times \boldsymbol{r}_{Aj}) \cdot (\boldsymbol{I}_B \times \boldsymbol{r}_{Bj})}{r_{Aj}^3 r_{Bj}^3} \psi_0^{(0)} \right\rangle.$$

Now, note that $(\boldsymbol{A} \times \boldsymbol{B}) \cdot \boldsymbol{C} = \boldsymbol{A} \cdot (\boldsymbol{B} \times \boldsymbol{C})$. Taking $\boldsymbol{A} = \boldsymbol{I}_A$, $\boldsymbol{B} = \boldsymbol{r}_{Aj}$, $\boldsymbol{C} = \boldsymbol{I}_B \times \boldsymbol{r}_{Bj}$ we first have the following

$$\langle \psi_0^{(0)} | \hat{B}_1 \psi_0^{(0)} \rangle = \frac{e^2}{2mc^2} \sum_{A,B} \sum_j \gamma_A \gamma_B \boldsymbol{I}_A \cdot \left\langle \psi_0^{(0)} \middle| \frac{\boldsymbol{r}_{Aj} \times (\boldsymbol{I}_B \times \boldsymbol{r}_{Bj})}{r_{Aj}^3 r_{Bj}^3} \psi_0^{(0)} \right\rangle.$$

Recalling that $\boldsymbol{A} \times (\boldsymbol{B} \times \boldsymbol{C}) = \boldsymbol{B}(\boldsymbol{A} \cdot \boldsymbol{C}) - \boldsymbol{C}(\boldsymbol{A} \cdot \boldsymbol{B})$ this term (called the *diamagnetic spin–orbit contribution, DSO*[77]) reads as

diamagnetic spin–orbit

$$E_{\text{DSO}} = \frac{e^2}{2mc^2} \sum_{A,B} \sum_j \gamma_A \gamma_B \boldsymbol{I}_A \cdot \left[\boldsymbol{I}_B \left\langle \psi_0^{(0)} \middle| \frac{\boldsymbol{r}_{Aj} \cdot \boldsymbol{r}_{Bj}}{r_{Aj}^3 r_{Bj}^3} \psi_0^{(0)} \right\rangle - \left\langle \psi_0^{(0)} \middle| \boldsymbol{r}_{Bj} \frac{\boldsymbol{r}_{Aj} \cdot \boldsymbol{I}_B}{r_{Aj}^3 r_{Bj}^3} \psi_0^{(0)} \right\rangle \right].$$

---

[77] The name comes, of course, from the nuclear spin–electronic orbit interaction.

We see that we need to calculate some integrals with mono-electronic operators, which is an easy task.

12.11.2 PARAMAGNETIC CONTRIBUTIONS

The paramagnetic contribution E_{para} to the energy resulting from the perturbation given in eq. (12.68) can be written in a simpler form using the reduced resolvent $\hat{R}_0$ of eq. (10.62):

$$\begin{aligned}E_{\text{para}} &= \sum_k{}' \frac{\langle\psi_0^{(0)}|(\hat{B}_3+\hat{B}_6+\hat{B}_7)\psi_k^{(0)}\rangle\langle\psi_k^{(0)}|(\hat{B}_3+\hat{B}_6+\hat{B}_7)\psi_0^{(0)}\rangle}{E_0^{(0)}-E_k^{(0)}}\\ &= \langle\psi_0^{(0)}\big|(\hat{B}_3+\hat{B}_6+\hat{B}_7)\hat{R}_0(\hat{B}_3+\hat{B}_6+\hat{B}_7)\psi_0^{(0)}\rangle\\ &= E_{\text{PSO}}+E_{\text{SD}}+E_{\text{FC}}+\text{mixed terms,}\end{aligned}$$

where

paramagnetic spin–orbit

- the *paramagnetic spin–orbit contribution*:

$$E_{\text{PSO}} = \langle\psi_0^{(0)}\big|\hat{B}_3\hat{R}_0\hat{B}_3\psi_0^{(0)}\rangle$$

with $\hat{B}_3$ meaning the interaction between the *nuclear spin* magnetic moment and the magnetic moment resulting from the *electronic angular momenta* of the individual electrons in an atom,

spin–dipole

- the *spin–dipole contribution*

$$E_{\text{SD}} = \langle\psi_0^{(0)}\big|\hat{B}_6\hat{R}_0\hat{B}_6\psi_0^{(0)}\rangle,$$

which describes the interaction energy of the magnetic spin dipoles: the nuclear with the electronic dipole,

Fermi contact

- the *Fermi contact interaction*

$$E_{\text{FC}} = \langle\psi_0^{(0)}\big|\hat{B}_7\hat{R}_0\hat{B}_7\psi_0^{(0)}\rangle,$$

which is related to the electronic spin–nuclear spin interaction with zero distance between them.

- the mixed terms contain $\langle\psi_0^{(0)}|\hat{B}_i\hat{R}_0\hat{B}_j\psi_0^{(0)}\rangle$ for $i,j=3,6,7$ and $i\neq j$. These terms are either exactly zero or (in most cases, not always) small.[78]

---

[78] Let us consider all cross terms. First, let us check that $\langle\psi_0^{(0)}|\hat{B}_3\hat{R}_0\hat{B}_6\psi_0^{(0)}\rangle = \langle\psi_0^{(0)}|\hat{B}_6\hat{R}_0\hat{B}_3\psi_0^{(0)}\rangle = \langle\psi_0^{(0)}|\hat{B}_3\hat{R}_0\hat{B}_7\psi_0^{(0)}\rangle = \langle\psi_0^{(0)}|\hat{B}_7\hat{R}_0\hat{B}_3\psi_0^{(0)}\rangle = 0$. Note, that $\hat{B}_6$ and $\hat{B}_7$ both contain electron spin operators, while $\hat{B}_3$ does not. Let us assume that, as it is usually the case, $\psi_0^{(0)}$ is a singlet function. Recalling eq. (10.62) this implies that for $\langle\psi_0^{(0)}|\hat{B}_3\psi_k^{(0)}\rangle$ to survive, the function $\psi_k^{(0)}$ has to have the same multiplicity as $\psi_0^{(0)}$. This however kills the other factors: $\langle\psi_0^{(0)}|\hat{B}_6\psi_k^{(0)}\rangle$ and $\langle\psi_0^{(0)}|\hat{B}_7\psi_k^{(0)}\rangle$, terms describing the magnetic interaction of nuclei with exactly the same role played by electrons with α and β spins. Thus, the products $\langle\psi_0^{(0)}|\hat{B}_3\psi_k^{(0)}\rangle\langle\psi_k^{(0)}|\hat{B}_6\psi_0^{(0)}\rangle$ and $\langle\psi_0^{(0)}|\hat{B}_3\psi_k^{(0)}\rangle\langle\psi_k^{(0)}|\hat{B}_7\psi_0^{(0)}\rangle$ are zero.

12.11.3 COUPLING CONSTANTS

The energy contributions have to be averaged over rotations of the molecule and the coupling constants are to be extracted from the resulting formulae. How this is performed is shown in Appendix W on p. 1032.

Finally, the nuclear spin–spin coupling constant is calculated as the sum of the diamagnetic (J_{AB}^{DSO}) and paramagnetic contributions ($J_{AB}^{\mathrm{PSO}}, J_{AB}^{\mathrm{SD}}, J_{AB}^{\mathrm{FC}}, J_{AB}^{\mathrm{mixed}}$)

$$J_{AB} = J_{AB}^{\mathrm{dia}} + J_{AB}^{\mathrm{para}}, \tag{12.94}$$

$$J_{AB}^{\mathrm{dia}} \equiv J_{AB}^{\mathrm{DSO}}, \tag{12.95}$$

$$J_{AB}^{\mathrm{para}} = J_{AB}^{\mathrm{PSO}} + J_{AB}^{\mathrm{SD}} + J_{AB}^{\mathrm{FC}} + J_{AB}^{\mathrm{mixed}}, \tag{12.96}$$

where the particular contributions to the coupling constant are:[79]

$$J_{AB}^{\mathrm{DSO}} = \frac{e^2\hbar}{3\pi mc^2}\gamma_A\gamma_B \sum_j \left\langle \psi_0^{(0)} \middle| \frac{\boldsymbol{r}_{Aj}\cdot\boldsymbol{r}_{Bj}}{r_{Aj}^3 r_{Bj}^3}\psi_0^{(0)} \right\rangle,$$

$$J_{AB}^{\mathrm{PSO}} = \frac{1}{3\pi}\hbar\left(\frac{e}{mc}\right)^2 \gamma_A\gamma_B \sum_{j,l,Aj} \langle \psi_0^{(0)} | \hat{\boldsymbol{L}}_{Aj}\hat{R}_0\hat{\boldsymbol{L}}_{Bl}\psi_0^{(0)}\rangle,$$

$$J_{AB}^{\mathrm{SD}} = \frac{1}{3\pi}\hbar\gamma_{\mathrm{el}}^2\gamma_A\gamma_B \times \sum_{j,l=1}^{N} \left\langle \psi_0^{(0)} \middle| \left[\frac{\hat{\boldsymbol{s}}_j}{r_{Aj}^3} - 3\frac{(\hat{\boldsymbol{s}}_j\cdot\boldsymbol{r}_{Aj})\boldsymbol{r}_{Aj}}{r_{Aj}^5}\right]\hat{R}_0\left[\frac{\hat{\boldsymbol{s}}_l}{r_{Bl}^3} - 3\frac{(\hat{\boldsymbol{s}}_l\cdot\boldsymbol{r}_{Bl})(\boldsymbol{r}_{Bl})}{r_{Bl}^5}\right]\psi_0^{(0)} \right\rangle,$$

$$J_{AB}^{\mathrm{FC}} = \frac{1}{3\pi}\hbar\left(\frac{8\pi}{3}\right)^2 \gamma_{\mathrm{el}}^2\gamma_A\gamma_B \sum_{j,l=1} \langle \psi_0^{(0)} | \delta(\boldsymbol{r}_{Aj})\hat{\boldsymbol{s}}_j\hat{R}_0\delta(\boldsymbol{r}_{Bl})\hat{\boldsymbol{s}}_l\psi_0^{(0)}\rangle.$$

Thus, the nuclear spin magnetic moments are coupled *via* their magnetic interaction with the *electronic* magnetic moments:

- the $J_{AB}^{\mathrm{DSO}} + J_{AB}^{\mathrm{PSO}}$ results from the electronic *orbital* magnetic dipole moments,
- while $J_{AB}^{\mathrm{SD}} + J_{AB}^{\mathrm{FC}}$ corresponds to such interactions with the electronic *spin* magnetic dipole moments.

The mixed term $\langle\psi_0^{(0)}|\hat{B}_6\hat{R}_0\hat{B}_7\psi_0^{(0)}\rangle$ vanishes for the isotropic electron cloud around the nucleus, because in the product $\langle\psi_0^{(0)}|\hat{B}_6\psi_k^{(0)}\rangle\langle\psi_k^{(0)}|\hat{B}_7\psi_0^{(0)}\rangle$ the Fermi term $\langle\psi_k^{(0)}|\hat{B}_7\psi_0^{(0)}\rangle$ survives, if $\psi_0^{(0)}\psi_k^{(0)}$ calculated at the nucleus is non-zero. This kills, however, $\langle\psi_0^{(0)}|\hat{B}_6\psi_k^{(0)}\rangle$, because for $\psi_0^{(0)}\psi_k^{(0)} \neq 0$ (which as a rule comes from a 1*s* orbital, this is isotropic situation) the electron–nucleus dipole–dipole magnetic interaction averages to zero when different positions of the electron are considered. For non-isotropic cases this mixed contribution can be of importance.

[79] The empirical Hamiltonian (12.83) contains only the $A > B$ contributions, therefore the factor 2 appears in J.

As to the integrals involved, the Fermi contact contribution J_{AB}^{FC} (just the value of the wave function at the nucleus position) is the easiest to compute. Assuming that $\psi_k^{(0)}$ states are Slater determinants, the diamagnetic spin–orbit contribution J_{AB}^{DSO} requires some (easy) one-electron integrals of the type $\langle\psi_1|\frac{x_{Aj}}{r_{Aj}^3}\psi_2\rangle$, the paramagnetic spin–orbit contribution J_{AB}^{PSO} needs some one-electron integrals involving $\hat{\boldsymbol{L}}_{Aj}$ operators, which require differentiation of the orbitals, the spin–dipole contribution J_{AB}^{SD} leads also to some simple one-electron integrals, but handling the spin operators is needed (see p. 28), as for J_{AB}^{FC}. All the formulae require an infinite summation over states (due to the presence of $\hat{R}_0$), which is very tedious. This is why, in contemporary computational technique, some other approaches, mainly what is called propagator theory, are used.[80]

12.11.4 THE FERMI CONTACT COUPLING MECHANISM

There are no simple rules, but usually the most important contribution to J_{AB} comes from the Fermi contact term (J_{AB}^{FC}), the next most important is paramagnetic spin-orbit term J_{AB}^{PSO}, other terms, including the mixed contributions J_{AB}^{mixed}, are of small importance. Let us consider the Fermi contact coupling mechanism between two protons through a single bond (the coupling constant J_{AB} denoted as ${}^1J_{HH}$). The proton and the electron close to it prefer to have opposite spins. Then the other electron of the bond (being closer to the other nucleus) shows the other nucleus the spin of first nucleus, therefore the second nucleus prefers to have the opposite spin with respect to the first nucleus. According to eq. (12.83), since $m_{I,A}m_{I,B} < 0$, this means $J_{AB} \equiv {}^1J_{HH} > 0$. What about ${}^2J_{HH}$? This time, to have a through-bond interaction we have to have a central atom, like carbon ${}^{12}\mathrm{C}$ (i.e. with zero magnetic moment), Fig. 12.15. The key point now is what happens at the central atom: whether it is preferable to have on it two parallel or two antiparallel electron spins? We do not know, but we may have a suggestion. Hund's rule says that, in case of orbital degeneracy (in our case: this corresponds to two *equivalent* C–H bonds), the electrons prefer to have parallel spins. This suggests that the two distant proton spins have a negative coupling constant, i.e. ${}^2J_{HH} < 0$, which is indeed the case. The same argument suggests that ${}^3J_{HH} > 0$, etc.[81]

---

[80] J. Linderberg, Y. Öhrn, "*Propagators in Quantum Chemistry*", 2nd edition, John Wiley & Sons, Ltd, 2004.

[81] Thus, although calculation of the coupling constants is certainly complex, we have in mind a simple model of the nuclear spin–spin interaction that seems to work. We love such models, because they enable us to predict numbers knowing other numbers, or to predict new phenomena. This gives the impression that we understand what happens. This is by no means true. What the electrons are doing and how the spin magnetic moments interact is too complicated, but nevertheless we may suspect the main principles of the game. Such models help us to discuss things with others, to communicate some conjectures, to verify them, and to get more and more confidence in ourselves. Until one day something goes wrong. Then we try to understand why it happened. This *may* require a revision of our model, i.e. a new model, etc.

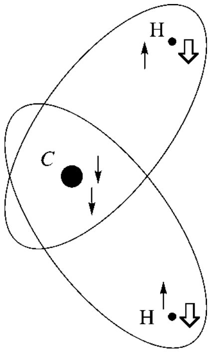


Fig. 12.15. Is the proton–proton coupling constant through two bonds (H–C–H), i.e. is $^2J_{HH}$ positive or negative? Please recall that $^1J_{HH} > 0$, see Fig. 12.13a, where the induction mechanism is described. The interaction of proton spins (wide arrows) through two bonds depends on what happens at the central carbon atom: are the spins of the two electrons there (one from each bond C–H) parallel or antiparallel? Hund's rule suggests they prefer to be parallel. This means that the situation with the two proton spins parallel is more energetically favourable, and this means $^2J_{HH} < 0$. This rule of thumb may fail when the carbon atom participates in multiple bonds, as in ethylene, see section "From the research front".

12.12 GAUGE INVARIANT ATOMIC ORBITALS (GIAO)

The coupling constants in practical applications may depend on the choice of vector and scalar potentials. The arbitrariness in the choice of the potentials $\boldsymbol{A}$ and ϕ ("gauge choice") does not represent any problem for an *atom*, because it is natural to locate the origin (related to the formula (G.13) on p. 964) on the nucleus. The same reasoning however means a serious problem for a molecule, because even though any choice is equally justified, this justification is only theoretical, not practical. Should the origin be chosen at the centre of mass, at the centre of the electron cloud, halfway between them, or at another point? *An unfortunate (although mathematically fully justified) choice of the vector potential origin would lead to correct results, but only after calculating and summing up all the contributions to infinity, including application of the complete set of atomic orbitals*. These requirements are too demanding.

12.12.1 LONDON ORBITALS

Atomic orbitals are used in quantum chemistry as the building blocks of many-electron functions (cf. p. 357). Where to centre the orbitals sometimes represents a serious problem. On top of this, in the case of a magnetic field, there is, additionally, the above mentioned arbitrariness of choice of the vector potential origin. A remedy to the second problem was found by Fritz London[82] in the form of atomic orbitals that depend explicitly on the field applied. Each atomic orbital $\chi(\boldsymbol{r}-\boldsymbol{R}_C)$ centred on nucleus C (with position shown by vector $\boldsymbol{R}_C$) and describing an electron pointed by vector $\boldsymbol{r}$, is replaced by the *London orbital* in the following way

London orbital

LONDON ATOMIC ORBITALS

$$\chi_L(\boldsymbol{r}-\boldsymbol{R}_C;\boldsymbol{A}_C)=\exp(-i\boldsymbol{A}_C\cdot\boldsymbol{r})\chi(\boldsymbol{r}-\boldsymbol{R}_C), \tag{12.97}$$

---

[82] F. London, *J. Phys. Radium* 8 (1937) 397.

where A_C stands for the value of vector field A at nucleus C, and A corresponds to the origin O according to formula (G.13) on p. 964, where H denotes the intensity of a homogeneous magnetic field (no contribution from the magnetic field created by the nuclei, etc.).

> As seen, the London orbitals are not invariant with respect to the choice of vector potential origin,

e.g., with respect to shifting the origin of the coordinate system in formula (G.13) by vector $\boldsymbol{R}$:

$$\boldsymbol{A}'(\boldsymbol{r}) = \frac{1}{2}\big[\boldsymbol{H} \times (\boldsymbol{r} - \boldsymbol{R})\big] = \boldsymbol{A}(\boldsymbol{r}) - \frac{1}{2}[\boldsymbol{H} \times \boldsymbol{R}]. \tag{12.98}$$

Indeed,

$$\begin{aligned}\chi_L(\boldsymbol{r} - \boldsymbol{R}_C; \boldsymbol{A}'_C) &= \exp(-i\boldsymbol{A}'_C \cdot \boldsymbol{r})\chi(\boldsymbol{r} - \boldsymbol{R}_C) \\ &= \exp(-i\boldsymbol{A}_C \cdot \boldsymbol{r})\exp\left(i\frac{1}{2}[\boldsymbol{H} \times \boldsymbol{R}] \cdot \boldsymbol{r}\right)\chi(\boldsymbol{r} - \boldsymbol{R}_C) \\ &= \exp\left(i\frac{1}{2}[\boldsymbol{H} \times \boldsymbol{R}] \cdot \boldsymbol{r}\right)\chi_L(\boldsymbol{r} - \boldsymbol{R}_C; \boldsymbol{A}_C) \neq \chi_L(\boldsymbol{r} - \boldsymbol{R}_C; \boldsymbol{A}_C).\end{aligned}$$

Despite this property the London orbitals are also known as Gauge Invariant Atomic Orbitals (GIAO).

12.12.2 INTEGRALS ARE INVARIANT

Let us calculate the overlap integral S between two London orbitals centred at points C and D. After shifting the origin of the coordinate system in eq. (G.13) by vector $\boldsymbol{R}$ we get

$$\begin{aligned}S &= \big\langle \chi_{L,1}\big(\boldsymbol{r} - \boldsymbol{R}_C; \boldsymbol{A}'_C\big)\big|\chi_{L,2}\big(\boldsymbol{r} - \boldsymbol{R}_D; \boldsymbol{A}'_D\big)\big\rangle \\ &= \big\langle \exp\big(-i\boldsymbol{A}'_C \cdot \boldsymbol{r}\big)\chi_1\big|\exp\big(-i\boldsymbol{A}'_D \cdot \boldsymbol{r}\big)\chi_2\big\rangle \\ &= \big\langle \chi_1\big|\exp\big(-i\big(\boldsymbol{A}'_D - \boldsymbol{A}'_C\big) \cdot \boldsymbol{r}\big)\chi_2\big\rangle,\end{aligned}$$

i.e. *the result is independent of* $\boldsymbol{R}$. It turns out[83] that all the integrals needed – those of kinetic energy, nuclear attraction and electron repulsion (cf. Appendix P on p. 1004) – are invariant with respect to an arbitrary shift of the origin of vector potential $\boldsymbol{A}$.

> This means that when we use the London orbitals the results do not depend on the choice of vector potential origin.

Summary

- The Hellmann–Feynman theorem tells us about the rate the energy changes, when we increase parameter P in the Hamiltonian (e.g., the intensity of the electric field). This

---

[83] T. Helgaker, P. Jørgensen, *J. Chem. Phys.* 95 (1991) 2595.

rate is $\frac{\partial E}{\partial P} = \langle \psi(P)|\frac{\partial \hat{H}}{\partial P}|\psi(P)\rangle$, where $\psi(P)$ means the *exact* solution to the Schrödinger equation [with energy $E(P)$] at value P of the parameter.

ELECTRIC PHENOMENA

- When a molecule is located in an inhomogeneous electric field the perturbation operator has the form $\hat{H}^{(1)} = -\sum_q \hat{\mu}_q \mathcal{E}_q - \frac{1}{3}\sum_{qq'} \hat{\Theta}_{qq'} \mathcal{E}_{qq'} \cdots$, where $\mathcal{E}_q$ for $q = x, y, z$ denote the electric field components along the corresponding axes of a Cartesian coordinate system, $\mathcal{E}_{qq'}$ is the q' component of the gradient of $\mathcal{E}_q$, while $\hat{\mu}_q$, $\hat{\Theta}_{qq'}$ stand for the operators of the corresponding components of the dipole and quadrupole moments. In a homogeneous electric field ($\mathcal{E}_{qq'} = 0$) this reduces to $\hat{H}^{(1)} = -\sum_q \hat{\mu}_q \mathcal{E}_q$.
- After using the last expression in the Hellmann–Feynman theorem we obtain the dependence of the dipole moment components on the (weak) field intensity: $\mu_q = \mu_{0q} + \sum_{q'} \alpha_{qq'} \mathcal{E}_{q'} + \frac{1}{2}\sum_{qq'} \beta_{qq'q''} \mathcal{E}_q \mathcal{E}_{q'} \mathcal{E}_{q''} + \cdots$, where μ_{0q} stands for the component corresponding to the isolated molecule, $\alpha_{qq'}$ denotes the q, q' component of the (dipole) polarizability tensor, $\beta_{qq'q''}$ is the corresponding component of the (dipole) first hyperpolarizability tensor, etc. The quantities μ_{0q}, $\alpha_{qq'}$, $\beta_{qq'q''}$ in a given Cartesian coordinate system characterize the *isolated* molecule (no electric field) and represent the target of the calculation methods.
- Reversing the electric field direction may in general give different absolute values of the induced dipole moment only because of non-zero hyperpolarizability $\beta_{qq'q''}$ and higher-order hyperpolarizabilities.
- In an inhomogeneous field we have the following interactions:
  - of the *permanent dipole moment* of the molecule with the electric field $-\boldsymbol{\mu}_0 \boldsymbol{\mathcal{E}}$,
  - of the *induced dipole moment* proportional to the field ($\sum_{q'} \alpha_{qq'} \mathcal{E}_{q'}$) with the field plus higher-order terms proportional to higher powers of the field intensity involving *dipole hyperpolarizabilities*,
  - of the *permanent quadrupole moment* $\Theta_{qq'}$ of the molecule with the field gradient: $-\frac{1}{3}\sum_{qq'} \Theta_{qq'} \mathcal{E}_{qq'}$,
  - of the *induced quadrupole moment* proportional to the field gradient with the field gradient ($-\frac{1}{4}\sum_{qq'q''q'''} C_{qq'q''q'''} \mathcal{E}_{qq'} \mathcal{E}_{q''q'''}$, the quantity C is called the *quadrupole polarizability*) + higher-order terms containing *quadrupole hyperpolarizabilities*,
  - higher multipole interactions.
- In the LCAO MO approximation, the dipole moment of the molecule can be divided into the sum of the atomic dipole moments and the dipole moments of the atomic pairs.
- The dipole polarizability may be computed by:
  - the *Sum Over States method* (SOS), which is based on second-order correction to the energy in the perturbational approach;
  - the *finite field method*, e.g., a variational approach in which the interaction with a weak homogeneous electric field is included in the Hamiltonian. The components of the polarizability are computed as the second derivatives of the energy with respect to the corresponding field components (the derivatives are calculated at zero field). In practical calculations within the LCAO MO approximation we often use the Sadlej relation that connects the shift of a Gaussian atomic orbital with its exponent and the electric field intensity.
- In laser fields we may obtain a series of non-linear effects (proportional to higher powers of field intensity), including the doubling and tripling of the incident light frequency.

MAGNETIC PHENOMENA

- An elementary particle has a magnetic dipole moment $\boldsymbol{M}$ proportional to its spin angular momentum $\boldsymbol{I}$, i.e. $\boldsymbol{M} = \gamma \boldsymbol{I}$, where γ stands for what is called the gyromagnetic factor (characteristic for the kind of particle).
- The magnetic dipole of a particle with spin $\boldsymbol{I}$ (corresponding to spin quantum number I) in homogeneous magnetic field $\boldsymbol{H}$ has $2I+1$ energy states $E_{m_I} = -\gamma m_I \hbar H$, where $m_I = -I, -I+1, \dots, +I$. Thus, the energy is proportional to H.
- The Hamiltonian of a system in an electromagnetic field has the form

$$\hat{H} = \sum_{j=1} \left[\frac{1}{2m_j} \left(\hat{\boldsymbol{p}}_j - \frac{q_i}{c} \boldsymbol{A}_j \right)^2 + q_j \phi_j \right] + \hat{V},$$

  where $\boldsymbol{A}_j$ and ϕ_j denote the vector and scalar fields at particle j (both are functions of position in the 3D space) that characterize the external electromagnetic field.
- $\boldsymbol{A}$ and ϕ potentials contain, in principle (see Appendix G), the same information as the magnetic and electric field $\boldsymbol{H}$ and $\boldsymbol{\mathcal{E}}$. There is an arbitrariness in the choice of $\boldsymbol{A}$ and ϕ.
- In order to calculate the energy states of a system of nuclei (detectable in NMR spectroscopy) we have to use the Hamiltonian $\hat{H}$ given above, supplemented by the interaction of all magnetic moments related to the orbital and spin of the electrons and the nuclei.
- The refinement is based on classical electrodynamics and the usual quantum mechanical rules for forming operators (Chapter 1) or, alternatively, on the relativistic Breit Hamiltonian (p. 131). This is how we get the Hamiltonian (12.66) which contains the usual non-relativistic Hamiltonian (12.67) plus the perturbation (12.68) with a number of terms (p. 657).
- Experimentalists use an empirical Hamiltonian (eq. (12.82)), in which they have the interaction of the nuclear spin magnetic moments with the magnetic field (the Zeeman effect), the latter weakened by the shielding of the nuclei by the electrons plus the dot products of the nuclear magnetic moments weighted by the coupling constants. The experiment gives both the shielding (σ_A) and the coupling (J_{AB}) constants.
- Nuclear spin coupling takes place through the induction mechanism in the chemical bond (cf. Figs. 12.13, 12.15). Of key importance for this induction is high electron density at the position of the nuclei (the so called Fermi contact term, Fig. 12.13).
- The theory of shielding and coupling constants was given by Ramsey. According to the theory, each quantity consists of diamagnetic and paramagnetic contributions. The diamagnetic term is easy to calculate, the paramagnetic one is more demanding.
- Each of the contributions to the shielding constant individually depends on the choice of the origin of the vector potential $\boldsymbol{A}$, while their sum is invariant with respect to this choice.
- The London atomic orbitals $\chi_L = \exp(-i\boldsymbol{A}_C \cdot \boldsymbol{r})\chi(\boldsymbol{r} - \boldsymbol{R}_C)$ used in calculations for a molecule in a magnetic field depend explicitly on that field, through the value $\boldsymbol{A}_C$ of the vector potential $\boldsymbol{A}$ calculated at the centre $\boldsymbol{R}_C$ of the usual atomic orbital $\chi(\boldsymbol{r} - \boldsymbol{R}_C)$.
- The most important feature of London orbitals is that all the integrals appearing in calculations are invariant with respect to the origin of the vector potential. This is why results obtained using London orbitals are also independent of that choice.

Main concepts, new terms

Hellmann–Feynman theorem (p. 618)
Cartesian multipole moments (p. 624)
dipole, quadrupole, octupole moments (p. 624)
multipole polarizability (p. 628)
multipole hyperpolarizability (p. 628)
induced dipole moment (p. 628)
ZDO (p. 635)
sum over states method (p. 635)
finite field method (p. 639)
Sadlej relation (p. 640)
second/third harmonic generation (p. 646)
nuclear magnetic dipole (p. 648)
spin magnetic moment (p. 648)
gyromagnetic factor (p. 648)
Bohr magneton (p. 648)
nuclear magneton (p. 648)
Maxwell equations (p. 962)
NMR (p. 658)
NMR Hamiltonian (p. 658)
shielding constants (p. 659)
spin–spin intermediate coupling (p. 659)
local field (p. 659)
chemical shift (p. 664)
Ramsey theory (p. 666)
diamagnetic effect (p. 668)
paramagnetic effect (p. 668)
coupling constant (p. 668)
direct spin–spin interaction (p. 669)
diamagnetic spin–orbit contribution (p. 669)
paramagnetic spin–orbit (p. 670)
spin–dipole contribution (p. 670)
Fermi contact contribution (p. 670)
coupling mechanism (p. 672)
London orbitals (p. 673)
GIAO (p. 673)

From the research front

The electric dipole (hyper)polarizabilities are not easy to calculate, because:

- the sum over states method (SOS) converges slowly, i.e. a huge number of states have to be taken into account, including those belonging to a continuum;
- the finite field method requires a large quantity of atomic orbitals with small exponents (they describe the lion's share of the electron cloud deformation), although, being diffuse, they do not contribute much to the minimized energy (and lowering the energy is the only indicator that tells us whether a particular function is important or not).

More and more often in their experiments chemists investigate *large* molecules. Such large objects cannot be described by "global" polarizabilities and hyperpolarizabilities (except perhaps optical properties, where the wave length is often much larger than size of molecule). How such large molecules function (interacting with other molecules) depends first of all on their local properties. We have to replace such characteristics by new ones offering atomic resolution, similar to those proposed in the techniques of Stone or Sokalski (p. 1018), where individual atoms are characterized by their multipole moments, polarizabilities, etc.

Even a few years ago, the shielding and especially spin–spin coupling constants were very hard to calculate with reasonable accuracy. Nowadays these quantities are computed routinely using commercial software with atomic London orbitals (or other than GIAO basis sets).

The current possibilities of the theory in predicting the nuclear shielding constants and the nuclear spin–spin coupling constants are shown in Tables 12.1 and 12.2. Note that the accuracy of the theoretical results for shielding constants is nearly the same as that of experiment. As to the spin–spin coupling constants, the theoretical results are only slightly off experimental values.

Table 12.1. Comparison of theoretical and experimental shielding constants. The shielding constant σ_A (unitless quantity) is (as usual) expressed in ppm, i.e. the number given has to be multiplied by 10^{-6} to obtain σ_A of eq. (12.83). The Hartree–Fock, MP2, MP4 results are calculated in J. Gauss, *Chem. Phys. Letters* 229 (1994) 198; the CCSD(T) in J. Gauss, J.F. Stanton, *J. Chem. Phys.* 104 (1996) 2574, and the CASSCF in K. Ruud, T. Helgaker, R. Kobayashi, P. Jørgensen, K.L. Bak, H.J. Jensen, *J. Chem. Phys.* 100 (1994) 8178. For the Hartree–Fock method see Chapter 8, for the other methods mentioned here, see Chapter 10. The references to the corresponding experimental papers are given in T. Helgaker, M. Jaszuński, K. Ruud, *Chem. Rev*. 99 (1999) 293. The experimental error is estimated for σ_H in ammonia as ±1.0, for σ_O as ±17.2, for σ_H in water as ±0.015, for σ_F as ±6, for σ_H in hydrogen fluoride as ±0.2

	CH_4		NH_3		H_2O		HF	
Method	σ_C	σ_H	σ_N	σ_H	σ_O	σ_H	σ_F	σ_H
Hartree–Fock	194.8	31.7	262.3	31.7	328.1	30.7	413.6	28.4
MP2	201.0	31.4	276.5	31.4	346.1	30.7	424.2	28.9
MP4	198.6	31.5	269.9	31.6	337.5	30.9	418.7	29.1
CCSD(T)	198.9	31.6	270.7	31.6	337.9	30.9	418.6	29.2
CASSCF	200.4	31.19	269.6	31.02	335.3	30.21	419.6	28.49
experiment	**198.7**	**30.61**	**264.54**	**31.2**	**344.0**	**30.052**	**410**	**28.5**

Table 12.2. Comparison of theoretical and experimental spin–spin coupling constants ${}^nJ_{AB}$ for ethylene (n denotes the number of separating bonds), in Hz. For the methods used see Chapter 10. All references to experimental and theoretical results are in T. Helgaker, M. Jaszuński, K. Ruud, *Chem. Rev*. 99 (1999) 293

	Spin–spin coupling constants J_{AB} for ethylene, in Hz					
Method	${}^1J_{CC}$	${}^1J_{CH}$	${}^2J_{CH}$	${}^2J_{HH}$	${}^3J_{HH\text{-cis}}$	${}^3J_{HH\text{-trans}}$
MC SCF	71.9	146.6	−3.0	−2.7	10.9	18.1
EOM-CCSD	70.1	153.23	−2.95	0.44	11.57	17.80
experiment	**67.457**	**156.302**	**−2.403**	**2.394**	**11.657**	**19.015**

Ad futurum...

It seems that the SOS method will be gradually sent out of business. The finite field method (in the electric field responses) will become more and more important, due to its simplicity. It remains however to solve the problem, how to process the information we get from such computations and translate it into the above mentioned local characteristics of the molecule.

Contemporary numerical methods allow routine calculation of polarizability. It is difficult with the hyperpolarizabilities that are much more sensitive to the quality of the atomic basis set used. The hyperpolarizabilities relate to non-linear properties, which are in high demand in new materials for technological applications.

Such problems as the dependence of the molecular spectra and of the molecular conformations and structure on the external electric field (created by our equipment or by a neighbouring molecule) will become more and more important.

The theory of the molecular response to an electric field and the theory of the molecular response to a magnetic field, despite some similarities, look as if they were "from another story". One of the reasons is that the electric field response can be described by solving the Schrödinger equation, while that corresponding to the magnetic field is based inherently on relativistic effects. The latter ones are much less investigated except some quite simple examples. Another reason may be the scale difference: the electric effects are much larger than the magnetic ones.

However, the theory for the interaction of matter with the electromagnetic field has to be coherent. The finite field method, so gloriously successful in electric field effects, is in the "stone age" stage for magnetic field effects. The propagator methods[84] look the most promising, these allow for easier calculation of NMR parameters than the sum-over-states methods.

Additional literature

A.D. Buckingham, *Advan. Chem. Phys*. 12 (1967) 107.

A classical paper on molecules in a static or periodic electric field.

H.F. Hameka, "Advanced Quantum Chemistry. Theory of Interactions between Molecules and Electromagnetic Fields", Addison-Wesley Publishing Co., Reading, Massachusetts, USA (1965).

This is a first class book, although it presents the state of the art before the *ab initio* methods for calculating the magnetic properties of molecules.

T. Helgaker, M. Jaszuński, K. Ruud, *Chem. Rev*. 99 (1999) 293.

A competent review article on the magnetic properties of molecules (NMR) with presentation of suitable contemporary theoretical methods.

Questions

1. The Hellmann–Feynman theorem says that ($\hat{H}$ means the Hamiltonian depending on the parameter P):
   a) $\frac{\partial E}{\partial P} = \langle\psi|\frac{\partial \hat{H}}{\partial P}|\psi\rangle$, if ψ is the variational trial function;
   b) $\frac{\partial E}{\partial P} = \langle\psi|\frac{\partial \hat{H}}{\partial P}|\psi\rangle$, if ψ is the Hartree–Fock wave function;
   c) $\frac{\partial E}{\partial P} = \langle\psi|\frac{\partial \hat{H}}{\partial P}|\psi\rangle$, for any normalized ψ;
   d) $\frac{\partial E}{\partial P} = \langle\psi|\frac{\partial \hat{H}}{\partial P}|\psi\rangle$, if ψ is an eigenfunction of $\hat{H}$.
2. The proportionality constant at the third-power term (the powers of electric field intensity) in the expression for the energy of a molecule in a homogeneous electric field is:
   a) a component of the quadrupole moment;
   b) a component of the dipole moment;
   c) $-\frac{1}{3!}\beta$, where β denotes a hyperpolarizability component;
   d) a component of the octupole moment.
3. A non-polar molecule (with a non-zero quadrupolar moment) in an electric field with a non-zero gradient:
   a) does not interact with the field;
   b) will rotate to align its dipole moment along the field;
   c) will orient to align its longer quadrupole axis along the field;
   d) will orient to align its longer quadrupole axis along the gradient of the field.

---

[84] J. Linderberg, Y. Öhrn, "*Propagators in Quantum Chemistry*", 2nd edition, John Wiley & Sons, Ltd, 2004.

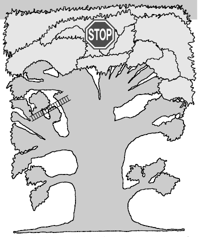
4. Second harmonic generation requires that the molecule has:
   a) large dipole hyperpolarizabilities;
   b) large quadrupole and octupole polarizabilities;
   c) large octupole and higher order hyperpolarizabilities;
   d) a large quadrupole hyperpolarizability.
5. In variational calculations for the benzene molecule (zero electric field) the GTO exponents and positions have been optimized. A small shift of the GTOs when using a finite field method:
   a) will always lower the energy;
   b) always increases the energy;
   c) will increase the energy if the GTOs move in the direction of the field and decrease if they move in the opposite direction;
   d) will always give a polarizability greater than zero.
6. The magnetic moment $\boldsymbol{M}$ of a particle:
   a) always has the direction of the particle's spin angular momentum;
   b) its length is always an integer or half-integer;
   c) interacts with a homogeneous magnetic field $\boldsymbol{H}$, and the interaction energy is equal to $\frac{1}{2}MH^2$;
   d) interacts with a homogeneous magnetic field $\boldsymbol{H}$, and the interaction energy is equal to $-\boldsymbol{H}\cdot\boldsymbol{M}$.
7. If we choose the vector potential $\boldsymbol{A}(\boldsymbol{r}) = \frac{1}{2}[\boldsymbol{H}\times\boldsymbol{r}]$, where $\boldsymbol{H}$ is the magnetic field intensity, then:
   a) we have $\nabla\boldsymbol{A} = 0$ and $\boldsymbol{A}$ agrees with the Maxwell equations;
   b) $\boldsymbol{A}(\boldsymbol{r})$ is a homogeneous field;
   c) $\boldsymbol{A}(\boldsymbol{r})$ is directed towards the origin;
   d) $\boldsymbol{A}(\boldsymbol{r})$ is parallel to $\boldsymbol{H}$.
8. The vector potential $\boldsymbol{A}(\boldsymbol{r})$ of electromagnetic field corresponds to homogeneous magnetic field $\boldsymbol{H}$. Then $\boldsymbol{A}$:
   a) is uniquely determined from the Maxwell equation;
   b) is uniquely determined from the Maxwell equation $\boldsymbol{A} = \operatorname{curl}\boldsymbol{H}$;
   c) is also a homogeneous field;
   d) $\operatorname{curl}[\boldsymbol{A} - \nabla(x^2 + y^3 + z^4)] = \boldsymbol{H}$.
9. A nuclear shielding constant consists of diamagnetic and paramagnetic parts. Each of these parts:
   a) represents a second-order effect in perturbation theory;
   b) represents a first-order effect in perturbation theory;
   c) changes when the origin of the vector potential $\boldsymbol{A}$ changes;
   d) represents the Fermi contact term.
10. The London or Gauge-Invariant Atomic Orbital $\chi_L(\boldsymbol{r}-\boldsymbol{R})$:
    a) depends on the vector potential calculated at position $\boldsymbol{R}$;
    b) depends on the vector potential calculated at position $\boldsymbol{r}$;
    c) does not depend on the vector potential;
    d) depends on the vector potential at the point shown by $\boldsymbol{r}-\boldsymbol{R}$.

Answers

1d, 2c, 3d, 4a, 5c, 6d, 7a, 8d, 9c, 10a

Chapter 13

INTERMOLECULAR INTERACTIONS



WHERE ARE WE?

Where are we?

We are already in the crown of the TREE.

An example

Why does liquid water exist? Why do molecules stick together at low temperatures? Visibly they attract each other *for some reason*. The interaction is not however very strong since water evaporates when heated (without destroying the water molecules).

What is it all about

THEORY OF INTERMOLECULAR INTERACTIONS

ENGINEERING OF INTERMOLECULAR INTERACTIONS

Chapter 8 dealt with the question of why atoms form molecules. Electrons and nuclei attract each other, and this results in almost exact neutralization of matter. Despite this, atoms and molecules interact, because

- two atoms or molecules cannot occupy the same space,
- electrons and nuclei in an atom or molecule may still interact with those in other atoms or molecules.

This chapter will tell us about the very reason for this and will give details of the interaction.

Why is this important?

What is the most important message humanity ever learned about matter? According to Richard Feynman the message would be: "*The world is built of atoms, which repel each other at short distances and attract at longer ones*". If the intermolecular interactions were suddenly switched off, the world would disintegrate in about a femtosecond, that is in a single period of atomic vibration (the atoms simply would not come back when shifted from their equilibrium positions). Soon after, everything would evaporate and a sphere of gas, the remainder of the Earth, would be held by gravitational forces. Isn't it enough?

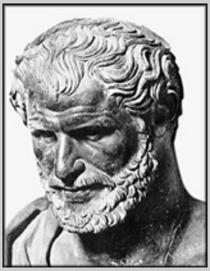
What is needed?

- Perturbation theory (Chapter 5, absolutely).

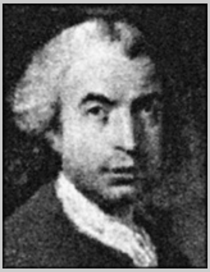
- Variational method (Chapter 5, recommended).
- Appendix X, p. 1038 (absolutely).
- Many-Body Perturbation Theory (MBPT) (Chapter 10, p. 554, necessary).
- Reduced resolvent (Chapter 10, p. 554, necessary).
- Appendix Y, p. 1050 (recommended).
- Appendix T (mentioned).

Classical works

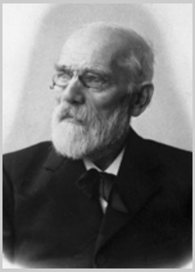
Such an important subject was recognized very early. The idea that the cohesion of matter stems from the interaction of small indivisible particles ("atoms") comes from Democritus. ★ An idea similar to that cited by Feynman was first stated clearly by the Croat scientist Rudjer Bosković in "*Theoria Philosophiae naturalis*", Venice, 1763. ★ Padé approximants were first proposed in the PhD thesis of Henri Padé entitled "*Sur la représentation approchée d'une fonction pour des fractions rationnelles*", which was published in *Annales des Sciences d'Ecole Normale Superieure, Suppl. [3]*, 9 (1892) 1. ★ The role of intermolecular interactions was highlighted in the work of Johannes Diderik van der Waals, especially in "*Die Kontinuität des gasformigen und flüssigen Zustandes*", Barth, Leipzig (1899, 1900). From that time on, intermolecular interactions are often called van der Waals interactions. ★ The concept of ionic radii was first proposed by Linus Pauling in "*The Sizes of Ions and the Structure of Ionic Crystals*", *Journal of the American Chemical Society*, 49 (1927) 765. ★ The quantum mechanical explanation of intermolecular forces, including the ubiquitous dispersion interactions, was given by Fritz London in "*Zur Theorie und Systematik der Molekularkräfte*", *Zeitschrift für Physik*, 63 (1930) 245 and in "*Über einige Eigenschaften und Anwendungen der Molekularkräfte*" from *Zeitschrift für Physikalische Chemie (B)*, 11 (1930) 222. ★ Linus Pauling, invited Baker Lecturer to Cornell University, wrote one of the most seminal books in chemistry "*The Nature of the Chemical Bond*", Cornell Univ. Press, Ithaca, 1948, where *inter alia* he elucidated the role of hydrogen bonds in forming structures. ★ The hydrophobic effect was first highlighted by Walter Kauzmann in a paper "*Some Factors in the Interpretation of Protein*



Democritus of Abdera (ca. 460 B.C. – ca. 370 B.C.), Greek philosopher, founder of the first atomic theory. According to him, nature represents a constant motion of indivisible and permanent particles (atoms), whose interactions result in various materials. *It turned out after almost 25 centuries that this hypothesis was basically correct!* All the written works of Democritus have been lost, but his ideas continued to have an important impact on science for centuries.



Rudjer Josip Bosković (1711–1787), a Croat physicist, mathematician, astronomer and philosopher from beautiful Dubrovnik.



Johannes Diderik van der Waals (1837–1923), Dutch physicist, professor at the University of Amsterdam. His research topic was the influence of intermolecular forces on the properties of gases (equation of state of the real gas, 1873) and liquids. In 1910 van der Waals received a Nobel Prize "*for his work on the equation of state for gases and liquids*".

Denaturation”, in *Advances in Protein Chemistry*, 14 (1959) 1, the effect was further elaborated by George Nemethy, Harold Scheraga, Frank Stillinger and David Chandler among others. ★ Resonance interactions were first described by Robert S. Mulliken in an article “*The Interaction of Differently Excited Like Atoms at Large Distances*”, in *Physical Reviews*, 120 (1960) 1674. ★ Bogumił Jeziorski and Włodzimierz Kołos extended the existing theory of intermolecular forces to intermediate distances (“*On the Symmetry Forcing in the Perturbation Theory of Weak Intermolecular Interactions*”, *International Journal of Quantum Chemistry*, 12 Suppl. 1 (1977) 91).

THEORY OF INTERMOLECULAR INTERACTIONS

There are two principal methods of calculating the intermolecular interactions: the supermolecular method and the perturbational method. Both assume the Born–Oppenheimer approximation.

13.1 INTERACTION ENERGY CONCEPT

The idea of interaction energy is based on the Born–Oppenheimer (clamped nuclei, see eq. (6.4)) approximation. Let us define interaction energy at the configuration $\boldsymbol{R}$ of the nuclei as

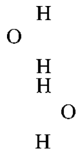
$$E_{\text{int}}(\boldsymbol{R}) = E_{ABC\ldots}(\boldsymbol{R}) - \left[E_A(\boldsymbol{R}) + E_B(\boldsymbol{R}) + E_C(\boldsymbol{R}) + \cdots\right], \tag{13.1}$$

where $E_{ABC\ldots}(\boldsymbol{R})$ is the electronic energy (corresponding to $E_0^{(0)}$ from eq. (6.21)) of the total system, and $E_A(\boldsymbol{R})$, $E_B(\boldsymbol{R})$, $E_C(\boldsymbol{R})$, … are the electronic energies of the interacting subsystems, calculated *at the same positions of the nuclei* as those in the total system.

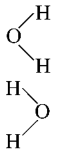
13.1.1 NATURAL DIVISION AND ITS GRADATION

Although the notion of interaction energy is of great practical value, its theoretical meaning is a little bit fuzzy. Right at the beginning we have a question: interaction *of what*? We view the system *as composed of particular subsystems*, that once isolated, then have to be put together.

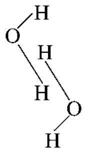
For instance, the supersystem



may be considered as two interacting water molecules, but even then we still have an uncertainty, whether the two molecules correspond to (I) or to (II):

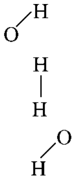


I



II

In addition the system might be considered as composed of a hydrogen molecule interacting with two OH radicals:



etc.

Choice of subsystems is of no importance from the point of view of mathematics, but is of crucial importance from the point of view of calculations in theoretical chemistry.

The particular choice of subsystem should depend on the kind of experiment with which we wish to compare our calculations:

- we are interested in the interaction of water molecules when studying water evaporation or freezing;
- we are interested in the interaction of atoms and ions that exist in the system when heating water to 1000 °C.

Let us stress, that in any case when choosing subsystems we are forced to single out *particular atoms* belonging to subsystem[1] A and B. It is not sufficient to define the kind of molecules participating in the interaction, see our examples I and II.

If when dividing a system into n subsystems in two ways (I and II), we obtain $|E_{\text{int}}|_{\text{I}} < |E_{\text{int}}|_{\text{II}}$, division I will be called more *natural* than division II.

natural subsystems

13.1.2 WHAT IS MOST NATURAL?

Which division is *most* natural? We do not have any experience in answering such questions. What? Why should we have any difficulties? It is sufficient to consider all possible divisions and to choose the one which requires lowest energy. Unfortunately, this is not so obvious. Let us consider two widely separated water molecules (Fig. 13.1.a).

---

[1] This means that the interaction energy idea belongs to classical concepts. In a quantum system, particles of the same kind are indistinguishable. A quantum system does not allow us to separate a part from the system. Despite this, the interaction energy idea is important and useful.

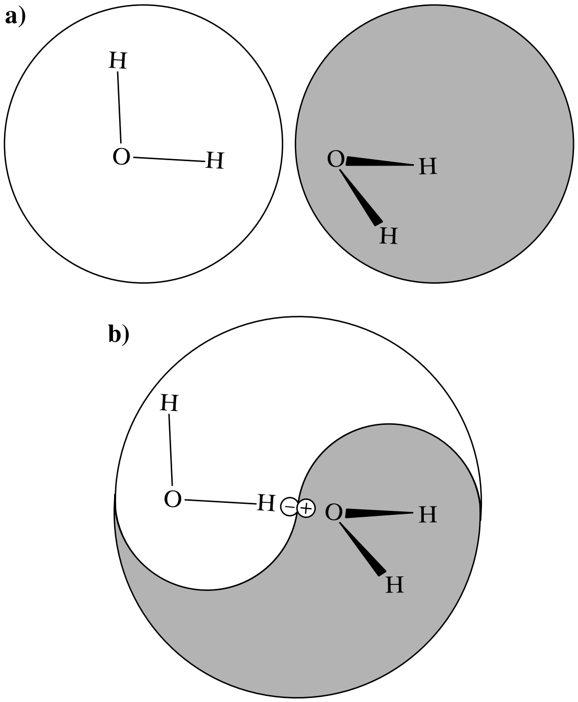


Fig. 13.1. Part-entity relationship. Two distinct ways of dividing the $(H_2O)_2$ system into subsystems. Division (a) is traditional. The interacting objects are two isolated water molecules and the interaction energy is equal to about -5 kcal/mol (attraction). Division (b) is more subtle – a certain point in space is treated as being composed of two fictitious charges $q > 0$ and $-q$, and one of the charges is ascribed to one, and the other to the second molecule. In this way two *new subsystems* are defined, each of them composed of a water molecule *and* the corresponding point charge. The value of q may be chosen in such a way as to produce the interaction energy of the new subsystems close to 0. Therefore this is a more natural choice of subsystems than the traditional one. The total interaction energy of the two water molecules is now absorbed within the interactions of the fictitious point charges with "their" water molecules. Each of the point charges takes over the interaction of "its" water molecule with the rest of the Universe. Hence, I have permitted myself (with the necessary *licentia poetica*) to use the *yin* and *yang* symbols – the two basic elements of ancient Chinese philosophy.

Right between the molecules, e.g., in the middle of the OO separation, we place two point charges $q > 0$ and $-q$, i.e. we place nothing, since the charges cancel each other (please compare a similar trick on p. 492). We therefore have just two water molecules. Now we start our game. We say that the charges are *real*: one belongs to one of the molecules and the other to the second (Fig. 13.1.b). The charge q could be anything, but we want to use it for a very special goal: to construct the two subsystems in a *more natural* way than just two water molecules. It is interesting that after the choice is made, any of the subsystems has *lower* energy than that of isolated water, since the molecules are oriented in such a way as to attract each other. This means that the value of q can be chosen from an interval making the choice of subsystems more natural. For a certain $q = q_{\text{opt}}$ we would obtain as the interaction energy of the new subsystems: $E'_{\text{int}} = 0$. This certainly would be the most natural choice,[2] with the "dressed" water molecules, not seeing each other.[3]

Below in this chapter we will not use any fictitious charges.

13.2 BINDING ENERGY

Interaction energy can be calculated at any configuration $\boldsymbol{R}$ of the nuclei. We may ask whether any "privileged" configuration exists with respect to the interaction energy. This was our subject in Chapter 7, and it turned out that the electronic energy may have many minima (equilibria) as a function of $\boldsymbol{R}$. For each of such configurations we may define the *binding energy with respect to a particular dissociation channel* as the difference of the corresponding interaction energies (all subsystems at the optimal positions $\boldsymbol{R}_{\text{opt}(j)}$ of the nuclei with respect to the electronic energy E_j):

equilibrium geometry

$$E_{\text{bind}} = E_{ABC\ldots}(\boldsymbol{R}_{\text{opt(tot)}}) - \sum_{j=A,B,C,\ldots} E_j(\boldsymbol{R}_{\text{opt}(j)}). \tag{13.2}$$

At a given configuration $\boldsymbol{R}_{\text{opt(tot)}}$ we usually have many dissociation channels.

13.3 DISSOCIATION ENERGY

The calculated interaction energy of eq. (13.1), as well as the binding energies are only theoretical quantities and cannot be measured. What is measurable is the closely connected *dissociation energy*

$$E_{\text{diss}} = E_{\text{bind}} - \left[\Delta E_{0,\text{tot}} - \sum_{j=A,B,C,\ldots} \Delta E_{0,j}\right], \tag{13.3}$$

where $\Delta E_{0,\text{tot}}$ stands for what is known as the *zero vibration energy* of the total system (cf. p. 304) at the equilibrium geometry $\boldsymbol{R}_{\text{opt(tot)}}$, and $\Delta E_{0,j}$ for $j = A, B, C, \ldots$ representing the zero vibration energies for the subsystems. In the harmonic approximation $\Delta E_{0,\text{tot}} = \frac{1}{2}\sum_i h\nu_{1,\text{tot}}$, $\Delta E_{0,A} = \frac{1}{2}\sum_i h\nu_{i,A}$, $\Delta E_{0,B} = \frac{1}{2}\sum_i h\nu_{i,B}$, $\Delta E_{0,C} = \frac{1}{2}\sum_i h\nu_{i,C}, \ldots$ at their equilibrium geometries $\boldsymbol{R}_{\text{opt}(A)}$, $\boldsymbol{R}_{\text{opt}(B)}$, $\boldsymbol{R}_{\text{opt}(C)}$, $\ldots$, respectively.

13.4 DISSOCIATION BARRIER

If a molecule receives dissociation energy it is most often a sufficient condition for its dissociation. Sometimes however the energy is too low, and the reason is that there is an energy barrier to be overcome. Sometimes the barrier is very high and

---

[2] Although not unique, since the charges could be chosen at different points in space, and we could also use point multipoles, etc.

[3] Allusion to the elementary particles "dressed" by interactions, see section "*What is it all about?*", Chapter 8. It is worth noting that we have to superpose the subsystems *first* (then the fictitious charges disappear), and *then* calculate the interaction energy of the water molecules deformed by the charges.

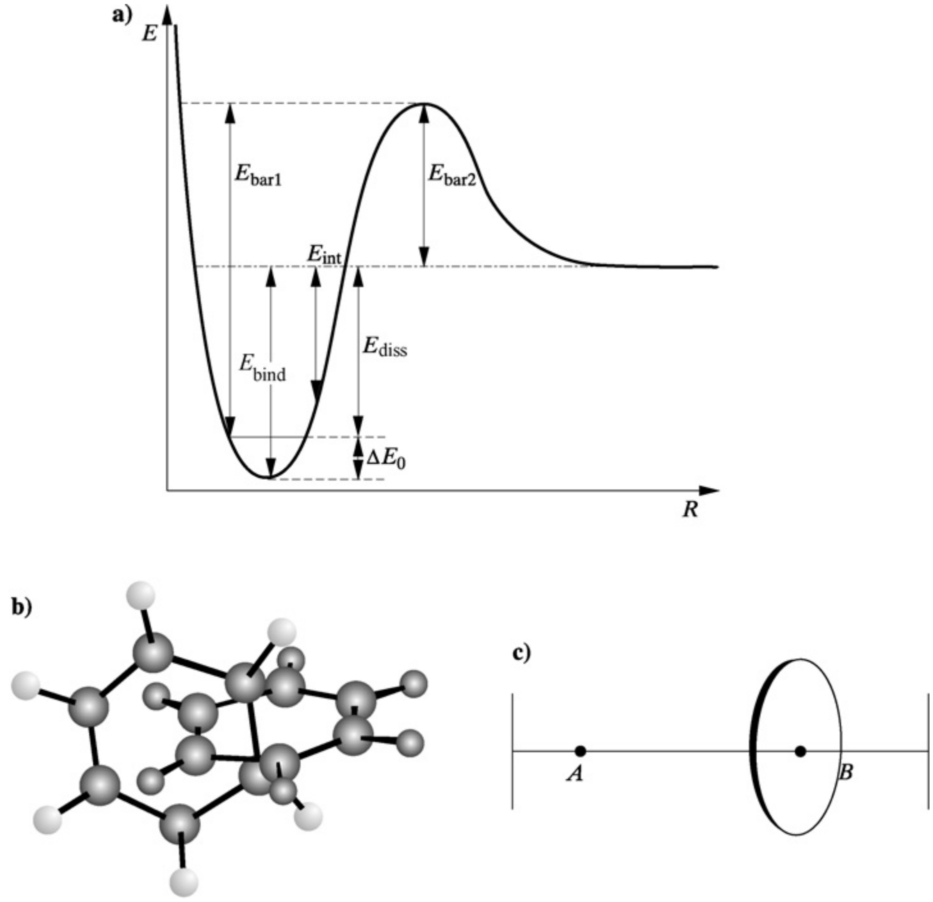


Fig. 13.2. (a) Interaction energy E_{int}, binding energy E_{bind}, dissociation energy E_{diss} and barrier energy E_{bar}. ΔE_0 is the zero vibration energy. Please note that the height of the barrier is different (E_{bar1} or E_{bar2}) depending on the starting point considered. The first is connected with the energy cost required to go from jail to freedom (large), while the second is the energetic price for going to jail (much easier). Figs. (b), (c), (d) represent rather exceptional cases of intermolecular interactions, when a part of the total system is somehow confined (Fig. (a)) by the rest of the system. The catenan shown in Fig. (b) consists of two intertwined rings, the rotaxan (scheme) is composed of a "molecular stick" with a molecular ring on it, Fig. (c), the latter having two stable positions ("stations A and B" to be used in future molecular computers). Fig. (d) shows what is called an endohedral complex. In this particular case a water molecule is confined in the fullerene cage. The systems have been trapped in metastable states, as they were formed. To free the subsystems, high energy has to be used to break some chemical bonds. One of the advantages of theory is that we can consider compounds which sometimes would be difficult to obtain in experiment.

the system is stable even if the dissociation products have (much) lower energy. Catenans, rotaxans and endohedral complexes shown in Fig. 13.2 may serve as examples.

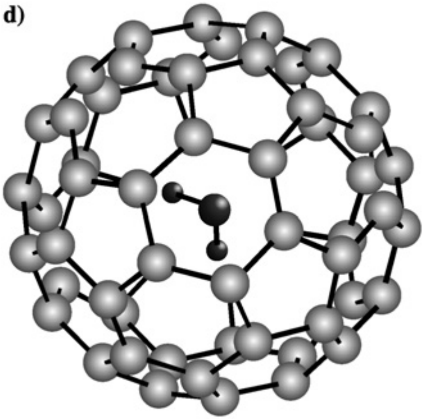


Fig. 13.2. *Continued.*

The energy necessary to overcome the barrier from the trap side is equal to

$$E_{\text{bar}} = E_{\#} - \left(E_{\min} + \frac{1}{2}\sum_j h\nu_j\right),$$

where $E_{\min}$ is the energy of the bottom of the well, $E_{\#}$ represents the barrier top energy, and $\frac{1}{2}\sum_j h\nu_j$ is the zero vibration energy of the well.

13.5 SUPERMOLECULAR APPROACH

In the supermolecular method the interaction energy is calculated from its definition (13.1) using any reliable method of electronic energy calculations. For the sake of brevity we will consider the interaction of two subsystems: A and B.

13.5.1 ACCURACY SHOULD BE THE SAME

There is a problem though. The trouble is that we are unable to solve the Schrödinger equation exactly either for AB or for A or for B. We have to use approximations. If so, we have to worry about the same accuracy of calculation for AB as for A and B. From this we may expect that

> in determining E_{AB} as well as E_A and E_B the same theoretical method should be preferred, because any method introduces its own systematic error and we may hope that these errors will cancel at least partially in the above formula.

This problem is already encountered at the stage of basis set choice. For example, suppose we have decided to carry out the calculations within the Hartree–Fock method in the LCAO MO approximation, p. 360. The same method has to be used for AB, A and B. However what does this really mean? Should we use the following protocol:

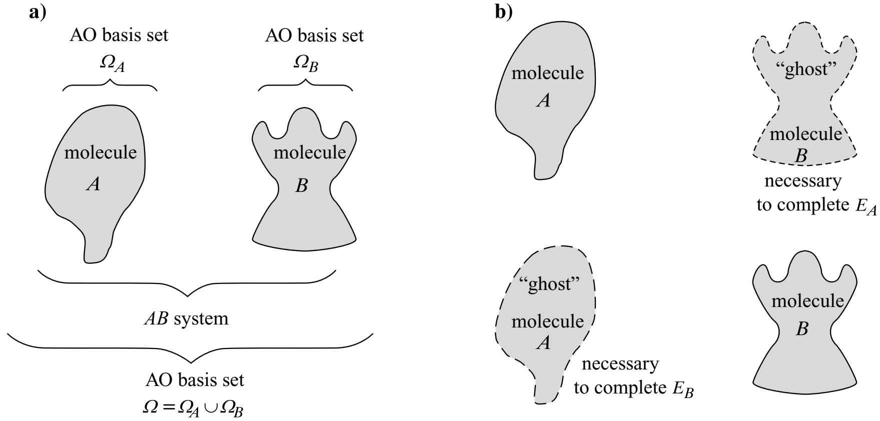


Fig. 13.3. (a) Basis set superposition problem (BSSE). Each of the molecules offers its own atomic orbitals to the total basis set $\Omega = \Omega_A \cup \Omega_B$. Fig. (b) illustrates the counter-poise method, in which the calculations for a single subsystem are performed within the full atomic basis set Ω: the atomic orbitals centred on it and what are called ghost orbitals centred on the partner.

1. Consider the atomic basis set Ω that consists of the atomic orbitals centred on the nuclei of A (set Ω_A) and on the nuclei of B (set Ω_B), i.e. $\Omega = \Omega_A \cup \Omega_B$.
2. Calculate E_{AB} using Ω, E_A using Ω_A and E_B using Ω_B (Fig. 13.3a). Apparently everything looks logical, but *we did not use the same method* when calculating the energies of AB, A and B. The basis set used has been different depending on what we wanted to calculate.

Thus, it seems more appropriate to calculate all three quantities using the same basis set Ω.

13.5.2 BASIS SET SUPERPOSITION ERROR (BSSE)

Such an approach is supported by the following reasoning. When the calculations are performed for E_{AB} within the basis set Ω we calculate implicitly not only the interaction energy, but also we allow the *individual subsystems* to lower their energy. Conclusion: by subtracting from E_{AB} the energies: E_A calculated with Ω_A and E_B with Ω_B, we are left not only with the interaction energy (as should be), but also with an unwanted and non-physical extra term (an error) connected with the artificial lowering of the subsystems' energies, when calculating E_{AB}. This error is called the BSSE (*Basis Set Superposition Error*).

...and the remedy

counter-poise method

"ghost orbitals"

To remove the BSSE we may consider the use of the basis set Ω not only for E_{AB} but also for E_A and E_B. This procedure called the *counter-poise method*, was first introduced by Boys and Bernardi.[4] Application of the full basis set Ω when calculating E_A results in the wave function of A containing not only its own atomic orbitals, but also the atomic orbitals of the ("absent") partner B, the "*ghost orbitals*" (Fig. 13.3b). As a by-product, the charge density of A exhibits broken symmetry

---

[4] S.F. Boys, F. Bernardi, *Mol. Phys*. 19 (1970) 553.

with respect to the symmetry of A itself (if any), e.g., the helium atom would have a small dipole moment, etc. Unfortunately, the counter-poise method depends on which channel of dissociation is considered. If several channels are considered at one time, not only are we confronted with an ambiguity, but also this inevitably leads to discontinuities in the calculated energies. This problem is not yet solved in the literature.

13.5.3 GOOD AND BAD NEWS ABOUT THE SUPERMOLECULAR METHOD

Two deficiencies

When performing the subtraction in formula (13.1), we obtain a number representing the interaction energy at a certain distance and orientation of the two subsystems.

> The resulting E_{int} has two disadvantages: it is less precise than E_{AB}, E_A and E_B, and it does not tell us anything about why the particular value is obtained.

The first disadvantage could be compared (following Coulson[5]) to weighing the captain's hat by first weighing the ship with the captain wearing his hat and the ship with the captain without his hat (Fig. 13.4). Formally everything is perfect, but there is a cancellation of significant digits in E_{AB} and $(E_A + E_B)$, that may lead to a very poor interaction energy.

The second deficiency deals with the fact that the interaction energy obtained is just a number and we will have no idea why the number is of such magnitude.[6] Both deficiencies will be removed in the perturbational approach to intermolecular

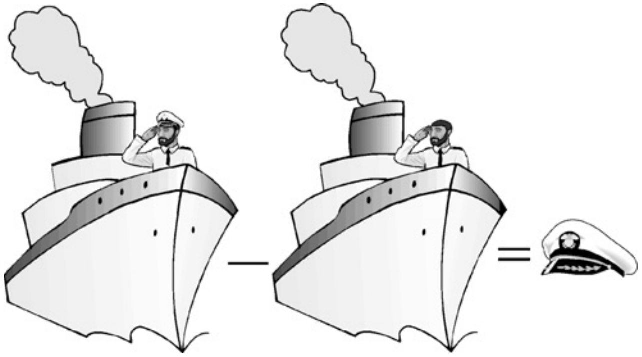


Fig. 13.4. In the supermolecular method we subtract two large numbers that differ only slightly and lose accuracy in this way. It resembles determining the weight of the captain's hat by weighing first the ship with the captain wearing his hat, then repeating the same with the captain without his hat and subtracting the two results. In order not to obtain a result like 240 kg or so, we have to have at our disposal a very accurate method of weighing things.

---

[5] C.A. Coulson, "*Valence*", Oxford University Press (1952).

[6] The severity of this can be diminished by analyzing the supermolecular interaction energy expression (using molecular orbitals of A and B) and identifying the physically distinguishable terms by the kind of

interaction. Then, the interaction energy will be calculated directly and we will be able to tell which physical contributions it consists of.

Important advantage

> A big advantage of the supermolecular method is its applicability at any intermolecular distance, i.e. independently of how strong the interaction is.

13.6 PERTURBATIONAL APPROACH

13.6.1 INTERMOLECULAR DISTANCE – WHAT DOES IT MEAN?

What is the distance (in kilometers) between the Polish and German populations, or what does the distance between two buses mean? Because of the non-zero dimensions of both objects, it is difficult to tell what the distance could be and any measure of it will be arbitrary. It is the same story with molecules. Up till now we did not need a notion for the intermolecular distance, the positions of the nuclei were sufficient. At the beginning we need only an infinite distance and therefore any definition will be acceptable. Later, however, we will be forced to specify the intermolecular distance (cf. p. 698 and Appendix X on p. 1038). The final numerical values should not depend on this choice, but intermediate results could depend on it. It will turn out that despite the existing arbitrariness, we will prefer those definitions which are based upon the charge barycentre distance or similar.

13.6.2 POLARIZATION APPROXIMATION (TWO MOLECULES)

According to the Rayleigh–Schrödinger perturbation theory (Chapter 5) the unperturbed Hamiltonian $\hat{H}^{(0)}$ is a sum of the isolated molecules' Hamiltonians: $\hat{H}^{(0)} = \hat{H}_A + \hat{H}_B$. Following quantum theory tradition in the present chapter the symbol for the perturbation operator will be changed (when compared to Chapter 5): $\hat{H}^{(1)} \equiv V$.

> Despite the fact that we may also formulate the perturbation theory for excited states, we will assume that we are dealing with the ground state (and denote it by subscript "0"). In what is called the *polarization approximation*, the zeroth-order wave function will be taken as a product
>
> $$\psi_0^{(0)} = \psi_{A,0}\psi_{B,0}, \tag{13.4}$$
>
> where $\psi_{A,0}$ and $\psi_{B,0}$ are the exact ground state wave functions for the isolated molecules A and B, respectively, i.e.
>
> $$\hat{H}_A\psi_{A,0} = E_{A,0}\psi_{A,0},$$
> $$\hat{H}_B\psi_{B,0} = E_{B,0}\psi_{B,0}.$$

---

molecular integrals of which they are composed (K. Kitaura, K. Morokuma, *Intern. J. Quantum Chem.* 10 (1976) 325).

We will assume that, because of the large separation of the two molecules, the electrons of molecule A are distinguishable from the electrons of molecule B. We have to stress the classical flavour of this approximation. Secondly, we assume that the exact wave functions of both isolated molecules:[7] $\psi_{A,0}$ and $\psi_{B,0}$ are at our disposal.

Of course, function $\psi_0^{(0)}$ is only an approximation to the exact wave function of the total system. Intuition tells us that this approximation is probably very good, because we assume the perturbation is small and the product function $\psi_0^{(0)} = \psi_{A,0}\,\psi_{B,0}$ is an exact wave function for the *non-interacting* system.

> The chosen $\psi_0^{(0)}$ has a wonderful feature, namely it represents an eigenfunction of the $\hat{H}^{(0)}$ operator, as is required by the Rayleigh–Schrödinger perturbation theory (Chapter 5).

The function has also an unpleasant feature: it differs from the exact wave function by symmetry. For example, it is easy to see that

> the function $\psi_0^{(0)}$ *is not* antisymmetric with respect to the electron exchanges between molecules, while the exact function has to be antisymmetric with respect to any exchange of electron labels.

This deficiency exists for any intermolecular distance.[8] We will soon pay a high price for this.

First order effect: electrostatic energy

The first order correction (see eq. (5.22), p. 207)

$$E_0^{(1)} \equiv E_{\text{elst}} \equiv E_{\text{pol}}^{(1)} = \langle \psi_0^{(0)} | V \psi_0^{(0)} \rangle, \tag{13.5}$$

---

[7] We will eliminate an additional complication which sometimes may occur. The n-th state of the two non-interacting molecules comes, of course, from *some* states of the isolated molecules A and B. It may happen (most often when the two molecules are identical), that two different sets of the states give the same energy $E_n^{(0)}$, typically, this may happen upon the exchange of excitations of both molecules. Then, $\psi_n^{(0)}$ has to be taken as a linear combination of these two possibilities, which leads to profound changes of the formulae with respect to the usual cases. Such an effect is called the *resonance interaction* (R.S. Mulliken, *Phys. Rev.* 120 (1960) 1674). The resulting interaction decays with the distance as R^{-3}, i.e. quite slowly, making possible an excitation energy transfer through long distances between the interacting molecules. The resonance interaction turns out to be very important (e.g., in biology). An interested reader may find more in the review article J.O. Hirschfelder, W.J. Meath, *Advan. Chem. Phys.* 12 (1967) 3.

resonance interaction

[8] We may say that the range of the Pauli principle is infinity. If somebody paints some electrons green and others red (distinguishable electrons, we do this in the perturbational method), they are in no man's land, between the classical and quantum worlds. Since the wave function $\psi_0^{(0)}$ does not have the proper symmetry, the corresponding operator $\hat{H}^{(0)} = \hat{H}_A + \hat{H}_B$ is just a mathematical object having little relation to the total system under study.

represents what is called the electrostatic interaction energy (E_{elst}). To stress that E_{elst} is the first-order correction to the energy in the polarization approximation, the quantity will alternatively be denoted by $E^{(1)}_{\text{pol}}$. The electrostatic energy represents the Coulombic interaction of two "frozen" charge distributions corresponding to the isolated molecules A and B, because it is the mean value of the Coulombic interaction energy operator V calculated with the wave function $\psi_0^{(0)}$ being the product of the wave functions of the isolated molecules[9] $\psi_0^{(0)} = \psi_{A,0}\psi_{B,0}$.

electrostatic energy

Second-order energy: induction and dispersion energies

The second-order energy (p. 208) in the polarization approximation approach can be expressed in a slightly different way. The n-th state of the total system at long intermolecular distances corresponds to some states n_A and n_B of the individual molecules, i.e.

$$\psi_n^{(0)} = \psi_{A,n_A}\psi_{B,n_B} \tag{13.6}$$

and[10]

$$E_n^{(0)} = E_{A,n_A} + E_{B,n_B}. \tag{13.7}$$

Using this assumption, the second-order correction to the ground-state energy (we assume $n=0$ and $\psi_0^{(0)} = \psi_{A,0}\psi_{B,0}$) can be expressed as (see Chapter 5, p. 208)

$$E_0^{(2)} = \sum_{n_A}\sum_{n_B}{}' \frac{|\langle\psi_{A,n_A}\psi_{B,n_B}|V\psi_{A,0}\psi_{B,0}\rangle|^2}{(E_{A,0}-E_{A,n_A})+(E_{B,0}-E_{B,n_B})}, \tag{13.8}$$

where "prime" in the summation means excluding $n=0$, or $(n_A, n_B) = (0,0)$. The quantity $E_0^{(2)}$ can be divided in the following way

$$E_0^{(2)} = \sum_{n_A}\sum_{n_B}{}' \cdots = \sum_{(n_A=0,\ n_B\neq 0)} \cdots + \sum_{(n_A\neq 0,\ n_B=0)} \cdots + \sum_{(n_A\neq 0,\ n_B\neq 0)} \cdots. \tag{13.9}$$

Let us construct a matrix $\boldsymbol{A}$ (of infinite dimension) composed of the element $A_{00} = 0$ and the other elements calculated from the formula

$$A_{n_A,n_B} = \frac{|\langle\psi_{A,n_A}\psi_{B,n_B}|V\psi_{A,0}\psi_{B,0}\rangle|^2}{(E_{A,0}-E_{A,n_A})+(E_{B,0}-E_{B,n_B})} \tag{13.10}$$

and divide it into the following parts (I, II, III on the scheme)

---

[9] We will see later that taking the zero-order wave function with the proper symmetry leads to the first order energy containing what is called the valence repulsion, besides the $E^{(1)}_{\text{pol}}$ term.

[10] Also in this case we exclude the resonance interaction.

$$\begin{array}{cc|c|cccccc}
 & & \multicolumn{7}{l}{n_A \rightarrow} \\
 & & 0 & 1 & 2 & 3 & 4 & 5 & \ldots \\ \cline{3-9}
n_B & 0 & 0 & & & \text{II} & & & \\ \cline{3-9}
\downarrow & 1 & & & & & & & \\
 & 2 & & & & & & & \\
 & 3 & \text{I} & & & \text{III} & & & \\
 & 4 & & & & & & & \\
 & 5 & & & & & & & \\
 & \vdots & & & & & & &
\end{array}$$

The quantity $E_0^{(2)}$ is a sum of all the elements of $\boldsymbol{A}$. This summation will be carried out in three steps. First, the sum of all the elements of column 0 (part I, $n_A = 0$) represents the *induction energy* associated with forcing a change in the charge distribution of the molecule B by the charge distribution of the isolated ("frozen") molecule A. Second, the sum of all the elements of row 0 (part II, $n_B = 0$) has a similar meaning, but the roles of the molecules are interchanged. Finally, the sum of all the elements of the "interior" of the matrix (part III, n_A and n_B not equal to zero) represents the *dispersion energy*. Therefore,

induction energy

dispersion energy

$$E_0^{(2)} = \underset{\text{I}}{E_{\text{ind}}(A \rightarrow B)} + \underset{\text{II}}{E_{\text{ind}}(B \rightarrow A)} + \underset{\text{III}}{E_{\text{disp}}}, \tag{13.11}$$

where

$$E_{\text{ind}}(A \rightarrow B) = \sum_{n_B}{}' \frac{|\langle \psi_{A,0}\psi_{B,n_B}|V\psi_{A,0}\psi_{B,0}\rangle|^2}{(E_{B,0} - E_{B,n_B})},$$

$$E_{\text{ind}}(B \rightarrow A) = \sum_{n_A}{}' \frac{|\langle \psi_{A,n_A}\psi_{B,0}|V\psi_{A,0}\psi_{B,0}\rangle|^2}{(E_{A,0} - E_{A,n_A})},$$

$$E_{\text{disp}} = \sum_{n_A}{}' \sum_{n_B}{}' \frac{|\langle \psi_{A,n_A}\psi_{B,n_B}|V\psi_{A,0}\psi_{B,0}\rangle|^2}{(E_{A,0} - E_{A,n_A}) + (E_{B,0} - E_{B,n_B})}. \tag{13.12}$$

What do these formulae tell us?

One thing has to be made clear. In formula (13.12) we sometimes see arguments for the interacting molecules undergoing excitations. We have to recall however that all the time we are interested in the ground state of the total system, and calculating its energy and wave function. The excited state wave functions appearing in the formulas are the consequence of the fact that the first-order correction to the wave function is expanded in a complete basis set chosen deliberately as $\{\psi_n^{(0)}\}$. If we took another basis set, e.g., the wave functions of another isoelectronic molecule, we would obtain the same numerical results (although formulae (13.12) will not hold), but the argument would be removed. From the mathematical point of view, the very essence of the perturbation theory means a small deformation of the

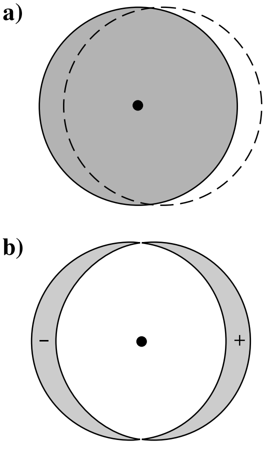


Fig. 13.5. A perturbation of the wave function is a *small* correction. Fig. (a) shows in a schematic way, how a wave function, spherically symmetric with respect to the nucleus, can be transformed into a function that is shifted off the nucleus. The function representing the correction is shown schematically in Fig. (b). Please note the function has symmetry of a p orbital.

starting $\psi_0^{(0)}$ function. This tiny deformation is the target of the expansion in the basis set $\{\psi_n^{(0)}\}$. In other words, the perturbation theory involves just a cosmetic adjustment of the $\psi_0^{(0)}$: add a small hump here (Fig. 13.5), subtract a small function there, etc. Therefore, the presence of the excited wave functions in the formulae is not an argument for observing some physical excitations. We may say that the system took what we have prepared for it, and we have prepared excited states.

This does not mean that the energy eigenvalues of the molecule have no influence on its induction or dispersion interactions with other molecules.[11] However this is a different story. It has to do with whether the small deformation we have been talking about does or does not depend on the energy eigenvalues spectrum of the individual molecules. The denominators in the expressions for the induction and dispersion energies suggest that the lower excitation energies of the molecules, the larger their deformation, induction and dispersion energies.

13.6.3 INTERMOLECULAR INTERACTIONS: PHYSICAL INTERPRETATION

Now the author would like to recommend the reader to study the multipole expansion concept (Appendix X on p. 1038, also cf. Chapter 12, p. 624).

“intermolecular distance”

The very essence of the multipole expansion is a replacement of the Coulombic interaction of two particles (one from molecule A, the other from the molecule B) by an infinite sum of interactions of what are called multipoles, where each interaction term has in the denominator an integer power of the distance (called the intermolecular distance R) between the origins of the two coordinate systems localized in the individual molecules.

---

[11] The smaller the gap between the ground and excited states of the molecule, the larger the polarizability, see Chapter 12.

In other words, multipole expansion describes the intermolecular interaction of two non-spherically symmetric, distant objects by the "interaction" of deviations (multipoles) from spherical symmetry.

To prepare ourselves for the application of the multipole expansion, let us introduce two Cartesian coordinate systems with x and y axes in one system parallel to the corresponding axes in the other system, and with the z axes collinear (see Fig. X.1 on p. 1039). One of the systems is connected to molecule A, the other one to molecule B, and the distance between the origins is R ("intermolecular distance").[12]

The operator V of the interaction energy of two molecules may be written as

$$V = -\sum_j \sum_a \frac{Z_a}{r_{aj}} - \sum_i \sum_b \frac{Z_b}{r_{bi}} + \sum_{ij} \frac{1}{r_{ij}} + \sum_a \sum_b \frac{Z_a Z_b}{R_{ab}}, \tag{13.13}$$

where we have used the convention that the summations over i and a correspond to all electrons and nuclei of molecule A, and over j and b of molecule B. Since the molecules are assumed to be distant, we have a *practical* guarantee that the interacting particles are distant too. In V many terms with inverse interparticle distance are present. For any such term we may write the corresponding multipole expansion (Appendix X, p. 1039, s is smaller of numbers k and l):

$$-\frac{Z_a}{r_{aj}} = \sum_{k=0} \sum_{l=0} \sum_{m=-s}^{m=s} A_{kl|m|} R^{-(k+l+1)} \hat{M}_A^{(k,m)}(a)^* \hat{M}_B^{(l,m)}(j),$$

$$-\frac{Z_b}{r_{bi}} = \sum_{k=0} \sum_{l=0} \sum_{m=-s}^{m=s} A_{kl|m|} R^{-(k+l+1)} \hat{M}_A^{(k,m)}(i)^* \hat{M}_B^{(l,m)}(b),$$

$$\frac{1}{r_{ij}} = \sum_{k=0} \sum_{l=0} \sum_{m=-s}^{m=s} A_{kl|m|} R^{-(k+l+1)} \hat{M}_A^{(k,m)}(i)^* \hat{M}_B^{(l,m)}(j),$$

$$\frac{Z_a Z_b}{R_{ab}} = \sum_{k=0} \sum_{l=0} \sum_{m=-s}^{m=s} A_{kl|m|} R^{-(k+l+1)} \hat{M}_A^{(k,m)}(a)^* \hat{M}_B^{(l,m)}(b),$$

where

$$A_{kl|m|} = (-1)^{l+m} \frac{(k+l)!}{(k+|m|)!(l+|m|)!}, \tag{13.14}$$

---

[12] A sufficient condition for the multipole expansion convergence is such a separation of the charge distributions of both molecules, that they could be enclosed in two non-penetrating spheres located at the origins of the two coordinate systems. This condition cannot be fulfilled with molecules, because their electronic charge density distribution extends to infinity. The consequences of this are described in Appendix X. However, the better the sphere condition is fulfilled (by a proper choice of the origins) the more effective in describing the interaction energy are the first terms of the multipole expansion.

The very fact that we use closed sets (like the spheres) in the theory, indicates that in the polarization approximation we are in no man's land between the quantum and classical worlds.

and the multipole moment $M_C^{(k,m)}(n)$ pertains to particle n and is calculated in "its" coordinate system $C = A, B$. For example,

$$\hat{M}_A^{(k,m)}(a) = Z_a R_a^k P_k^{|m|}(\cos\theta_a)\exp(im\phi_a), \tag{13.15}$$

where R_a, θ_a, ϕ_a are the polar coordinates of nucleus a (with charge Z_a) of molecule A taken in the coordinate system of molecule A. When all such expansions are inserted into the formula for V, we may perform the following chain of transformations

$$\begin{aligned}
V &= -\sum_j\sum_a \frac{Z_a}{r_{aj}} - \sum_i\sum_b \frac{Z_b}{r_{bi}} + \sum_{ij}\frac{1}{r_{ij}} + \sum_a\sum_b \frac{Z_a Z_b}{R_{ab}} \\
&\cong \sum_j\sum_a\sum_{k=0}\sum_{l=0}\sum_{m=-s}^{m=s} A_{kl|m|}R^{-(k+l+1)}\hat{M}_A^{(k,m)}(a)^*\hat{M}_B^{(l,m)}(j) \\
&\quad + \sum_i\sum_b\sum_{k=0}\sum_{l=0}\sum_{m=-s}^{m=s} A_{kl|m|}R^{-(k+l+1)}\hat{M}_A^{(k,m)}(i)^*\hat{M}_B^{(l,m)}(b) \\
&\quad + \sum_{ij}\sum_{k=0}\sum_{l=0}\sum_{m=-s}^{m=s} A_{kl|m|}R^{-(k+l+1)}\hat{M}_A^{(k,m)}(i)^*\hat{M}_B^{(l,m)}(j) \\
&\quad + \sum_a\sum_b\sum_{k=0}\sum_{l=0}\sum_{m=-s}^{m=s} A_{kl|m|}R^{-(k+l+1)}\hat{M}_A^{(k,m)}(a)^*\hat{M}_B^{(l,m)}(b) \\
&= \sum_{k=0}\sum_{l=0}\sum_{m=-s}^{m=s} A_{kl|m|}R^{-(k+l+1)}\Biggl\{\Bigl[\sum_a \hat{M}_A^{(k,m)}(a)\Bigr]^*\Bigl[\sum_j \hat{M}_B^{(l,m)}(j)\Bigr] \\
&\quad + \Bigl[\sum_i \hat{M}_A^{(k,m)}(i)\Bigr]^*\Bigl[\sum_b \hat{M}_B^{(l,m)}(b)\Bigr] + \Bigl[\sum_i \hat{M}_A^{(k,m)}(i)\Bigr]^*\Bigl[\sum_j \hat{M}_B^{(l,m)}(j)\Bigr] \\
&\quad + \Bigl[\sum_a \hat{M}_A^{(k,m)}(a)\Bigr]^*\Bigl[\sum_b \hat{M}_B^{(l,m)}(b)\Bigr]\Biggr\} \\
&= \sum_{k=0}\sum_{l=0}\sum_{m=-s}^{m=s} A_{kl|m|}R^{-(k+l+1)}\Bigl[\sum_a \hat{M}_A^{(k,m)}(a) + \sum_i \hat{M}_A^{(k,m)}(i)\Bigr]^* \\
&\quad \times \Bigl[\sum_b \hat{M}_B^{(l,m)}(b) + \sum_j \hat{M}_B^{(l,m)}(j)\Bigr] \\
&= \sum_{k=0}\sum_{l=0}\sum_{m=-s}^{m=s} A_{kl|m|}R^{-(k+l+1)}\hat{M}_A^{(k,m)*}\hat{M}_B^{(l,m)}.
\end{aligned} \tag{13.16}$$

In the square brackets we can recognize the multipole moment operators for the total molecules calculated in "their" coordinate systems

$$\hat{M}_A^{(k,m)} = \sum_a \hat{M}_A^{(k,m)}(a) + \sum_i \hat{M}_A^{(k,m)}(i),$$
$$\hat{M}_B^{(l,m)} = \sum_b \hat{M}_B^{(l,m)}(b) + \sum_j \hat{M}_B^{(l,m)}(j).$$

Eq. (13.16) has the form of a single multipole expansion, but this time the multipole moment operators correspond to *entire molecules*.

Using the table of multipoles (p. 1042), we may easily write down the multipole operators for the individual molecules. The lowest moment is the net charge (monopole) of the molecules

$$\hat{M}_A^{(0,0)} = q_A = (Z_A - n_A),$$
$$\hat{M}_B^{(0,0)} = q_B = (Z_B - n_B),$$

where Z_A is the sum of all the nuclear charges of molecule A, and n_A is its number of electrons (similarly for B). The next moment is $\hat{M}_A^{(1,0)}$, which is a component of the dipole operator equal to

$$\hat{M}_A^{(1,0)} = -\sum_i z_i + \sum_a Z_a z_a, \tag{13.17}$$

where the small letters z denote the z coordinates of the corresponding particles measured in the coordinate system A (the capital Z denotes the nuclear charge). Similarly, we could very easily write other multipole moments and the operator V takes the form (see Appendix X)

$$V = \frac{q_A q_B}{R} - R^{-2}\big(q_A \hat{\mu}_{Bz} - q_B \hat{\mu}_{Az}\big) + R^{-3}\big(\hat{\mu}_{Ax}\hat{\mu}_{Bx} + \hat{\mu}_{Ay}\hat{\mu}_{By} - 2\hat{\mu}_{Az}\hat{\mu}_{Bz}\big)$$
$$+ R^{-3}\big(q_A \hat{Q}_{B,z^2} + q_B \hat{Q}_{A,z^2}\big) + \cdots,$$

where

$$\hat{\mu}_{Ax} = -\sum_i x_i + \sum_a Z_a x_a,$$
$$\hat{Q}_{A,z^2} = -\sum_i \frac{1}{2}\big(3z_i^2 - r_i^2\big) + \sum_a Z_a \frac{1}{2}\big(3z_a^2 - R_a^2\big)$$

and symbol A means that all these moments are measured in coordinate system A. The other quantities have similar definitions, and are easy to derive. There is one thing that may bother us, namely that $\hat{\mu}_{Bz}$ and $\hat{\mu}_{Az}$ appear in the charge–dipole interaction terms with opposite signs, so are not on equal footing. The reason is that the two coordinate systems are also not on equal footing, because the z coordinate of the coordinate system A points to B, whereas the opposite is not true (see Appendix X).

13.6.4 ELECTROSTATIC ENERGY IN THE MULTIPOLE REPRESENTATION AND THE PENETRATION ENERGY

Electrostatic energy (p. 693) represents the first-order correction in polarization perturbational theory and is the mean value of V with the product wave function $\psi_0^{(0)} = \psi_{A,0}\psi_{B,0}$. Because we have the multipole representation of V, we may insert it into formula (13.5).

Let us stress, for the sake of clarity, that V is an *operator* that contains the *operators* of the molecular multipole moments, and that the integration is, as usual, carried out over the x, y, z, σ coordinates of all electrons (the nuclei have positions fixed in space according to the Born–Oppenheimer approximation), i.e. over the coordinates of electrons 1, 2, 3, etc. Since in the polarization approximation we know perfectly well which electrons belong to molecule A ("we have painted them green"), and which belong to B ("red"), therefore we perform the integration separately over the electrons of molecule A and those of molecule B. We have a comfortable situation, because every term in V represents a *product* of an operator depending on the coordinates of the electrons belonging to A and of an operator depending on the coordinates of the electrons of molecule B. This (together with the fact that in the integral we have a *product* of $|\psi_{A,0}|^2$ and $|\psi_{B,0}|^2$) results in a product of two integrals: one over the electronic coordinates of A and the other one over the electronic coordinates of B. This is the reason why we like multipoles so much.

Therefore,

> the expression for $E_0^{(1)} = E_{\text{elst}}$ *formally* has to be of exactly the same form as the multipole representation of V, the only difference being that in V we have the molecular multipole *operators*, whereas in E_{elst} we have the molecular multipoles themselves as the *mean values* of the corresponding molecular multipole operators in the ground state (the index "0" has been omitted on the right-hand side).

However, the operator V from the formula (13.13) and the operator in the multipole form (13.16) are equivalent only when the multipole form converges. It does so when the interacting objects are non-overlapping, which is *not* the case here. The electronic charge distributions penetrate and this causes a small difference (*penetration energy* E_{penetr}) between the E_{elst} calculated with and without the multipole expansion. The penetration energy vanishes very fast with intermolecular distance R, cf. Appendix R, p. 1009.

$$E_{\text{elst}} = E_{\text{multipol}} + E_{\text{penetr}}, \tag{13.18}$$

where E_{multipol} contains all the terms of the multipole expansion

$$\begin{aligned}E_{\text{multipol}} = {} & \frac{q_A q_B}{R} - R^{-2}(q_A \mu_{Bz} - q_B \mu_{Az}) \\ & + R^{-3}(\mu_{Ax}\mu_{Bx} + \mu_{Ay}\mu_{By} - 2\mu_{Az}\mu_{Bz}) \\ & + R^{-3}(q_A Q_{B,z^2} + q_B Q_{A,z^2}) + \cdots .\end{aligned}$$

The molecular multipoles are

$$\begin{aligned}q_A &= \langle \psi_{A,0}| - \sum_i 1 + \sum_a Z_a |\psi_{A,0}\rangle = \Bigl(-\sum_i 1 + \sum_a Z_a\Bigr)\langle \psi_{A,0}|\psi_{A,0}\rangle \\ &= \sum_a Z_a - n_A = \text{the same as operator } q_A, \\ \mu_{Ax} &= \langle \psi_{A,0}|\hat{\mu}_{Ax}\psi_{A,0}\rangle = \langle \psi_{A,0}| - \sum_i x_i + \sum_a Z_a x_a |\psi_{A,0}\rangle \\ &= \langle \psi_{A,0}| - \sum_i x_i |\psi_{A,0}\rangle + \sum_a Z_a x_a \qquad (13.19)\end{aligned}$$

and similarly the other multipoles.

Since the multipoles in the formula for E_{multipol} pertain to the isolated molecules, we may say that the electrostatic interaction represents the interaction of the permanent multipoles.

permanent multipoles

The above multipole expansion also represents a useful source for the expressions for particular multipole–multipole interactions.

Dipole–dipole

Let us take as an example of the important case of the dipole–dipole interaction.

From the above formulae the dipole–dipole interaction reads as

$$E_{\text{dip–dip}} = \frac{1}{R^3}(\mu_{Ax}\mu_{Bx} + \mu_{Ay}\mu_{By} - 2\mu_{Az}\mu_{Bz}).$$

This is a short and easy to memorize formula, and we might be completely satisfied in using it *provided we always remember the particular coordinate system used for its derivation*. This may end up badly one day for those who have a short memory. Therefore, we will write down the same formula in a "waterproof" form.

Taking into account our coordinate system, the vector (pointing the coordinate system origin a from b) is $\boldsymbol{R} = (0, 0, R)$. Then we can express $E_{\text{dip–dip}}$ in a very useful form *independent of any choice of coordinate system* (cf., e.g., pp. 131, 655):

DIPOLE–DIPOLE INTERACTION:

$$E_{\text{dip–dip}} = \frac{\boldsymbol{\mu}_A \cdot \boldsymbol{\mu}_B}{R^3} - 3\frac{(\boldsymbol{\mu}_A \cdot \boldsymbol{R})(\boldsymbol{\mu}_B \cdot \boldsymbol{R})}{R^5}. \qquad (13.20)$$

This form of the dipole–dipole interaction has been used in Chapters 3 and 12.

Is the electrostatic interaction important?

Electrostatic interaction can be attractive or repulsive. For example, in the electrostatic interaction of Na^+ and Cl^- the main role will be played by the charge–

charge interaction, which is negative and therefore represents attraction, while for $Na^+ \ldots Na^+$ the electrostatic energy will be positive (repulsion). For neutral molecules the electrostatic interaction may depend on their *orientation* to such an extent that the sign may change. This is an exceptional feature peculiar only to electrostatic interaction.

When the distance R is small when compared to size of the interacting subsystems, multipole expansion gives bad results. To overcome this the total charge distribution may be divided into *atomic* segments (Appendix S). Each atom would carry its charge and other multipoles, and the electrostatic energy would be the sum of the atom–atom contributions, any of which would represent a series similar[13] to $E_0^{(1)}$.

Reality or fantasy?

In principle, this part (about electrostatic interactions) may be considered as completed. I am tempted, however, to enter some "obvious" subjects, which will turn out to lead us far away from the usual track of intermolecular interactions.

Let us consider the Coulomb interaction of two point charges q_1 on molecule A and q_2 on molecule B, both charges separated by distance r

$$E_{\text{elst}} = \frac{q_1 q_2}{r}. \qquad (13.21)$$

This is an outstanding formula:

- first of all we have the amazing exponent of the *exact* value -1;
- second, change of the charge *sign* does not make any profound changes in the formula, except the change of *sign* of the interaction energy;
- third, the formula is bound to be false (it has to be only an approximation), since instantaneous interaction is assumed, whereas the interaction has to have time to travel between the interacting objects and during that time the objects change their distance (see Chapter 3, p. 131).

From these remarks follow some apparently obvious observations, that E_{elst} is invariant with respect to the following operations:

II $q_1' = -q_1, \quad q_2' = -q_2$ (charge conjugation, Chapter 2, 2.1.8),
III $q_1' = q_2, \quad q_2' = q_1$ (*exchange* of charge positions),
IV $q_1' = -q_2, \quad q_2' = -q_1$ (charge conjugation *and* exchange of charge positions).

These invariance relations, when treated literally and rigorously, are not of particular usefulness in theoretical chemistry. They may, however, open new possibilities when considered as some limiting cases. Chemical reaction mechanisms very often involve the interaction of molecular ions. Suppose we have a particular reaction mechanism. Now, let us make the charge conjugation of all the objects involved in the reaction (this would require the change of matter to antimatter).

---

[13] A.J. Stone, *Chem. Phys. Lett.* 83 (1981) 233; A.J. Stone, M. Alderton, *Mol. Phys.* 56 (1985) 1047; W.A. Sokalski, R. Poirier, *Chem. Phys. Lett.* 98 (1983) 86; W.A. Sokalski, A. Sawaryn, *J. Chem. Phys.* 87 (1987) 526.

This will preserve the reaction mechanism. We cannot do such changes in chemistry. However, we may think of some *other molecular systems*, which have similar geometry but opposite overall charge pattern ("counter pattern"). The new reaction has a chance to run in a similar direction as before. This concept is parallel to the idea of *Umpolung* functioning in organic chemistry. It seems that nobody has looked, from that point of view, at all known reaction mechanisms.[14] Umpolung

13.6.5 INDUCTION ENERGY IN THE MULTIPOLE REPRESENTATION

> The induction energy contribution consists of two parts: $E_{\text{ind}}(A \to B)$ and $E_{\text{ind}}(B \to A)$ or, respectively, the polarization energy of molecule B in the electric field of the unperturbed molecule A and *vice versa*.

The goal of the present section is to take apart the induction mechanism by showing its multipole components. If we insert the multipole representation of V into the induction energy $E_{\text{ind}}(A \to B)$ then

$$
\begin{aligned}
E_{\text{ind}}(A \to B) &= \sum_{n_B}{}' \frac{|\langle \psi_{A,0}\psi_{B,n_B}|V\psi_{A,0}\psi_{B,0}\rangle|^2}{E_{B,0}-E_{B,n_B}} \\
&= \sum_{n_B}{}' \frac{1}{E_{B,0}-E_{B,n_B}} \bigl\{ \bigl| R^{-1}q_A \cdot 0 - R^{-2}q_A \langle \psi_{B,n_B}|\hat{\mu}_{Bz}\psi_{B,0}\rangle + R^{-2}\cdot 0 \\
&\quad + R^{-3}\bigl[\mu_{Ax}\langle \psi_{B,n_B}|\hat{\mu}_{Bx}\psi_{B,0}\rangle + \mu_{Ay}\langle \psi_{B,n_B}|\hat{\mu}_{By}\psi_{B,0}\rangle \\
&\quad - 2\mu_{Az}\langle \psi_{B,n_B}|\hat{\mu}_{Bz}\psi_{B,0}\rangle\bigr] + \cdots \bigr| \bigr\}^2 \\
&= \sum_{n_B}{}' \frac{1}{E_{B,0}-E_{B,n_B}} \bigl\{ \bigl| -R^{-2}q_A \langle \psi_{B,n_B}|\hat{\mu}_{Bz}\psi_{B,0}\rangle \\
&\quad + R^{-3}\bigl[\mu_{Ax}\langle \psi_{B,n_B}|\hat{\mu}_{Bx}\psi_{B,0}\rangle + \mu_{Ay}\langle \psi_{B,n_B}|\hat{\mu}_{By}\psi_{B,0}\rangle \\
&\quad - 2\mu_{Az}\langle \psi_{B,n_B}|\hat{\mu}_{Bz}\psi_{B,0}\rangle\bigr] + \cdots \bigr| \bigr\}^2 \\
&= -\frac{1}{2}\frac{1}{R^4}q_A^2\alpha_{B,zz} + \cdots,
\end{aligned}
$$

where

- the zeros appearing in the first part of the derivation come from the orthogonality of the eigenstates of the isolated molecule B,
- symbol "$+\cdots$" stands for higher powers of R^{-1},
- $\alpha_{B,zz}$ represents the zz component of the dipole polarizability tensor of the molecule B, which absorbed the summation over the excited states of B according to definition (12.40).

---

[14] The author is aware of only a single example of such a pair of counter patterns: the Friedel–Crafts reaction and what is called the vicarious nucleophilic substitution discovered by Mieczysław Mąkosza (M. Mąkosza, A. Kwast, *J. Phys. Org. Chem.* 11 (1998) 341).

A molecule in the electric field of another molecule

Note that $\frac{1}{R^4}q_A^2$ represents the square of the electric field intensity $\mathcal{E}_z(A \to B) = \frac{q_A}{R^2}$ measured on molecule B and created by the net charge of molecule A. Therefore, we have

$$E_{\text{ind}}(A \to B) = -\frac{1}{2}\alpha_{B,zz}\mathcal{E}_z^2(A \to B) + \cdots$$

according to formula (12.24) describing the molecule in an electric field. For molecule B its partner – molecule A (and *vice versa*…) represents an external world creating the electric field, and molecule B has to behave as described in Chapter 12. The net charge of A created the electric field $\mathcal{E}_z(A \to B)$ on molecule B, which as a consequence induced on B a dipole moment $\mu_{B,\text{ind}} = \alpha_{B,zz}\ \mathcal{E}_z(A \to B)$ according to formula (12.19). This is associated with the interaction energy term $-\frac{1}{2}\alpha_{B,zz}\mathcal{E}_z^2(A \to B)$, see eq. (12.24), p. 628.

There is however a small problem. Why is the induced moment proportional only to the net charge of molecule A? This would be absurd. Molecule B does not know anything about multipoles of molecule A, it only knows about the *local* electric field that acts on it and has to react to that field by a suitable polarization. Everything is all right, though. The rest of the problem is in the formula for $E_{\text{ind}}(A \to B)$. So far we have analyzed the electric field on B coming from the net charge of A, but the other terms of the formula will give contributions to the electric field coming from *all other* multipole moments of A. Then, the response of B will pertain to the total electric field created by "frozen" A on B, as it should be. A similar story can be given for $E_{\text{ind}}(B \to A)$. This is all we have in the induction energy (second-order perturbation theory). Interaction of the induced multipoles of A and B is a subject of the third-order terms.

13.6.6 DISPERSION ENERGY IN THE MULTIPOLE REPRESENTATION

After inserting V in the multipole representation (p. 701) into the expression for the dispersion energy we obtain

$$\begin{aligned}
E_{\text{disp}} &= \sum_{n_A}{}' \sum_{n_B}{}' \frac{1}{(E_{A,0} - E_{A,n_A}) + (E_{B,0} - E_{B,n_B})} \\
&\quad \times \big| R^{-1} q_A q_B \cdot 0 \cdot 0 - R^{-2} q_A \cdot 0 \cdot (\mu_{Bz})_{n_B,0} \\
&\quad - R^{-2} q_B \cdot 0 \cdot (\mu_{Az})_{n_A,0} + R^{-3}\big[(\mu_{Ax})_{n_A,0}(\mu_{Bx})_{n_B,0} + (\mu_{Ay})_{n_A,0}(\mu_{By})_{n_B,0} \\
&\quad - 2(\mu_{Az})_{n_A,0}(\mu_{Bz})_{n_B,0}\big] + \cdots \big|^2 \\
&= \sum_{n_A}{}' \sum_{n_B}{}' \big| R^{-3}\big[(\mu_{Ax})_{n_A,0}(\mu_{Bx})_{n_B,0} + (\mu_{Ay})_{n_A,0}(\mu_{By})_{n_B,0} \\
&\quad - 2(\mu_{Az})_{n_A,0}(\mu_{Bz})_{n_B,0}\big] + \cdots \big|^2 \big[(E_{A,0} - E_{A,n_A}) + (E_{B,0} - E_{B,n_B})\big]^{-1}
\end{aligned}$$

where $(\mu_{Ax})_{n_A,0} = \langle\psi_{A,n_A}|\hat{\mu}_{Ax}\psi_{A,0}\rangle$, $(\mu_{Bx})_{n_B,0} = \langle\psi_{B,n_B}|\hat{\mu}_{Bx}\psi_{B,0}\rangle$ and similarly the other quantities. The zeros in the first part of the equality chain come from the orthogonality of the eigenstates of each of the molecules.

The square in the formula pertains to all terms. The other terms, not shown in the formula, have the powers of R^{-1} higher than R^{-3}.

Hence, if we squared the total expression, the most important term would be the dipole–dipole contribution with the asymptotic R^{-6} distance dependence.

As we can see from formula (13.12), its calculation requires *double* electronic excitations (one on the first, the other one on the second interacting molecules), and these already belong to the correlation effect (cf. Chapter 10, p. 558).

The dispersion interaction is a pure correlation effect and therefore the methods used in a supermolecular approach, that do not take into account the electronic correlation (as for example the Hartree–Fock method) are unable to produce any non-zero dispersion contribution.

Where does this physical effect come from?

Imagine we have two hydrogen atoms, each in its ground state, i.e. 1*s* state, and with a long internuclear distance R. Let us simplify things as much as possible and give only the possibility of two positions for each of the two electrons: one closer to the other proton and the opposite (crosses in Fig. 13.6), the electron–proton distance being $a \ll R$. Let us calculate the instantaneous dipole–dipole interactions for all four possible situations from formula (13.20) assuming the local coordinate systems on the protons (Table 13.1).

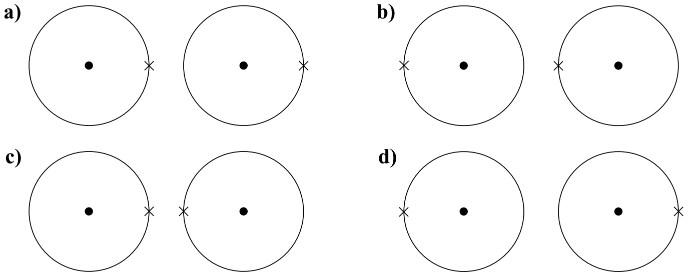


Fig. 13.6. Dispersion energy origin shown schematically for two hydrogen atoms. A popular explanation for the dispersion interaction is that, due to electron repulsion: the situations (a) and (b) occur more often than situation (c) and this is why the dispersion interaction represents a net attraction of dipoles. The positions of the electrons that correspond to (a) and (b) represent two favourable instantaneous dipole – instantaneous dipole interactions, while (c) corresponds to a non-favourable instantaneous dipole – instantaneous dipole interaction. The trouble with this explanation is that there is also the possibility of having electrons far apart as in (d). This most favourable situation (the longest distance between the electrons) means, however, repulsion of the resulting dipoles. It may be shown, though, that the net result (dispersion interaction) is still an attraction (see the text) as it should be.

Table 13.1.

Situation, i	Fig. 13.6	Interaction energy $E_{\text{int}}(i)$
1	a	$-2\frac{\mu^2}{R^3}$
2	b	$-2\frac{\mu^2}{R^3}$
3	c	$+2\frac{\mu^2}{R^3}$
4	d	$+2\frac{\mu^2}{R^3}$

Here $\boldsymbol{\mu} = (0, 0, \pm a)$ for electrons $i = 1, 2$ according to definition (13.17), and $\mu \equiv a$ in a.u. Note that if we assume the same probability for each situation, the net energy would be zero, i.e. $\sum_i E_{\text{int}}(i) = 0$. These situations have, however, different probabilities (p_i), because the electrons repel each other, and the total potential energy depends on where they actually are. *Note, that the probabilities should be different only because of the electron correlation*. In this total energy, there is a common contribution, identical in all the four situations: the interaction within the individual atoms [the remainder is the interaction energy $E_{\text{int}}(i)$]. If we could somehow guess these probabilities p_i, $i = 1, 2, 3, 4$, then we could calculate the mean interaction energy of our model one-dimensional atoms as $\bar{E}_{\text{int}} = \sum_i p_i E_{\text{int}}(i)$. In this way we could see whether it corresponds to net attraction ($\bar{E}_{\text{int}} < 0$) or repulsion ($\bar{E}_{\text{int}} > 0$), which is most interesting for us. Well, but how to calculate them?[15] We may suspect that for the ground state (we are interested in the ground state of our system) the lower the potential energy $V(x)$ the higher the probability density $p(x)$. This is what happens for the harmonic oscillator, for the Morse oscillator, for the hydrogen-like atom, etc. Is there any tip that could help us work out what such a dependence might be? If you do not know where to begin, then think of the harmonic oscillator model as a starting point! This is what people usually do as a first guess. As seen from eq. (4.16), the ground-state wave function for the harmonic oscillator may be written as $\psi_0 = A\exp[-BV(x)]$, where $B > 0$, and $V(x)$ stands for the potential energy for the harmonic oscillator. Therefore the probability density changes as $A^2\exp[-2BV(x)]$. Interesting… Let us assume that a similar thing happens[16] for the probabilities p_i of finding the electrons 1 and 2 in small cubes of volumes dV_1 and dV_2, respectively, i.e. they may be reasonably estimated as

$$p_i = NA^2\exp\bigl[-2BE_{\text{int}}(i)\bigr]dV_1 dV_2,$$

where E_{int} plays a role of potential energy, and

$$N = 1\Big/\Bigl(\sum_i A^2\exp\bigl[-2BE_{\text{int}}(i)\bigr]dV_1 dV_2\Bigr)$$

---

[15] In principle we could look at what people have calculated in the most sophisticated calculations for the hydrogen molecule at a large R, and assign the p_i's as the squares of the wave function value for the corresponding four positions of both electrons. Since these wave functions are awfully complex, we leave this path without regret.

[16] This is like having the electron attached to the nucleus by a harmonic spring (instead of Coulombic attraction).

is the normalization constant assuring that in our model $\sum_i p_i = 1$. For long distances R [small $E_{\mathrm{int}}(i)$] we may expand this expression in a Taylor series and obtain

$$\begin{aligned} p_i &= \frac{A^2[1-2BE_{\mathrm{int}}(i)]\,\mathrm{d}V_1\,\mathrm{d}V_2}{\sum_j A^2\exp[-2BE_{\mathrm{int}}(j)]\,\mathrm{d}V_1\,\mathrm{d}V_2} \approx \frac{1-2BE_{\mathrm{int}}(i)+\cdots}{\sum_j(1-2BE_{\mathrm{int}}(j)+\cdots)} \\ &= \frac{1-2BE_{\mathrm{int}}(i)+\cdots}{4-2B\cdot 0+\sum_j \frac{1}{2}[2BE_{\mathrm{int}}(j)]^2+\cdots} \approx \frac{1}{4}-\frac{B}{2}E_{\mathrm{int}}(i), \end{aligned}$$

where the Taylor series has been truncated to the accuracy of the linear terms in the interaction. Then, the mean interaction energy

$$\begin{aligned} \bar{E}_{\mathrm{int}} &= \sum_i p_i E_{\mathrm{int}}(i) \approx \sum_i \left[\frac{1}{4}-\frac{B}{2}E_{\mathrm{int}}(i)\right]E_{\mathrm{int}}(i) \\ &= \frac{1}{4}\sum_i E_{\mathrm{int}}(i)-\frac{B}{2}\sum_i \left[E_{\mathrm{int}}(i)\right]^2 = 0-\frac{B}{2}\frac{16\mu^4}{R^6} = -8B\frac{\mu^4}{R^6} < 0. \end{aligned}$$

> We may not expect our approximation to be extremely accurate, but it is worth noting that we have grasped two important features of the correct dispersion energy: that it corresponds to attractive interaction and that it vanishes with distance as R^{-6}.

Examples

The electrostatic interaction energy of two molecules can be calculated from formula (13.5). However, it is very important for a chemist to be able to predict the main features of the electrostatic interaction *without any calculation* at all, based on some general rules. This will create chemical intuition or chemical common sense so important in planning, performing and understanding experiments. The data of Table 13.2 were obtained assuming a long intermolecular distance and the molecular orientations as shown in the table.

In composing Table 13.2 some helpful rules have been used:

- *Induction and dispersion energies always represent attraction, except in some special cases when they are zero*. These special cases are obvious, e.g., it is impossible to induce some changes on molecule B, if molecule A does not have any non-zero permanent multipoles. Also, the dispersion energy is zero if an interacting subsystem has no electrons on it.
- *Electrostatic energy is non-zero, if both interacting molecules have some non-zero permanent multipoles*.
- *Electrostatic energy is negative (positive), if the lowest non-vanishing multipoles of the interacting partners attract (repel) themselves*.[17] How to recognize that a particular multipole–multipole interaction represents attraction or repulsion? First we replace the molecules by their lowest non-zero multipoles represented by

---

[17] This statement is true for sufficiently long distances.

Table 13.2. The table pertains to two molecules in their electronic ground states. For each pair of molecules a short characteristic of their electrostatic, induction and dispersion interactions is given. It consists of the sign of the corresponding interaction type (the minus sign means attraction, the plus sign means repulsion and 0 corresponds to the absence of such an interaction, the penetration terms have been neglected)

System	Electrost.	Induc.	Disper.
$\mathrm{He}\cdots\mathrm{He}$	0	0	−
$\mathrm{He}\cdots\mathrm{H}^+$	0	−	0
$\mathrm{He}\cdots\mathrm{HCl}$	0	−	−
$\mathrm{H}^+\cdots\mathrm{HCl}$	+	−	0
$\mathrm{HCl}\cdots\mathrm{ClH}$	+	−	−
$\mathrm{HCl}\cdots\mathrm{HCl}$	−	−	−
$\mathrm{H{-}H}\cdots\mathrm{He}$	0	−	−
$\mathrm{H{-}H}\cdots\mathrm{H{-}H}$	+	−	−
$\begin{smallmatrix}\mathrm{H}\\\mathrm{H}\end{smallmatrix}\cdots\mathrm{H{-}H}$	−	−	−
$\begin{smallmatrix}\mathrm{H}\diagdown\\\mathrm{H}\diagup\end{smallmatrix}\mathrm{O}\cdots\mathrm{H{-}O}_{\diagdown\mathrm{H}}$	−	−	−
$\begin{smallmatrix}\mathrm{H}\diagdown\\\mathrm{H}\diagup\end{smallmatrix}\mathrm{O}\cdots\mathrm{O}\begin{smallmatrix}\diagup\mathrm{H}\\\diagdown\mathrm{H}\end{smallmatrix}$	+	−	−

Table 13.3. The exponent m in the asymptotic dependence R^{-m} of the electrostatic (column 2), induction (column 3) and dispersion (column 4) contributions for the systems given in column 1. Zero denotes that the corresponding contribution is equal to zero in the multipole approximation

System	Electrost.	Induc.	Disper.
$\mathrm{He}\cdots\mathrm{He}$	0	0	6
$\mathrm{He}\cdots\mathrm{H}^+$	0	4	0
$\mathrm{He}\cdots\mathrm{HCl}$	0	6	6
$\mathrm{H}^+\cdots\mathrm{HCl}$	2	4	0
$\mathrm{HCl}\cdots\mathrm{ClH}$	3	6	6
$\mathrm{HCl}\cdots\mathrm{HCl}$	3	6	6
$\mathrm{H{-}H}\cdots\mathrm{He}$	0	8	6
$\mathrm{H{-}H}\cdots\mathrm{H{-}H}$	5	8	6
$\begin{smallmatrix}\mathrm{H}\\\mathrm{H}\end{smallmatrix}\cdots\mathrm{H{-}H}$	5	8	6
$\begin{smallmatrix}\mathrm{H}\diagdown\\\mathrm{H}\diagup\end{smallmatrix}\mathrm{O}\cdots\mathrm{H{-}O}_{\diagdown\mathrm{H}}$	3	6	6
$\begin{smallmatrix}\mathrm{H}\diagdown\\\mathrm{H}\diagup\end{smallmatrix}\mathrm{O}\cdots\mathrm{O}\begin{smallmatrix}\diagup\mathrm{H}\\\diagdown\mathrm{H}\end{smallmatrix}$	3	6	6

point charges, e.g., ions by $+$ or $-$, dipolar molecules by $+-$, quadrupoles by $+^{-}_{-}+$, etc. In order to do this we have to know which atoms are electronegative and which electropositive.[18] After doing this we replace the two molecules by the multipoles. If the nearest neighbour charges in the two multipoles are of opposite sign, the multipoles attract each other, otherwise they repel (Fig. 13.7).

---

[18] This is common knowledge in chemistry and is derived from experiments as well as from quantum mechanical calculations. The later provides the partial atomic charges from what is called population analysis (see Appendix S). Despite its non-uniqueness it would satisfy our needs. A unique and elegant method of calculation of atomic partial charges is related to the Bader analysis described on p. 573.

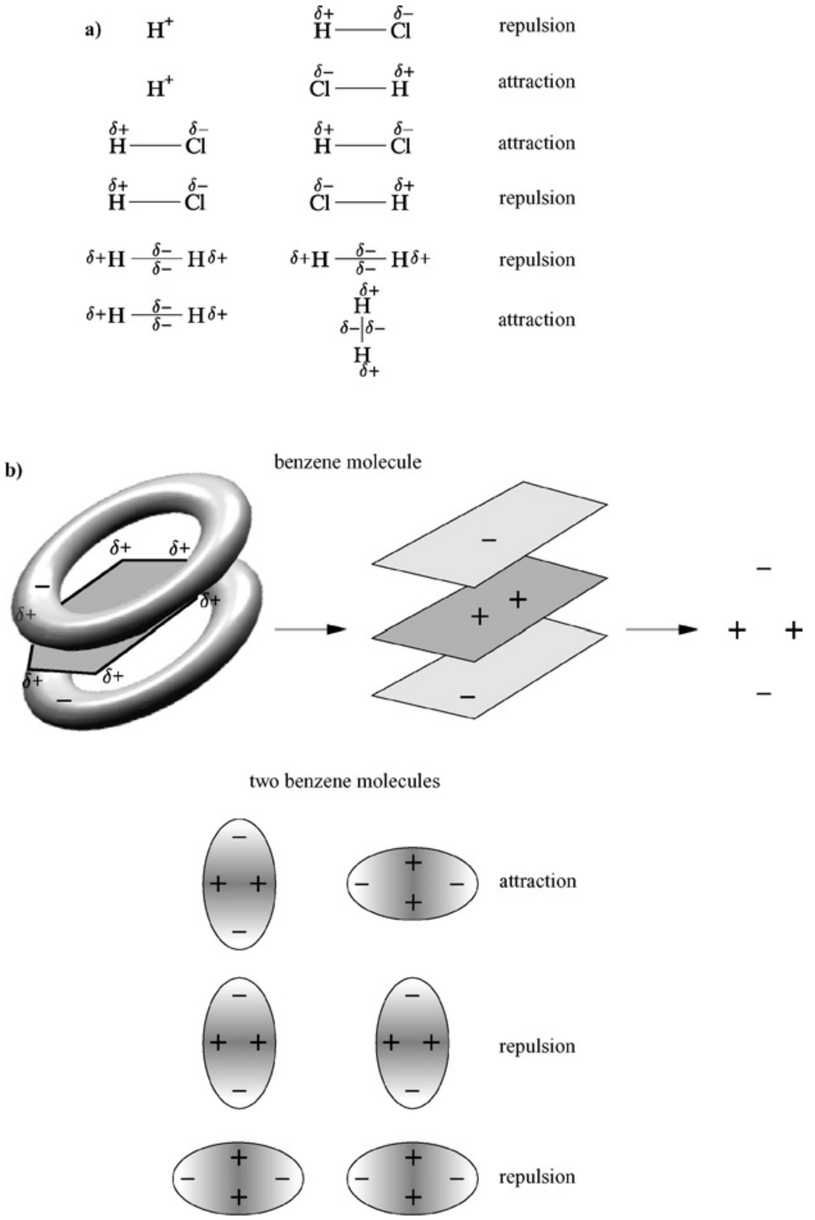


Fig. 13.7. For sufficiently large intermolecular separations the interaction of the lowest non-vanishing multipoles dominates. Whether this is an attraction or repulsion can be recognized by representing the molecular charge distributions by non-point-like multipoles (clusters of point charges). If such multipoles point to each other by point charges of the opposite (same) sign, then the electrostatic interaction of the molecules is attraction (repulsion). (a) A few examples of simple molecules and the atomic partial charges. (b) Even the interaction of the two benzene molecules obeys this rule: in the face-to-face configuration they repel, while they attract each other in the perpendicular configuration.

Since we can establish which effect dominates, its asymptotic dependence (Table 13.3), as the intermolecular distance R tends to ∞, can be established.

Table 13.3 was composed using a few simple and useful rules:

1. The dispersion energy always decays as R^{-6}.
2. The electrostatic energy vanishes as $R^{-(k+l+1)}$, where the 2^k-pole and 2^l-pole represent the lowest non-vanishing multipoles of the interacting subsystems.
3. The induction energy vanishes as $R^{-2(k+2)}$, where the 2^k-pole is the lower of the two lowest non-zero permanent multipoles of the molecules A and B. The formula is easy to understand if we take into account that the lowest *induced* multipole is always a dipole ($l = 1$), and that the induction effect is of the second order (hence 2 in the exponent).

13.7 SYMMETRY ADAPTED PERTURBATION THEORIES (SAPT)

The SAPT approach is applicable for intermediate intermolecular separations, where the electron clouds of both molecules overlap to such an extent, that

- the polarization approximation, i.e. ignoring the Pauli principle (p. 692), becomes a very poor approximation,
- the multipole expansion becomes invalid.

13.7.1 POLARIZATION APPROXIMATION IS ILLEGAL

First, the polarization approximation zero-order wave function $\psi_{A,0}\psi_{B,0}$ will be deprived of the privilege of being the unperturbed function $\psi_0^{(0)}$ in a perturbation theory. Since it will still play an important role in the theory, let us denote it by $\varphi^{(0)} = \psi_{A,0}\psi_{B,0}$.

> The polarization approximation seems to have (at first glimpse) a very strong foundation, because at long intermolecular distances R, the zero-order *energy* is close to the exact one. The trouble is, however, that a similar statement is not true for the zero-order *wave function* $\varphi^{(0)}$ and the exact wave-function at any intermolecular distance (even at infinity).

Let us take an example of two ground-state hydrogen atoms. The polarization approximation zero-order wave function

$$\varphi^{(0)}(1,2) = 1s_a(1)\alpha(1)1s_b(2)\beta(2), \tag{13.22}$$

where the spin functions have been introduced (the Pauli principle is ignored[19])

> This function is neither symmetric (since $\varphi^{(0)}(1,2) \neq \varphi^{(0)}(2,1)$), nor antisymmetric (since $\varphi^{(0)}(1,2) \neq -\varphi^{(0)}(2,1)$), and therefore is “illegal” and in principle not acceptable.

---

[19] This is the essence of the polarization approximation.

13.7.2 CONSTRUCTING A SYMMETRY ADAPTED FUNCTION

zero-order wave function

In the Born–Oppenheimer approximation the electronic ground-state wave function of H_2 has to be the eigenfunction of the nuclear inversion symmetry operator $\hat{I}$ interchanging nuclei a and b (cf. Appendix C). Since $\hat{I}^2 = 1$, the eigenvalues can be either -1 (called u symmetry) or $+1$ (g symmetry).[20] The ground-state is of g symmetry, therefore the projection operator $\frac{1}{2}(1+\hat{I})$ will take care of that (it says: make fifty-fifty combination of a function and its counterpart coming from the exchange of nuclei a and b).[21] On top of this, the wave function has to fulfil the Pauli exclusion principle, which we will ensure with the antisymmetrizer $\hat{A}$ (cf. p. 986). Altogether the proper symmetry will be assured by projecting $\varphi^{(0)}$ using the projection operator

$$\hat{\mathcal{A}} = \frac{1}{2}(1+\hat{I})\hat{A}. \tag{13.23}$$

We obtain as a zero-order approximation to the wave function (N ensures normalization)

$$\begin{aligned}
\psi_0^{(0)} &= N\hat{A}\frac{1}{2}(1+\hat{I})\varphi^{(0)} = \frac{1}{2!}N\frac{1}{2}(1+\hat{I})\sum_P(-1)^p\hat{P}\big[1s_a(1)\alpha(1)1s_b(2)\beta(2)\big]\\
&= \frac{1}{2}N\frac{1}{2}(1+\hat{I})\big[1s_a(1)\alpha(1)1s_b(2)\beta(2) - 1s_a(2)\alpha(2)1s_b(1)\beta(1)\big]\\
&= \frac{1}{2}N\frac{1}{2}\big[1s_a(1)\alpha(1)1s_b(2)\beta(2) - 1s_a(2)\alpha(2)1s_b(1)\beta(1)\\
&\quad + 1s_b(1)\alpha(1)1s_a(2)\beta(2) - 1s_b(2)\alpha(2)1s_a(1)\beta(1)\big]\\
&= N\frac{1}{2}\big[1s_a(1)1s_b(2) + 1s_a(2)1s_b(1)\big]\left\{\frac{1}{2}\big[\alpha(1)\beta(2) - \alpha(2)\beta(1)\big]\right\}.
\end{aligned}$$

Heitler–London wave function

This is precisely the Heitler–London wave function from p. 521, where its important role in chemistry has been highlighted:

$$\psi_{\mathrm{HL}} \equiv \psi_0^{(0)} = N\big[1s_a(1)1s_b(2) + 1s_a(2)1s_b(1)\big]\left\{\frac{1}{2}\big[\alpha(1)\beta(2) - \alpha(2)\beta(1)\big]\right\}. \tag{13.24}$$

The function is of the same symmetry as the exact solution to the Schrödinger equation (antisymmetric with respect to the exchange of electrons and symmetric with respect to the exchange of protons). It is easy to calculate,[22] that normaliza-

---

[20] The symbols come from German: g or *gerade* (even) and u or *ungerade* (odd).

[21] We ignore the proton spins.

[22]

$$\begin{aligned}
\int |\psi_0^{(0)}|^2\,\mathrm{d}\tau_1\,\mathrm{d}\tau_2 &= |N|^2\frac{1}{4}\frac{1}{2}\sum_{\sigma_1}\sum_{\sigma_2}\frac{1}{2}\big[\alpha(1)\beta(2) - \alpha(2)\beta(1)\big]^2\big[2+2S^2\big]\\
&= |N|^2\frac{1}{4}(1+S^2) = 1,
\end{aligned} \tag{13.25}$$

tion of $\psi^{(0)}$ means $N = 2[(1+S^2)]^{-1/2}$, where $S = (1s_a|1s_b)$ stands for the overlap integral of the atomic orbitals $1s_a$ and $1s_b$.

13.7.3 THE PERTURBATION IS ALWAYS LARGE IN POLARIZATION APPROXIMATION

Let us check (Appendix B) how distant are functions $\varphi^{(0)}$ and $\psi^{(0)}$ in the Hilbert space (they are both normalized, i.e. they are unit vectors in the Hilbert space). We will calculate the norm of difference $\varphi^{(0)} - \psi_0^{(0)}$. If the norm were small, then the two functions would be close in the Hilbert space. Let us see:

$$\begin{aligned}
&\|\varphi^{(0)} - \psi_0^{(0)}\| \\
&\quad \equiv \left[\int \left(\varphi^{(0)} - \psi_0^{(0)}\right)^* \left(\varphi^{(0)} - \psi_0^{(0)}\right) \mathrm{d}\tau\right]^{\frac{1}{2}} = \left[1 + 1 - 2\int \psi_0^{(0)} \varphi^{(0)} \,\mathrm{d}\tau\right]^{\frac{1}{2}} \\
&\quad = \left\{2 - 2\int \left[1s_a(1)\alpha(1)1s_b(2)\beta(2)\right] N\frac{1}{2}\left[1s_a(1)1s_b(2) + 1s_a(2)1s_b(1)\right]\right. \\
&\qquad \left. \times \left\{\frac{1}{2}\left[\alpha(1)\beta(2) - \alpha(2)\beta(1)\right]\right\} \mathrm{d}\tau\right\}^{\frac{1}{2}} \\
&\quad = \left\{2 - N\frac{1}{2}\int \left[1s_a(1)1s_b(2)\right]\left[1s_a(1)1s_b(2) + 1s_a(2)1s_b(1)\right] \mathrm{d}v\right\}^{\frac{1}{2}} \\
&\quad = \left\{2 - \frac{1}{\sqrt{1+S^2}}\left(1+S^2\right)\right\}^{\frac{1}{2}} = \left\{2 - \sqrt{1+S^2}\right\}^{1/2}
\end{aligned}$$

where we have assumed that the functions are real. When $R \to \infty$, then $S \to 0$ and

$$\lim_{R\to\infty} \|\varphi^{(0)} - \psi_0^{(0)}\| = 1 \neq 0. \tag{13.27}$$

Thus, the Heitler–London wave function differs from $\varphi^{(0)}$, this difference is huge and *does not vanish*, when $R \to \infty$.

The two normalized functions $\varphi^{(0)}$ and $\psi_0^{(0)}$ represent two unit vectors in the Hilbert space. The scalar product of the two unit vectors $\langle\varphi^{(0)}|\psi_0^{(0)}\rangle$ is equal to $\cos\theta$. Let us calculate this angle θ_{lim} which corresponds to R tending to ∞. The quantity

$$\begin{aligned}
\lim_{R\to\infty} \|\varphi^{(0)} - \psi_0^{(0)}\|^2 &= \lim_{R\to\infty} \int \left(\varphi^{(0)} - \psi_0^{(0)}\right)^* \left(\varphi^{(0)} - \psi_0^{(0)}\right) \mathrm{d}\tau \\
&= \lim_{R\to\infty} \left[2 - 2\cos\theta\right] = 1.
\end{aligned}$$

$$N = \frac{2}{\sqrt{1+S^2}}. \tag{13.26}$$

In a moment we will need function $\psi_0^{(0)}$ with the intermediate normalization with respect to $\varphi^{(0)}$, i.e. satisfying $\langle\psi_0^{(0)}|\varphi^{(0)}\rangle = 1$. Then N will be different and equal to $\langle\varphi^{(0)}|\hat{\mathcal{A}}\varphi^{(0)}\rangle^{-1}$.

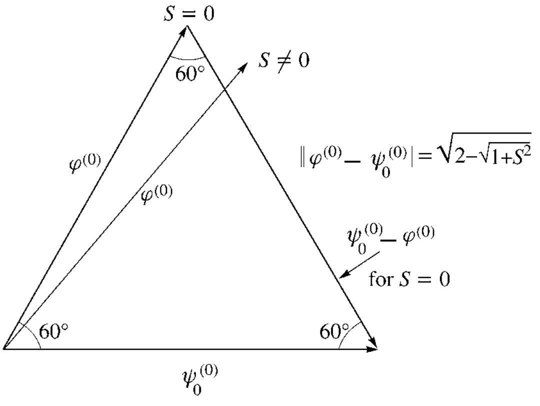


Fig. 13.8. The normalized functions $\varphi^{(0)}$ and $\psi_0^{(0)}$ for the hydrogen molecule as unit vectors belonging to the Hilbert space. The functions differ widely at any intermolecular distance R. For $S=0$, i.e. for long internuclear distances the difference $\psi_0^{(0)}-\varphi^{(0)}$ represents a vector of the Hilbert space having the length 1. Therefore, for $R=\infty$ the three vectors $\varphi^{(0)}$, $\psi_0^{(0)}$ and $\psi_0^{(0)}-\varphi^{(0)}$ form an equilateral angle. For shorter distances the angle between $\varphi^{(0)}$ and $\psi_0^{(0)}$ becomes smaller than 60°.

Hence, $\cos\theta_{\text{lim}}=\frac{1}{2}$, and therefore $\theta_{\text{lim}}=60°$, see Fig. 13.8. This means that the three unit vectors: $\varphi^{(0)}$, $\psi_0^{(0)}$ and $\varphi^{(0)}-\psi_0^{(0)}$ for $R\to\infty$ form an equilateral triangle, and therefore, $\varphi^{(0)}$ represents a highly "handicapped" function, which lacks about a half with respect to a function of the proper symmetry.[23] This is certainly bad news.

Therefore, the perturbation V has to be treated as *always large*, because it is responsible for a huge wave function change: from the unperturbed one of bad symmetry to the exact one of the correct symmetry.

In contrast to this, there would be no problem at all with the vanishing of the $\|\psi_0^{(0)}-\psi_0\|$ as $R\to\infty$, *where* ψ_0 *represents the ground state solution of the Schrödinger equation. Indeed,* $\psi_0^{(0)}$ *correctly describes the dissociation of the molecule into two hydrogen atoms (both in the 1s state), as well as both functions having the same symmetry for all interatomic distances.* Therefore,

the Heitler–London wave function represents a good approximation to the exact function for long (and we hope medium) intermolecular distances. *Unfortunately, it is not the eigenfunction of the* $\hat{H}^{(0)}$ and therefore we cannot construct the usual Rayleigh–Schrödinger perturbation theory.

And this is the second item of bad news today...

13.7.4 ITERATIVE SCHEME OF THE SYMMETRY ADAPTED PERTURBATION THEORY

We now have two issues: either to construct another zero-order Hamiltonian, for which the $\psi_0^{(0)}$ function would be an eigenfunction (then the perturbation would be small and the Rayleigh–Schrödinger perturbation theory might be applied), or to abandon any Rayleigh–Schrödinger perturbation scheme and replace it by

---

[23] In Appendix Y, p. 1050, we show, how the charge distribution changes when the Pauli exclusion principle is forced by a proper projection of the $\varphi^{(0)}$ wave function.

something else. The first of these possibilities was developed intensively in many laboratories. The approach had the deficiency that the operators appearing in the theories depended explicitly on the basis set used, and therefore there was no guarantee that a basis independent theory exists.

The second possibility relies on an iterative solution of the Schrödinger equation, forcing the proper symmetry of the intermediate functions. The method was proposed mainly by Bogumił Jeziorski and Włodzimierz Kołos.

Bloch equations

Claude Bloch was probably the first to write the Schrödinger equation in the form shown in formulae[24] (10.76) and (10.59). Let us recall them in a notation adapted to the present situation:

$$\begin{aligned} \psi_0 &= \varphi^{(0)} + \hat{R}_0\big(E_0^{(0)} - E_0 + V\big)\psi_0, \\ E_0 &= E_0^{(0)} + \langle \varphi^{(0)} | V\psi_0 \rangle, \end{aligned}$$

where we assume that $\varphi^{(0)}$ satisfies

$$\hat{H}^{(0)}\varphi^{(0)} = E_0^{(0)}\varphi^{(0)}$$

with the eigenvalues of the unperturbed Hamiltonian $\hat{H}^{(0)} = \hat{H}_A + \hat{H}_B$ given as the sum of the energies of the isolated molecules A and B:

$$E_0^{(0)} = E_{A,0} + E_{B,0},$$

and ψ_0 is the exact ground-state solution to the Schrödinger equation with the total non-relativistic Hamiltonian $\hat{H}$ of the system:

$$\hat{H}\psi_0 = E_0\psi_0.$$

We focus our attention on the difference $\mathcal{E}_0$ between E_0, which is our target and $E_0^{(0)}$, which is at our disposal as the unperturbed energy. We may write the Bloch equations in a form exposing the interaction energy $\mathcal{E}_0 = E_0 - E_0^{(0)}$

$$\begin{aligned} \psi_0 &= \varphi^{(0)} + \hat{R}_0(-\mathcal{E}_0 + V)\psi_0, \\ \mathcal{E}_0 &= \langle \varphi^{(0)} | V\psi_0 \rangle, \end{aligned}$$

the equations are valid for intermediate normalization $\langle \varphi^{(0)} | \psi_0 \rangle = 1$. This system of equations for $\mathcal{E}_0$ and ψ_0 might be solved by an iterative method:[25]

ITERATIVE SCHEME:

$$\psi_0(n) = \varphi^{(0)} + \hat{R}_0\big[-\mathcal{E}(n) + V\big]\psi_0(n-1), \tag{13.28}$$

$$\mathcal{E}_0(n) = \langle \varphi^{(0)} | V\psi_0(n-1) \rangle, \tag{13.29}$$

where the *iteration* number n is in the parentheses.

---

[24] C. Bloch, *Nucl. Phys*. 6 (1958) 329.

[25] In such a method we have freedom in choosing the starting point – this is one of its most beautiful features.

Polarization scheme replaced

We start in the zeroth iteration with $\psi_0(0) = \varphi^{(0)}$.

> When repeating the above iterative scheme and grouping the individual terms according to the powers of V, at each turn we obtain the exact expression appearing in the Rayleigh–Schrödinger polarization approximation (Chapter 5) plus some higher order terms.

It is worth noting that $\mathcal{E}_0(n)$ is the *sum* of corrections of the Rayleigh–Schrödinger *up to the* n-th order with respect to V (*not* the n-th perturbation correction). For large R, the quantity $\mathcal{E}_0(n)$ is an arbitrarily good approximation of the exact interaction energy.

Of course, the rate, at which the iterative procedure converges depends very much on the starting point chosen. From this point of view, the start from $\psi_0(0) = \varphi^{(0)}$ is particularly unfortunate, because the remaining (roughly) 50% of the *wave function* has to be restored by the hard work of the perturbational series (high-order corrections are needed). This will be especially pronounced for long intermolecular distances, where the exchange interaction energy will not be obtained in any finite order.

Murrell–Shaw and Musher–Amos (MS–MA) perturbation theory

A much more promising starting point in eq. (13.28) seems to be $\psi_0(0) = \psi_0^{(0)}$, because the symmetry of the wave function is already correct. For convenience the intermediate normalization is used (see p. 204) $\langle\varphi^{(0)}|\psi_0^{(0)}\rangle = 1$, i.e. $\psi_0^{(0)} = N\hat{\mathcal{A}}\varphi^{(0)}$ with $N = \langle\varphi^{(0)}|\hat{\mathcal{A}}\varphi^{(0)}\rangle^{-1}$. The first iteration of eqs. (13.28) and (13.29) gives the first-order correction to the energy

intermediate normalization

$$\mathcal{E}_0(1) = N\langle\varphi^{(0)}|V\hat{\mathcal{A}}\varphi^{(0)}\rangle = E^{(1)}_{\text{pol}} + E^{(1)}_{\text{exch}},$$
$$E^{(1)}_{\text{pol}} \equiv E_{\text{elst}} = \langle\varphi^{(0)}|V\varphi^{(0)}\rangle.$$

We have obtained the electrostatic energy already known plus a correction $E^{(1)}_{\text{exch}}$ which we will discuss in a minute.

The first-iteration wave function will be obtained in the following way. First, we will use the commutation relation $\hat{\mathcal{A}}\hat{H} = \hat{H}\hat{\mathcal{A}}$ or

$$\hat{\mathcal{A}}\big(\hat{H}^{(0)} + V\big) = \big(\hat{H}_0 + V\big)\hat{\mathcal{A}}. \tag{13.30}$$

Of course

$$\hat{\mathcal{A}}\big(\hat{H}^{(0)} - E_0^{(0)} + V\big) = \big(\hat{H}^{(0)} - E_0^{(0)} + V\big)\hat{\mathcal{A}}, \tag{13.31}$$

which gives[26] $V\hat{\mathcal{A}} - \hat{\mathcal{A}}V = [\hat{\mathcal{A}}, \hat{H}^{(0)} - E_0^{(0)}]$, as well as $(V - \mathcal{E}_1)\hat{\mathcal{A}} = \hat{\mathcal{A}}(V - \mathcal{E}_1) +$

---

[26] Let us stress *en passant* that the left-hand side is of the first order in V, while the right-hand side is of the zeroth order. Therefore, in symmetry adapted perturbation theory, the order is not a well defined quantity, its role is taken over by the iteration number.

$[\hat{\mathcal{A}}, \hat{H}^0 - E_0^{(0)}]$. Now we are ready to use formula (13.28) with $n = 1$:

$$\begin{aligned}\psi_0(1) &= \varphi^{(0)} + \hat{R}_0\big(V - \mathcal{E}_0(1)\big)\psi_0^{(0)} = \varphi^{(0)} + N\hat{R}_0\big(V - \mathcal{E}_0(1)\big)\hat{\mathcal{A}}\varphi^{(0)} \\ &= \varphi^{(0)} + N\hat{R}_0\big\{\hat{\mathcal{A}}\big(V - \mathcal{E}_0(1)\big) + \hat{\mathcal{A}}\big(\hat{H}^{(0)} - E_0^{(0)}\big) - \big(\hat{H}^{(0)} - E_0^{(0)}\big)\hat{\mathcal{A}}\big\}\varphi^{(0)} \\ &= \varphi^{(0)} + N\hat{R}_0\hat{\mathcal{A}}\big(V - \mathcal{E}_0(1)\big)\varphi^{(0)} + N\hat{R}_0\hat{\mathcal{A}}\big(\hat{H}^{(0)} - E_0^{(0)}\big)\varphi^{(0)} \\ &\quad - N\hat{R}_0\big(\hat{H}^{(0)} - E_0^{(0)}\big)\hat{\mathcal{A}}\varphi^{(0)}.\end{aligned}$$

The third term is equal to 0, because $\varphi^{(0)}$ is an eigenfunction of $\hat{H}^{(0)}$ with an eigenvalue $E_0^{(0)}$. The fourth term may be transformed by decomposing $\hat{\mathcal{A}}\varphi^{(0)}$ into the vector (in the Hilbert space) parallel to $\varphi^{(0)}$ or $\langle\hat{\mathcal{A}}\varphi^{(0)}|\varphi^{(0)}\rangle\varphi^{(0)}$ and the vector orthogonal to $\varphi^{(0)}$, or $(1 - |\varphi^{(0)}\rangle\langle\varphi^{(0)}|)\mathcal{A}\varphi^{(0)}$. The result of $\hat{R}_0(\hat{H}^{(0)} - E_0^{(0)})$ acting on the first vector is zero (p. 554), while the second vector gives $(1 - |\varphi^{(0)}\rangle\langle\varphi^{(0)}|)\mathcal{A}\varphi^{(0)}$. This gives as the first iteration ground-state wave function $\psi_0(1)$:

$$\begin{aligned}\psi_0(1) &= \varphi^{(0)} + N\hat{R}_0\hat{\mathcal{A}}\big(V - \mathcal{E}_0(1)\big)\varphi^{(0)} + N\hat{\mathcal{A}}\varphi^{(0)} - N\big\langle\varphi^{(0)}\big|\hat{\mathcal{A}}\varphi^{(0)}\big\rangle\varphi^{(0)} \\ &= \frac{\hat{\mathcal{A}}\varphi^{(0)}}{\langle\varphi^{(0)}|\hat{\mathcal{A}}\varphi^{(0)}\rangle} + N\hat{R}_0\hat{\mathcal{A}}\big(V - \mathcal{E}_0(1)\big)\varphi^{(0)} \\ &= \hat{\mathcal{B}}\varphi^{(0)} - N\hat{R}_0\hat{\mathcal{A}}\big(\mathcal{E}_0(1) - V\big)\varphi^{(0)},\end{aligned}$$

where

$$\hat{\mathcal{B}}\varphi^{(0)} = \frac{\hat{\mathcal{A}}\varphi^{(0)}}{\langle\varphi^{(0)}|\hat{\mathcal{A}}\varphi^{(0)}\rangle}. \tag{13.32}$$

After inserting $\psi_0(1)$ into the iterative scheme (13.29) with $n = 2$ we obtain the second-iteration energy

$$\begin{aligned}\mathcal{E}_0(2) &= \big\langle\varphi^{(0)}\big|V\psi_0(1)\big\rangle \\ &= \frac{\langle\varphi^{(0)}|V\hat{\mathcal{A}}\varphi^{(0)}\rangle}{\langle\varphi^{(0)}|\hat{\mathcal{A}}\varphi^{(0)}\rangle} - N\big\langle\varphi^{(0)}\big|V\hat{R}_0\hat{\mathcal{A}}\big[\mathcal{E}_0(1) - V\big]\varphi^{(0)}\big\rangle.\end{aligned} \tag{13.33}$$

These equations are identical to the corresponding corrections in perturbation theories derived by Murrell and Shaw[27] and by Musher and Amos[28] (MS–MA).

13.7.5 SYMMETRY FORCING

Finally, there is good news. It turns out that we may formulate a general iterative scheme which is able to produce various perturbation procedures, known and unknown in the literature. In addition the scheme has been designed by my nearest-neighbour colleagues (Jeziorski and Kołos). This scheme reads as:

---

[27] J.N. Murrell, G. Shaw, *J. Chem. Phys.* 46 (1967) 1768.

[28] J.I. Musher, A.T. Amos, *Phys. Rev.* 164 (1967) 31.

Table 13.4. Symmetry forcing in various perturbation schemes. The operator $\hat{\mathcal{B}}$ is defined by: $\hat{\mathcal{B}}\chi = \hat{\mathcal{A}}\chi / \langle \varphi^{(0)} | \hat{\mathcal{A}}\chi \rangle$

Perturbation scheme	$\psi^{(0)}$	$\hat{\mathcal{F}}$	$\hat{\mathcal{G}}$
polarization	$\varphi^{(0)}$	1	1
symmetrized polarization[a]	$\varphi^{(0)}$	1	$\hat{\mathcal{B}}$
MS–MA	$\hat{\mathcal{B}}\varphi^{(0)}$	1	1
Jeziorski-Kołos scheme[b]	$\hat{\mathcal{B}}\varphi^{(0)}$	$\hat{\mathcal{A}}$	1
EL–HAV[c]	$\hat{\mathcal{B}}\varphi^{(0)}$	$\hat{\mathcal{A}}$	$\hat{\mathcal{B}}$

[a] B. Jeziorski, K. Szalewicz, G. Chałasiński, *Int. J. Quantum Chem.* 14 (1978) 271; in the expression for the energy in the polarization perturbation theory all corrections to the wave function are first subject to the operator $\hat{\mathcal{B}}$.

[b] B. Jeziorski, W. Kołos, *Int. J. Quantum Chem.* 12 (1977) 91.

[c] Eisenschitz–London and Hirschfelder–van der Avoird perturbation theory: R. Eisenschitz, F. London, *Zeit. Phys.* 60 (1930) 491; J.O. Hirschfelder, *Chem. Phys. Letters* 1 (1967) 363; A. van der Avoird, *J. Chem. Phys.* 47 (1967) 3649.

$$\psi_0(n) = \varphi^{(0)} + \hat{R}_0\big[-\mathcal{E}_0(n) + V\big]\hat{\mathcal{F}}\psi_0(n-1),$$
$$\mathcal{E}_0(n) = \big\langle \varphi^{(0)} \big| V\hat{\mathcal{G}}\psi_0(n-1)\big\rangle$$

where in eqs. (13.28) and (13.29) we have inserted operators $\hat{\mathcal{F}}$ and $\hat{\mathcal{G}}$ which have to fulfil the obvious condition

$$\hat{\mathcal{F}}\psi_0 = \hat{\mathcal{G}}\psi_0 = \psi_0, \tag{13.34}$$

where ψ_0 is the solution to the Schrödinger equation.

WHY FORCE THE SYMMETRY?
At the end of the iterative scheme (convergence) the insertion of the operators $\hat{\mathcal{F}}$ and $\hat{\mathcal{G}}$ has no effect at all, but *before that* their presence may be crucial for the numerical convergence. This is the goal of symmetry forcing.

This method of generating perturbation theories has been called by the authors the *symmetry forcing* method in symmetry adapted perturbation theory (SAPT).

Polarization collapse removed

The corrections obtained in SAPT differ from those of the polarization perturbational method. The first-order energy correction is already different.

To show the relation between the results of the two approaches, let us first introduce some new quantities. The first is an idempotent antisymmetrizer

$$\hat{\mathcal{A}} = C\hat{\mathcal{A}}^A\hat{\mathcal{A}}^B(1+\hat{P}) \quad \text{with } C = \frac{N_A!N_B!}{(N_A+N_B)!},$$

where $\hat{\mathcal{A}}^A$, $\hat{\mathcal{A}}^B$ are idempotent antisymmetrizers for molecules A and B, each molecule contributing N_A and N_B electrons. Permutation operator $\hat{P}$ contains all the electron exchanges between molecules A and B:

$$\hat{P} = \hat{P}^{AB} + \hat{P}', \quad \hat{P}^{AB} = -\sum_{i \in A}\sum_{j \in B} \hat{P}_{ij},$$

with $\hat{P}^{AB}$ denoting the single exchanges only, and $\hat{P}'$ the rest of the permutations, i.e. the double, triple, etc. exchanges. Let us stress that $\varphi^{(0)} = \psi_{A,0}\psi_{B,0}$ represents a product of two antisymmetric functions[29] and therefore $\hat{\mathcal{A}}\varphi^{(0)} = C(1 + \hat{P}^{AB} + \hat{P}')\psi_{A,0}\psi_{B,0}$. Taking into account the operator $\hat{P}$ in $\langle\varphi^{(0)}|V\hat{\mathcal{A}}\varphi^{(0)}\rangle$ and $\langle\varphi^{(0)}|\hat{\mathcal{A}}\varphi^{(0)}\rangle$ produces (p. 715, $E^{(1)} \equiv \mathcal{E}_0(1)$):

$$E^{(1)} = \frac{\langle\psi_{A,0}\psi_{B,0}|V\psi_{A,0}\psi_{B,0}\rangle + \langle\psi_{A,0}\psi_{B,0}|V\hat{P}^{AB}\psi_{A,0}\psi_{B,0}\rangle + \mathrm{O}(S^4)}{1 + \langle\psi_{A,0}\psi_{B,0}|\hat{P}^{AB}\psi_{A,0}\psi_{B,0}\rangle + \mathrm{O}(S^4)}, \quad (13.35)$$

exchange interaction

where the integrals with $\hat{P}^{AB}$ are of the order[30] of S^2.

In the polarization approximation

$$E^{(1)}_{\mathrm{pol}} \equiv E_{\mathrm{elst}} = \langle\varphi^{(0)}|V\varphi^{(0)}\rangle \quad (13.36)$$

while in the symmetry adapted perturbation theory

$$E^{(1)} = \frac{\langle\varphi^{(0)}|V\hat{\mathcal{A}}\varphi^{(0)}\rangle}{\langle\varphi^{(0)}|\hat{\mathcal{A}}\varphi^{(0)}\rangle}, \quad (13.37)$$

$$E^{(1)} = E^{(1)}_{\mathrm{pol}} + E^{(1)}_{\mathrm{exch}}, \quad (13.38)$$

where *the exchange interaction* in first-order perturbation theory

$$\begin{aligned} E^{(1)}_{\mathrm{exch}} &= \langle\psi_{A,0}\psi_{B,0}|VP^{AB}\psi_{A,0}\psi_{B,0}\rangle \\ &\quad - \langle\psi_{A,0}\psi_{B,0}|V\psi_{A,0}\psi_{B,0}\rangle\langle\psi_{A,0}\psi_{B,0}|P^{AB}\psi_{A,0}\psi_{B,0}\rangle + \mathrm{O}(S^4). \quad (13.39) \end{aligned}$$

In the most commonly encountered interaction of closed shell molecules the $E^{(1)}_{\mathrm{exch}}$ term represents the *valence repulsion.*

valence repulsion

The symbol $\mathrm{O}(S^4)$ stands for all the terms that vanish with the fourth power of the overlap integrals or faster. The valence repulsion already appears (besides the

---

[29] The product itself does not have this symmetry.

[30] This means that we also take into account such a decay in other than overlap integrals S, e.g., $(1s_a 1s_b|1s_b 1s_a)$ is of the order S^2, where $S = (1s_a|1s_b)$. Thus the criterion is the differential overlap rather than the overlap integral.

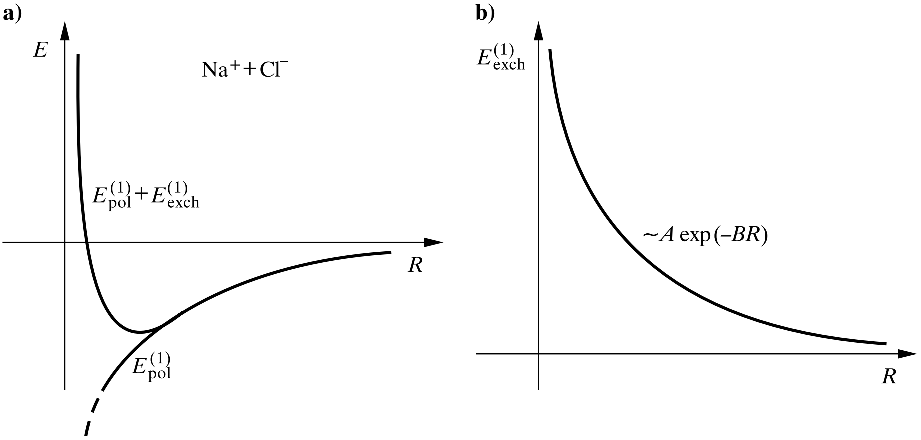


Fig. 13.9. Interaction energy of Na$^+$ and Cl$^-$. The polarization approximation gives an absurdity for small separations: the subsystems attract very strongly (mainly because of the electrostatic interaction), while they have had to repel very strongly. The absurdity is removed when the valence repulsion is taken into account (a). Fig. (b) shows the valence repulsion alone modelled by the term $A\exp(-BR)$, where A and B are positive constants.

electrostatic energy $E^{(1)}_{\text{pol}}$) in the first order of the perturbation theory as a result of the Pauli exclusion principle.[31]

We have gained a remarkable thing, which may be seen by taking the example of two interacting subsystems: Na$^+$ and Cl$^-$. In the polarization approximation the electrostatic, induction and dispersion contributions to the interaction energy are negative, the total energy will go down and we would soon have a catastrophe: both subsystems would occupy the same place in space and according to the energy calculated (Fig. 13.9) the system would be extremely happy (very low energy). This is absurd.

If this were true, we could not exist. Indeed, sitting safely on a chair we have an equilibrium of the gravitational force and..., well, and what? First of all, the force coming from valence repulsion. It is claimed sometimes that quantum effects are peculiar to small objects (electrons, nuclei, atoms, molecules) and are visible only when dealing with such particles. We see, however, that we owe even sitting on a chair to the Pauli exclusion principle (a quantum effect).

The valence repulsion removes the absurdity of the polarization approximation, which made the collapse of the two subsystems possible.

---

[31] An intriguing idea: the polarization approximation should be an extremely good approximation for the interaction of a molecule with an antimolecule (built from antimatter). Indeed, in the molecule we have electrons, in the antimolecule positrons and no antisymmetrization (between the systems) is needed. Therefore a product wave function should be a very good starting point. No valence repulsion will appear, the two molecules will penetrate like ghosts. Soon after, the tremendous lightning will be seen and the terrible thunder of annihilation will be heard. The system will disappear.

13.7.6 A LINK TO THE VARIATIONAL METHOD – THE HEITLER–LONDON INTERACTION ENERGY

Since the $\hat{\mathcal{A}}\varphi^{(0)}$ wave function is a good approximation of the exact ground state wave function at high values of R, we may calculate what is called the Heitler–London interaction energy ($E_{\text{int}}^{\text{HL}}$) as the mean value of the total (electronic) Hamiltonian minus the energies of the isolated subsystems

$$E_{\text{int}}^{\text{HL}} = \frac{\langle \hat{\mathcal{A}}\varphi^{(0)} | \hat{H} \hat{\mathcal{A}}\varphi^{(0)} \rangle}{\langle \hat{\mathcal{A}}\varphi^{(0)} | \hat{\mathcal{A}}\varphi^{(0)} \rangle} - (E_{A,0} + E_{B,0}).$$

This expression may be transformed in the following way

$$\begin{aligned}
E_{\text{int}}^{\text{HL}} &= \frac{\langle \varphi^{(0)} | \hat{H} \hat{\mathcal{A}}\varphi^{(0)} \rangle}{\langle \varphi^{(0)} | \hat{\mathcal{A}}\varphi^{(0)} \rangle} - (E_{A,0} + E_{B,0}) \\
&= \frac{\langle \varphi^{(0)} | \hat{H}^{(0)} \hat{\mathcal{A}}\varphi^{(0)} \rangle + \langle \varphi^{(0)} | V \hat{\mathcal{A}}\varphi^{(0)} \rangle}{\langle \varphi^{(0)} | \hat{\mathcal{A}}\varphi^{(0)} \rangle} - (E_{A,0} + E_{B,0}) \\
&= \frac{(E_{A,0} + E_{B,0})\langle \varphi^{(0)} | \hat{\mathcal{A}}\varphi^{(0)} \rangle + \langle \varphi^{(0)} | V \hat{\mathcal{A}}\varphi^{(0)} \rangle}{\langle \varphi^{(0)} | \hat{\mathcal{A}}\varphi^{(0)} \rangle} - (E_{A,0} + E_{B,0}) \\
&= \frac{\langle \varphi^{(0)} | V \hat{\mathcal{A}}\varphi^{(0)} \rangle}{\langle \varphi^{(0)} | \hat{\mathcal{A}}\varphi^{(0)} \rangle}.
\end{aligned}$$

Therefore, the Heitler–London interaction energy is equal to the first order SAPT energy

$$E_{\text{int}}^{\text{HL}} = E^{(1)}.$$

13.7.7 WHEN WE DO NOT HAVE AT OUR DISPOSAL THE IDEAL $\psi_{A,0}$ AND $\psi_{B,0}$

Up till now we have assumed that the ideal ground-state solutions of the Schrödinger equation for molecules A and B are at our disposal. In practice this will never happen. Instead of $\psi_{A,0}$ and $\psi_{B,0}$ we will have some approximate functions, $\tilde{\psi}_{A,0}$ and $\tilde{\psi}_{B,0}$, respectively. In such a case

$$E_{\text{int}}^{\text{HL}} \neq E^{(1)}.$$

Let us assume that $\tilde{\psi}_{A,0}$ and $\tilde{\psi}_{B,0}$, respectively, represent *Hartree–Fock solutions* for the subsystems A and B. Then the corresponding Heitler–London interaction energy equal to $\tilde{E}_{\text{int}}^{\text{HL}}$ may be written as

$$\tilde{E}_{\text{int}}^{\text{HL}} = \tilde{E}^{(1)} + \Delta_L + \Delta_M,$$

where $\tilde{E}^{(1)}$ is what the old formula gives in the new situation

$$\tilde{E}^{(1)} = \frac{\langle \tilde{\psi}_{A,0} \tilde{\psi}_{B,0} | V \hat{\mathcal{A}} \tilde{\psi}_{A,0} \tilde{\psi}_{B,0} \rangle}{\langle \tilde{\psi}_{A,0} \tilde{\psi}_{B,0} | \hat{\mathcal{A}} \tilde{\psi}_{A,0} \tilde{\psi}_{B,0} \rangle}$$

and Δ_L denotes a correction – called the *Landshoff delta*

Landshoff Δ

$$\Delta_L = \Delta_L^A + \Delta_L^B$$

with the *Landshoff's delta* for individual molecules[32]

$$\Delta_L^A = \frac{\langle \tilde{\psi}_{A,0}\tilde{\psi}_{B,0}|\hat{\mathcal{A}}(\hat{F}_A - \langle \hat{F}_A\rangle)\tilde{\psi}_{A,0}\tilde{\psi}_{B,0}\rangle}{\langle \tilde{\psi}_{A,0}\tilde{\psi}_{B,0}|\hat{\mathcal{A}}(\tilde{\psi}_{A,0}\tilde{\psi}_{B,0})\rangle}$$

and similar definition for Δ_L^B. The other correction – called the *Murrell delta*[33] – is defined as[34]

Murrell Δ

$$\Delta_M = \Delta_M^A + \Delta_M^B$$

with

$$\Delta_M^A = \frac{\langle \tilde{\psi}_{A,0}\tilde{\psi}_{B,0}|\hat{\mathcal{A}}(\hat{W}_A - \langle \hat{W}_A\rangle)\tilde{\psi}_{A,0}\tilde{\psi}_{B,0}\rangle}{\langle \tilde{\psi}_{A,0}\tilde{\psi}_{B,0}|\hat{\mathcal{A}}(\tilde{\psi}_{A,0}\tilde{\psi}_{B,0})\rangle}$$

where $\hat{F}_A$ and $\hat{F}_B$ are the sums of the Fock operators for molecules A and B, respectively, whereas $\hat{W}_A = \hat{H}_A - \hat{F}_A$ and $\hat{W}_B = \hat{H}_B - \hat{F}_B$ are the corresponding *fluctuation potentials* (see p. 558), i.e. $\hat{H}^{(0)} = \hat{F}_A + \hat{F}_B + \hat{W}_A + \hat{W}_B$. The symbols $\langle \hat{F}_A\rangle$ and $\langle \hat{W}_A\rangle$ denote the mean values of the corresponding operators calculated with the approximate wave functions: $\langle \hat{F}_A\rangle \equiv \langle \tilde{\psi}_{A,0}|\hat{F}_A\tilde{\psi}_{A,0}\rangle$ and $\langle \hat{W}_A\rangle \equiv \langle \tilde{\psi}_{A,0}|\hat{W}_A\tilde{\psi}_{A,0}\rangle$, and similarly for B.

fluctuation potential

13.8 CONVERGENCE PROBLEMS

In perturbation theories all calculated corrections are simply *added* together. This may lead to partial sums that do not converge. This pertains also to the symmetry adapted perturbation theories. Why? Let us see Table 13.4. One of the perturbational schemes given there, namely that called the symmetrized polarization approximation, is based on the calculation of the wave function exactly as in the polarization approximation scheme, but just before the calculation of the corrections to the energy, the polarization wave function is projected on the antisymmetrized space. This procedure is bound to have trouble. The system changes its charge distribution without paying any attention to the Pauli exclusion principle (thus allow-

---

[32] It has been shown that the Landshoff's deltas Δ_L^A and Δ_L^B vanish for the Hartree–Fock solutions for individual molecules A and B (R. Landshoff, *Zeit. Phys.* 102 (1936) 201). They vanish as well for the SCF solutions (i.e. for finite basis sets) for individual molecules calculated in the basis of *all atomic orbitals of the total system* (B. Jeziorski, M. Bulski, L. Piela, *Intern. J. Quantum Chem.* 10 (1976) 281; M. Gutowski, G. Chałasiński, J. van Duijneveldt-van de Rijdt, *Intern. J. Quantum Chem.* 26 (1984) 971).

[33] J.N. Murrell, A.J.C. Varandas, *Mol. Phys.* 30 (1975) 223.

[34] It has been shown (B. Jeziorski, M. Bulski, L. Piela, *Intern. J. Quantum Chem.* 10 (1976) 281) that Δ_M^A and Δ_M^B are of the order of O(S^4).

over-polarization

Pauli blockade

ing it to polarize itself in a non-physical way), while it turns out that it has to fulfil a principle[35] (the Pauli principle). This may be described as "*overpolarization*".

This became evident after a study called the *Pauli blockade*.[36] It was shown that, if the Pauli exclusion principle is not obeyed, the electrons of the subsystem A can flow, without any penalty and totally unphysically, to the low-energy orbitals of B. This may lead to occupation of that orbital by, e.g., four electrons, whereas the Pauli principle admits only a maximum of a double occupation.

Thus, any realistic deformation of the electron clouds has to take into account simultaneously the exchange interaction (valence repulsion), or the Pauli principle. Because of this, we have introduced what is called the *deformation–exchange interaction energy* as

deformation–exchange interaction energy

$$E_{\text{def–exch}} = E^{(2)} - (E_{\text{elst}} + E_{\text{disp}}). \tag{13.40}$$

Padé approximants may improve convergence

Any perturbational correction carries information. Summing up (this is the way we calculate the total effect) these corrections means a certain processing of the information. We may ask an amazing question: *is there any possibility of taking the same corrections and squeezing out more information*[37] *than just making the sum*?

In 1892 Henri Padé[38] wrote his doctoral dissertation in mathematics and presented some fascinating results.

For a power series

$$f(x) = \sum_{j=0}^{\infty} a_j x^j \tag{13.41}$$

we may define a *Padé approximant* $[L/M]$ as the ratio of two polynomials:

$$[L/M] = \frac{P_L(x)}{Q_M(x)} \tag{13.42}$$

where $P_L(x)$ is a polynomial of at most L-th degree, while $Q_M(x)$ is a polynomial of M-th degree. The coefficients of the polynomials P_L and Q_M will be determined by the following condition

$$f(x) - [L/M] = \text{terms of higher degree than } x^{L+M}. \tag{13.43}$$

In this way it will be guaranteed that for $x = 0$ the Padé approximant $[L, M]$ will have the derivatives up to the $(L + M)$-th degree identical with those of the original function $f(x)$. In other words,

---

[35] This is similar to letting all plants grow as they want and just after harvesting everything selecting the wheat alone.

[36] M. Gutowski, L. Piela, *Mol. Phys*. 64 (1988) 337.

[37] That is, a more accurate result.

[38] H. Padé, *Ann. Sci. Ecole Norm. Sup., Suppl. [3]* 9 (1892) 1.

the first $L+M$ terms of the Taylor expansion for a function $f(x)$ and for its Padé approximant are identical.

Since the nominator and denominator of the approximant can be harmlessly multiplied by any non-zero number, we may set, without losing anything, the following normalization condition

$$Q_M(0) = 1. \tag{13.44}$$

Let us assume also that $P_L(x)$ and $Q_M(x)$ do not have any common factor.

If we now write the polynomials as:

$$\begin{aligned} P_L(x) &= p_0 + p_1 x + p_2 x^2 + \cdots + p_L x^L, \\ Q_M(x) &= 1 + q_1 x + q_2 x^2 + \cdots + q_M x^M, \end{aligned}$$

then multiplying eq. (13.42) by Q_M and forcing the coefficients at the same powers of x being equal we obtain the following system of equations for the unknowns p_i and q_i (there are $L+M+1$ of them, the number of equations is the same):

$$\begin{aligned} &a_0 && = p_0, \\ &a_1 + a_0 q_1 && = p_1, \\ &a_2 + a_1 q_1 + a_0 q_2 && = p_2, \\ &a_L + a_{L-1} q_1 + \cdots + a_0 q_L && = p_L, \\ &a_{L+1} + a_L q_1 + \cdots + a_{L-M+1} q_M && = 0, \\ &\vdots && \\ &a_{L+M} + a_{L+M-1} q_1 + \cdots + a_L q_M && = 0. \end{aligned} \tag{13.45}$$

Note please, that the sum of the subscripts in either term is a constant {from the range $[0, L+M]$}, which is connected to the above mentioned equal powers of x.

Example 1 ***(Mathematical)*****.** The method is best illustrated in action. Let us take a function

$$f(x) = \frac{1}{\sqrt{1-x}}. \tag{13.46}$$

Suppose we have an inspiration to calculate $f(\frac{1}{2})$. We get of course $\sqrt{2} = 1.414213562\ldots$ Let us expand f in a Taylor series:

$$f(x) = 1 + \frac{1}{2}x + \frac{3}{8}x^2 + \frac{5}{16}x^3 + \frac{35}{128}x^4 + \cdots. \tag{13.47}$$

Therefore, $a_0 = 1$; $a_1 = \frac{1}{2}$; $a_2 = \frac{3}{8}$; $a_3 = \frac{5}{16}$; $a_4 = \frac{35}{128}$. Now let us forget that these coefficients came from the Taylor expansion of $f(x)$. Many other functions may have the same *beginning* of the Taylor series. Let us calculate some partial sums of the right-hand side of eq. (13.47):

$f(\frac{1}{2})$	sum up to the n-th term
$n = 1$	1.00000
$n = 2$	1.25000
$n = 3$	1.34375
$n = 4$	1.38281
$n = 5$	1.39990

We see that the Taylor series "works very hard", it succeeds but not without pain and effort.

Now let us check out how one of the simplest Padé approximants, namely, [1/1], performs the same job. By definition

$$\frac{(p_0 + p_1 x)}{(1 + q_1 x)}. \tag{13.48}$$

Solving (13.45) gives as the approximant:[39]

$$\frac{(1 - \frac{1}{4}x)}{(1 - \frac{3}{4}x)}. \tag{13.49}$$

Let us stress that information contained in the power series (13.41) has been limited to a_0, a_1, a_2 (all other coefficients have not been used). For $x = \frac{1}{2}$ the Padé approximant has the value

$$\frac{(1 - \frac{1}{4}\frac{1}{2})}{(1 - \frac{3}{4}\frac{1}{2})} = \frac{7}{5} = 1.4, \tag{13.50}$$

which is *more effective* than the painful efforts of the Taylor series that used a coefficients up to a_4 (this gave 1.39990). To be fair, we have to compare the Taylor series result that used only a_0, a_1, a_2 and this gives only 1.34375! Therefore, the approximant failed by 0.01, while the Taylor series failed by 0.07. The Padé approximant [2/2] has the form:

$$[2, 2] = \frac{(1 - \frac{3}{4}x + \frac{1}{16}x^2)}{(1 - \frac{5}{4}x + \frac{5}{16}x^2)}. \tag{13.51}$$

For $x = \frac{1}{2}$ its value is equal to $\frac{41}{29} = 1.414$, which means accuracy of 10^{-4}, while *without Padé approximants, but using the same information contained in the coefficients, we get accuracy two orders of magnitude worse.*

Our procedure did not have the information that the function expanded is $(1 - x)^{-\frac{1}{2}}$, for we gave the first five terms of the Taylor expansion only. Despite this, the procedure determined, with high accuracy, what will give higher terms of the expansion.

---

[39] Indeed, $L = M = 1$, and therefore the equations for the coefficients p and q are the following: $p_0 = 1$, $\frac{1}{2} + q_1 = p_1$, $\frac{3}{8} + \frac{1}{2}q_1 = 0$. This gives the solution: $p_0 = 1$, $q_1 = -\frac{3}{4}$, $p_1 = -\frac{1}{4}$.

Example 2 ***(Quantum mechanical).*** This is not the end of the story yet. The reader will see in a minute some things which will be even stranger. Perturbation theory also represents a power series (with respect to λ) with coefficients that are energy corrections. If perturbation is small, the corrections are small as well. In general the higher the perturbation order, the smaller the corrections. As a result, a partial sum of a few low-order corrections, usually gives sufficient accuracy. However, the higher the order the more difficult are the corrections to calculate. Therefore, we may ask if there is any possibility of obtaining good results and at a low price by using the Padé approximants. In Table 13.5 some results of a study by Jeziorski et al. are collected.[40]

For $R = 12.5$ a.u., we see that the approximants had a very difficult task to do. First of all they "recognized" the series limit, only at about $2L + 1 = 17$. Before that, they have been less effective than the original series. It has to be stressed, however, that they "recognized" it *extremely* well (see $2L + 1 = 21$). In contrast to this, the (traditional) partial sums ceased to improve when L increased. This means that either the partial sum series converges to a false limit or it converges to the correct limit, but does it extremely slowly. We see from the variational result (the error is calculated with respect to this) that the convergence *is* false. If the variational result had not been known, we would say that the series has already converged. However, the Padé approximants said: "*no, this is a false convergence*" and they were right.

For $R = 3.0$ a.u. (see Table 13.5) the original series represents a real tragedy. For this distance, the perturbation is too large and the perturbational series just evidently *diverges*. The greater our effort, the greater the error of our result. The error is equal to 13% for $2L+1=17$, then to 22% for $2L+1=19$ and attains 36% for $2L + 1 = 21$. Despite of these hopeless results, it turns out that the problem

Table 13.5. Convergence of the MS–MA (p. 715) perturbational series for the hydrogen atom in the field of a proton (state $2p\sigma_u$) for internuclear distance R (a.u.). The error (in %) is given for the sum of the original perturbational series and for the Padé $[L + 1, L]$ approximant, and is calculated with respect to the variational method (i.e. the best for the basis set used)

$2L+1$	$R=12.5$		$R=3.0$	
	pert. series	$[L+1,L]$	pert. series	$[L+1,L]$
3	0.287968	0.321460	0.265189	0.265736
5	0.080973	−0.303293	0.552202	−1.768582
7	0.012785	−0.003388	0.948070	0.184829
9	−0.000596	−0.004147	1.597343	0.003259
11	−0.003351	−0.004090	2.686945	0.002699
13	−0.003932	−0.004088	4.520280	0.000464
15	−0.004056	−0.004210	7.606607	0.000009
17	−0.004084	−0.001779	12.803908	0.000007
19	−0.004090	0.000337	21.558604	−0.000002
21	−0.004092	−0.000003	36.309897	0.000001

---

[40] B. Jeziorski, K. Szalewicz, M. Jaszuński, *Chem. Phys. Letters* 61 (1979) 391.

represent “peanuts” for the Padé approximants.[41] They were already much better for $L = 3$.

Why are the Padé approximants so effective?

The apparent garbage produced by the perturbational series represented for the Padé approximants precise information that the absurd perturbational corrections pertain the energy of the $2p\sigma_u$ state of the hydrogen atom in the electric field of the proton. How come? Low-order perturbational corrections, even if absolutely crazy, somehow carry information about the physics of the problem. The convergence properties of the Rayleigh–Schrödinger perturbation theory depend critically on the poles of the function approximated (see discussion on p. 210). A pole cannot be described by any power series (as happens in perturbation theories), whereas the Padé approximants have poles built in the very essence of their construction (the denominator as a polynomial). This is why they may fit so well with the nature of the problems under study.[42]

13.9 NON-ADDITIVITY OF INTERMOLECULAR INTERACTIONS

Interaction energy represents the non-additivity of the total energy

The *total* energy of interacting molecules is not an additive quantity, i.e. does not represent the sum of the energies of the isolated molecules. The reason for this non-additivity is the interaction energy.

Let see, whether the interaction energy itself has some additive properties. First of all the interaction energy requires the declaration of which fragments of the total system we treat as (interacting) molecules (see beginning of this chapter). The only real system is the total system, not these fragments. The fragments or subsystems can be chosen in many ways (Fig. 13.10).

If the theory is exact, the total system can be described at any such choice (cf. p. 492).

A theory has to be invariant with respect to any choice of subsystems in the system under consideration. Such a choice (however in many cases apparently evident) represents an arbitrary operation, similar to the choice of coordinate system.

Only the supermolecular theory is invariant with respect to such choices.[43] The perturbation theory so far has no such powerful feature (this problem is not even raised in the literature), because it requires the intra and intermolecular interac-

---

[41] Similar findings are reported in T.M. Perrine, R.K. Chaudhuri, K.F. Freed, *Intern. J. Quantum Chem.* 105 (2005) 18.

[42] There are cases, however, where Padé approximants may fail in a spectacular way.

[43] However for rather trivial reasons, i.e. interaction energy represents a by-product of the method. The main goal is the total energy, which by definition is independent of the choice of subsystems.

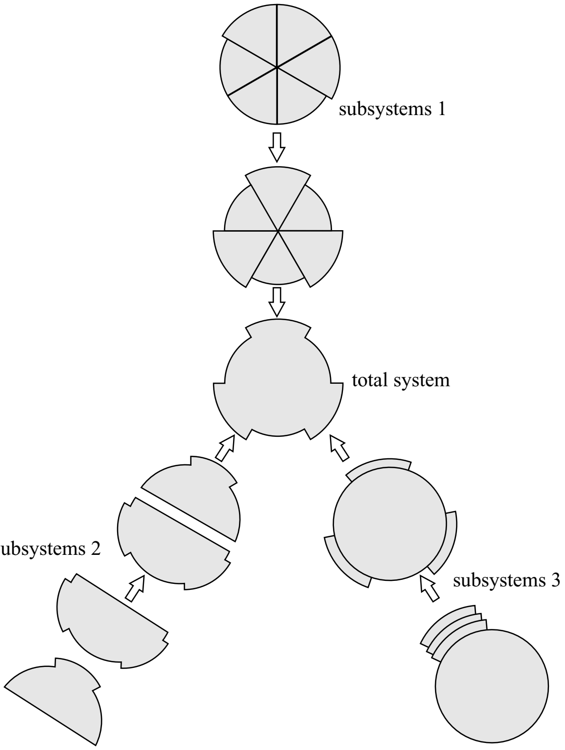


Fig. 13.10. Schematic illustration of arbitrariness behind the selection of subsystems within the total system. The total system under study is in the centre of the figure and can be divided into subsystems in many different ways. The isolated subsystems may differ from those incorporated in the total system (e.g., by shape). Of course, the sum of the energies of the isolated molecules depends on the choice made. The rest of the energy represents the interaction energy and depends on choice too. A correct theory has to be invariant with respect to these choices, which is an extreme condition to fulfil. The problem is even more complex. Using isolated subsystems does not tell us anything about the kind of complex they are going to make. We may imagine several stable aggregates (our system in the centre of the figure is only one of them). In this way we encounter the fundamental and so far unsolved problem of the most stable structure (cf. Chapter 7).

tions to be treated on the same footing. However this is extremely difficult in such a theory, because the assumption that the perturbation is small is inherent to perturbational theories.[44] Of course, choice of subsystems as with choice of coordinate systems, influences very strongly the mathematical difficulties and therefore the economy of the solution to be reached. Before performing calculations, a scientist already has some intuitive knowledge as to which choice is the most realistic. The intuition is applied when considering different ways in which our system may disintegrate and concentrating on those that require the least energy. The smaller the changes in the subsystems when going from isolated to bound, the smaller the interaction energy and the easier the application of the perturbational theory (cf. p. 685). The smaller the intermolecular distance(s) the more difficult and ambiguous the problem of subsystem choice becomes. In Chapter 9 probably the only example of the invariance of a quantum mechanical method is described.

13.9.1 MANY-BODY EXPANSION OF INTERACTION ENERGY

A next question could be: *is the interaction energy pair-wise additive*, i.e.

> is the interaction energy a sum of *pairwise* interactions?

---

[44] It has to be an infinite order perturbation theory with a large radius of convergence.

If this were true, it would be sufficient to calculate all possible interactions of pairs of molecules in the configuration identical to that of the total system[45] and our problem would be solved.

For the time being let us take the example of a stone, a shoe and a foot. The owner of the foot will certainly remember the three-body interaction, while nothing special happens when you put a stone into the shoe, or your foot into the shoe, or a small stone on your foot (two-body interactions). The molecules behave like this –

their interactions are not pairwise additive.

In the case of three interacting molecules, there is an effect of a strictly three-body character, which cannot be reduced to any two-body interactions. Similarly for larger numbers of molecules, there is a non-zero four-body effect, because all cannot be calculated as two- and three-body interactions, etc.

In what is called the many-body expansion for N molecules $A_1, A_2, \dots, A_N$ the interaction energy $E_{\text{int}}(A_1 A_2 \dots A_N)$, i.e. the difference between the total energy $E_{A_1 A_2 \dots A_N}$ and the sum of the energies of the isolated molecules $\sum_i E_{A_i}$ can be represented as a series of m-body terms $\Delta E(m,N)$, $m = 2, 3, \dots, N$:

$$E_{\text{int}} = E_{A_1 A_2 \dots A_N} - \sum_{i=1}^{N} E_{A_i} = \sum_{i>j}^{N} \Delta E_{A_i A_j}(2,N) + \sum_{i>j>k}^{N} \Delta E_{A_i A_j A_k}(3,N) + \dots + \Delta E_{A_1 A_2 \dots A_N}(N,N). \quad (13.52)$$

The $\Delta E(m,N)$ contribution to the interaction energy of N molecules ($m \leqslant N$) represents the sum of the interactions of m molecules (all possible combinations of m molecules among N molecules keeping their configurations fixed as in the total system) inexplicable by the interactions of $m' < m$ molecules.

One more question. Should we optimize the geometry, when calculating the individual many-body terms? In principle, we should not do this, because we are interested in the interaction energy at a given configuration of the nuclei. However, we may present the opposite point of view. For instance, we may be interested in how the geometry of the AB complex changes in the presence of molecule C. This is also a three-body interaction. These dilemmas have not yet been solved in the literature.

Example 3. ***Four molecules.*** The many-body expansion concept is easiest to understand by taking an example. Suppose we have four (point-like, for the sake of simplicity) molecules: A, B, C and D lying on a straight line. Their distances (in arbitrary units) are equal to the number of "stars": A*B***C**D. Let us assume

---

[45] This would be much less expensive than the calculation for the total system.

Table 13.6.

Three molecules	Interaction energy	Pairwise interactions	Difference
A*B***C	−8	−10	+2
A*B*****D	−5	−6	+1
A****C**D	−7	−8	+1
B***C**D	−9	−10	+1

that the total energy calculated for this configuration equals to −3000 kcal/mol, whereas the sum of the energies of the isolated molecules is −2990 kcal/mol. Hence, the interaction energy of the four molecules is −10 kcal/mol. The negative sign means that the interaction corresponds to attraction, i.e. the system is stable (as far as the binding energy is concerned) with respect to dissociation on $A + B + C + D$. Now we want to analyze the many-body decomposition of this interaction energy. First, we calculate the two-body contribution, let us take all possible pairs of molecules and calculate the corresponding interaction energies (the results are in parentheses, kcal/mol): $A^{*}B$ (−4), $A^{****}C$ (−2), $A^{******}D$ (−1), $B^{***}C$ (−4), $B^{*****}D$ (−1), $C^{**}D$ (−5). As we can see, the sum of all the pairwise interaction energies is $\Delta E(2,4) = -17$ kcal/mol. We did not obtain −10 kcal/mol, because *the interactions are not pairwise additive*. Now let us turn to the three-body contribution $\Delta E(3,4)$. To calculate this we consider all possible three-molecule systems in a configuration identical to that in the total system: $A^{*}B^{***}C$, $A^{*}B^{*****}D$, $A^{****}C^{**}D$, $B^{***}C^{**}D$, and calculate, in each case, the interaction energy of three molecules minus the interaction energies of all pairwise interactions involved. In Table 13.6 we list all the three-body systems possible and in each case give three numbers (in kcal/mol): the interaction energy of the three bodies (with respect to the isolated molecules), the sum of the pairwise interactions and the difference of these two numbers, i.e. the contribution of these three molecules to $\Delta E(3,4)$.

Hence, the three-body contribution to the interaction energy $\Delta E(3,4) = 2 + 1 + 1 + 1 = +5$ kcal/mol. The last step in the example is to calculate the four-body contribution. This can be done by subtracting from the interaction energy (−10) the two-body contribution (−17) and the three-body contribution (+5). We obtain $\Delta E(4,4) = -10 + 17 - 5 = 2$ kcal/mol.

We may conclude that in our (fictitious) example, at the given configuration, the many-body expansion of the interaction energy $E_{\text{int}} = -10$ kcal/mol represents a series decaying rather quickly: $\Delta E(2,4) = -17$ kcal/mol for the two-body, $\Delta E(3,4) = +5$ for the three-body and $\Delta E(4,4) = +2$ for the four-body interactions.

Are non-additivities large?

Already a vast experience has been accumulated and some generalizations are possible.[46] The many-body expansion usually converges faster than in our fictitious example.[47] For three argon atoms in an equilibrium configuration, the three-body

---

[46] V. Lotrich, K. Szalewicz, *Phys. Rev. Letters* 79 (1997) 1301.

[47] In quantum chemistry this almost always means a numerical convergence, i.e. a fast decay of individual contributions.

term is of the order of 1%. It should be noted, however, that in the argon *crystal* there is a lot of three-body interactions and the three-body effect increases to about 7%. On the other hand, for liquid water the three-body effect is of the order of 20%, and the higher contributions are about 5%. Three-body effects are sometimes able to determine the crystal structure and have significant influence on the physical properties of the system close to a phase transition ("critical region").[48]

In the case of the interaction of metal atoms, the non-additivity is much larger than that for the noble gases, and the three-body effects may attain a few tens of percent. This is important information since the force fields widely used in molecular mechanics (see p. 284) are based almost exclusively on effective pairwise interactions (neglecting the three- and more-body contributions).[49]

> Although the intermolecular interactions are non-additive, we may ask whether individual contributions to the interaction energy (electrostatic, induction, dispersion, valence repulsion) are additive?

Let us begin from the electrostatic interaction.

13.9.2 ADDITIVITY OF THE ELECTROSTATIC INTERACTION

Suppose we have three molecules A, B, C, intermolecular distances are long and therefore it is possible to use the polarization perturbation theory, in a very similar way to that presented in the case of two molecules (p. 692). In this approach, the unperturbed Hamiltonian $\hat{H}^{(0)}$ represents the sum of the Hamiltonians for the isolated molecules A, B, C. Let us change the abbreviations a little bit to be more concise for the case of three molecules. A *product* function $\psi_{A,n_A}\psi_{B,n_B}\psi_{C,n_C}$ will be denoted by $|n_A n_B n_C\rangle = |n_A\rangle|n_B\rangle|n_C\rangle$, where n_A, n_B, n_C $(= 0, 1, 2, \ldots)$ stand for the quantum numbers corresponding to the orthonormal wave functions for the molecules A, B, C, respectively. The functions $|n_A n_B n_C\rangle = |n_A\rangle|n_B\rangle|n_C\rangle$ are the eigenfunctions of $\hat{H}^{(0)}$:

$$\hat{H}^{(0)}|n_A n_B n_C\rangle = \big[E_A(n_A) + E_B(n_B) + E_C(n_C)\big]|n_A n_B n_C\rangle.$$

The perturbation is equal to $\hat{H} - \hat{H}^{(0)} = V = V_{AB} + V_{BC} + V_{AC}$, where the operators V_{XY} contain all the Coulomb interaction operators involving the nuclei and electrons of molecule X and those of molecule Y.

Let us recall that the electrostatic interaction energy $E_{\text{elst}}(ABC)$ of the ground-state ($n_A = 0$, $n_B = 0$, $n_C = 0$) molecules is defined as the first-order correction to the energy in the polarization approximation perturbation theory[50]

---

[48] R. Bukowski, K. Szalewicz, *J. Chem. Phys.* 114 (2001) 9518.

[49] That is, the effectivity of a force field relies on such a choice of interaction parameters, that the experimental data are reproduced (in such a way the parameters implicitly contain part of the higher-order terms).

[50] The $E_{\text{elst}}(ABC)$ term in symmetry-adapted perturbation theory represents only part of the first-order correction to the energy (the rest being the valence repulsion).

$$E^{(1)}_{\text{pol}} \equiv E_{\text{elst}}(ABC) = \langle 0_A 0_B 0_C | V | 0_A 0_B 0_C \rangle = \langle 0_A 0_B 0_C | V_{AB} + V_{BC} + V_{AC} | 0_A 0_B 0_C \rangle,$$

where the quantum numbers 000 have been supplied (maybe because of my excessive caution) by the redundant and self-explanatory indices $(0_A, 0_B, 0_C)$.

The integration in the last formula goes over the coordinates of all electrons. In the polarization approximation, the electrons can be unambiguously divided into three groups: those belonging to A, B and C. Because the zero-order wave function $|0_A 0_B 0_C\rangle$ represents a product $|0_A\rangle|0_B\rangle|0_C\rangle$, the integration over the electron coordinates of one molecule can be easily performed and yields

$$E_{\text{elst}} = \langle 0_A 0_B | V_{AB} | 0_A 0_B \rangle + \langle 0_B 0_C | V_{BC} | 0_B 0_C \rangle + \langle 0_A 0_C | V_{AC} | 0_A 0_C \rangle,$$

where, in the first term, the integration was performed over the electrons of C, in the second over the electrons of B, and in the third over those of C.

Now, let us look at the last formula. We easily see that the individual terms simply represent the electrostatic interaction energies of pairs of molecules: AB, BC and AC, that we would obtain in the perturbational theory (within the polarization approximation) for the interaction of AB, BC and AC, respectively. Conclusion:

the electrostatic interaction is pairwise additive.

13.9.3 EXCHANGE NON-ADDITIVITY

What about the exchange contribution? This contribution *does not exist in the polarization approximation*. It appears only in symmetry-adapted perturbation theory, in pure form in the first-order energy correction and coupled to other effects in higher order energy corrections.[51] The exchange interaction is difficult to interpret, because it appears as a result of the antisymmetry of the wave function (Pauli exclusion principle). The antisymmetry is forced by one of the postulates of quantum mechanics (see Chapter 1) and its immediate consequence is that the probability density of finding two electrons with the same spin and space coordinates is equal to zero.

A CONSEQUENCE OF THE PAULI EXCLUSION PRINCIPLE

In an atom or molecule, the Pauli exclusion principle results in a shell-like electronic structure (electrons with the same spin coordinates hate each other and try to occupy different regions in space). The valence repulsion may be seen as the same effect manifesting itself in the intermolecular interaction. Any attempt to make the molecular charge distributions overlap or occupy the same space ("pushing") leads to a violent increase in the energy.

---

[51] Such terms are bound to appear. For example, the induction effect is connected to deformation of the electron density distribution. The interaction (electrostatic, exchange, dispersive, etc.) of such a deformed object will change with respect to that of the isolated object. The coupling terms take care of this change.

PAULI DEFORMATION
The Pauli exclusion principle leads to a *deformation* of the wave functions describing the two molecules (by projecting the product-like wave function by the antisymmetrizer $\hat{A}$) with respect to the product-like wave function. The Pauli deformation (cf. Appendix Y) appears in the zeroth order of perturbation theory, whereas in the polarization approximation, the deformation of the wave function appears in the first order and is not related to the Pauli exclusion principle.

The antisymmetrizer pertains to the permutation symmetry of the wave function with respect to the coordinates of all electrons and therefore is different for a pair of molecules and for a system of three molecules.

The expression for the three-body non-additivity of the valence repulsion[52] [given by formula (13.39), based on definition (13.37) of the first-order correction in symmetry-adapted perturbation theory[53] and from definition (13.52) of the three-body contribution] is:

$$\begin{aligned} E^{(1)}_{\text{exch},ABC} &= N_{ABC}\langle 0_A 0_B 0_C | V_{AB} + V_{BC} + V_{AC} | \hat{A}_{ABC}(0_A 0_B 0_C)\rangle \\ &\quad - \sum_{(XY)=(AB),(AC),(BC)} N_{XY}\langle 0_X 0_Y | V_{XY} | \hat{A}_{XY}(0_X 0_Y)\rangle, \qquad (13.53) \end{aligned}$$

where $N_{ABC}\hat{A}_{ABC}|0_A 0_B 0_C\rangle$ and $N_{AB}\hat{A}_{AB}|0_A 0_B\rangle$, and so forth represent the normalized (N_{ABC} etc. are the normalization coefficients) antisymmetrized product-like wave function of the systems ABC, AB, etc. The antisymmetrizer $\hat{A}_{ABC}$ pertains to subsystems A, B, C, similarly $\hat{A}_{AB}$ pertains to A and B, etc., all antisymmetrizers containing only the intersystem electron exchanges and the summation goes over all pairs of molecules.

There is no chance of proving that the exchange interaction is additive, i.e. that eq. (13.53) gives 0. Let us consider the simplest possible example: each molecule has only a single electron: $|0_A(1)0_B(2)0_C(3)\rangle$. The operator $\hat{A}_{ABC}$ (see p. 986) makes (besides other terms) the following permutation:

$$\hat{A}_{ABC}|0_A(1)0_B(2)0_C(3)\rangle = \cdots - \frac{1}{(N_A+N_B+N_C)!}|0_A(3)0_B(2)0_C(1)\rangle + \cdots,$$

which according to eq. (13.53) leads to the integral

$$\begin{aligned} &-\frac{1}{(N_A+N_B+N_C)!}N_{ABC}\langle 0_A(1)0_B(2)0_C(3)|\frac{1}{r_{12}}|0_A(3)0_B(2)0_C(1)\rangle \\ &= -\frac{1}{(N_A+N_B+N_C)!}N_{ABC}\langle 0_A(1)0_B(2)|\frac{1}{r_{12}}|0_B(2)0_C(1)\rangle\langle 0_C(3)|0_A(3)\rangle \end{aligned}$$

---

[52] B. Jeziorski, M. Bulski, L. Piela, *Intern. J. Quantum Chem*. 10 (1976) 281.

[53] Because, as we have already proved, the rest, i.e. the electrostatic energy, is an additive quantity.

involving the wave functions centred on A, B and C. This means that the term belongs to the *three-body* effect.

The permutation operators of which the $\hat{A}_{ABC}$ operator is composed, correspond to the identity permutation[54] as well as to the exchange of one or more electrons between the interacting subsystems: $\hat{A}_{ABC} = 1+$ single exchanges $+$ double exchanges $+\cdots$.

It is easy to demonstrate,[55] that

> *the larger the number of electrons exchanged, the less important such exchanges are*, because the resulting contributions would be proportional to higher and higher powers of the (as a rule small) overlap integrals (S).

Single Exchange (SE) Mechanism

The smallest non-zero number of electron exchanges in $\hat{A}_{ABC}$ is equal to 1 (two electrons involved). Such an exchange may only take place between two molecules, say, AB.[56] This results in terms of the order of S^2 in the three-body expression. The third molecule *does not participate in the electron exchanges, but is not just a spectator in the interaction* (Fig. 13.11.a,b,c). If it were, the interaction would not be three-body.

> SE MECHANISM
> Molecule C interacts electrostatically with the *Pauli deformation of molecules A and B (i.e. with the multipoles that represent the deformation)*. Such a mixed interaction is called the SE mechanism.

It would be nice to have a simple formula which could replace the tedious calculations involving the above equations. The three-body energy may be approximated[57] by the product of the exponential term $b\exp(-aR_{AB})$ and the electric field produced by C, calculated, e.g., in the middle of the distance R_{AB} between molecules A and B. The goal of the exponential terms is to grasp the idea that the overlap integrals (and their squares) vanish exponentially with distance. The exponent a should depend on molecules A and B as well as on their mutual orientation and reflects the hardness of both molecules. These kind of model formulae have low scientific value but are of practical use.

---

[54] The operator reproduces the polarization approximation expressions in SAPT.

[55]
- First, we write down the exact expression for the first-order exchange non-additivity.
- Then, we expand the expression in the Taylor series with respect to those terms that arise from all electron exchanges except the identity permutation.
- Next, we see that the exchange non-additivity expression contains terms of the order of S^2 and higher, where S stands for the overlap integrals between the orbitals of the interacting molecules. S decays very fast (exponentially), when intermolecular distance increases.

[56] After that we have to consider AC and BC.

[57] Three-body effects are difficult to calculate. Researchers would like to understand the main mechanism and then capture it by designing a simple approximate formula ready to use in complex situations.

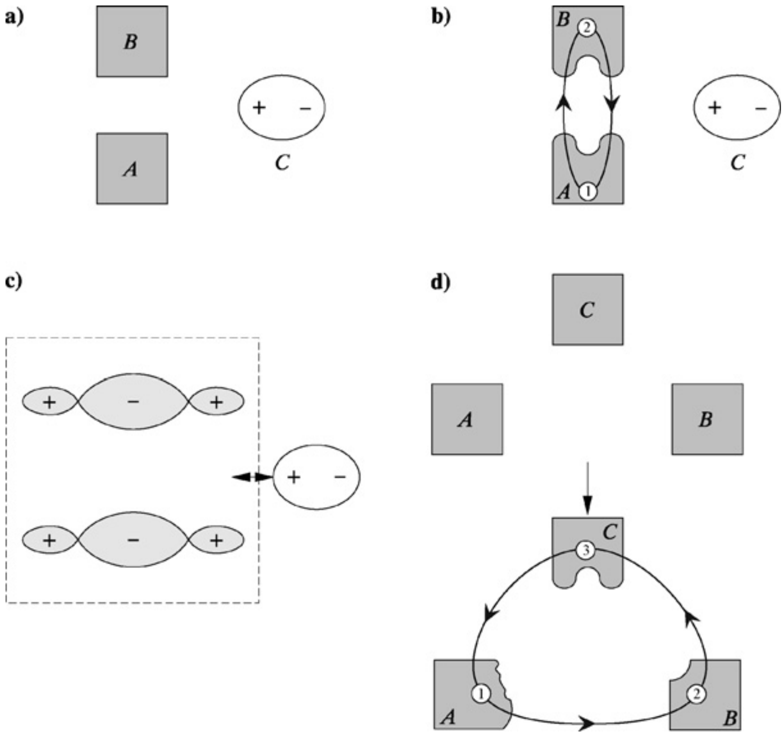


Fig. 13.11. A scheme of the SE and TE exchange non-additivities. Figs. (a), (b), (c) show the single exchange mechanism (SE). (a) Three non-interacting molecules (schematic representation of electron densities). (b) Pauli deformation of molecules A and B. (c) Electrostatic interaction of the Pauli deformation (resulting from exchange of electrons 1 and 2 between A and B) with the dipole moment of C. (d) The TE mechanism: molecules A and B exchange an electron with the mediation of molecule C.

When the double electron exchanges are switched on, we would obtain part of the three-body effect of the order of S^4. Since S is usually of the order of 10^{-2}, this contribution is expected to be small, although caution is advised, because the *number* of such terms is much more important.

Triple Exchange (TE) Mechanism

Is there any contribution of the order of S^3? Yes. The antisymmetrizer $\hat{A}_{ABC}$ is able to make the single electron exchange between, e.g., A and B, but by mediation of C. The situation is schematically depicted in Fig. 13.11.d.

TE MECHANISM

This effect is sometimes modelled as a product of three exponential functions: const $\exp(-a_{AB}R_{AB})\exp(-a_{BC}R_{BC})\exp(-a_{AC}R_{AC})$ and is misleadingly called a triple electron exchange. A molecule is involved in a single exchange with another molecule by mediation of a third.

Let us imagine that molecule B is very long and the configuration corresponds to: A B C. When C is far from A, the three-body effect is extremely small, because almost everything in the interaction is of the two-body character, Appendix Y. If molecule C approaches A and has some non-zero low-order multipoles, e.g., a charge, then it may interact by the SE mechanism even from a far. The TE mechanism operates only at short intermolecular distances.

The exchange interaction is non-additive, but the effects pertain to the contact region of both molecules.[58] The Pauli exclusion principle does not have any finite range in space, i.e. after being introduced it has serious implications for the wave function even at infinite intermolecular distance (cf. p. 712). Despite this, it always leads to the differential overlap of atomic orbitals (as in overlap or exchange integrals), which decays exponentially with increasing intermolecular distance (the SE mechanism has a partly long-range character).

13.9.4 INDUCTION ENERGY NON-ADDITIVITY

> The non-additivity of the intermolecular interaction results mainly from the non-additivity of the induction contribution.

How do we convince ourselves about the non-additivity? This is very easy. It will be sufficient to write the expression for the induction energy for the case of three molecules and to see whether it gives the sum of the pairwise induction interactions. Before we do this, let us write the formula for the total second order energy correction (similar to the case of two molecules on p. 694):

$$E^{(2)}(ABC) = {\sum_{n_A,n_B,n_C}}' \frac{|\langle n_A n_B n_C|V|0_A 0_B 0_C\rangle|^2}{[E_A(0_A)-E_A(n_A)]+[E_B(0_B)-E_B(n_B)]+[E_C(0_C)-E_C(n_C)]}. \tag{13.54}$$

According to perturbation theory, the term with all the indices equal to zero has to be omitted in the above expression. It is much better like this, because otherwise the denominator would "explode". The terms with all non-zero indices are equal to zero. Indeed, let us recall that V is the sum of the Coulomb potentials corresponding to all three *pairs* of the three molecules. This is the reason why it is easy to perform the integration over the electron coordinates of the third molecule (not involved in the pair). A similar operation was performed for the electrostatic interaction. This time, however, the integration makes the term equal to zero, because of the orthogonality of the ground and excited states of the third molecule. All this leads to the conclusion that to have a non-zero term in the summation, among the three indices, *one or two of them have to be of zero value*. Let us perform the summation in two stages: all the terms with only-two-zeros (or a single

---

[58] See Appendix Y.

non-zero) indices will make a contribution to $E_{\text{ind}}(ABC)$, while all the terms with only-one-zero (or, two non-zero) indices will sum to $E_{\text{disp}}(ABC)$:

$$E^{(2)}(ABC) = E_{\text{ind}}(ABC) + E_{\text{disp}}(ABC), \tag{13.55}$$

where the first term represents the *induction energy*:

$$E_{\text{ind}}(ABC) = E_{\text{ind}}(AB \to C) + E_{\text{ind}}(AC \to B) + E_{\text{ind}}(BC \to A),$$

where

$$E_{\text{ind}}(BC \to A) \equiv \sum_{n_A \neq 0} \frac{|\langle n_A 0_B 0_C | V | 0_A 0_B 0_C \rangle|^2}{[E_A(0_A) - E_A(n_A)]}$$

means that the "frozen" molecules B and C *acting together* polarize molecule A, etc. The second term in (13.55) represents the *dispersion energy* (this will be considered later on, see p. 740).

For the time being let us consider the induction energy $E_{\text{ind}}(ABC)$. Writing V as the sum of the Coulomb interactions of the pairs of molecules we have

$$\begin{aligned}
&E_{\text{ind}}(BC \to A) \\
&\quad = \sum_{n_A \neq 0} \langle n_A 0_B 0_C | V_{AB} + V_{BC} + V_{AC} | 0_A 0_B 0_C \rangle \langle 0_A 0_B 0_C | V_{AB} + V_{BC} + V_{AC} | n_A 0_B 0_C \rangle \\
&\qquad \times \left[E_A(0_A) - E_A(n_A)\right]^{-1} \\
&\quad = \sum_{n_A \neq 0} \left[\langle n_A 0_B | V_{AB} | 0_A 0_B \rangle + \langle n_A 0_C | V_{AC} | 0_A 0_C \rangle\right] \\
&\qquad \times \left[\langle 0_A 0_B | V_{AB} | n_A 0_B \rangle + \langle 0_A 0_C | V_{AC} | n_A 0_C \rangle\right] \\
&\qquad \times \left[E_A(0_A) - E_A(n_A)\right]^{-1}.
\end{aligned}$$

Look at the product in the nominator. The induction non-additivity arises just because of this product. If the product (being the square of the absolute value of $\langle n_A 0_B | V_{AB} | 0_A 0_B \rangle + \langle n_A 0_C | V_{AC} | 0_A 0_C \rangle$) were equal to the square of the absolute values of the first and second component, the total expression shown explicitly would be equal to the induction energy corresponding to the polarization of A by the frozen charge distribution of B plus a similar term corresponding to the polarization of A by C, i.e. the polarization occurring *separately*. Together with the other terms in $E_{\text{ind}}(AB \to C) + E_{\text{ind}}(AC \to B)$ *we would obtain the additivity* of the induction energy $E_{\text{ind}}(ABC)$. However, besides the sum of squares we also have the mixed terms. They will produce the non-additivity of the induction energy:

$$\boxed{E_{\text{ind}}(ABC) = E_{\text{ind}}(AB) + E_{\text{ind}}(BC) + E_{\text{ind}}(AC) + \Delta_{\text{ind}}(ABC). \quad (13.56)}$$

Thus, we obtain the following expression for the *induction non-additivity* $\Delta_{\text{ind}}(ABC)$:

$$\Delta_{\text{ind}}(ABC) = 2\,\text{Re} \sum_{n_A \neq 0} \frac{\langle n_A 0_B | V_{AB} | 0_A 0_B \rangle \langle n_A 0_C | V_{AC} | 0_A 0_C \rangle}{[E_A(0_A) - E_A(n_A)]} + \cdots, \tag{13.57}$$

where "$+\cdots$" stands for the non-additivities of $E_{\text{ind}}(AB \to C) + E_{\text{ind}}(AC \to B)$.

Example 4. ***Induction non-additivity.*** To show that the induction interaction of two molecules depends on the presence of the third molecule let us consider the system shown in Fig. 13.12.

Let molecule B be placed half-way between A^+ and C^+, thus the configuration of the system is: A^+.......B.........C^+ with long distances between the subsystems. In such a situation, the total interaction energy is practically represented by the induction contribution plus the constant electrostatic repulsion of A^+ and C^+. Is the three-body term (induction non-additivity) large? We will show in a minute that this term is large and positive (destabilizing). Since the electric field intensities nearly cancel within molecule B, then despite the high polarizability of the latter, the induction energy will be small. On the contrary, the opposite is true when considering two-body interaction energies. Indeed, A^+ polarizes B very strongly, C^+ does the same, resulting in high stabilization due to high two-body induction energy. Since the total effect is nearly zero, the induction non-additivity is bound to be a large positive number.[59]

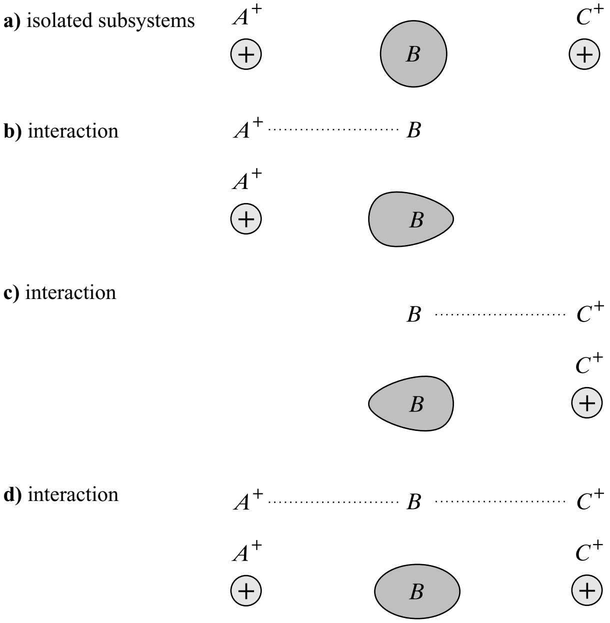


Fig. 13.12. The induction interaction may produce a large non-additivity. (a) Two distant non-polarizable cations: A^+, C^+ and a small, polarizable neutral molecule B placed exactly in the middle between AC. (b) The two-body induction interaction A^+B, a strong polarization. (c) The two-body induction interaction BC^+, a strong polarization. (d) The two cations polarize molecule B. Their electric field vectors cancel each other in the middle of B and give a small electric field intensity within B (a weak polarization).

---

[59] If the intermolecular distances were small, B were not in the middle of AC or molecule B were of large spatial dimension, the strength of our conclusion would diminish.

Self-consistency and polarization catastrophe

The second-order induction effects pertain to polarization by the charge distributions corresponding to the isolated molecules. However, the induced multipoles introduce a change in the electric field and in this way contribute to further changes in charge distribution. These effects already belong to the third[60] and higher orders of perturbation theory.

It is therefore evident that a two-body interaction model cannot manage the induction interaction energy. This is because we have to ensure that any subsystem, e.g., A, should experience polarization in an electric field, which is the *vector sum* of the electric fields from all its partner subsystems (B, C, ...) calculated at the position of A. The calculated induced dipole moment of A (we focus on the lowest multipole) creates the electric field that produces some changes in the dipole moments of B, C, ..., which in turn change the electric field acting on all the molecules, including A. The circle closes and the polarization procedure has to be performed till self-consistency is reached. This can often be done, although such a simplified interaction model does not allow for geometry optimization, which may lead to a *polarization catastrophe* ending up with induction energy equal to $-\infty$ (due to excessive approach and lack of the Pauli blockade described on p. 722).

polarization catastrophe

Three-body polarization amplifier

After recognizing that self-consistency may be achieved safely within a variational method (with the Pauli exclusion principle satisfied), this may end the story. However, it would be instructive to get a feeling for the polarization machinery. Let us pose a simple question: is it possible that the polarization of molecule B by molecule A, both separated by distance R, is amplified by the presence of molecule C which is unable by itself to polarize B? An interesting problem. If it appears that the mediator C might increase the polarization of A or B, C would play the role of an amplifier based on three-body induction non-additivity.

Suppose A is represented by a non-polarizable cation A^+, molecule B separated from A by R, is medium-polarizable, and a strongly polarizable molecule C (to-be-amplifier) enters between A^+ and B: $A^+ \ldots C \ldots B$, Fig. 13.13.a. How do we measure the polarization of B due to the presence of C? We might propose the electric dipole moment induced on B (of course, there will be nothing induced on A^+). Let R_{BC} be the BC separation. A simple calculation of the electric field intensities gives the following result[61] for the ratio of the induced dipole moments on B with and without the presence of C:

---

[60] Each of the induced multipoles is proportional to V, their interaction introduces another V; altogether this gives a term proportional to VVV, i.e. indeed of the third-order.

[61] Let us denote the distance AB by R. The axis x is directed from A to B, C is between A and B. If C possessed dipole moment μ_C, then the unit positive charge $+1$ on B would feel the potential (the charge–dipole term from the multipole expansion): $\frac{\mu_C}{R_{BC}^2}$. The corresponding electric field

$$\mathcal{E}_{C\to B} = -\frac{\partial}{\partial R_{BC}}\left(\frac{\mu_C}{R_{BC}^2}\right) = 2\frac{\mu_C}{R_{BC}^3}.$$

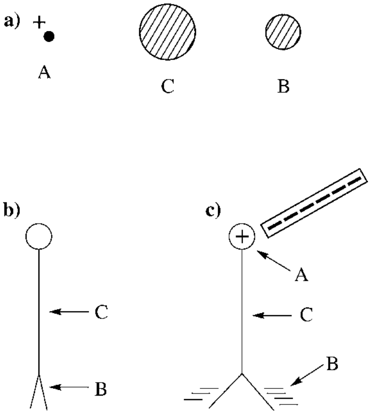


Fig. 13.13. A three-body polarization amplifier. (a) A represents a non-polarizable cation, B and C are polarizable molecules (C is the amplifier). (b) An electroscope. C is a metal wire, B represents two strips of paper. (c) The molecular amplifier resembles an electroscope with A being a positively charged metal ball, C a metal wire (i.e. a body with high polarizability) and B undergoes a huge charge redistribution due to the mediation of wire C.

$$\delta = 1 + 2\alpha_C \frac{R^2}{(R - R_{BC})^2 R_{BC}^3}. \tag{13.58}$$

For the polarizabilities chosen, polarization amplification ($\delta - 1$) takes place and at any R_{BC} exceeds 60%, see Fig. 13.14. When the amplifier is about in the middle of the AB distance the amplification is about 60%. When the amplifier approaches the electric field source (i.e. A^+), the amplification increases to about 100%. When the amplifier is close to B, it increases to about 200%.

This seems to be an interesting three-body effect we could investigate both theoretically and experimentally. Let us go a little crazy and assume C is made of a metal plate perpendicular to the AB line. Why a metal plate? Because the polarizability of a piece of metal is huge.[62] There is trouble though. The dipoles induced

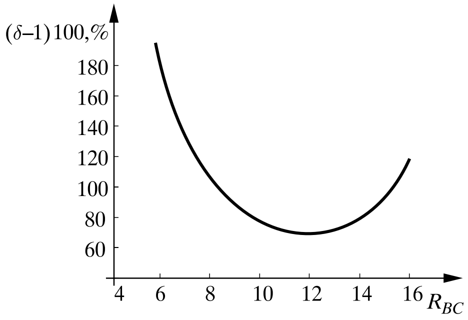


Fig. 13.14. Polarization amplification ($\delta - 1$ in %) on molecule B due to the mediation (three-body effect) of a polarizable molecule C ($\alpha_C = 100$ a.u.). The distance $R = 20$ a.u. The cation A^+ strongly polarizes molecule C. The dipole moment induced in this way on C, creates an additional electric field on B. This leads to polarization amplification on B.

Molecule C has the dipole moment (induced by the electric field from A^+). Let us calculate it as follows. The electric field created on mediator C by A^+ equals $\mathcal{E}_{A\to C} = \frac{1}{(R-R_{BC})^2}$, and therefore the corresponding induced dipole moment (component x) on C is $\mu_C = \alpha_C \frac{1}{(R-R_{BC})^2}$.

In the absence of the mediator C, the electric field on B would equal $\mathcal{E}_{A\to B} = \frac{1}{R^2}$, while with it (neglecting the self-consistency of the dipole moments on B and C) $\mathcal{E}_{A\to B} + \mathcal{E}_{C\to B}$. The ratio of the second and the first is given by eq. (13.58).

[62] Let us recall the description of the metallic state of Chapter 9, p. 454 (HOMO-LUMO degeneracy) and then the definition of polarizability in Chapter 12, p. 635. Since the HOMO-LUMO separation is 0, the polarizability of a metal gives ∞.

in the metal plate perpendicular to AB will be parallel to each other (side by side), which is energetically unfavourable. However, if the metal-plate is replaced by a metal wire oriented along line AB, everything would be amplified: the elementary dipoles would form a chain thus giving a big dipole within the wire. This means that the cation A^+ would attract a lot of the electrons within wire, so that on the opposite side of the wire we would have a sort of copy of it. Since the copy of A^+ would be very close to B, the polarization of B would increase very much.[63]

13.9.5 ADDITIVITY OF THE SECOND-ORDER DISPERSION ENERGY

The dispersion energy is a second-order correction, eq. (13.12) on p. 695 gives the formula for the interaction of two molecules. For *three* molecules we obtain the following formula for the dispersion part of the second-order effect (cf. the discussion on the induction energy on p. 736)

$$E_{\text{disp}}(ABC) = \sum_{n_A, n_B \neq (0_A, 0_B)} \frac{|\langle n_A n_B 0_C | V_{AB} + V_{BC} + V_{AC} | 0_A 0_B 0_C \rangle|^2}{[E_A(0_A) - E_A(n_A)] + [E_B(0_B) - E_B(n_B)]} + \cdots,$$

where $+\cdots$ denotes analogous terms with summations over n_A, n_C as well as n_B, n_C. Among three integrals in the nominator only the first one will survive, since the other vanish due to the integration over the coordinates of the electrons of molecule Z not involved in the interaction V_{XY}:

$$\begin{aligned} &E_{\text{disp}}(ABC) \\ &\quad = \sum_{n_A, n_B \neq (0_A, 0_B)} \frac{|\langle n_A n_B 0_C | V_{AB} | 0_A 0_B 0_C \rangle + 0 + 0|^2}{[E_A(0_A) - E_A(n_A)] + [E_B(0_B) - E_B(n_B)]} + \sum_{n_A, n_C \neq (0_A, 0_C)} \cdots \\ &\quad + \sum_{n_B, n_C \neq (0_B, 0_C)} \cdots. \end{aligned}$$

In the first term we can integrate over the coordinates of C. Then the first term displayed in the above formula turns out to be the dispersion interaction of A and B,

$$\begin{aligned} E_{\text{disp}}(ABC) &= \sum_{n_A, n_B \neq (0,0)} \frac{|\langle n_A n_B | V_{AB} | 0_A 0_B \rangle|^2}{[E_A(0_A) - E_A(n_A)] + [E_B(0_B) - E_B(n_B)]} \\ &\quad + \sum_{n_A, n_C \neq (0_A, 0_C)} \cdots + \sum_{n_B, n_C \neq (0_B, 0_C)} \cdots \\ &= E_{\text{disp}}(AB) + E_{\text{disp}}(AC) + E_{\text{disp}}(BC). \end{aligned}$$

---

[63] Is it something (Fig. 13.13.b) you may recall from a lesson in physics with an electroscope in your school? A glass rod (Fig. 13.13.c) rubbed by fur acquires a charge (an analogue of A^+), then it approaches a metal (analogue of C) protruding from a glass vessel it causes repulsion of two pieces of paper attached to the metal in the vessel. The induction has reached distant regions of space. If, instead of the pieces of paper we have molecule B, it would exhibit a large induced dipole moment.

Thus, we have proved that

> the dispersion interaction (second-order of the perturbation theory) *is additive*:
> $$E_{\text{disp}}(ABC) = E_{\text{disp}}(AB) + E_{\text{disp}}(AC) + E_{\text{disp}}(BC).$$

13.9.6 NON-ADDITIVITY OF THE THIRD-ORDER DISPERSION INTERACTION

One of the third-order energy terms represents a correction to the dispersion energy. The correction as shown by Axilrod and Teller[64] has a three-body character. The part connected to the interaction of three distant instantaneous dipoles on A, B and C reads as

$$E^{(3)}_{\text{disp}} = 3C^{(3)}_{\text{ddd}} \frac{1 + 3\cos\theta_A \cos\theta_B \cos\theta_C}{R^3_{AB} R^3_{AC} R^3_{BC}}, \tag{13.59}$$

where R_{XY} and θ_X denote the sides and the angles of the ABC triangle, and $C^{(3)}_{\text{ddd}} > 0$ represents a constant. The formula shows that

> when the ABC system is in a linear configuration, the dispersion contribution is negative, i.e. *stabilizing*, while the equilateral triangle configuration corresponds to a *destabilization*.

ENGINEERING OF INTERMOLECULAR INTERACTIONS

13.10 NOBLE GAS INTERACTION

Theoretical description of the noble gas interaction requires quite advanced computational techniques, because here the binding effect comes from the dispersion interaction, which represents an electronic correlation effect. Such an effect is inaccessible in Hartree–Fock calculations. Some very expensive post-Hartree–Fock methods have to be used. The larger the number of electrons (N), the more expensive the calculations quickly become as N increases (as we have seen in Chapter 10): proportionally to N^5 for the MP2 method, and even as N^7 for the CCSD(T) method. Therefore, whereas He_2 CCSD(T) calculations would take a minute, similar Xe_2 calculations would take about $(\frac{108}{4})^7 = 26^7$ minutes, i.e. about 3000 years. No wonder, the xenon atom has 54 electrons, and in a system of 108 electrons there are plenty of events to correlate, but because of the 3000 years this

---

[64] B.M. Axilrod, E. Teller, *J. Chem. Phys*. 11 (1943) 299.

is scary. To complete the horror, the calculations would have to be performed for many interatomic distances.

We may, however, make use of the following. First the calculations may be performed for He_2, Ne_2, Ar_2, Kr_2, Xe_2 using some reasonably poor basis sets. For each of the systems we obtain the equilibrium distance R_0 and the corresponding binding energy ε. Then, every curve $E(R)$ will be transformed (energy in ε units, distance in R_0 units) to $\frac{E(\frac{R}{R_0})}{\varepsilon}$. Every curve (independently of the system considered) has therefore depth 1 and minimum at $\frac{R}{R_0} = 1$.

> It turns out that all the curves coincide to good accuracy.[65]

Thus, all these objects are made out of the same matrix, despite the fact that this is so difficult to reveal using our computers. If we assume that this property were preserved for larger basis sets, we would be able to foresee the curve $E(R)$ for Xe_2 from good quality calculations for smaller noble gas dimers calculating $E(R_{\text{min}})$.

13.11 VAN DER WAALS SURFACE AND RADII

It would be of practical importance to know how close two molecules can approach each other. We will not enter this question too seriously, because this problem cannot have an elegant solution: it depends on the direction of approach, and the atoms involved, as well as how strongly the two molecules collide. Searching for the effective radii of atoms would be nonsense, if the valence repulsion were not a sort of "soft wall" or if the atom sizes were very sensitive to molecular details. Fortunately, it turns out that an atom, despite different roles played in molecules, can be characterized by its approximate radius, called the *van der Waals radius*. The radius may be determined in a naive, but quite effective, way. For example, we may approach two HF molecules axially with the fluorine atoms heading on, then find the distance[66] R_{FF} at which the interaction energy is equal to, say, 5 kcal/mol (repulsion). The proposed fluorine atom radius would be $r_{\text{F}} = \frac{R_{\text{FF}}}{2}$. A similar procedure may be repeated with two HCl molecules with the resulting r_{Cl}. Now, let us consider an axial complex H–F....Cl–H with the intermolecular distance corresponding to 5 kcal/mol. What F...Cl distance are we expecting? Of course, something close to $r_{\text{F}} + r_{\text{Cl}}$. It turns out that we are about right. This is why the atomic van der Waals radius concept is so attractive from the practical point of view.

van der Waals radius

We may define a superposition of atomic van der Waals spheres. This defines what is called the *van der Waals surface* of the molecule,[67] or a *molecular shape* – a concept of great importance and of the same arbitrariness as the radii themselves.

molecular shape

---

[65] Similar results have been obtained for the noble gas atom and sulphur atom interactions [J. Kłos, G. Chałasiński, R.V. Krems, A.A. Buchachenko, V. Aquilanti, F. Pirani, D. Cappelletti, *J. Chem. Phys.* 116 (2002) 9269].

[66] Using a reliable quantum mechanical method.

[67] The van der Waals surface of a molecule may sometimes be very complex, e.g., a molecule may have two or more surfaces (like fullerenes).

In a similar way we may define *ionic radii*,[68] to reproduce the ion packing in ionic crystals, as well as *covalent radii* to foresee chemical bond lengths.

13.11.1 PAULI HARDNESS OF THE VAN DER WAALS SURFACE

How would an atom penetrate the van der Waals surface? It depends on the particular molecule, surface point and atom. The helium atom seems to be a good probe, because of its simplicity and small size. The question may be more specific: what is the value of the valence repulsion gradient or, alternatively, the interaction energy gradient, when the atomic probe penetrates perpendicularly at a given point of the van der Waals isosurface? Such hardness depends on the particular spot on the isosurface and exhibits the symmetry of the molecule.[69]

The van der Waals surface might be modelled as one of the isosurfaces of the function $D(\boldsymbol{r}) = \sum_i A_i \exp(-B_i|\boldsymbol{r} - \boldsymbol{R}_i|)$, where the summation goes over the atoms of the molecule and the coefficients A_i and B_i depend not only on their kind (element), but also on their neighbourhood in the molecule. Therefore, we may propose

$$T(\boldsymbol{r}_0) = \left|(\nabla D)_{\boldsymbol{r}=\boldsymbol{r}_0}\right|$$

as the Pauli hardness at point $\boldsymbol{r}_0$ of the isosurface. Any point of the isosurface defined this way corresponds to a Pauli deformation of the wave function (Appendix Y) of the system: molecule and probe. *This represents another kind of deformation than that corresponding to the polarization of the molecule in an external electric field.* In one case the perturbation corresponds to a mechanical pushing, while in the other it pertains to the external electric field. The Pauli deformation will have complex anisotropic characteristic, when the probe penetrates the molecule. It is intriguing that, while the deformation due to the electric field results in an energy contribution of the second and higher orders, the Pauli deformation already appears in the first order energy correction.

13.11.2 QUANTUM CHEMISTRY OF CONFINED SPACE – THE NANOVESSELS

Molecules at long distances interact through the mediation of the electric fields created by them. The valence repulsion is of a different character, since it results from the Pauli exclusion principle, and may be interpreted as an energy penalty for an attempt by electrons of the same spin coordinate to occupy the same space (cf. Chapter 1 and p. 516).

Luty and Eckhardt[70] have highlighted the role of pushing one molecule by another. Let us imagine an atomic probe, e.g., a helium atom. The pushing by the probe deforms the molecular electronic wave function (Pauli deformation), but

---

[68] This concept was introduced by Pauling, based on crystallographic data (L. Pauling, *J. Amer. Chem. Soc.* 49 (1927) 765).

[69] Interestingly, water molecule is the hardest when approached in its plane about 44° off the OH direction, and the softest normal to the plane right above (and below) the oxygen atom. Data from E. Małolepsza, L. Piela, *J. Phys. Chem.* 107 (2003) 5356.

[70] T. Luty, C.J. Eckhardt, in "*Reactivity of Molecular Solids*", eds. E. Boldyreva, V. Boldyrev, Wiley, 1999, p. 51.

motion of the electrons is accompanied by the motion of the nuclei. Both motions may lead to dramatic events. For example, we may wonder how an explosive reaction takes place. Nothing happens during tens of years, and suddenly: boom! The spike hitting the material in its metastable chemical state is similar to the helium atom probe pushing a molecule. Due to the pushing, the molecule distorts to such an extent that the HOMO-LUMO separation vanishes and the system rolls down (see Chapter 14) to a deep potential energy minimum on the corresponding potential energy hypersurface. The HOMO-LUMO gap closing takes place within the reaction barrier. Since the total energy is conserved, the large reaction net energy gain goes to highly excited vibrational states (in the classical approximation corresponding to large amplitude vibrations). The amplitude may be sufficiently large to assure the pushing of the next molecules in the neighbourhood and a chain reaction starts with exponential growth.

Now imagine a lot of atomic probes confining the space (like a cage or template) available to a molecule under study. In such a case the molecule will behave differently from a free one. For example,

- a protein molecule, when confined, *will fold to another conformation*;[71]
- some photochemical reactions that require a space for the rearrangement of molecular fragments *will not occur*, if the space is not accessible;
- in a restricted space some other chemical reactions will take place (new chemistry – *chemistry in "nanovessels"*);
- some unstable molecules may *become stable* when enclosed in a nanovessel.

These are fascinating and little explored topics.

13.12 SYNTHONS AND SUPRAMOLECULAR CHEMISTRY

To make complex chemical structures, synthetic chemists take advantage of the large scale of the atom–atom binding energies: from strong chemical bonds (of the order of 100 kcal/mol) to weak intermolecular interactions (of the order of a fraction of kcal/mol). For over one hundred and fifty years (since the time of Butlerov and Kekulé) chemists have used theory (of various levels) to plan and then build chemical structures with some chemical bonds to be broken and others to be created. Often the substances do not resemble the reagents, and the structure is held together by

Alexandr Butlerov (1828–1886) Russian chemist, professor at the University of Kazan and Saint Petersburg. In 1861 Butlerov presented a concept of molecular *spatial structure*, where the atoms are bound by atom-to-atom chemical bonds, with properties characteristic for the atoms involved, an atom being able to bind only a few nearest neighbour atoms. Kazan University may be proud of several excellent scholars. Besides Butlerov, among others, there are one of the founders of the non-Euclidean geometry Nicolai Lobachevsky as well as the inventor of electronic paramagnetic resonance Evgeniy Zavoiski.

---

[73] For example, in E. Małolepsza, M. Boniecki, A. Koliński, L. Piela, *Proc. Nat. Acad. Sciences* 102 (2005) 7835 a theoretical model of the conformational autocatalysis is investigated. The native conformation of a protein becomes unstable in presence of a misfolded conformation of another molecule of the protein. The native conformation unfolds and refolds to the metastable conformation.

strong chemical bonds, and therefore may be called "hard architecture". The use of intermolecular interactions in synthesis ("soft architecture", supramolecular chemistry) has arisen only during the last few decades (since Cram, Pedersen and Lehn[72]). The supramolecular structures contain (as bricks) some loosely bound molecules, which therefore do not lose their individual properties.[73]

Friedrich August Kekulé von Stradonitz (1829–1896), German organic and theoretical chemist, professor at the universities in Gent and Bonn. In 1858 Kekulé proved, that carbon has valency four and in 1865 proposed the correct ring-like formula for benzene after a peculiar dream about a serpent eating its own tail.

It would seem that these "soft" structures are not interesting as they are unstable (it is sufficient to increase the temperature to make the structure disappear). The opposite is true, because such structures, after performing their function, may be destroyed without any significant energy expense.

13.12.1 BOUND OR NOT BOUND

Do the confined complexes such as catenans, rotaxans[74] and endohedral complexes (see Fig. 13.2, p. 688) represent intermolecular or intramolecular complexes? Certainly, when the distance between the subsystems, still within the structure of the complex, is large enough (this might be achieved by synthesis) the interaction is weak, as in any typical intermolecular interaction.

And what about the interaction of fragments of the same macromolecule that are close in space and at the same time distant, when walking through the framework of the chemical bonds? In this case we will also have some constraint for approaching two fragments, but chemists treat the interaction of two fragments of the DNA as if they were separate molecules. In such a way we have a coupling of the present section of the book with Chapter 7, where the force field contained the electrostatic interaction energy (of the net atomic charges, thus also taking into account higher-order molecular multipoles), valence repulsion and dispersion interaction (e.g., *via* terms r^{-12} and r^{-6} in the Lennard-Jones, p. 287). Among important contributions, only the induction energy is neglected in typical force fields.[75]

---

[72] Three scholars shared the 1987 Nobel Prize in chemistry for creating supramolecular chemistry, in particular "*for their development and use of molecules with structure-specific interactions of high selectivity*". Donald James Cram (b. 1919), American chemist, professor at the University of California–Berkeley; Charles John Pedersen (1904–1989), American chemist, employee of Dupont; Jean-Marie Lehn (b. 1939), French chemist, professor at the Université de Strasbourg and College de France in Paris.

[73] Although small modifications still take place.

[74] See a review article A.B. Braunschweig, B.H. Northrop, J.F. Stoddart, *J. Materials Chem.* 16 (2006) 32.

[75] Although the new generation of force fields take it into account, see W.D. Cornell, P. Cieplak, C.I. Bayly, I.R. Gould, K.M. Merz, Jr., D.M. Ferguson, D.C. Spellmeyer, T. Fox, J.W. Caldwell, P.A. Kollman, *J. Am. Chem. Soc.* 117 (1995) 5179.

13.12.2 DISTINGUISHED ROLE OF THE ELECTROSTATIC INTERACTION AND THE VALENCE REPULSION

> The electrostatic contribution plays a prominent role in intermolecular interaction. The electrostatic forces already operate effectively at long intermolecular distances (their range may, however, be reduced in polar solvents).

The induction and dispersion contributions, even if sometimes larger than the electrostatic interaction, usually play a less important role. This is because only the electrostatics may change the sign of the energy contribution when the molecules reorient, thus playing the pivotal role in the interaction energy.

> The induction and dispersion contributions are negative (at any orientation of the molecules), and we may say, as a rule of thumb, that their role is to make the configurations (already being stabilized by the electrostatics) more stable.

> The valence repulsion plays the role of a hard wall (covered by a "soft blanket") that forbids the closed-shell molecules to approach too closely. This represents a very important factor, since those molecules that do not fit together receive an energy penalty.

13.12.3 HYDROGEN BOND

Among the electrostatic interactions, the most important are those having a strong dependence on orientation, the most representative being the *hydrogen bonds X–H... Y, where an electronegative atom X plays the role of a proton donor, while an electronegative atom Y – plays the role of a proton acceptor. Most often the hydrogen bond X–H... Y deviates only a little from linearity.* Additionally, the XY separation usually falls into a narrow range: 2.5–3.2 Å, at least for the most important $X, Y \in \{O, N\}$. The hydrogen bond features are unique, because of the extraordinary properties of the hydrogen atom itself. This is the only atom which occasionally may attain the partial charge equal to +0.45 e, which means it represents a nucleus devoid to a large extent of an electron density. This is one of the reasons why the hydrogen bond is so strong when compared with other types of intermolecular interactions.

Example 5. ***Water–water dimer.*** Let us take the example of two water molecules to show the dominant role of electrostatics in the hydrogen bond.

As it is seen, while at the equilibrium distance $R_{OO} = 3.00$ Å all the contributions are of equal importance (although the electrostatics dominates), all the contributions except electrostatics, diminish considerably after increasing separation by only about 0.70 Å. For the largest separation ($R_{OO} = 4.76$), the electrostatics

Table 13.7. Energy contributions to the interaction energy E_{int} in the system HO–H…OH$_2$ (hydrogen bond) calculated[a] within the SAPT method: electrostatic energy E_{elst}, valence repulsion energy $E^{(1)}_{\text{exch}}$, induction energy E_{ind} and dispersion energy E_{disp} for three O…O distances: equilibrium distance $R_{\text{OO}} = 3.00$ and two distances a little larger: medium 3.70 Å and large 4.76 Å

Contributions to E_{int} (in kcal/mol)				
R_{OO} (Å)	E_{elst}	$E^{(1)}_{\text{exch}}$	E_{ind}	E_{disp}
3.00	−7.12	4.90	−1.63	−1.54
3.70	−2.79	0.30	−0.18	−0.31
4.76	−1.12	0.00	−0.02	−0.05

[a] B. Jeziorski, M. van Hemert, *Mol. Phys.* 31 (1976) 713.

dominates by far. This is why the hydrogen bond is said to have a mainly electrostatic character.[76]

13.12.4 COORDINATION INTERACTION

Coordination interaction appears if an electronic pair of one subsystem (electron donor) lowers its energy by interacting[77] with an electron acceptor offering an empty orbital, e.g., a cation (acceptor) interacts with an atom or atoms (donors) offering lone electronic pairs. This may be also seen as a special kind of electrostatic interaction.[78] Fig. 13.15.a shows a derivative of porphyrin as well as a cryptand (the name comes from the ritual of burying the dead in crypts), Fig. 13.15.b, the compounds offering lone pairs for the interaction with a cation.

cryptands

When concentrating on the ligands we can see that in principle they represent a negatively charged cavity (lone pairs) waiting for a monoatomic cation with *dimensions of a certain range only*. The interaction of such a cation with the ligand would be exceptionally large and therefore "specific" for such a pair of interacting moieties, which is related to the selectivity of the interaction.

Let us consider a water solution containing ions: Li$^+$, Na$^+$, K$^+$, Rb$^+$, Cs$^+$. After adding the above mentioned cryptand and after the equilibrium state is attained (ions/cryptand, ions/water and cryptand/water solvation), only for K$^+$ will the equilibrium be shifted towards the K$^+$/cryptand complex. For the other ions the equilibrium will be shifted towards their association with water molecules, not the cryptand.[79] This is remarkable information.

---

[76] It has been proved that covalent structures (cf. p. 520) also contribute to the properties of the hydrogen bond, but their role decreases dramatically when the molecules move apart.

[77] Forming a molecular orbital.

[78] A lone pair has a large dipole moment (see Appendix T), which interacts with the positive charge of the acceptor.

[79] J.-M. Lehn, "*Supramolecular Chemistry*", Institute of Physical Chemistry Publications, 1993, p. 88: the equilibrium constants of the ion/cryptand association reactions are: for Li$^+$, Na$^+$, K$^+$, Rb$^+$, Cs$^+$

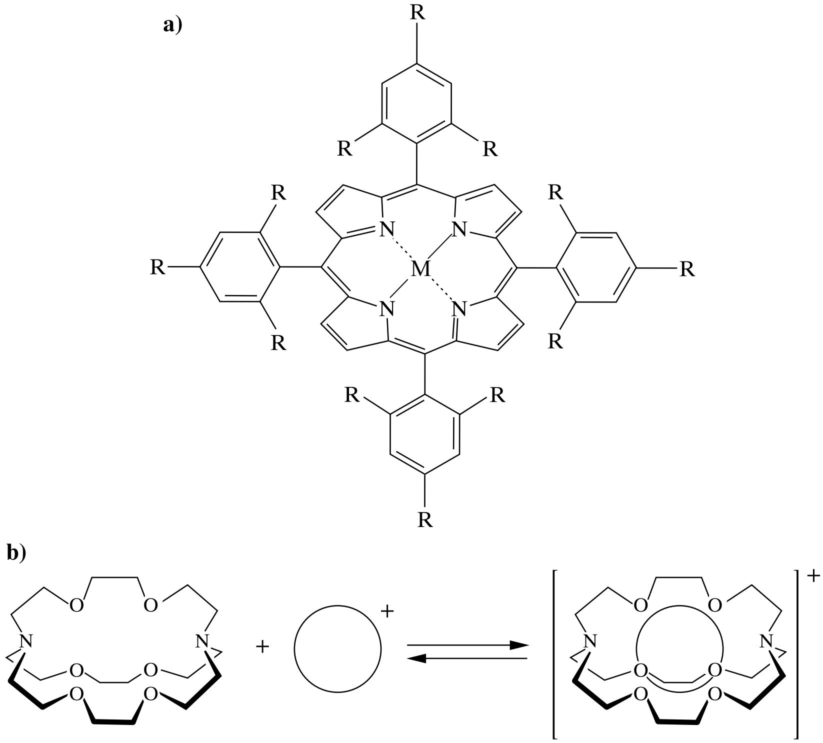


Fig. 13.15. A cation fits (a) the porphyrin ring or (b) the cryptand.

> We are able to selectively extract objects of some particular shape and dimensions (recognition).

13.12.5 HYDROPHOBIC EFFECT

This is quite a peculiar type of interaction, which appears mainly (not only) in water solutions.[80] The hydrophobic interaction does not represent any particular new interaction (beyond those we have already considered), because at least potentially they could be explained by the electrostatic, induction, dispersion, valence repulsion and other interactions already discussed, cf. pp. 718 and 695.

The problem may be seen from a different point of view. The basic interactions have been derived as if operating in vacuum. However, in a medium the molecules interact with one another through the mediation of other molecules, including those of the solvent. In particular, a water medium creates a strong network of

---

(only the order of magnitude is given): $10^2, 10^7, 10^{10}, 10^8, 10^4$, respectively. As seen the cryptand's cavity only fits well to the potassium cation.

[80] W. Kauzmann, *Advan. Protein Chem.* 14 (1959) 1. A contemporary theory is given in K. Lum, D. Chandler, J.D. Weeks, *J. Phys. Chem.* 103 (1999) 4570.

the hydrogen bonds that surround the hydrophobic moieties expelling them from the solvent[81] and pushing together which imitates their mutual attraction, resulting in the formation of a sort of "oil drop".

> We may say in a rather simplistic way that hydrophobic molecules aggregate not because they attract particularly strongly, but because water strongly prefers them to be out of its hydrogen bond net structure.

Hydrophobic interactions have a complex character and are not yet fully understood. The interaction depends strongly on the size of the hydrophobic synthons. For small sizes, e.g., such as two methane molecules in water, the hydrophobic interaction is small, increasing considerably for larger synthons. The hydrophobic effects become especially important for what is called the *amphiphilic* macromolecules with their van der Waals surfaces differing in hydrophobic character (hydrophobic/hydrophilic). The amphiphilic molecules are able to self-organize, forming structures up to the nanometer scale ("nanostructures").

amphiphilic molecules

self-organization

nano-structures

Fig. 13.16 shows an example of the hierarchic ("multi-level") character of a molecular architecture:

- The chemical binding of the amino acids into the oligopeptides is the first level ("hard architecture").
- The second level ("soft architecture") corresponds to a beautiful network of hydrogen bonds responsible for forming the α-helical conformation of each of the two oligopeptides.
- The third level corresponds to an extremely effective hydrophobic interaction, the *leucine-valine zipper*. Two α-helices form a very stable structure[82] winding up around each other and thus forming a kind of a superhelix, known as *coiled-coil*, due to the hydrophobic leucine-valine zipper.[83]

leucine-valine zipper

coiled-coil

The molecular architecture described above was first planned by a chemist. The system fulfilled all the points of the plan and self-organized in a spontaneous process.[84]

---

[81] Hydrophobic interactions involve not only the molecules on which we focus our attention, but also, to an important extent, the water molecules of the solvent. The hydrogen bond network keeps the hydrophobic objects together, as a shopping bag keeps lard slabs together.

The idea of solvent-dependent interactions represents a general and fascinating topic of research. Imagine the interaction of solutes in mercury, in liquid gallium, liquid sodium, in a highly polarizable organic solvent, etc. Due to the peculiarities of these solvents, we will have different chemistry going on in them.

[82] B. Tripet, L. Yu, D.L. Bautista, W.Y. Wong, T.R. Irvin, R.S. Hodges, *Prot. Engin*. 9 (1996) 1029.

[83] Leucine may be called the "flag ship" of the hydrophobic amino acids, although this is not the most polite compliment for a hydrophobe.

[84] One day I said to my friend Leszek Stolarczyk: "*If those organic chemists wanted to, they could synthesize anything you might dream of. They are even able to cook up in their flasks a molecule composed of the carbon atoms that would form the shape of a cavalry man on his horse*". Leszek answered: "*Of course! And the cavalry man would have a little sabre, made of iron atoms.*"

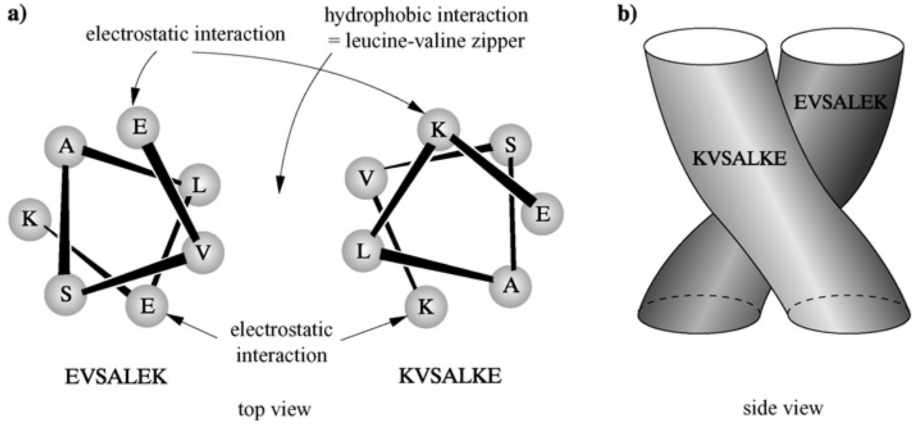


Fig. 13.16. An example of formation of the coiled-coil in the case of two oligopeptide chains (a): $(\mathrm{EVSALEK})_n$ with $(\mathrm{KVSALKE})_n$, with E standing for the glutamic acid, V for valine, S for serine, A for alanine, L for leucine, K for lysine. This is an example of a multi-level molecular architecture. First, each of the two oligopeptide chains form α-helices, which afterwards form a strong hydrophobic complex due to a perfect matching (leucine and valine of one of the α-helices with valine and leucine of the second one, known as the leucine-valine zipper (b)). The complex is made stronger additionally by two salt bridges (COO^- and NH_3^+ electrostatic interaction) involving pairs of glutamic acid (E) and lysine (K). The resulting complex (b) is so strong that it serves in analytical chemistry for the separation of some proteins.

13.12.6 MOLECULAR RECOGNITION – SYNTHONS

Organic molecules often have quite a few donor and acceptor substituents. These names may pertain to donating/accepting electrons or protons (cf. the charge conjugation described on p. 702). Sometimes a particular side of a molecule displays a system of donors and acceptors. Such a system “awaiting” interaction with a complementary object is called a synthon,[85] and their matching represents the *molecular recognition*. The cryptand in Fig. 13.15.b therefore contains a synthon able to recognize a narrow class of cations (with sizes within a certain range).

In Fig. 13.17 we show another example of synthons based on hydrogen bonds. Due to the particular geometry of the molecules as well as to the above mentioned weak dependence of the XY distance on X and Y, both synthons are complementary. The example is of immense importance, because it pertains to guanine (G), cytosine (C), adenine (A) and thymine (T). Thanks to these two pairs of synthons (GC and AT) we exist, because the G, C together with the A and T represent the four letters which are sufficient to write the Book of Life word by word in a single molecule of DNA. The words, the sentences and the chapters of this Book decide the majority of the very essence of your (and my) personality. The whole DNA strand may be considered as a large single synthon. The synthon has its important counterpart which fits the DNA perfectly because of the complementarity. The molecular machine which synthesizes this counterpart molecule (a “negative”) is

---

[85] G.R. Desiraju, “*Crystal Engineering, The Design of Organic Solids*”, Elsevier, Amsterdam, 1989.

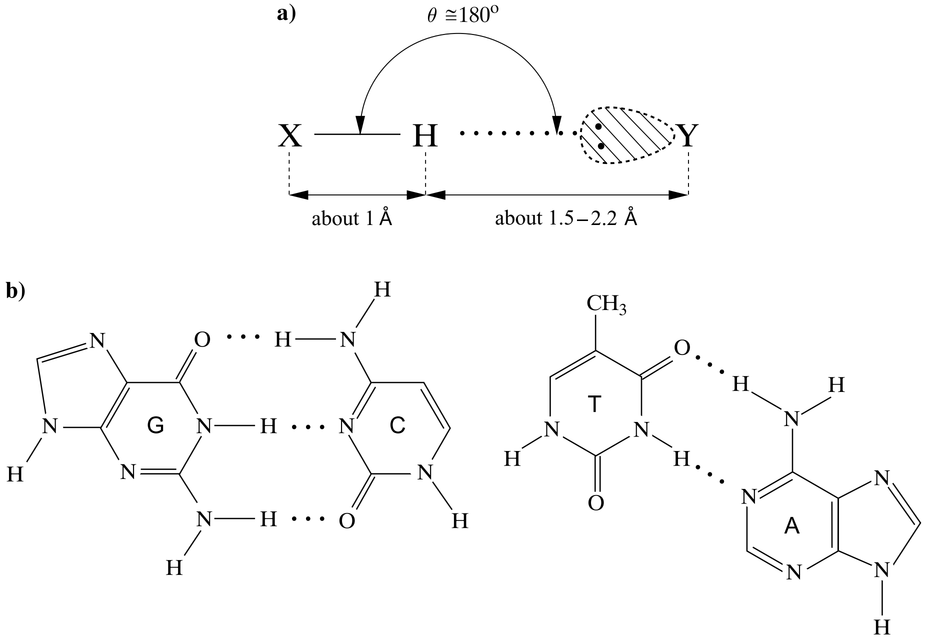


Fig. 13.17. Synthons are often based on a hydrogen bond pattern (a). The synthon of guanine (G) fits the synthon of cytosine (C), while the synthon of adenine (A) fits that of the thymine (T) (b).

the polymerase, a wonderful molecule (you will read about in Chapter 15). Any error in this complementarity results in a mutation.[86]

13.12.7 "KEY-LOCK", TEMPLATE AND "HAND-GLOVE" SYNTHON INTERACTIONS

The energy spectrum of a molecule represents something like its finger print. The particular energy levels correspond to various electronic, vibrational and rotational states (Chapter 6). Different electronic states[87] may be viewed as representing different chemical bond patterns. Different vibrational states[88] form series, each series for an energy well on the PES. The energy level pattern is completed by the rotational states of the molecule as a whole. Since the electronic excitations are of the highest energy, the PES of the ground electronic state is most important. For flexible molecules such a PES is characterized by a lot of potential energy wells corresponding to the conformational states. If the bottoms of the excited conformational wells are of high energy (with respect to the lowest-energy well, Fig. 13.18.a), then the molecule in its ground state may be called "rigid", because high energy is needed to change the molecular conformation.

---

[86] Representing a potential or real danger, as well as a chance for evolution.

[87] In the Born–Oppenheimer approximation, each corresponding to a potential energy hypersurface, PES.

[88] Including internal rotations, such as those of the methyl group.

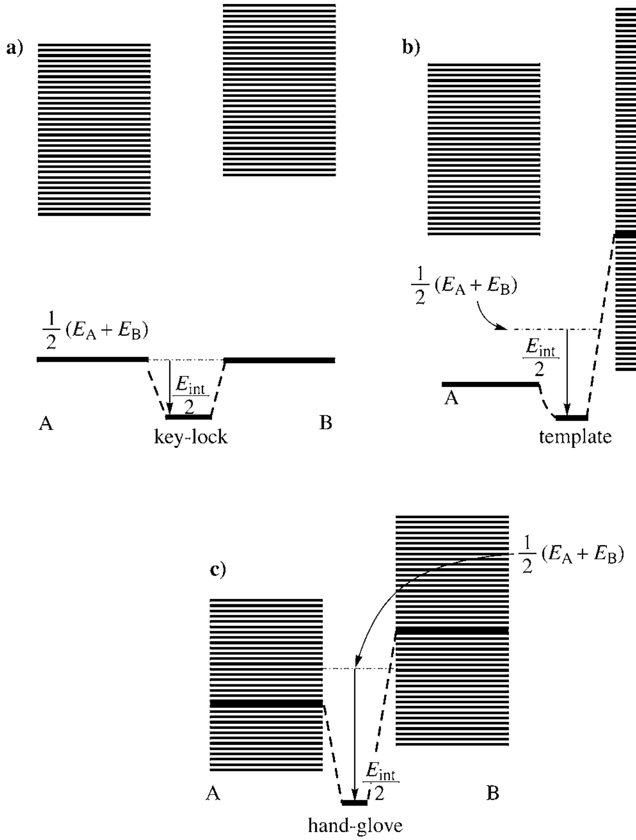


Fig. 13.18. The key-lock, template and hand-glove molecular recognition. Any molecule may be characterized by a spectrum of its energy states. (a) In the key-lock type interaction of two rigid molecules A and B their low-energy conformational states are separated from the quasi-continuum high-energy conformational states (including possibly those of some excited electronic states) by an energy gap, in general different for A and B. Due to the effective synthon interactions the energy per molecule lowers substantially with respect to that of the isolated molecules leading to the molecular recognition without significant changes of molecular shape. (b) In the template-like interaction one of the molecules is rigid (large energy gap), while the other one offers a quasi-continuum of conformational states. Among the later, there is one that (despite of being a conformational excited state), due to the perfect matching of synthons results in considerable energy lowering, much below the energy of isolated molecules. Thus, one of the molecules has to distort in order to get perfect matching. (c) In the hand-glove type of interaction the two interacting molecules offer quasi-continua of their conformational states. Two of the excited conformational states correspond to such molecular shapes as match each other perfectly and lower the total energy considerably. This lowering is so large that it is able to overcome the conformational excitation energy (an energy cost of molecular recognition).

If such rigid molecules A and B match each other, this corresponds to the *key-lock* type of molecular recognition. To match, the interacting molecules sometimes have only to orient properly in space when approaching one another and then dock (the AT or GC pairs may serve as an example). The key-lock concept of Fischer from 100 years ago (concerning enzyme–substrate interaction) is considered as the foundation of supramolecular chemistry – the chemistry that deals with the complementarity and matching of molecules.

One of the molecules, if rigid enough, may serve as a *template* for another molecule, which is flexible, Fig. 13.18.b, and together they form a strong complex. Finally two flexible molecules (Fig. 13.18.c) may pay an energy penalty for acquiring higher-energy conformations, but such ones which lead to a very strong interaction of the molecules in the *hand-glove* type of molecular recognition.[89]

Still another variation of this interaction comes into play, when during the approach, a new type of synthon appears, and the synthons match afterwards. For example in the Hodges superhelical structure (Fig. 13.16), only after formation of the α-helices does it turn out that the leucine and valine side chains of one helix match perfectly similar synthons of the second helix ("leucine-valine zipper").

Nature has done it routinely for millions of years. Endonuclease (EcoRV) represents an enzyme whose function is selective chemical bond breaking between nucleotides (linking the adenine and thymine) in a single DNA strand. Fig. 13.19 shows a model of the complex of EcorV with a fragment of DNA,[90] altogether about 62000 atoms. Fig. 13.19 highlights some aspects of the interaction.

Hermann Emil Fischer (1852–1919), German chemist, professor at the universities in Strasbourg, Munich, Erlangen, Würzburg, Berlin. Known mainly for his excellent works on the structure of sugar compounds. His (recognized decades later) correct determination of the absolute conformation of sugars was based solely on the analysis of their chemical properties. Even today this would require advanced physicochemical investigations. In 1902 Fischer received the Nobel Prize "*for his work on sugar and purine syntheses*".

"host–guest" complex

Note the hierarchic structure of the *host–guest* complex (DNA-EcoRV). DNA is a double-helix (Fig. 13.19.a) and this shape results mainly from the intermolecular A...T and G...C interactions through mediation of the hydrogen bonds. The enzyme EcoRV (Fig. 13.19.b) also has a highly organized structure, in particular six α-helices and a few β strands exhibit their characteristic hydrogen bond patterns (not displayed in the figure), these secondary structure elements fit together through the mediation of hydrophobic interactions. As we can see, the cavity of the EcorV is too small, but becomes larger when the guest molecule is accommodated ("hand-glove" effect), thus enabling an effective host–guest interaction. This example shows how important valence repulsion is. *If the EcoRV cavity differed much from that suitable to accommodate the guest molecule, the host when deforming would pay a too high an energetic price* and the energetic gain connected to docking would become too small to compensate for this energy expense.

---

[89] Our immunological system represents an excellent example. When a foreign agent enters the blood system, it is bound by an antibody that is *able to adapt its shape to practically any agent*. Moreover, a complex mechanism transmits the information about the agent's size and shape, and all this results in mass production of antibodies with the particular shape needed to bind the invader.

[90] L. Wróblewska, Master thesis, University of Warsaw, 2000.

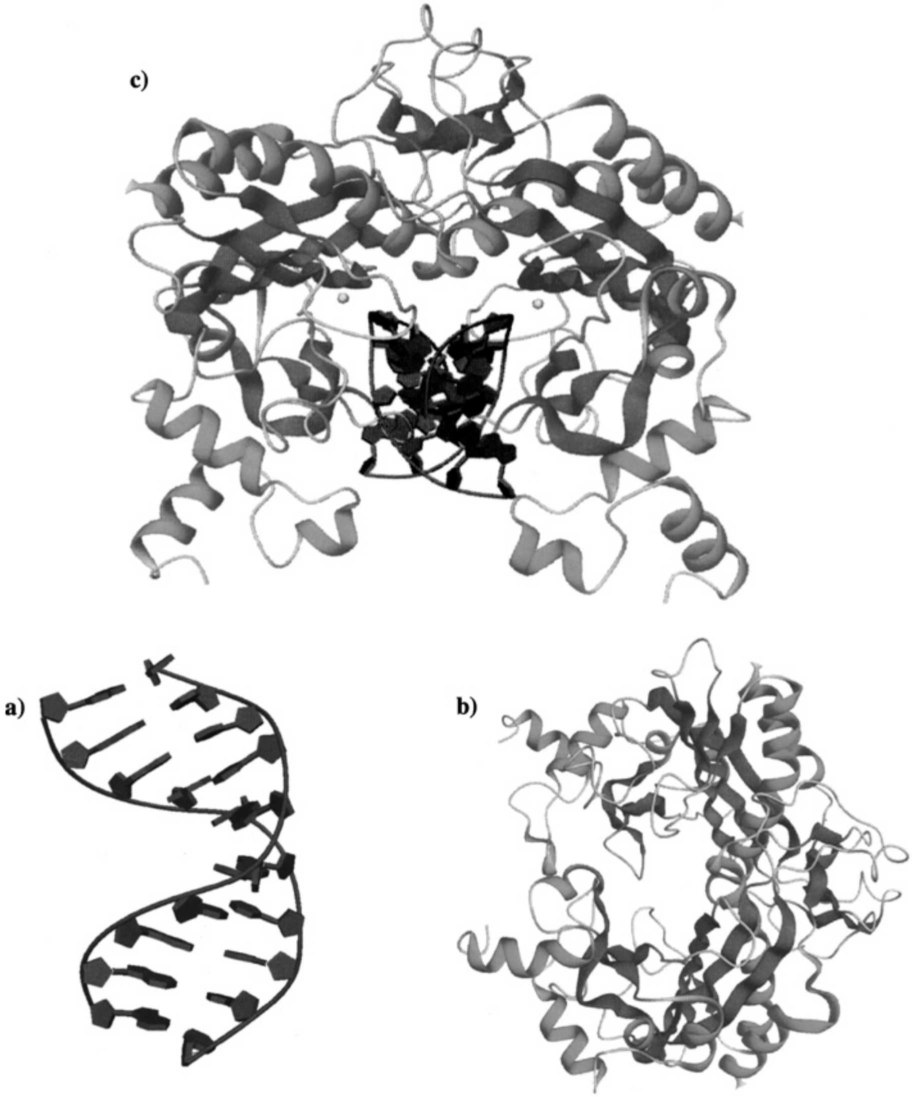


Fig. 13.19. The DNA fragment ("guest") fits the cavity in the enzyme EcoRV ("host") structure very well. (a) A fragment of the double-strand DNA helix (side view). (b) EcoRV. (c) Host–guest complex (the DNA molecule shown in the top view). Besides the geometric fitting (i.e. a lack of considerable valence repulsion) there is also an electrostatic and amphiphilic fitting of both subsystems.

Another masterpiece of nature – self-organization of the tobacco virus is shown in Fig. 13.20. Such a complex system self-assembles, because its parts not only fit one another (synthons), but also found themselves in solution and made perfect matching accompanied by an energy gain. Even more spectacular is the structure and functioning of bacteriophage T (Fig. 13.21).

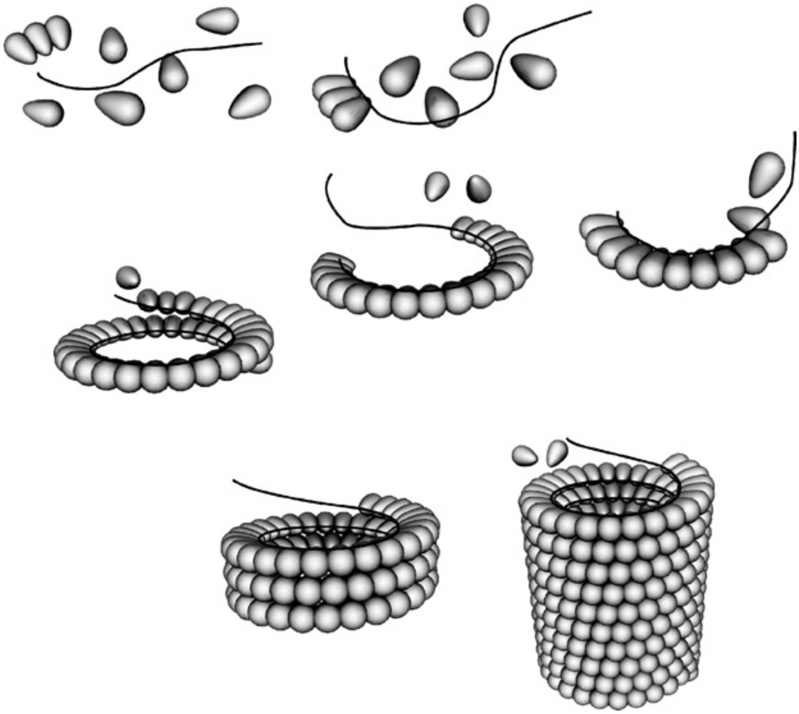


Fig. 13.20. Self-organization of the tobacco virus. The virus consists of an RNA helix (shown as a single strand) containing about 7000 nucleotides – sufficient genetic material to code the production of 5–10 proteins (first level of supramolecular self-organization). The RNA strand interacts very effectively with a certain protein (shown as a "drop"; the second level). The protein molecules associate with the RNA strand forming a kind of necklace, and then the system folds (third level) into a rod-like shape, typical for this virus. The rods are able to form a crystal (level four, not shown here), which melts after heating, but is restored when cooled down.

Summary

- Interaction energy of two molecules (at a given geometry) may be calculated within any reliable quantum mechanical method by subtracting from the total system energy the sum of the energies of the subsystems. This is called a *supermolecular method*.
- The supermolecular method has at least one important advantage: *it works independently of the interaction strength and of the intermolecular distance.* The method has the disadvantage that due to the subtraction, a loss of accuracy occurs and no information is obtained about the structure of the interaction energy.
- In the supermolecular method there is a need to compensate for what is called the basis set superposition error (BSSE). The error appears because due to the incompleteness of the atomic basis set (Ω_A, Ω_B), the individual subsystem A with the interaction switched off profits from the Ω_A basis set only, while when interacting lowers its energy due to the total $\Omega_A \cup \Omega_B$ basis set (the same pertains to any of the subsystems). As a result a part of the calculated interaction energy does not come from the interaction, but from the problem of the basis set used (BSSE) described above. A remedy is called the counterpoise method, in which all quantities (including the energies of the individual subsystems) are calculated within the $\Omega_A \cup \Omega_B$ basis set.

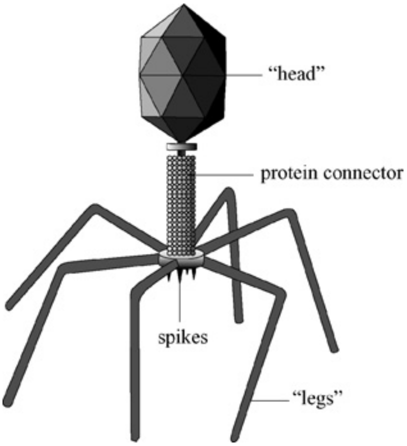


Fig. 13.21. Bacteriophage T represents a supramolecular construction that terrorizes bacteria. The hexagonal "head" contains a tightly packed double helix of DNA (the virus genetic material) wrapped in a coat build of protein subunits. The head is attached to a tube-like molecular connector built of 144 contractible protein molecules. On the other side of the connector there is a plate with six spikes protruding from it as well as six long, kinked "legs" made of several different protein molecules. The legs represent a "landing aparatus" which, using intermolecular interactions, attaches to a particular receptor on the bacterium cell wall. This reaction is reversible, but what happens next is highly irreversible. First, an enzyme belonging to the monster makes a hole in the cell wall of the bacterium. Then the 144 protein molecules contract probably at the expense of energy from hydrolysis of the ATP molecule (adenosine triphosphate – a universal energy source in biology), which the monster has at its disposal. This makes the head collapse and the whole monster serves as a syringe. The bacteriophage's genetic material enters the bacterium body almost in no time. That is the end of the bacterium.

- Perturbational method has limited applicability:
  - at long intermolecular separations what is called the polarization approximation may be used,
  - at medium distances a more general formalism called the symmetry adapted perturbation theory may be applied,
  - at short distances (of the order of chemical bond lengths) perturbational approaches are inapplicable.
- One of the advantages of a low-order perturbational approach is the possibility of dividing the interaction energy into well defined physically distinct energy contributions.
- In a polarization approximation approach, the unperturbed wave function is assumed as a *product* of the exact wave functions of the individual subsystems: $\psi_0^{(0)} = \psi_{A,0}\psi_{B,0}$. The corresponding zero-order energy is $E_0^{(0)} = E_{A,0} + E_{B,0}$.
- Then, the first-order correction to the energy represents what is called the *electrostatic* interaction energy: $E_0^{(1)} = E_{\text{elst}} = \langle\psi_{A,0}\psi_{B,0}|V\psi_{A,0}\psi_{B,0}\rangle$, which is the Coulombic interaction (at a given intermolecular distance) of the frozen charge density distributions of the individual, non-interacting molecules. After using the multipole expansion E_{elst} can be divided into the sum of the multipole–multipole interactions plus a remainder, called the penetration energy. A multipole–multipole interaction corresponds to the permanent multipoles of the isolated molecules. An individual multipole–multipole interaction term (2^k-pole with 2^l-pole) vanishes asymptotically as $R^{-(k+l+1)}$, e.g., the dipole–dipole term decreases as $R^{-(1+1+1)} = R^{-3}$.

- In the second order we obtain the sum of the induction and dispersion terms: $E^{(2)} = E_{\text{ind}} + E_{\text{disp}}$.
- The induction energy splits into

$$E_{\text{ind}}(A \to B) = \sum_{n_B}{}' \frac{|\langle \psi_{A,0}\psi_{B,n_B}|V\psi_{A,0}\psi_{B,0}\rangle|^2}{(E_{B,0} - E_{B,n_B})},$$

  which pertains to polarization of molecule B by the unperturbed molecule A, and

$$E_{\text{ind}}(B \to A) = \sum_{n_A}{}' \frac{|\langle \psi_{A,n_A}\psi_{B,0}|V|\psi_{a,0}\psi_{b,0}\rangle|^2}{(E_{A,0} - E_{A,n_A})},$$

  with the roles of the molecules exchanged. The induction energy can be represented as the permanent multipole – induced multipole interaction, where the interaction of the 2^k-pole with the 2^l-pole vanishes as $R^{-2(k+l+1)}$.
- The dispersion energy is defined as

$$E_{\text{disp}} = \sum_{n_A}{}' \sum_{n_B}{}' \frac{|\langle \psi_{A,n_A}\psi_{B,n_B}|V\psi_{A,0}\psi_{B,0}\rangle|^2}{(E_{A,0} - E_{A,n_A}) + (E_{B,0} - E_{B,n_B})}$$

  and represents a result of the electronic correlation. After applying the multipole expansion, the effect can be described as a series of instantaneous multipole – instantaneous multipole interactions, with the individual terms decaying asymptotically as $R^{-2(k+l+1)}$. The most important contribution is the dipole–dipole ($k = l = 1$), which vanishes as R^{-6}.
- The polarization approximation fails for medium and short distances. For medium separations we may use *symmetry-adapted perturbation theory* (SAPT). The unperturbed wave function is symmetry-adapted, i.e. has the same symmetry as the exact function. This is not true for the polarization approximation, where the product-like $\varphi^{(0)}$ does not exhibit the proper symmetry with respect to electron exchanges between the interacting molecules. The symmetry-adaptation is achieved by a projection of $\varphi^{(0)}$.
- Symmetry-adapted perturbation theory *reproduces all the energy corrections that appear in the polarization approximation* ($E_{\text{elst}}, E_{\text{ind}}, E_{\text{disp}}, \ldots$) *plus provides some exchange-type terms* (in any order of the perturbation except the zeroth).
- The most important exchange term is the valence repulsion appearing in the first-order correction to the energy:

$$\begin{aligned} E^{(1)}_{\text{exch}} &= \langle \psi_{A,0}\psi_{B,0}|V\hat{P}^{AB}\psi_{A,0}\psi_{B,0}\rangle \\ &\quad - \langle \psi_{A,0}\psi_{B,0}|V\psi_{A,0}\psi_{B,0}\rangle\langle \psi_{A,0}\psi_{B,0}|\hat{P}^{AB}\psi_{A,0}\psi_{B,0}\rangle + \mathrm{O}(S^4), \end{aligned}$$

  where $\mathrm{O}(S^4)$ represents all the terms decaying as the fourth power of the overlap integral(s) or faster, $\hat{P}^{AB}$ stands for the single exchanges' permutation operator.
- The interaction energy of N molecules *is not pairwise additive*, i.e. is not the sum of the interactions of all possible pairs of molecules. Among the energy corrections up to the second order, the exchange and, first of all, the induction terms contribute to the non-additivity. The electrostatic and dispersion (in the second order) contributions are pairwise additive.
- The non-additivity is highlighted in what is called the *many-body expansion of the interaction energy*, where the interaction energy is expressed as the sum of two-body, three-body, etc. energy contributions. The m-body interaction is defined as that part of the interaction energy that is non-explicable by any interactions of $m' < m$ molecules, but explicable by the interactions among m molecules.

- The dispersion interaction in the third-order perturbation theory contributes to the three-body non-additivity and is called the Axilrod–Teller energy. The term represents a correlation effect. Note that the effect is negative for three bodies in a linear configuration.
- The most important contributions: electrostatic, valence repulsion, induction and dispersion lead to a richness of supramolecular structures.
- The electrostatic interaction plays a particularly important role, because it is of a long-range character as well as very sensible to relative orientation of the subsystems. The hydrogen bond X–H…Y represents an example of the domination of the electrostatic interaction. This results in its directionality, linearity and a small (as compared to typical chemical bonds) interaction energy of the order of 5 kcal/mol.
- Also the valence repulsion is one of the most important energy contributions, because it controls how the interacting molecules fit together in space.
- The induction and dispersion interactions for polar systems, although contributing significantly to the binding energy, in most cases do not have a decisive role and only slightly modify the geometry of the resulting structures.
- In aqueous solutions the solvent structure contributes very strongly to the intermolecular interaction, thus leading to what is called the hydrophobic effect. The effect expels the non-polar subsystems from the solvent, thus causing them to approach, which *looks* like an attraction.
- A molecule may have such a shape that it fits that of another molecule (synthons, small valence repulsion and a large number of attractive atom–atom interactions).
- In this way molecular recognition may be achieved by the key-lock-type fit (the molecules non-distorted), template fit (one molecule distorted) or by the hand-glove-type fit (both molecules distorted).
- Molecular recognition may be planned by chemists and used to build complex molecular architectures, in a way similar to that in which living matter operates.

Main concepts, new terms

interaction energy (p. 684)
natural division (p. 684)
binding energy (p. 687)
dissociation energy (p. 687)
dissociation barrier (p. 687)
catenans (p. 688)
rotaxans (p. 688)
endohedral complexes (p. 688)
supermolecular method (p. 689)
basis set superposition error (BSSE) (p. 690)
ghosts (p. 690)
polarization perturbation theory (p. 692)
electrostatic energy (p. 693)
induction energy (p. 694)
dispersion energy (p. 694)
multipole moments (p. 698)
permanent multipoles (p. 701)
SAPT (p. 710)
function with adapted symmetry (p. 711)
symmetry forcing (p. 716)
polarization collapse (p. 717)
symmetrized polarization approximation (p. 717)
MS–MA perturbation theory (p. 717)
Jeziorski–Kołos perturbation theory (p. 717)
valence repulsion (p. 718)
Padé approximants (p. 722)
Pauli blockade (p. 722)
exchange–deformation interaction (p. 722)
interaction non-additivity (p. 726)
many-body expansion (p. 727)
SE mechanism (p. 733)
TE mechanism (p. 734)
polarization catastrophe (p. 738)
three-body polarization amplifier (p. 738)
Axilrod–Teller dispersion energy (p. 741)
van der Waals radius (p. 742)
van der Waals surface (p. 742)
supramolecular chemistry (p. 744)
hydrogen bond (p. 746)

hydrophobic effect (p. 748)
amphiphilicity (p. 749)
nanostructures (p. 749)
leucine-valine zipper (p. 749)
synthon (p. 750)
template (p. 751)
"key-lock" interaction (p. 751)
"hand-glove" interaction (p. 751)

From the research front

Intermolecular interactions influence any liquid and solid state measurements. Physicochemical measurement techniques give only some indications of the shape of a molecule, except NMR, X-ray and neutron analyses, which provide the atomic positions in space, but are very expensive. This is why there is a need for theoretical tools which may offer such information in a less-expensive way. For very large molecules, such an analysis uses the force fields described in Chapter 7. This is currently the most powerful theoretical tool for determining the approximate shape of molecules with numbers of atoms even of the order of hundreds of thousands. To obtain more reliable information about intermolecular interactions we may perform calculations within a supermolecular approach, necessarily of an *ab initio* type, because other methods give rather poor quality results. The DFT method popular nowadays fails at its present stage of development, because the intermolecular interactions area, especially the dispersion interaction, is a particularly weak point of the method. If the particular method chosen is the Hartree–Fock approach (currently limited to about 300 atoms), we have to remember that it cannot take into account any dispersion contribution to the interaction energy by *definition*.[91] *Ab initio* calculations of the correlation energy still represent a challenge. Good quality calculations for a molecule with a dozen atoms may be carried out using the MP2 method. Still more time consuming are the CCSD(T) or SAPT calculations, which are feasible for only a few atom systems, but offer an accuracy of 1 kcal/mol required for chemical applications.

Ad futurum...

No doubt the computational techniques will continue to push the limits mentioned above. The more coarse the method used, the more spectacular this pushing will be. The most difficult to imagine would be a great progress in methods using explicitly correlated wave functions. It seems that pushing the experimental demands and calculation time required will cause experimentalists (they will perform the calculations[92]) to prefer a rough estimation using primitive methods rather than wait too long for a precise result (still not very appropriate, because the results are obtained without taking the influence of solvent, etc. into account). It seems that in the near future we may expect theoretical methods exploiting the synthon concept. It is evident that a theoretician has to treat the synthons on an equal footing with other atoms, but a practice-oriented theoretician cannot do that, otherwise he would wait virtually forever for something to happen in the computer, while in reality the reaction takes only a picosecond or so. Still further into the future we will see the planning of hierarchic multi-level supramolecular systems, taking into account the kinetics and competitiveness among such structures. In the still more distant future, functions performed by such supramolecular structures, as well as their sensitivity to changing external conditions, will be investigated.

---

[91] The dispersion energy represents an electronic correlation effect, absent in the Hartree–Fock energy.

[92] What will theoreticians do? My answer is given in Chapter 15.

Additional literature

J.O. Hirschfelder, C.F. Curtiss, R.B. Bird, “Molecular Theory of Gases and Liquids”, Wiley, New York, 1964.

A thick “bible” (1249 pages) of intermolecular interactions. We find there everything before the advent of computers and of the symmetry-adapted perturbation theory.

H. Margenau, N.R. Kestner, “Theory of Intermolecular Forces”, Pergamon, Oxford, 1969.

A lot of detailed derivations. There is also a chapter devoted to the non-additivity of the interaction energy – a *rara avis* in textbooks.

A.J. Stone, “The Theory of Intermolecular Forces”, Oxford Univ. Press, Oxford, 1996.

The book contains the basic knowledge in the field of intermolecular interactions given in the language of perturbation theory as well as the multipole expansion (a lot of useful formulae for the electrostatic, induction and dispersion contributions). This very well written book presents many important problems in a clear and comprehensive way.

“Molecular Interactions”, ed. S. Scheiner, Wiley, Chichester, 1997.

A selection of articles written by experts.

P. Hobza, R. Zahradnik, “Intermolecular Complexes”, Elsevier, Amsterdam, 1988.

Many useful details.

I.G. Kaplan, “Intermolecular Interactions”, Wiley, Chichester, 2006.

A very well written book, covering broad and fundamental aspects of the field.

Questions

1. In order to avoid the basis set superposition error using the counter-poise method:
   a) all quantities have to be calculated within the total joint atomic basis set;
   b) the energy of the total system has to be calculated within the total joint atomic basis set, while the subsystem energies have to be calculated within their individual atomic basis sets;
   c) the total energy has to be calculated within a modest quality basis set, while for the subsystems we may allow ourselves a better quality basis set;
   d) all the atomic orbitals have to be centred in a single point in space.

2. In the polarization approximation the zero-order wave function is:
   a) the sum of the wave functions of the polarized subsystems;
   b) the product of the wave functions of the polarized subsystems;
   c) a linear combination of the wave functions for the isolated subsystems;
   d) the product of the wave functions of the isolated subsystems.

3. Please find the false statement:
   a) dissociation energy depends on the force constants of the normal modes of the molecule;
   b) interaction energy calculated for the optimal geometry is called the binding energy;
   c) the absolute value of the dissociation energy is larger than the absolute value of the binding energy;
   d) some systems with energy higher than the dissociation limit may be experimentally observed as stable.

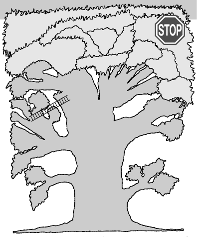
4. Please find the false statement. The induction energy (R stands for the intermolecular distance):
   a) represents an electronic correlation effect;
   b) for two water molecules decays asymptotically as R^{-6};
   c) is always negative (attraction);
   d) for the water molecule and the argon atom vanishes as R^{-6}.
5. Dispersion energy (R stands for the intermolecular distance):
   a) is non-zero only for the noble gas atom interaction;
   b) is equal to zero for the interaction of the two water molecules calculated with the minimal basis set;
   c) represents an electronic correlation effect;
   d) decays asymptotically as R^{-4}.
6. A particle has electric charge equal to q and Cartesian coordinates x, y, z. Please find the false statement:
   a) the multipole moments of the particle depend in general on the choice of the coordinate system;
   b) the operator of the z component of the dipole moment (μ_z) of the particle is equal to qz;
   c) a point-like particle does not have any dipole moment, therefore $\mu_z = 0$;
   d) the octupole moment operator of the particle is a polynomial of the third degree.
7. In symmetry adapted perturbation theory for He_2:
   a) we obtain the dispersion energy in the first order of the perturbation theory;
   b) the electronic energy exhibits a minimum as a function the interatomic distance;
   c) the binding energy is non-zero and is caused by the induction interaction;
   d) the unperturbed wave function is asymmetric (do not mix with antisymmetry) with respect to every electron permutation.
8. Please choose the false statement. In symmetry adapted perturbation theory:
   a) the zeroth order wave function is a product of the wave functions of the individual molecules;
   b) the electrostatic energy together with the valence repulsion appears in the first order;
   c) the exchange corrections appear in all (non-zero) orders;
   d) the Pauli exclusion principle is taken into account.
9. The non-additivity of the interaction energy has the following features:
   a) the electrostatic energy is additive, while the induction energy is not;
   b) the valence repulsion is additive for single exchanges;
   c) the dispersion interaction being a correlation effect is non-additive;
   d) the Axilrod–Teller interaction is additive.
10. Typical distances X…Y and H…Y for a hydrogen bond X–H…Y are closest to the following values (Å):
    a) 1.5, 1; b) 2.8, 1.8; c) 3.5, 2.5; d) 2.8, 1.

Answers

1a, 2d, 3c, 4a, 5c, 6c, 7b, 8a, 9a, 10b

Chapter 14

INTERMOLECULAR MOTION OF ELECTRONS AND NUCLEI: CHEMICAL REACTIONS



WHERE ARE WE?

Where are we?

We are already picking fruit in the crown of the TREE.

Example

Why do two substances react and another two do not? Why does increasing the temperature often start a reaction? Why does a reaction mixture change colour? As we know from Chapter 6, this tells us about some important electronic structure changes. On the other hand the products (as opposed to the reactants) tell us about profound changes in the positions of the nuclei that take place simultaneously. Something dramatic is going on. But what?

What is it all about

How atom A eliminates atom C from diatomic molecule BC? How can a chemical reaction be described as a molecular event? Where does the reaction barrier come from? Such questions will be discussed in this chapter.

The structure of the chapter is the following.

- Energy close to IRC
- Vibrationally adiabatic approximation
- Vibrationally non-adiabatic model
- Application of the reaction path Hamiltonian method to the reaction $H_2 + OH \rightarrow H_2O + H$

Acceptor–donor (AD) theory of chemical reactions (Ʊ♦) **p. 798**

- Maps of the molecular electrostatic potential
- Where does the barrier come from?
- MO, AD and VB formalisms
- Reaction stages
- Contributions of the structures as the reaction proceeds
- Nucleophilic attack $H^- +$ ethylene $\rightarrow$ ethylene $+ H^-$
- Electrophilic attack $H^+ + H_2 \rightarrow H_2 + H^+$
- Nucleophilic attack on the polarized chemical bond in the VB picture
- What is going on in chemist' flask?
- Role of symmetry
- Barrier means a cost of opening the closed-shells

Barrier for the electron transfer reaction (ƱⓈ) **p. 828**

- Diabatic and adiabatic potential
- Marcus theory

We are already acquainted with the toolbox for describing the electronic structure at *any* position of the nuclei. It is time now to look at possible *large* changes of the electronic structure at *large* changes of nuclear positions. The two motions: of the electrons and nuclei will be coupled together (especially in a small region of the configurational space).

Our plan consists of four parts:

- In the first part (after using the Born–Oppenheimer approximation, fundamental to this chapter), we assume that we have calculated the ground-state electronic energy, i.e. the potential energy for the nuclear motion. It will turn out that the hypersurface has a *characteristic "drain-pipe" shape, and the bottom in the central section, in many cases, exhibits a barrier*. Taking a three-atom example, we will show how the problem *could* be solved, if we *were capable* of calculating the quantum dynamics of the system accurately.
- In the second part we will concentrate on a specific representation of the system's energy that takes *explicitly* into account the above mentioned reaction drain-pipe ("*reaction path Hamiltonian*"). Then we will focus on describing how a chemical reaction proceeds. Just to be more specific, an example will be shown in detail.
- In the third part (acceptor–donor theory of chemical reactions) we will find the answer to the question, of *where the reaction barrier comes from and what happens to the electronic structure when the reaction proceeds*.
- The fourth part will pertain to the *reaction barrier height* in electron transfer (a subject closely related to the second and the third parts).

Why is this important?

Chemical reactions are at the heart of chemistry, making possible the achievement of its ultimate goals, which include synthesizing materials with desired properties. What happens in the chemist's flask is a complex phenomenon[1] which consists of an astronomical number of elementary reactions of individual molecules. In order to control the reactions in the flask, it would be good to *first understand the rules which govern these elementary reaction acts*.

What is needed?

- Hartree–Fock method (Chapter 8, necessary).
- Conical intersection (Chapter 6, necessary).
- Normal modes (Chapter 7, necessary).
- Appendices M (recommended), E (recommended), Z (necessary), I (recommended), G (just mentioned).
- Elementary statistical thermodynamics or even phenomenological thermodynamics: entropy, free energy (necessary).

Classical works

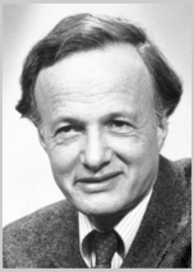
Everything in chemistry began in the twenties of the twentieth century.

The first publications that considered conical intersection – a key concept for chemical reactions – were two articles from the Budapest schoolmates: Janos (John) von Neumann and Jenó Pál (Eugene) Wigner "*Über merkwürdige diskrete Eigenwerte*" published in *Physikalische Zeitschrift*, 30 (1929) 465 and "*Über das Verhalten von Eigenwerten bei adiabatischen Prozessen*" which also appeared in *Physikalische Zeitschrift*, 30 (1929) 467. ★ Then a paper "*The Crossing of Potential Surfaces*" by their younger schoolmate Edward Teller was published in the *Journal of Chemical Physics*, 41 (1937) 109. ★ A classical theory of the "reaction drain-pipe" with entrance and exit channels was first proposed by Henry Eyring, Harold Gershinowitz and Cheng E. Sun in "*Potential Energy Surface for Linear* H_3", the *Journal of Chemical Physics*, 3 (1935) 786, and then by Joseph O. Hirschfelder, Henry Eyring and Bryan Topley in an article "*Reactions Involving Hydrogen Molecules and Atoms*" in *Journal of Chemical Physics*, 4 (1936) 170 and by Meredith G. Evans and Michael Polanyi in "*Inertia and Driving Force of Chemical Reactions*" which appeared in *Transactions of the Faraday Society*, 34 (1938) 11. ★ Hugh Christopher Longuet-Higgins, U. Öpik, Maurice H.L. Pryce and Robert A. Sack in a splendid paper "*Studies of the Jahn–Teller Effect*", *Proceedings of the Royal Society of London*, A244 (1958) 1 noted for the first time, that the wave function changes its phase close to a conical intersection, which later on became known as the Berry phase. ★ The acceptor–donor description of chemical reactions was first proposed by Robert S.J. Mulliken in "*Molecular Compounds and their Spectra*", *Journal of the American Chemical Society*, 74 (1952) 811. ★ The idea of the intrinsic reaction coordinate (IRC) was first given by Isaiah Shavitt in "*The Tunnel Effect Corrections in the Rates of Reactions with Parabolic and Eckart Barriers*", Report WIS-AEC-23, Theoretical Chemistry Lab., University of Wisconsin (1959) as well as by Morton A. Eliason and Joseph O. Hirschfelder in the *Journal of the Chemical Physics*, 30 (1959) 1426 in an article "*General Collision Theory Treatment for the Rate of Bimolecular, Gas Phase Reactions*". ★ The symmetry rules allowing some reactions and forbidding others were first proposed by Robert B. Woodward and Roald Hoffmann in two letters to the editor: "*Stereochemistry of Electrocyclic Reactions*" and "*Selection Rules for Sigmatropic Reactions*", *Journal of American Chemical Society*, 87 (1965) 395, 2511 as well as by Kenichi Fukui and Hiroshi Fujimoto in an article published

---

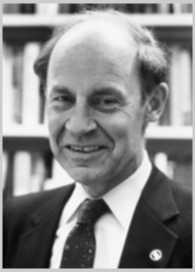
[1] See Chapter 15.

in the *Bulletin of the Chemical Society of Japan*, 41 (1968) 1989. ★ The concept of the steepest descent method was formulated by Kenichi Fukui in "*A Formulation of the Reaction Coordinate*", which appeared in the *Journal of Physical Chemistry*, 74 (1970) 4161, although the idea seems to have a longer history. ★ Other classical papers include a seminal article by Sason S. Shaik "*What Happens to Molecules as They React? Valence Bond Approach to Reactivity*" in *Journal of the American Chemical Society*, 103 (1981) 3692. ★ The Hamiltonian path method was formulated by William H. Miller, Nicolas C. Handy and John E. Adams, in an article "*Reaction Path Hamiltonian for Polyatomic Molecules*" in the *Journal of the Chemical Physics*, 72 (1980) 99. ★ The first quantum dynamics simulation was performed by a PhD student George C. Schatz (under the supervision of Aron Kupperman) for the reaction $H_2 + H \rightarrow H + H_2$, reported in "*Role of Direct and Resonant Processes and of their Interferences in the Quantum Dynamics of the Collinear* $H + H_2$ *Exchange Reaction*", in *Journal of Chemical Physics*, 59 (1973) 964. ★ John Polanyi, Dudley Herschbach and Yuan Lee proved that the lion's share of the reaction energy is delivered through the rotational degrees of freedom of the products, e.g., J.D. Barnwell, J.G. Loeser, D.R. Herschbach, "*Angular Correlations in Chemical Reactions. Statistical Theory for Four-Vector Correlations*" published in the *Journal of Physical Chemistry*, 87 (1983) 2781. ★ Ahmed Zewail (Egypt/USA) developed an amazing experimental technique known as femtosecond spectroscopy, which for the first time allowed the study of the reacting molecules at different stages of an ongoing reaction ("*Femtochemistry – Ultrafast Dynamics of The Chemical Bond*", vol. I and II, A.H. Zewail, World Scientific, New Jersey, Singapore (1994)). ★ Among others, Josef Michl, Lionel Salem, Donald G. Truhlar, Robert E. Wyatt, and W. Ronald Gentry contributed to the theory of chemical reactions.



John Charles Polanyi (born 1929), Canadian chemist of Hungarian origin, son of Michael Polanyi (one of the pioneers in the field of chemical reaction dynamics), professor at the University of Toronto. John was attracted to chemistry by Meredith G. Evans, who was a student of his father. Three scholars: John Polanyi, Yuan Lee and Dudley Herschbach shared the 1986 Nobel prize "*for their contributions concerning the dynamics of chemical elementary processes*".

Yuan T. Lee is a native of Taiwan, called by his colleagues "a Mozart of physical chemistry". He wrote that he was deeply impressed by a biography of Mme Curie and that her idealism decided his own path.



Dudley Herschbach writes in his CV, that he spent his childhood in a village close to San Jose, picking fruit, milking cows, etc. Thanks to his wonderful teacher he became interested in chemistry. He graduated from Harvard University (physical chemistry), where as he says, he has found "an exhilarating academic environment". In 1959 he became professor at University of California at Berkeley. In 1967 the group was joined by Yuan Lee and constructed a "supermachine" for studying crossing molecular beams and the reactions in them. One of the topics was the alkali metal atom – iodine collisions. These investigations were supported by John Polanyi, who studied the chemiluminescence in IR, i.e. the heat radiation of chemical reactions.

14.1 HYPERSURFACE OF THE POTENTIAL ENERGY FOR NUCLEAR MOTION

Theoretical chemistry is still in a stage which experts in the field characterized as "the primitive beginnings of chemical *ab initio* dynamics".[2] The majority of the systems studied so far are *three-atomic* systems.[3]

The Born–Oppenheimer approximation works wonders, as it is possible to consider the (classical or quantum) dynamics of the *nuclei*, while the electrons disappear from the scene (their role became, after determining the potential energy for the motion of the nuclei, described in the electronic energy, the quantity corresponding to $E_0^0(R)$ from eq. (6.8) on p. 225).

Even with this approximation our job is not simple:

- The reactants as well as the products may be quite large systems and the many-dimensional ground-state potential energy hypersurface $E_0^0(\boldsymbol{R})$ may have a very complex shape, whereas we are most often interested in the small fragment of the hypersurface that pertains to a particular one of many possible chemical reactions.
- We have many such hypersurfaces $E_k^0(\boldsymbol{R})$, $k = 0, 1, 2, \ldots$, each corresponding to an electronic state: $k = 0$ means the ground state, $k = 1, 2, \ldots$ correspond to the excited states. There are processes which take place on a single hypersurface without changing the chemical bond pattern,[4] but the very essence of chemical reaction is to change the bond pattern, and therefore excited states come into play.

It is quite easy to see where the fundamental difficulty is. Each of the hypersurfaces $E_k^0(\boldsymbol{R})$ for the motion of $N > 2$ nuclei depends on $3N - 6$ atomic coordinates (the number of translational and rotational degrees of freedom was subtracted).

Determining the hypersurface is not an easy matter:

- A high accuracy of 1 kcal/mol is required, which is (for a fixed configuration) very difficult to achieve for *ab initio* methods,[5] and even more difficult for the semi-empirical or empirical methods.

---

[2] R.D. Levine and R.B. Bernstein, "*Molecular Reaction Dynamics and Chemical Reactivity*", Oxford University Press, 1987.

[3] John Polanyi recalls that the reaction dynamics specialists used to write as the first equation on the blackboard A + BC → AB + C, which made any audience burst out laughing. However, one of the outstanding specialists (Richard Zare) said about the simplest of such reactions (H_3) (*Chem. Engin. News*, June 4 (1990) 32): "*I am smiling, when somebody calls this reaction the simplest one. Experiments are extremely difficult, because one does not have atomic hydrogen in the stockroom, especially the high speed hydrogen atoms (only these react). Then, we have to detect the product, i.e. the hydrogen, which is a transparent gas. On top of that it is not sufficient to detect the product in a definite spot, but we have to know which quantum state it is in*".

[4] Strictly speaking a change of conformation or formation of an intermolecular complex represents a chemical reaction. Chemists, however, reserve this notion for more profound changes of electronic structure.

[5] We have seen in Chapter 10, that the correlation energy is very difficult to calculate.

- The *number of points* on the hypersurface which have to be calculated is extremely large and increases exponentially with the system size.[6]
- There is no general methodology telling us what to do with the calculated points. There is a consensus that we should approximate the hypersurface by a smooth analytical function, but no general solution has yet been offered.[7]

14.1.1 POTENTIAL ENERGY MINIMA AND SADDLE POINTS

Let us denote $E_0^0(\boldsymbol{R}) \equiv V$. The most interesting points of the hypersurface V are its *critical points*, i.e. the points for which the gradient ∇V is equal to zero:

critical points

$$G_i = \frac{\partial V}{\partial X_i} = 0 \quad \text{for } i = 1, 2, \ldots, 3N, \tag{14.1}$$

where X_i denote the Cartesian coordinates that describe the configurations of N nuclei. Since $-G_i$ represents the force acting along the axis X_i, therefore no forces act on the atoms in the configuration of a critical point.

There are several types of critical points. Each type can be identified after considering the *Hessian*, i.e. the matrix with elements

Hessian

$$V_{ij} = \frac{\partial^2 V}{\partial X_i \partial X_j} \tag{14.2}$$

calculated for the critical point. There are three types of critical points: maxima, minima and saddle points (cf. Chapter 7 and Fig. 7.11, as well as the Bader analysis, p. 573). The saddle points, as will be shown in a while, are of several classes depending on the signs of the Hessian eigenvalues. Six of the eigenvalues are equal to zero (rotations and translations of the total system, see p. 294), because this type of motion proceeds without any change of the potential energy V.

We will concentrate on the remaining $3N - 6$ eigenvalues:

- In the minimum the $3N - 6$ Hessian eigenvalues $\lambda_k \equiv \omega_k^2$ (ω is the angular momentum of the corresponding normal modes) are all positive,
- In the maximum – all are negative.
- For a saddle point of the n-th order, $n = 1, 2, \ldots, 3N - 7$, the n eigenvalues are negative, the rest are positive. Thus, a first-order saddle point corresponds to all

---

[6] Indeed, if we assume that ten values for each coordinate axis is sufficient (and this looks like a rather poor representation), then for N atoms we have 10^{3N-6} quantum mechanical calculations of good quality to perform. This means that for $N = 3$ we may still pull it off, but for larger N everybody has to give up. For example, for the reaction $HCl + NH_3 \rightarrow NH_4Cl$ we would have to calculate 10^{12} points in the configurational space, while even a single point is a computational problem.

[7] Such an approximation is attractive for two reasons: first, we dispose of the (approximate) values of the potential energy for *all* points in the configuration space (not only those for which the calculations were performed), and second, the analytical formula may be differentiated and the derivatives give the forces acting on the atoms.

It is advisable to construct the above mentioned analytical functions following some theoretical arguments. These are supplied by intermolecular interaction theory (see Chapter 13).

but one the Hessian eigenvalues positive, i.e. one of the angular frequencies ω is therefore imaginary.

saddle

The eigenvalues were obtained by diagonalization of the Hessian. Such diagonalization corresponds to a rotation of the local coordinate system (cf. p. 297). Imagine a two-dimensional surface that at the minimum could be locally approximated by an ellipsoidal valley. The diagonalization means such a rotation of the coordinate system x, y that both axes of the ellipse coincide with the new axes x', y' (Chapter 7). On the other hand, if our surface locally resembled a cavalry saddle, diagonalization would lead to such a rotation of the coordinate system that one axis would be directed along the horse, and the other across.[8]

femtosecond spectroscopy

IR and Raman spectroscopies providing the vibration frequencies and force constants tell us a lot about how the energy hypersurface close to minima, looks, both for the reactants and the products. On the other hand theory and recently also femtosecond spectroscopy,[9] are the only source of information about the first order saddle points. However, the latter are extremely important for determining reaction rates since any saddle point is a kind of pivot point – it is as important for the reaction as the Rubicon was for Caesar.[10]

The simplest chemical reactions are those which do not require crossing any reaction barrier. For example, the reaction $\text{Na}^+ + \text{Cl}^- \rightarrow \text{NaCl}$ or other similar reactions (like recombination of radicals) that are not accompanied by bond breaking take place *without any barrier*.[11]

After the barrierless reactions, there is a group of reactions in which the reactants and the products are separated by a single first-order saddle point (no intermediate products). *How do we describe such a reaction in a continuous way?*

14.1.2 DISTINGUISHED REACTION COORDINATE (DRC)

We often define a reaction path in the following way. First, we make

- a choice of a particular distance (s) between the reacting molecules (e.g., an interatomic distance, one of the atoms belongs to molecule A, the other to B);
- then we minimize the potential energy by optimization of all atomic positions, while keeping the s distance fixed;
- change s by small increments from its reactant value until the product value is obtained (for each s optimizing all other distances);

---

[8] A cavalry saddle represents a good example of the first order saddle of a two-dimensional surface.

[9] In this spectroscopy we hit a molecule with a laser pulse of a few femtoseconds. The pulse perturbs the system, and when relaxing it is probed by a series of new pulses, each giving a spectroscopic fingerprint of the system. A femtosecond is an incredibly short time, light is able to move only about $3 \cdot 10^{-5}$ cm. Ahmed Zewail, the discoverer of this spectroscopy received the Nobel prize 1999.

[10] In 49 B.C. Julius Caesar with his Roman legions crossed the Rubicon river (the border of his province of Gaul), and this initiated a civil war with the central power in Rome. His words, "*alea iacta est*" (the die is cast) became a symbol of a final and irreversible decision.

[11] As a matter of fact, the formation of van der Waals complexes may also belong to this group. However in large systems, when precise docking of molecules take place, the final docking may occur with a steric barrier.

- this defines a path (DRC) in the configurational space, the progress along the path is measured by s.

A deficiency of the DRC is an arbitrary choice of the distance. The energy profile obtained (the potential energy vs s) depends on the choice. Often the DRC is reasonably close to the reactant geometry and becomes misleading when close to the product value (or *vice versa*). There is no guarantee that such a reaction path passes through the saddle point. On top of this other coordinates may undergo discontinuities, which feels a little catastrophic.

14.1.3 STEEPEST DESCENT PATH (SDP)

Because of the Boltzmann distribution the potential energy *minima* are most important, mainly low-energy ones.[12]

The saddle points of the first order are also important, because we may prove that any two minima may be connected by at least one saddle point[13] which corresponds to the highest energy *point* on the lowest-energy *path* from one minimum to the other (pass). Thus, *the least energy-demanding path from the reactants to products goes via a saddle point of the first order*. This steepest descent path (SDP) is determined by the direction $-\nabla V$. First, we choose a first-order saddle point $\boldsymbol{R}_0$, then diagonalize the Hessian matrix calculated at this point and the eigenvector $\boldsymbol{L}$ corresponding to the single negative eigenvalue of the Hessian (cf. p. 297). Now, let us move *a little* from position $\boldsymbol{R}_0$ in the direction indicated by $\boldsymbol{L}$, and then let us follow vector $-\nabla V$ until it reduces to zero (then we are at the minimum). In this way we have traced half the SDP. The other half will be determined starting down from the other side of the saddle point and following the $-\boldsymbol{L}$ vector first.

In a moment we will note a certain disadvantage of the SDP, which causes us to prefer another definition of the reaction path (see p. 781).

14.1.4 OUR GOAL

We would like to present a theory of elementary chemical reactions within the Born–Oppenheimer approximation, i.e. which describes nuclear motion on the potential energy hypersurface.

We have the following alternatives:

1. To perform *molecular dynamics*[14] on the hypersurface V (a point on the hypersurface represents the system under consideration).

---

[12] Putting aside some subtleties (e.g., does the minimum support a vibrational level), the minima correspond to stable structures, since a small deviation from the minimum position causes a gradient of the potential to become non-zero, and this means a force pushing the system back towards the minimum position.

[13] Several first-order saddle points to pass mean a multi-stage reaction that consists of several steps, each one representing a pass through a single first-order saddle point (elementary reaction).

[14] A classical approach. We have to assume that the bonds may break – this is a very non-typical molecular dynamics problem.

2. To solve the *time-independent Schrödinger equation* $\hat{H}\psi = E\psi$ for the motion of the nuclei with potential energy V.
3. To solve the time-dependent *Schrödinger equation* with the boundary condition for $\psi(x, t = 0)$ in the form of a wave packet.[15] The wave packet may be directed into the entrance channel towards the reaction barrier (from various starting conditions). In the barrier range, the wave packet splits into a wave packet crossing the barrier and a wave packet reflected from the barrier (cf. p. 153).
4. To perform a semi-classical analysis that highlights the existence of the SDP, or a similar path, leading from the reactant to the product configuration.

Before going to more advanced approaches let us consider possibility 1.

14.1.5 CHEMICAL REACTION DYNAMICS (A PIONEERS' APPROACH)

The SDP does not represent the only possible reaction path. It is only the *least-energy expensive path* from reactants to products. In real systems, the point representing the system will attempt to get through the pass in many different ways. Many such attempts are unsuccessful (*non-reactive trajectories*). If the system leaves the entrance channel (*reactive trajectories*), it will not necessarily pass through the saddle point, because it may have some extra kinetic energy, which may allow it to go with a higher energy than that of the barrier. Everything depends on the starting position and velocity of the point running through the entrance channel.

reactive trajectories

In the simplest case of a three-atom reaction

$$\mathrm{A} + \mathrm{BC} \rightarrow \mathrm{AB} + \mathrm{C}$$

the potential energy hypersurface represents a function of $3N - 6 = 3$ coordinates (the translations and rotations of the total system were separated). Therefore, even in such a simple case, it is difficult to draw this dependence. We may simplify the problem by considering only a limited set of geometries, e.g., the three atoms in a linear configuration. In such a case we have only two independent variables[16] R_{AB}

---

[15] For example, a Gaussian function (in the nuclear coordinate space) moving from a position in this space with a starting velocity.

[16] After separating the centre-of-mass motion. The separation may be done in the following way. The kinetic energy operator has the form

$$\hat{T} = -\frac{\hbar^2}{2M_A}\frac{\partial^2}{\partial X_A^2} - \frac{\hbar^2}{2M_B}\frac{\partial^2}{\partial X_B^2} - \frac{\hbar^2}{2M_C}\frac{\partial^2}{\partial X_C^2}.$$

We introduce some new coordinates:

- the centre-of-mass coordinate $X_{CM} = (M_A X_A + M_B X_B + M_C X_C)/M$ with the total mass $M = M_A + M_B + M_C$,
- $R_{AB} = X_B - X_A$,
- $R_{BC} = X_C - X_B$.

and R_{BC} and the function $V(R_{AB}, R_{BC})$ may be visualized by a map quite similar to those used in geography. The map has a characteristic shape shown in Fig. 14.1.

- *Reaction map*. First of all we can see the characteristic *"drain-pipe"* shape of the potential energy V for the motion of the nuclei, i.e. the function $V(R_{AB}, R_{BC}) \to \infty$ for $R_{AB} \to 0$ or for $R_{BC} \to 0$, therefore we have a high energy wall along the axes. When R_{AB} and R_{BC} are both large we have a kind of plateau that goes gently downhill towards the bottom of the curved drain-pipe extending nearly parallel to the axes. The chemical reaction A + BC → AB + C means a motion close to the bottom of the "drain-pipe" from a point corresponding to a large R_{AB}, while R_{BC} has a value corresponding to the equilibrium BC length to a point, corresponding to a large R_{BC} and R_{AB} with a value corresponding to the length of the isolated molecule AB (arrows in Fig. 14.1).

reaction drain-pipe

- *Barrier.* A projection of the "drain-pipe" bottom on the R_{AB} R_{BC} plane gives the SDP. Therefore, the SDP represents one of the important features of the "landscape topography". Travel on the potential energy surface along the SDP

reaction barrier

To write the kinetic energy operator in the new coordinates we start with relations

$$\frac{\partial}{\partial X_A} = \frac{\partial R_{AB}}{\partial X_A}\frac{\partial}{\partial R_{AB}} + \frac{\partial X_{CM}}{\partial X_A}\frac{\partial}{\partial X_{CM}} = -\frac{\partial}{\partial R_{AB}} + \frac{M_A}{M}\frac{\partial}{\partial X_{CM}},$$

$$\frac{\partial}{\partial X_B} = \frac{\partial R_{AB}}{\partial X_B}\frac{\partial}{\partial R_{AB}} + \frac{\partial R_{BC}}{\partial X_B}\frac{\partial}{\partial R_{BC}} + \frac{\partial X_{CM}}{\partial X_B}\frac{\partial}{\partial X_{CM}} = \frac{\partial}{\partial R_{AB}} - \frac{\partial}{\partial R_{BC}} + \frac{M_B}{M}\frac{\partial}{\partial X_{CM}},$$

$$\frac{\partial}{\partial X_C} = \frac{\partial R_{BC}}{\partial X_C}\frac{\partial}{\partial R_{BC}} + \frac{\partial X_{CM}}{\partial X_C}\frac{\partial}{\partial X_{CM}} = \frac{\partial}{\partial R_{BC}} + \frac{M_C}{M}\frac{\partial}{\partial X_{CM}}.$$

After squaring these operators and substituting them into $\hat{T}$ we obtain, after a brief derivation,

$$\hat{T} = -\frac{\hbar^2}{2M}\frac{\partial^2}{\partial X_{CM}^2} - \frac{\hbar^2}{2\mu_{AB}}\frac{\partial^2}{\partial R_{AB}^2} - \frac{\hbar^2}{2\mu_{BC}}\frac{\partial^2}{\partial R_{BC}^2} + \hat{T}_{ABC},$$

where the reduced masses

$$\frac{1}{\mu_{AB}} = \frac{1}{M_A} + \frac{1}{M_B}, \frac{1}{\mu_{BC}} = \frac{1}{M_B} + \frac{1}{M_C},$$

whereas $\hat{T}_{ABC}$ stands for the mixed term

$$\hat{T}_{ABC} = -\frac{\hbar^2}{M_B}\frac{\partial^2}{\partial R_{AB}\partial R_{BC}}.$$

In this way we obtain the centre-of-mass motion separation (the first term). The next two terms represent the kinetic energy operators for the independent pairs AB and BC, while the last one is the mixed term $\hat{T}_{ABC}$, whose presence is understandable: atom B participates in two motions, those associated with: R_{AB} and R_{BC}. We may eventually get rid of $\hat{T}_{ABC}$ after introducing a skew coordinate system with the R_{AB} and R_{BC} axes (the coordinates are determined by projections parallel to the axes). After a little derivation, we obtain the following condition for the angle θ between the two axes, which assures the mixed term:

$$\cos\theta_{\text{opt}} = \frac{2}{M_B}\frac{\mu_{AB}\mu_{BC}}{\mu_{AB} + \mu_{BC}}$$

vanish. If all the atoms have their masses equal, we obtain $\theta_{\text{opt}} = 60°$.

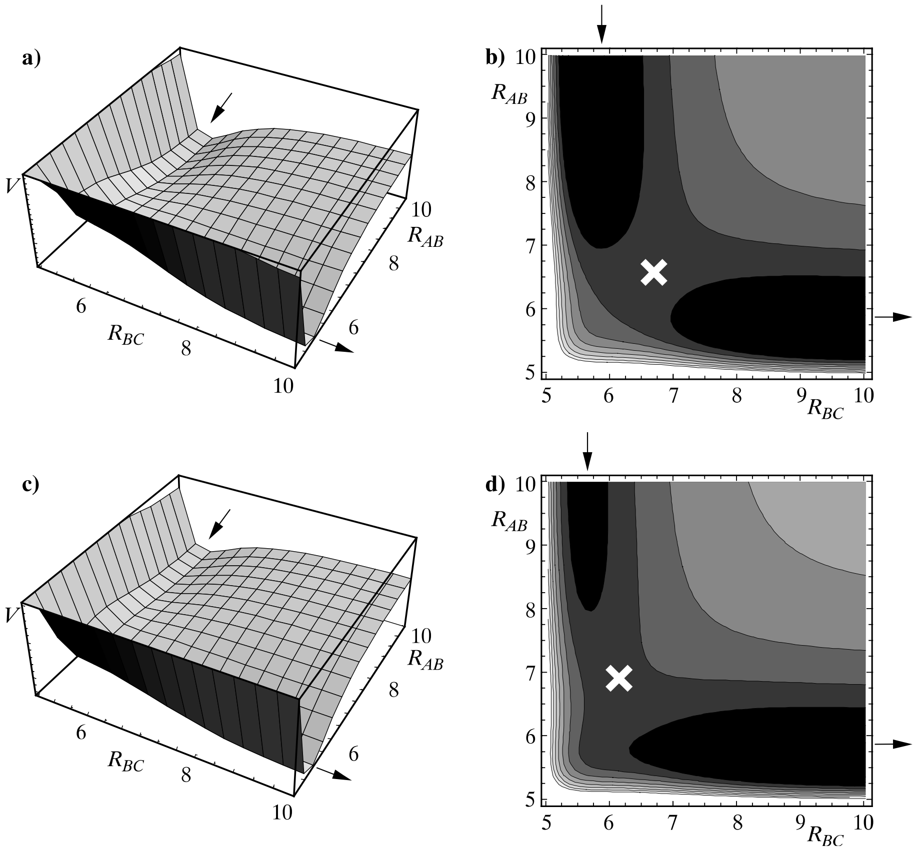


Fig. 14.1. The "drain-pipe" A + BC → AB + C (for a fictitious system). The surface of the potential energy for the motion of the nuclei is a function of distances R_{AB} and R_{BC}. On the left-hand side there is the view of the surface, while on the right-hand side the corresponding maps are shown. The barrier positions are given by the crosses on the right-hand figures. Figs. (a) and (b) show the symmetric entrance and exit channels with the separating barrier. Figs. (c) and (d) correspond to an exothermic reaction with the barrier in the entrance channel ("an early barrier"). Figs. (e) and (f) correspond to an endothermic reaction with the barrier in the exit channel ("a late barrier"). This endothermic reaction will not proceed spontaneously, because due to the equal width of the two channels, the reactant free energy is lower than the product free energy. Figs. (g) and (h) correspond to a spontaneous endothermic reaction, because due to the much wider exit channel (as compared to the entrance channel) the free energy is lower for the products. There is a van der Waals complex well in the entrance channel just before the barrier. There is no such well in the exit channel.

entrance and exit channel

is not a flat trip, because the drain-pipe consists of two valleys: the reactant valley (*entrance channel*) and the product valley (*exit channel*) separated by a pass (*saddle point*), which causes the reaction barrier. The saddle point corresponds to the situation, in which the old chemical bond is already weakened (but still exists), while the new bond is just emerging. This explains (as has been shown by Henry Eyring, Michael Polanyi and Meredith Evans) why the energy required to go from the entrance to the exit barrier is much smaller than the dissociation

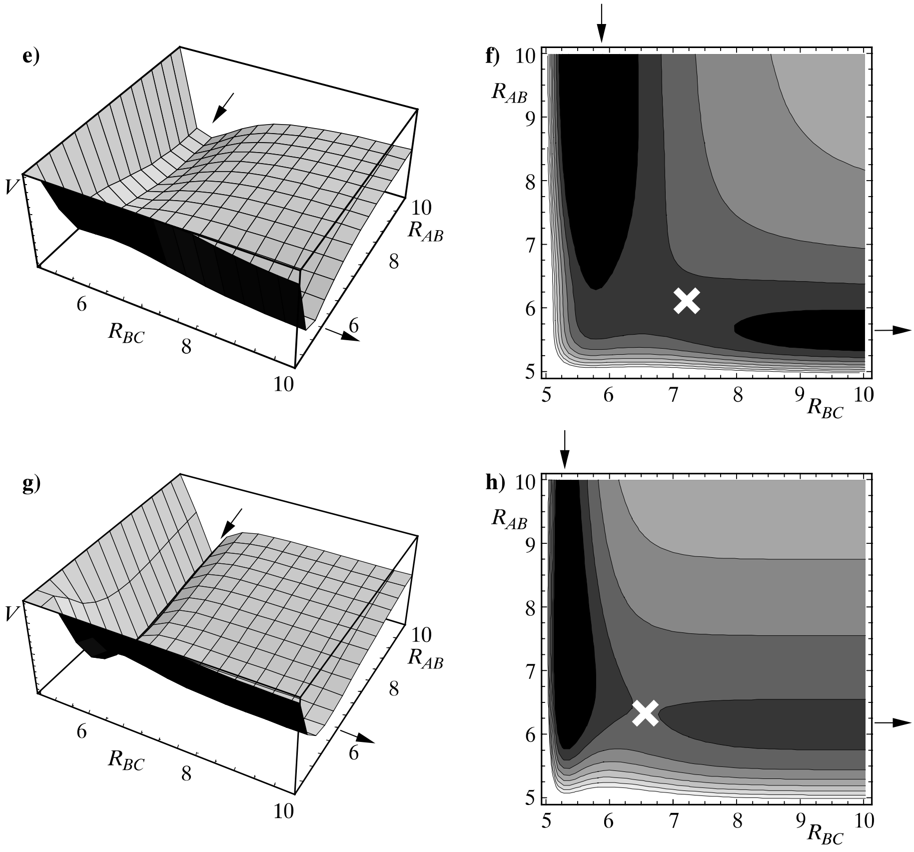


Fig. 14.1. *Continued.*

energy of BC, e.g., for the reaction $H + H_2 \rightarrow H_2 + H$ the activation energy (to overcome the reaction barrier) amounts only to about 10% of the hydrogen molecule binding energy. Simply, when the BC bond breaks, a new bond AB forms at the same time compensating for the energy cost needed to break the BC bond.

The barrier may have different positions in the reaction "drain-pipe", e.g., it may be in the entrance channel (*early barrier*), Fig. 14.1.c,d, or in the exit channel (*late barrier*), Fig. 14.1.e,f, or, it may be inbetween (symmetric case, Fig. 14.1.a,b). The barrier position influences the course of the reaction.

early or late barrier

When determining the SDP, kinetic energy was neglected, i.e. the motion of the point representing the system resembles a "crawling". A chemical reaction does not, however, represent any crawling over the energy hypersurface, but rather a dynamics that begins in the entrance channel and ends in the exit channel, including motion "uphill" against the potential energy V. Overcoming the barrier thus is possible only, when the system has an excess of kinetic energy.

What will happen, if we have an early barrier? A possible reactive trajectory for such a case is shown in Fig. 14.2.a.

bobsleigh effect

It is seen that the most effective way to pass the barrier is to set the point (representing the system) in fast motion along the entrance channel. This means that atom A has to have lots of kinetic energy when attacking the molecule BC. After passing the barrier the point slides downhill, entering the exit channel. Since, after sliding down, it has large kinetic energy, a *bobsleigh effect* takes place, i.e. the point climbs up the potential wall (as a result of the repulsion of atoms A and B) and then moves by making zigzags similar to a bobsleigh team. This zigzag means, of course, that strong oscillations of AB take place (and the C atom leaves the rest of the system). Thus,

> early location of a reaction barrier may result in a vibrationally excited product.

A different thing happens when the barrier is *late*. A possible reactive (i.e. successful) trajectory is shown in Fig. 14.2.b. For the point to overcome the barrier it has to have a large momentum along the BC axis, because otherwise it would climb up the potential energy wall in vain as the energy cost is too large. This may happen if the point moves along a zigzag-like way *in the entrance channel* (as shown in Fig. 14.2.b). This means that

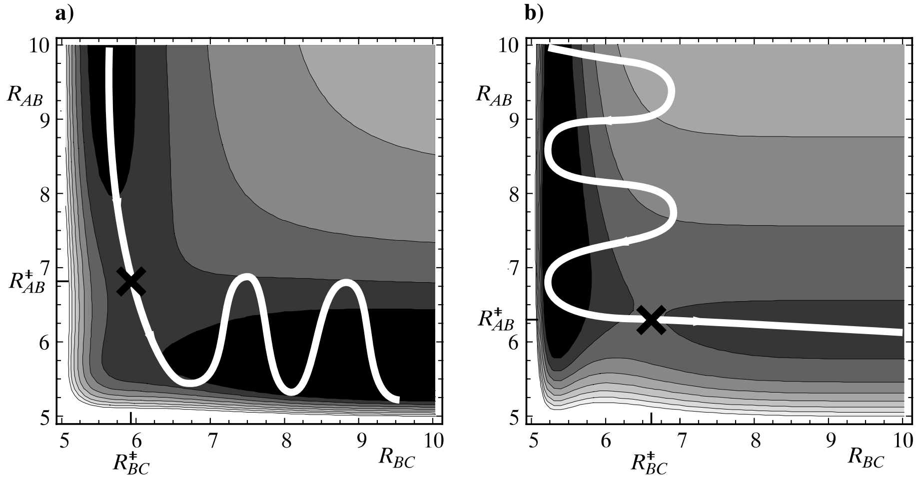


Fig. 14.2. A potential energy map for the collinear reaction $A + BC \rightarrow AB + C$ as a function of R_{AB} and R_{BC}. The distances $R^{\#}_{AB}$ and $R^{\#}_{BC}$ determine the saddle point position. Fig. (a) shows a reactive trajectory. If the point that represents the system runs sufficiently fast along the entrance channel towards the barrier, it will overcome the barrier by a "*charge ahead*". Then, in the exit channel the point has to oscillate, which means product vibrations. Fig. (b) shows a reaction with a late barrier. In the entrance channel a promising reactive trajectory is shown as the wavy line. This means the system oscillates in the entrance channel in order to be able to attack the barrier directly after passing the corner area (bobsleigh effect).

to overcome a late barrier, the *vibrational excitation* of the reactant BC is effective,

because an increase in the kinetic energy of A will not produce much. Of course, the conditions for the reaction to occur matter less for high collision energies of the reactants. On the other hand, a too fast a collision may lead to unwanted reactions occurring, e.g., dissociation of the system into $A+B+C$. Thus there is an energy window for any given reaction.

AB INITIO APPROACH

14.2 ACCURATE SOLUTIONS FOR THE REACTION HYPERSURFACE (THREE ATOMS[17])

14.2.1 COORDINATE SYSTEM AND HAMILTONIAN

This approach to the chemical reaction problem corresponds to point 2 on p. 770.

Jacobi coordinates

For three atoms of masses M_1, M_2, M_3, with total mass $M = M_1 + M_2 + M_3$ we may introduce the Jacobi coordinates (see p. 279) in three different ways (Fig. 14.3.a).

Each of the coordinate systems (let us label them $k = 1, 2, 3$) highlights two atoms "close" to each other (i, j) and a third "distant" (k). Now, let us choose a pair of vectors $\boldsymbol{r}_k, \boldsymbol{R}_k$ for each of the choices of the Jacobi coordinates by the following procedure ($\boldsymbol{X}_i$ represents the vector identifying nucleus i in a space-fixed coordinate system, SFCS, cf. Appendix I). First, let us define $\boldsymbol{r}_k$:

$$\boldsymbol{r}_k = \frac{1}{d_k}(\boldsymbol{X}_j - \boldsymbol{X}_i), \tag{14.3}$$

where the square of the *mass scaling parameter* equals

mass scaling parameter

$$d_k^2 = \left(1 - \frac{M_k}{M}\right)\frac{M_k}{\mu}, \tag{14.4}$$

while μ represents the *reduced mass* (for three masses)

reduced mass

$$\mu = \sqrt{\frac{M_1 M_2 M_3}{M}}. \tag{14.5}$$

Now the second vector needed for the Jacobi coordinates is chosen as

$$\boldsymbol{R}_k = d_k\left[\boldsymbol{X}_k - \frac{M_i\boldsymbol{X}_i + M_j\boldsymbol{X}_j}{M_i + M_j}\right]. \tag{14.6}$$

---

[17] The method was generalized for an arbitrary number of atoms [D. Blume, C.H. Greene, "*Monte Carlo Hyperspherical Description of Helium Cluster Excited States*", 2000].

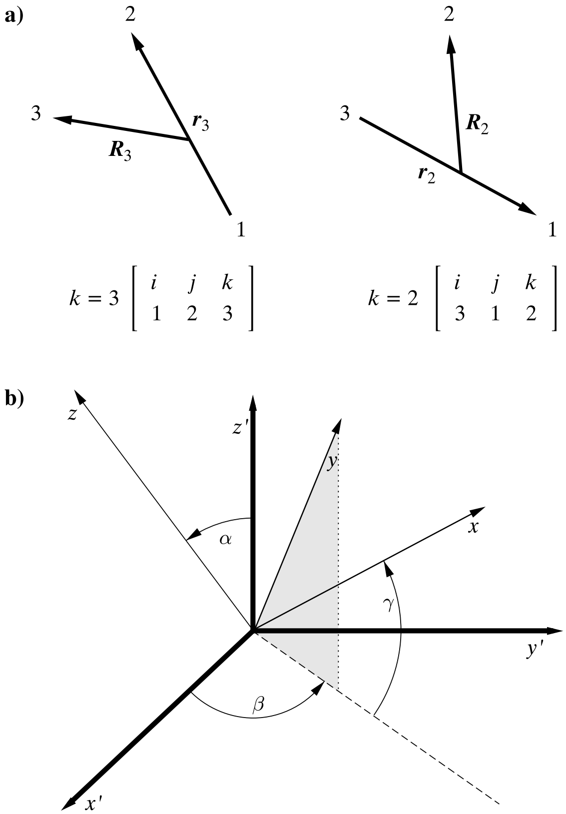


Fig. 14.3. (a) The three equivalent Jacobi coordinate systems. (b) The Euler angles show the mutual orientation of the two Cartesian coordinate systems. First, we project the y axis on the x', y' plane (the result is the dashed line). The first angle α is the angle between axes z' and z, the two other (β and γ) use the projection line described above. The relations among the coordinates are given by H. Eyring, J. Walter, G.E. Kimball, "*Quantum Chemistry*", John Wiley, New York, 1967.

The three Jacobi coordinate systems are related by the following formulae (cf. Fig. 14.3):

$$\begin{pmatrix} \boldsymbol{r}_i \\ \boldsymbol{R}_i \end{pmatrix} = \begin{pmatrix} \cos\beta_{ij} & \sin\beta_{ij} \\ -\sin\beta_{ij} & \cos\beta_{ij} \end{pmatrix} \begin{pmatrix} \boldsymbol{r}_j \\ \boldsymbol{R}_j \end{pmatrix}, \tag{14.7}$$

$$\tan\beta_{ij} = -\frac{M_k}{\mu}, \tag{14.8}$$

$$\beta_{ij} = -\beta_{ji}.$$

The Jacobi coordinates will now be used to define what is called the (more convenient) hyperspherical democratic coordinates.

Democratic hyperspherical coordinates

When a chemical reaction proceeds, the role of the atoms changes and using the same Jacobi coordinate system all the time leads to technical problems. In order not to favour any of the three atoms despite possible differences in their masses, we introduce *democratic hyperspherical coordinates*.

democratic hyperspherical coordinates

First, let us define the axis z of a Cartesian coordinate system, which is perpendicular to the molecular plane at the centre of mass, i.e. parallel to $\boldsymbol{A} = \frac{1}{2}\boldsymbol{r} \times \boldsymbol{R}$, where $\boldsymbol{r}$ and $\boldsymbol{R}$ are *any* (just democracy, the result is the same) of the vectors $\boldsymbol{r}_k, \boldsymbol{R}_k$. Note that by definition $|\boldsymbol{A}|$ represents the area of the triangle built of the atoms. Now, let us construct the axes x and y of the rotating with molecule coordinate system (RMCS, cf. p. 245) in the plane of the molecule taking care that:

- the Cartesian coordinate system is right-handed,
- the axes are oriented along the main axes of the moments of inertia,[18] with $I_{yy} = \mu(r_y^2 + R_y^2) \geqslant I_{xx} = \mu(r_x^2 + R_x^2)$.

Finally, we introduce democratic hyperspherical coordinates equivalent to RMCS:

- the first coordinate measures the *size of the system*, or its “radius”:

$$\rho = \sqrt{R_k^2 + r_k^2}, \tag{14.9}$$

where ρ has no subscript, because the result is independent of k (to check this use eq. (14.7)),

- the second coordinate describes the system’s *shape*:

$$\cos\theta = \frac{2|\boldsymbol{A}|}{\rho^2} \equiv u. \tag{14.10}$$

Since $|\boldsymbol{A}|$ is the area of the triangle, $2|\boldsymbol{A}|$ means, therefore, the area of the corresponding parallelogram. The last area (in the nominator) is compared to the area of a square with side ρ (in the denominator; if u is small, the system is elongated like an ellipse with three atoms on its circumference)

- the third coordinate represents the angle ϕ_k for any of the atoms (in this way we determine, where the k-th atom is on the ellipse)

$$\cos\phi_k = \frac{2(\boldsymbol{R}_k \cdot \boldsymbol{r}_k)}{\rho^2 \sin\theta} \equiv \cos\phi. \tag{14.11}$$

As chosen, the hyperspherical democratic coordinates (which cover all possible atomic positions within the plane $z = 0$) have the following ranges: $0 \leqslant \rho < \infty$, $0 \leqslant \theta \leqslant \frac{\pi}{2}$, $0 \leqslant \phi \leqslant 4\pi$.

Hamiltonian in these coordinates

The hyperspherical democratic coordinates represent a useful alternative for RMCS from Appendix I (they themselves form another RMCS), and therefore

---

[18] These directions are determined by diagonalization of the inertia moment matrix (cf. Appendix K).

do not depend on the orientation with respect to the body-fixed coordinate system (BFCS). However, the molecule has somehow to "be informed" that it rotates (preserving the length and the direction of the total angular momentum), because a centrifugal force acts on its parts and the Hamiltonian expressed in BFCS (cf. Appendix I) has to contain information about this rotation.

The exact kinetic energy expression for a polyatomic molecule in a space fixed coordinate system (SFCS, cf. Appendix I) has been derived in Chapter 6 (eq. (6.34)). After separation of the centre-of-mass motion, the Hamiltonian is equal to $\hat{H} = \hat{T} + V$, where V represents the electronic energy playing the role of the potential energy for the motion of the nuclei (an analogue of $E_0^0(R)$ from eq. (6.8), we assume the Born–Oppenheimer approximation). In the democratic hyperspherical coordinates we obtain[19]

$$\hat{H} = -\frac{\hbar^2}{2\mu\rho^5}\frac{\partial}{\partial\rho}\rho^5\frac{\partial}{\partial\rho} + \hat{\mathcal{H}} + \hat{\mathcal{C}} + V(\rho,\theta,\phi), \tag{14.12}$$

with

$$\hat{\mathcal{H}} = \frac{\hbar^2}{2\mu\rho^2}\left[-\frac{4}{u}\frac{\partial}{\partial u}u\left(1-u^2\right)\frac{\partial}{\partial u} - \frac{1}{1-u^2}\left(4\frac{\partial^2}{\partial\phi^2} - \hat{J}_z^2\right)\right], \tag{14.13}$$

$$\hat{\mathcal{C}} = \frac{\hbar^2}{2\mu\rho^2}\left[\frac{1}{1-u^2}4i\hat{J}_z u\frac{\partial}{\partial\phi} + \frac{2}{u^2}\left[\hat{J}_x^2 + \hat{J}_y^2 + \sqrt{1-u^2}\left(\hat{J}_x^2 - \hat{J}_y^2\right)\right]\right], \tag{14.14}$$

where the first part, and the term with $\frac{\partial^2}{\partial\phi^2}$ in $\hat{\mathcal{H}}$, represent what are called deformation terms, the term with $\hat{J}_z^2$ and the terms in $\hat{\mathcal{C}}$ describe the rotation of the system.

14.2.2 SOLUTION TO THE SCHRÖDINGER EQUATION

Soon we will need some basis functions that depend on the angles θ and ϕ, preferentially each of them somehow adapted to the problem we are solving. These basis functions will be generated as the eigenfunctions of $\hat{\mathcal{H}}$ obtained at a fixed value $\rho = \rho_p$:

$$\hat{\mathcal{H}}(\rho_p)\Phi_{k\Omega}(\theta,\phi;\rho_p) = \varepsilon_{k\Omega}(\rho_p)\Phi_{k\Omega}(\theta,\phi;\rho_p), \tag{14.15}$$

where, because of two variables θ, ϕ we have two quantum numbers k and Ω (numbering the solutions of the equations).

The total wave function that also takes into account rotational degrees of freedom (θ,ϕ) is constructed as (the quantum number $J = 0, 1, 2, \ldots$ determines the length of the angular momentum of the system, while the quantum number $M = -J, -J+1, \ldots, 0, \ldots, J$ gives the z component of the angular momentum)

---

[19] J.G. Frey, B.J. Howard, *Chem. Phys.* 99 (1985) 415.

a linear combination of the basis functions $U_{k\Omega} = D^{JM}_{\Omega}(\alpha, \beta, \gamma)\Phi_{k\Omega}(\theta, \phi; \rho_p)$:

$$\psi^{JM} = \rho^{-\frac{5}{2}} \sum_{k\Omega} F^J_{k\Omega}(\rho; \rho_p) U_{k\Omega}(\alpha, \beta, \gamma, \theta, \phi; \rho_p), \tag{14.16}$$

where α, β, γ are the three Euler angles (Fig. 14.3.b) that define the orientation of the molecule with respect to the distant stars, $D^{JM}_{\Omega}(\alpha, \beta, \gamma)$ represent the eigenfunctions of the symmetric top,[20] $\Phi_{k\Omega}$ are the solutions to eq. (14.15), while $F^J_{k\Omega}(\rho; \rho_p)$ stand for the ρ-dependent expansion coefficients, i.e. functions of ρ (centred at point ρ_p). Thanks to $D^{JM}_{\Omega}(\alpha, \beta, \gamma)$ the function ψ^{JM} is the eigenfunction of the operators $\hat{J}^2$ and $\hat{J}_z$.

close coupling method

In what is known as the *close coupling method* the function from eq. (14.16) is inserted into the Schrödinger equation $\hat{H}\psi^{JM} = E_J\psi^{JM}$. Then, the resulting equation is multiplied by a function $U_{k'\Omega'} = D^{JM}_{\Omega'}(\alpha, \beta, \gamma)\Phi_{k'\Omega'}(\theta, \phi; \rho_p)$ and integrated over angles $\alpha, \beta, \gamma, \theta, \phi$, which means taking into account all possible orientations of the molecule in space (α, β, γ) and all possible shapes of the molecule (θ, ϕ) which are allowed for a given size ρ. We obtain a set of linear equations for the unknowns $F^J_{k\Omega}(\rho; \rho_p)$:

$$\rho^{-\frac{5}{2}} \sum_{k\Omega} F^J_{k\Omega}(\rho; \rho_p) \langle U_{k'\Omega'} | (\hat{H} - E_J) U_{k\Omega} \rangle_{\omega} = 0. \tag{14.17}$$

The summation extends over some assumed set of k, Ω (the number of k, Ω pairs is equal to the number of equations). The symbol $\omega \equiv (\alpha, \beta, \gamma, \theta, \phi)$ means integration over the angles. The system of equations is solved numerically.

state-to-state reaction

If, when solving the equations, we apply the boundary conditions suitable for a discrete spectrum (vanishing for $\rho = \infty$), we obtain the stationary states of the three-atomic molecule. We are interested in chemical reactions, in which one of the atoms comes to a diatomic molecule, and after a while another atom flies out leaving (after reaction) the remaining diatomic molecule. Therefore, we have to apply suitable boundary conditions. As a matter of fact we are not interested in details of the collision, we are positively interested in what comes to our detector from the spot where the reaction takes place. What may happen at a certain energy E to a given reactant state (i.e. what the product state is; such a reaction is called "*state-to-state*") is determined by the corresponding *cross section*[21] $\sigma(E)$. The cross section can be calculated from what is called the $\boldsymbol{S}$ matrix, whose elements are constructed from the coefficients $F^J_{k\Omega}(\rho; \rho_p)$ found from eqs. (14.17). The $\boldsymbol{S}$ matrix plays a role of an energy dependent dispatcher: such a reactant state changes to such a product state with such and such probability.

cross section

reaction rate constant

We calculate the *reaction rate* k assuming all possible energies E of the system (satisfying the Boltzmann distribution) and taking into account that fast products

---

[20] D.M. Brink, G.R. Satchler, "*Angular Momentum*", Clarendon Press, Oxford, 1975.

[21] After summing up the experimental results over all the angles, this is ready to be compared with the result of the above mentioned integration over angles.

arrive more often at the detector when counting per unit time

$$k = \text{const} \int \mathrm{d}E\, E\sigma(E) \exp\left(-\frac{E}{k_B T}\right), \tag{14.18}$$

where k_B is the Boltzmann constant.

The calculated reaction rate constant k may be compared with the result of the corresponding "state-to-state" experiment.

14.2.3 BERRY PHASE

When considering accurate quantum dynamics calculations (point 3 on p. 770) we encounter the problem of what is called Berry phase.

In Chapter 6 wave function (6.19) corresponding to the adiabatic approximation was assumed. In this approximation the electronic wave function depends parametrically on the positions of the nuclei. Let us imagine we take one (or more) of the nuclei on an excursion. We set off, go slowly (in order to allow the electrons to adjust), the wave function deforms, and then, we are back home and put the nucleus exactly in place. Did the wave function come back exactly too? Not necessarily. By definition (cf. Chapter 2) a class Q function has to be a unique function of coordinates. This, however, does not pertain to a parameter. What certainly came back is the probability density $\psi_k(\boldsymbol{r};\boldsymbol{R})^*\psi_k(\boldsymbol{r};\boldsymbol{R})$, because it decides that we cannot distinguish the starting and the final situations. *The wave function itself might undergo a phase change, i.e. the starting function is equal to $\psi_k(\boldsymbol{r};\boldsymbol{R}_0)$, while the final function is $\psi_k(\boldsymbol{r};\boldsymbol{R}_0)\exp(i\phi)$ and $\phi \neq 0$.* This phase shift is called the Berry phase.[22] Did it happen or not? Sometimes we can tell.

Let us consider a quantum dynamics description of a chemical reaction according to point 3 from p. 770. For example, let us imagine a molecule BC fixed in space, with atom B directed to us. Now, atom A, represented by a wave packet, rushes towards atom B. We may imagine that the atom A approaches the molecule and makes a bond with the atom B (atom C leaves the diatomic molecule) or atom A may first approach atom C, then turn back and make a bond with atom B (as before). The two possibilities correspond to two waves, which finally meet and interfere. If the phases of the two waves differed, we would see this in the results of the interference. The scientific community was surprised that some details of the reaction $H + H_2 \rightarrow H_2 + H$ at higher energies are impossible to explain without taking the Berry phase[23] into account. One of the waves described above made a turn around the conical intersection point (because it had to by-pass the equilateral triangle configuration, cf. Chapter 6). As it was shown in the work of Longuet-Higgins *et al.* mentioned above, this is precisely the reason why the function acquires a phase shift. We have shown in Chapter 6 (p. 264) that such a trip

---

[22] The discoverers of this effect were H.C. Longuet-Higgins, U. Öpik, M.H.L. Pryce and R.A. Sack, *Proc. Roy. Soc. London, A* 244 (1958) 1. The problem of this geometric phase diffused into the consciousness of physicists much later after an article by M.V. Berry, *Proc. Roy. Soc. London A*392 (1984) 45.

[23] Y.-S.M. Wu, A. Kupperman, *Chem. Phys. Letters* 201 (1993) 178.

around a conical intersection point results in changing the phase of the function by π.

The phase appears, when the system makes a "trip" in configurational space. We may make the problem of the Berry phase more familiar by taking an example from everyday life. Let us take a 3D space. Please put your arm down against your body with the thumb directed forward. During the operations described below, please do not move the thumb with respect to the arm. Now stretch your arm horizontally sideways, rotate it to your front and then put down along your body. Note that now your thumb is not directed towards your front anymore, but towards your body. When your arm has come back, the thumb had made a rotation of 90°.

Your thumb corresponds to $\psi_k(\boldsymbol{r};\boldsymbol{R})$, i.e. a vector in the Hilbert space, which is coupled with a slowly varying neighbourhood ($\boldsymbol{R}$ corresponds to the hand positions). When the neighbourhood returns, the vector may have been rotated in the Hilbert space [i.e. multiplied by a phase $\exp(i\phi)$].

APPROXIMATE METHODS

14.3 INTRINSIC REACTION COORDINATE (IRC) OR STATICS

This section addresses point 4 of our plan from p. 770.

On p. 770 two reaction coordinates were proposed: DRC and SDP. Use of the first of them may lead to some serious difficulties (like energy discontinuities). The second reaction coordinate will undergo in a moment a useful modification and will be replaced by the so called *intrinsic reaction coordinate* (IRC).

What the IRC is?

Let us use the Cartesian coordinate system once more with $3N$ coordinates for the N nuclei: X_i, $i=1,\dots,3N$, where X_1, X_2, X_3 denote the x, y, z coordinates of atom 1 of mass M_1, etc. The i-th coordinate is therefore associated with mass M_i of the corresponding atom. The classical Newtonian equation of motion for an atom of mass M_i and coordinate X_i is:[24]

$$M_i\ddot{X}_i = -\frac{\partial V}{\partial X_i} \quad \text{for } i=1,\dots,3N. \tag{14.19}$$

Let us introduce what are called *mass-weighted coordinates* (or, more precisely, weighted by the square root of mass)

mass-weighted coordinates

$$x_i = \sqrt{M_i}X_i. \tag{14.20}$$

In such a case we have

$$\sqrt{M_i}\sqrt{M_i}\ddot{X}_i = -\frac{\partial V}{\partial x_i}\frac{\partial x_i}{\partial X_i} = \sqrt{M_i}\left(-\frac{\partial V}{\partial x_i}\right) \tag{14.21}$$

---

[24] Mass × acceleration equals force; a dot over the symbol means a time derivative.

or

$$\ddot{x}_i = -\frac{\partial V}{\partial x_i} \equiv -g_i, \tag{14.22}$$

where g_i stands for the i-th component of the gradient of potential energy V calculated in mass-weighted coordinates. This equation can easily be integrated and we obtain

$$\dot{x}_i = -g_i t + v_{0,i} \tag{14.23}$$

or, for a small time increment $\mathrm{d}t$ and initial speed $v_{0,i} = 0$ (for the definition of the IRC as a path characteristic for potential energy V we want to neglect the influence of the kinetic energy) we obtain

$$\frac{\mathrm{d}x_i}{-g_i} = t\,\mathrm{d}t = \textit{independent} \text{ of } i. \tag{14.24}$$

Thus,

> in the coordinates weighted by the square roots of the masses, a displacement of atom number i is proportional to the potential gradient (and does not depend on the atom mass).

If mass-weighted coordinates were not introduced, a displacement of the point representing the system on the potential energy map *would not follow the direction of the negative gradient or the steepest descent* (on a geographic map such a motion would look natural, because slow rivers flow this way). Indeed, the formula analogous to (14.24) would have the form: $\frac{\mathrm{d}X_i}{-G_i} = \frac{t}{M_i}\,\mathrm{d}t$, and therefore, during a single watch tick $\mathrm{d}t$, light atoms would travel long distances while heavy atoms short distances.

Thus, after introducing mass-weighted coordinates, we may forget about masses, in particular about the atomic and the total mass, or equivalently, we may treat these as unit masses. The atomic displacements in this space will be measured in units of $\sqrt{\text{mass}} \times$ length, usually in: $\sqrt{u}a_0$, where $12u = {}^{12}\text{C}$ atomic mass, $u = 1822.887m$ (m is the electron mass), and sometimes also in units of $\sqrt{u}$ Å.

trajectory-in-molasses

Eq. (14.24) takes into account our assumption about the zero initial speed of the atom in any of the integration steps (also called "*trajectory-in-molasses*"), because otherwise we would have an additional term in $\mathrm{d}x_i$: the initial velocity times time. Broadly speaking, when the watch ticks,

> the system, represented by a point in $3N$-dimensional space, crawls over the potential energy hypersurface along the negative gradient of the hypersurface (in mass weighted coordinates). When the system starts from a saddle point of the first order, a small deviation of the position makes the system slide down on one or the other side of the saddle. The trajectory of the nuclei during such a motion is called the *intrinsic reaction coordinate* or IRC.

The point that represents the system slides down with infinitesimal speed along the IRC.

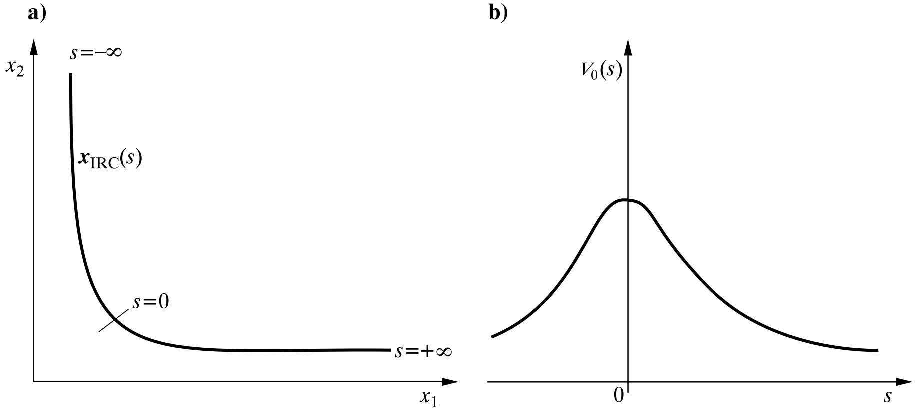


Fig. 14.4. A schematic representation of the IRC: (a) curve $\boldsymbol{x}_{\mathrm{IRC}}(s)$ and (b) energy profile when moving along the IRC [i.e. curve $V_0(\boldsymbol{x}_{\mathrm{IRC}}(s))$] in the case of two mass-weighted coordinates x_1, x_2.

Measuring the travel along the IRC

In the space of the mass-weighted coordinates, trajectory IRC represents a certain curve $\boldsymbol{x}_{\mathrm{IRC}}$ that depends on a parameter s: $\boldsymbol{x}_{\mathrm{IRC}}(s)$.

The parameter s measures the length along the reaction path IRC

(e.g., in $\sqrt{u}a_0$ or $\sqrt{u}$ Å). Let us take two close points on the IRC and construct the vector: $\boldsymbol{\xi}(s) = \boldsymbol{x}_{\mathrm{IRC}}(s + \mathrm{d}s) - \boldsymbol{x}_{\mathrm{IRC}}(s)$, then

$$(\mathrm{d}s)^2 = \sum_i \left[\xi_i(s)\right]^2. \tag{14.25}$$

We assume that $s = 0$ corresponds to the saddle point, $s = -\infty$ to the reactants, and $s = \infty$ to the products (Fig. 14.4).

For each point on the IRC, i.e. on the curve $\boldsymbol{x}_{\mathrm{IRC}}(s)$ we may read the mass-weighted coordinates, and use them to calculate the coordinates of each atom. Therefore, each point on the IRC corresponds to a certain structure of the system.

14.4 REACTION PATH HAMILTONIAN METHOD

14.4.1 ENERGY CLOSE TO IRC

A hypersurface of the potential energy represents an expensive product. We have first to calculate the potential energy for a grid of points. If we assume that ten points per coordinate is a sufficient number, then we have to perform 10^{3N-6}

advanced quantum mechanical calculations, for $N = 10$ atoms this gives $\ldots 10^{24}$ calculations, which is an unreasonable task. Now you see why specialists so much prefer three-atomic systems.

Are all the points necessary? For example, if we assume low energies, the system will in practice, stay close to the IRC. Why, therefore, worry about other points? This idea was exploited by Miller, Handy and Adams.[25] They decided to introduce the coordinates that are natural for the problem of motion in the reaction "drain-pipe". The approach corresponds to point 4 from p. 770.

The authors derived the

REACTION PATH HAMILTONIAN:
an approximate expression for the energy of the reacting system in the form, that stresses the existence of the IRC and of deviations from it.

This formula (*Hamilton function of the reaction path*) has the following form:

$$H\big(s, p_s, \{Q_k, P_k\}\big) = T\big(s, p_s, \{Q_k, P_k\}\big) + V\big(s, \{Q_k\}\big), \tag{14.26}$$

where T is the kinetic energy, V stands for the potential energy, s denotes the reaction coordinate along the IRC, $p_s = \frac{ds}{dt}$ represents the momentum coupled with s (mass = 1), $\{Q_k\}$, $k = 1, 2, \ldots, 3N - 7$, stand for other coordinates orthogonal to the reaction path $\boldsymbol{x}_{\mathrm{IRC}}(s)$ (this is why Q_k will depend on s) and the momenta $\{P_k\}$ conjugated with them.

We obtain the coordinates Q_k in the following way. At point s on the reaction path we diagonalize the Hessian, i.e. the matrix of the second derivatives of the potential energy and consider all the resulting normal modes ($\omega_k(s)$ are the corresponding frequencies; cf. Chapter 7) *other* than that, which corresponds to the reaction coordinate s (the later corresponds to the "imaginary"[26] frequency ω_k). The diagonalization also gives the normal vectors $\boldsymbol{L}_k(s)$, each having a direction in the $(3N - 6)$-dimensional configurational space (the mass-weighted coordinate system). *The coordinate $Q_k \in (-\infty, +\infty)$ measures the displacement along the direction of $\boldsymbol{L}_k(s)$*. The coordinates s and $\{Q_k\}$ are called the *natural coordinates*. To stress that Q_k is related to $\boldsymbol{L}_k(s)$, we will write it as $Q_k(s)$.

natural coordinates

The potential energy, close to the IRC, can be approximated (*harmonic approximation*) by

$$V\big(s, \{Q_k\}\big) \cong V_0(s) + \frac{1}{2} \sum_{k=1}^{3N-7} \omega_k(s)^2 Q_k(s)^2, \tag{14.27}$$

where the term $V_0(s)$ represents the potential energy that corresponds to the bottom of the reaction "drain-pipe" at a point s along the IRC, while the second term tells us what will happen to the potential energy if we displace the point (i.e. the

---

[25] W.H. Miller, N.C. Handy, J.E. Adams, *J. Chem. Phys.* 72 (1980) 99.

[26] For large $|s|$ the corresponding ω^2 is close to zero. When $|s|$ decreases (we approach the saddle point), ω^2 becomes negative (i.e. ω is imaginary). For simplicity we will call this the "imaginary frequency" for any s.

system) perpendicular to $\boldsymbol{x}_{\mathrm{IRC}}(s)$ along all the normal oscillator coordinates. In the *harmonic approximation* for the oscillator k, the energy goes up by half the force constant × the square of the normal coordinate Q_k^2. The force constant is equal to ω_k^2, because the mass is equal to 1.

The kinetic energy turns out to be more complicated

$$T\big(s, p_s, \{Q_k, P_k\}\big) = \frac{1}{2}\frac{\big[p_s - \sum_{k=1}^{3N-7}\sum_{k'=1}^{3N-7} B_{kk'}Q_{k'}P_k\big]^2}{\big[1+\sum_{k=1}^{3N-7} B_{ks}Q_k\big]^2} + \sum_{k=1}^{3N-7}\frac{P_k^2}{2}. \qquad (14.28)$$

The last term is recognized as the vibrational kinetic energy for the oscillations perpendicular to the reaction path (recall that the mass is treated as equal to 1). If in the first term we insert $B_{kk'} = 0$ and $B_{ks} = 0$, the term would be equal to $\frac{1}{2}p_s^2$ and, therefore, would describe the kinetic energy of a point moving as if the reaction coordinate were a straight line.

CORIOLIS AND CURVATURE COUPLINGS:
$B_{kk'}$ are called the *Coriolis coupling constants*. They couple the normal modes perpendicular to the IRC.
The B_{ks} are called the *curvature coupling constants*, because they would be equal zero if the IRC was a straight line. They couple the translational motion along the reaction coordinate with the vibrational modes orthogonal to it. All the above coupling constants B depend on s.

Coriolis coupling constant

curvature coupling constant

Therefore, in the reaction path Hamiltonian we have the following quantities that characterize the reaction "drain-pipe":

- The reaction coordinate s that measures the progress of the reaction along the "drain-pipe".
- The value $V_0(s) \equiv V_0(\boldsymbol{x}_{\mathrm{IRC}}(s))$ represents the energy that corresponds to the bottom of the "drain-pipe"[27] at the reaction coordinate s.
- The width of the "drain-pipe" is characterized by $\{\omega_k(s)\}$.[28]
- The curvature of the "drain-pipe" is hidden in constants B, their definition will be given later in this chapter. Coefficient $B_{kk'}(s)$ tells us how normal modes k and k' are coupled together, while $B_{ks}(s)$ is responsible for a similar coupling between reaction path $\boldsymbol{x}_{\mathrm{IRC}}(s)$ and vibration k perpendicular to it.

14.4.2 VIBRATIONALLY ADIABATIC APPROXIMATION

Most often when moving along the bottom of the "drain-pipe", potential energy $V_0(s)$ only changes moderately when compared to the potential energy changes

---

[27] I.e. the classical potential energy corresponding to the point of the IRC given by s (this gives an idea of how the potential energy changes when walking along the IRC).

[28] A small ω corresponds to a wide valley, when measured along a given normal mode coordinate ("soft" vibration), a large ω means a narrow valley ("hard" vibration).

the molecule undergoes when oscillating perpendicularly[29] to $\boldsymbol{x}_{\mathrm{IRC}}(s)$. Simply, the valley bottom profile results from the fact that the molecule hardly holds together when moving along the reaction coordinate s, *a chemical bond breaks*, while *other bonds remain strong* and it is not so easy to stretch their strings. This suggests that there is *slow* motion along s and *fast* oscillatory motion along the coordinates Q_k.

Since we are mostly interested in the slow motion along s, we may average over the fast motion.

The philosophy behind the idea is that while the system moves slowly along s, it undergoes a large number of oscillations along Q_k. After such vibrational averaging the only information that remains about the oscillations are the vibrational quantum levels for each of the oscillators (the levels will depend on s).

VIBRATIONALLY ADIABATIC APPROXIMATION:
The fast vibrational motions will be treated quantum mechanically and their total energy will enter the potential energy for the classical motion along s.

This approximation parallels the adiabatic approximation made in Chapter 6, where the fast motion of electrons was separated from the slow motion of the nuclei. There the total electronic energy became the potential energy for the motion of nuclei, here the total vibrational energy (the energy of the corresponding harmonic oscillators in their quantum states) becomes the potential energy for the slow motion along s. This concept is called the *vibrationally adiabatic approximation*.

In this approximation, to determine the stage of the reaction we give two *classical* quantities: *where* the system is on the reaction path (s), and *how fast* the system moves along the reaction path (p_s). Also we need the *quantum states* of the oscillators vibrating perpendicularly to the reaction path (vibrational quantum number $v_k = 0, 1, 2, \dots$ for each of the oscillators). Therefore, the potential energy for the (slow) motion along the reaction coordinate s is:[30]

$$\begin{aligned} &V_{\mathrm{adiab}}(s; v_1, v_2, \dots, v_{3N-7}) \\ &\quad = V_0(s) + \sum_{k=1}^{3N-7} \left(v_k + \frac{1}{2}\right)\hbar\big[\omega_k(s) - \omega_k(-\infty)\big], \end{aligned} \tag{14.29}$$

where we have chosen an additive constant in the potential as equal to the vibrational energy of the reactants (with minus sign): $-\sum_{k=1}^{3N-7}(v_k + \frac{1}{2})\hbar\omega_k(-\infty)$. Note that even though $v_k = 0$ for each of the oscillators, there is a non-zero vibrational correction to the classical potential energy $V_0(s)$, because the zero-vibrational energy changes if s changes.

---

[29] I.e. when moving along the coordinates Q_k, $k = 1, 2, \dots, 3N - 7$.

[30] Even if (according to the vibrationally adiabatic approximation) the vibrational quantum numbers were kept constant during the reaction, their energies as depending on s through ω would change.

The vibrationally adiabatic potential V_{adiab} was created for a given set of vibrational quantum numbers v_k, fixed during the reaction process. Therefore, it is impossible to exchange energy between the vibrational modes (we assume therefore the Coriolis coupling constants $B_{kk'} = 0$), as well as between the vibrational modes and the reaction path (we assume the curvature coupling constants $B_{ks} = 0$). This would mean a change of v_k's.

From eq. (14.29) we may draw the following conclusion.

When during the reaction the frequency of a normal mode decreases dramatically (which corresponds to breaking of a chemical bond), the square bracket becomes negative. This means that an excitation of the bond before the reaction decreases the (vibrationally adiabatic) reaction barrier and the reaction rate will increase.

As a matter of fact, this is a quite obvious: a vibrational excitation that engages the chemical bond to be broken already weakens the bond before the reaction.

Why do chemical reactions proceed?

Exothermic reactions. When the reactants (products) have a kinetic energy higher than the barrier and the corresponding momentum p_s is large enough, with a high probability the barrier will be overcome (cf. p. 155). Even if the energy is lower than the barrier there is still a non-zero probability of passing to the other side because of the tunnelling effect. In both cases (passing over and under the barrier) it is easier when the kinetic energy is large.

The barrier *height* is usually different for the reaction reactants → products and for the products → reactants (Fig. 14.1). If the barrier height is smaller for the reactants, this *may* result in an excess of the product concentration over the reactant concentration.[31] Since the reactants have higher energy than the products, the potential energy excess will change into the kinetic energy[32] of the products (which is observed as a temperature increase – the reaction is exothermic). This may happen if the system has the possibility to pump the potential (i.e. electronic) energy excess into the translational and rotational degrees of freedom or to a "third body or bodies" (through collisions, e.g., with the solvent molecules) or has the possibility to emit light quanta. *If the system has no such possibilities the reaction will not take place*.

Endothermic reactions. The barrier height does not always decide.

Besides the barrier height the *widths* of the entrance and exit channels also count.

---

[31] Because a lower barrier is easier to overcome.

[32] Most often rotational energy.

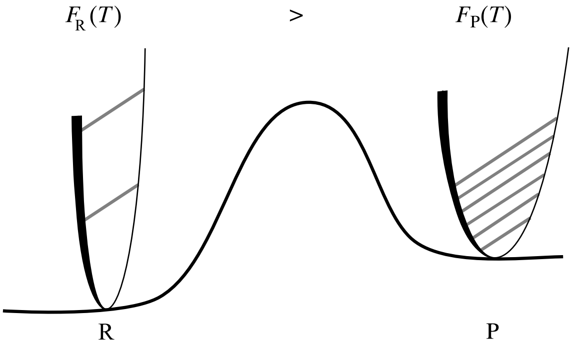


Fig. 14.5. Why do some endothermic reactions proceed spontaneously? The figure shows the energy profile as a function of s, i.e. along the intrinsic reaction coordinate (IRC). As we can see, the reactants have lower energy than the products. Yet it is not V that decides the reaction to proceed, but the free energy $F = E - TS$, where T is the temperature and S the entropy. The free energy depends on the density of the vibrational states of the reactants and products. The more numerous the low-energy vibrational levels the larger the entropy and the lower the free energy, if $T > 0$. As we can see, the reactant vibrational levels are scarce, while on the product side they are densely distributed. When the energy gain related to the entropy overcomes the potential energy loss, then the (endothermic) reaction will proceed spontaneously.

For the time being let us take an example with $V_0(-\infty) = V_0(\infty)$, i.e. the barrier calculated from IRC is *the same* in both directions. Imagine a narrow entrance channel, i.e. large force constants for the oscillators, and a wide exit channel, i.e. low force constants.

The vibrational energies in the entrance channel are high, while in the exit channel they are low. This results in $V_{\text{adiab}}(-\infty; v_1, v_2, \ldots, v_{3N-7}) > V_{\text{adiab}}(\infty; v_1, v_2, \ldots, v_{3N-7})$, i.e. the barrier for the reaction reactants → products is low, while for the reaction products → reactants it is high. The products will form more often than the reactants.

On top of that if the entrance channel is narrow, while the exit channel is wide, the density of the vibrational states will be small in the entrance channel and large in the exit channel (Figs. 14.1.g,h and 14.5). Therefore, for $T > 0$ there will be a lot of possibilities to occupy the low-energy vibrational levels for the products, while only a few possibilities for the reactants. This means a high entropy of the products and a small entropy of the reactants, i.e. the products will be more stabilized by the entropy than the reactants.[33] Once again we can see, that

> while the energy in the endothermic reaction *increases*, the decisive factor is the free energy which *decreases*. The reactants → products reaction occurs "uphill" in potential energy, but "downhill" in free energy.

Kinetic and thermodynamic pictures

- In a macroscopic reaction carried out in a chemist's flask we have a statistical ensemble of the systems that are in different microscopic stages of the reaction.

---

[33] It pertains to the term $-TS$ in the free energy.

- The ensemble may be modelled by a reaction "drain-pipe" (we assume the barrier) with a lot of points, each representing one of the reacting systems.
- When the macroscopic reaction begins (e.g., we mix two reactants) a huge number of points appear in the entrance channel, i.e. we have the reactants only. As the reactant molecules assemble or dissociate the points appear or disappear.
- If the barrier were high (no tunnelling) and temperature $T = 0$, all the reactants would be in their zero-vibrational states[34] and in their ground rotational states. This state would be stable even when the product valley corresponded to a lower energy (this would be a metastable state).
- Raising the temperature causes some of the ensemble elements in the entrance channel to acquire energy comparable to the barrier height. Those elements might have a chance to pass the barrier either by tunnelling (for energy smaller than the barrier) or by going over the barrier. Not all of the elements with sufficient energies would pass the barrier, only those with reactive trajectories.
- After passing the barrier the energy is conserved, but changes into the translational, vibrational and rotational energy of products or may be transferred to a "third body" (e.g., the solvent molecules) or changed into electromagnetic radiation.
- The probability of the reactive trajectories might be calculated in a way similar to that described in Chapter 4 (tunnelling[35]), with additional taking into account the initial vibrational and rotational states of the reactants as well as averaging over the energy level populations.
- The products would have also a chance to pass the barrier back to the reactant side, but at the beginning the number of the elements passing the barrier in the reactant-to-product direction would be larger (non-equilibrium state).
- However the higher the product concentration, the more often the products transform into the reactants. As an outcome we arrive at the *thermodynamic equilibrium state*, in which the average numbers of the elements passing the barrier per unit time in either direction are equal.
- If the barrier is high and the energies considered low, then the stationary states of the system could be divided into those (of energy $E_{i,R}$), which have high amplitudes in the entrance channel (the reactant states) and those (of energy $E_{i,P}$) with high amplitudes in the exit channel (product states). In such a case we may calculate the partition function for the reactants:

  $$Z_R(T) = \sum_i g_i \exp\left(-\frac{E_{i,R} - E_{0,R}}{k_B T}\right)$$

  and for the products

  $$Z_P(T) = \sum_i g_i \exp\left(-\frac{E_{i,P} - E_{0,R}}{k_B T}\right) = \sum_i g_i \exp\left(-\frac{E_{i,P} - E_{0,P} - \Delta E}{k_B T}\right),$$

---

[34] Please note, that even in this case ($T = 0$) the energy of these states would not only depend on the bottom of the valley, $V_0(s)$, but also on the valley's width through $\omega_k(s)$, according to eq. (14.29).

[35] See also H. Eyring, J. Walter, G.E. Kimball, "*Quantum Chemistry*", John Wiley, New York, 1967.

where g_i stands for the degeneracy of the i-th energy level and the difference of the ground-state levels is $\Delta E = E_{0,R} - E_{0,P}$.

- Having the partition functions, we may calculate (at a given temperature, volume and a fixed number of particles[36]) the free or Helmholtz energy (F) corresponding to the entrance and to the exit channels (in thermodynamic equilibrium)

$$F_R(T) = -k_B T \frac{\partial}{\partial T} \ln Z_R(T), \tag{14.30}$$

$$F_P(T) = -k_B T \frac{\partial}{\partial T} \ln Z_P(T). \tag{14.31}$$

- The reaction goes in such a direction as to attain the minimum of free energy F.
- The higher the density of states in a given channel (this corresponds to higher entropy) the lower F. The density of the vibrational states is higher for wider channels (see Fig. 14.5).

14.4.3 VIBRATIONALLY NON-ADIABATIC MODEL

Coriolis coupling

The vibrationally adiabatic approximation is hardly justified, because the reaction channel is curved. This means that motion along s couples with some vibrational modes, and also the vibrational modes couple among themselves. We have therefore to use the non-adiabatic theory and this means we need coupling coefficients B. The Miller–Handy–Adams reaction path Hamiltonian theory gives the following expression for the $B_{kk'}$:

$$B_{kk'}(s) = \frac{\partial \boldsymbol{L}_k(s)}{\partial s} \cdot \boldsymbol{L}_{k'}(s), \tag{14.32}$$

where $\boldsymbol{L}_k$, $k = 1, 2, \ldots, 3N - 7$, represent the orthonormal eigenvectors ($3N$-dimensional, cf. Chapter 7, p. 297) of the normal modes Q_k of frequency ω_k (Fig. 14.6).

If the derivative in the above formula is multiplied by an increment of the reaction path Δs, we obtain $\frac{\partial \boldsymbol{L}_k(s)}{\partial s}\Delta s$ which represents a change of normal mode vector $\boldsymbol{L}_k$ when the system moved along the IRC by the increment Δs. This change might be similar to normal mode eigenvector $\boldsymbol{L}_{k'}$. This means that $B_{kk'}$ measures how much eigenvector $\boldsymbol{L}_{k'}(s)$ resembles the *change* of eigenvector $\boldsymbol{L}_k(s)$ (when the system moves along the reaction path).[37] Coupling coefficient $B_{kk'}$ is usually

---

[36] Similar considerations may be performed for a constant pressure (instead of volume). The quantity that then attains the minimum at the equilibrium state is the Gibbs potential G.

[37] From differentiating the orthonormality condition $\boldsymbol{L}_k(s) \cdot \boldsymbol{L}_{k'}(s) = \delta_{kk'}$ we obtain

$$\begin{aligned}\frac{\partial}{\partial s}\left[\boldsymbol{L}_k(s) \cdot \boldsymbol{L}_{k'}(s)\right] &= \left[\frac{\partial \boldsymbol{L}_k(s)}{\partial s} \cdot \boldsymbol{L}_{k'}(s) + \boldsymbol{L}_k(s) \cdot \frac{\partial \boldsymbol{L}_{k'}(s)}{\partial s}\right] \\ &= \left[\frac{\partial \boldsymbol{L}_k(s)}{\partial s} \cdot \boldsymbol{L}_{k'}(s) + \frac{\partial \boldsymbol{L}_{k'}(s)}{\partial s} \cdot \boldsymbol{L}_k(s)\right] \\ &= B_{kk'} + B_{k'k} = 0.\end{aligned}$$

Hence, $B_{kk'} = -B_{k'k}$.

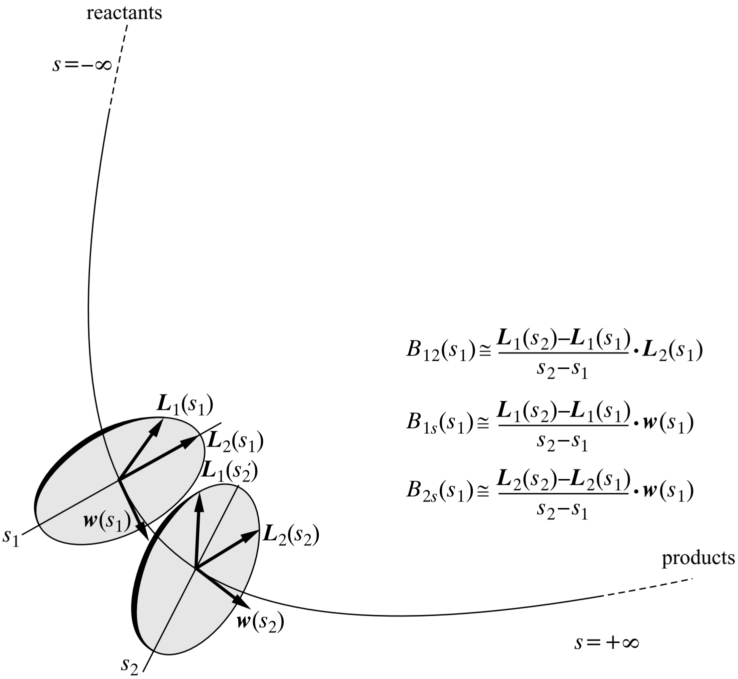


Fig. 14.6. Calculation of the Coriolis coupling coefficient (B_{12}) and the curvature coefficients (B_{1s} and B_{2s}) related to the normal modes 1 and 2 and reaction coordinate s. Diagonalization of the two Hessians calculated at points $s = s_1$ and $s = s_2$ gives two corresponding normal mode eigenvectors $\boldsymbol{L}_1(s_1)$ and $\boldsymbol{L}_2(s_1)$ as well as $\boldsymbol{L}_1(s_2)$ and $\boldsymbol{L}_2(s_2)$. At both points s_1 and s_2 we also calculate the versors $\boldsymbol{w}(s_1)$ and $\boldsymbol{w}(s_1)$ that are tangent to the IRC. The calculated vectors inserted into the formulae give the approximations to B_{1s}, B_{2s} and B_{12}.

especially large close to those values of s, for which $\omega_k \cong \omega_{k'}$, i.e. for the crossing points of the vibrational frequency (or energy) curves $\omega_k(s)$. These are the points where we may expect an important energy flow from one normal mode to another, because the energy quanta match ($\hbar\omega_k(s) \cong \hbar\omega_{k'}(s)$). Coriolis coupling means that the directions of $\boldsymbol{L}_k$ and $\boldsymbol{L}_{k'}$ change, when the reaction proceeds and this resembles a rotation in the configurational space about the IRC.

Curvature couplings

Curvature coupling constant B_{ks} links the motion along the reaction valley with the normal modes orthogonal to the IRC (Fig. 14.6):

$$B_{ks}(s) = \frac{\partial \boldsymbol{L}_k(s)}{\partial s} \cdot \boldsymbol{w}(s), \tag{14.33}$$

where $\boldsymbol{w}(s)$ represents the unit vector tangent to the intrinsic reaction path $\boldsymbol{x}_{\mathrm{IRC}}(s)$ at point s. Coefficient $B_{ks}(s)$ therefore represents a measure of how the *change* in the normal mode eigenvector $\boldsymbol{L}_k(s)$ resembles a motion along the IRC. Large

$B_{ks}(s)$ makes energy flow from the normal mode to the reaction path (or *vice versa*) much easier.

DONATING MODES:
The modes with large $B_{ks}(s)$ in the *entrance* channel are called the *donating modes*, because an excitation of such modes makes possible an energy transfer to the reaction coordinate degree of freedom (an increase of the kinetic energy along the reaction path). This will make the reaction rate increase.

donating modes

In the vibrationally adiabatic approximation, coefficients B_{ks} are equal to zero. This means that in such an approximation an exothermic reaction would transform the net reaction energy (defined as the difference between the energy of the reactants and the products) into the kinetic energy of translational motion of products, because the energy of the system in the entrance channel could not be changed into the vibrational energy of the products (including the "vibrations" of a rotational character). However, as was shown by John Polanyi and Dudley Herschbach, the reactions do not go this way – a majority of the reaction energy goes into the rotational degrees of freedom (excited states of some modes). The rotations are hidden in the vibrations at $s = 0$ which are similar to the internal rotations and in the limit of $s \to +\infty$ transform into product rotations. Next, the excited products emit infrared quanta in the process of infrared fluorescence (the chemist's test tube gets hot). This means that in order to have a realistic description of reaction we have to abandon the vibrationally adiabatic approximation.

14.4.4 APPLICATION OF THE REACTION PATH HAMILTONIAN METHOD TO THE REACTION $H_2 + OH \to H_2O + H$

The reaction represents one of a few polyatomic systems for which precise calculations were performed.[38] It may be instructive to see how a practical implementation of the reaction path Hamiltonian method looks.

Potential energy hypersurface

The *ab initio* configuration interaction calculations (Chapter 10) of the potential energy hypersurface for the system under study were performed by Walsh and Dunning[39] within the Born–Oppenheimer ("clamped nuclei") approximation described in Chapter 6. The electronic energy obtained as a function of the nuclear configuration plays the role of the potential energy for the motion of the nuclei. The calculation gave the electronic energy for a relatively scarce set of configu-

---

[38] G.C.J. Schatz, *J. Chem. Phys.* 74 (1981) 113; D.G. Truhlar, A.D. Isaacson, *J. Chem. Phys.* 77 (1982) 3516; A.D. Isaacson, D.G. Truhlar, *J. Chem. Phys.* 76 (1982) 380 and above all the paper by Thom Dunning Jr. and Elfi Kraka in "*Advances in Molecular Electronic Structure Theory: The Calculation and Characterization of Molecular Potential Energy Surfaces*", ed. T.H. Dunning, Jr., JAI Press, Inc., Greenwich, CN (1989) 1.

[39] S.P. Walsh, T.H. Dunning, Jr., *J. Chem. Phys.* 72 (1980) 1303.

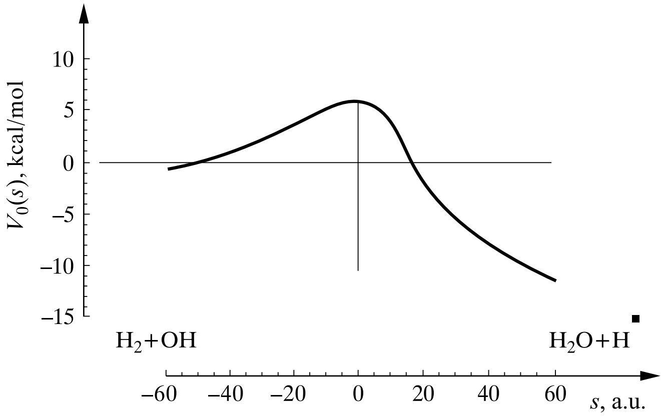


Fig. 14.7. The reaction $H_2 + OH \rightarrow H_2O + H$ energy profile $V_0(s)$ for $-\infty \leqslant s \leqslant \infty$. The value of the reaction coordinate $s = -\infty$ corresponds to the reactants, while $s = \infty$ corresponds to the products. It turns out that the product energy is lower than the energy of the reactants (i.e. the reaction is exothermic). The barrier height in the entrance channel calculated as the difference of the top of the barrier and the lowest point of the entrance channel amounts to 6.2 kcal/mol. According to T. Dunning, Jr. and E. Kraka, from "*Advances in Molecular Electronic Structure Theory*", ed. T. Dunning, Jr., JAI Press, Greenwich, CN (1989), courtesy of the authors.

rations of the nuclei, but then the numerical results were fitted by an analytical function.[40] The IRC energy profile is shown in Fig. 14.7.

It is seen from Fig. 14.7 that the barrier height for the reactants is equal to about 6.2 kcal/mol, while the reaction energy calculated as the difference of the products minus the energy of the reactants is equal to about −15.2 kcal/mol (an exothermic reaction). What happens to the atoms when the system moves along the reaction path? This is shown in Fig. 14.8.

The saddle point configuration of the nuclei when compared to those corresponding to the reactants and to products tells us whether the barrier is early or late. The difference of the OH distances for the saddle point and for the product (H_2O) amounts to 0.26 Å, which represents $\frac{0.26}{0.97} = 27\%$, while the HH distance difference for the saddle point and of the reactant (H_2) is equal to 0.11 Å, which corresponds to $\frac{0.11}{0.74} = 15\%$. In conclusion, the saddle point resembles the reactants more than the products, i.e. the barrier is *early*.

Normal mode analysis

Let us see what the normal mode analysis gives when performed for some selected points along the IRC. The calculated frequencies are shown in Fig. 14.9 as wave numbers $\bar{\nu} = \omega/(2\pi c)$.

As we can see, before the reaction takes place we have two normal mode frequencies ω_{HH} and ω_{OH}. When the two reacting subsystems approach one another

---

[40] G.C. Schatz, H. Elgersma, *Chem. Phys. Letters* 73 (1980) 21.

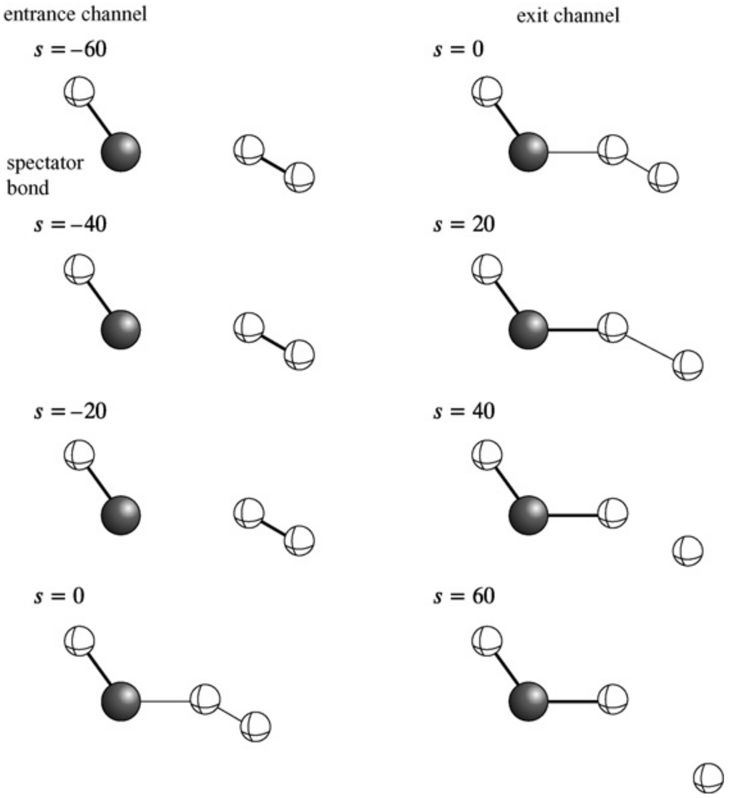


Fig. 14.8. The optimum atomic positions in the reacting system $H_2 + OH \rightarrow H_2O + H$ as functions of the reaction coordinate s. According to T. Dunning, Jr. and E. Kraka, from "*Advances in Molecular Electronic Structure Theory*", ed. T. Dunning, Jr., JAI Press, Greenwich, CN (1989), courtesy of the authors.

we have to treat them as an entity. The corresponding number of vibrations is $3N - 6 = 3 \times 4 - 6 = 6$ normal modes. Two of them have frequencies close to those of HH and OH, three others have frequencies close to zero and correspond to the vibrational and rotational motions of the loosely bound reactants,[41] the last "vibrational mode" is connected with a motion along the reaction path and has an imaginary frequency. Such a frequency means that the corresponding curvature of the potential energy is negative.[42] For example, at the saddle point, when moving along the reaction path, we have a potential energy maximum instead of minimum as would be for a regular oscillator. Fig. 14.9 shows five (real) frequencies. The frequency ω_{HH} drops down close to the saddle point. This is precisely the bond to be broken. Interestingly, the frequency minimum is

---

[41] van der Waals interactions, see Chapter 13.

[42] Note that $\omega = \sqrt{k/m}$, where the force constant k stands for the second derivative of the potential energy, i.e. its curvature.

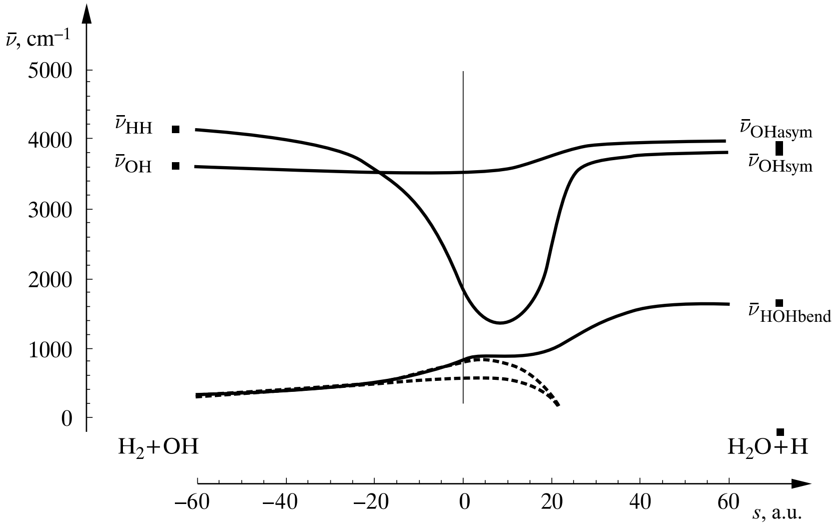


Fig. 14.9. The reaction $H_2 + OH \rightarrow H_2O + H$. The vibrational frequencies (in wave numbers $\bar{\nu} = \omega/(2\pi c)$.) for the normal modes along the coordinate s. Only the real wave numbers are given (the "vibration" along s is imaginary and not given). According to T. Dunning, Jr. and E. Kraka, from "*Advances in Molecular Electronic Structure Theory*", ed. T. Dunning, Jr., JAI Press, Greenwich, CN (1989), courtesy of the authors.

attained at 9 a.u. beyond the saddle point. Afterwards the frequency increases fast and when the reaction is completed it turns out to be the OH symmetric stretching frequency of the water molecule. Knowing only this, we can tell what has happened: the HH bond was broken and a new OH bond was formed. At the end of the reaction path we have, in addition, the antisymmetric stretching mode of the H_2O ($\bar{\nu}_{OHasym}$), which evolved from the starting value of ω_{OH} while changing only a little (this reflects that one OH bond exists all the time and in fact represents a "spectator" to the reaction) as well as the HOH bending mode ($\bar{\nu}_{HOHbend}$), which appeared as a result of the strengthening of an intermolecular interaction in H_2 + OH when the reaction proceeded. The calculations have shown that this vibration corresponds to the *symmetric* stretching mode[43] of the H_2O ($\bar{\nu}_{OHsym}$). The two other modes at the beginning of the reaction have almost negligible frequencies, and after an occasional increasing of their frequencies near the saddle point end up with zero frequencies for large s. Of course, at the end we have to have $3 \times 3 - 6 = 3$ vibrational modes of the H_2O and so we do.

spectator bond

---

[43] At first sight this looks like contradicting chemical intuition since the antisymmetric mode is apparently compatible to the reaction path (one hydrogen atom being far away while the other is close to the oxygen atom). However, everything is all right. The SCF LCAO MO calculations for the water molecule within a medium size basis set give the OH bond length equal to 0.95 Å, whereas the OH radical bond length is equal to 1.01 Å. This means that when the hydrogen atom approaches the OH radical (making the water molecule), the hydrogen atom of the radical has to get *closer* to the oxygen atom. The resulting motion of both hydrogen atoms is similar to the symmetric (not antisymmetric) mode.

Example 1. *Vibrationally adiabatic approximation*

Let us consider several versions of the reaction that differ by assuming various vibrational states of the reactants.[44] Using eq. (14.29), for each set of the vibrational quantum numbers we obtain the vibrationally adiabatic potential V_{adiab} as a function of s (Fig. 14.10).

The adiabatic potentials obtained are instructive. It turns out that:

- The adiabatic potential corresponding to the vibrational ground state $(v_{\mathrm{OH}}, v_{\mathrm{HH}}) = (0, 0)$ gives lower barrier height than the classical potential $V_0(s)$ (5.9 kcal/mol vs 6.1). The reason for this is the lower zero-vibration energy for the saddle point configuration than for the reactants.[45]
- The adiabatic potential for the vibrational ground state has its maximum at $s = -5$ a.u., not at the saddle point $s = 0$.
- Excitation of the OH stretching vibration does not significantly change the energy profile, in particular the barrier is lowered by only about 0.3 kcal/mol. Thus, the OH is definitely a spectator bond.
- This contrasts with what happens when the H_2 molecule is excited. In such a case the barrier is lowered by as much as about 3 kcal/mol. This suggests that the HH stretching vibration is a "donating mode". donating mode

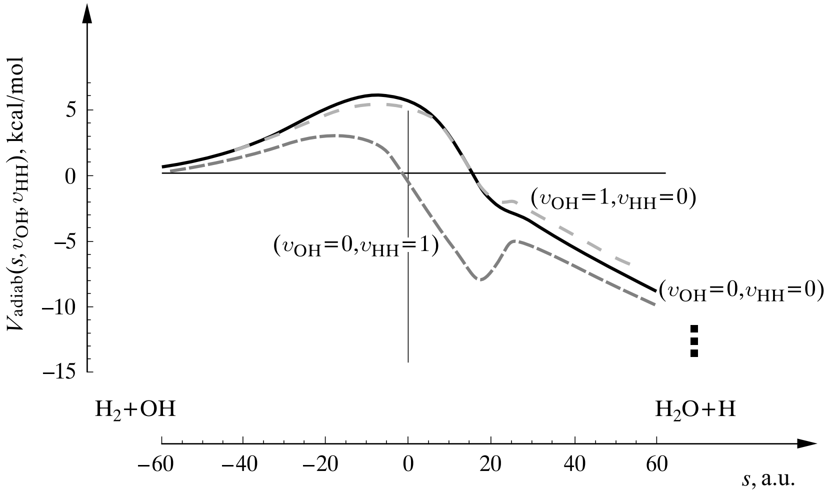


Fig. 14.10. The reaction $H_2 + OH \rightarrow H_2O + H$ (within the vibrationally adiabatic approximation). Three sets of the vibrational numbers $(v_{\mathrm{OH}}, v_{\mathrm{HH}}) = (0,0), (1,0), (0,1)$ were chosen. Note, that the height and position of the barrier depend on the vibrational quantum numbers assumed. An excitation of H_2 considerably decreases the barrier height. The small squares on the right show the limiting values. According to T. Dunning, Jr. and E. Kraka, from "*Advances in Molecular Electronic Structure Theory*", ed. T. Dunning, Jr., JAI Press, Greenwich, CN (1989), courtesy of the authors.

---

[44] We need the frequencies of the modes which are orthogonal to the reaction path.

[45] This stands to reason, because when the Rubicon is crossed, all the bonds are weakened with respect to the reactants.

Example 2. *Non-adiabatic theory*

Now let us consider the vibrationally non-adiabatic procedure. To do this we have to include the coupling constants B. This is done in the following way. Moving along the reaction coordinate s we perform the normal mode analysis resulting in the vibrational eigenvectors $\boldsymbol{L}_k(s)$. This enables us to calculate how these vectors change and to determine the derivatives $\partial \boldsymbol{L}_k/\partial s$. Now we may calculate the corresponding dot products (see eqs. (14.32) and (14.33)) and obtain the coupling constants $B_{kk'}(s)$ and $B_{ks}(s)$ at each selected point s. A role of the coupling constants B in the reaction rate can be determined after dynamic studies assuming various starting conditions (the theory behind this approach will not be presented in this book). Yet some important information may be extracted just by inspecting functions $B(s)$. The functions $B_{ks}(s)$ are shown in Fig. 14.11.

As we can see:

- In the entrance channel the value of $B_{\mathrm{OH},s}$ is close to zero, therefore, there is practically no coupling between the OH stretching vibrations and the reaction path and hence there will be practically no energy flow between those degrees of freedom. This might be expected from a weak dependence of ω_{OH} as a function of s. Once more we see that the OH bond plays only the role of a reaction spectator.
- This is not the case for $B_{\mathrm{HH},s}$. This quantity attains maximum just before the saddle point (let us recall that the barrier is early). The energy may, therefore, flow from the vibrational mode of H_2 to the reaction path (and *vice versa*) and a

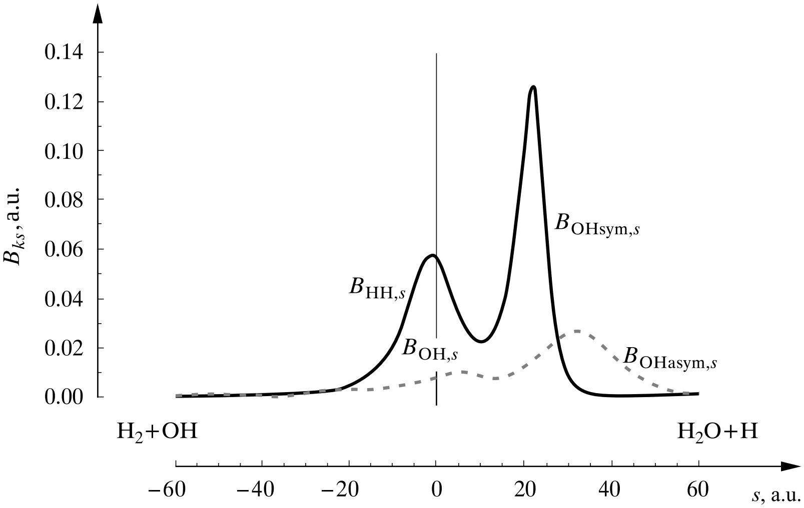


Fig. 14.11. The reaction $H_2 + OH \rightarrow H_2O + H$. The curvature coupling constants $B_{ks}(s)$ as functions of s. The $B_{ks}(s)$ characterize the coupling of the k-th normal mode with the reaction coordinate s. According to T. Dunning, Jr. and E. Kraka, from "*Advances in Molecular Electronic Structure Theory*", ed. T. Dunning, Jr., JAI Press, Greenwich, CN (1989), courtesy of the authors.

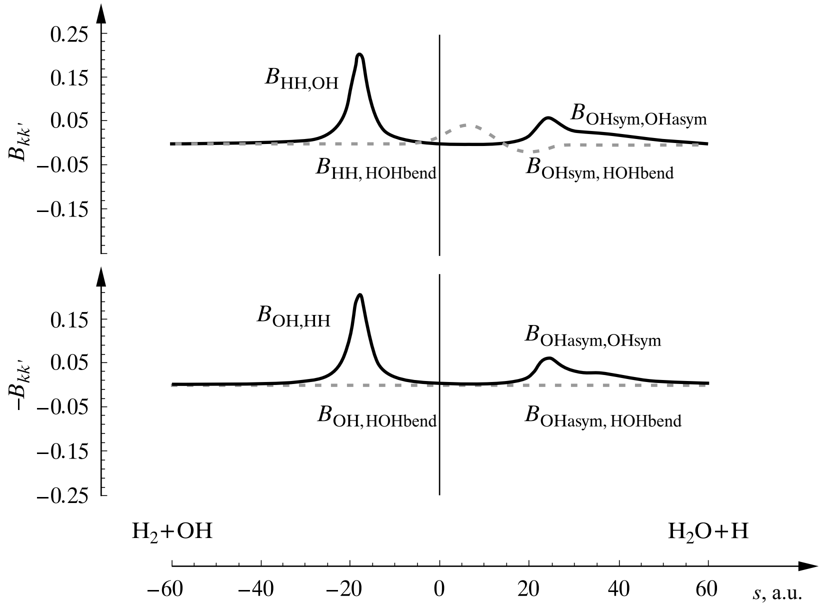


Fig. 14.12. The reaction $H_2 + OH \rightarrow H_2O + H$. The Coriolis coupling constants $B_{kk'}(s)$ as functions of s. A high value of $B_{kk'}(s)$ means that close to reaction coordinate s the changes of the k-th normal mode eigenvector resemble eigenvector k'. According to T. Dunning, Jr. and E. Kraka, from "*Advances in Molecular Electronic Structure Theory*", ed. T. Dunning, Jr., JAI Press, Greenwich, CN (1989), courtesy of the authors.

vibrational excitation of H_2 may have an important impact on the reaction rate (recall please the lowering of the adiabatic barrier when this mode is excited).

The Coriolis coupling constants $B_{kk'}$ as functions of s are plotted in Fig. 14.12 (only for the OH and HH stretching and HOH bending modes).

The first part of Fig. 14.12 pertains to the HH vibrational mode, the second to the OH vibrational mode. As we can see:

- the maximum coupling for the HH and OH modes occurs long before the saddle point (close to $s = -18$ a.u.) enabling the system to exchange energy between the two vibrational modes;
- in the exit channel we have quite significant couplings between the symmetric and antisymmetric OH modes and the HOH bending mode.

14.5 ACCEPTOR–DONOR (AD) THEORY OF CHEMICAL REACTIONS

14.5.1 MAPS OF THE MOLECULAR ELECTROSTATIC POTENTIAL

Chemical reaction dynamics is possible only for very simple systems. Chemists, however, have most often to do with medium-size or large molecules. Would it be possible to tell anything about the barriers for chemical reactions in such systems? Most of chemical reactions start from a situation when the molecules are far away, but already interact. The main contribution is the electrostatic interaction energy, which is of long-range character (Chapter 13). Electrostatic interaction depends strongly on the mutual orientation of the two molecules (*steric effect*). Therefore, the orientations are biased towards the privileged ones (energetically favourable). There is quite a lot of experimental data suggesting that privileged orientations lead, at smaller distances, to low reaction barriers. There is no guarantee of this, but it often happens for electrophilic and nucleophilic reactions, because the attacking molecule prefers those parts of the partner that correspond to high electron density (for electrophilic attack) or to low electron density (for nucleophilic attack).

steric effect

electrophilic attack

We may use an electrostatic probe (e.g., a unit positive charge) to detect, which parts of the molecule "like" the approaching charge (energy lowering), and which do not (energy increasing).

The electrostatic interaction energy of the point-like probe in position $\boldsymbol{r}$ with molecule A is described by the formula (the definition of the electrostatic potential produced by molecule A, see Fig. 14.13.a):

nucleophilic attack

$$V_A(\boldsymbol{r}) = +\sum_a \frac{Z_a}{|\boldsymbol{r}_a - \boldsymbol{r}|} - \int \frac{\rho_A(\boldsymbol{r}')}{|\boldsymbol{r}' - \boldsymbol{r}|}\, \mathrm{d}^3\boldsymbol{r}', \tag{14.34}$$

where the first term describes the interaction of the probe with the nuclei denoted by index a, and the second means the interaction of the probe with the electron density distribution of the molecule A denoted by ρ_A (according to Chapter 11).[46]

In the Hartree–Fock or Kohn–Sham approximation (Chapter 11, p. 570; we assume the n_i-tuple occupation of the molecular orbital $\varphi_{A,i}$, $n_i = 0, 1, 2$)

$$\rho_A(\boldsymbol{r}) = \sum_i n_i \left|\varphi_{A,i}(\boldsymbol{r})\right|^2. \tag{14.35}$$

In order to obtain $V_A(\boldsymbol{r})$ at point $\boldsymbol{r}$ it is sufficient to calculate the distances of the point from any of the nuclei (trivial) as well as the one-electron integrals, which appear after inserting into (14.34) $\rho_A(\boldsymbol{r}') = 2\sum_i |\varphi_{A,i}(\boldsymbol{r}')|^2$. Within the LCAO MO approximation the electron density distribution ρ_A represents the sum of products of two atomic orbitals (in general centred at two different points). As a result the task reduces to calculating typical one-electron three-centre integrals of the nuclear attraction type (cf. Chapter 8 and Appendix P), because the third centre corresponds to the point $\boldsymbol{r}$ (Fig. 14.13). There is no computational problem with this for contemporary quantum chemistry.

---

[46] By the way, to calculate the electrostatic interaction energy of the molecules A and B we have to take (instead of a probe) the nuclei of the molecule B and sum of the corresponding contributions, and then to do the same with the electronic cloud of B. This corresponds to the following formula: $E_{\text{elst}} = \sum_b Z_b V_A(\boldsymbol{r}_b) - \int \mathrm{d}^3\boldsymbol{r}\, \rho_B(\boldsymbol{r}) V_A(\boldsymbol{r})$, where b goes over the nuclei of B, and ρ_B represents its electronic cloud.

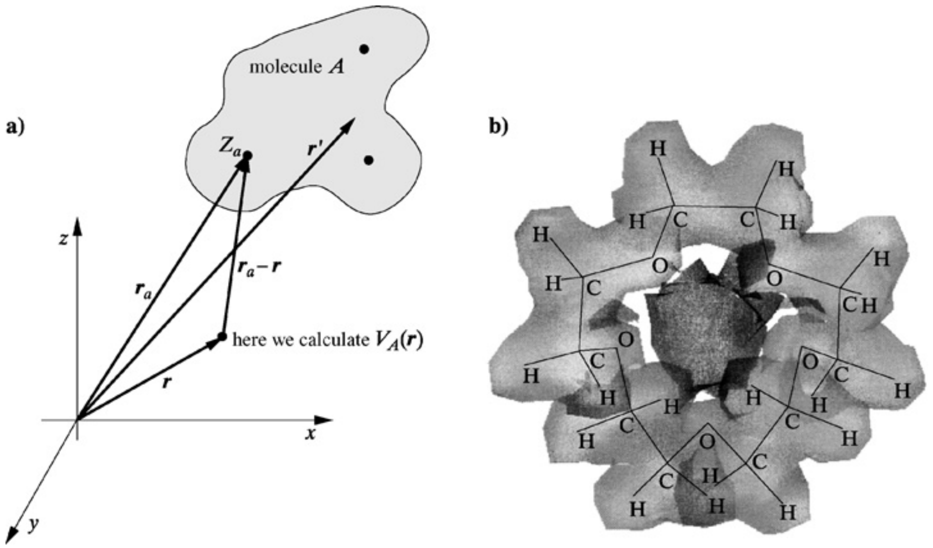


Fig. 14.13. Interaction of the positive unit charge (probe) with molecule A. Fig. (a) shows the coordinate system and the vectors used in eq. (14.34). Fig. (b) shows the equipotential surfaces $|V_A(\boldsymbol{r})|$ for the crown ether molecule. The light shadowed surface corresponds to $V_A(\boldsymbol{r}) > 0$, and the darker one to $V_A(\boldsymbol{r}) < 0$ (in more expensive books this is shown by using different colours). It is seen that the crown ether cavity corresponds to the negative potential, i.e. it would attract strongly cations.

In order to represent $V_A(\boldsymbol{r})$ graphically we usually choose to show an equipotential surface corresponding to a given absolute value of the potential, while additionally indicating its sign (Fig. 14.13.b). The sign tells us which parts of the molecule are preferred for the probe-like object to attack and which not. In this way we obtain basic information about the reactivity of different parts of the molecule.[47]

ESP charges

Who attacks whom?

In chemistry a probe will not be a point charge, but rather a neutral molecule or an ion. Nevertheless our new tool (electrostatic potential) will still be useful:

- If the probe represents a cation, it will attack those parts of the molecule A which are electron-rich (electrophilic reaction).
- If the probe represents an anion, it will attack the electron-deficient parts (nucleophilic reaction).
- If the probe represents a molecule (B), *its* electrostatic potential V_B is the most interesting. Those AB configurations that correspond to the contacts of the associated sections of V_A and V_B with the opposite signs are the most (electrostatically) stable.

The site of the molecular probe (B) which attacks an electron-rich site of A itself has to be of electron-deficient character (and *vice versa*). Therefore, from

---

[47] Having the potential calculated according to (14.34) we may ask about the set of atomic charges that reproduce it. Such charges are known as ESP (ElectroStatic Potential).

the point of view of the attacked molecule (A), everything looks "upside down": an electrophilic reaction becomes nucleophilic and *vice versa*. When two objects exhibit an affinity to each other, who attacks whom represents a delicate and ambiguous problem and let it be that way. Therefore where does such nomenclature in chemistry come from? Well, it comes from the concept of didactics.

The problem considered is related to the Umpolung problem from p. 703. A change of sign or an exchange of charges on the interacting molecules (both operations may have only a limited meaning in chemistry) should not influence the key features of some reaction mechanisms involving intermediate ionic species.

We may ask whether there is any difference between a reaction taking place in a vacuum and the same reaction proceeding in the electric field resulting from the neighbouring point charges. Why might this be of interest? Well, many important chemical reactions proceed in the presence of catalysts, e.g., in the active centre of enzymes. To proceed, many chemical reactions require a chemist to heat the flask to very high temperatures, while in enzymes the same reaction proceeds in mild conditions. For example, in order to synthesize ammonia from atmospheric nitrogen chemists use the hellish conditions of an electric arc. However a similar reaction takes place in lupin roots. The enzymes are proteins, which Nature took care to make of a self-assembling character with a nearly unique, final conformation,[48] assuring active centre formation. Only in this *native* conformation does the active centre work as a catalyst: the reactant is recognized (cf. p. 750), and then docked in the reaction cavity, a particular bond is broken, the products are released and the enzyme comes back to the initial state. Well, why does the bond break? The majority of amino acids in an enzyme play an important yet passive role: just to allow a few important amino acids to make the reaction cavity as well as the reaction site. The role of the reaction cavity is to assure the reactant is properly oriented in space with respect to those amino acids that form the reaction site. The role of the latter is to create a specific electric field at the reactant position. Such a field lowers the reaction barrier, thus making it easier. Andrzej Sokalski,[49] reversing the problem, asked a very simple question: for a given reaction *how should the field-producing charges look in order to lower (most-effectively) the reaction barrier*? This is what we will need first when planning artificial enzymes for the reactions desired.

native conformation

reaction cavity

As we have seen, for computational reasons we are able to take into account only reactions involving *a few* nuclei. However, chemists are interested in the reactivity of *much* larger molecules. We would like to know what particular features of the electronic structure make a chemical reaction proceed. This is the type of question that will be answered in the acceptor–donor theory.[50]

Molecules attract each other at long distances

Let us assume that two molecules approach each other, say, because of their chaotic thermic motion. When the molecules are still far away, they already un-

---

[48] Among myriads of other conformations, cf. Chapter 7.

[49] W.A. Sokalski, *J. Mol. Catalysis* 30 (1985) 395.

[50] More details about the acceptor–donor theory may be found in excellent paper by S. Shaik, *J. Am. Chem. Soc.* 103 (1981) 3692. The results reported in Tables 14.2–14.6 also originate from this paper.

dergo tiny changes due to the electric field produced by their partner (Chapters 12 and 13). The very fact that each of the molecules having permanent multipole moments is now immersed in a non-homogeneous electric field, means the multipole moments interact with the field and the resulting electrostatic interaction energy.

The electric field also distorts the partner's electronic cloud, and as a consequence of the Hellmann–Feynman theorem (Chapter 12) this creates a distortion of the nuclear framework. Thus, the multipole moments of a distorted molecule are a little changed ("induced moments") with respect to the permanent ones. If we take the induced moment interaction with the field into account, then apart from the electrostatic interaction energy we obtain the induction energy contribution.[51]

Besides this, each of the molecules feels electric field fluctuations coming from motion of the electrons in the partner and adjusts to that motion. This leads to the dispersion interaction (Chapter 13).

> Even at long intermolecular distances the dominating electrostatic interaction orients the molecules to make them attract each other. The induction and dispersion energies are always attractive. Therefore, in such a case all important energy contributions mean *attraction*.

They are already close...

What happens when the molecules are closer? Besides the effects just described a new one appears – the valence repulsion coming from the electron clouds overlap. Such an interaction vanishes exponentially with intermolecular distance. This is why we were able to neglect this interaction at longer distances.

steric factor

Two molecules, even simple ones, may undergo different reactions depending not only on their collision energy, but also on their mutual orientation with respect to one another (*steric factor)*. The steric factor often assures selectivity, since only molecules of a certain shape (that fit together) may get their active centres close enough (cf. Chapter 13).

van der Waals complex

Suppose that two molecules under consideration collide with a proper orientation in space.[52] What will happen next? As we will see later, it depends on the molecules involved. Very often at the beginning there will be an energetic barrier to overcome. When the barrier is too high compared with the collision energy, then a *van der Waals* complex is usually formed, otherwise the barrier is overcome and the chemical reaction occurs. On the other hand for some reactants no reaction barrier exists.

---

[51] If one or both molecules do not have any non-zero permanent multipole moments, their electrostatic interaction energy is zero. If at least one of them has a non-zero permanent moment (and the partner has electrons), then there is a non-zero induction energy contribution.

[52] A desired reaction to occur.

14.5.2 WHERE DOES THE BARRIER COME FROM?

The barrier always results from the intersection of diabatic potential energy hypersurfaces. We may think of diabatic states as preserving the electronic state (e.g., the system of chemical bonds): I and II, respectively.

Sometimes it is said that the barrier results from an avoided crossing (cf. Chapter 6) of two diabatic hypersurfaces that belong to the same irreducible representation of the symmetry group of the Hamiltonian (in short: "of the same symmetry"). This, however, cannot be taken literally, because, as we know from Chapter 6, the non-crossing rule is valid for diatomics only. The solution to this dilemma is the conical intersection described in Chapter 6 (cf. Fig. 6.15).[53] Instead of diabatic we have two adiabatic hypersurfaces ("upper" and "lower"[54]), each consisting of the diabatic part I and the diabatic part II. A thermic reaction takes place as a rule on the lower hypersurface and corresponds to crossing the border between I and II.

14.5.3 MO, AD AND VB FORMALISMS

Let us take an example of a simple substitution reaction:

$$\mathrm{H}:^- + \mathrm{H}-\mathrm{H} \rightarrow \mathrm{H}-\mathrm{H} + \mathrm{H}:^- \qquad (14.36)$$

and consider the acceptor–donor formalism (AD). The formalism may be treated as intermediate between the configuration interaction (CI) and the valence bond (VB) formalisms. Any of the three formalisms is equivalent to the two others, provided they differ only by a linear transformation of many-electron basis functions.

> In the CI formalism the Slater determinants are built of the molecular spinorbitals
> In the VB formalism the Slater determinants are built of the atomic spinorbitals
> In the AD formalism the Slater determinants are built of the acceptor and donor spinorbitals

MO picture → AD picture

Molecular orbitals for the total system φ_1, φ_2, φ_3 in a minimal basis set may be expressed (Fig. 14.14) using the molecular orbital of the donor (n, in our case the 1s atomic orbital of H^-) and the acceptor molecular orbitals (bonding χ and antibonding χ^*):

$$\begin{aligned} \varphi_1 &= a_1 n + b_1 \chi - c_1 \chi^*, \\ \varphi_2 &= a_2 n - b_2 \chi - c_2 \chi^*, \\ \varphi_3 &= -a_3 n + b_3 \chi - c_3 \chi^*, \end{aligned} \qquad (14.37)$$

---

[53] The term "conical" stems from a *linear* (or "conical-like") dependence of the two adiabatic energy hypersurfaces on the distance from the conical intersection point.

[54] On the energy scale.

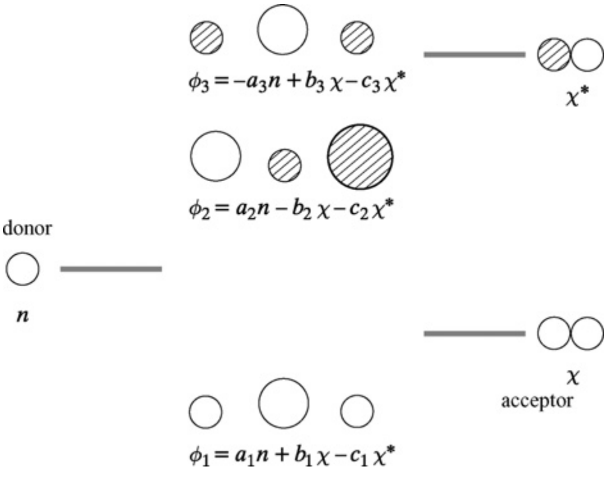


Fig. 14.14. A schematic representation of the molecular orbitals and their energies: of the donor (n representing the hydrogen atom 1s orbital), of the acceptor (bonding χ and antibonding χ^* of the hydrogen molecule) as well as of the total system H_3 in a linear configuration (centre of the figure). The lowest-energy molecular orbital of H_3 does not have any node, the higher has one while the highest has two nodes. In all cases we use the approximation that the molecular orbitals are built of the three 1s hydrogen atomic orbitals.

where $a_i, b_i, c_i > 0$, for $i = 1, 2, 3$. This convention comes from the fact that φ_1 is of the lowest energy and therefore exhibits no node, φ_2 has to play the role of the orbital second in energy scale and therefore has a single node, while φ_3 is the highest in energy and therefore has two nodes.[55]

Any N-electron Slater determinant Ψ composed of the molecular spinorbitals $\{\phi_i\}$, $i = 1, 2, \ldots$ (cf. eq. (M.1) on p. 986) may be written as a linear combination of the Slater determinants Ψ_i^{AD} composed of the spinorbitals u_i, $i = 1, 2, \ldots$, of the *acceptor and donor*[56] (AD picture)

$$\Psi_k^{\mathrm{MO}} = \sum_i C_k(i)\Psi_i^{\mathrm{AD}}. \tag{14.38}$$

A similar expansion can also be written for the *atomic* spinorbitals (VB picture) instead of the donors and acceptors (AD picture).

---

[55] Positive a, b, c make possible the node structure described above.

[56] We start from the Slater determinant built of N molecular spinorbitals. Any of these is a linear combination of the spinorbitals of the donor and acceptor. We insert these combinations into the Slater determinant and expands the determinant according to the first row (Laplace expansion, see Appendix A on p. 889). As a result we obtain a linear combination of the Slater determinants all having the *donor or acceptor* spinorbitals in the first row. For each of the Slater determinants we repeat the procedure, but focusing on the second row, then the third row, etc. We end up with a linear combination of the Slater determinants that contain *only* the donor or acceptor spinorbitals. We concentrate on one of them, which contains some particular donor and acceptor orbitals. We are interested in the coefficient $C_k(i)$ that multiplies this Slater determinant.

In a moment we will be interested in some of the coefficients $C_k(i)$. For example, the expansion for the ground-state Slater determinant (in the MO picture)

$$\Psi_0 = N_0|\varphi_1\bar{\varphi}_1\varphi_2\bar{\varphi}_2| \tag{14.39}$$

gives

$$\Psi_0 = C_0(\mathrm{DA})\Psi_{\mathrm{DA}} + C_0(\mathrm{D^+A^-})\Psi_{\mathrm{D^+A^-}} + \cdots, \tag{14.40}$$

where $\bar{\varphi}_i$ denotes the spinorbital with spin function β, and φ_i – the spinorbital with spin function α, N_0 stands for the normalization coefficient, while Ψ_{DA}, $\Psi_{\mathrm{D^+A^-}}$ represent the normalized Slater determinants with the following electronic configurations, in $\Psi_{\mathrm{DA}} : n^2\chi^2$, in $\Psi_{\mathrm{D^+A^-}} : n^1\chi^2(\chi^*)^1$, etc.

We are first interested in the coefficient $C_0(\mathrm{DA})$. As shown by Fukui, Fujimoto and Hoffmann (cf. Appendix Z, p. 1058)[57]

$$C_0(\mathrm{DA}) \approx \langle\Psi_{\mathrm{DA}}|\Psi_0\rangle = \begin{vmatrix} a_1 & b_1 \\ a_2 & -b_2 \end{vmatrix}^2 = (a_1b_2 + a_2b_1)^2, \tag{14.41}$$

where in the determinant, the coefficients of the donor and acceptor orbitals appear in those molecular orbitals φ_i of the total system that are occupied in ground-state Slater determinant Ψ_0 (the coefficients of n and χ in φ_1 are a_1 and b_1, respectively, while those in φ_2 are a_2 and $-b_2$, respectively, see eqs. (14.37)).

Roald Hoffmann, American chemist, born 1937 in Złoczów (then Poland) to a Jewish family, professor at Cornell University in Ithaca, USA. Hoffmann discovered the symmetry rules that pertain to some reactions of organic compounds. In 1981 he shared the Nobel Prize with Kenichi Fukui "*for their theories, developed independently, concerning the course of chemical reactions*". Roald Hoffmann is also a poet and playwright. His poetry is influenced by chemistry, in which, as he wrote, was inspired by Marie Curie.

His CV reads like a film script. When in 1941 the Germans entered Złoczów, the four year old Roald was taken with his mother to a labour camp. One of the Jewish detainees betrayed a camp conspiration network to the Germans. They massacred the camp, but Roald and his mother had earlier been smuggled out of the camp by his father and hidden in a Ukrainian teacher's house. Soon after, his father was killed. The Red Army pushed the Germans out in 1944 and Roald and his mother went via Przemyśl to Cracow. In 1949 they finally reached America. Roald Hoffmann graduated from Stuyvesant High School, Columbia University and Harvard University. In Harvard Roald met the excellent chemist Professor Robert Burns Woodward (syntheses of chlorophyll, quinine, strychnine, cholesterol, penicillin structure, vitamins), a Nobel Prize winner in 1965. Woodward alerted Hoffmann to a mysterious behaviour of polyens in substitution reactions. Roald Hoffmann clarified the problem using the symmetry properties of the molecular orbitals (now known as the Woodward–Hoffmann symmetry rules, cf. p. 825).

---

[57] We assume that the orbitals n, χ and χ^* are orthogonal (approximation).

Kenichi Fukui (1918–1998), Japanese chemist, professor at the Kyoto University. One of the first scholars who stressed the importance of the IRC, and introduced what is called the frontier orbitals (mainly HOMO and LUMO), which govern practically all chemical processes. Fukui received the Nobel Prize in chemistry in 1981.

Now instead of Ψ_0 let us take two doubly excited configurations of the total system:[58]

$$\Psi_{2d} = N_2|\varphi_1\bar{\varphi}_1\varphi_3\bar{\varphi}_3| \tag{14.42}$$

and

$$\Psi_{3d} = N_3|\varphi_2\bar{\varphi}_2\varphi_3\bar{\varphi}_3|, \tag{14.43}$$

where N_i stand for the normalization coefficients. Let us ask about the coefficients that *they* produce for the DA configuration (let us call these coefficients $C_2(\mathrm{DA})$ for Ψ_{2d} and $C_3(\mathrm{DA})$ for Ψ_{3d}), i.e.

$$\Psi_{2d} = C_2(\mathrm{DA})\Psi_{\mathrm{DA}} + C_2(\mathrm{D^+A^-})\Psi_{\mathrm{D^+A^-}} + \cdots, \tag{14.44}$$

$$\Psi_{3d} = C_3(\mathrm{DA})\Psi_{\mathrm{DA}} + C_3(\mathrm{D^+A^-})\Psi_{\mathrm{D^+A^-}} + \cdots. \tag{14.45}$$

According to the result described above (see p. 1058) we obtain:

$$C_2(\mathrm{DA}) = \begin{vmatrix} a_1 & b_1 \\ -a_3 & b_3 \end{vmatrix}^2 = (a_1b_3 + a_3b_1)^2, \tag{14.46}$$

$$C_3(\mathrm{DA}) = \begin{vmatrix} a_2 & -b_2 \\ -a_3 & b_3 \end{vmatrix}^2 = (a_2b_3 - a_3b_2)^2. \tag{14.47}$$

Such formulae enable us to calculate the contributions of the particular donor-acceptor resonance structures (e.g., DA, $\mathrm{D^+A^-}$, etc., cf. p. 520) in the Slater determinants built of the molecular orbitals (14.37) of the total system. If one of these structures prevailed at a given stage of the reaction, this would represent important information about what has happened in the course of the reaction.

Please recall that at every reaction stage the main object of interest will be the ground-state of the system. The ground-state will be dominated[59] by various resonance structures. As usual the resonance structures are associated with the corresponding chemical structural formulae with the proper chemical bond pattern. *If at a reaction stage a particular structure dominated, then we would say that the system is characterized by the corresponding chemical bond pattern.*

14.5.4 REACTION STAGES

We would like to know the a, b, c values at *various reaction stages,* because we could then calculate the coefficients C_0, C_2 and C_3 for the DA as well as for other donor-acceptor structures (e.g., $\mathrm{D^+A^-}$, see below) and deduce what really happens during the reaction.

---

[58] We will need this information later to estimate the configuration interaction role in calculating the CI ground state.

[59] I.e. these structures will correspond to the highest expansion coefficients.

Reactant stage (R)

The simplest situation is at the starting point. When H^- is far away from H–H, then of course (Fig. 14.14) $\varphi_1 = \chi$, $\varphi_2 = n$, $\varphi_3 = -\chi^*$. Hence, we have $b_1 = a_2 = c_3 = 1$, while the other $a, b, c = 0$, therefore:

i	a_i	b_i	c_i
1	0	1	0
2	1	0	0
3	0	0	1

Using formulae (14.41), (14.46) and (14.47) (the superscript R recalls that the results correspond to reactants):

$$C_0^R(\mathrm{DA}) = (0 \cdot 1 + 1 \cdot 1)^2 = 1, \tag{14.48}$$

$$C_2^R(\mathrm{DA}) = 0, \tag{14.49}$$

$$C_3^R(\mathrm{DA}) = (1 \cdot 0 - 0 \cdot 0)^2 = 0. \tag{14.50}$$

When the reaction begins, the reactants are correctly described as a Slater determinant with doubly occupied n and χ orbitals, which corresponds to the DA structure.

This is, of course, what we expected to obtain for the electronic configuration of the non-interacting reactants.

Intermediate stage (I)

What happens at the intermediate stage (I)?

It will be useful to express the atomic orbitals $1s_a$, $1s_b$, $1s_c$ through orbitals n, χ, χ^* (they span the same space). From Chapter 8, p. 371, we obtain

$$1s_a = n, \tag{14.51}$$

$$1s_b = \frac{1}{\sqrt{2}}(\chi - \chi^*), \tag{14.52}$$

$$1s_c = \frac{1}{\sqrt{2}}(\chi + \chi^*), \tag{14.53}$$

where we have assumed that the overlap integrals between different atomic orbitals are equal to zero.

The intermediate stage corresponds to the situation in which the hydrogen atom in the middle (b) is at the same distance from a as from c, and therefore the two atoms are equivalent. This implies that the nodeless, one-node and two-node orbitals have the following form (where ○ stands for the $1s$ orbital and ● for the $-1s$

orbital)

$$\begin{aligned}
\varphi_1 &= \bigcirc\ \bigcirc\ \bigcirc = \frac{1}{\sqrt{3}}(1s_a + 1s_b + 1s_c),\\
\varphi_2 &= \bigcirc\ \cdot\ \bullet = \frac{1}{\sqrt{2}}(1s_a - 1s_c),\\
\varphi_3 &= \bullet\ \bigcirc\ \bullet = \frac{1}{\sqrt{3}}(-1s_a + 1s_b - 1s_c).
\end{aligned} \tag{14.54}$$

Inserting formulae (14.52) we obtain:

$$\begin{aligned}
\varphi_1 &= \frac{1}{\sqrt{3}}\Big(n + \sqrt{2}\chi + 0\cdot\chi^*\Big),\\
\varphi_2 &= \frac{1}{\sqrt{2}}\Big(n - \frac{1}{\sqrt{2}}(\chi + \chi^*)\Big),\\
\varphi_3 &= \frac{1}{\sqrt{3}}\Big(-n + 0\cdot\chi - \sqrt{2}\chi^*\Big),
\end{aligned} \tag{14.55}$$

	a_i	b_i	c_i
$i=1$	$\frac{1}{\sqrt{3}}$	$\sqrt{\frac{2}{3}}$	0
$i=2$	$\frac{1}{\sqrt{2}}$	$\frac{1}{2}$	$\frac{1}{2}$
$i=3$	$\frac{1}{\sqrt{3}}$	0	$\sqrt{\frac{2}{3}}$

(14.56)

From eq. (14.41) we have

$$C_0^I(\mathrm{DA}) = \left(\frac{1}{\sqrt{3}}\frac{1}{2} + \frac{1}{\sqrt{2}}\sqrt{\frac{2}{3}}\right)^2 = \frac{3}{4} = 0.75, \tag{14.57}$$

$$C_2^I(\mathrm{DA}) = \left(\frac{1}{\sqrt{3}}\cdot 0 + \sqrt{\frac{2}{3}}\frac{1}{\sqrt{3}}\right)^2 = \frac{2}{9} = 0.22, \tag{14.58}$$

$$C_3^I(\mathrm{DA}) = \left(\frac{1}{\sqrt{2}}\cdot 0 - \frac{1}{2}\frac{1}{\sqrt{3}}\right)^2 = \frac{1}{12} = 0.08. \tag{14.59}$$

The first of these three numbers is the most important. Something happens to the electronic ground-state of the system. At the starting point, the ground-state wave function had a DA contribution equal to $C_0^R(\mathrm{DA}) = 1$, while now this contribution has decreased to $C_0^I(\mathrm{DA}) = 0.75$. Let us see what will happen next.

Product stage (P)

How does the reaction end up?

Let us see how molecular orbitals φ corresponding to the products are expressed by n, χ and χ^* (they were defined for the starting point). At the end we have the molecule H–H (made of the middle and left hydrogen atoms) and the outgoing ion H^- (made of the right hydrogen atom).

Therefore the lowest-energy orbital at the end of the reaction has the form

$$\varphi_1 = \frac{1}{\sqrt{2}}(1s_a + 1s_b) = \frac{1}{\sqrt{2}}n + \frac{1}{2}\chi - \frac{1}{2}\chi^*, \tag{14.60}$$

which corresponds to $a_1 = \frac{1}{\sqrt{2}}$, $b_1 = \frac{1}{2}$, $c_1 = \frac{1}{2}$.

Since the φ_2 orbital is identified with $1s_c$, we obtain from eqs. (14.52): $a_2 = 0$, $b_2 = c_2 = \frac{1}{\sqrt{2}}$ (never mind that all the coefficients are multiplied by -1) and finally as φ_3 we obtain the antibonding orbital

$$\varphi_3 = \frac{1}{\sqrt{2}}(1s_a - 1s_b) = \frac{1}{\sqrt{2}}n - \frac{1}{2}\chi + \frac{1}{2}\chi^*, \tag{14.61}$$

i.e. $a_3 = \frac{1}{\sqrt{2}}$, $b_3 = \frac{1}{2}$, $c_3 = \frac{1}{2}$ (the sign is reversed as well). This leads to

i	a_i	b_i	c_i
1	$\frac{1}{\sqrt{2}}$	$\frac{1}{2}$	$\frac{1}{2}$
2	0	$\frac{1}{\sqrt{2}}$	$\frac{1}{\sqrt{2}}$
3	$\frac{1}{\sqrt{2}}$	$\frac{1}{2}$	$\frac{1}{2}$

(14.62)

Having a_i, b_i, c_i for the end of reaction, we may easily calculate $C_0^P(\mathrm{DA})$ of eq. (14.41) as well as $C_2^P(\mathrm{DA})$ and $C_3^P(\mathrm{DA})$ from eqs. (14.46) and (14.47), respectively, for the reaction products

$$C_0^P(\mathrm{DA}) = \left(\frac{1}{\sqrt{2}} \cdot \frac{1}{\sqrt{2}} + 0 \cdot \frac{1}{2}\right)^2 = \frac{1}{4}, \tag{14.63}$$

$$C_2^P(\mathrm{DA}) = \left(\frac{1}{\sqrt{2}} \cdot \frac{1}{2} + \frac{1}{\sqrt{2}} \cdot \frac{1}{2}\right)^2 = \frac{1}{2}, \tag{14.64}$$

$$C_3^P(\mathrm{DA}) = \left(0 \cdot \frac{1}{2} - \frac{1}{\sqrt{2}} \cdot \frac{1}{\sqrt{2}}\right)^2 = \frac{1}{4}. \tag{14.65}$$

Now we can reflect for a while. It is seen that during the reaction some important changes occur, namely

> when the reaction begins, the system is 100% described by the structure DA, while after the reaction it resembles this structure only by 25%.

Role of the configuration interaction

We may object that our conclusions look quite naive. Indeed, there is something to worry about. We have assumed that, independent of the reaction stage, the ground-state wave function represents a single Slater determinant Ψ_0, whereas we should rather use a configuration interaction expansion. In such an expansion, besides the dominant contribution of Ψ_0, double excitations would be the most important (p. 560), which in our simple approximation of the three φ orbitals means a leading role for Ψ_{2d} and Ψ_{3d}:

$$\Psi_{CI} = \Psi_0 + \kappa_1 \Psi_{2d} + \kappa_2 \Psi_{3d} + \cdots.$$

The two configurations would be multiplied by some *small* coefficients (because all the time we deal with the electronic ground-state dominated by Ψ_0). It will be shown that the κ coefficients in the CI expansion $\Psi = \Psi_0 + \kappa_1 \Psi_{2d} + \kappa_2 \Psi_{3d}$ are *negative*. This will serve us to make a more detailed analysis (than that performed so far) of the role of the DA structure at the beginning and end of the reaction.

The coefficients κ_1 and κ_2 may be estimated using perturbation theory with Ψ_0 as unperturbed wave function. The first-order correction to the wave function is given by formula (5.25) on p. 208, where we may safely insert the total Hamiltonian $\hat{H}$ instead of the operator[60] $\hat{H}^{(1)}$ (this frees us from saying what $\hat{H}^{(1)}$ looks like). Then we obtain

$$\kappa_1 \cong \frac{\langle \varphi_2 \bar{\varphi}_2 | \varphi_3 \bar{\varphi}_3 \rangle}{E_0 - E_{2d}} < 0, \tag{14.66}$$

$$\kappa_2 \cong \frac{\langle \varphi_1 \bar{\varphi}_1 | \varphi_3 \bar{\varphi}_3 \rangle}{E_0 - E_{3d}} < 0, \tag{14.67}$$

because from the Slater–Condon rules (Appendix M) we have $\langle \Psi_0 | \hat{H} \Psi_{2d} \rangle = \langle \varphi_2 \bar{\varphi}_2 | \varphi_3 \bar{\varphi}_3 \rangle - \langle \varphi_2 \bar{\varphi}_2 | \bar{\varphi}_3 \varphi_3 \rangle = \langle \varphi_2 \bar{\varphi}_2 | \varphi_3 \bar{\varphi}_3 \rangle - 0 = \langle \varphi_2 \bar{\varphi}_2 | \varphi_3 \bar{\varphi}_3 \rangle$ and, similarly, $\langle \Psi_0 | \hat{H} \Psi_{3d} \rangle = \langle \varphi_1 \bar{\varphi}_1 | \varphi_3 \bar{\varphi}_3 \rangle$, where E_0, E_{2d}, E_{3d} represent the energies of the corresponding states. The integrals $\langle \varphi_2 \bar{\varphi}_2 | \varphi_3 \bar{\varphi}_3 \rangle$ and $\langle \varphi_1 \bar{\varphi}_1 | \varphi_3 \bar{\varphi}_3 \rangle$ are Coulombic repulsions of a certain electron density distribution with *the same* charge distribution, *therefore*, $\langle \varphi_2 \bar{\varphi}_2 | \varphi_3 \bar{\varphi}_3 \rangle > 0$ *and* $\langle \varphi_1 \bar{\varphi}_1 | \varphi_3 \bar{\varphi}_3 \rangle > 0$.

Thus, the contribution of the DA structure to the ground-state CI function results mainly from its contribution to the single Slater determinant Ψ_0 [*coefficient* $C_0(\mathrm{DA})$], *but is modified by a small correction* $\kappa_1 C_2(\mathrm{DA}) + \kappa_2 C_3(\mathrm{DA})$, *where* $\kappa < 0$.

What are the values of $C_2(\mathrm{DA})$ and $C_3(\mathrm{DA})$ at the beginning and at the end of the reaction? At the beginning our calculations gave: $C_2^R(\mathrm{DA}) = 0$ and $C_3^R(\mathrm{DA}) = 0$. Note that $C_0^R(\mathrm{DA}) = 1$. Thus the electronic ground-state at the start of the reaction mainly represents the DA structure.

And what about the end of the reaction? We have calculated that $C_2^P(\mathrm{DA}) = \frac{1}{2} > 0$ and $C_3^P(\mathrm{DA}) = \frac{1}{4} > 0$. This means that at the end of the reaction the coefficient corresponding to the DA structure will be *certainly smaller* than $C_0^P(\mathrm{DA}) =$

---

[60] Because the unperturbed wave function Ψ_0 is an eigenfunction of the $\hat{H}^{(0)}$ Hamiltonian and is orthogonal to any of the expansion functions.

0.25, the value obtained for the single determinant approximation for the ground-state wave function.

Thus, taking the CI expansion into account *makes our conclusion* based on the single Slater determinant even *sharper*.

> When the reaction starts, the wave function means the DA structure, while when it ends, this contribution is very strongly reduced.

14.5.5 CONTRIBUTIONS OF THE STRUCTURES AS REACTION PROCEEDS

What therefore represents the ground-state wave function at the end of the reaction? To answer this question let us consider first all possible occupations of the three energy levels (corresponding to n, χ, χ^*) by four electrons. As before we assume for the orbital energy levels: $\varepsilon_\chi < \varepsilon_n < \varepsilon_{\chi^*}$. The number of such singlet-type occupations is equal to six, Table 14.1 and Fig. 14.15.

Now, let us ask what is the contribution of each of these structures[61] in Ψ_0, Ψ_{2d} and Ψ_{3d} in the three stages of the reaction. This question is especially important for Ψ_0, because this Slater determinant is dominant for the ground-state wave function. The corresponding contributions in Ψ_{2d} and Ψ_{3d} are less important, because these configurations enter the ground-state CI wave function multiplied by the tiny coefficients κ. We have already calculated these contributions for the DA structure. The contributions of all the structures are given[62] in Table 14.2.

First, let us focus on which structures contribute to Ψ_0 (because this determines the main contribution to the ground-state wave function) at the three stages of the reaction. As has been determined,

> at point R we have only the contribution of the DA structure.

Table 14.1. All possible singlet-type occupations of the orbitals: n, χ and χ^* by four electrons

ground state	DA	$(n)^2(\chi)^2$
singly excited state	D^+A^-	$(n)^1(\chi)^2(\chi^*)^1$
singly excited state	DA^*	$(n)^2(\chi)^1(\chi^*)^1$
doubly excited state	D^+A^{-*}	$(n)^1(\chi)^1(\chi^*)^2$
doubly excited state	$D^{+2}A^{-2}$	$(\chi)^2(\chi^*)^2$
doubly excited state	DA^{**}	$(n)^2(\chi^*)^2$

---

[61] We have already calculated some of these contributions.

[62] Our calculations gave $C_0^I(\mathrm{DA}) = 0.75$, $C_2^I(\mathrm{DA}) = 0.22$, $C_3^I(\mathrm{DA}) = 0.08$. In Table 14.2 these quantities are equal: 0.729, 0.250, 0.020. The only reason for the discrepancy may be the non-zero overlap integrals, which were neglected in our calculations and were taken into account in those given in Table 14.2.

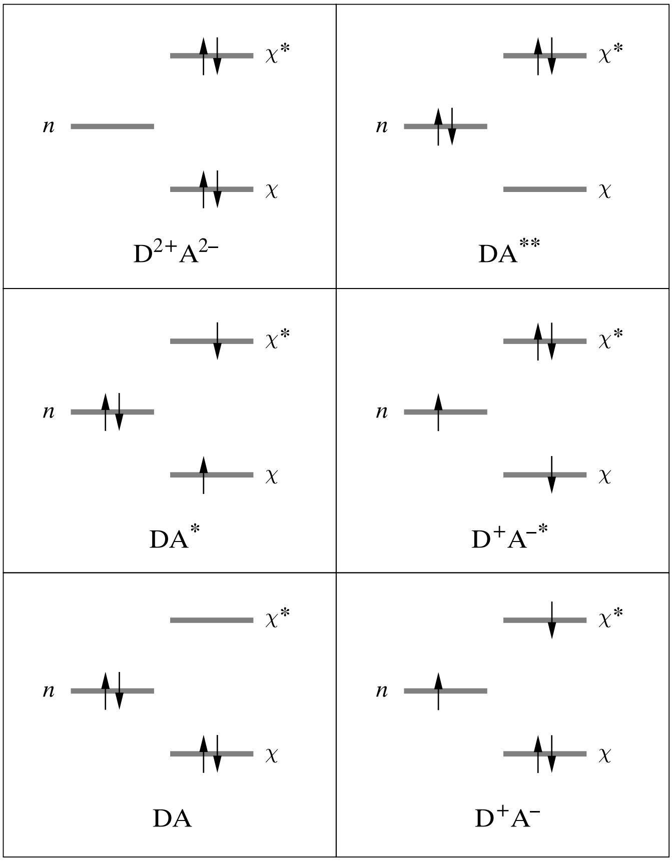


Fig. 14.15. The complete set of the six singlet wave functions ("structures"), that arise from occupation of the donor orbital n and of the two acceptor orbitals (χ and χ^*).

However, as we can see (main contributions in bold in Table 14.2),

when the reaction advances along the reaction path to point I, the contribution of DA decreases to 0.729, other structures come into play with the dominant D^+A^- (the coefficient equal to -0.604).

At point P there are three dominant structures: D^+A^-, D^+A^{-*} and $D^{+2}A^{-2}$.

Now we may think of going beyond the single determinant approximation by performing the CI. In the R stage the DA structure dominates as before, but has some small admixtures of DA^{**} (because of Ψ_{3d}) and $D^{+2}A^{-2}$ (because of Ψ_{2d}), while at the product stage the contribution of the DA structure almost vanishes. Instead, some important contributions of the excited states appear, mainly of the

Table 14.2. The contribution of the six donor–acceptor structures in the three Slater determinants Ψ_0, Ψ_{2d} and Ψ_{3d} built of molecular orbitals at the three reaction stages: reactant (R), intermediate (I) and product (P) [S. Shaik, *J. Am. Chem. Soc.* 103 (1981) 3692. Adapted with permission from the American Chemical Society. Courtesy of the author.]

Structure	MO determinant		R	I	P
DA	Ψ_0	C_0(DA)	1	**0.729**	0.250
	Ψ_{2d}	C_2(DA)	0	0.250	0.500
	Ψ_{3d}	C_3(DA)	0	0.020	0.250
D^+A^-	Ψ_0	$C_0(D^+A^-)$	0	**−0.604**	**−0.500**
	Ψ_{2d}	$C_2(D^+A^-)$	0	0.500	0.000
	Ψ_{3d}	$C_3(D^+A^-)$	0	0.103	0.500
DA^*	Ψ_0	$C_0(DA^*)$	0	0.177	0.354
	Ψ_{2d}	$C_2(DA^*)$	0	0.354	−0.707
	Ψ_{3d}	$C_3(DA^*)$	0	0.177	0.354
D^+A^{-*}	Ψ_0	$C_0(D^+A^{-*})$	0	0.103	**0.500**
	Ψ_{2d}	$C_2(D^+A^{-*})$	0	0.500	0.000
	Ψ_{3d}	$C_3(D^+A^{-*})$	0	−0.604	−0.500
DA^{**}	Ψ_0	$C_0(DA^{**})$	0	0.021	0.250
	Ψ_{2d}	$C_2(DA^{**})$	0	0.250	0.500
	Ψ_{3d}	$C_3(DA^{**})$	1	0.729	0.250
$D^{+2}A^{-2}$	Ψ_0	$C_0(D^{+2}A^{-2})$	0	0.250	**0.500**
	Ψ_{2d}	$C_2(D^{+2}A^{-2})$	1	0.500	0.000
	Ψ_{3d}	$C_3(D^{+2}A^{-2})$	0	0.250	0.500

D^+A^-, D^+A^{-*} and $D^{+2}A^{-2}$ structures, but also other structures of smaller importance.

> The value of the qualitative conclusions comes from the fact that they do not depend on the approximation used, e.g., on the atomic basis set, neglecting the overlap integrals, etc.

For example, the contributions of the six structures in Ψ_0 calculated using the Gaussian atomic basis set STO-3G and within the extended Hückel method are given in Table 14.3 (main contributions in bold). Despite the fact that even the geometries used for the R, I, P stages are slightly different, the qualitative results are the same. It is rewarding to learn things that do not depend on detail.

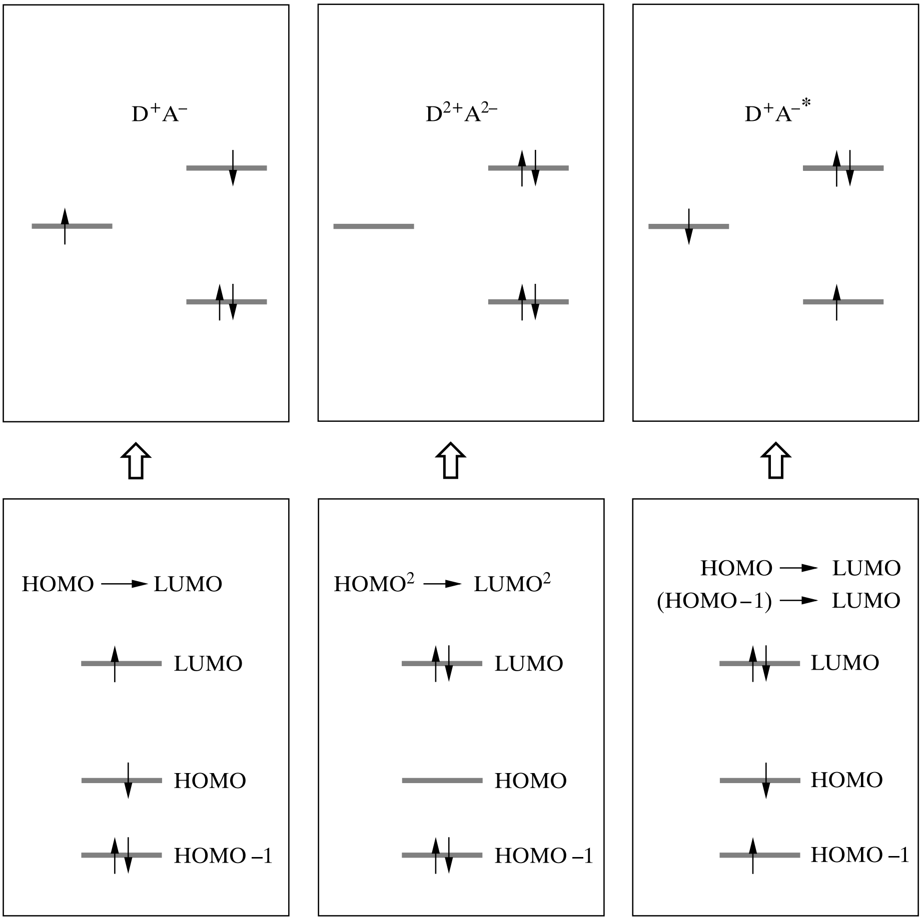
Where do the final structures D^+A^-, D^+A^{-*} and $D^{+2}A^{-2}$ come from?

As seen from Table 14.2, the main contributions at the end of the reaction come from the D^+A^-, D^+A^{-*} and $D^{+2}A^{-2}$ structures. What do they correspond to *when the reaction starts*? From Table 14.2 it follows that the $D^{+2}A^{-2}$ structure simply represents Slater determinant Ψ_{2d} (Fig. 14.16). But where do the D^+A^- and D^+A^{-*} structures come from? There are no such contributions either in Ψ_0, or in Ψ_{2d} or in Ψ_{3d}. It turns out however that a similar analysis applied to the

Table 14.3. Contributions of the six donor–acceptor structures in the Ψ_0 Slater determinant at three different stages (R, I, P) of the reaction [S. Shaik, *J. Am. Chem. Soc.* 103 (1981) 3692. Adapted with permission from the American Chemical Society. Courtesy of the author.]

	STO-3G			Extended Hückel		
Structure	R	I	P	R	I	P
DA	**1.000**	**0.620**	0.122	**1.000**	**0.669**	0.130
D^+A^-	0.000	**−0.410**	**−0.304**	−0.012	**−0.492**	**−0.316**
DA^*	0.000	0.203	0.177	0.000	0.137	0.179
D^+A^{-*}	0.000	0.125	**0.300**	0.000	0.072	**0.298**
DA^{**}	0.000	0.117	**0.302**	0.000	0.176	**0.301**
$D^{+2}A^{-2}$	0.000	0.035	0.120	0.000	0.014	0.166

Most important acceptor–donor structures at P



These structures correspond to the following MO configurations at R

Fig. 14.16. What final structures are represented at the starting point?

normalized configuration[63] $N|\varphi_1\bar{\varphi}_1\varphi_2\bar{\varphi}_3|$ at stage R gives *exclusively* the D^+A^- structure, while applied to the $N|\varphi_1\bar{\varphi}_2\varphi_3\bar{\varphi}_3|$ determinant, it gives *exclusively* the D^+A^{-*} structure (Fig. 14.16). So we have traced them back. The first of these configurations corresponds to a single-electron excitation from HOMO to LUMO – this is, therefore, the lowest excited state of the reactants. Our picture is clarified:

> the reaction starts from DA, at the intermediate stage (transition state) we have a large contribution of the first excited state that at the starting point was the D^+A^- structure related to the excitation of an electron from HOMO to LUMO.

The states DA and D^+A^- undergo the "quasi-avoided crossing" in the sense described on p. 262. This means that at a certain geometry, the roles played by HOMO and LUMO interchange, i.e. what was HOMO becomes LUMO and *vice versa*.[64]

Donor and acceptor orbital populations at stages R, I, P

Linear combinations of orbitals n, χ and χ^* construct the molecular orbitals of the system in full analogy with the LCAO expansion of the molecular orbitals. Therefore we may perform a similar population analysis as that described in Appendix S, p. 1015. The analysis will tell us where the four key electrons of the system are (more precisely how many of them occupy n, χ and χ^*), and since the population analysis may be performed at different stages of the reaction, we may obtain information as to what happens to the electrons when the reaction proceeds. The object to analyze is the wave function Ψ. We will report the population analysis results for its dominant component, namely Ψ_0. The results of the population analysis are reported in Table 14.4. The content of this table confirms our previous conclusions.

Table 14.4. Electronic population of the donor and acceptor orbitals at different reaction stages (R, I, P) [S. Shaik, *J. Am. Chem. Soc.* 103 (1981) 3692. Adapted with permission from the American Chemical Society. Courtesy of the author.]

	Population		
Orbital	R	I	P
n	2.000	1.513	1.000
χ	2.000	1.950	1.520
χ^*	0.000	0.537	1.479

---

[63] N stands for the normalization coefficient.

[64] The two configurations differ by a single spinorbital and the resonance integral $\langle DA|\hat{H}|D^+A^-\rangle$ when reduced using the Slater–Condon rules is dominated by the one-electron integral involving HOMO (or n) and the LUMO (or χ^*). Such an integral is of the order of the overlap integral between these orbitals. The energy gap between the two states is equal to twice the absolute value of the resonance integral (the reason is similar to the bonding-antibonding orbital separation in the hydrogen molecule).

- The starting point (R) has occupation: $(n)^2(\chi)^2$, and that is fine, because we are dealing with the DA structure.
- The intermediate stage represents a mixture of two structures mainly (with almost equal contributions): $(n)^2(\chi)^2$, i.e. DA and $(n)^1(\chi)^2(\chi^*)^1$, i.e. D^+A^-. Therefore we may expect that the population of χ is close to 2, of n is about 1.5, while of χ^* is about 0.5. This is indeed the case.
- The final stage (P) is a mixture of the D^+A^- structure, which corresponds to the occupation $(n)^1(\chi)^2(\chi^*)^1$, of the structure D^+A^{-*} corresponding to $(n)^1(\chi)^1(\chi^*)^2$ and the structure $D^{+2}A^{-2}$ with occupation $(n)^0(\chi)^2(\chi^*)^2$. Equal contributions of these structures should therefore give the occupations of n, χ, χ^* equal to $\frac{2}{3}$, $\frac{5}{3}$ and $\frac{5}{3}$, respectively. The population analysis gives similar numbers, see Table 14.4, last column.

14.5.6 NUCLEOPHILIC ATTACK H^-+ ETHYLENE → ETHYLENE + H^-

Maybe the acceptor–donor theory described above pertains only to the H^- + H–H reaction? Fortunately enough, its applicability goes far beyond. Let us consider a nucleophilic attack of the H^- ion on the ethylene molecule (Fig. 14.17), perpendicular to the ethylene plane towards the position of one of the carbon atoms. The arriving ion binds to the carbon atom forming the CH bond, while another proton with two electrons (i.e. H^- ion) leaves the system. Such a reaction looks like it is of academic interest only (except some isotopic molecules are involved, e.g., when one of the protons is replaced by a deuteron), but comprehension comes from the simplest examples possible, when the least number of things change.

The LCAO MO calculations for the ethylene molecule give the following result. The HOMO orbital is of the π bonding character, while the LUMO represents the antibonding π^* orbital (both are linear combinations of mainly carbon $2p_z$ atomic orbitals, z being the axis perpendicular to the ethylene plane). On the energy scale the H^- 1*s* orbital goes between the π and π^* energies, similarly as happened with the χ and χ^* orbitals in the H^- + H–H reaction. The virtual orbitals (let us call them $2\chi^*, 3\chi^*, 4\chi^*$) are far away up in the energy scale, while the occupied σ-type orbitals are far down in the energy scale. Thus, the H^- $n = 1s$ orbital energy is close to that of χ and χ^*, while other orbitals are well separated from them.

This energy level scheme allows for many possible excitations, far more numerous than considered before. *Despite this*, because of the effective mixing of only those donor and acceptor orbitals that are of comparable energies, *the key partners are, as before, n, χ and χ^**. The role of the other orbitals is only marginal: their admixtures will only slightly deform the shape of the main actors of the drama n, χ and χ^* known as the *frontier orbitals*. The coefficients at various acceptor–donor structures in the expansion of Ψ_0 are shown in Table 14.5. The calculations were performed using the extended Hückel method[65] at three stages of the reaction (R, in which the H^- ion is at a distance of 3 Å from the attacked carbon atom; I, with a

frontier orbitals

---

[65] Introduced to chemistry by Roald Hoffmann. He often says that he cultivates chemistry with an old, primitive tool, which because of this ensures access to the wealth of the complete Mendeleev periodic table.

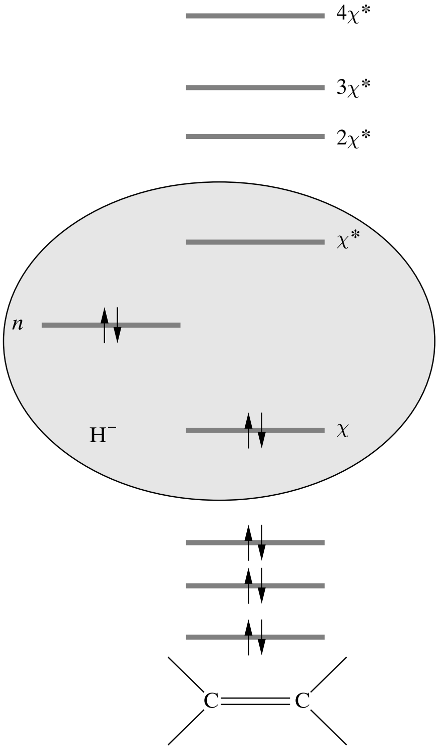


Fig. 14.17. Nucleophilic substitution of ethylene by H^-. The figure aims to demonstrate that, despite considering a more complex system than the $H^- + H_2 \rightarrow H_2 + H^-$ reaction discussed so far, the machinery behind the scene works in the same way. The attack of H^- goes perpendicularly to the ethylene plane, onto one of the carbon atoms. The figure shows the (orbital) energy levels of the donor (H^-, left hand side) and of the acceptor (ethylene, right hand side). Similarly as for $H^- + H_2$ the orbital energy of the donor orbital n is between the acceptor orbital energies χ and χ^* corresponding to the bonding π and antibonding π^* orbitals. Other molecular orbitals of the ethylene (occupied and virtual: $2\chi^*, 3\chi^*, \ldots$) play a marginal role, due to high energetic separation from the energy level of n.

Table 14.5. Expansion coefficients at the acceptor–donor structures in the ground-state wave function at various stages of the reaction: reactant (R), intermediate (I) and product (P) [S. Shaik, *J. Am. Chem. Soc.* 103 (1981) 3692. Adapted with permission from the American Chemical Society. Courtesy of the author.]

	Coefficients		
Structure	R	I	P
DA	**1.000**	**0.432**	0.140
$D^+A^-(n \rightarrow \pi^*)$	0.080	**0.454**	**0.380**
DA^* $(\pi \rightarrow \pi^*)$	-0.110	-0.272	-0.191
$D^+A^{-*}(n \rightarrow \pi^*, \pi \rightarrow \pi^*)$	-0.006	-0.126	$\mathbf{-0.278}$
$D^+A^-(n \rightarrow 2\sigma^*)$	$<10^{-4}$	0.006	0.004
$D^+A^-(n \rightarrow 3\sigma^*)$	$<10^{-4}$	-0.070	-0.040

distance 1.5 Å and P, with a distance equal to 1.08 Å; in all cases the planar geometry of the ethylene was preserved). It is seen that, despite the fact that a more complex method was used, the emerging picture is quite similar: at the beginning the DA structure prevails, at the intermediate stage we have a “hybrid” of the DA and D^+A^- structures, while at the end we have a major role for the D^+A^- and D^+A^{-*} structures. We can see also that even if some higher excitations were taken into account (to the orbitals $2\sigma^*, 3\sigma^*$) they play only a marginal role. The cor-

responding population analysis (not reported here) indicates a basically identical mechanism. This resemblance extends also to the S_N2 nucleophilic substitutions in aromatic compounds.

14.5.7 ELECTROPHILIC ATTACK $H^+ + H_2 \to H_2 + H^+$

Let us see whether this mechanism is even more general and consider the *electrophilic* substitution in the model reaction $H^+ + H\text{–}H \to H\text{–}H + H^+$. This time the role of the donor is played by the hydrogen molecule, while that of the acceptor is taken over by the proton. The total number of electrons is only two. The DA structure corresponds to $(\chi)^2(n)^0(\chi^*)^0$. Other structures are defined in an analogous way to the previous case of the H_3 system: structure D^+A^- means $(\chi)^1(n)^1(\chi^*)^0$, structure $D^{+*}A^-$ obviously corresponds to $(\chi)^0(n)^1(\chi^*)^1$, structure D^*A to $(\chi)^1(n)^0(\chi^*)^1$, $D^{**}A$ to $(\chi)^0(n)^0(\chi^*)^2$ and $D^{+2}A^{-2}$ to $(\chi)^0(n)^2(\chi^*)^0$. As before, the ground-state Slater determinant may be expanded into the contributions of these structures. The results (the overlap neglected) are collected in Table 14.6.

Table 14.6. Expansion coefficients at the acceptor–donor structures for the reaction of proton with the hydrogen molecule at three different stages of the reaction: reactant (R), intermediate (I) and product (P) [S. Shaik, *J. Am. Chem. Soc.* 103 (1981) 3692. Adapted with permission from the American Chemical Society. Courtesy of the author.]

	Coefficients		
Structure	R	I	P
DA	**1.000**	**0.729**	0.250
D^+A^-	0	**0.604**	**0.500**
$D^{+*}A^-$	0	−0.104	**−0.500**
D^*A	0	−0.177	−0.354
$D^{**}A$	0	0.021	0.250
$D^{+2}A^{-2}$	0	0.250	0.500

It is worth stressing that we obtain the same reaction machinery as before. First, at stage R the DA structure prevails, next at intermediate stage I we have a mixture of the DA and D^+A^- structures, and we end up (stage P) with D^+A^- and $D^{+*}A^-$ (the energy levels for the donor are the same as the energy levels were previously for the acceptor, hence we have $D^{+*}A^-$, and not D^+A^{-*} as before). This picture would not change qualitatively if we considered electrophilic substitution of the ethylene or benzene.

14.5.8 NUCLEOPHILIC ATTACK ON THE POLARIZED CHEMICAL BOND IN THE VB PICTURE

$$X^- + \rangle C{=}Y \to \rangle C{=}X + Y^-$$

What does the quasi-avoided crossing described above really mean? At the beginning we have the DA structure almost exclusively. The DA structure obviously corresponds to a chemical bond in the acceptor and the lack of any bond between the donor and acceptor. In the D^+A^- structure which comes into play at the intermediate stage, we have two paired electrons: one on D^+ occupying orbital n, and the second on A^- occupying χ^*. *These electrons represent the pair that will be responsible for formation of the new bond, the D–A bond*. At the same time, the old bond in A is positively weakened, because one of the electrons in A occupies the antibonding χ^* orbital. Therefore, the quasi-avoided crossing between the diabatic hypersurfaces DA and D^+A^- represents the key region, in which breaking of the old bond in A and formation of new bond D–A are taking place.

The above theory is based on the acceptor/donor expansion functions (AD formalism). As has already been mentioned (p. 803), the third possibility (apart from AD and MO) is VB, in which the Slater determinants (playing the role of the expansion functions for Ψ_0) are built of the *atomic orbitals* of the interacting species (Chapter 10, p. 520). How does such a VB picture look? Let us consider a nucleophilic attack of the species X on the polarized double bond $\rangle C{=}Y$, where Y represents an atom more electronegative than carbon (say, oxygen). Our goal is to expand the AD structures into the VB. The arguments of the kind already used for ethylene make it possible to limit ourselves exclusively to the frontier orbitals n, π and π^* (Fig. 14.18).

The bonding π orbital may be approximated as a linear combination of the $2p_z$ atomic orbitals of Y and C[66]

$$\pi = a \cdot (2p_z)_{\mathrm{C}} + b \cdot (2p_z)_{\mathrm{Y}}, \tag{14.68}$$

where we assume (by convention) that the coefficients satisfy: $a, b > 0$. Note that the orbital π is *polarized* this time, and due to a higher electronegativity of Y we

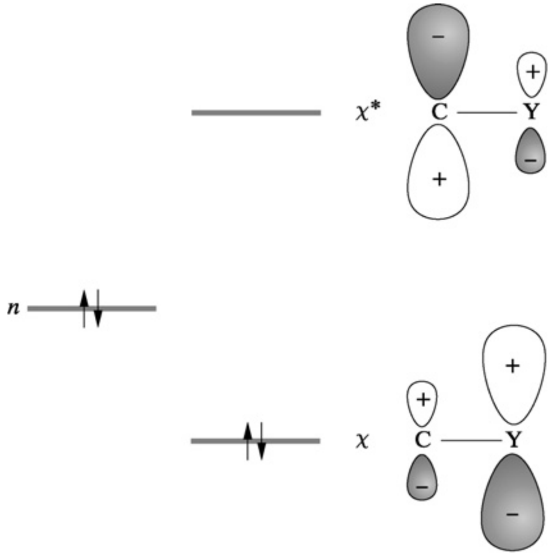


Fig. 14.18. Nucleophilic attack $X^- + \rangle C{=}Y \rightarrow \rangle C{=}X + Y^-$. The orbitals π and π^* are polarized (their polarizations are *opposite*). [S. Shaik, *J. Am. Chem. Soc.* 103 (1981) 3692. Adapted with permission from the American Chemical Society. Courtesy of the author.]

---

[66] The z axis is perpendicular to the $\rangle C{=}Y$ plane.

have $b > a$ with the normalization condition $a^2 + b^2 = 1$ (we neglect the overlap integrals between the atomic orbitals). In this situation the antibonding orbital π^* may be obtained from the orthogonality condition of the orbital π as:[67]

$$\pi^* = b \cdot (2p_z)_{\mathrm{C}} - a \cdot (2p_z)_{\mathrm{Y}}. \tag{14.69}$$

The role of the donor orbital n will be played by $(2p_z)_{\mathrm{X}}$. Note that π^* has the opposite polarization to that of π, i.e. the electron described by π^* prefers to be close to the less electronegative carbon atom, Fig. 14.18.

At the starting point the DA structure which corresponds to the double occupation of n and χ turned out to be the most important. In Chapter 8 on p. 371, a Slater determinant was analyzed that corresponded to double occupation of bonding orbital $\sigma 1s$ of the hydrogen molecule. In the present situation this corresponds to a double occupation of the π orbital of the acceptor. The Slater determinant was then expanded onto the VB structures (eq. (10.18), p. 521), and as it turns out, there are three of them. The first was the Heitler–London structure, which described a covalent bond: if one electron is close to the first nucleus then the other (with opposite spin) will be close to the second nucleus. Both electrons played exactly the same role with respect to the two nuclei. The second and third structures were of the ionic character, because both electrons were at the same nucleus (one or the other). The two ionic structures had equal coefficients and together with the Heitler–London structure, this led to treating both nuclei on an equal footing. If one of the nuclei were a bit different to the other (e.g., by increasing its charge, which would simulate its higher electronegativity), as is the case in a polarized bond ($a \neq b$), then the *Heitler–London function would continue to treat the nuclei in the same way*, but the polarity would be correctly restored by making the coefficients of the ionic structures different. The reason for this asymmetry is setting $b > a$.

This is why the chemical bond pattern corresponding to the VB picture may be expressed by the pictorial description shown in Fig. 14.19.a.

What happens at the intermediate stage, when the $\mathrm{D^+A^-}$ structure enters into play? In this structure one electron occupying n goes to χ^*. The Slater determinant

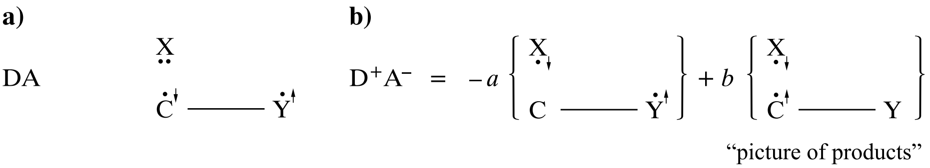


Fig. 14.19. Pictorial description of the DA and $\mathrm{D^+A^-}$ structures. For a large donor–acceptor distance the electronic ground state is described by the DA structure (a). Structure $\mathrm{D^+A^-}$ already becomes very important for the intermediate stage (I). This structure, belonging to the acceptor–donor picture, is shown (b) in the VB representation, where the opposite spins of the electrons remind us that we have the corresponding covalent structure.

---

[67] Let us check:

$$\langle \pi | \pi^* \rangle = \langle a \cdot (2p_z)_{\mathrm{C}} + b \cdot (2p_z)_{\mathrm{Y}} | b \cdot (2p_z)_{\mathrm{C}} - a \cdot (2p_z)_{\mathrm{Y}} \rangle = ab + 0 + 0 - ba = 0.$$

that corresponds to this structure [one of its rows contains a spinorbital corresponding to orbital $\pi^* = b(2p_z)_C - a(2p_z)_Y$] may be expanded as a linear combination of the two Slater determinants, which instead of the spinorbital mentioned above have the corresponding atomic spinorbital [$(2p_z)_C$ or $(2p_z)_Y$]. The corresponding pictorial notation is shown in Fig. 14.19.b. Note that the weight of the second structure is higher. Therefore, we see what is going on in the VB picture. The DA structure corresponds to the “old” CY bond, while the D^+A^- structure becomes more and more important (when the reaction proceeds) and mainly represents the Heitler–London structure for the new CX covalent bond.

The avoided crossing is needed to cause such a change of the electronic structure as to break the old bond and form the new one. Taking the leading VB structures only, we may say that

> the avoided crossing appears between two hypersurfaces, from which one corresponds to the old bond pattern (the first diabatic hypersurface) and the other to the new bond pattern (the second diabatic hypersurface).

We see from the VB results, why the variational method has chosen the D^+A^- structure among the six possible ones.

> This configuration was chosen, because it corresponds exactly to the formation of the new bond: the two unpaired electrons with opposite spins localized on those atoms that are going to bind.

> The mechanism given is general and applies wherever at least one of the reactants has a closed shell. When both the reacting molecules are of the open-shell type, there will be no avoided crossing and no reaction barrier: the reactants are already prepared for the reaction.

When we have a closed-shell system among the reactants, for the reaction to happen we have to reorganize the electronic structure. The reorganization may happen only *via* an avoided crossing and that means the appearance of a reaction barrier.

14.5.9 WHAT IS GOING ON IN THE CHEMIST’S FLASK?

Let us imagine the molecular dynamics on energy hypersurface calculated using a quantum-mechanical method (classical force fields are not appropriate since they offer non-breakable chemical bonds). The system is represented by a point that slides downhill (with an increasing velocity) and climbs uphill (with deceasing velocity). The system has a certain kinetic energy, because chemists often heat their flasks.

Let us assume that, first the system wanders through those regions of the hypersurface which are far from other electronic states in the energy scale. In such a situation, the adiabatic approximation (Chapter 6) is justified and the electronic energy (corresponding to the hypersurface) represents potential energy for the motion of the nuclei. The system corresponds to a given chemical bond pattern (we may work out a structural formula). Bond lengths vibrate as do bond angles, torsional angles also change, even leading to new isomers (conformers), but a single bond remains single, double remains double, etc.

After a while the system climbs to a region of the configurational space in which another diabatic hypersurface (corresponding to another electronic state) lowers its energy to such an extent that the two hypersurfaces tend to intersect. In this region the adiabatic approximation fails, since we have *two* electronic states of comparable energies (both have to be taken into account), and the wave function cannot be taken as the product of an electronic function and a function describing the nuclear motion (as is required by the adiabatic approximation). As a result of mixing, electronic states crossing is avoided, and two *adiabatic* hypersurfaces (upper and lower) appear. Each is composed of two parts. One part corresponds to a molecule looking as if it had one bond pattern, while the other part pertains to a different bond pattern. The bond pattern on each of the adiabatic hypersurfaces changes and the Rubicon for this change is represented by the boundary, i.e. the region of the quasi-avoided crossing that separates the two diabatic parts of the adiabatic hypersurface. Therefore, when the system in its dynamics goes uphill and enters the boundary region, the corresponding bond pattern becomes fuzzy, and changes to another pattern after crossing the boundary. The reaction is completed.

What will happen next? The point representing the system in the configurational space continues to move and it may happen to arrive at another avoided-crossing region[68] and its energy is sufficient to overcome the corresponding barrier. This is the way multistep chemical reactions happen. It is important to realize that, in experiments, we have to do with an ensemble of such points rather than one. The points differ by their positions (configurations of the nuclei) and momenta. Only a fraction of them has sufficiently high kinetic energy to cross the reaction barrier. The rest wander through a superbasin (composed of numerous basins) of the initial region thus undergoing vibrations, rotations including internal rotations, etc. Of those which cross a barrier, only a fraction crosses the same barrier again (i.e. the barrier of the same reaction). Others, depending on the initial conditions (nuclear positions and momenta) may cross other barriers. The art of chemistry means that in such a complicated situation it is still possible to perform reactions with nearly 100% yield and obtain a pure chemical compound – the chemist's target.

14.5.10 ROLE OF SYMMETRY

A VB point of view is simple and beautiful, but sometimes the machinery gets stuck. For example, this may happen when the described mechanism has to be rejected, because it does not meet some symmetry requirements. Imagine that instead of a linear approach of H^- to H_2, we consider a T-shape configuration. In

---

[68] This new avoided crossing may turn out to be the old one. In such a case the system will cross the barrier in the opposite direction. Any chemical reaction is reversible (to different extents).

such a case the all-important D^+A^- structure becomes *useless* for us, because the resonance integral which is proportional to the overlap integral between the 1*s* orbital of H^- (HOMO of the donor) and χ^* (LUMO of the acceptor) is equal to zero for symmetry reasons. If the reaction were to proceed, we would have had to form molecular orbitals from the above orbitals and this is impossible.

Yet there is an emergency exit from this situation. Let us turn our attention to the D^+A^{-*} structure, which corresponds to a doubly occupied χ^*, but a singly occupied χ. *This* structure would lead to the avoided crossing, because the overlap integral of 1*s* H^- and χ H–H has a non-zero value. In this way,

> a forbidden symmetry will simply cause the system to choose as the lead, another structure, such that it allows the formation of new bonds in *this situation*.

The above example shows that symmetry can play an important role in chemical reactions. The role of symmetry will be highlighted in what follows.

The cycloaddition reaction

Let us take the example of the cycloaddition of two ethylene molecules when they bind together forming the cyclobutane. The frontier orbitals of the ground-state ethylene molecule are: the doubly occupied π (HOMO) and the empty π^* (LUMO) molecular orbitals.

The right-hand side of Fig. 14.20.a shows that the reaction would go towards the products, if we prepared the reacting ethylene molecules in their triplet states. Such a triplet state has to be stabilized during the reaction, while the state corresponding to the old bond pattern should loose its importance. Is it reasonable to expect the triplet state to be of low energy in order to have the chance to be pulled sufficiently down the energy scale? Yes, it is, because the triplet state arises by exciting an electron from the HOMO (i.e. π) to the LUMO (i.e. π^*), and this energy cost is the lowest possible (in the orbital picture). Within the π-electron approximation the Slater determinant corresponding to the triplet state (and representing the corresponding molecular orbitals as linear combination of the carbon $2p_z$ atomic orbitals denoted simply as a and b) has the form

$$N\det\bigl(\pi(1)\alpha(1)\pi^*(2)\alpha(2)\bigr) \tag{14.70}$$

$$= N\bigl[\pi(1)\alpha(1)\pi^*(2)\alpha(2) - \pi(2)\alpha(2)\pi^*(1)\alpha(1)\bigr] \tag{14.71}$$

$$\begin{aligned} &= N\alpha(1)\alpha(2)\bigl[\bigl(a(1)+b(1)\bigr)\bigl(a(2)-b(2)\bigr) \\ &\quad - \bigl(a(2)+b(2)\bigr)\bigl(a(1)-b(1)\bigr)\bigr] \end{aligned} \tag{14.72}$$

$$= -2N\alpha(1)\alpha(2)\bigl[a(1)b(2) - a(2)b(1)\bigr]. \tag{14.73}$$

This means that when one electron is on the first carbon atom, the other is on the second carbon atom (no ionic structures!). The “parallel” electron spins of

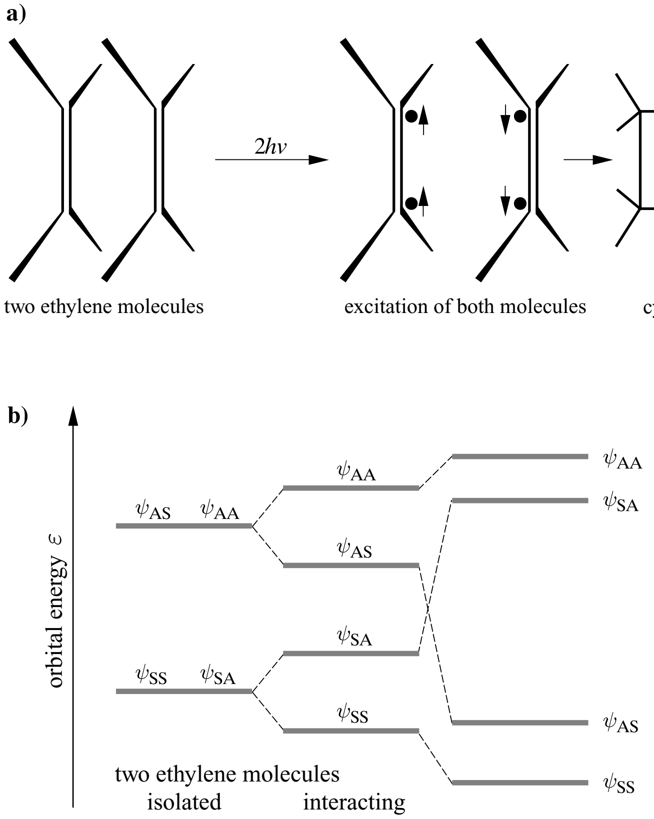


Fig. 14.20. Two equivalent schemes for the cycloaddition reaction of ethylene. Two ethylene molecules, after excitation to the triplet state, dimerize forming cyclobutane (a), because everything is prepared for electron pairing and formation of the new bonds (see text). We obtain the same from the Woodward–Hoffmann rules (Fig. (b), (c), (d)). According to these rules we assume that the ethylene molecules preserve two planes of symmetry: P_1 and P_2 during all stages of the reaction. We concentrate on four π electrons – the main actors in the drama. At the beginning the lowest-energy molecular orbital of the total system (b,c) is of the SS type (i.e. symmetric with respect to P_1 and P_2). The other three orbitals (not shown in Fig. (c)) are of higher energies that increases in the following order: SA, AS, AA. Hence, the four electrons occupy SS and SA, (b). Fig. (d) shows the situation after the reaction. The four electrons are no longer of the π type, we now call them the σ type, and they occupy the hybrid orbitals shown in the figure. Once more, the lowest energy (b) corresponds to the SS symmetry orbital (d). The three others (not shown in Fig. (d)) have higher energy, but their order *is different than before* (b): AS, SA, AA. The four electrons should occupy, therefore, the SS and AS type orbitals, whereas (according to the Woodward–Hoffmann rule) they still occupy SS and SA. This is energetically unfavourable and such a thermic reaction does not proceed. Yet, if before the reaction the ethylene molecules were excited to the triplet state $(\pi)^1(\pi^*)^1$, then at the end of the reaction they would correspond to the configuration: $(SS)^2(AS)^2$, of very low energy, and the photochemical reaction proceeds.

one molecule may be in the opposite direction to the similar electron spins of the second molecule. Everything is prepared for the cycloaddition, i.e. formation of the new chemical bonds.

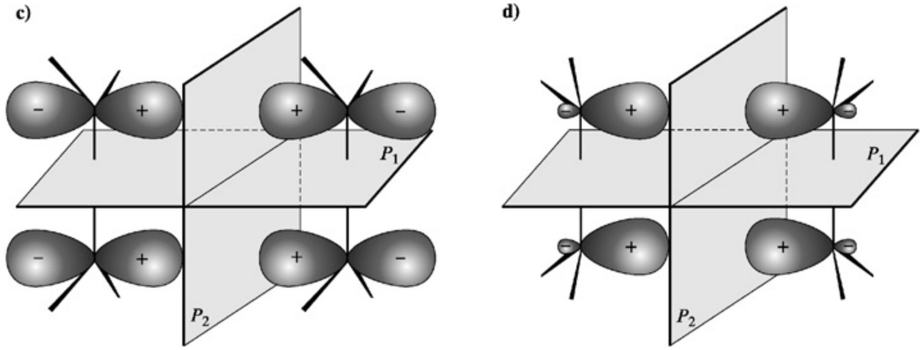


Fig. 14.20. *Continued.*

Similar conclusions can be drawn from the Woodward–Hoffmann symmetry rules.

Woodward–Hoffmann symmetry rules

The rules pertain to such an approach of two molecules that all the time some symmetry elements of the nuclear framework are preserved (there is a symmetry group associated with the reaction, see Appendix C). Then,

- the molecular orbitals belong to the irreducible representations of the group,
- we assume that during the approach the orbital energies change, *but their electron occupancies do not*,
- the reaction is allowed when the sum of the (occupied) orbital energies lowers, otherwise it is forbidden.

Example 3. *Two ethylene molecules – Diels–Alder reaction*

The two ethylene molecules are oriented as shown in Fig. 14.20.c. Let us focus on the frontier (HOMO and LUMO) orbitals at long intermolecular distances. All are built of symmetry orbitals composed of four 2p carbon atomic orbitals (perpendicular to the planes corresponding to the individual molecules) and can be classified as symmetric (S) or antisymmetric (A) with respect to the symmetry planes P_1 and P_2. Moreover, by recognizing the bonding or antibonding interactions, without performing any calculations, we can tell that the SS-symmetry orbital is of the lowest energy (because of the bonding character of the intra- as well as intermolecular interactions), then the SA-symmetry (the bonding intramolecular – the π orbitals and the antibonding intermolecular), next the AS-symmetry (the antibonding intramolecular and bonding intermolecular orbitals π^*), and the highest-energy orbital AA (the antibonding intra- and intermolecular). Since the number of electrons involved is four, they occupy the SS and SA orbitals.[69] This is what we have at the beginning of the reaction.

What do we have at the end of the reaction? At the end there are no π-electrons whatsoever, instead we have two new σ chemical bonds, each built from the two

---

[69] What a nasty historical association.

sp^3 hybrids (Fig. 14.20.d) oriented towards the other ethylene molecule.[70] Therefore, we may form the symmetry orbitals once again, recognize their bonding and antibonding character and hence the order of their orbital energies without any calculations, just by inspection (Fig. 14.20.b). The lowest energy corresponds, of course, to SS (because the newly formed σ chemical bonds correspond to the bonding combination and the lateral overlap of the hybrids is also of the bonding character), *the next in energy however is the* AS (because of the bonding interactions in the newly formed σ bonds, while the lateral interaction is weakly antibonding), then follows the SA-symmetry orbital (antibonding interaction along the bonds that is only slightly compensated by the lateral bonding overlap of the hybrids), and finally, the highest-energy corresponds to the totally antibonding orbital of the AA-symmetry.

According to the Woodward–Hoffmann rules, the four π electrons, on which we focus, occupy the SS and SA orbitals from the beginning to the end of the reaction. This corresponds to low energy at the beginning of the reaction (R), but is very unfavourable at its end (P), because the unoccupied AS orbital is lower in the energy scale. And what if we were smart and excited the reactants by laser? This would allow double occupation of the AS orbital right at the beginning of the reaction and end up with a low energy configuration. To excite an electron per molecule, means to put one on orbital π^*, while the second electron stays on orbital π. Of two possible spin states (singlet and triplet) the triplet state is lower in energy (see Chapter 8, p. 391). This situation was described by eq. (14.73) and the result is that *when one electron sits on nucleus a, the other sits on b. These electrons have parallel spins – everything is prepared for the reaction*.

Therefore, the two ethylene molecules, when excited to the triplet state, open their closed-shells in such a way that favours cycloaddition.

14.5.11 BARRIER MEANS A COST OF OPENING THE CLOSED-SHELLS

Now we can answer more precisely the question of what happens when two molecules react. When the molecules are of the closed-shell character, first a change of their electronic structure has to take place. For that to happen, the kinetic energy of molecular collisions (the temperature plays important role) has to be sufficiently high in order to push and distort[71] the nuclear framework, together with the electron cloud of each of the partners (kinetic energy contra valence repulsion described in Chapter 13), to such an extent that the new configuration already corresponds to that behind the reaction barrier. For example, in the case of an electrophilic or nucleophilic attack, these changes correspond to the transformation $D \to D^+$ and $A \to A^-$, while in the case of the cycloaddition to the excitation of the reacting molecules, to their triplet states. *These changes make the unpaired*

---

[70] We have to do with a four-membered ring, therefore the sp^3 hybrids match the bond directions only roughly.

[71] Two molecules cannot occupy the same volume due to the Pauli exclusion principle, cf. p. 744.

electrons move to the proper reaction centres. As long as this state is not achieved, the changes within the molecules are small and, at most, a molecular complex forms, in which the partners preserve their integrity and their main properties. The profound changes follow from a quasi-avoided crossing of the DA diabatic hypersurface with an excited-state diabatic hypersurface, the excited state being to a large extent a "*picture of the product*". Even the noble gases open their electronic shells when subject to extreme conditions. For example, xenon atoms under pressure of about 150 GPa[72] change their electronic structure so much,[73] that their famous closed-shell electronic structure ceases to be the ground-state. The energy of some excited states lowers so much that the xenon atoms begin to exist in the *metallic state*.

> Reaction barriers appear because the reactants have to open their valence shells and prepare themselves to form new bonds. This means their energy goes up until the "right" excited structure (i.e. the one which resembles the products) lowers its energy so much that the system slides down the new diabatic hypersurface to the product configuration.

The right structure means the electronic distribution in which, for each to-be-formed chemical bond, there is a set of two properly localized unpaired electrons. *The barrier height depends on the energetic gap between the starting structure and the excited state which is the "picture" of the products*. By proper distortion of the geometry (due to the valence repulsion with neighbours) we achieve a "pulling down" of the excited state mentioned, but the same distortion causes the ground state to go up. The larger the initial energy gap, the harder to make the two states interchange their order. The reasoning is supported by the observation that the barrier height for electrophilic or nucleophilic attacks is roughly proportional to the *difference between the donor ionization energy and the acceptor electronic affinity, while the barrier for cycloaddition increases with the excitation energies of the donor and acceptor to their lowest triplet states*. Such relations show the great interpretative power of the acceptor–donor formalism. We would not see this in the VB picture, because it would be difficult to correlate the VB structures based on the atomic orbitals with the ionization potentials or the electron affinities of the molecules involved. The best choice is to look at all three pictures (MO, AD and VB) simultaneously. This is what we have done.

---

[72] M.I. Eremetz, E.A. Gregoryantz, V.V. Struzhkin, H. Mao, R.J. Hemley, N. Mulders, N.M. Zimmerman, *Phys. Rev. Letters* 85 (2000) 2797. The xenon was metallic in the temperature range 300 K–25 mK. The pioneers of these investigations were R. Reichlin, K.E. Brister, A.K. McMahan, M. Ross, S. Martin, Y.K. Vohra, A.L. Ruoff, *Phys. Rev. Letters* 62 (1989) 669.

[73] Please recall the Pauli Blockade, p. 722. Space restrictions for an atom or molecule by the excluded volume of other atoms, i.e. mechanical pushing leads to changes in its electronic structure. These changes may be very large under high pressure.

14.6 BARRIER FOR THE ELECTRON-TRANSFER REACTION

In the AD theory, a chemical reaction of two closed-shell entities means opening their electronic shells (accompanied by an energy cost), and then forming the new bonds (accompanied by an energy gain). The electronic shell opening might have been achieved in two ways: either an electron transfer from the donor to the acceptor, or an excitation of each molecule to the triplet state and subsequent electron pairing between the molecules.

Now we will be interested in the barrier *height* when the first of these possibilities occurs.

14.6.1 DIABATIC AND ADIABATIC POTENTIAL

Example 4. *Electron transfer in* $H_2^+ + H_2$

Let us imagine two molecules, H_2^+ and H_2, in a parallel configuration[74] at distance R from one another and having identical length 1.75 a.u. (Fig. 14.21.a). The value chosen is the arithmetic mean of the two equilibrium separations (2.1 a.u. for H_2^+, 1.4 a.u. for H_2).

There are two geometry parameters to change (Fig. 14.21): the length q_1 of the left (or "first") molecule and the length q_2 of the right (or "second") molecule. Instead of these two variables we may consider the other two: their sum and their difference. Since our goal is to be as simple as possible, we will assume,[75] that $q_1 + q_2 = \text{const}$, and therefore the geometry of the total nuclear framework may be described by a single variable: $q = q_1 - q_2$, with $q \in (-\infty, \infty)$.

It is quite easy to imagine, what happens when q changes from 0 (i.e. from both bonds of equal length) to a somewhat larger value. Variable $q = q_1 - q_2 > 0$ means that $q_1 > q_2$, therefore when q increases, the energy of the system will decrease, because the H_2^+ molecule elongates, while the H_2 shortens, i.e. both approach their equilibrium geometries. If q increases further, it will soon reach the value $q = q_0 = 2.1 - 1.4 = 0.7$ a.u., the optimum value for both molecules. A further increase of q will mean, however, a kind of discomfort for each of the molecules and the energy will go up, for large q – very much up. This means that the potential energy $E(q)$ has a parabola-like shape.

And what will happen for $q < 0$? *It depends on where the extra electron resides.* If it is still on the second molecule (which means it is H_2), then $q < 0$ means an elongation of an already-too-long H_2 and a shortening of an already-too-short H_2^+. The potential energy goes up and the total plot is similar to a parabola with the minimum at $q = q_0 > 0$. If, however, we assume that the extra electron resides all the time on the first of the molecules, then we will obtain the identical parabola-like curve as before, but with the minimum position at $q = -q_0 < 0$.

---

[74] We freeze all the translations and rotations.

[75] The assumption stands to reason, because a shortening of one molecule will be accompanied by an almost identical lengthening of the other, when they exchange an electron.

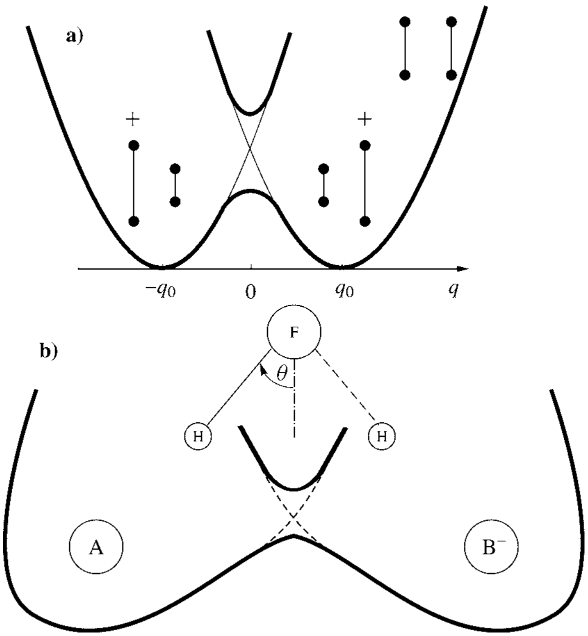


Fig. 14.21. An electron transfer is accompanied by a geometry change. (a) When H_2 molecule gives an electron to H_2^+, both undergo some geometry changes. Variable q equals the difference of the bond lengths of both molecules. At $q = \pm q_0$ both molecules have their optimum bond lengths. (b) The HF pendulum oscillates between two sites, A and B, which accommodate an extra electron becoming either A^-B or AB^-. The curves similar to parabolas denote the energies of the diabatic states as functions of the pendulum angle θ. The thick solid line indicates the adiabatic curves.

DIABATIC AND ADIABATIC POTENTIALS:
Each of these curves with a single minimum represents the *diabatic* potential energy curve for the motion of the nuclei. If, when the donor-acceptor distance changes, the electron *keeps pace* with it and jumps on the acceptor, then increasing or decreasing q from 0 gives a similar result: we obtain a single electronic ground-state potential energy curve with *two* minima in positions $\pm q_0$. This is the *adiabatic* curve.

Whether the adiabatic or diabatic potential has to be applied is equivalent to asking *whether the electron will keep pace (adiabatic) or not (diabatic) with the motion of the nuclei.*[76] This is within the spirit of the adiabatic approximation, cf. Chapter 6, p. 253. Also, a diabatic curve corresponding to the same electronic

---

[76] In the reaction $H_2^+ + H_2 \rightarrow H_2 + H_2^+$ the energy of the reactants is equal to the energy of the products, because the reactants and the products represent the same system. Is it therefore a kind of fiction? Is there any reaction at all taking place? From the point of view of a bookkeeper (thermodynamics) no reaction took place, but from the point of view of a molecular observer (kinetics) – such a reaction may take place. It is especially visible, when instead of one of the hydrogen atoms we use deuterium, then the reaction $HD^+ + H_2 \rightarrow HD + H_2^+$ becomes real even for the bookkeeper (mainly because of the difference in the zero-vibration energies of the reactants and products).

structure (the extra electron sitting on one of the molecules all the time) is an analogue of the diabatic hypersurface that preserved the same chemical bond pattern encountered before.

Example 5. *The "HF pendulum"*

Similar conclusions come from another ideal system, namely the hydrogen fluoride molecule treated as the pendulum of a grandfather clock (the hydrogen atom down, the clock axis going through the fluorine atom) moving over two molecules: A and B, one of them accommodates an extra electron (Fig. 14.21.b).

The electron is negatively charged, the hydrogen atom in the HF molecule carries a partial positive charge, and both objects attract each other. If the electron sits on the left-hand molecule and during the pendulum motion does not keep pace,[77] the potential energy has a single minimum for the angle $-\theta_0$ (the diabatic potential might be approximated by a parabola-like curve with the minimum at $-\theta_0$). An analogous curve with the minimum at θ_0 arises, when the electron resides on B all the time. When the electron keeps pace with any position of the pendulum, we have a single adiabatic potential energy curve with two minima: at $-\theta_0$ and θ_0.

14.6.2 MARCUS THEORY

Rudolph Arthur Marcus (b. 1923), American chemist, professor at the University of Illinois in Urbana and at California Institute of Technology in Pasadena. In 1992 Marcus received the Nobel Prize "*for his contribution to the theory of electron transfer reactions in chemical systems*".

The contemporary theory of the electron transfer reaction was proposed by Rudolph Marcus.[78] The theory is based to a large extent on the harmonic approximation for the diabatic potentials involved, i.e. the diabatic curves represent parabolas. One of the parabolas corresponds to the reactants $V_R(q)$, the other to the products $V_P(q)$ of the electron transfer reaction (Fig. 14.22).[79]

Now, let us assume that both parabolas have the same curvature (force constant f).[80] The reactants correspond to the parabola with the minimum at q_R (without loosing generality we adopt a convention that at $q_R = 0$ the energy is equal zero)

$$V_R(q) = \frac{1}{2} f (q - q_R)^2,$$

while the parabola with the minimum at q_P is shifted in the energy scale by ΔG^0

---

[77] That is, does not jump over to the right-hand side molecule.

[78] The reader may find a good description of the theory in a review article by P.F. Barbara, T.J. Meyer, M.A. Ratner, *J. Phys. Chem.* 100 (1996) 13148.

[79] Let the mysterious q be a single variable for a while, whose deeper meaning will be given later. In order to make the story more concrete let us think about two reactant molecules (R) that transform into the product molecules (P): $A^- + B \rightarrow A + B^-$.

[80] This widely used assumption is better fulfilled for large molecules when one electron more or less does not change much.

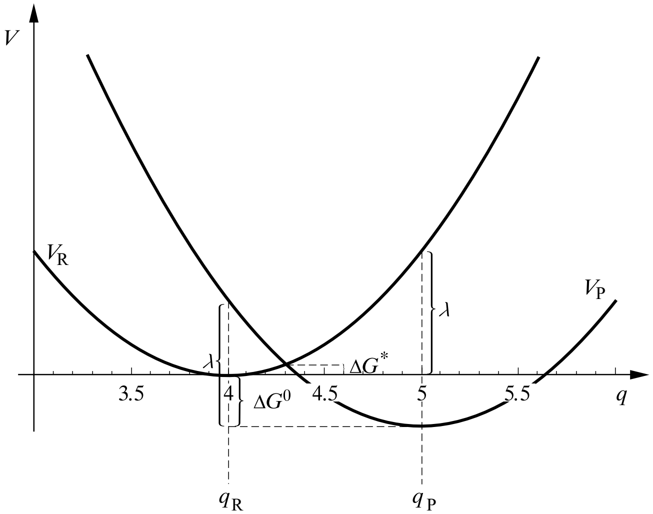


Fig. 14.22. The Marcus theory is based on two parabolic diabatic potentials $V_R(q)$ and $V_P(q)$ for the reactants and products, having minima at q_R and q_P, respectively. The quantity $\Delta G^0 \equiv V_P(q_P) - V_R(q_R)$ represents the energy difference between the products and the reactants, the reaction barrier $\Delta G^* \equiv V_R(q_c) - V_R(q_R) = V_R(q_c)$, where q_c corresponds to the intersection of the parabolas. The reorganization energy $\lambda \equiv V_R(q_P) - V_R(q_R) = V_R(q_P)$ represents the energy expense for making the geometry of the reactants identical with that of the products (and *vice versa*).

($\Delta G^0 < 0$ corresponds to an exothermic reaction[81]).

$$V_P(q) = \frac{1}{2} f (q - q_P)^2 + \Delta G^0.$$

So far we just treat the quantity ΔG^0 *as a potential energy difference* $V_P(q_P) - V_R(q_R)$ *of the model system under consideration* ($H_2^+ + H_2$ or the “pendulum” HF), *even though the symbol suggests that this interpretation will be generalized in the future.*

Such parabolas represent a simple situation.[82] The parabolas’ intersection point q_c satisfies by definition $V_R(q_c) = V_P(q_c)$. This gives

$$q_c = \frac{\Delta G^0}{f} \frac{1}{q_P - q_R} + \frac{q_P + q_R}{2}.$$

Of course on the parabola diagram, the two minima are the most important, the intersection point q_c and the corresponding energy, which represents the reaction barrier reactants → products.

ET barrier

MARCUS FORMULA:
The electron-transfer reaction barrier is calculated as

$$\Delta G^* = V_R(q_c) = \frac{1}{4\lambda} \left(\lambda + \Delta G^0\right)^2, \qquad (14.74)$$

---

[81] That is, the energy of the reactants is higher than the energy of the products (as in Fig. 14.22).

[82] If the curves did not represent parabolas, we might have serious difficulties. This is why we need harmonicity.

reorganization energy

where the energy λ (*reorganization energy*) represents the energy difference between the energies of the products in the equilibrium configuration of the reactants $V_P(q_R)$ and the energy in the equilibrium configuration of the products $V_P(q_P)$:

$$\lambda = V_P(q_R) - V_P(q_P) = \frac{1}{2} f(q_R - q_P)^2 + \Delta G^0 - \Delta G^0 = \frac{1}{2} f(q_R - q_P)^2.$$

The reorganization energy is therefore always positive (energy expense).

REORGANIZATION ENERGY:
Reorganization energy is the energy cost needed for making products in the nuclear configuration of the reactants.

If we ask about the energy needed to transform the optimal geometry of the reactants into the optimal geometry of the products, we obtain the same number. Indeed, we immediately obtain $V_R(q_P) - V_R(q_R) = \frac{1}{2} f(q_R - q_P)^2$, which is the same as before. Such a result is a consequence of the harmonic approximation and the same force constant assumed for V_R and V_P, and shows that this is the energy cost needed to stretch a harmonic string from the equilibrium q_P position to the final q_R position (or *vice versa*). It is seen that the barrier for the thermic electron transfer reaction is higher if the geometry change is wider for the electron transfer [large $(q_R - q_P)^2$] and if the system is stiffer (large f).

Svante August Arrhenius (1859–1927), Swedish physical chemist and astrophysicist, professor at the Stockholm University, originator of the electrolytic theory of ionic dissociation, measurements of the temperature of planets and of the solar corona, also of the theory deriving life on Earth from outer space. In 1903 he received the Nobel Prize in chemistry "*for the services he has rendered to the advancement of chemistry by his electrolytic theory of dissociation*".

From the Arrhenius theory the electron transfer reaction rate constant reads as

$$k_{\mathrm{ET}} = A e^{-\frac{(\lambda + \Delta G^0)^2}{4\lambda k_B T}}. \tag{14.75}$$

How would the reaction rate change, if parabola $V_R(q)$ stays in place, while parabola $V_P(q)$ moves with respect to it? In experimental chemistry this corresponds to a *class* of the chemical reactions $A^- + B \rightarrow A + B^-$, with A (or B) from a homological series of compounds. The homology suggests that the parabolas are similar, because the mechanism is the same (the reactions proceed similarly), and the situations considered differ only by a lowering the second parabola with respect to the first. We may have four qualitatively different cases, eq. (14.74):

Case 1: If the lowering is zero, i.e. $\Delta G^0 = 0$, the reaction barrier is equal to $\lambda/4$ (Fig. 14.23.a).

Case 2: Let us consider an exothermic electron transfer reaction ($\Delta G^0 < 0$, $|\Delta G^0| < \lambda$). In this case the reaction barrier is lower, because of the subtraction in the exponent, and the reaction rate *increases* (Fig. 14.23.b). Therefore the $-\Delta G^0$ is the "driving force" in such reactions.

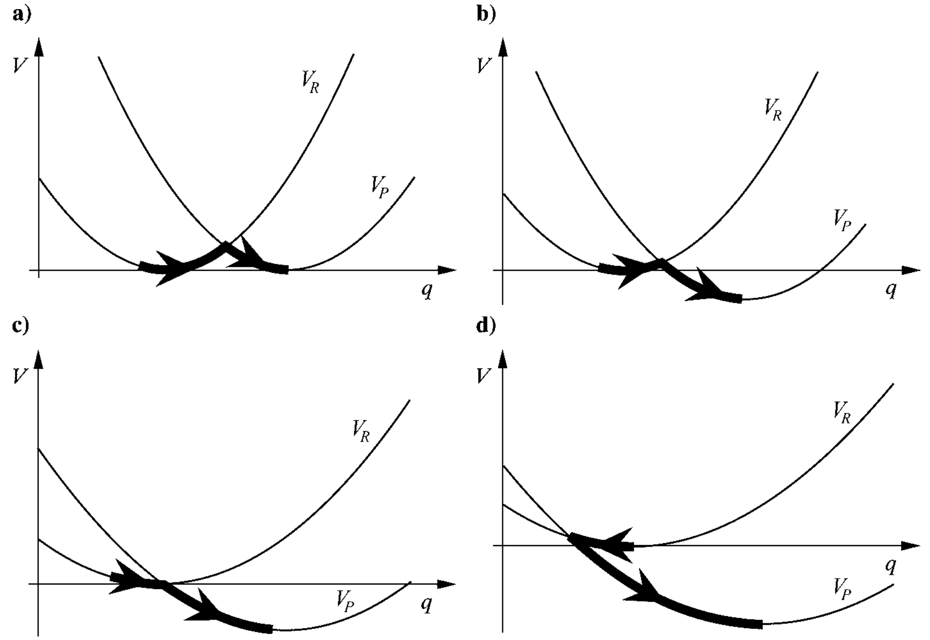


Fig. 14.23. Four qualitatively different cases in the Marcus theory. (a) $\Delta G^0 = 0$, hence $\Delta G^* = \frac{\lambda}{4}$. (b) $|\Delta G^0| < \lambda$ (c) $|\Delta G^0| = \lambda$ (d) inverse Marcus region $|\Delta G^0| > \lambda$.

Case 3: When the $|\Delta G^0|$ keeps increasing, at $|\Delta G^0| = \lambda$ the reorganization energy cancels the driving force, and the barrier vanishes to zero. Note that this represents the highest reaction rate possible (Fig. 14.23.c).

Case 4: Inverse Marcus region (Fig. 14.23.d). Let us imagine now that we keep increasing the driving force. We have a reaction for which $\Delta G^0 < 0$ and $|\Delta G^0| > \lambda$. Compared to the previous case, the *driving force has increased, whereas the reaction rate decreases.* This might look like a possible surprise for experimentalists. A case like this is called the *inverse Marcus region*, foreseen by Marcus in the sixties, using the two parabola model. People could not believe this prediction until experimental proof[83] in 1984.

inverse Marcus region

New meaning of the variable q

Let us make a subtraction:

$$\begin{aligned} V_R(q) - V_P(q) &= f(q-q_R)^2/2 - f(q-q_P)^2/2 - \Delta G^0 \\ &= \frac{f}{2}[2q - q_R - q_P][q_P - q_R] - \Delta G^0 = Aq + B \qquad (14.76) \end{aligned}$$

where A and B represent constants. This means that

---

[83] J.R. Miller, L.T. Calcaterra, G.L. Closs, *J. Am. Chem. Soc.* 97 (1984) 3047.

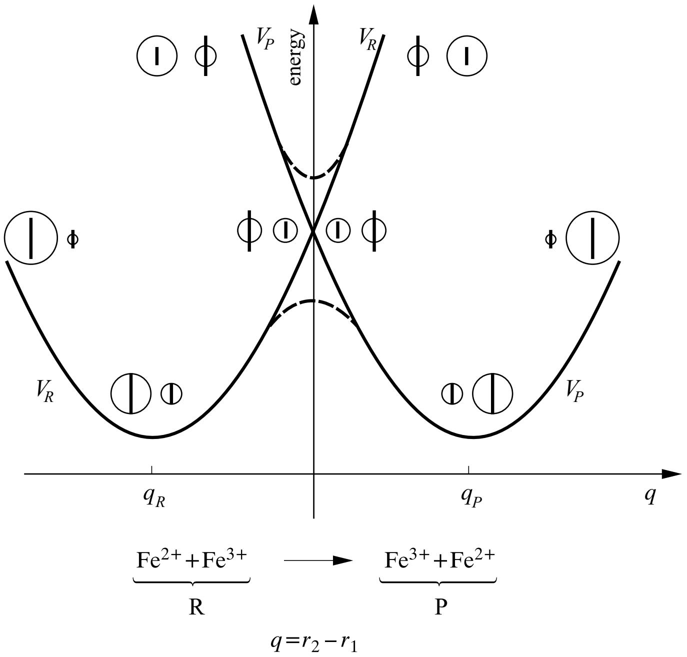


Fig. 14.24. The diabatic potential energy curves (V_R for the reactants and V_P for the products) pertaining to the electron transfer reaction $Fe^{2+} + Fe^{3+} \rightarrow Fe^{3+} + Fe^{2+}$ in aqueous solution. The curves depend on the variable $q = r_2 - r_1$ that describes the *solvent*, which is characterized by the radius r_1 of the cavity for the first (say, left) ion and by the radius r_2 of the cavity for the second ion. For the sake of simplicity we assume $r_1 + r_2 = \text{const}$ and equal to the sum of the ionic radii of Fe^{2+} and Fe^{3+}. For several points q the cavities were drawn as well as the vertical sections that symbolize the diameters of the left and right ions. In this situation, the plots V_R and V_P have to differ widely. The dashed lines represent the adiabatic curves (in the peripheral sections they coincide with the diabatic curves).

the diabatic potential energy difference depends *linearly* on coordinate q.

In other words for a given electron transfer reaction either q *or* $V_R(q) - V_P(q)$ *represents the same information.*

The above examples and derivations pertain to a one-dimensional model of electron transfer (a single variable q), while in reality (imagine a solution) the problem pertains to a huge number of variables. What happens here? Let us take the example of electron transfer between Fe^{2+} and Fe^{3+} ions in an aqueous solution $Fe^{2+} + Fe^{3+} \rightarrow Fe^{3+} + Fe^{2+}$ (Fig. 14.24)[84]

The solvent behaviour is of key importance for the electron-transfer process.

---

[84] In this example $\Delta G^0 = 0$, i.e. case 1 considered above.

The ions Fe^{2+} and Fe^{3+} are hydrated. For the reaction to proceed, the solvent has to *reorganize* itself next to both ions. The hydration shell of Fe^{2+} ion is of larger radius than the hydration shell of Fe^{3+} ion, because Fe^{3+} is smaller than Fe^{2+} and, in addition, creates a stronger electric field due to its higher charge (Fig. 14.24). Both factors add to a stronger association of the water molecules with the Fe^{3+} ion than with Fe^{2+}. In a crude approximation, the state of the solvent may be characterized by two quasi-rigid cavities, say: left and right (or, numbers 1 and 2) that could accommodate the ions. Let us assume the cavities have radii r_1 and r_2, whereas the ionic radii are $r_{\mathrm{Fe2+}}$ and $r_{\mathrm{Fe3+}}$ with $r_{\mathrm{Fe2+}} > r_{\mathrm{Fe3+}}$. Let us assume, for the sake of simplicity, that $r_1 + r_2 = r_{\mathrm{Fe2+}} + r_{\mathrm{Fe3+}} = \mathrm{const}$ and introduce a single variable $q = r_2 - r_1$ that in this situation characterizes the state of the solvent. Let us see what happens when q changes.

We first consider that the extra electron sits on the left ion all the time (reactant curve V_R) and the variable q is a negative number (with a high absolute value, i.e. $r_1 \gg r_2$). As seen from Fig. 14.24, the energy is very high, because the solvent squeezes the Fe^{3+} ion out (the second cavity is too small). It does not help that the Fe^{2+} ion has a lot of space in its cavity. Now we begin to move towards higher values of q. The first cavity begins to shrink, for a while without any resistance from the Fe^{2+} ion, the second cavity begins to lose its pressure thus making Fe^{3+} ion more and more happy. The energy decreases. Finally we reach the minimum of V_R, at $q = q_R$ and the radii of the cavities match the ions perfectly. Meanwhile variable q continues to increase. Now the solvent squeezes the Fe^{2+} ion out, while the cavity for Fe^{3+} becomes too large. The energy increases again, mainly because of the first effect. We arrive at $q = 0$. The cavities are of equal size, but do not match either of the ions. This time the Fe^{2+} ion experiences some discomfort, and after passing the point $q = 0$ the pain increases more and more, and the energy continues to increase. The whole story pertains to extra electron sitting on the left ion all the time (no jump, i.e. the reactant situation). A similar dramatic story can be told when the electron is sitting all the time on the right ion (products situation). In this case we obtain the V_P plot.

The V_R and V_P plots just described represent the diabatic potential energy curves for the motion of the nuclei, valid for the extra electron residing on the same ion all the time. Fig. 14.24 also shows the adiabatic curve (dashed line) when the extra electron has enough time to adjust to the motion of the approaching nuclei and the solvent, and jumps at the partner ion.

Taking a single parameter q to describe the electron transfer process in a solvent is certainly a crude simplification. Actually there are billions of variables in the game describing the degrees of freedom of the water molecules in the first and further hydration shells. One of the important steps towards successful description of the electron transfer reaction was the Marcus postulate,[85] that

---

[85] Such collective variables are used very often in every-day life. Who cares about all the atomic positions when studying a ball rolling down an inclined plane? Instead, we use a single variable (the position of the centre of the ball), which gives us a perfect description of the system in a certain energy range.

> despite the multidimensionality of the problem, eq. (14.76) *is still valid*, i.e. $V_R - V_P$ is a *single* variable describing the position of system on the electron-transfer reaction path (it is therefore a collective coordinate that describes the positions of the solvent molecules).

No doubt the potential energy value is important, but how often can this value be reached by the system, is equally important. This is connected to the *width* of the low-energy basin associated with the *entropy*[86] and to the *free energy*. In statistical thermodynamics we introduce the idea of the *potential of the mean force*, related to the free energy. Imagine a system in which we have two motions on different time scales: fast (e.g., of small solvent molecules) and slow (e.g., which change the shape of a macromolecule). To focus on the slow motion, we average the energy over the fast motions (the Boltzmann factor will be needed, which will introduce a temperature dependence on the resulting energy). In this way, from the potential energy we obtain the mean force potential depending only on the slow variables, sometimes called the free energy (which is a function of geometry of the macromolecule), cf. p. 293.

mean force potential

> The second Marcus assumption is that the ordinate axis should be treated as the *mean force potential*, or the free energy rather than just potential energy.

It is very rare in theoretical chemistry[87] that a many-dimensional problem can be transformed to a single variable problem. This is why the Marcus idea described above of a collective coordinate, provokes the reaction: “no way”. However, as it turned out later, this simple postulate lead to a solution that grasps the essential features of electron transfer.

What do the Marcus parabolas mean?

The example just considered of the electron transfer reaction: $\mathrm{Fe}^{2+} + \mathrm{Fe}^{3+} \rightarrow \mathrm{Fe}^{3+} + \mathrm{Fe}^{2+}$ reveals that in this case the reaction barrier is controlled by the solvent, i.e. by billions of coordinates. As shown by Marcus, this plethora can be effectively replaced by a single collective variable. Only after this approximation, may we draw the diabatic parabola-like curves. The intersection point of the two

---

[86] A wide potential energy well can accommodate a lot of closely lying vibrational levels and therefore the number of possible states of the system in a given energy range may be huge (large entropy). Please recall the particle-in-a-box problem: the longer the box the closer the energy levels.

[87] The free energy is defined as $F(T) = -kT\frac{\partial}{\partial T}\ln Z$, where $Z = \sum_i \exp(-\frac{E_i}{kT})$ represents the partition function, E_i stands for the i-th energy level. In the classical approach this energy level corresponds to the potential energy $V(x)$, where x represents a point in configurational space, and the sum corresponds to an integral over the total configurational space $Z = \int \mathrm{d}x \exp(-\frac{V}{kT})$. Note that the free energy is a function of temperature only, not of the spatial coordinates x. If however, the integration were only carried out over part of the variables, say, only the fast variables, then Z, and therefore also F, would become a function of the slow variables and of temperature (mean force potential). Despite the incomplete integration, we sometimes use the name “free energy” for this mean force potential by saying that “the free energy is a function of coordinates...”.

diabatic curves can easily be found only after assuming their parabolic character. And yet any collective variable means motion along a line in an extremely complex configurational space (solvent molecules plus reactants). Moving along this line means that, according to Marcus, we encounter the intersection of the ground and excited electronic states. As shown in Chapter 6, such a crossing occurs at the conical intersection. Is it therefore that during the electron transfer reaction, the system goes through the conical intersection point? How to put together such notions as reaction barrier, reaction path, entrance and exit channels, not to speak of acceptor–donor theory? Fig. 14.25.a shows the paraboloid model of the diabatic DA and D^+A^- surfaces, while Fig. 14.25.b shows them in a more realistic way.

- The *diabatic hypersurfaces*, one corresponds to DA (i.e. the extra electron is on the donor all the time) and the second to D^+A^- (i.e. the extra electron resides on the acceptor), undergo the conical intersection. For conical intersection to happen at least three atoms are required. Imagine a simple model, with a diatomic acceptor A and an atom D as donor. Atom D has a dilemma: either to transfer the electron to the first or the second atom of A. This dilemma means conical intersection. Like the coordinate system shown in Fig. 14.25, the variables ξ_1 and ξ_2 described in Chapter 6 were chosen (they lead to splitting of the adiabatic hypersurfaces), which measure the deviation of the donor D with respect to the corner of the equilateral triangle of side equal to the length of the diatomic molecule A. The conical intersection point, i.e. $(0, 0)$ corresponds to the equilateral triangle configuration. The figure also shows the upper and lower cones touching at $(0, 0)$.
- The conical intersection led to two *adiabatic* hypersurfaces: lower (electronic ground state) and upper (electronic excited state). Each of the adiabatic hypersurfaces shown in Fig. 14.25.b consists of the "reactant half" (the diabatic state of the reactants, DA) and the "product half" (the diabatic state of the products, D^+A^-). The border between them reveals the intersection of the two diabatic states and represents the line of change of the electronic structure reactants/products. Crossing the line means the chemical reaction happens.
- The "avoided crossing" occurs everywhere along the border except at the conical intersection. It is improbable that the reactive trajectory passes through the conical intersection, because it usually corresponds to higher energy. It will usually pass at a distance from the conical intersection and this resembles an avoided crossing. This is why we speak of the avoided crossing in a polyatomic molecule, whereas the concept pertains to diatomics only.
- Passing the border is easiest at *two* points. These are the two saddle points (barriers I and II). A thermic electron transfer reaction goes through one of them, the corresponding IRCs are denoted by dotted lines. In each case we obtain different products. Both saddle points differ in that D, when attacking A has the choice of joining either of the two ends of A, usually forming *two different* products. We therefore usually have *two* barriers. In the example given (H_3) they are identical, but in general they may differ. When the barrier heights are equal because of symmetry, it does not matter which is overcome. When they are dif-

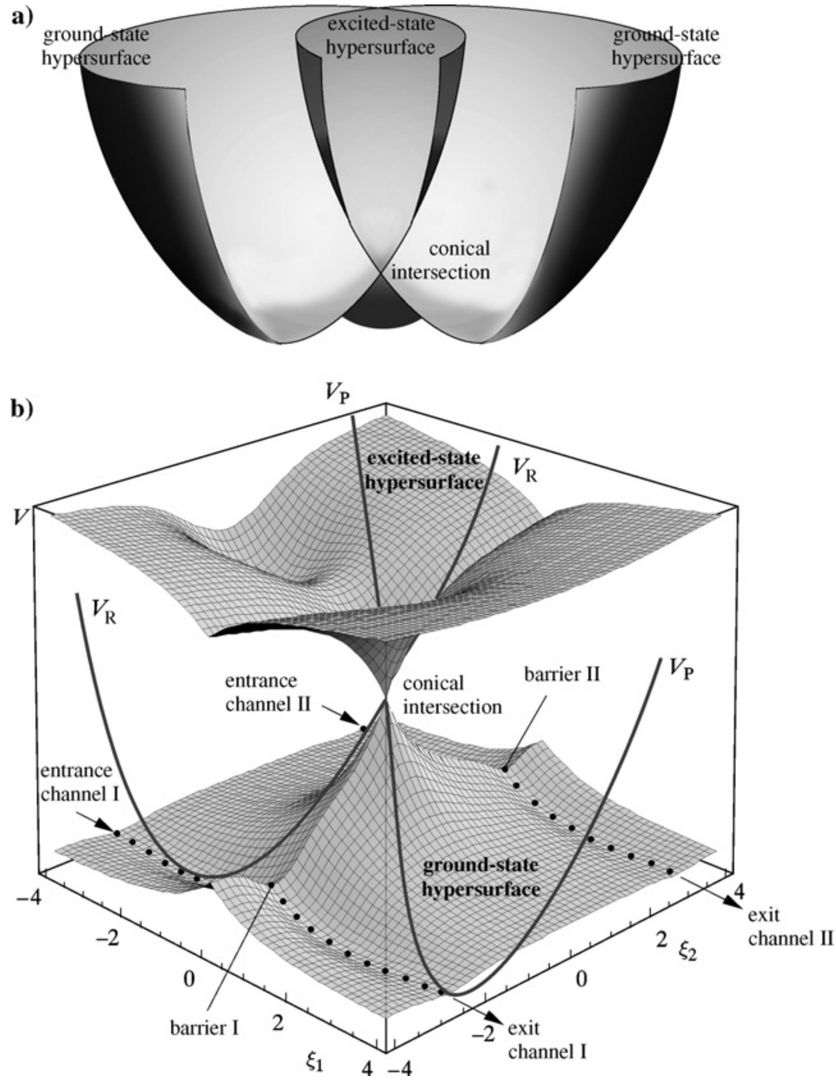


Fig. 14.25. Electron transfer in the reaction $DA \rightarrow D^+A^-$ as well as the relation of the Marcus parabolas to the concepts of the conical intersection, diabatic and adiabatic states, entrance and exit channels and the reaction barrier. Fig. (a) shows two diabatic (and adiabatic) surfaces of the electronic energy as functions of the ξ_1 and ξ_2 variables that describe the deviation from the conical intersection point (cf. p. 262). Both diabatic surfaces are shown schematically in the form of the two paraboloids: one for the reactants (DA), the second for products (D^+A^-). The region of the conical intersection is also indicated. Fig. (b) also shows the conical intersection, but the surfaces are presented more realistically. The upper and lower parts of Fig. (b) touch at the conical intersection point. On the lower part of the surface we can see two reaction channels each with its reaction barrier (see the text), on the upper part (b) an energy valley is shown that symbolizes a bound state that is separated from the conical intersection by a reaction barrier.

ferent, one of them dominates (usually the lower barrier[88]). The channels shown in the figure are not curved, because we use a coordinate system different from that used in the collinear reaction.

- The Marcus parabolas represent a special section (along the collective variable) of the hypersurfaces passing through the conical intersection (parabolas V_R and V_P in Fig. 14.25.b). Each parabola represents a diabatic state, therefore a part of each reactant parabola is on the lower hypersurface, while the other one on the upper hypersurface. We see that the parabolas are only an approximation to the hypersurface profile. The reaction is of a thermic character, and as a consequence, the parabolas should not pass through the conical intersection, because it corresponds to high energy, instead it passes through one of the saddle points.
- The "product half" of the excited state hypersurface runs up to the "reactant half" of the ground state hypersurface and *vice versa*. This means that photoexcitation (following the Franck–Condon rule this corresponds to a vertical excitation) means a profound change: the system looks as if it has already reacted (*photoreaction*). photoreaction

Quantum mechanical modification

In Marcus formula (14.74) we assume that in order to make the electron transfer effective, we have to supply at least the energy equal to the barrier height. The formula does not obviously take into account the quantum nature of the transfer. The system may overcome the barrier not only by having energy higher than the barrier, but also by tunnelling, when its energy is lower than the barrier height (cf. p. 153). Besides, the reactant and product energies are quantized (vibrational-rotational levels[89]). The reactants may be excited to one of such levels. The reactant vibrational levels will have different abilities to tunnel.

According to Chapter 2 only a time-dependent perturbation is able to change the system's energy. Such a perturbation may serve the electric field of the electromagnetic wave. When the perturbation is periodic, with the angular frequency ω matching the energy difference of initial state k and one of the states of higher energy (n), then the transition probability between these states is equal to: $P_k^n(t) = \frac{2\pi t}{\hbar}|v_{kn}|^2\ \delta(E_n^{(0)} - E_k^{(0)} - \hbar\omega)$ (the Fermi golden rule, eq. (2.23), p. 85 is valid for relatively short times t), where $v_{kn} = \langle k|v|n\rangle$, with $v(\boldsymbol{r})$ representing the perturbation amplitude,[90] $V(\boldsymbol{r}, t) = v(\boldsymbol{r})e^{i\omega t}$. The Dirac delta function δ is a quantum-mechanical way of saying that the total energy has to be conserved. In phototransfer of the electron, state "k" represents the quantum mechanical state of the reactants, and "n" – a product state, each of diabatic character.[91] In practice the adiabatic approximation is used, in which the reactant and product wave

---

[88] There may be some surprises. Barrier height is not all that matters. Sometimes it may happen that what decides is access to the barrier region, in the sense of its width (this is where the entropy and free energy matter).

[89] For large molecules, we may forget the rotational spectrum, since, because of the large moment of inertia, the rotational states form a quasi-continuum ("no quantization").

[90] $\boldsymbol{r}$ stands for those variables on which the wave functions depend.

[91] They will be denoted by the subscripts R and P.

functions are products of the electronic wave functions (which depend on the electronic coordinates $\boldsymbol{r}$ and, parametrically, on the nuclear configuration $\boldsymbol{R}$) and the vibrational functions $f(\boldsymbol{R})$ describing the motion of the nuclei: $\psi_{k,R}(\boldsymbol{r};\boldsymbol{R})f_{v_1,R}(\boldsymbol{R})$ and $\psi_{n,P}(\boldsymbol{r};\boldsymbol{R})f_{v_2,P}(\boldsymbol{R})$. The indices v_1 and v_2 in functions f denote the vibrational quantum numbers.

Then, the transition probability depends on the integral (Chapter 2)

$$v_{kn}=\langle\psi_{k,R}(\boldsymbol{r};\boldsymbol{R})f_{v_1,R}(\boldsymbol{R})|v(\boldsymbol{r})|\psi_{n,P}(\boldsymbol{r};\boldsymbol{R})f_{v_2,P}(\boldsymbol{R})\rangle.$$

Let us rewrite it, making the integration over the nuclear and electronic coordinates explicit (where $\mathrm{d}V_{\mathrm{nucl}}$ and $\mathrm{d}\tau_e$ mean that the integrations is over the nuclear and electronic coordinates, respectively)

$$v_{kn}=\int \mathrm{d}V_{\mathrm{nucl}}\, f^*_{v_1,R}(\boldsymbol{R})f_{v_2,P}(\boldsymbol{R})\int \mathrm{d}\tau_e\, \psi^*_{k,R}(\boldsymbol{r};\boldsymbol{R})v(\boldsymbol{r})\psi_{n,P}(\boldsymbol{r};\boldsymbol{R}).$$

Now, let us use the Franck–Condon approximation that the optical perturbation makes the electrons move instantaneously while the nuclei do not keep pace with the electrons and stay in the same positions (we assume therefore equilibrium positions of the nuclei $\boldsymbol{R}_0$ in the reactants):

$$v_{kn}\approx\int \mathrm{d}V_{\mathrm{nucl}}\, f^*_{v_1,R}(\boldsymbol{R})f_{v_2,P}(\boldsymbol{R})\int \mathrm{d}\tau_e\, \psi^*_{kR}(\boldsymbol{r};\boldsymbol{R}_0)v(\boldsymbol{r})\psi_{n,P}(\boldsymbol{r};\boldsymbol{R}_0).$$

The last integral therefore represents a constant and therefore

$$v_{kn}=V_{\mathrm{RP}}S_{\mathrm{osc}}(v_1,v_2),$$

where

$$V_{\mathrm{RP}}=\int \mathrm{d}\tau_e\, \psi^*_{k,R}(\boldsymbol{r};\boldsymbol{R}_0)v(\boldsymbol{r})\psi_{n,P}(\boldsymbol{r};\boldsymbol{R}_0),$$

$$S_{\mathrm{osc}}(v_1,v_2)=\int \mathrm{d}V_{\mathrm{nucl}}\, f^*_{v_1,R}(\boldsymbol{R})f_{v_2,P}(\boldsymbol{R}). \tag{14.77}$$

The last integral is called the *Franck–Condon factor*.

Franck–Condon factor

FRANCK–CONDON FACTOR:
A Franck–Condon factor is the overlap integral of the vibrational wave functions: one pertaining to the reactants with $\mathrm{d}V_{\mathrm{nucl}}$ vibrational quantum number v_1 and the second, pertaining to the products with vibrational quantum number v_2.

The calculation of V_{RP} is not an easy matter, we prefer often therefore an empirical approach by modelling the integral as[92]

$$V_{RP} = V_0 \exp\bigl[-\beta(R - R_0)\bigr],$$

where R_0 stands for the van der Waals distance of the donor and acceptor, R represents their distance, $\beta > 0$ represents a constant and V_0 means V_{RP} for the van der Waals distance.[93]

A large Franck–Condon factor means that by exciting the reactants to the vibrational state v_1 there is a particularly high probability for the electron transfer (by tunnelling) with the products in vibrational state v_2.

Reorganization energy

In the Marcus formula, reorganization energy plays an important role. This energy is the main reason for the electron-transfer reaction barrier.

The reorganization pertains to the neighbourhood of the transferred electron,[94] i.e. to the solvent molecules, but also to the donors and acceptors themselves.[95] This is why the reorganization energy, in the first approximation, consists of the internal reorganization energy (λ_i) that pertains to the donor and acceptor molecules, and of the solvent reorganization energy (λ_0):

$$\lambda = \lambda_i + \lambda_0.$$

Internal reorganization energy. For the electron to have the chance of jumping from molecule A^- to molecule[96] B, it has to have the neighbourhood reorganized in a special way. The changes should make the extra electron's life hard on A^- (together with solvation shells) and seduce it by the alluring shape of molecule B and its solvation shells. To do this, work has to be done. First, this is an energy cost for the proper deformation of A^- to the geometry of molecule A, i.e. already without the extra electron (the electron obviously does not like this – this is how it is forced out). Next, molecule B is deformed to the geometry of B^- (this is what

---

[92] Sometimes the dependence is different. For example, in *Twisted Intramolecular Charge Transfer* (TICT), after the electron is transferred between the donor and acceptor moieties (a large V_{RP}) the molecule undergoes an internal rotation of the moieties, which causes an important decreasing of the V_{RP} [K. Rotkiewicz, K.H. Grellmann, Z.R. Grabowski, *Chem. Phys. Letters* 19 (1973) 315].

[93] As a matter of fact, such formulae only contain a simple message: V_{RP} decreases *very* fast when the donor and acceptor distance increases.

[94] The neighbourhood is adjusted perfectly to the extra electron (to be transferred) in the reactant situation, and very unfavourable for its future position in the products. Thus the neighbourhood has to be reorganized to be adjusted for the electron transfer products.

[95] It does not matter for an electron what in particular prevents it from jumping.

[96] "Minus" denotes the site of the extra electron. It does not necessarily mean that A^- represents an anion.

makes B attractive to the extra electron – everything is prepared for it in B). These two energy effects correspond to λ_i.

Calculation of λ_i is simple:

$$\lambda_i = E(\mathrm{A^-B};\ \mathrm{geom\ AB^-}) - \mathrm{E}(\mathrm{A^-B};\ \mathrm{geom\ A^-B})$$

where $E(\mathrm{A^-B}; \mathrm{geom\ AB^-})$ denotes the energy of $\mathrm{A^-B}$ calculated for the equilibrium geometry of another species, namely $\mathrm{AB^-}$, while $E(\mathrm{A^-B}; \mathrm{geom\ A^-B})$ stands for the energy of $\mathrm{A^-B}$ at its optimum geometry.

Usually the geometry changes in $\mathrm{AB^-}$ and $\mathrm{A^-B}$ attain several percent of the bond lengths or the bond angles. The change is therefore relatively small and we may represent it by a superposition of the normal mode vectors[97] $\boldsymbol{L}_k, k = 1, 2, \ldots, 3N$, described in Chapter 7. We may use the normal modes of the molecule $\mathrm{A^-B}$ (when we are interested in electron transfer from $\mathrm{A^-}$ to B) or of the molecule $\mathrm{AB^-}$ (back transfer). What for? Because some normal modes are more effective than others in facilitating electron transfer. The normal mode analysis would show[98] that

> the *most effective normal mode of the reactants deforms them in such a way as to resemble the products*. This vibration reorganizes the neighbourhood in the desired direction (for electron transfer to occur), and therefore effectively lowers the reaction barrier.

Solvent reorganization energy. Spectroscopic investigations are unable to distinguish between the internal or solvent reorganization, because Nature does not distinguish between the solvent and the rest of the neighbourhood. An approximation to the solvent reorganization energy may be calculated by assuming a continuous solvent model. Assuming that the mutual configuration of the donor and acceptor (separated by distance R) allows for enclosing them in non-overlapping spheres of radii a_1 and a_2, the following formula was derived by Marcus:

$$\lambda_0 = (\Delta e)^2 \left\{ \frac{1}{2a_1} + \frac{1}{2a_2} - \frac{1}{R} \right\} \left\{ \frac{1}{\epsilon_\infty} - \frac{1}{\epsilon_0} \right\},$$

where ϵ_∞ and ϵ_0 denote the dielectric screening constants measured at infinite and zero electromagnetic field frequency, respectively, and Δe is equal to the effective electric charge transferred between the donor and acceptor. The dielectric screening constant is related to the polarization of the medium. The value ϵ_0 is

---

[97] Yet the normal modes are linear combinations of the Cartesian displacements.

[98] It usually turns out that there are several such vibrations. They will help electron transfer from $\mathrm{A^-}$ to B. The reason is quite obvious, e.g., the empirical formula for V_{RP} implies that a vibration that makes the AB distance smaller will increase the transfer probability. This can be seen in what is known as resonance Raman spectroscopy close to a charge transfer optical transition. In such spectroscopy, we have the opportunity to observe particular vibronic transitions. The intensity of the vibrational transitions (usually from $v = 0$ to $v = 1$) of those normal modes which facilitate electron transfer will be highest.

larger than ϵ_∞, because, at a constant electric field, the electrons as well as the nuclei (mainly an effect of the reorientation of the molecules) keep pace to adjust to the electric field. At high frequency only the electrons keep pace, hence $\epsilon_\infty < \epsilon_0$. The last parenthesis takes care of the difference, i.e. of the reorientation of the molecules in space (cf. Chapter 12).

Summary

- A chemical reaction represents a molecular catastrophe, in which the electronic structure, as well as the nuclear framework of the system changes qualitatively. Most often a chemical reaction corresponds to the breaking of an old and creation of a new bond.
- Simplest chemical reactions correspond to overcoming single reaction barrier on the way from reactants to products through saddle point along the intrinsic reaction coordinate (IRC). The IRC corresponds to the steepest descent trajectory (in the mass-weighted coordinates) from the saddle point to configurations of reactants and products.
- Such a process may be described as the system passing from the entrance channel (reactants) to the exit channel (products) on the electronic energy map as a function of the nuclear coordinates. For a collinear reaction A + BC → AB + C the map shows a characteristic reaction "drain-pipe". Passing along the "drain-pipe" bottom usually requires overcoming a reaction barrier, its height being a fraction of the energy of breaking the "old" chemical bond.
- The reaction barrier reactants → products, is as a rule, of different height to the corresponding barrier for the reverse reaction.
- We have shown how to obtain an accurate solution for three atom reaction. After introducing the democratic hyperspherical coordinates it is possible to solve the Schrödinger equation (within the Ritz approach). We obtain the rate constant for the state-to-state elementary chemical reaction.

A chemical reaction may be described by the reaction path Hamiltonian in order to focus on the intrinsic reaction coordinate (IRC) measuring the motion along the "drain-pipe" bottom (reaction path) and the normal mode coordinates orthogonal to the IRC.

- During the reaction, energy may be exchanged between the vibrational normal modes, as well as between the vibrational modes and the motion along the IRC.
- Two atoms or molecules may react in many different ways (reaction channels). Even if under some conditions they do not react (e.g., the noble gases), the reason for this is that their kinetic energy is too low with respect to the corresponding reaction barrier, and the opening of their electronic closed shells is prohibitively expensive on the energy scale. If the kinetic energy increases, more and more reaction channels open up, because it is possible for higher and higher energy barriers to be overcome.
- A reaction barrier is a consequence of the "quasi-avoided crossing" of the corresponding diabatic hypersurfaces, as a result we obtain two adiabatic hypersurfaces ("lower" or electronic ground state, and "upper" or electronic excited state). Each of the adiabatic hypersurfaces consists of two diabatic parts stitched along the border passing through the conical intersection point. On both sides of the conical intersection there are usually two saddle points along the border line leading in general to two different reaction products (Fig. 14.25).
- The two intersecting diabatic hypersurfaces (at the reactant configuration) represent (a) the electronic ground state DA (b) and *that electronic excited state that resembles the electronic charge distribution of the products*, usually D^+A^-.

- The barrier appears therefore as the cost of opening the closed shell in such a way as to prepare the reactants for the formation of new bond(s).
- In Marcus electron transfer theory, the barrier also arises as a consequence of the intersection of the two diabatic potential energy curves. The barrier height depends mainly on the (solvent and reactant) reorganization energy.

Main concepts, new terms

critical points (p. 767)
femtosecond spectroscopy (p. 768)
saddle point (p. 768)
steepest descent path (SDP) (p. 769)
reactive and non-reactive trajectories (p. 770)
skew coordinate system (p. 770)
reaction "drain-pipe" (p. 772)
entrance and exit channels (p. 772)
early and late reaction barriers (p. 773)
bobsleigh effect (p. 774)
democratic coordinates (p. 776)
cross section (p. 779)
reaction rate (p. 779)
Berry phase (p. 780)
mass-weighted coordinates (p. 781)
intrinsic reaction coordinate (IRC) (p. 781)
"trajectory-in-molasses" (p. 782)
reaction path Hamiltonian (p. 783)
natural coordinates (p. 784)
vibrationally adiabatic approximation (p. 785)
vibrationally adiabatic potential (p. 786)
Coriolis coupling (p. 785 and 791)
curvature coupling (p. 785 and 791)
exo- and endothermic reactions (p. 787)
donating mode (p. 792)
spectator bond (p. 795)
molecular electrostatic potential (p. 798)
steric effect (p. 799)
acceptor–donor (AD) reaction theory (p. 803)
MO and AD pictures (p. 805)
reaction stages (p. 806)
role of states DA, D$^+$A$^-$, D$^+$A^{-*} (p. 811)
HOMO-LUMO crossing (p. 815)
nucleophilic attack (p. 816)
electrophilic attack (p. 818)
cycloaddition reaction (p. 823)
Woodward–Hoffmann rules (p. 825)
Diels–Alder reaction (p. 825)
diabatic and adiabatic potentials (p. 828)
inverse Marcus region (p. 833)
collective coordinate (p. 836)
mean force potential (p. 836)
Franck–Condon factors (p. 840)
reorganization energy (p. 841)

From the research front

Chemical reactions represent a very difficult problem for quantum chemistry, because:

- There are a lot of possible reaction channels. Imagine the number of all combinations of atoms in a monomolecular dissociation reaction, also in their various electronic states. We have to select first which reaction to choose and a good clue may be the lowest possible reaction barrier.
- A huge change in the electronic structure is usually quite demanding for standard quantum mechanical methods.
- Given a chosen single reaction channel we confront the problem of calculating the potential energy hypersurface. Let us recall (Chapters 6 and 7) the number of quantum mechanical calculations to perform this is of the order of 10^{3N-6}. For as small number of nuclei as $N = 4$ we already have a million computation tasks to perform.
- Despite unprecedented progress in the computational technique, the cutting edge possibilities are limited in *ab initio* calculations to two diatomic molecules.

On the other hand, a chemist always has some additional information on which chemical reactions are expected to occur. Very often the most important changes happen in a limited set of atoms, e.g., in functional groups, their reactivity being quite well understood. Freezing the positions of those atoms which are reaction spectators only, allows us to limit the number of degrees of freedom to consider.

Ad futurum...

Chemical reactions with the reactants precisely oriented in space will be more and more important in chemical experiments of the future. Here it will be helpful to favour some reactions by supramolecular recognition, docking in reaction cavities or reactions on prepared surfaces. For theoreticians, such control of orientation will mean the reduction of certain degrees of freedom. This, together with eliminating or simulating the spectator bonds, may reduce the task to manageable size. State-to-state calculations and experiments that will describe an effective chemical reaction that starts from a given quantum mechanical state of the reactants and ends up with another well defined quantum mechanical state of the products will become more and more important. Even now, we may design with great precision practically any sequence of laser pulses (a superposition of the electromagnetic waves, each of a given duration, amplitude, frequency and phase). For a chemist, this means that we are able to change the shape of the hypersurfaces (ground and excited states) in a controllable way, because every nuclear configuration corresponds to a dipole moment that interacts with the electric field (cf. Chapter 12). The hypersurfaces may shake and undulate in such a way as to make the point representing the system move to the product region. In addition, there are possible excitations and the products may be obtained *via* excited hypersurfaces. As a result we may have selected bonds broken, and others created in a selective and highly efficient way. This technique demands important developments in the field of chemical reaction theory and experiment, because currently we are far from such a goal.

Note that the most important achievements in the chemical reaction theory pertained to concepts (von Neumann, Wigner, Teller, Woodward, Hoffmann, Fukui, Evans, Polanyi, Shaik) rather than computations. The potential energy hypersurfaces are so complicated that it took the scientists fifty years to elucidate their main machinery. Chemistry means first of all chemical reactions, and most chemical reactions still represent *terra incognita*. This will change considerably in the years to come. In the longer term this will be the main area of quantum chemistry.

Additional literature

R.D. Levine, R.B. Bernstein, "Molecular Reaction Dynamics and Chemical Reactivity", Oxford University Press, 1987.

An accurate approach to the reactions of small molecules.

H. Eyring, J. Walter, G.F. Kimball, "Quantum chemistry", John Wiley, New York, 1967.

A good old textbook, written by the outstanding specialists in the field. To my knowledge no later textbook has done it in more detail.

R.B. Woodward, R. Hoffmann, "The Conservation of Orbital Symmetry", Academic Press, New York, 1970.

A summary of the important discoveries made by these authors (Woodward–Hoffmann symmetry rules).

S.S. Shaik “What Happens to Molecules as They React? Valence Bond Approach to Reactivity”, *Journal of the American Chemical Society* 103 (1981) 3692.

An excellent paper that introduces many important concepts in a simple way.

Questions

1. The intrinsic reaction coordinate means:
   a) a trajectory of an atom when the reaction proceeds;
   b) the steepest descent path in the Cartesian space of the nuclear coordinates;
   c) the steepest descent path from a saddle point in the Cartesian space of the mass-weighted nuclear coordinates;
   d) a straight line in the Cartesian space of $3N - 6$ coordinates that connects the minima of the two basins.

2. In the vibrationally adiabatic approximation (reaction path Hamiltonian method) with all the normal modes in their ground states:
   a) the potential energy does not depend on the normal mode frequencies;
   b) the zero-vibrations depend on the reaction path coordinate s;
   c) the normal modes may exchange energy;
   d) the oscillators may exchange energy with the reaction path degree of freedom.

3. An endothermic reaction proceeds spontaneously ($T > 0$), because:
   a) the “drain-pipe” bottom potential energy plus the energies of the normal modes is lower in the entrance than in the exit channel;
   b) the oscillators are anharmonic;
   c) the “drain-pipe” bottom potential energy in the entrance channel is lower than that in the exit channel;
   d) the exit channel is wider than the entrance channel.

4. Donating mode:
   a) couples with the reaction path in the entrance channel;
   b) increases the reaction barrier;
   c) corresponds to high Coriolis couplings with other modes;
   d) corresponds to the lowest zero-vibration energy in the entrance channel.

5. In the acceptor–donor picture at the intermediate reaction stage (I) the following structures prevail:
   a) DA; b) D^+A^- and $D^{2+}A^{2-}$; c) D^+A^- and D^+A^{-*}; d) DA and D^+A^-.

6. In the acceptor–donor picture at the product reaction stage (P) the following structures prevail:
   a) DA; b) D^+A^-, $D^{2+}A^{2-}$ and D^+A^{-*}; c) D^+A^{-*}; d) DA and D^+A^-.

7. The ground-state adiabatic hypersurface in the neighbourhood of the conical intersection for three atoms:
   a) does not touch the excited-state adiabatic hypersurface;
   b) is a plane;
   c) consists of two diabatic parts of different electronic structures;
   d) does not touch a diabatic hypersurface.

8. In Marcus electron transfer theory:
   a) the reaction barrier is always equal to $\frac{1}{4}$ of the reorganization energy;

b) the larger the absolute value of the energy difference between products and reactants, the faster the reaction;
c) the activation energy is equal to the reorganization energy;
d) if the reactant energy is equal to the product energy, then the reaction barrier is equal to $\frac{1}{4}$ of the reorganization energy.

9. In Marcus theory of electron transfer:
a) we assume the same force constant for the reactants and products;
b) the reorganization energy in the reaction $Fe^{2+} + Fe^{3+} \rightarrow Fe^{3+} + Fe^{2+}$ in solution is equal to zero;
c) to have electron transfer we have to have the inverse Marcus region;
d) the solvent reorganization energy is equal to zero.

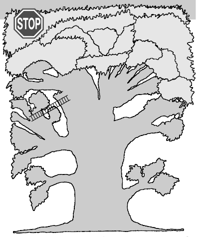
10. The reaction barrier:
a) has the same height from the reactant side and from the product side;
b) appears, because the hypersurface of an excited state that resembles the products intersects with the ground-state hypersurface for reactants;
c) means that the reactants have to have kinetic energy higher than its height;
d) results from the tunnelling effect.

Answers

1c, 2b, 3d, 4a, 5d, 6b, 7c, 8d, 9a, 10b

Chapter 15

INFORMATION PROCESSING – THE MISSION OF CHEMISTRY



WHERE ARE WE?

Where are we?

We have now explored almost the whole TREE.

An example

Chemistry has played, and continues to play, a prominent role in human civilization. If you doubt it, just touch *any* surface around you – most probably it represents a product of the chemical industry.[1] Pharmaceutical chemistry may be seen as a real benefactor, for it makes our lives longer and more comfortable. Is the mission of chemistry therefore to produce better dyes, polymers, semi-conductors, drugs? *No, its true mission is much, much more exciting.*

What is it all about

---

[1] Just a quick test around myself (random choice of surfaces): laptop (polymers), marble table (holes filled with a polymer), pencil (wood, but coated by a polymer), box of paper tissue (dyes and polymer coat), etc.

Why is this important?

In this book we have dealt with many problems in quantum chemistry. If this book were only about quantum chemistry, I would not write it. My goal was to focus on perspectives and images, rather than on pixel-like separate problems. Before we are quantum chemists we are scientists, happy eye-witnesses of miracles going on around us. We are also human beings, and have the right to ask ourselves, just what are we aiming for? *Why* is the Schrödinger equation to be solved? *Why* do we want to understand the chemical foundations of the world? Just for curiosity? Well, should curiosity legitimize *any* investigation?[2] What will the future role of chemistry be?

Chemistry is on the threshold of a big leap forward. Students of today will participate in this revolution. The limits will be set by our imagination, maybe by our responsibility as well. The direction we choose for the future progress in chemistry and biochemistry will determine the fate of human civilization. This *is* important…

What is needed?

- Elements of chemical kinetics.
- Elements of differential equations.
- Let us leave the traditional topics of chemistry, let us look around, let us look at how Nature operates.

Classical works

The classic papers pertain to three, at first sight unrelated, topics: molecular recognition, oscillatory solutions in mathematics and information flow. These topics evolved virtually separately within chemistry, mathematics and radio-communication, and only now[3] are beginning to converge. ★ Emil Hermann Fischer was the first to stress the importance of molecular recognition. In "*Einfluss der Konfiguration auf die Wirkung der Enzyme*" published in *Berichte*, 27 (1894) 2985 Fischer used the self-explanatory words "key-lock" for the perfect fit of an enzyme and its ligand. ★ In 1903 Jules Henri Poincaré published in *Journal de Mathematiques Pures et Appliques*, 7 (1881) 251 an article "*Mémoire sur les courbes définies par une équation différentielle*", where he showed that a wide class of two coupled non-linear differential equations leads to oscillating solutions that tend

Jules Henri Poincaré (1854–1912), French mathematician and physicist, professor at the Sorbonne, made important contributions to the theory of differential equations, topology, celestial mechanics, probability theory, and the theory of functions.

---

[2] Do not answer "yes" too easily, for it gives people the right to any experiments on you and me.

[3] The aim of the present chapter is to highlight these connections.

to a particular behaviour *independently of the initial conditions* (called the *limit cycle*). ★ It seems that the first experiment with an oscillatory chemical reaction was reported by Robert Boyle in the XVII century (oxidation of phosphorus). Then several new reports on chemical oscillations were published (including books). *All these results did not attract any significant interest in the scientific community, because they contradicted the widely known, all important, and successful equilibrium thermodynamics*. ★ The Soviet general Boris Belousov finally agreed to publish his only unclassified paper "*Periodichesky deystvouyoushchaya rieakcya i yeyo miekhanism*" in an obscure Soviet medical journal *Sbornik Riefieratow Radiacjonnoj Miediciny*, *Medgiz, Moskwa*, 1 (1959) 145 reporting spectacular colour oscillations in his test tube: yellow Ce^{4+} and then colourless Ce^{3+}, and again yellow, etc. (nowadays called the Belousov–Zhabotinsky reaction). ★ Independently, there was a continuing parallel progress in oscillatory solutions in mathematics. In 1910 Alfred J. Lotka in "*Contributions to the theory of chemical reactions*" published in the *Journal of Physical Chemistry*, 14 (1910) 271 proposed some differential equations that corresponded to the kinetics of an autocatalytic chemical reaction, and then with Vito Volterra gave a differential equation that describes a prey-predator *feedback* (oscillation) known as Lotka–Volterra model. ★ In February 1943, at the Dublin Institute for Advanced Studies,[4] Erwin Schrödinger gave several lectures trying to reconcile thermodynamics and biology. He stressed that biological systems are open: there is a flow of matter and energy. Independently of all these investigations there were attempts in radiocommunication to look quantitatively at information flow. ★ Ralph V.L. Hartley, published the first article on measuring information entitled "*Transmission of Information*" in *The Bell Systems Technical Journal*, 7 (1928) 535. ★ Twenty years later, the same topic was developed by Claude E. Shannon in "*A Math-*

Boris Pavlovich Belousov (1893–1970) looked for an inorganic analogue of the biochemical Krebs cycle. The investigations began in 1950 in a Soviet secret military institute. Belousov studied mixtures of potassium bromate with citric acid, and a small admixture of a catalyst: a salt of cerium ions. He expected a monotonic transformation of the yellow Ce^{4+} ions into the colourless Ce^{3+}. Instead, he found oscillations of the colour of the solvent (colourless-yellow-colourless-... etc., also called by Russians "vodka-cognac-vodka-...").

He wrote a paper and sent it to a Soviet journal, but the paper was rejected with a referee's remark that what the author had described was simply impossible. His involvement in classified research caused him to limit himself to bringing (by intermediacy of somebody) a piece of paper with reactants and his phone number written on it. He refused to meet anybody. Finally, Simon Schnoll persuaded him to publish his results. Neither Schnoll nor his PhD student Zhabotinsky ever met Belousov, though all they lived in Moscow.

Belousov's first chemistry experience was at the age of 12, while engaged in making bombs in the Marxist underground. Stalin thought of everything. When, formally underqualified, Belousov had problems as head of the lab, Stalin's handwriting in ordinary blue-pencil on a piece of paper: "*Has to be paid as a head of laboratory as long as he has this position*" worked miracles.

After S.E. Schnoll "*Geroi i zladiei rossiyskoi nauki*", Kron-Press, Moscow, 1997.

Ilya Prigogine (1917–2003) Belgian physicist, professor at the Université Libre de Bruxelles. In 1977 he received the Nobel prize "*for his contributions to non-equilibrium thermodynamics, particularly the theory of dissipative structures*".

---

[4] In that period of the war certainly looking like a tiny nucleus of civilization beyond the reach of barbarians. The lecture notes were published in 1944 by Cambridge University Press under the title "*What is Life?*"

ematical Theory of Communication" also published in *The Bell Systems Technical Journal*, 27 (1948) 379, 623, in which he related the notion of information and that of entropy. ★ The Belgian scientists Paul Glansdorff and Ilya Prigogine published a paper "*Sur les propriétés différentielles de la production d'entropie*" in *Physica*, 20 (1954) 773, that became the basis of irreversible thermodynamics. Ilya Prigogine and Gregoire Nicolis in an article "*On Symmetry-Breaking Instabilities in Dissipative Systems*", *Journal of Chemical Physics* 46 (1967) 3542 introduced the notion of dissipative structures. ★ Charles John Pedersen reopened (after the pioneering work of Emil Fischer) the field of supramolecular chemistry, publishing an article "*Cyclic Polyethers and their Complexes with Metal Salts*", which appeared in the *Journal of the American Chemical Society*, 89 (1967) 7017 and dealt with molecular recognition (cf. Chapter 13). ★ Manfred Eigen and Peter Schuster, in three articles "*The Hypercycle. A Principle of Natural Self-Organization*" in *Naturwissenschaften* 11 (1977), 1 (1978) and 7 (1978) introduced the idea of a hypercycle and of the natural selection of molecules to chemistry. ★ The mathematician Leonard Adleman published in *Science*, 266 (1994) 1021 "*Molecular Computation of Solutions to Combinatorial Problems*", in which he described *his own* chemical experiments that shed new light on the role molecules can play in processing information.

What are the most important problems in chemistry? Usually we have no time to compose such a list, not even to speak of presenting it to our students. The choice made reflects the author's personal point of view. The author tried to keep in mind that he is writing for mainly young (undergraduate and graduate) students, who are seeking not only for detailed research reports, but also for new *guidelines* in chemistry, for some general *trends* in it, and who want to establish strong and general *links* between mathematics, physics, chemistry and biology. An effort was made to expose the ideas, not only to students' minds but also to their hearts.

It is good to recall from time to time that all of us: physicists, chemists and biologists share the same electrons and nuclei as the objects of our investigation. It sounds trivial, but sometimes there is the impression that these disciplines investigate three different worlds. In the triad physics–chemistry–biology, chemistry plays a bridging role. By the middle of the twentieth century, chemistry had closed the

Kurt Gödel (1906–1978), German mathematician (then American, he was hardly persuaded in a taxi going to the ceremony of his naturalisation not to present inconsistencies in the US Constitution he had found). This mathematical genius proved a theorem now called Gödel's Undecidability Theorem that has shaken the foundations of mathematics (K. Gödel, *Monatshefte Math. Phys.*, 38 (1931) 173). Roughly speaking, the theorem says that any system of axioms leads to theorems neither true nor false. Gödel was probably inspired by old Greek paradoxes, like "*all Creteans lie – said a Cretean*".

Kurt Gödel was permanently afraid of being poisoned. After his wife's death, when nobody could persuade him that his food was safe, he died of hunger...

period of the exploration of its basic building blocks: elements, chemical bonds and their typical lengths, typical values of angles between chemical bonds, etc. Future discoveries in this field are not expected to change our ideas fundamentally. Now we are in a period of using this knowledge for the construction of what we only could dream of. In this Chapter I will refer now and then to mathematicians and mathematics, who deal with ideal worlds. For some strange reason at the foundation of (almost[5]) everything there is logic and mathematics. We have to notice, however, that after Kurt Gödel's proof of the incompleteness of *any* axiomatic system mathematics has become more like natural sciences. Physics, while describing the real rather than the ideal world, more than other natural sciences is symbiotic with mathematics.

Important cornerstones of this frontier region are given in brief below in three sections: Molecular Structures, Dynamics and Chemical Information Processing.

MOLECULAR STRUCTURES (STATICS)

15.1 COMPLEX SYSTEMS

Even a relatively simple system (e.g., an atom) often exhibits strange properties. Understanding simple objects seemed to represent a key for description of complex systems (e.g., molecules). Complexity *can* be explained using the first principles.[6] However, the complexity itself may add some important features. In a complex system some phenomena may occur, which would be extremely difficult to foresee from a knowledge of their component parts. Most importantly, sometimes the behaviour of a complex system is universal, i.e. independent of the properties of the parts of which it is composed (some of them will be mentioned in the present chapter) and related to the very fact that the system is composed of many small parts interacting in a simple way.

The behaviour of a large number of argon atoms represents a difficult task for theoretical description, but is still quite predictable. When the number of atoms increases, they pack together in compact clusters similar to those we would have with the densest packing of tennis balls (the maximum number of contacts). We may have to do here with complicated phenomena (similar to chemical reactions) and connected to the different stability of the clusters (e.g., "magic numbers" related to particularly robust closed shells[7]). Yet, the interaction of the argon atoms, however difficult for quantum mechanical description, comes from the quite primitive two-body, three-body etc. interactions (Chapter 13).

---

[5] Yes, almost: e.g., generosity is not included here.

[6] In the 20-ties of the twentieth century, after presenting his equation (see Chapter 3), Paul Dirac said that now chemistry is explained. Yet, from the equation to foreseeing the properties of complex organic molecules is a long, long way.

[7] Similar closed shells are observed in nuclear matter, where the "tennis balls" correspond to nucleons.

15.2 SELF-ORGANIZING COMPLEX SYSTEMS

Chemistry offers a plethora of intermolecular interactions.

Some intermolecular interactions are specific, i.e. a substrate A interacts with a particular molecule B_i from a set $B_1, B_2, \ldots, B_N$ (N is large) much more strongly than with others. The reasons for this are their shape, the electric field[8] compatibility, a favourable hydrophobic interaction etc. resulting either in the "key-lock" or "hand-glove" types of interaction, cf. Chapter 13. A molecule may provide a set of potential contacts localized in space (synthon, p. 744), which may fit to another synthon of another molecule. Two of nature's most important pairs of synthons are the hydrogen bond system of guanine and cytosine (GC) and of adenine and thymine (AT)[9] (see Fig. 13.17): in the case of *extended* synthons exhibiting an internal structure ("polysynthons" like, e.g., GAATC and CTTAG being sections of a DNA strand) finding in solution the first two matching synthons, e.g., in our case G and C, makes the next ones much easier, i.e. A and T etc., to fit, since they are already close in space and the entropy barrier is much easier to overcome.[10]

This idea is used in supramolecular chemistry. Suppose a particular reaction does not proceed with sufficient yield. Usually the reason is that, to run just this reaction the molecules have to find themselves in a very specific position in space (a huge entropy barrier to overcome), but before this happens they undergo some unwanted reactions. We may however "instruct" the reactants by substituting them with such synthons that the latter lock the reactants in the right position in space. The reaction we want to happen becomes inevitable. The driving force for all this is the particularly high interaction energy of the reactants. Very often however, the interaction energy has to be high, but not too high, in order to enable the reaction products to separate. This reversibility is one of the critically important features for "intelligent" molecules, which could adapt to external conditions in a flexible way. If a system with synthons is not flexible enough, we will still have to do with a relatively primitive structure.

If the system under consideration is relatively simple, even if the matching of corresponding synthons is completed, we would still have a relatively primitive spatial structure. However, we may imagine far more interesting situation, when:

- The molecules were *chosen* in such a way as to ensure that some intermolecular interaction is particularly attractive. A specific matching is known as molecular recognition.

molecular recognition

- The molecular complexes formed this way may recognize themselves again by using synthons previously existing or created *in situ*. In this way a multilevel structure can be formed, each level characterized by its own stability (cf. p. 744).

---

[8] Both molecules carry their charge distributions, their interaction at a certain geometry may considerably lower the Coulombic energy.

[9] G, C, A, T are four letters used by nature to compose the words, sentences, chapters, essays and poems of the Book of Life (the DNA code). The complementarity of the related synthons is of prime importance.

[10] The entropy barrier for A and B to make a complex AB is large when there are a lot of non-reactive A and B positions, and only a few that lead to formation of the complex.

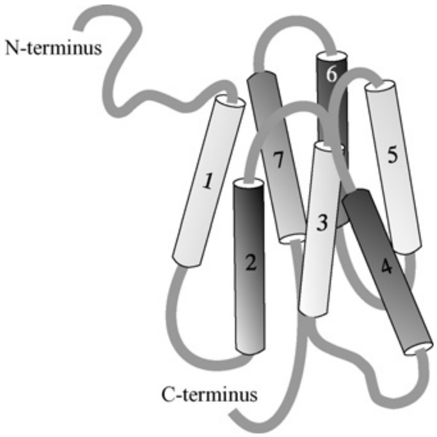


Fig. 15.1. A "universal" biological sensor based on rhodopsin. The sensor consists of seven α-helices connected by some oligopeptide links (a schematic view), the α-helices are shown as cylinders. The helices form a cavity, in which (in one of version of the sensor) there is a cis-retinal molecule (a chain of alternating single and double bonds), not shown in the figure, stretching between two helices. The cis-retinal is able to absorb a photon and change its conformation to trans. This triggers the cascade of processes responsible for our vision. The total system is hydrophobic outside, which makes it spontaneously anchor inside the cell walls composed of a lipid bilayer. The protruding protein loops exhibit specific interactions with some drugs. Such a system is at the basis of interaction with about 70% of drugs.

- The multilevel molecular structure may depend very strongly on its environment. When this changes, the structure may decompose, and eventually another structure may emerge.

> A *hierarchical* multilevel structure may be formed, where the levels exhibit different stability with regard to external perturbations. The stability differs due to the different binding energies of the synthons involved and/or on the steric constraints.

The coiled-coil structure of oligopeptides described on p. 748 may serve as an example of such a multilevel structure, or the spontaneous folding of enzymes to their native structure, e.g., rhodopsin is composed of seven α-helices linked by some oligopeptide links (Fig. 15.1).

There is nothing accidental in this system. The helices are composed of such amino acids, that ensure that the external surface of the helices is hydrophobic, and therefore enter the hydrophobic lipid bilayer of the cell walls. The peptide links serve to *recognize* and dock some particular signalling molecules. The 7-helix systems serve in biology as a universal sensor, with variations to make it specific for some particular molecular recognition and the processes that occur afterwards. After docking with a ligand or by undergoing photochemical isomerization of the retinal, some conformational changes take place, which after involving several intermediates, finally resulting in a signal arriving at a nerve cell. We see how wonderful things this sophisticated structure is able to do in a dynamic way.

15.3 COOPERATIVE INTERACTIONS

Some events may cooperate. Suppose we have an extended object, which may undergo a set of events: A, B, C, ..., each taking place separately and locally with a small probability. However, it may happen that for a less extended object the events cooperate, i.e. event A makes it easier for event B to occur, and when A and then B happens this makes it easier for event C to happen, etc.

Self-organization is possible without cooperativity, but cooperativity may greatly increase the effectiveness of self-organization. The hemoglobin molecule may serve as an example of cooperativity in intermolecular interactions, where its interaction with the first oxygen molecule makes its interaction with the second easier despite a considerable separation of the two binding events in space.

15.4 SENSITIVITY ANALYSIS

Sensitivity analysis represents a fast developing branch of applied mathematics. The essence of this approach is determining the response of a structure to a perturbation. The structure may represent a building or a molecule, and the perturbations may be of different kinds.[11] Experimental chemists very often introduce some substitutions, exchanging one functional group for another, and then observing the changes in the structure and properties of the system. Similarly, in biochemistry, both in experiment and theory (e.g., in molecular mechanics or dynamics), we make some artificial mutations. However, the current limitations of theory do not enable us to perform global molecular mechanics (cf. Chapter 7) and carry out sensitivity analysis when large responses of the system are admitted. It is very probable that this type of analysis will be of great importance in the future, because we will try to control the system globally, e.g., to foresee what will be the most stable structure after a perturbation is switched on.

15.5 COMBINATORIAL CHEMISTRY – MOLECULAR LIBRARIES

Chemistry is often regarded as dealing with pure substances,[12] which is obviously too demanding. This is difficult to achieve even for a pure compound, because of isomerization. In most cases we are interested in having a single isomer in the specimen. However, there are cases when the chemist is interested in a *mixture* of all possible isomers instead of a single isomer. Such a mixture is called a chemical library, and the chemistry that uses such libraries is called combinatorial chemistry. Thanks to the libraries we can search and find a given isomer. This is particularly spectacular in cases in which we have a labile equilibrium (i.e. easily shiftable) among the isomers.

A complex system may adjust itself to an external stimulus by changing its molecular structure. A good example is liquid water, which may be regarded as a "library" of different clusters, all of them being in an easy-to-shift equilibrium with others. This is why water is able to hydrate a nearly infinite variety of molecules, shifting the equilibrium towards the clusters that are needed to "wrap the solute by a water coat".

---

[11] Sensitivity analysis is *universal*. We apply it in everyday life (we see how our organism reacts to a perturbation by drug A, drug B, ...).

[12] This is stressed by the Dutch name for chemistry: "*scheikunde*" – i.e. the art of separation.

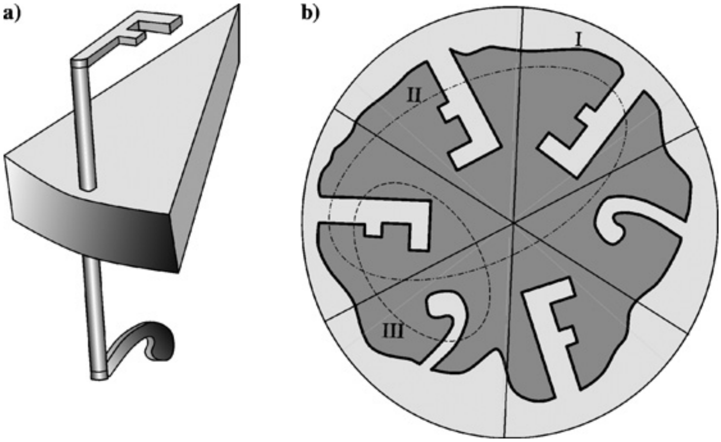


Fig. 15.2. A model of the immune system. (a) The figure shows schematically some monomers in a solvent. They have the shape of a slice of pie with two synthons: protruding up and protruding down, differing in shape. The monomers form some side-by-side aggregates containing from two to six monomers, each aggregate resulting in some pattern of synthons on one face and the complementary pattern on the other face. We have then a library of all possible associates in thermodynamical equilibrium. Say, there are plenty of monomers, a smaller number of dimers, even fewer trimers, etc. up to a tiny concentration of hexamers. (b) The attacking factor I (the irregular body shown) is best recognized and bound by one of the hexamers. If the concentration of I is sufficiently high, the *equilibrium among the aggregates shifts towards the hexamer mentioned above, which therefore binds all the molecules of I*, making them harmless. If the attacking factor was II and III, binding could be accomplished with some trimers or dimers (as well as some higher aggregates). The defence is highly specific and at the same time highly flexible (adjustable).

The immune system in our body is able to fight and win against practically any enemy, irrespective of its shape and molecular properties (charge distribution). How is it possible? Would the organism be prepared for everything? Well, yes and no.

Let us imagine a system of molecules (building blocks) having some synthons and able to create some van der Waals complexes, Fig. 15.2. Since the van der Waals forces are quite weak, the complexes are in dynamic equilibrium. All possible complexes are present in the solution, none of the complexes dominates.

Now, let us introduce some "enemy-molecules". The building blocks use part of their synthons for binding the enemies (that have complementary synthons), and at the same time bind among themselves in order to strengthen the interaction. Some of the complexes are especially effective in this binding. Now, the Le Chatelier rule comes into play and the equilibrium shifts to produce as many of the most effective binders as possible. On top of this, the most effective binder may undergo a chemical reaction that replaces the weak van der Waals forces by strong chemical forces (the reaction rate is enhanced by the supramolecular interaction). The enemy was tightly secured, the invasion is over.[13]

---

[13] A simple model of immunological defence, similar to that described above, was proposed by F. Cardullo, M. Crego Calama, B.H.M. Snelling-Ruël, J.-L. Weidmann, A. Bielejewska, R. Fokkens, N.M.M. Nibbering, P. Timmerman, D.N. Reinhoudt, *J. Chem. Soc. Chem. Commun*. 367 (2000).

DYNAMICS

15.6 NON-LINEARITY

Its origin is mathematical, where non-linearity is defined as opposed to linearity. Linearity, in the sense of the proportionality between a cause and an effect, is widely used in physics and technical sciences. There is common sense in this, since when a cause is small, the result is in most cases also small.[14] For instance, when a light object is hanging on a spring, the spring elongates in proportion to its weight (to high accuracy).[15] Similarly, when a homogeneous weak electric field is applied to the helium atom, its electrons will shift slightly towards the anode, while the nucleus will be displaced a little in the direction of the cathode, cf. Chapter 12. This results in an induced dipole moment, which to a high degree of accuracy is proportional to the electric field intensity, and the proportionality coefficient is the polarizability of the helium atom. Evidently, reversing the direction of the electric field would produce exactly the same magnitude of induced dipole moment, but its direction will be opposite. We can perform such an experiment with the HCl molecule (the molecule is fixed in space, the electric field directed along the H....Cl axis, from H to Cl).[16] When an electric field is applied, the dipole moment of the molecule will change slightly, and the change (an induced dipole moment) is to a good accuracy proportional to the field with the proportionality coefficient being the longitudinal polarizability of HCl. However, when the direction of the field is reversed, the absolute value of the induced dipole moment will be the same as before. Wait a minute! This is pure nonsense. The electrons move with the same facility towards the electron acceptor (chlorine) as to the electron donor (hydrogen)? Yes, as far as the polarizability (i.e. linearity) decides, this is true indeed. Only, when going beyond the linearity, i.e. when the induced dipole moment depends on higher powers of the electric field intensity, we recover common sense: electrons move more easily towards an electron acceptor than towards an electron donor... Thus, the non-linearity is there and is important.

Non-linearity was an unwanted child of physics. It sharply interfered with making equations easy to solve. Without it, the solutions often represent beautiful, concise expressions, with great interpretative value, whereas with it everything gets difficult, clumsy and most often impossible to treat. We are eventually left with numerical solutions, which have to be treated case by case with no hope of a nice generalization. Sometimes the non-linearity could be treated by perturbation theories,

---

[14] "Most" is a dangerous word. What about such things dice, roulette, etc.? There is a kind of "histerical" dependence of the result from the initial conditions. The same is true for the solution of the equation... $x^3 = -1$. Until the nineteen-eighties mathematicians thought that nothing new would be added to this solution. However, when they applied Newton's method to solve it numerically, a fractal dependence on the initial conditions appeared.

[15] Non-linearity is, however, entering into play if the object is heavy and/or if the spring is compressed with the same force instead of elongated.

[16] In this molecule, without any external electric field applied, the electrons are slightly shifted from the hydrogen (electron donor) to the chlorine atom (electron acceptor), which results in a permanent dipole moment.

where the linear case is considered as a reference and the non-linear corrections are taken into account and calculated. Nothing particularly important emerged from this. Now we know why. Perturbation theory requires a small perturbation (a weak non-linearity), while the most interesting phenomena discovered in the 1970-ties by Prigogine, emerged when non-linearity is large (large fluctuations exploring new possibilities of the system).

With the advent of computers that which was difficult to solve (numerically) before, often became easy. Without computers, we would understand much less about dissipative structures, chaos theory, attractors, etc. These subjects are of a mathematical nature, but have a direct relation to physics and chemistry, and most of all to biology. The relation happens on remarkably different scales and in remarkably different circumstances:[17] from chemical waves *in space* rationalizing the extraordinary pattern of the zebra skin to population waves of lynxes and rabbits as functions of time. In all these phenomena non-linearity plays a prominent role.

Quite surprisingly, it turns out that a few non-linear equations have analytical and simple solutions. One of such cases is a soliton, i.e. a solitary wave (a kind of hump). Today solitons already serve to process information, thanks to the *non-linear* change of the refractive index in a strong laser electric field. Conducting polymers turn out to be channels for another kind of solitons[18] (cf. Chapter 9).

15.7 ATTRACTORS

Mitchell Feigenbaum (b. 1944), American physicist, employee of the Los Alamos National Laboratory, then professor at the Cornell University and at the Rockefeller University. Feigenbaum discovered attractors after making some observations just playing with a pocket calculator.

Non-linearity in mathematics is connected to the notion of attractors.

The theory of attractors was created by Mitchell Feigenbaum. When applying an iterative method of finding a solution,[19] we first decide which operation is supposed to bring us closer to the solution as well as what represents a reasonable zero-order guess (starting point: a number, a function, a sequence of functions). Then we force an evolution (“dynamics”) of the approximate solutions by applying the operation first to the starting point, then to the result obtained by the operation on the starting point, and then again and again until convergence is achieved.

Let us take an example and choose as the operation on a number x the following $x_{n+1} = \sin(x_n^2 + 1)$, where n stands for the iteration number. The iterative scheme therefore means choosing any x_0, and then applying many times a sequence of

---

[17] This witnesses the universality of Nature’s strategy.

[18] The word “channel” has been used on purpose to allude to the first soliton wave observed in an irrigation channel.

[19] Cf. the SCF LCAO MO method, p. 364, or the iterative version of perturbational theory, p. 717.

four keys on the calculator keyboard. Here are the results of two different starting points: $x_0 = 1410$ and -2000.

1410	-2000
-0.0174524	0.656059
0.0174577	0.0249628
0.0174577	0.0174633
0.0174577	0.0174577

The *result is independent of the starting point chosen*. The number 0.0174577 represents an attractor or a *fixed point* for the operation. In the SCF method the fixed point is identical with the single Slater-determinant function (a point in the Hilbert space, cf. Appendix B) – a result of the SCF iterative procedure.

fixed point

Let us consider some other attractors. If we take the clamped-nuclei electronic energy $V(\mathbf{R})$ as a function of the nuclear configuration $\mathbf{R}$ ($V(\mathbf{R})$ represents a generalization of $E_0^0(R)$ from eq. (6.18), p. 227, that pertains to a diatomic molecule). The forces acting on atoms can be computed as the components of the vector $\mathbf{F} = -\nabla V$. Imagine we are looking for the most stable configurations of the nuclei, i.e. for the minima of $V(\mathbf{R})$. We know that when such a configuration is achieved, the forces acting on all the atoms are zero. When we start from an initial guess $\mathbf{R}_0$ and follow the computed force $\mathbf{F} = -\nabla V$ (this defines the operation in question), then it is hoped that we end up at a local minimum of V independent of the starting point, provided the point belongs to the basin corresponding to the minimum (cf. p. 769). If, however, the starting point were outside the basin, we would find another minimum (having its own basin, where the starts would all lead to the same result). Thus, we may have to do with many attractors at the same time. The positions of the *maxima* of V may be called *repellers* to stress their action opposite to the attractors. For a repeller the procedure of following the direction of $-\nabla V$ gets us further and further away from the repeller.

repellers

In thermodynamics, the equilibrium state of an isolated system (at some fixed external parameters) may be regarded as an attractor, that *any* non-equilibrium state attains after a sufficiently long time.

15.8 LIMIT CYCLES

Sometimes an attractor represents something other than just a point at which the evolution of the system definitely ends up.

Consider a set of two differential equations with time t as variable. Usually their solution [$x(t)$ and $y(t)$] depends on the initial conditions assumed, Fig. 15.3.a.

Now let us take a particular set of two *non-linear* differential equations. As seen from Fig. 15.3.b, this time the behaviour of the solution as a function of time is completely different: for high values of t the solution does not depend on the ini-

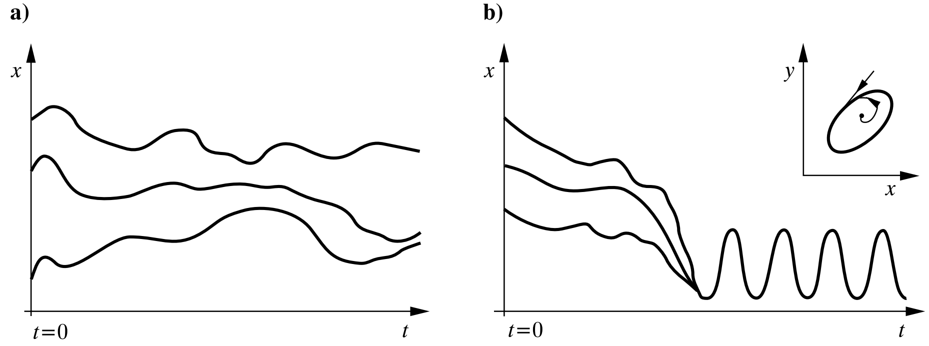


Fig. 15.3. Two different behaviours of solutions of differential equations, depending on initial conditions. (a) The plots represent $x(t)$ for three sets of initial conditions. As seen, the trajectories differ widely, i.e. the fate of the system depends very much on the initial conditions. Fig. (b) shows the idea of the limit cycle for a set of hypothetical non-linear differential equations. For large values of t, the three sets of initial conditions lead to *the same* trajectory.

tial conditions chosen. We obtain the $y(x)$ dependence in a form called the *limit cycle*, and the functions $x(t)$ and $y(t)$ exhibit periodic oscillations. The system is condemned to repeat forever the same sequence of positions – the limit cycle.

In chemistry x and y may correspond to the concentrations of two substances. The limit cycles play a prominent role in new chemistry, since they ensure that the system evolves to the same periodic oscillations independent of the initial conditions of some chemical reactions (with the non-linear dependence of their velocity on concentrations, cf. p. 872). Such reactions could, therefore,

chemical clock

- provide a stimulus for the periodic triggering of some chemical processes (*chemical clock*),
- provide chemical counting, which (similar to today's computers) could be related to chemical programming in the future.

15.9 BIFURCATIONS[20] AND CHAOS

Non-linear dynamics turned out to be extremely sensitive to coupling with some external parameters (representing the "neighbourhood").

logistic equation

Let us take what is called the *logistic equation*

$$x = Kx(1 - x),$$

where $K > 0$ is a constant. The Oxford biologist, Sir Robert May, gave a numerical exercise to his Australian graduate students. They had to calculate how a rabbit

---

[20] A bifurcation (corresponding to a parameter p) denotes in mathematics a doubling of an object when the parameter exceeds a value p_0. For example, when the object corresponds to the number of solutions of equation $x^2 + px + 1 = 0$, then the bifurcation point $p_0 = 2$. Another example of bifurcation is branching of roads, valleys, etc.

population evolves when we let it grow according to the rule

$$x_{n+1} = Kx_n(1 - x_n),$$

which is obviously related to the logistic equation. The natural number n denotes the current year, while x_n stands for the (relative) population of, say, rabbits in a field, $0 \leqslant x_n \leqslant 1$. The number of the rabbits in year $(n+1)$ is proportional to their population in the preceding year (x_n), because they reproduce very fast, but the rabbits eat grass and the field has a finite size. The larger x_n the less the amount of grass to eat, which makes the rabbits a bit weaker and less able to reproduce (this effect corresponds to $1 - x_n$).

> The logistic equation contains a feed back mechanism.

The constant K measures the population–grass coupling strength (low-quality grass means a small K). What interests us is the fixed point of this operation, i.e. the final population the rabbits develop after many years at a given coupling constant K. For example, for $K = 1$ the evolution leads to a steady self-reproducing population x_0, and x_0 depends on K (the larger K the larger x_0). The graduate students took various values of K. *Nobody* imagined this quadratic equation could hide a mystery.

If K were small ($0 \leqslant K < 1$, extremely poor grass), the rabbit population would simply vanish (the first part of Fig. 15.4). If K increased (the second part of the plot, $1 \leqslant K < 3$), the population would flourish. When K exceeded 3 this flourishing would give, however, a unexpected twist: instead of reaching a fixed point, the system would oscillate between two sizes of the population (every second year the population was the same, but two consecutive years have different populations). This resembles the limit cycle described above – the system just repeats the same cycle all the time.

This mathematical phenomenon was carefully investigated and the results were really amazing. Further increase in K introduces further qualitative changes. First, for $3 \leqslant K < 3.44948$ the oscillations have period two (*bifurcation*), then at $3.44948 \leqslant K < 3.5441$ the oscillations have period four (next bifurcation, the four-member limit cycle), then for $3.5441 \leqslant K < 3.5644$ the period is eight (next bifurcation).[21]

bifurcation

Then, the next surprise: exceeding $K = 3.56994$ we obtain populations that do not exhibit any regularity (no limit cycle, just *chaos*). A further surprise is that this is not the end of the surprises. Some sections of K began to exhibit *odd*-period behaviour, separated by some sections of chaotic behaviour.

chaos

---

[21] Mitchell Feigenbaum was interested to see at which value $K(n)$ the next bifurcation into 2^n branches occurs. It turned out that there is a certain regularity, namely, $\lim_{n\to\infty} \frac{K_{n+1}-K_n}{K_{n+2}-K_{n+1}} = 4.669201609\ldots \equiv \delta$. To the astonishment of scientists, the value of δ turned out to be "universal", i.e. characteristic for many *very different* mathematical problems and, therefore, reached a status similar to that of the numbers π and e. The numbers π and e satisfy the exact relation $e^{i\pi} = -1$, but so far no similar relation was found for the Feigenbaum constant. There is an *approximate* relation (used by physicists in phase transition theory) which is satisfied: $\pi + \tan^{-1} e^{\pi} = 4.669201932 \approx \delta$.

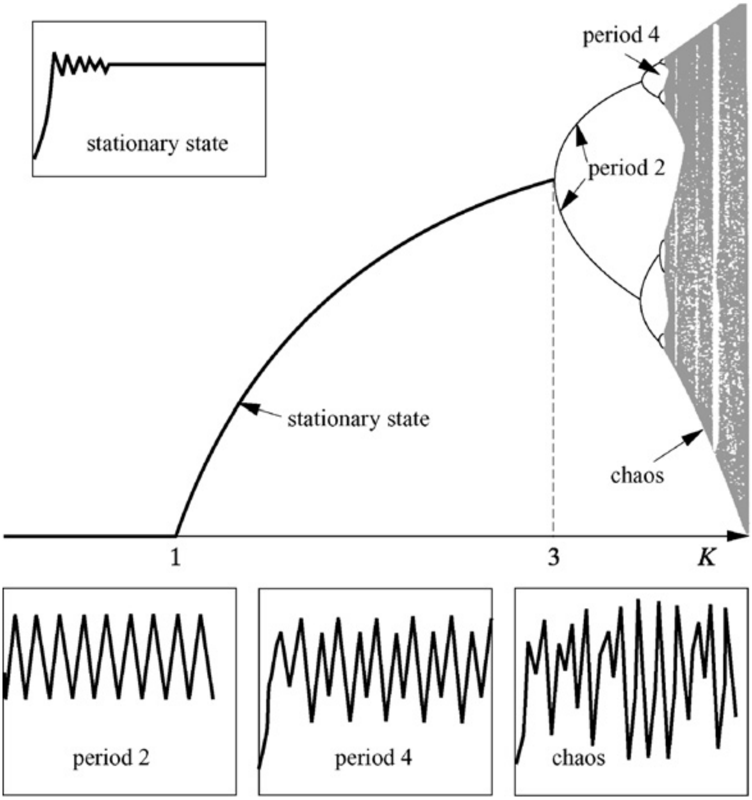


Fig. 15.4. The diagram of the fixed points and the limit cycles for the logistic equation as a function of the coupling constant K. From J. Gleick, "*Chaos*", Viking, New York, 1988, reproduced with permission of the author.

15.10 CATASTROPHES

The problems described above have to do with another important mathematical theory.

As has been shown for electronic energy $V(\mathbf{R})$, we may have several minima. Having a deterministic procedure that leads from a given point to a minimum means creating the dynamics of the system (along a trajectory), in which any minimum may be treated as an attractor (Chapter 6), with its basin meaning those points that, following the dynamics, produce trajectories that end up at the minimum. We can also imagine trajectories that do not end up at a point, but in a closed loop (limit cycle).

Imagine $V(\mathbf{R})$ depends on a parameter t. What would happen to the attractors and limit cycles if we changed the value of the parameter? When a change has a qualitative character (e.g., the number of basins changes), the founder of the theory, René Thom, called it a catastrophe.

15.11 COLLECTIVE PHENOMENA

Imagine some subunits cooperate so strongly that many events require less energy than a single one or a few. In such a case, a few events may trigger an avalanche of other events (*domino effect*). Numerous examples of this are phase transitions, where a change of the position, orientation or conformation of a few molecules requires energy, whereas when a barrier is overcome the changes occur spontaneously for all the molecules. Imagine a photoisomerization (such as that of azabenzene) in the solid state. If a *single* molecule in a crystal were to undergo the change, such an excitation might cost a lot of energy, because there might not be enough space to perform the trans to cis transition.[22] When, however, a lot of molecules undergo such a change in a concerted motion, the atomic collision would not necessarily take place and the cost in energy would be much smaller than the sum of all the single excitations.

domino effect

An example of electronic collectivity may also be the electronic bistability effect expected to occur in a rigid donor–acceptor oligomer; $(DA)_N$, composed of suitable electron donors (D) and acceptors (A) at a proper DA distance and orientation, Fig. 15.5.

15.11.1 SCALE SYMMETRY (RENORMALIZATION)

It turns out that different substances, when subject to phase transition, behave in exactly the same way exhibiting therefore a universal behaviour.

Imagine a system of N identical equidistant spin magnetic moments located on the z axis, each spin parallel or antiparallel to the axis.[23] The j-th spin has two components (cf. p. 28) $\sigma_j = 1, -1$. Often the Hamiltonian H of a system is approximated by taking into account nearest-neighbour interactions only (*Ising model*) in the following way (the constants K, h, C fully determine the Hamiltonian)

Ernst Ising (1900–1998), German mathematician and physicist. In 1939, after interrogation by the gestapo in Berlin, Ising emigrated to Luxemburg, and there in a German labour camp he held out until liberation by the Allies. From 1948 he became a professor at Bradley University (USA). His two-state chain model is very often used in mathematical physics.

$$H=K\sum_j \sigma_j\sigma_{j+1} + h\sum_j \sigma_j + C, \tag{15.1}$$

where the first term corresponds to dipole-dipole magnetic interactions like those described on p. 655, the second term takes care of the interactions with an external magnetic field (Zeeman effect, p. 659), and C is a constant.

---

[22] Some atoms would simply hit others, causing an enormous increase in energy resulting in an energy barrier.

[23] The objects need not be spins, they may represent two possible orientations of the molecules, etc.

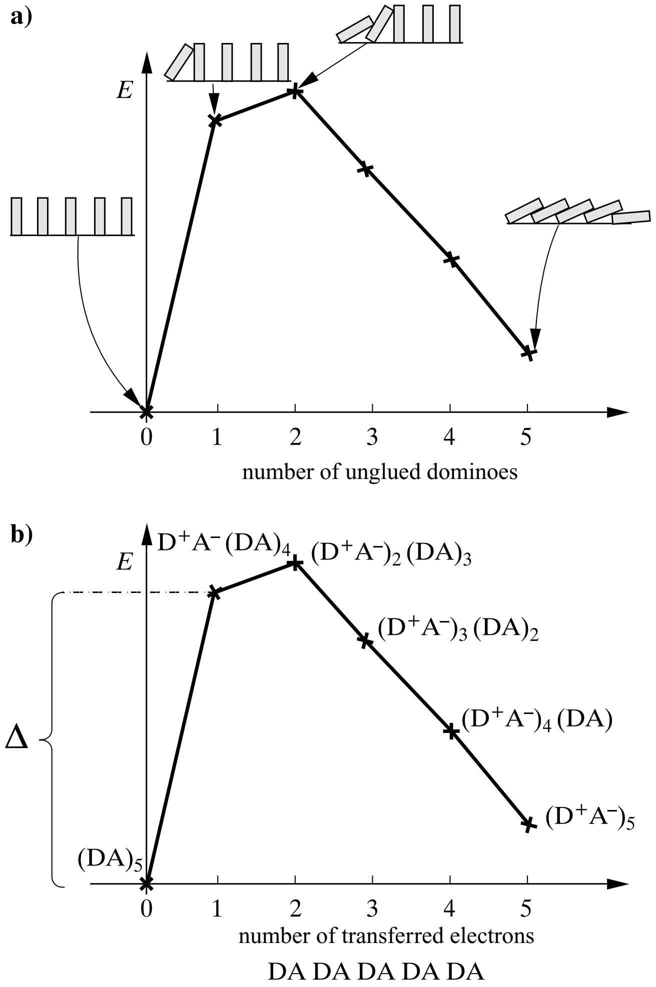


Fig. 15.5. Collective phenomena. (a) The domino principle. An energy cost corresponding to unglueing and knocking down the dominoes. (b) Hypothetical electronic domino (or "mnemon" – an element of molecular memory) composed of electron donors (D) and electron acceptors (A). In order to transfer the first electron we have to pay energy Δ. The second electron transfer (when the first is already transferred) needs less energy, because it is facilitated by the dipole created. The transfer of the third and further electrons does not need any energy at all (the energy actually decreases). The hypothetical electronic domino starts running (L.Z. Stolarczyk, L. Piela, *Chem. Phys.* 85 (1984) 451).

The partition function (which all the thermodynamic properties can be computed from) is defined as:

$$Z(T) = \frac{1}{2^N} \sum_{\sigma_1} \sum_{\sigma_2} \cdots \sum_{\sigma_N} \exp\left(-\frac{H(K,h,C)}{k_B T}\right). \tag{15.2}$$

Each of the N sums in eq. (15.2) pertains to a single spin. A trivial observation that the summation in eq. (15.2) *can* be carried out in (two) steps, leads to something extraordinary. We may first *sum over every other object*.[24] Then, the spins of the objects we have summed formally disappear from the formula, we have the summation over spins of the *remaining* objects only. Such a procedure is called *decimation*[25] from a form of collective capital punishment in the regulations of the Roman legions (very unpleasant for every tenth legionary). As a result of the procedure, the Hamiltonian H is changed and now corresponds to the interaction of the spins of the remaining objects. These spins, however, are "dressed" in the interaction with the other spins, which have been killed in the decimation procedure. What purpose may such a decimation serve? Well,

decimation

after this is done, the expression $Z(T)$ from formula (15.2) will look *similar* to that before the transformation (*self-similarity.*). Only the constants $K \to K'$, $h \to h'$, $C \to C'$ change.[26]

The two Hamiltonians are related by a self-similarity. The decimation may then be repeated again and again, leading to a trajectory in the space of the parameters K, h, C. It turns out that a system undergoing a phase transition is located on such a trajectory. By repeating the decimation, we may reach a fixed point (cf. p. 858), i.e. further decimations do not change the parameters, the system attains *self-similarity on all scales*. The *fixed point may be common for a series of substances*, because the trajectories (each for a given substance) may converge to a common fixed point. The substances may be different, may interact differently, may undergo different phase transitions, but since they share the fixed point, some features of their phase transitions are nevertheless identical.

self-similarity

This section links together several topics: attractors, self-similarity (renormalization group theory), catastrophe theory.

15.11.2 FRACTALS

Self-similarity, highlighted by renormalization, represents the essence of fractals. Let us consider what is called the *Sierpiński carpet* (Fig. 15.6.a).

Sierpiński carpet

---

[24] Here we follow D.R. Nelson and M.E. Fisher, *Ann. Phys. (N.Y.)* 91 (1975) 226.

[25] Although in this situation the name does not fit quite so well.

[26] It is a matter of fifteen minutes to show (e.g., M. Fisher, *Lecture Notes in Physics* 186 (1983)), that the new constants are expressed by the old ones as follows:

$$\exp\left(4K'\right) = \frac{\cosh(2K+h)\cosh(2K-h)}{\cosh^2 h},$$

$$\exp\left(2h'\right) = \exp(2h)\frac{\cosh(2K+h)}{\cosh(2K-h)},$$

$$\exp(4C') = \exp(8C)\cosh(2K+h)\cosh(2K-h)\cosh^2 h.$$

The self-similarity of this mathematical object (when we decide to use more and more magnifying glasses) is evident.

Wacław Sierpiński (1882–1969), Polish mathematician, from 1910 professor at the Jan Casimir University in Lwów, and from 1918 at the University of Warsaw. One of the founders of the famous Polish school of mathematics. His most important achievements are related to set theory, number theory, theory of real functions and topology (there is the carpet in question).

Benoit Mandelbrot, French mathematician, born in 1924 in Warsaw, first worked at the Centre National de la Recherche Scientifique in Paris, then at the Université de Lille, from 1974 an employee of the IBM Research Center in New York. When playing with a computer, Mandelbrot discovered the world of fractals.

On the other hand, it is striking that fractals of fantastic complexity and shape may be constructed in an amazingly simple way by using the dynamics of the iteration processes described on p. 858. Let us take, for example, the following operation defined on the complex plane: let us choose a complex number C, and then let us carry out the iterations

$$z_{n+1} = z_n^2 + C$$

for $n = 0, 1, 2, 3, \ldots$ starting from $z_0 = 0$. The point C will be counted as belonging to what is called the Mandelbrot set, if the points z_n do not exceed a circle of radius 1. The points of the Mandelbrot set will be denoted by black, the other points will be coloured depending on the velocity at which they flee the circle. Could anybody ever think that we would get the incredibly rich pattern shown in Fig. 15.6.b?

15.12 CHEMICAL FEEDBACK – NON-LINEAR CHEMICAL DYNAMICS

Could we construct chemical feedback? What for? Those who have ever seen feedback working know the answer[27] – this is the very basis of control. Such control of chemical concentrations is at the heart of how biological systems operate.

The first idea is to prepare such a system in which an increase in the concentration of species X triggers the process of its decreasing. The decreasing occurs by replacing X by a very special substance Y, each molecule of which, when disintegrating, produces several X molecules. Thus we would have a scheme (X denotes a large concentration of X, x denotes a small concentration of X; similarly for the species Y): (X, y) → (x, Y) → (X, y) or *oscillations of the concentration of X and Y in time*.[28]

---

[27] For example, an oven heats until the temperature exceeds an upper bound, then it switches off. When the temperature reaches a lower bound, the oven switches *itself* on (therefore, we have temperature *oscillations*).

[28] Similar to the temperature oscillations in the feedback of the oven.

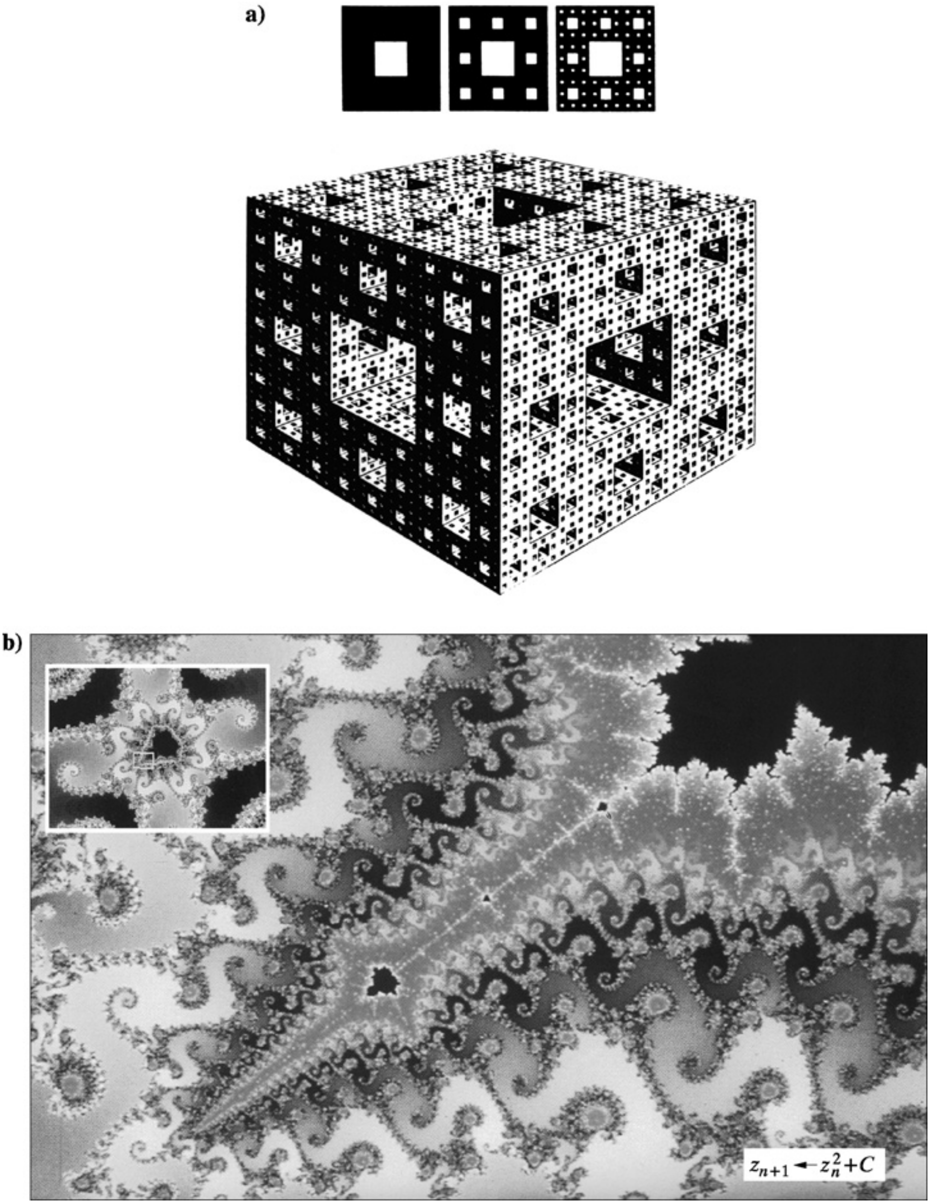


Fig. 15.6. Fractals. (a) Sierpiński carpet. (b) Mandelbrot set. Note that the incredibly complex (and beautiful) set exhibits some features of self-similarity, e.g., the central "turtle" is repeated many times in different scales and variations, as does the fantasy creature in the form of an S. On top of this, the system resembles the complexity of the Universe: using more and more powerful magnifying glasses, we encounter ever new elements that resemble (but not just copy) those we have already seen. From J. Gleick, "*Chaos*", Viking, New York, 1988, reproduced by permission of the author.

15.12.1 BRUSSELATOR – DISSIPATIVE STRUCTURES

Brusselator without diffusion

Imagine we carry out a complex chemical reaction in flow conditions,[29] i.e. the reactants A and B are pumped with a constant speed into a long narrow tube reactor, there is intensive stirring in the reactor, then the products flow out to the sink (Fig. 15.7). After a while a steady state is established.[30]

After the A and B are supplied, the substances[31] X and Y appear, which play the role of catalysts, i.e. they participate in the reaction, but in total their amounts do not change. To model such a situation let us assume the following chain of chemical reactions:

$$\begin{aligned} \mathrm{A} &\rightarrow \mathrm{X} \\ \mathrm{B}+\mathrm{X} &\rightarrow \mathrm{Y}+\mathrm{D} \\ 2\mathrm{X}+\mathrm{Y} &\rightarrow 3\mathrm{X} \\ \mathrm{X} &\rightarrow \mathrm{E} \\ \text{in total} &: \\ \mathrm{A}+\mathrm{B}+4\mathrm{X}+\mathrm{Y} &\rightarrow \mathrm{D}+\mathrm{E}+4\mathrm{X}+\mathrm{Y} \end{aligned}$$

This chain of reactions satisfies our feedback postulates. In step 1 the concentration of X increases, in step 2 Y is produced at the expense of X, in step 3 substance Y enhances the production of X (at the expense of itself, this is an *autocatalytic step*), then again X transforms to Y (step 2), etc.

autocatalysis

If we shut down the fluxes in and out, after a while a thermodynamic equilibrium is attained with all the concentrations of the six substances (A, B, D, E, X, Y; their concentrations will be denoted as A, B, D, E, X, Y, respectively) being constant

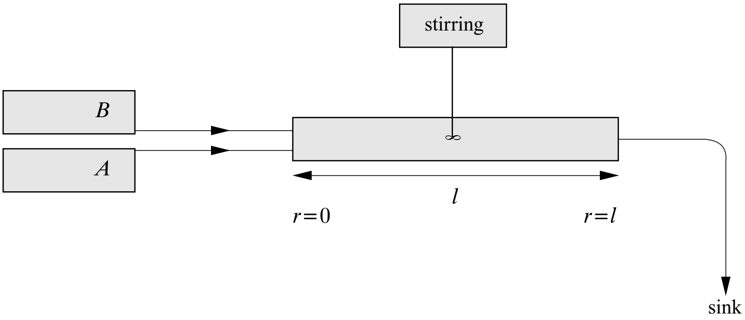


Fig. 15.7. A flow reactor (a narrow tube – in order to make a 1D description possible) with stirring (no space oscillations in the concentrations). The concentrations of A and B are kept constant at all times (the corresponding fluxes are constant).

---

[29] Such reaction conditions are typical for industry.

[30] To be distinguished from the thermodynamic equilibrium state, where the system is isolated (no energy or matter flows).

[31] Due to the chemical reactions running.

in space (along the reactor) and time. On the other hand, when we fix the in and out fluxes to be constant (but non-zero) for a long time, we force the system to be in a steady state and as far from thermodynamic equilibrium as we wish. In order to simplify the kinetic equations, let us assume the irreversibility of all the reactions considered (as shown in the reaction equations above) and put all the velocity constants equal to 1. This gives the kinetic equations for what is called the *Brusselator model* (of the reactor)

brusselator

$$\begin{aligned} \frac{\mathrm{d}X}{\mathrm{d}t} &= A - (B+1)X + X^2Y, \\ \frac{\mathrm{d}Y}{\mathrm{d}t} &= BX - X^2Y. \end{aligned} \tag{15.3}$$

These two equations, plus the initial concentrations of X and Y, totally determine the concentrations of all the species as functions of time (due to the stirring there will be no dependence on position in the reaction tube).

Steady state

A steady state (at constant fluxes of A and B) means $\frac{dX}{dt} = \frac{\mathrm{d}Y}{\mathrm{d}t} = 0$ and therefore we easily obtain the corresponding steady-state concentrations X_s, Y_s by solving eq. (15.3)

$$\begin{aligned} 0 &= A - (B+1)X_s + X_s^2 Y_s, \\ 0 &= BX_s - X_s^2 Y_s. \end{aligned}$$

Please check that these equations are satisfied by

$$\begin{aligned} X_s &= A, \\ Y_s &= \frac{B}{A}. \end{aligned}$$

Evolution of fluctuations from the steady state

> Any system undergoes some spontaneous concentration fluctuations, or we may perturb the system by injecting a small amount of X and/or Y. *What will happen to the stationary state found a while before, if such a fluctuation happens?*

Let us see. We have fluctuations x and y from the steady state

$$\begin{aligned} X(t) &= X_s + x(t), \\ Y(t) &= Y_s + y(t). \end{aligned} \tag{15.4}$$

What will happen next?

After inserting (15.4) in eqs. (15.3) we obtain the equations describing how the fluctuations evolve in time

$$\begin{aligned}\frac{\mathrm{d}x}{\mathrm{d}t} &= -(B+1)x + Y_s\big(2X_s x + x^2\big) + y\big(X_s^2 + 2xX_s + x^2\big),\\ \frac{\mathrm{d}y}{\mathrm{d}t} &= Bx - Y_s\big(2X_s x + x^2\big) - y\big(X_s^2 + 2xX_s + x^2\big).\end{aligned} \tag{15.5}$$

Since a mathematical theory for arbitrarily large fluctuations does not exist, we will limit ourselves to small x and y. Then, all the quadratic terms of these fluctuations can be neglected (*linearization* of (15.5)). We obtain

linearization

$$\begin{aligned}\frac{\mathrm{d}x}{\mathrm{d}t} &= -(B+1)x + Y_s(2X_s x) + yX_s^2,\\ \frac{\mathrm{d}y}{\mathrm{d}t} &= Bx - Y_s(2X_s x) - yX_s^2.\end{aligned} \tag{15.6}$$

Let us assume fluctuations of the form[32]

$$\begin{aligned}x &= x_0 \exp(\omega t),\\ y &= y_0 \exp(\omega t)\end{aligned} \tag{15.7}$$

and represent particular solutions to eqs. (15.6) provided the proper values of ω, x_0 and y_0 are chosen. After inserting (15.7) in eqs. (15.6) we obtain the following set of equations for the unknowns ω, x_0 and y_0

$$\begin{aligned}\omega x_0 &= (B-1)x_0 + A^2 y_0,\\ \omega y_0 &= -Bx_0 - A^2 y_0.\end{aligned} \tag{15.8}$$

This represents a set of homogeneous linear equations with respect to x_0 and y_0, and this means we have to ensure that the determinant, composed of the coefficients multiplying the unknowns x_0 and y_0, vanishes (*characteristic equation*, cf. secular equation, p. 202)

$$\begin{vmatrix}\omega - B + 1 & -A^2\\ B & \omega + A^2\end{vmatrix} = 0.$$

This equation is satisfied by some special values of[33] ω:

$$\omega_{1,2} = \frac{T \pm \sqrt{T^2 - 4\Delta}}{2}, \tag{15.9}$$

where

---

[32] Such a form allows for exponential growth ($\omega > 0$), decaying ($\omega < 0$) or staying constant ($\omega = 0$), as well as for periodic behaviour ($\mathrm{Re}\,\omega = 0, \mathrm{Im}\,\omega \neq 0$), quasiperiodic growth ($\mathrm{Re}\,\omega > 0, \mathrm{Im}\,\omega \neq 0$) or decay ($\mathrm{Re}\,\omega < 0, \mathrm{Im}\,\omega \neq 0$).

[33] They represent an analogue of the normal mode frequencies from Chapter 7.

$$T = -\left(A^2 - B + 1\right), \tag{15.10}$$

$$\Delta = A^2. \tag{15.11}$$

Fluctuation stability analysis

Now it is time to pick the fruits of our hard work.

How the fluctuations depend on time is characterized by the roots $\omega_1(t)$ and $\omega_2(t)$ of eq. (15.9), because x_0 and y_0 are nothing but some constant amplitudes of the changes. We have the following possibilities (Fig. 15.8, Table 15.1.):

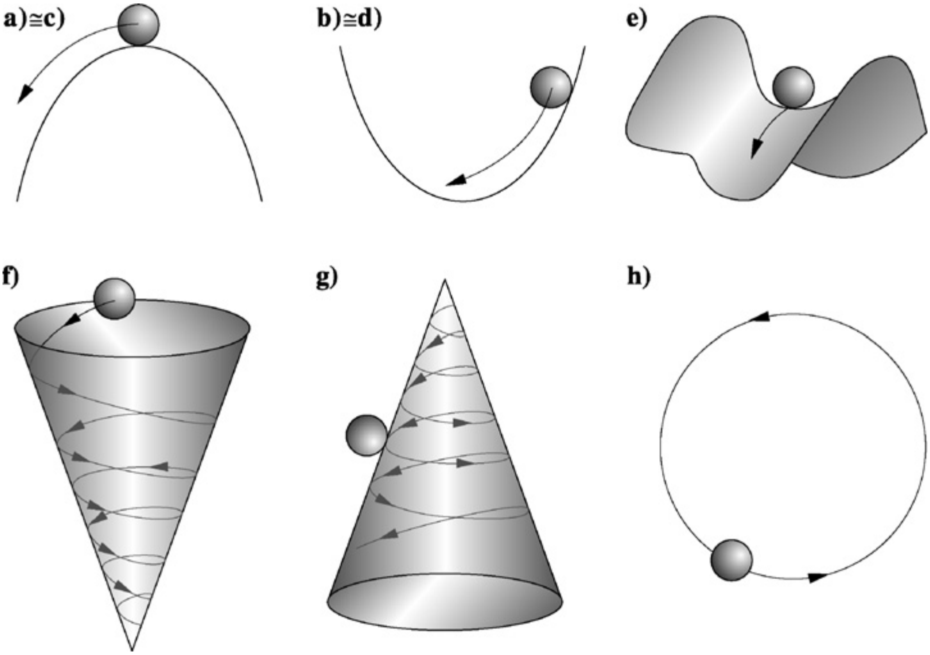


Fig. 15.8. Evolution types of fluctuations from the reaction steady state. The classification is based on the numbers ω_1 and ω_2 of eq. (15.9). The individual figures correspond to the rows of Table 15.1. The behaviour of the system (in the space of chemical concentrations) resembles sliding of a point or rolling a ball over certain surfaces in a gravitational field directed downward:

(a) unstable node resembles sliding from the top of a mountain;
(b) stable node resembles moving inside a bowl-like shape;
(c) the unstable stellar node is similar to case (a), with a slightly different mathematical reason behind it;
(d) similarly for the stable stellar node [resembles case (b)];
(e) saddle – the corresponding motion is similar to a ball rolling over a cavalry saddle (applicable for a more general model than the one considered so far);
(f) stable focus – the motion resembles rolling a ball over the interior surface of a cone pointing downward;
(g) unstable focus – a similar rolling but on the external surface of a cone that points up;
(h) centre marginal stability corresponds to a circular motion.

Table 15.1. Fluctuation stability analysis, i.e. what happens if the concentrations undergo a fluctuation from the steady state values. The analysis is based on the values of ω_1 and ω_2 from eq. (15.9); they may have real (subscript r) as well as imaginary (subscript i) parts, hence: $\omega_{r,1}$, $\omega_{i,1}$, $\omega_{r,2}$, $\omega_{i,2}$

T	Δ	$T^2-4\Delta$	$\omega_{r,1}$	$\omega_{i,1}$	$\omega_{r,2}$	$\omega_{i,2}$	Stability
$+$	$+$	$+$	$+$	0	$+$	0	unstable node
$-$	$+$	$+$	$-$	0	$-$	0	stable node
$-$	$+$	0	$-$	0	$-$	0	stable stellar node
$+$	$+$	0	$+$	0	$+$	0	unstable stellar node
$-$	$+$	$-$	$-$	$i\omega$	$-$	$-i\omega$	stable focus
$+$	$+$	$-$	$+$	$i\omega$	$+$	$-i\omega$	unstable focus
0	$+$	$-$	0	$i\omega$	0	$-i\omega$	centre marginal stability

unstable node

- Both roots are real, which happens only if $T^2 - 4\Delta \geqslant 0$. Since $\Delta > 0$, the two roots are of the same sign (sign of T). If $T > 0$, then both roots are positive, which means that the fluctuations $x = x_0 \exp(\omega t)$, $y = y_0 \exp(\omega t)$ *increase over time and the system will never return to the steady state ("unstable node")*. Thus the steady state represents a *repeller* of the concentrations X and Y.

stable node

- If, as in the previous case at $T^2 - 4\Delta \geqslant 0$, but this time $T < 0$, then both roots are negative, and this means that the fluctuations from the steady state will vanish ("*stable node*"). It looks as if we had in the steady state an *attractor* of the concentrations X and Y.

stable and unstable stellar nodes

- Now let us take $T^2 - 4\Delta = 0$, which means that the two roots are equal ("degeneracy"). This case is similar to the two previous ones. If the two roots are positive then the point is called the *stable stellar node* (attractor), if they are negative it is called the *unstable stellar node* (repeller).
- If $T^2 - 4\Delta < 0$, we have an interesting situation: both roots are complex conjugate $\omega_1 = \omega_r + i\omega_i$, $\omega_2 = \omega_r - i\omega_i$, or $\exp \omega_{1,2} t = \exp \omega_r t \exp(\pm i\omega_i t) = \exp \omega_r (\cos \omega_i t \pm i \sin \omega_i t)$. Note that $\omega_r = \frac{T}{2}$. We have therefore three special cases:

stable and unstable focuses

  - $T > 0$. Because of $\exp \omega_r t$ we have, therefore, a monotonic increase in the *fluctuations*, and at the same time because of $\cos \omega_i t \pm i \sin \omega_i t$ the two concentrations oscillate. Such a point is called the *unstable focus* (and represents a repeller).
  - $T < 0$. In a similar way we obtain the *stable focus*, which means some damped vanishing concentration oscillations (attractor).

centre marginal stability

  - $T = 0$. In this case $\exp \omega_{1,2} t = \exp(\pm i\omega_i t)$, i.e. we have the *undamped oscillations* of *X and Y about the stationary point* X_s, Y_s, which is called, in this case, the *centre marginal stability*.

Qualitative change

Can we qualitatively change the behaviour of the reaction? Yes. It is sufficient just to change the concentrations of A or B (i.e. to rotate the reactor taps). For example, let us gradually change B. Then, from eqs. (15.10), it follows that the key parameter T begins to change, which leads to an *abrupt qualitative change* in the behaviour (a catastrophe in the mathematical sense, p. 862). Such changes may be of great importance, and as the control switch may serve to regulate the concentrations of some substances in the reaction mixture.

Note that the *reaction is autocatalytic, because in step 3 the species X catalyzes the production of itself.*[34]

Brusselator with diffusion

If the stirrer were removed from the reactor, eqs. (15.3) have to be modified by adding diffusion terms

$$\frac{dX}{dt} = A - (B+1)X + X^2Y + D_X\frac{\partial^2 X}{\partial r^2}, \tag{15.12}$$

$$\frac{dY}{dt} = BX - X^2Y + D_Y\frac{\partial^2 Y}{\partial r^2}. \tag{15.13}$$

A stability analysis similar to that carried out a moment before results *not only in oscillations in time, but also in space, i.e. in the reaction tube there are waves of the concentrations* of X and Y *moving in space* (*dissipative structures*). Now, look at the photo of a zebra (Fig. 15.9) and at the bifurcation diagram in the logistic equation, Fig. 15.4.

dissipative structures

15.12.2 HYPERCYCLES

Let us imagine a system with a chain of consecutive chemical reactions. There are a lot of such reaction chains around, it is difficult to single out an elementary reaction without such a chain being involved. They end up with a final product and everything stops. What would happen however, if at a given point of the reaction chain, a substance X were created, the same as one of the reactants at a previous stage of the reaction chain? The X would take control over its own fate, by the Le Chatelier rule. In such a way, feedback would have been established, and instead of the chain, we would have a catalytic cycle. A system with feedback may adapt to changing external conditions, reaching a steady or oscillatory state. Moreover, in our system a number of such independent cycles may be present. However, when two of them share a common reactant X, both cycles would begin to cooperate, usually exhibiting a very complicated stability/instability pattern or an oscillatory character. We may think of coupling many such cycles in a *hypercycle*, etc.

hypercycle

Cooperating hypercycles based on multilevel supramolecular structures could behave in an extremely complex way when subject to variable fluxes of energy and matter.[35] No wonder, then, that a single photon produced by the prey hidden in the dark and absorbed by the retinal in the lynx's eye may trigger an enormous

---

[34] If autocatalysis were absent, our goal, i.e. concentration oscillations (dissipative structures), would not be achieved.

[35] Note that similar hypercycles function in economics...

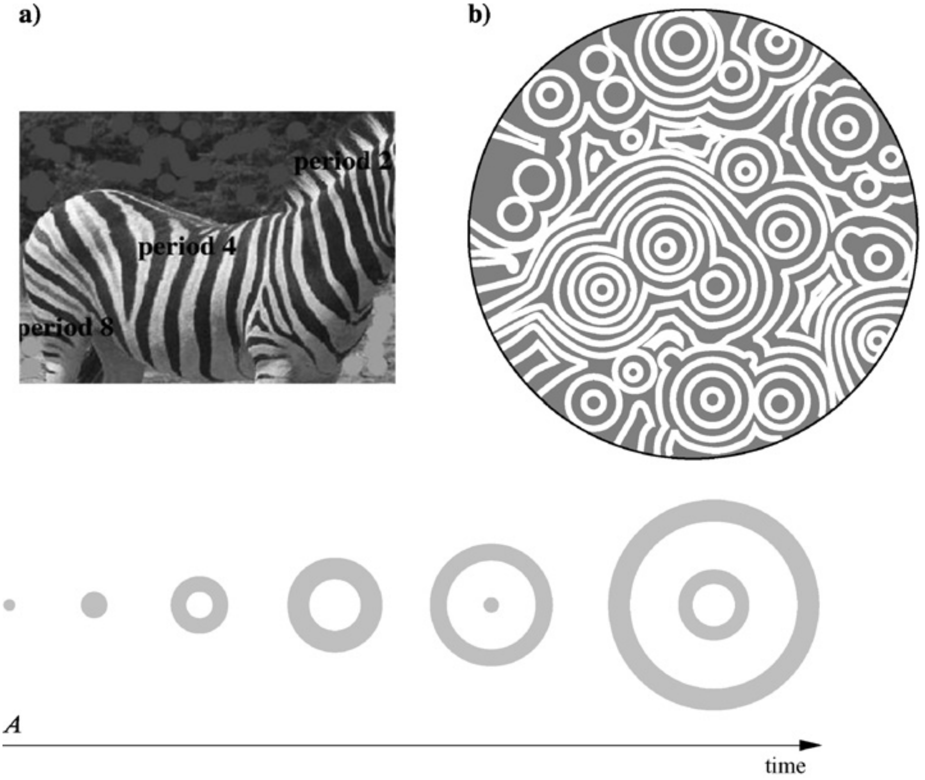


Fig. 15.9. (a) Such an animal "should not exist". Indeed, how did the *molecules* know that they have to make a beautiful pattern. I looked many times on zebras, but only recently I was struck by the observation that what I see on the zebra's skin is described by the logistic equation. The skin on the zebra's neck exhibits quasiperiodic oscillations of the black and white colour (period 2), in the middle of the zebra's body we have a *period doubling* (period 4), the zebra's back has period 8. Fig. (b) shows the waves of the chemical information (concentration oscillations in space and time) in the Belousov–Zhabotinski reaction from several sources in space. A "freezing" (for any reason) of the chemical waves leads to a striking similarity with the zebra's skin, from A. Babloyantz, "*Molecules, Dynamics and Life*", Wiley-Interscience Publ., New York, 1986, reproduced with permission from John Wiley and Sons, Inc. Fig. (c) shows similar waves of an epidemic in a rather immobile society. The epidemic broke out in centre A. Those who have contact with the sick person get sick, but after some time they regain their health, and *for some time* become immune. After the immune period is over these people get sick again, because there are a lot of microbes around. This is how the epidemic waves may propagate.

variety of hunting behaviours. Or, maybe from another domain: a single glimpse of a girl may change the fates of many people,[36] and sometimes the fate of the world. This is the retinal in the eye hit by the photon of a certain energy changes its conformation from cis to trans. This triggers a cascade of further processes, which end up as a nerve impulse travelling to the brain, and it is over.

---

[36] Well, think of a husband, children, grandchildren, etc.

CHEMICAL INFORMATION PROCESSING

15.13 FUNCTIONS AND THEIR SPACE-TIME ORGANIZATION

Using multi-level supramolecular architectures we may tailor new materials exhibiting desired properties, e.g., adapting themselves to changes in the neighbourhood ("smart materials"). Such materials have a function to perform, i.e. an action in time like ligand binding and/or releasing, transport of a ligand, an electron, a photon.[37]

A molecule may perform *several* functions. Sometimes these functions may be *coupled*, giving functional cooperation. The cooperation is most interesting when the system is far from thermodynamic equilibrium, and the equilibrium is most important when it is complex. In such a case the energy and matter fluxes result in structures with unique features.

Biology teaches us that an unbelievable effect is possible: molecules may spontaneously form some large aggregates with very complex dynamics and the whole system searches for energy-rich substances to keep itself running. However, one question evades answer: what is the goal of the system?

The molecular functions of very many molecules may be coupled in a complex space-time relationship on many time and space scales involving enormous transport problems at huge distances of the size of our body, engaging many structural levels, at the upper level the internal organs (heart, liver, etc.), which themselves have to cooperate[38] by exchanging *information*.

Chemists of the future will deal with molecular functions and their interactions. The achievements of today, such as molecular switches, molecular wires, etc. represent just simple elements of the big machinery of tomorrow.

15.14 THE MEASURE OF INFORMATION

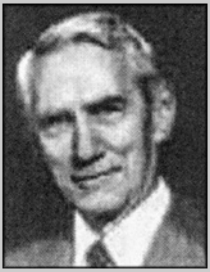
The TV News service presents a series of information items each evening. What kind of selection criteria are used by the TV managers? One of possible answers is that, for a given time period, they maximize the amount information given. A particular news bulletin contains a large amount of information, if it does not represent trivial common knowledge, but instead reports some unexpected facts. Claude Shannon defined the amount of information in a news bulletin as

$$\boxed{I = -\log_2 p, \qquad (15.14)}$$

---

[37] For example, a molecular antenna on one side of the molecule absorbs a photon, another antenna at the opposite end of the molecule emits another photon.

[38] This recalls the renormalization group or self-similarity problem in mathematics and physics.



Claude Elwood Shannon (1916–2001), American mathematician, professor at the Massachusetts Institute of Technology, his professional life was associated with the Bell Laboratories. His idea, now so obvious, that information may be transmitted as a sequence of "0" and "1" was shocking in 1948. It was said that Shannon used to understand problems '*in zero time*'.

where p stands for the probability of the event the information reports. How much information is contained in the news that in a single trial coin came down tails? Well, it is $I = -\log_2 \frac{1}{2} = 1$ bit. The news "*there is air in Paris*" is of no use in a TV news service,[39] because in this case $I = -\log_2 1 = 0$.

Claude Shannon introduced the notion of the average information associated with all possible N results of an event in the usual way

$$H = \sum_{i=1}^{N} p_i I_i = -\sum_{i=1}^{N} p_i \log_2 p_i, \qquad (15.15)$$

information entropy

where H is called the *entropy of information*, because a similar formula works in thermodynamics for entropy.

The quantity H (a measure of our ignorance) is largest, if all p_i are equal.

At a given instant we estimate the probabilities of all possible results of an event (we compute H_o), then reliable information arrives and the estimation changes (we compute the information entropy in the new situation H_f). Then, according to Shannon the measure of the information received is

$$I = H_o - H_f. \qquad (15.16)$$

codon

Example 1. ***Information flow in transcription.*** The sequence of three DNA bases (there are four bases possible: A, T, G, C), or a *codon*, codes for a single amino acid (there are 20 possible amino acids) in protein. Why three? Maybe three is too many? Let us see, what the problem looks like from the point of view of information flow.

A single codon carries the following information (in bits)

$$I_{\text{codon3}} = -\log_2\left(\frac{1}{4}\cdot\frac{1}{4}\cdot\frac{1}{4}\right) = 6,$$

---

[39] Now we know why the everyday TV news is full of thefts, catastrophes and unbridled crimes... Although it apparently looks upsetting, in fact it represents an optimistic signal: this is just incredibly rare.

while a single amino acid in a protein

$$I_{\mathrm{aa}} = -\log_2\left(\frac{1}{20}\right) = 4.23.$$

Thus, the codon information is sufficient to choose a particular amino acid from 20 possibles.

If nature designed the two-base codons, then such a codon would contain only $I_{\mathrm{codon2}} = -\log_2(\frac{1}{4}\cdot\frac{1}{4}) = 4$ *bits, which would be insufficient to code the amino acid.*

Thus, the protein coding that uses the information from the DNA sequence, takes place with the information excess of $6 - 4.23 = 1.77$ bits per amino acid.

15.15 THE MISSION OF CHEMISTRY

There is an impression that chemistry in biology is only a kind of substitute, a pretext, no more than a material carrier of the mission of the whole organism. Textbooks of biochemistry do not say much about chemistry, they talk about molecular functions to perform, in a sense about metachemistry. A particular molecule seems not to be so important. What counts is its function. A good example are enzymes. One *type* of enzyme may perform the same or similar functions in many different organisms (from fungi to man). The function is the same, but the composition of the enzyme changes from species to species: two species may differ by as much as 70% of the amino acids. However, those amino acids that are crucial for the enzyme function are preserved in all species.

We may perceive chemistry as a potential medium for information processing. This unbelievable chemical task would be collecting, transporting, changing, dispatching and transferring of information.

Chemistry, as we develop it, is far from such a masterpiece. What we are doing currently might be compared to chemical research by a Martian with a beautifully edited "Auguries of Innocence" by William Blake. The little green guy would perform a chemical analysis of the paper (he probably would even make a whole branch of science of that), examine the chemical composition of the printing dye; with other Martian Professors he would make some crazy hypotheses on the possible source of the leather cover, list the 26 different black signs as well as their perpendicular and horizontal clusters, analyze their frequencies, etc. He would, however, be very far from the *information* the book contains, including the boring matrix of black marks:

To see a world in a grain of sand
And heaven in a wild flower
Hold infinity in the palm of your hand
And eternity in an hour

and most importantly he could not even imagine his heart[40] beating any faster

---

[40] If any…

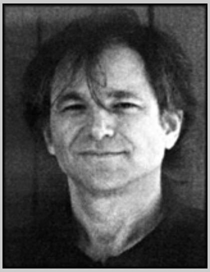
after reading this passage, because of thousands of associations *he could never have had*... We are close to what the Martian Professor would do. We have wonderful matter in our hands from which we could make chemical poems, but so far we are able to do only very little.

Molecules *could* play much more demanding roles than those, we have foreseen for them: *they can process information*. The first achievement in this direction came from Leonard Adleman – a mathematician.

15.16 MOLECULAR COMPUTERS BASED ON SYNTHON INTERACTIONS

Computers have changed human civilization. Their speed doubles every year or so, but the expectations are even greater. A possible solution is parallel processing, or making lots of computations at the same time, another is miniaturization. As will be seen in a moment, both these possibilities could be offered by molecular computers, in which the elementary devices would be the individual molecules chemists work with all the time. This stage of technology is not yet achieved. The highly elaborated silicon lithographic technology makes it possible to create electronic devices of size of the order of 1000 Å. Chemists would be able to go down to the hundreds or even tens of Å. Besides, the new technology would be based on self-organization (supramolecular chemistry) and self-assembling. In 1 cm^3 we could store the information of a huge number of todays CD-ROMs. People thought a computer had to have the form of a box with metallic and nonmetallic tools inside, as it is now. However,...

Leonard M. Adleman (b. 1945), American mathematician, professor of computer science and of molecular biology at the University of California, Los Angeles. As a young boy he dreamed of becoming a chemist, then a medical doctor. These dreams led him to the discovery described here.



polymerase

Turing machine

In 1994 mathematician Leonard M. Adleman[41] began his experiments in one of the American genetics labs, while learning the biological stuff in the evenings. Once, reading in bed Watson's textbook "*The Molecular Biology of the Gene*", he recognized that the features of the polymerase molecule interacting with the DNA strand described in the textbook perfectly match the features of what is called Turing machine, or, an *abstract representation of a computing device*, made just before the Second World War by Alan Turing.

Therefore, it was *certain* that the polymerase and the DNA (and certainly some other molecules) could be used as computers. If we think about it *now*, the computer in our head is more similar to *excusez le mot* water, than to a box with hard disks, etc. The achievement of Adleman was that he was able to translate a known and important mathematical problem into the language of laboratory recipes, and then using a chemical procedure he was able to solve the mathematical problem.

---

[41] L. Adleman, *Science* 266 (1994) 1021.

Alan Mathison Turing (1912–1954), British mathematical genius, in a paper in *Proc. London Math. Soc.* 42 (1937) 230), defined a simple device (known now as the Turing machine). The machine consists of a read/write head that scans a 1D tape divided into squares, each of which contains a "0" or "1". The behaviour of the machine is completely characterized by the current state of the machine, the content of the square it is just reading, and a table of instructions. Such a theoretical concept was of importance in considering the feasibility of any program coded on the tape. During the Second World War Turing continued Polish achievements by decoding further versions of the German Enigma code at Bletchley Park, the British wartime cryptanalytic headquarter. He was remembered for his eccentric habits. People saw him riding his bicycle with a gas mask on (he claimed it relieved his allergies). Alan Turing was found dead in his bed with a half eaten poisoned apple.

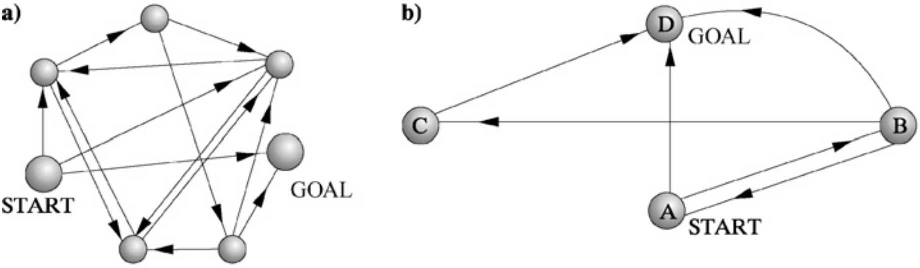


Fig. 15.10. A graph of airplane flights. Is the graph of the Hamilton type? This was a question for the molecular computer. (a) The graph from the Adleman's experiment. (b) A simplified graph described in this book.

Fig. 15.10 shows the original problem of Adleman: a graph with 14 airplane flights involving seven cities.

travelling salesman problem

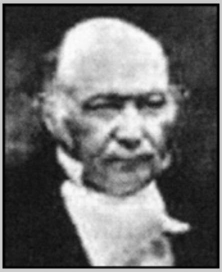
The task is called the *travelling salesman* problem, notorious in mathematics as extremely difficult.[42] The salesman begins his journey from the city START and wants to go to the city GOAL, visiting every other city precisely once. This is feasible only for some flight patterns. Those graphs for which it is feasible are called the *Hamilton graphs*. When the number of cities is small, such a problem may be quite effectively solved by the computer in our head. For seven cities it takes on average 56 s, as stated by Adleman, for a little larger number we need a desk computer, but *for a hundred* cities *all the computers of the world would be unable to provide the answer*. But,… a molecular computer would have the answer within a second.

Hamilton graphs

---

[42] The problem belongs to what is called NP-hard (NP from *non-polynomial*), in which the difficulties increase faster than any polynomial with the size of the problem.

William Rowan Hamilton (1805–1865) was a Astronomer Royal in Ireland. At the age of 17 he found an error in the famous “Celestial Mechanics” by Laplace. This drew the attention of scientists and was the beginning of the Hamilton’s scientific career. In the present book his name is repeated many times (because of Hamiltonian).



How does a molecular computer work?

Let us recall two important examples of complementary synthons: guanine and cytosine (GC) and adenine with tymine, see p. 751.

Let us repeat Adleman’s algorithm for a much simpler graph (Fig. 15.10.b). What Adleman did was the following.

1. He assigned for every city some particular piece of DNA (sequence) composed of eight nucleic bases:

City A	A	C	T	T	G	C	A	G
City B	T	C	G	G	A	C	T	G
City C	G	G	C	T	A	T	G	T
City D	C	C	G	A	G	C	A	A

2. Then to each existing flight X→Y, another eight-base DNA sequence was assigned: composed of the second half of the sequence of X and the first part of the sequence of Y:

Flight A→B	G	C	A	G	T	C	G	G
Flight A→D	G	C	A	G	C	C	G	A
Flight B→C	A	C	T	G	G	G	C	T
Flight B→D	A	C	T	G	C	C	G	A
Flight B→A	A	C	T	G	A	C	T	T
Flight C→D	A	T	G	T	C	C	G	A

3. Then, Adleman ordered the synthesis of the DNA sequences of the flights and the DNA sequences complementary to the cities, i.e.

co-City A	T	G	A	A	C	G	T	C
co-City B	A	G	C	C	T	G	A	C
co-City C	C	C	G	A	T	A	C	A
co-City D	G	G	C	T	C	G	T	T

4. All these substances are to be mixed together, dissolved in water, add a bit of salt and an enzyme called ligase.[43]

---

[43] To be as effective as Nature, we want to have conditions similar to those in living cells.

How to read the solution

What happened in the test tube? First of all matching and pairing of the corresponding synthons took place. For example, the DNA strand that codes the AB-flight (i.e. GCAGTCGG) found in the solution the complementary synthon of city B (i.e. the co-City AGCCTGAC) and because of the molecular recognition mechanism made a strong intermolecular complex:

$$\begin{array}{cccccccccccc}
G & C & A & G & T & C & G & G & & & & \\
 & & & & \vdots & \vdots & \vdots & \vdots & & & & \\
 & & & & A & G & C & C & T & G & A & C
\end{array}$$

where the upper part is flights, and the lower part is co-Cities. Note that the flights are the only feasible ones, because only feasible flights' DNA sequences were synthesized. The role of a co-City's DNA is to provide the information that there is the possibility to land and take-off in this particular city.

In the example just given, the complex will also find the synthon that corresponds to flight B → C, i.e. ACTGGGCT, and we obtain a more extended strand

$$\begin{array}{cccccccccccccccc}
G & C & A & G & T & C & G & G & |A & C & T & G & G & G & C & T \\
 & & & & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & & & & \\
 & & & & A & G & C & C & T & G & A & C & & & &
\end{array}$$

In this way from the upper part[44] of the intermolecular complexes we can read a particular itinerary. The ligase was needed, because this enzyme binds the loose ends of the DNA strands (thus removing the perpendicular separators above). *Therefore, every possible itinerary is represented by a DNA oligomer. If the graph were Hamiltonian, then in the solution there would be the DNA molecule composed of 24 nucleotides that codes the proper itinerary*:

GCAGTCGGACTGGGCTATGTCCGA.

Eliminating wrong trajectories...

Practically, independent of how large N is, after a second the solution to the travelling salesman problem is ready. The only problem now is to be able to read the solution. This will currently take much more than a second, but in principle only depends linearly on the number of cities.

To get the solution we use three techniques: polymerase chain reaction (PCR), electrophoresis and separation through affinity. The machinery behind all this is recognition of synthons and co-synthons (known in biochemistry as hybridization, it has nothing to do with hybridization described in Chapter 8).

---

[44] From the lower part as well.

The itineraries coded by the hybridization are mostly wrong. *One* of the reasons is that they do not start from the START CITY (A) and do not end up at the GOAL CITY (D). Using the PCR technique[45] it is possible to increase the concentration of only those itineraries, which start from START and end at GOAL to such an extent that all other concentrations may be treated as marginal.

Still there are a lot of wrong itineraries. First of all there are a lot of itineraries that are too long or too short. This problem may be fixed by electrophoresis,[46] which allows the separation of DNA strands of a given length, in our case the 24-city itineraries. In this way we have itineraries starting from START and ending at GOAL and having 24 cities. They can be copied again by PCR.

Now we have to eliminate more wrong itineraries: those which repeat some transit cities and leave others unvisited. This is done by the affinity separation method.[47] First, the co-synthon for the first transit city (in our case: C) on the list of transit cities (in our case: C and D) is prepared and attached to the surface of iron balls. The iron balls are then added to the solution and after allowing a second to bind to those itineraries that contain the city, they are picked out using a magnet. The balls are then placed in another test tube, the attached "itineraries" released from the surface of the iron balls and the empty iron balls are separated. Thus, we have in a test tube the "itineraries" that begin and end correctly, have the correct number of 24 nucleotides and certainly go through the first transit city (C) on our list of transit cities.

The process is repeated for the second etc. transit cities. If, in the last test tube, there is an "itinerary", the answer to the salesman problem is positive and the corresponding "itinerary" is identified (after copying by PCR and sequencing). Otherwise the answer is negative.

Thus, a mathematical problem was solved using a kind of molecular biocomputer. From the information processing point of view, this was possible because parallel processing was under way – a lot of DNA oligomers interacted with themselves at the same time. The number of such molecular processors was of the order of 10^{23}. This number is so huge, that such a biocomputer is able to check (virtually) all possibilities and to find the solution.

---

[45] The PCR technique is able to copy a chosen DNA sequence and to grow its population even from a single molecule to a high concentration by using the repeated action of an enzyme, a polymerase.

The reaction was invented by Kary B. Mullis (b. 1944), American technical chemist in an industrial company. In 1983 Mullis was driving to his favourite California surfing area, when the idea of a DNA copying molecular machine struck him suddenly. He stopped the car and made a note of the reaction. His company gave him a prize of $10 000 and sold the patent to another company for $300 000 000. In 1993 Kary Mullis received the Nobel Prize in chemistry "*for his invention of the polymerase chain reaction (PCTR) method*".

[46] Electrophoresis is able to physically separate DNA sequences according to their length. It is based on the electrolysis of a gel. Since DNA is an anion, it will travel through the gel to anode. The shorter the molecule, the longer distance it will reach. The DNA molecules of a given length can then be picked out by cutting the particular piece of gel and then they can be multiplied by PCR.

[47] Affinity separation method makes possible to separate particular sequences from a mixture of DNA sequences. This is achieved by providing its co-synthon attached to iron spheres. The particular sequence we are looking for binds to the surface of the iron ball, which may afterwards be separated from the solution using a magnet.

Summary

Chemistry has attained such a stage that soon a new quality can be achieved:

- chemistry entered the second half of the twentieth century with detailed knowledge of the main building blocks of molecular structures: atoms, chemical bonds, bond angles and intermolecular interactions;
- the accumulated knowledge now serves to build more and more complex molecular architectures;
- in these architectures we may use *chemical bonds* (with energy of the order of 50–150 kcal/mol) to build the molecules as well as *intermolecular interactions* (with energy of about 1–20 kcal/mol) to construct supramolecular structures from them;
- in supramolecular chemistry we operate with synthons, i.e. some special systems of functional groups that fit together perfectly when rigid ("key-lock" mechanism) or flexible ("hand-glove" mechanism), giving rise to molecular recognition;
- the interaction leads to a molecular complex that facilitates further evolution of the system: either by a chemical reaction going on selectively at such a configuration of the molecules, or by further self-organization due to next-step molecular recognition of the newly formed synthons;
- this may result in forming complex systems of multilevel architecture, each level characterized by its own stability;
- the self-organization may take place with significant interaction non-additivity effects ("non-linearity" in mathematical terms) that may lead to cooperation in forming the multilevel structure;
- high cooperation may lead to spontaneous transformation of the structure, called collective transformation, to another state ("domino effect");
- the self-organized structures may interact with other such structures (chemical reactions or association);
- in particular they may create the autocatalytic cycle which represents chemical feed back;
- such cycles may couple in a higher-order cycle forming hypercycles;
- a dynamic system with hypercycles, when perturbed by an external stimulus, reacts in a complex and non-linear way;
- one of the possibilities in non-equilibrium conditions are the limit cycles, which lead to dissipative structures, which may exhibit periodicity (in space and time) as well as chaotic behaviour;
- some dynamic systems may represent molecular libraries with the proportions of species strongly depending on external conditions (cf. the immune system);
- molecules may act (e.g., transfer photon, electron, proton, ion, conformational change, etc.) thus performing a function;
- several functions may cooperate exhibiting a space/time organization of the individual functions;
- some molecules may serve for effective information processing;
- information processing seems to represent the ultimate goal of the future chemistry.

Main concepts, new terms

complex systems (p. 852)
self-organization (p. 853)
cooperativity (p. 854)
combinatorial chemistry (p. 855)
molecular libraries (p. 855)
non-linearity (p. 857)
attractors (p. 858)
repellers (p. 858)

fixed point (p. 858)
limit cycle (p. 858)
logistic equation (p. 860)
chaos (p. 860)
bifurcation (p. 861)
catastrophe (p. 862)
domino (p. 863)
renormalization (p. 863)
collectivity (p. 863)
decimation (p. 865)
self-similarity (p. 865)
fractals (p. 865)
feed-back (p. 866)
autocatalysis (p. 868)
brusselator (p. 868)
nodes (stable and unstable, p. 872)
saddle point of reaction (p. 872)
stellar nodes (stable and unstable, p. 872)
focus (stable and unstable, p. 872)
reaction centre (p. 872)
dissipative structures (p. 873)
hypercycles (p. 873)
molecular function (p. 875)
information (p. 876)
information entropy (p. 876)
DNA computing (p. 878)
Turing machine (p. 878)
Hamilton graph (p. 879)
travelling salesman problem (p. 879)
NP-hard problem (p. 879)
DNA hybridization (p. 881)
PCR (p. 882)
separation by affinity (p. 882)

From the research front

To say that organic chemists are able to synthesize almost any molecule one may think of is certainly an exaggeration, but the statement seems sometimes to be very close to reality. Chemists were able to synthesize the five-olympic-ring molecule, the three interlocked Borromean rings, the football made of carbon atoms, the "cuban" – a hydrocarbon cube, "basketan" – in the form of an apple basket, the rotaxans shown in Fig. 13.2, a molecular in the form of Möbius band, etc. Now we may ask *why* the enormous synthetic effort was undertaken and what these molecules were synthesized for. Well, the answer seems to be that contemporary chemists are fascinated by their art of making complex and yet perfect and beautiful molecular objects. The main goal apparently was to demonstrate the mastership of modern chemistry. However, high symmetry does not necessarily means a particular usefulness. The synthetic targets should be identified by the careful planning of molecular *functions*, rather than molecular beauty.

Ad futurum...

We may expect that more and more often chemical research will focus on molecular function, and (later) on the space/time cooperation of the functions. Research projects will be formulated in a way that will highlight the role of the molecular function, and will consist of several (interrelated) steps:

- first, the technical goal will be defined,
- the molecular functions will be identified which will make this goal achievable,
- theoreticians will design and test in computers ("in silico") the molecules which will exhibit the above functions,
- synthetic chemists will synthesize the molecules designed,
- physicochemists will check whether the molecular functions are there,
- finally, the material will be checked against the technical goal.

We will be able to produce "smart" materials which will respond to external conditions in a previously designed, complex, yet we hope, predictable way. The materials that will be created this way will not resemble the materials of today, which are mostly carrying out one primitive function. The drugs of today are usually quite simple molecules, which enter the extremely complex system of our body. The drugs of tomorrow will involve much larger molecules (like proteins). Will we be clever enough to avoid unpredictable interactions with our body? What in principle do we want to achieve?

What will the motivation of our work be? Will we take into account the psychological needs of the human being, equilibrium of their minds?

What will the future of the human family be, which was able in the past to create such wonderful music, Chartres cathedral, breathtaking painting, moving poetry, abstract mathematics, proudly landed on other celestial bodies? In the past nothing could stop their curiosity and ingeniousness, they were able to resist the harshest conditions on their planet. Humans have reached nowadays the technical level that probably will assure avoiding the next glaciation,[48] maybe allow a *small* asteroid be pushed off the target by nuclear warheads if it were aimed dangerously at the Earth, also … erasing in nuclear war most of its own population together with the wonders of our civilization.

What is the goal of these beings and what will be the final limit of their existence? What are they aiming at? Do we want to know the smell of fresh bread, to be charmed by Chartres cathedral with all it has in it, to use our knowledge to cherish the friendship of the human family, or will it be sufficient to pack a newborn into a personal container and make computers inject substances that will make his neural system as happy as in Seventh Heaven?

Which of the goals we do want, as chemists, to participate in?

Additional literature

M. Eigen, P. Schuster, "The Hypercycle. A Principle of Natural Organization", Springer Verlag, Berlin, 1979.

An excellent, comprehensible book, written by the leading specialists in the domain of molecular evolution.

I. Prigogine, "From Being to Becoming. Time and Complexity in Physical Sciences", Freeman, San Francisco, 1980.

A book written by the most prominent specialist in the field.

A. Babloyantz, "Molecules, Dynamics and Life", Wiley, New York, 1987.

The author describes the scientific achievements of Prigogine and his group, which she participated in. An excellent, competent book, the most comprehensible among the first three recommended books.

J.-M. Lehn, "Supramolecular chemistry: Concept and Perspectives", VCH, 1995.

A vision of supramolecular chemistry given by one of its founders.

Questions

1. Decimation means:
   a) bifurcation; b) renormalization of the Hamiltonian and reaching self-similarity; c) scaling all the distances by a factor of ten; d) taking explicitly every tenth electron in a wave function.
2. A dissipative structure in a complex system:
   a) appears in a system far from equilibrium;
   b) means the largest molecular complex in the system;

---

[48] Well, it is expected within the next 500 years.

c) is independent of external conditions;
d) is the least stable structure appearing in equilibrium conditions.

3. A molecular library composed of the associates of the molecules A and B represents:
a) a mixture of the complexes AB; b) a mixture of all possible complexes of the A and B species; c) the complete physicochemical characterization of A and B; d) a mixture of all A_nB_n.

4. The self-organization of molecules is the spontaneous formation of:
a) molecular complexes only in equilibrium conditions; b) a structure with minimum entropy; c) a structure with maximum entropy; d) complexes of molecules with synthons.

5. In the iterative solution of the logistic equation $x_{n+1} = Kx_n(1 - x_n)$:
a) there is a fixed point at any K; b) at any attempt to increase of K we obtain a bifurcation; c) some values of K lead to chaotic behaviour; d) at no value of K do we have extinction of the population.

6. In the Brusselator without diffusion the stable focus means:
a) monotonic decreasing of the fluctuations x and y; b) dumped oscillations of the fluctuations x and y; c) non-vanishing oscillations of the fluctuations x and y; d) a limit circle.
7. In the thermodynamic equilibrium of an isolated system:
a) the entropy increases; b) we may have a non-zero gradient of temperature; c) we may have a non-zero gradient of concentration; d) no dissipative structures are possible.

8. The bifurcation point for the number of solutions of $x^2 - px + 2 = 0$ corresponds to:
a)$p = 2\sqrt{2}$; b) $p = 1$; c) $p = -1$; d) $p = \sqrt{2}$.

9. An event has only four possible outputs with *a priori* probabilities $p_1 = p_2 = p_3 = p_4 = \frac{1}{4}$. Reliable information comes that in fact the probabilities are different: $p_1 = \frac{1}{2}$, $p_2 = \frac{1}{4}$, $p_3 = \frac{1}{8}$, $p_4 = \frac{1}{8}$. The information had I_1 bits and I_1 is equal to:
a) 1 bit; b) 0.5 bit; c) 2 bits; d) 0.25 bit.

10. The situation corresponds to Question 9, but a second piece of reliable information coming says that the situation has changed once more and now: $p_1 = \frac{1}{2}$, $p_2 = 0$, $p_3 = 0$, $p_4 = \frac{1}{2}$. The second piece of information had I_2 bits. We pay for information in proportion to its quantity. Therefore, for the second piece of information we have to pay:
a) the same as for the first piece of information; b) twice as much as for the first piece of information; c) half of the prize for the first piece of information; d) three times more than for the first piece of information.

Answers

1b, 2a, 3b, 4d, 5c, 6b, 7d, 8a, 9d, 10d

APPENDICES

A. A REMINDER: MATRICES AND DETERMINANTS

1 MATRICES

Definition

A $n \times m$ matrix $\boldsymbol{A}$ represents a rectangular table of numbers[1] A_{ij} standing like soldiers in n perfect rows and m columns (index i tells us in which row, and index j tells in which column the number A_{ij} is located)

$$\boldsymbol{A} = \begin{pmatrix} A_{11} & A_{12} & \dots & A_{1m} \\ A_{21} & A_{22} & \dots & A_{2m} \\ \dots & \dots & \dots & \dots \\ A_{n1} & A_{n2} & \dots & A_{nm} \end{pmatrix}.$$

Such a notation allows us to operate whole matrices (like troops), instead of specifying what happens to each number ("soldier") separately. If matrices were not invented, the equations would be very long and clumsy, instead they are short and clear.

Addition

Two matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ may be *added* if their dimensions n and m match. The result is matrix $\boldsymbol{C} = \boldsymbol{A} + \boldsymbol{B}$ (of the same dimensions as $\boldsymbol{A}$ and $\boldsymbol{B}$), where each element of $\boldsymbol{C}$ is the sum of the corresponding elements of $\boldsymbol{A}$ and $\boldsymbol{B}$:

$$C_{ij} = A_{ij} + B_{ij},$$

e.g.,

$$\begin{pmatrix} 1 & -1 \\ -3 & 4 \end{pmatrix} + \begin{pmatrix} 2 & 1 \\ -2 & 3 \end{pmatrix} = \begin{pmatrix} 3 & 0 \\ -5 & 7 \end{pmatrix}.$$

Multiplying by a number

A matrix may be multiplied by a number by multiplying every element of the matrix by this number: $c\boldsymbol{A} = \boldsymbol{B}$ with $B_{ij} = cA_{ij}$. For example, $2\begin{pmatrix} 1 & -1 \\ 3 & -2 \end{pmatrix} = \begin{pmatrix} 2 & -2 \\ 6 & -4 \end{pmatrix}$.

---

[1] If instead of numbers a matrix contained functions, everything below would remain valid (at particular values of the variables instead of functions we would have their values).

Matrix product

The product of two matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ is matrix $\boldsymbol{C}$ denoted by $\boldsymbol{C} = \boldsymbol{A}\boldsymbol{B}$, its elements are calculated using elements of $\boldsymbol{A}$ and $\boldsymbol{B}$:

$$C_{ij} = \sum_{k=1}^{N} A_{ik} B_{kj},$$

where the number of columns (N) of matrix $\boldsymbol{A}$ has to be equal to the number of rows of matrix $\boldsymbol{B}$. The resulting matrix $\boldsymbol{C}$ has the number of rows equal to the number of rows of $\boldsymbol{A}$ and the number of columns equal to the number of columns of $\boldsymbol{B}$. Let us see how it works in an example. The product $\boldsymbol{A}\boldsymbol{B} = \boldsymbol{C}$:

$$\begin{pmatrix} A_{11} & A_{12} & A_{13} & A_{14} \\ \boldsymbol{A_{21}} & \boldsymbol{A_{22}} & \boldsymbol{A_{23}} & \boldsymbol{A_{24}} \\ A_{31} & A_{32} & A_{33} & A_{34} \end{pmatrix} \begin{pmatrix} B_{11} & B_{12} & \boldsymbol{B_{13}} & B_{14} & B_{15} & B_{16} & B_{17} \\ B_{21} & B_{22} & \boldsymbol{B_{23}} & B_{24} & B_{25} & B_{26} & B_{27} \\ B_{31} & B_{32} & \boldsymbol{B_{33}} & B_{34} & B_{35} & B_{36} & B_{37} \\ B_{41} & B_{42} & \boldsymbol{B_{43}} & B_{44} & B_{45} & B_{46} & B_{47} \end{pmatrix}$$

$$= \begin{pmatrix} C_{11} & C_{12} & C_{13} & C_{14} & C_{15} & C_{16} & C_{17} \\ C_{21} & C_{22} & \boldsymbol{C_{23}} & C_{24} & C_{25} & C_{26} & C_{27} \\ C_{31} & C_{32} & C_{33} & C_{34} & C_{35} & C_{36} & C_{37} \end{pmatrix},$$

e.g., C_{23} is simply the dot product of two vectors or in matrix notation

$$\begin{aligned} C_{23} &= \begin{pmatrix} A_{21} & A_{22} & A_{23} & A_{24} \end{pmatrix} \cdot \begin{pmatrix} B_{13} \\ B_{23} \\ B_{33} \\ B_{43} \end{pmatrix} \\ &= A_{21}B_{13} + A_{22}B_{23} + A_{23}B_{33} + A_{24}B_{43}. \end{aligned}$$

Some remarks:

- The result of matrix multiplication depends in general on whether we have to multiply $\boldsymbol{A}\boldsymbol{B}$ or $\boldsymbol{B}\boldsymbol{A}$, i.e. $\boldsymbol{A}\boldsymbol{B} \neq \boldsymbol{B}\boldsymbol{A}$.[2]
- Matrix multiplication satisfies the relation (easy to check): $\boldsymbol{A}(\boldsymbol{B}\boldsymbol{C}) = (\boldsymbol{A}\boldsymbol{B})\boldsymbol{C}$, i.e. the parentheses do not count and we can simply write: $\boldsymbol{A}\boldsymbol{B}\boldsymbol{C}$.
- Often we will have multiplication of a square matrix $\boldsymbol{A}$ by a matrix $\boldsymbol{B}$ composed of one column. Then, using the rule of matrix multiplication, we obtain matrix $\boldsymbol{C}$ in the form of a single column (with the number of elements identical to the dimension of $\boldsymbol{A}$):

$$\begin{pmatrix} A_{11} & A_{12} & \ldots & A_{1m} \\ A_{21} & A_{22} & \ldots & A_{2m} \\ \ldots & \ldots & \ldots & \ldots \\ A_{m1} & A_{m2} & \ldots & A_{mm} \end{pmatrix} \begin{pmatrix} B_1 \\ B_2 \\ \ldots \\ B_m \end{pmatrix} = \begin{pmatrix} C_1 \\ C_2 \\ \ldots \\ C_m \end{pmatrix}.$$

---

[2] Although it may happen, that $\boldsymbol{A}\boldsymbol{B} = \boldsymbol{B}\boldsymbol{A}$.

Transposed matrix

For a given matrix $\boldsymbol{A}$ we may define the transposed matrix $\boldsymbol{A}^T$ defined as $(\boldsymbol{A}^T)_{ij} = \boldsymbol{A}_{ji}$.

For example,

$$\text{if } \boldsymbol{A} = \begin{pmatrix} 1 & 2 \\ -2 & 3 \end{pmatrix}, \text{ then } \boldsymbol{A}^T = \begin{pmatrix} 1 & -2 \\ 2 & 3 \end{pmatrix}.$$

If matrix $\boldsymbol{A} = \boldsymbol{BC}$, then $\boldsymbol{A}^T = \boldsymbol{C}^T\boldsymbol{B}^T$, i.e. the order of multiplication is reversed. Indeed, $(\boldsymbol{C}^T\boldsymbol{B}^T)_{ij} = \sum_k (\boldsymbol{C}^T)_{ik}\, (\boldsymbol{B}^T)_{kj} = \sum_k \boldsymbol{C}_{ki}\, \boldsymbol{B}_{jk} = \sum_k \boldsymbol{B}_{jk}\boldsymbol{C}_{ki} = (\boldsymbol{BC})_{ji} = (\boldsymbol{A}^T)_{ij}$.

Inverse matrix

For some square matrices $\boldsymbol{A}$ (which will be called non-singular) we can define what is called the inverse matrix denoted as $\boldsymbol{A}^{-1}$, which has the following property: $\boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{A}^{-1}\boldsymbol{A} = \mathbf{1}$, where $\mathbf{1}$ stands for the unit matrix:

$$\mathbf{1} = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{pmatrix}.$$

For example, for matrix $\boldsymbol{A} = \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}$ we can find $\boldsymbol{A}^{-1} = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{3} \end{pmatrix}$.

For square matrices $\boldsymbol{A}\mathbf{1} = \mathbf{1}\boldsymbol{A} = \boldsymbol{A}$.

If we cannot find $\boldsymbol{A}^{-1}$ (because it does not exist), $\boldsymbol{A}$ is called a singular matrix. For example, matrix $\boldsymbol{A} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$ is singular. The inverse matrix for $\boldsymbol{A} = \boldsymbol{BC}$ is $\boldsymbol{A}^{-1} = \boldsymbol{C}^{-1}\boldsymbol{B}^{-1}$. Indeed, $\boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{BC}\boldsymbol{C}^{-1}\boldsymbol{B}^{-1} = \boldsymbol{B}\mathbf{1}\boldsymbol{B}^{-1} = \boldsymbol{B}\boldsymbol{B}^{-1} = \mathbf{1}$.

singular matrix

Adjoint, Hermitian, symmetric matrices

If matrix $\boldsymbol{A}$ is transposed and in addition all its elements are changed to their complex conjugate, we obtain the *adjoint matrix* denoted as $\boldsymbol{A}^\dagger = (\boldsymbol{A}^T)^* = (\boldsymbol{A}^*)^T$. If, for a square matrix, we have $\boldsymbol{A}^\dagger = \boldsymbol{A}$, $\boldsymbol{A}$ is called *Hermitian*. If $\boldsymbol{A}$ is real, then, of course, $\boldsymbol{A}^\dagger = \boldsymbol{A}^T$. If, in addition, for a real square matrix $\boldsymbol{A}^T = \boldsymbol{A}$, then $\boldsymbol{A}$ is called *symmetric*. Examples:

$$\boldsymbol{A} = \begin{pmatrix} 1+i & 3-2i \\ 2+i & 3-i \end{pmatrix};\ \boldsymbol{A}^T = \begin{pmatrix} 1+i & 2+i \\ 3-2i & 3-i \end{pmatrix};\ \boldsymbol{A}^\dagger = \begin{pmatrix} 1-i & 2-i \\ 3+2i & 3+i \end{pmatrix}.$$

Matrix $\boldsymbol{A} = \begin{pmatrix} 1 & -i \\ i & -2 \end{pmatrix}$ is an example of a Hermitian matrix, because $\boldsymbol{A}^\dagger = \boldsymbol{A}$. Matrix $\boldsymbol{A} = \begin{pmatrix} 1 & -5 \\ -5 & -2 \end{pmatrix}$ is a symmetric matrix.

Unitary and orthogonal matrices

If for a square matrix $\boldsymbol{A}$ we have $\boldsymbol{A}^\dagger = \boldsymbol{A}^{-1}$, $\boldsymbol{A}$ is called a *unitary* matrix. If $\boldsymbol{B}$ is Hermitian, the matrix $\exp(i\boldsymbol{B})$ is unitary, where we define $\exp(i\boldsymbol{B})$ by using the Taylor expansion: $\exp(i\boldsymbol{B}) = \mathbf{1} + i\boldsymbol{B} + \frac{1}{2!}(i\boldsymbol{B})^2 + \cdots$. Indeed, $[\exp(i\boldsymbol{B})]^\dagger = \mathbf{1} - i\boldsymbol{B}^T + \frac{1}{2!}(-i\boldsymbol{B}^T)^2 + \cdots = \mathbf{1} - i\boldsymbol{B} + \frac{1}{2!}(-i\boldsymbol{B})^2 + \cdots = \exp(-i\boldsymbol{B})$, while $\exp(i\boldsymbol{B})\exp(-i\boldsymbol{B}) = \mathbf{1}$.

If $\boldsymbol{A}$ is a real unitary matrix $\boldsymbol{A}^\dagger = \boldsymbol{A}^T$, it is called *orthogonal* with the property $\boldsymbol{A}^T = \boldsymbol{A}^{-1}$. For example, if

$$\boldsymbol{A} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}, \text{ then } \boldsymbol{A}^T = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} = \boldsymbol{A}^{-1}.$$

Indeed,

$$\boldsymbol{A}\boldsymbol{A}^T = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

2 DETERMINANTS

Definition

For any square matrix $\boldsymbol{A} = \{A_{ij}\}$ we may calculate a number called its determinant and denoted by $\det\boldsymbol{A}$ or $|\boldsymbol{A}|$. The determinant is calculated by using the Laplace expansion

$$\det\boldsymbol{A} = \sum_i^N (-1)^{i+j} A_{ij}\bar{A}_{ij} = \sum_j^N (-1)^{i+j} A_{ij}\bar{A}_{ij},$$

where (N is the dimension of the matrix) the result does not depend on which column j has been chosen in the first expression or which row i in the second expression. The symbol $\bar{A}_{ij}$ stands for the determinant of the matrix, which is obtained from $\boldsymbol{A}$ by removing the i-th row and the j-th column. Thus we have defined a determinant (of dimension N) by saying that it is a certain linear combination of determinants (of dimension $N-1$). It is then sufficient to say what we mean by a determinant that contains only one number c (i.e. has only one row and one column), this is simply $\det c \equiv c$.

For example, for matrix

$$\boldsymbol{A} = \begin{pmatrix} 1 & 0 & -1 \\ 2 & 2 & 4 \\ 3 & -2 & -3 \end{pmatrix},$$

its determinant is

$$\det\boldsymbol{A} = \begin{vmatrix} 1 & 0 & -1 \\ 2 & 2 & 4 \\ 3 & -2 & -3 \end{vmatrix} = (-1)^{1+1} \times 1 \times \begin{vmatrix} 2 & 4 \\ -2 & -3 \end{vmatrix} + (-1)^{1+2} \times 0 \times \begin{vmatrix} 2 & 4 \\ 3 & -3 \end{vmatrix}$$

$$+ (-1)^{1+3} \times (-1) \times \begin{vmatrix} 2 & 2 \\ 3 & -2 \end{vmatrix} = \begin{vmatrix} 2 & 4 \\ -2 & -3 \end{vmatrix} - \begin{vmatrix} 2 & 2 \\ 3 & -2 \end{vmatrix}$$

$$\begin{aligned}&= \big(2\times(-3)-4\times(-2)\big)-\big(2\times(-2)-2\times 3\big)\\&= 2+10=12.\end{aligned}$$

In particular, $\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad-bc.$

By repeating the Laplace expansion again and again (i.e. expanding $\bar{A}_{ij}$, etc.) we finally arrive at a linear combination of the products of the elements

$$\det \boldsymbol{A} = \sum_P (-1)^p \hat{P}[A_{1\mathbf{1}}A_{2\mathbf{2}}\cdots A_{N\mathbf{N}}],$$

where the permutation operator $\hat{P}$ pertains to the second indices (shown in bold), and p is the parity of permutation $\hat{P}$.

Slater determinant

In this book we will most often have to do with determinants of matrices whose elements are functions, not numbers. In particular what are called Slater determinants will be the most important. A Slater determinant for the N -electron system is built of functions called spinorbitals $\phi_i(j)$, $i=1,2,\dots,N$, where the *symbol* j denotes the space and spin coordinates $(x_j, y_j, z_j, \sigma_j)$ of electron j:

$$\psi(1,2,\dots,N) = \begin{vmatrix} \phi_1(1) & \phi_1(2) & \dots & \phi_1(N) \\ \phi_2(1) & \phi_2(2) & \dots & \phi_2(N) \\ \dots & \dots & \dots & \dots \\ \phi_N(1) & \phi_N(2) & \dots & \phi_N(N) \end{vmatrix}.$$

After this is done the Laplace expansion gives

$$\psi(1,2,\dots,N) = \sum_P (-1)^p \hat{P}\big[\phi_1(1)\phi_2(2)\dots\phi_N(N)\big],$$

where the summation is over $N!$ permutations of the N electrons, $\hat{P}$ stands for the permutation operator that acts on the *arguments* of the product of the spinorbitals $[\phi_1(1)\phi_2(2)\dots\phi_N(N)]$, p is the parity of permutation $\hat{P}$ (i.e. the number of the transpositions that change $[\phi_1(1)\phi_2(2)\dots\phi_N(N)]$ into $\hat{P}[\phi_1(1)\phi_2(2)\dots\phi_N(N)]$.

All properties of determinants pertain also to Slater determinants.

Some useful properties

- $\det \boldsymbol{A}^T = \det \boldsymbol{A}$.
- From the Laplace expansion it follows that if one of the spinorbitals is composed of two functions $\phi_i = \xi + \zeta$, the Slater determinant is the sum of the two Slater determinants, one with ξ instead of ϕ_i, the second with ζ instead of ϕ_i.
- If we add to a row (column) any linear combination of other rows (columns), the value of the determinant does not change.

- If a row (column) is a linear combination of other rows (columns), then $\det \boldsymbol{A} = 0$. In particular, if two rows (columns) are identical then $\det \boldsymbol{A} = 0$. Conclusion: in a Slater determinant the spinorbitals have to be linearly independent, otherwise the Slater determinant is equal zero.
- If in a matrix $\boldsymbol{A}$ we exchange two rows (columns), then $\det \boldsymbol{A}$ changes sign. Conclusion: the exchange of the coordinates of any two electrons leads to a change of sign of the Slater determinant (Pauli exclusion principle).
- $\det(\boldsymbol{AB}) = \det \boldsymbol{A} \det \boldsymbol{B}$.
- From the Laplace expansion it follows that multiplying the determinant by a number is equivalent to multiplication of an arbitrary row (column) by this number. Therefore, $\det(c\boldsymbol{A}) = c^N \det \boldsymbol{A}$, where N is the matrix dimension.[3]
- If matrix $\boldsymbol{U}$ is unitary then $\det \boldsymbol{U} = \exp(i\phi)$, where ϕ is a real number. This means that if $\boldsymbol{U}$ is an orthogonal matrix, $\det \boldsymbol{U} = \pm 1$.

---

[3] Note, that to multiply a matrix by a number we have to multiply every element of the matrix by this number. However, to multiply a determinant by a number means multiplication of one row (column) by this number.

B. A FEW WORDS ON SPACES, VECTORS AND FUNCTIONS

1 VECTOR SPACE

A vector space means a set V of elements $x, y, \ldots$, that form an Abelian group and can be "added" together[1] and "multiplied" by numbers $z = \alpha x + \beta y$ thus producing $z \in V$. The multiplication (α, β are, in general, complex numbers) satisfies the usual rules (the group is Abelian, because $x + y = y + x$):

$1 \cdot x = x,$

$\alpha(\beta x) = (\alpha\beta)x,$

$\alpha(x + y) = \alpha x + \alpha y,$

$(\alpha + \beta)x = \alpha x + \beta x.$

Example 1. ***Integers.*** The elements $x, y, \ldots$ are integers, the "addition" means simply the usual addition of integers, the numbers $\alpha, \beta, \ldots$ are also integers, "multiplication" means just usual multiplication. Does the set of integers form a vector space? Let us see. The integers form a group (with the addition as the operation in the group). Checking all the above axioms, we can easily prove that they are satisfied by integers. Thus, the integers (with the operations defined above) form a vector space.

Example 2. ***Integers with real multipliers.*** If, in the previous example, we admitted α, β to be real, the multiplication of integers x, y by real numbers would give real numbers (not necessarily integers). Therefore, in this case $x, y, \ldots$ *do not* represent any vector space.

Example 3. ***Vectors.*** Suppose $x, y, \ldots$ are vectors, each represented by a N-element sequence of real numbers (they are called the vector "components") $x = (a_1, a_2, \ldots, a_N)$, $y = (b_1, b_2, \ldots, b_N)$, etc. Their addition $x + y$ is an operation that produces the vector $z = (a_1 + b_1, a_2 + b_2, \ldots, a_N + b_N)$. The vectors form an Abelian group, because $x + y = y + x$, the unit ("neutral") element is $(0, 0, \ldots, 0)$, the inverse element to $(a_1, a_2, \ldots, a_N)$ is equal to $(-a_1, -a_2, \ldots, -a_N)$. Thus, the vectors form a group. "Multiplication" of a vector by a real number α means $\alpha(a_1, a_2, \ldots, a_N) = (\alpha a_1, \alpha a_2, \ldots, \alpha a_N)$. Please check that the four axioms above are satisfied. Conclusion: the vectors form a vector space.

---

[1] See Appendix C, to form a group any pair of the elements can be "added" (operation in the group), the addition is associative, there exists a unit element and for each element an inverse exists.

Note that if only the positive vector components were allowed, they would not form an Abelian group (no neutral element), and on top of this their addition (which might mean a subtraction of components, because α, β could be negative) could produce vectors with non-positive components. Thus vectors with all positive components do not form a vector space.

Example 4. ***Functions.*** This example is important in the context of this book. This time the vectors have real components.[2] Their "addition" means the addition of two functions $f(x) = f_1(x) + f_2(x)$. The "multiplication" means multiplication by a real number. The unit ("neutral") function means $f = 0$, the "inverse" function to f is $-f(x)$. Therefore, the functions form an Abelian group. A few seconds are needed to show that the four axioms above are satisfied. Such functions form a vector space.

Linear independence. A set of vectors is called a set of linearly independent vectors if no vector of the set can be expressed as a linear combination of the other vectors of the set. The number of linearly independent vectors in a vector space is called the dimension of the space.

Basis means a set of n linearly independent vectors in n-dimensional space.

2 EUCLIDEAN SPACE

A vector space (with multiplying real numbers α, β) represents the Euclidean space, if for any two vectors x, y of the space we assign a real number called an *inner* (*or scalar*) *product* $\langle x|y\rangle$ with the following properties:

- $\langle x|y\rangle = \langle y|x\rangle$,
- $\langle \alpha x|y\rangle = \alpha\langle x|y\rangle$,
- $\langle x_1 + x_2|y\rangle = \langle x_1|y\rangle + \langle x_2|y\rangle$,
- $\langle x|x\rangle = 0$, only if $x = 0$.

Inner product and distance. The concept of the inner product is used to introduce

- the *length of the vector* x defined as $\|x\| \equiv \sqrt{x|x}$, and
- the *distance between two vectors* x and y as a non-negative number $\|x - y\| = \sqrt{\langle x - y|x - y\rangle}$. The distance satisfies some conditions which we treat as obvious from everyday experience:
- the distance from Paris to Paris has to equal zero (just insert $x = y$);
- the distance from Paris to Rome has to be the same as from Rome to Paris (just exchange $x \leftrightarrow y$);
- the Paris–Rome distance is equal to or shorter than the sum of two distances: Paris–X and X–Rome for any town X (a little more difficult to show).

---

[2] Note the similarity of the present example with the previous one: a function $f(x)$ may be treated as a vector with an infinite number of components. The components are listed in the sequence of increasing $x \in R$, the component $f(x)$ corresponding to x.

Schwarz inequality. For any two vectors belonging to the Euclidean space the Schwarz inequality holds[3]

$$\left|\langle x|y\rangle\right| \leqslant \|x\|\|y\| \tag{B.1}$$

or equivalently $|\langle x|y\rangle|^2 \leqslant \|x\|^2\|y\|^2$.

Orthogonal basis means that all basis vectors x_j, $j = 1, 2, \ldots, N$, are orthogonal to each other: $\langle x_i|x_j\rangle = 0$ for $i \neq j$.

Orthonormal basis is an orthogonal basis set with all the basis vectors of length $\|x_i\| = 1$. Thus for the orthonormal basis set we have $\langle x_i|x_j\rangle = \delta_{ij}$, where $\delta_{ij} = 1$ for $i = j$ and $\delta_{ij} = 0$ for $i \neq j$ (Kronecker delta).

Example 5. ***Dot product.*** Let us take the vector space from Example 3 and introduce the dot product (representing the inner product) defined as

$$\langle x|y\rangle = \sum_{i=1}^{N} a_i b_i. \tag{B.2}$$

Let us check whether this definition satisfies the properties required for a inner product:

- $\langle x|y\rangle = \langle y|x\rangle$, because the order of a and b in the product is irrelevant.
- $\langle \alpha x|y\rangle = \alpha\langle x|y\rangle$, because the sum says that multiplication of each a_i by α is equivalent to multiplying the inner product by α.
- $\langle x_1 + x_2|y\rangle = \langle x_1|y\rangle + \langle x_2|y\rangle$, because if vector x is decomposed into two vectors $x = x_1 + x_2$ in such a way that $a_i = a_{i1} + a_{i2}$ (with a_{i1}, a_{i2} being the components of x_1, x_2, respectively), the summation of $\langle x_1|y\rangle + \langle x_2|y\rangle$ gives $\langle x|y\rangle$.
- $\langle x|x\rangle = \sum_{i=1}^{N}(a_i)^2$, and this equals zero if, and only if, all components $a_i = 0$. Therefore, the proposed formula operates as the inner product definition requires.

3 UNITARY SPACE

If three changes were introduced into the definition of the Euclidean space, we would obtain another space: the unitary space. These changes are as follows:

- the numbers $\alpha, \beta, \ldots$ instead of being real are complex;
- the inner product, instead of $\langle x|y\rangle = \langle y|x\rangle$ has the property $\langle x|y\rangle = \langle y|x\rangle^*$;
- instead of $\langle \alpha x|y\rangle = \alpha\langle x|y\rangle$ we have:[4] $\langle \alpha x|y\rangle = \alpha^*\langle x|y\rangle$.

---

[3] The Schwarz inequality agrees with what everyone recalls about the dot product of two vectors: $\langle x|y\rangle = \|x\|\|y\|\cos\theta$, where θ is the angle between the two vectors. Taking the absolute value of both sides, we obtain $|\langle x|y\rangle| = \|x\|\|y\||\cos\theta| \leqslant \|x\|\|y\|$.

[4] While we still have $\langle x|\alpha y\rangle = \alpha\langle x|y\rangle$.

After the new inner product definition is introduced, related quantities: the length of a vector and the distance between the vectors are defined in exactly the same way as in the Euclidean space. Also the definitions of orthogonality and of the Schwarz inequality remain unchanged.

4 HILBERT SPACE

This is for us the most important unitary space – its elements are wave functions, which instead of $x, y, \ldots$ will be often denoted as $f, g, \ldots, \phi, \chi, \psi, \ldots$ etc. The wave functions which we are dealing with in quantum mechanics (according to John von Neumann) are the elements (i.e. vectors) of the Hilbert space. The inner product of two functions f and g means $\langle f|g\rangle \equiv \int f^* g \, d\tau$, where the integration is over the whole space of variables, on which both functions depend. The length of vector f is denoted by $\|f\| = \sqrt{\langle f|f\rangle}$. Consequently, the orthogonality of two functions f and g means $\langle f|g\rangle = 0$, i.e. an integral $\int f^* g \, d\tau = 0$ over the whole range of the coordinates on which the function f depends. The Dirac notation (1.9) is in fact the inner product of such functions in a unitary space.

Let us imagine an infinite sequence of functions (i.e. vectors) $f_1, f_2, f_3, \ldots$ in a unitary space, Fig. B.1. The sequence will be called a Cauchy sequence, if for a given $\varepsilon > 0$ a natural number N can be found, such that for $i > N$ we will have $\|f_{i+1} - f_i\| < \varepsilon$. In other words, in a Cauchy sequence the distances between consecutive vectors (functions) decrease when we go to sufficiently large indices, i.e. the functions become more and more similar to each other. *If the converging Cauchy sequences have their limits (functions) which belong to the unitary space, such a space is called a Hilbert space*.

A basis in the Hilbert space is such a set of linearly independent functions (vectors) that any function belonging to the space can be expressed as a linear combination of the basis set functions. Because of the infinite number of dimensions, the number of the basis set functions is infinite. This is difficult to imagine. In a way analogous to a 3D Euclidean space, we may imagine an orthonormal basis as the unit vectors protruding from the origin in an infinite number of directions (like a "hedgehog", Fig. B.2).

David Hilbert (1862–1943), German mathematician, professor at the University of Göttingen. At the II Congress of Mathematicians in Paris Hilbert formulated 23 goals for mathematics he considered to be very important. This had a great impact on mathematics and led to some unexpected results (e.g., Gödel theorem, cf. p. 851). Hilbert's investigations in 1900–1910 on integral equations resulted in the concept of the Hilbert space. Hilbert also worked on the foundations of mathematics, on mathematical physics, number theory, variational calculus, etc. This hard working and extremely prolific mathematician was deeply depressed by Hitler's seizure of power. He regularly came to his office, but did not write a single sheet of paper.

Each vector (function) can be represented as a linear combination of the hedgehog functions. We see that we may rotate the "hedgehog" (i.e. the basis set)[5] and the completeness of the basis will be preserved, i.e. any vector of the Hilbert space can be represented as a linear combination of the new basis set vectors.

---

[5] The new orthonormal basis set is obtained by a unitary transformation of the old one.

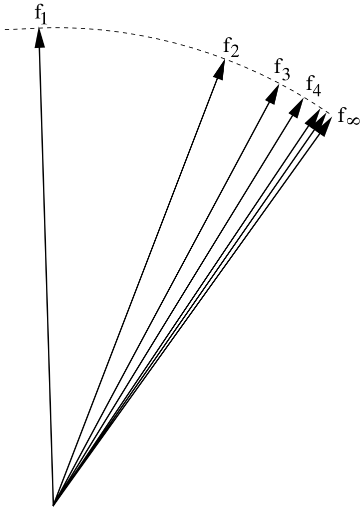


Fig. B.1. A pictorial representation of the Hilbert space. We have a vector space (each vector represents a wave function) and a series of unit vectors f_i that differ less and less (Cauchy series). If any convergent Cauchy series has its limit belonging to the vector space, the space represents the *Hilbert space.*

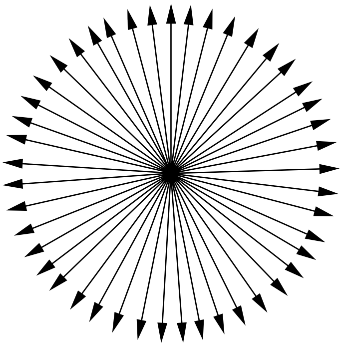


Fig. B.2. A pictorial representation of something that surely cannot be represented: an orthonormal basis in the Hilbert space looks like a hedgehog of the unit vectors (their number equal to ∞), each pair of them orthogonal. This is analogous to a 2D or 3D basis set, where the hedgehog has two or three orthogonal unit vectors.

Linear operator

Operator $\hat{A}$ transforms any vector x from the operator's domain into vector y (both vectors x, y belong to the unitary space): $\hat{A}(x) = y$, which is written as $\hat{A}x = y$. A *linear operator* satisfies $\hat{A}(c_1x_1 + c_2x_2) = c_1\hat{A}x_1 + c_2\hat{A}x_2$, where c_1 and c_2 stand for complex numbers.

We define:

- Sum of operators: $\hat{A} + \hat{B} = \hat{C}$ as $\hat{C}x = \hat{A}x + \hat{B}x$.
- Product of operators: $\hat{A}\hat{B} = \hat{C}$ as $\hat{C}x = \hat{A}(\hat{B}(x))$.

commutation

If, for two operators, we have $\hat{A}\hat{B} = \hat{B}\hat{A}$, we say they *commute*, or their *commutator* $[\hat{A}, \hat{B}] \equiv \hat{A}\hat{B} - \hat{B}\hat{A} = 0$. In general $\hat{A}\hat{B} \neq \hat{B}\hat{A}$, i.e. the operators do not commute.

- Inverse operator (if it exists): $\hat{A}^{-1}(\hat{A}x) = x$.

Adjoint operator

If, for an operator $\hat{A}$, we can find a new operator $\hat{A}^{\dagger}$, such that for any two vectors x and y of the unitary space[6] we have[7]

$$\langle x|\hat{A}y\rangle = \langle \hat{A}^{\dagger}x|y\rangle \tag{B.3}$$

then we say that $\hat{A}^{\dagger}$ is the *adjoint* operator to $\hat{A}$.

Hermitian operator

If $\hat{A}^{\dagger} = \hat{A}$, we will call operator $\hat{A}$ a *self-adjoint* or *Hermitian* operator:[8]

$$\langle x|\hat{A}y\rangle = \langle \hat{A}x|y\rangle. \tag{B.4}$$

Unitary operator

A unitary operator $\hat{U}$ transforms a vector x into $y = \hat{U}x$ both belonging to the unitary space (the domain is the unitary space) and the *inner product is preserved*:

$$\langle \hat{U}x|\hat{U}y\rangle = \langle x|y\rangle.$$

This means that any unitary transformation preserves the angle between the vectors x and y, i.e. the angle between x and y is the same as the angle between $\hat{U}x$ and $\hat{U}y$. The transformation also preserves the length of the vector, because $\langle \hat{U}x|\hat{U}x\rangle = \langle x|x\rangle$. This is why operator $\hat{U}$ can be thought of as a *transformation related to a motion in the unitary space* (rotation, reflection, etc.). For a unitary operator we have $\hat{U}^{\dagger}\hat{U} = 1$, because $\langle \hat{U}x|\hat{U}y\rangle = \langle x|\hat{U}^{\dagger}\hat{U}y\rangle = \langle x|y\rangle$.

5 EIGENVALUE EQUATION

If, for a particular vector x, we have

$$\hat{A}x = ax, \tag{B.5}$$

where a is a complex number and $x \neq 0$, x is called the eigenvector[9] of operator $\hat{A}$ corresponding to eigenvalue a. Operator $\hat{A}$ may have an infinite number, a finite number including number zero of the eigenvalues, labelled by subscript i:

---

[6] The formal definition is less restrictive and the domains of the operators $\hat{A}^{\dagger}$ and $\hat{A}$ do not need to extend over the whole unitary space.

[7] Sometimes we make a useful modification in the Dirac notation: $\langle x|\hat{A}y\rangle \equiv \langle x|\hat{A}|y\rangle$.

[8] The self-adjoint and Hermitian operators differ in mathematics (a matter of domains), but we will ignore this difference in the present book.

[9] In quantum mechanics, vector x will correspond to a function (a vector in the Hilbert space) and therefore is called the eigenfunction.

$$\hat{A}x_i = a_i x_i.$$

Hermitian operators have the following important properties:[10]

> If $\hat{A}$ represents a Hermitian operator, its eigenvalues a_i are real numbers, and its eigenvectors x_i, which correspond to different eigenvalues, are orthogonal.

The number of linearly independent eigenvectors which correspond to a given eigenvalue a is called the degree of degeneracy of the eigenvalue. Such vectors form the basis of the *invariant space of operator* $\hat{A}$, i.e. any linear combination of the vectors represents a vector that is also an eigenvector (with the same eigenvalue a). If the eigenvectors corresponded to different eigenvalues, their linear combination *is not* an eigenvector of $\hat{A}$. Both things need a few seconds to show.

degeneracy

One can show that the eigenvectors of a Hermitian operator form the complete basis set[11] in Hilbert space, i.e. any function of class Q[12] can be expanded in a linear combination of the basis set.

Sometimes an eigenvector x of operator $\hat{A}$ (with eigenvalue a) is subject to an operator $f(\hat{A})$, where f is an analytic function. Then,[13]

$$f(\hat{A})x = f(a)x. \tag{B.6}$$

Commutation and eigenvalues

We will sometimes use the theorem that, if two linear and Hermitian operators $\hat{A}$ and $\hat{B}$ commute, they have a common set of eigenfunctions and *vice versa*.

---

[10] We have the eigenvalue problem $\hat{A}x = ax$. Making a complex conjugate of both sides, we obtain $(\hat{A}x)^* = a^*x^*$. Multiplying the first of the equations by x^* and integrating, and then using x and doing the same with the second equation, we get: $\langle x|\hat{A}x\rangle = a\langle x|x\rangle$ and $\langle \hat{A}x|x\rangle = a^*\langle x|x\rangle$. But $\hat{A}$ is Hermitian, and therefore the left-hand sides of both equations are equal. Subtracting them we have $(a - a^*)\langle x|x\rangle = 0$. Since $\langle x|x\rangle \neq 0$, because $x \neq 0$, then $a = a^*$. This is what we wanted to show.

The orthogonality of the eigenfunctions of a Hermitian operator (corresponding to different eigenvalues) may be proved as follows. We have $\hat{A}x_1 = a_1x_1$, $\hat{A}x_2 = a_2x_2$, with $a_1 \neq a_2$. Multiplying the first equation by x_2^* and integrating, we obtain $\langle x_2|\hat{A}x_1\rangle = a_1\langle x_2|x_1\rangle$. Then, let us make the complex conjugate of the second equation: $(\hat{A}x_2)^* = a_2x_2^*$, where we have used $a_2 = a_2^*$ (this was proved above). Then let us multiply by x_1 and integrate: $\langle \hat{A}x_2|x_1\rangle = a_2\langle x_2|x_1\rangle$. Subtracting the two equations, we have $0 = (a_1 - a_2)\langle x_2|x_1\rangle$, and taking into account $a_1 - a_2 \neq 0$ this gives $\langle x_2|x_1\rangle = 0$.

[11] This basis set may be assumed to be orthonormal, because the eigenfunctions

- as square-integrable can be normalized;
- if they correspond to different eigenvalues, are automatically orthogonal;
- if they correspond to the same eigenvalue, they can be orthogonalized (still remaining eigenfunctions) by a method described in Appendix J.

[12] That is, continuous, single-valued and square integrable, see Fig. 2.5.

[13] The operator $f(\hat{A})$ is defined through the Taylor expansion of the function f: $f(\hat{A}) = c_0 + c_1\hat{A} + c_2\hat{A}^2 + \cdots$. If the operator $f(\hat{A})$ now acts on an eigenfunction of $\hat{A}$, then, because $\hat{A}^n x = a^n x$, we obtain the result.

We will prove this theorem in the case of no degeneracy (i.e. there is only one linearly independent vector for a given eigenvalue). We have an eigenvalue equation $\hat{B}y_n = b_n y_n$. Applying this to both sides of operator $\hat{A}$ and using the commutation relation $\hat{A}\hat{B} = \hat{B}\hat{A}$ we have: $\hat{B}(\hat{A}y_n) = b_n(\hat{A}y_n)$. This means that $\hat{A}y_n$ is an eigenvector of $\hat{B}$ corresponding to the eigenvalue b_n. But, we already know such a vector, this is y_n. The two vectors have to be proportional: $\hat{A}y_n = a_n y_n$, which means that y_n is an eigenvector of $\hat{A}$.

Now, the inverse theorem. We have two operators and any eigenvector of $\hat{A}$ is also an eigenvector of $\hat{B}$. We want to show that the two operators commute. Let us write the two eigenvalue equations: $\hat{A}y_n = a_n y_n$ and $\hat{B}y_n = b_n y_n$. Let us take a vector ϕ. Since the eigenvectors $\{y_n\}$ form the complete set, then

$$\phi = \sum_n c_n y_n.$$

Applying the commutator $[\hat{A}, \hat{B}] = \hat{A}\hat{B} - \hat{B}\hat{A}$ we have

$$\begin{aligned}
[\hat{A}, \hat{B}]\phi &= \hat{A}\hat{B}\phi - \hat{B}\hat{A}\phi = \hat{A}\hat{B}\sum_n c_n y_n - \hat{B}\hat{A}\sum_n c_n y_n \\
&= \hat{A}\sum_n c_n \hat{B}y_n - \hat{B}\sum_n c_n \hat{A}y_n = \hat{A}\sum_n c_n b_n y_n - \hat{B}\sum_n c_n a_n y_n \\
&= \sum_n c_n b_n \hat{A}y_n - \sum_n c_n a_n \hat{B}y_n = \sum_n c_n b_n a_n y_n - \sum_n c_n a_n b_n y_n = 0.
\end{aligned}$$

This means that the two operators commute.

C. GROUP THEORY IN SPECTROSCOPY

Quite a lot of what we will be talking about in this Appendix was invented by Evariste Galois. He was only 21 when he died in a duel (*cherchez la femme*!). Galois spent his last night writing down his group theory.

Evariste Galois (1811–1832), French mathematician, also created many fundamental ideas in the theory of algebraic equations.

The group theory in this textbook will be treated in a practical way, as one of useful tools.[1]

> Our goal will be to predict the selection rules in ultraviolet (UV), visual (VIS) and infrared (IR) molecular spectra.

We will try to be concise, but examples need explanations, there are few lovers of dry formulae.

1 GROUP

Imagine a set of elements $\hat{R}_i$, $i = 1, 2, \dots, g$. We say that they form a group G of order[2] g, if the following four conditions are satisfied:

1. An operation exists called "*multiplication*", $\hat{R}_i \cdot \hat{R}_j$, which associates every pair of the elements of G with another element of G, i.e. $\hat{R}_i \cdot \hat{R}_j = \hat{R}_k$. Hereafter

---

[1] Rather than as a field of abstract mathematics. Symmetry may be viewed either as something beautiful or primitive. It seems that, from the psychological point of view, symmetry stresses people's longing for simplicity, order and understanding. On the other hand, symmetry means less information and hence often a kind of wearingly dull stimuli. Possibly the interplay between these two opposite features leads us to consider *broken* symmetry as beautiful. Happily enough, trees and leaves exhibit broken symmetry and look beautiful. Ancient architects knew the secrets of creating beautiful buildings, which relied on breaking symmetry, in a substantial way, but almost invisible from a distance.

[2] g may be finite or infinite. In most applications of the present Appendix, g will be finite.

the multiplication $\hat{R}_i \cdot \hat{R}_j$ will be denoted simply as $\hat{R}_i\hat{R}_j$. Thus the elements can multiply each other and the result always belongs to the group.

2. The multiplication is associative,[3] i.e. for any three elements of G we have $\hat{R}_i(\hat{R}_j\hat{R}_k) = (\hat{R}_i\hat{R}_j)\hat{R}_k$.
3. Among $\hat{R}_i \in G$ an *identity* element exists, denoted by $\hat{E}$, with a nice property: $\hat{R}_i\hat{E} = \hat{R}_i$ and $\hat{E}\hat{R}_i = \hat{R}_i$ for any i.
4. For each $\hat{R}_i$ we can find an element of G (denoted as $\hat{R}_i^{-1}$, called the *inverse* element with respect to $\hat{R}_i$), such that $\hat{R}_i\hat{R}_i^{-1} = \hat{E}$, also $\hat{R}_i^{-1}\hat{R}_i = \hat{E}$.

Example 1. ***A four-element group.*** The elements $1, -1, i, -i$ with the chosen operation the ordinary multiplication of numbers, form a group of order 4. Indeed, any product of these numbers gives one of them. Here is the corresponding "multiplication table"

		second in the product			
		1	-1	i	$-i$
first in the product	1	1	-1	i	$-i$
	-1	-1	1	$-i$	i
	i	i	$-i$	-1	1
	$-i$	$-i$	i	1	-1

Note that

ABELIAN GROUP:
The table is *symmetric* with respect to the diagonal. A group with a symmetric multiplication table is called Abelian.

Abelian group

The associativity requirement is of course satisfied. The unit element is 1. You can always find an inverse element. Indeed, for 1 it is 1, for -1 it is -1, for i it is $-i$, for $-i$ it is i. Thus, all conditions are fulfilled and $g = 4$.

Example 2. ***Group of integers.*** Let us take as G the set of integers with the "multiplication" being the regular addition of numbers. Let us check. The sum of two integers is an integer, so requirement 1 is satisfied. The operation is associative, because addition is. The unit element is, of course, 0. The inverse element to an integer means the opposite number. Thus, G is a group of order $g = \infty$.

Example 3. ***Group of non-singular matrices.*** All non-singular $n \times n$ matrices[4] with matrix multiplication as the operation, form a group. Let us see. Multiplication of a non-singular matrix $\boldsymbol{A}$ (i.e. $\det \boldsymbol{A} \neq 0$) by a non-singular matrix $\boldsymbol{B}$ gives a non-singular matrix $\boldsymbol{C} = \boldsymbol{AB}$, because $\det \boldsymbol{C} = \det \boldsymbol{A} \det \boldsymbol{B} \neq 0$. The unit element is the unit matrix $\mathbf{1}$, the inverse element exists (this is why we needed the non-singularity) and is equal to $\boldsymbol{A}^{-1}$. Also from the matrix multiplication rule we have $(\boldsymbol{AB})\boldsymbol{C} = \boldsymbol{A}(\boldsymbol{BC})$. This is a group of order ∞.

---

[3] Thanks to this, expressions similar to $\hat{R}_i\hat{R}_j\hat{R}_k$ have unambiguous meaning.

[4] See Appendix A.

Example 4. ***Group of unitary matrices*** U(n). In particular, all the unitary $n \times n$ matrices form a group with matrix multiplication as the group multiplication operation. Let us check. Any such multiplication is feasible and the product represents a unitary matrix (if matrices $\boldsymbol{U}_1$ and $\boldsymbol{U}_2$ are unitary, i.e. $\boldsymbol{U}_1^\dagger = \boldsymbol{U}_1^{-1}$ and $\boldsymbol{U}_2^\dagger = \boldsymbol{U}_2^{-1}$, then $\boldsymbol{U} = \boldsymbol{U}_1\boldsymbol{U}_2$ is also unitary, because $\boldsymbol{U}^{-1} = \boldsymbol{U}_2^{-1}\boldsymbol{U}_1^{-1} = \boldsymbol{U}_2^\dagger\boldsymbol{U}_1^\dagger = (\boldsymbol{U}_1\boldsymbol{U}_2)^\dagger = \boldsymbol{U}^\dagger$), matrix multiplication is associative, the identity element means the $n \times n$ unit matrix, and the inverse matrix $\boldsymbol{U}^{-1} = \boldsymbol{U}^\dagger \equiv (\boldsymbol{U}^T)^*$ always exists. In physics this group is called U(n).

Example 5. SU(n) ***group.*** The group (famous in physics) SU(n) for $n \geqslant 2$ is defined as the subset of U(n) of such matrices $\boldsymbol{U}$ that $\det \boldsymbol{U} = 1$ with the same multiplication operation. Indeed, since $\det(\boldsymbol{U}_1\boldsymbol{U}_2) = \det \boldsymbol{U}_1 \det \boldsymbol{U}_2$, then multiplication of any two elements of the SU(n) gives an element of SU(n). Also of great importance in physics is the SO(n) group, that is the SU(n) group with real (i.e. orthogonal) matrices.[5] SO(3)

Unitary vs symmetry operation

Let us take the SO(3) group of all rotations of the coordinate system in 3D (the *Cartesian 3D Euclidean space*, see Appendix B, p. 895). The rotation operators acting in this space will be denoted by $\hat{R}$ and defined as follows: operator $\hat{R}$ acting on a vector $\boldsymbol{r}$ produces vector $\hat{R}\boldsymbol{r}$:

$$\hat{R}\boldsymbol{r} = \boldsymbol{R}\boldsymbol{r}, \tag{C.1}$$

where[6] $\boldsymbol{R}$ represents an orthogonal matrix of dimension 3. The orthogonality guarantees that the transformation preserves the vector dot (or scalar) products (and therefore their lengths as well).

Let us take an arbitrary function $f(\boldsymbol{r})$ of position $\boldsymbol{r}$. Now, for each of the operators $\hat{R}$ let us construct the corresponding operator $\hat{\mathcal{R}}$ that *moves the function in space without its deformation*. Generally, we obtain another function, which means that $\hat{\mathcal{R}}$ *operates in the Hilbert space*. The construction of operator $\hat{\mathcal{R}}$ is based on the following description

$$\boxed{\hat{\mathcal{R}} f(\boldsymbol{r}) = f\big(\hat{R}^{-1}\boldsymbol{r}\big). \qquad \text{(C.2)}}$$

---

[5] Recall (Appendix A) that for a unitary matrix $\boldsymbol{U}$ we have $\det \boldsymbol{U} = \exp(i\phi)$. For orthogonal matrices (i.e. unitary ones with all the elements real) $\det \boldsymbol{U} = \pm 1$. This does not mean that the SU(n) is composed of the orthogonal matrices only. For example, all four 2×2 matrices:

$$\begin{Bmatrix} 1 & 0 \\ 0 & 1 \end{Bmatrix}, \begin{Bmatrix} -1 & 0 \\ 0 & -1 \end{Bmatrix}, \begin{Bmatrix} 0 & i \\ i & 0 \end{Bmatrix}, \begin{Bmatrix} 0 & -i \\ -i & 0 \end{Bmatrix}$$

have determinants equal to 1 and belong to SU(2), while only the first two belong to SO(2).

[6] The point in 3D space is indicated by vector $\boldsymbol{r} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$.

This means that the displacement in space of function $f(\boldsymbol{r})$ is simply equivalent to leaving the function intact, but instead *inversing the displacement of the coordinate system*.[7]

Operators $\hat{\mathcal{R}}$ rotate functions without their deformation, therefore they preserve the scalar products in the Hilbert space and are unitary. They form a group isomorphic with the group of operators $\hat{R}$, because they have the same multiplication table as operators $\hat{R}$: if $\hat{R} = \hat{R}_1\hat{R}_2$, then $\hat{\mathcal{R}} = \hat{\mathcal{R}}_1\hat{\mathcal{R}}_2$, where $\hat{\mathcal{R}}_1 f(\boldsymbol{r}) = f(\hat{R}_1^{-1}\boldsymbol{r})$ and $\hat{\mathcal{R}}_2 f(\boldsymbol{r}) = f(\hat{R}_2^{-1}\boldsymbol{r})$. Indeed,[8] $\hat{\mathcal{R}} f = (\hat{\mathcal{R}}_1\hat{\mathcal{R}}_2) f(\boldsymbol{r}) = f(\hat{R}_2^{-1}\hat{R}_1^{-1}\boldsymbol{r}) = f(\hat{R}^{-1}\boldsymbol{r})$.

UNITARY VS SYMMETRY OPERATION

A unitary operation is a symmetry operation of function $f(\boldsymbol{r})$, when $\hat{\mathcal{R}} f(\boldsymbol{r}) = f(\boldsymbol{r})$.

Example 6. *Rotation of a point.* Operator $\hat{R}(\alpha; z)$ of the rotation of a point with coordinates x, y, z by angle α about axis z gives a point with coordinates x', y', z' (Fig. C.1.a)

$$x' = r\cos(\phi + \alpha) = r\cos\phi\cos\alpha - r\sin\phi\sin\alpha = x\cos\alpha - y\sin\alpha,$$

$$y' = r\sin(\phi + \alpha) = r\sin\phi\cos\alpha + r\cos\phi\sin\alpha = x\sin\alpha + y\cos\alpha,$$

$$z' = z,$$

the corresponding transformation matrix of the old to the new coordinates, therefore, is

$$\boldsymbol{U} = \begin{bmatrix} \cos\alpha & -\sin\alpha & 0 \\ \sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

We obtain the same new coordinates if the point remains still while the coordinate system rotates in the opposite direction (i.e. by angle $-\alpha$).

Example 7. *Rotation of an atomic orbital.* Let us construct a single spherically symmetric Gaussian orbital $f(\boldsymbol{r}) = \exp[-|\boldsymbol{r} - \boldsymbol{r}_0|^2]$ in Hilbert space for one electron. Let the atomic orbital be centred on the point indicated by vector $\boldsymbol{r}_0$. Operator $\hat{\mathcal{R}}(\alpha; z)$ has to perform the rotation of a *function*[9] by angle α about axis z

---

[7] Motion is relative. Let us concentrate on a rotation by angle α. The result is the same if:

- the coordinate system stays still, but the point rotates by angle α
- or, the point does not move, while the coordinate system rotates by angle $-\alpha$.

What would happen if function $f(\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_N)$ is rotated? Then we will do the following: $\hat{\mathcal{R}} f(\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_N) = f(\hat{R}^{-1}\boldsymbol{r}_1, \hat{R}^{-1}\boldsymbol{r}_2, \ldots, \hat{R}^{-1}\boldsymbol{r}_N)$.

[8] This result is correct, but the routine notation works in a quite misleading way here when suggesting that $(\hat{\mathcal{R}}_1\hat{\mathcal{R}}_2) f(\boldsymbol{r})$ and $f(\hat{R}_1^{-1}\hat{R}_2^{-1}\boldsymbol{r})$ mean the same. However, we derive the correct result in the following way. First, from the definition we have $\hat{\mathcal{R}}_2 f(\boldsymbol{r}) = (R_2^{-1}\boldsymbol{r}) \equiv g_2(\boldsymbol{r})$. Then we get $(\hat{\mathcal{R}}_1\hat{\mathcal{R}}_2) f(\boldsymbol{r}) = \hat{\mathcal{R}}_1[\hat{\mathcal{R}}_2 f(\boldsymbol{r})] = \hat{\mathcal{R}}_1 g_2(\boldsymbol{r}) = g_2(R_1^{-1}\boldsymbol{r}) = \hat{\mathcal{R}}_2 f(R_1^{-1}\boldsymbol{r}) = f(R_2^{-1}R_1^{-1}\boldsymbol{r})$.

[9] This orbital represents our *object* rotating by α. The coordinate system remains unchanged while the object moves. The job will be done by operator $\hat{\mathcal{R}}(\alpha; z)$.

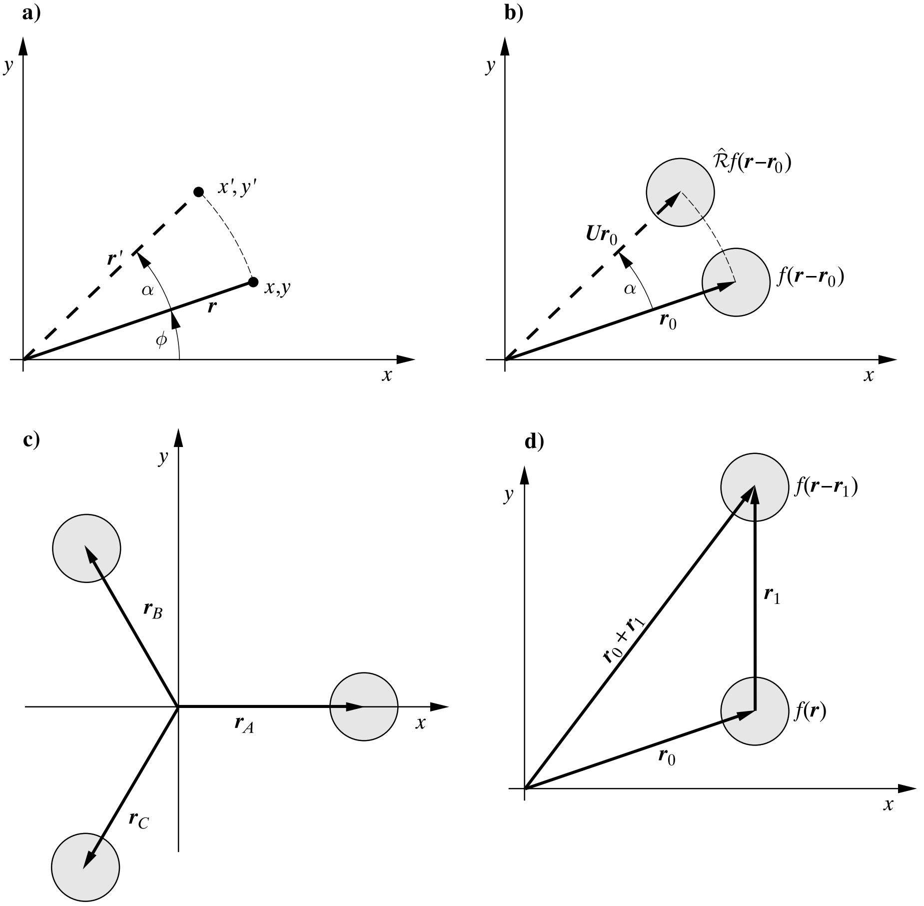


Fig. C.1. Examples of an isometric operation. (a) Unitary operation: rotation of a point by angle α about axis z. The old position of the point is indicated by the vector $\boldsymbol{r}$, the new position by $\boldsymbol{r}'$ (of the same length). (b) Unitary operation: rotation of function $f(\boldsymbol{r}-\boldsymbol{r}_0)$ by angle α about axis z. As a result we have function $f(\boldsymbol{r}-\boldsymbol{U}\boldsymbol{r}_0)$, which in general represents a function which differs from $f(\boldsymbol{r}-\boldsymbol{r}_0)$. (c) The unitary operation which represents a symmetry operation: rotation by angle $\alpha = 120°$ of function $f(\boldsymbol{r}) = \exp[-|\boldsymbol{r}-\boldsymbol{r}_A|^2] + \exp[-|\boldsymbol{r}-\boldsymbol{r}_B|^2] + \exp[-|\boldsymbol{r}-\boldsymbol{r}_C|^2]$, where vectors $\boldsymbol{r}_A, \boldsymbol{r}_B, \boldsymbol{r}_C$ are of the same length and form a "Mercedes trademark" (angle 120°). The new function is identical to the old one. (d) Translational operator by vector $\boldsymbol{r}_1$: $\hat{\mathcal{R}}(\boldsymbol{r}_1)$ applied to the Gaussian function $f(\boldsymbol{r}) = \exp[-|\boldsymbol{r}-\boldsymbol{r}_0|^2]$ gives $\hat{\mathcal{R}}(\boldsymbol{r}_1)f(\boldsymbol{r}) = f(\hat{R}^{-1}\boldsymbol{r}) = \exp[-|\hat{R}^{-1}\boldsymbol{r}-\boldsymbol{r}_0|^2] = \exp[-|\boldsymbol{r}-\boldsymbol{r}_1-\boldsymbol{r}_0|^2] = \exp[-|\boldsymbol{r}-(\boldsymbol{r}_1+\boldsymbol{r}_0)|^2] = f(\boldsymbol{r}-\boldsymbol{r}_1)$, i.e. the function is shifted in space by vector $\boldsymbol{r}_1$ with respect to the original function.

(Fig. C.1.b), which corresponds to a rotation in Hilbert space.[10] According to the definition of a rotation, what we need is $\hat{\mathcal{R}}f(\boldsymbol{r}) = f(\hat{R}^{-1}\boldsymbol{r})$. Since operator $\hat{R}$ corresponds to matrix $\boldsymbol{U}$, then $\hat{R}^{-1}$ corresponds to $\boldsymbol{U}^{-1}$. The last matrix is simply

---

[10] We will obtain another (because differently centred) function.

$$U^{-1} = U^T = \begin{bmatrix} \cos\alpha & \sin\alpha & 0 \\ -\sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

We obtain the following chain of transformations

$$\begin{aligned} f(\hat{R}^{-1}\boldsymbol{r}) &= \exp\left[-\left|\hat{R}^{-1}\boldsymbol{r} - \boldsymbol{r}_0\right|^2\right] = \exp\left[-\left|\hat{R}^{-1}\boldsymbol{r} - \hat{R}^{-1}\hat{R}\boldsymbol{r}_0\right|^2\right] \\ &= \exp\left[-\langle\hat{R}^{-1}\boldsymbol{r} - \hat{R}^{-1}\hat{R}\boldsymbol{r}_0|\hat{R}^{-1}\boldsymbol{r} - \hat{R}^{-1}\hat{R}\boldsymbol{r}_0\rangle\right] \\ &= \exp\left[-\langle\hat{R}\hat{R}^{-1}\boldsymbol{r} - \hat{R}\hat{R}^{-1}\hat{R}\boldsymbol{r}_0|\boldsymbol{r} - \hat{R}\boldsymbol{r}_0\rangle\right] \\ &= \exp\left[-\langle\boldsymbol{r} - \hat{R}\boldsymbol{r}_0|\boldsymbol{r} - \hat{R}\boldsymbol{r}_0\rangle\right] = \exp\left[-\left|\boldsymbol{r} - \hat{R}\boldsymbol{r}_0\right|^2\right]. \end{aligned}$$

Thus, the centre of the orbital undergoes rotation and therefore $\hat{\mathcal{R}}f(\boldsymbol{r})$ indeed represents the spherically symmetric orbital[11] displaced by angle α.

> Since *in general* for any value of angle α function $\exp[-|\boldsymbol{r} - \boldsymbol{U}\boldsymbol{r}_0|^2]$ is not equal to $\exp[-|\boldsymbol{r} - \boldsymbol{r}_0|^2]$, unitary operation $\hat{\mathcal{R}}$ *is not* a symmetry operation on the object.

If, however, $\alpha = 2\pi n$, $n = 0, \pm 1, \pm 2, \ldots$, then $\hat{\mathcal{R}}f(\boldsymbol{r}) = f(\boldsymbol{r})$ and $\hat{\mathcal{R}}(2\pi n; z)$ *is*[12] a symmetry operation.

Example 8. ***Rotation of a particular sum of atomic orbitals.*** Let us take the example of the *sum* of three spherically symmetric Gaussian orbitals:

$$f(\boldsymbol{r}) = \exp\left[-|\boldsymbol{r} - \boldsymbol{r}_A|^2\right] + \exp\left[-|\boldsymbol{r} - \boldsymbol{r}_B|^2\right] + \exp\left[-|\boldsymbol{r} - \boldsymbol{r}_C|^2\right],$$

where vectors $\boldsymbol{r}_A, \boldsymbol{r}_B, \boldsymbol{r}_C$ are of the same length and form the "Mercedes sign" (angles equal to 120°), Fig. C.1.c. Let us take operator $\hat{\mathcal{R}}(\alpha = 120°; z)$ corresponding to matrix $\boldsymbol{U}$. Application of $\hat{\mathcal{R}}$ to function $f(\boldsymbol{r})$ is equivalent to[13]

$$\begin{aligned} f(\hat{R}^{-1}\boldsymbol{r}) &= \exp\left[-\left|\hat{R}^{-1}\boldsymbol{r} - \boldsymbol{r}_A\right|^2\right] + \exp\left[-\left|\hat{R}^{-1}\boldsymbol{r} - \boldsymbol{r}_B\right|^2\right] + \exp\left[-\left|\hat{R}^{-1}\boldsymbol{r} - \boldsymbol{r}_C\right|^2\right] \\ &= \exp\left[-\left|\boldsymbol{r} - \hat{R}\boldsymbol{r}_A\right|^2\right] + \exp\left[-\left|\boldsymbol{r} - \hat{R}\boldsymbol{r}_B\right|^2\right] + \exp\left[-\left|\boldsymbol{r} - \hat{R}\boldsymbol{r}_C\right|^2\right]. \end{aligned}$$

---

[11] The definition $\hat{\mathcal{R}}f(\boldsymbol{r}) = f(\hat{R}^{-1}\boldsymbol{r})$ can transform anything: from the spherically symmetric Gaussian orbital through a molecular orbital (please recall it can be represented by the LCAO expansion) to the Statue of Liberty. Indeed, do you want to rotate the Statue of Liberty? Then leave the Statue in peace, but transform (in the opposite way) your Cartesian coordinate system.

More general transformations, allowing deformation of objects, could also be described by this formula $\hat{\mathcal{R}}f(\boldsymbol{r}) = f(\hat{R}^{-1}\boldsymbol{r})$, but operator $\hat{R}$ would be non-unitary.

[12] The transformed and non-transformed orbitals coincide.

[13] We use the result from the last example.

From the figure (or from the matrix) we have $\hat{R}\boldsymbol{r}_A = \boldsymbol{r}_B$; $\hat{R}\boldsymbol{r}_B = \boldsymbol{r}_C$; $\hat{R}\boldsymbol{r}_C = \boldsymbol{r}_A$. This gives

$$\hat{\mathcal{R}}f(\boldsymbol{r}) = \exp\left[-|\boldsymbol{r}-\boldsymbol{r}_B|^2\right] + \exp\left[-|\boldsymbol{r}-\boldsymbol{r}_C|^2\right] + \exp\left[-|\boldsymbol{r}-\boldsymbol{r}_A|^2\right] = f(\boldsymbol{r}).$$

We have obtained our old function. $\hat{\mathcal{R}}(\alpha = 120°; z)$ *is* therefore the *symmetry* operation[14] $f(\boldsymbol{r})$.

> $\hat{\mathcal{R}}(\alpha = 120°; z)$ represents a symmetry operation not only for function f, but also for other objects which have the *symmetry of an equilateral triangle*.

Example 9. ***Rotation of a many-electron wave function.*** If, in the last example, we had taken a three-electronic function as the *product* of the Gaussian orbitals

$$f(\boldsymbol{r}_1, \boldsymbol{r}_2, \boldsymbol{r}_3) = \exp\left[-|\boldsymbol{r}_1-\boldsymbol{r}_A|^2\right] \cdot \exp\left[-|\boldsymbol{r}_2-\boldsymbol{r}_B|^2\right] \cdot \exp\left[-|\boldsymbol{r}_3-\boldsymbol{r}_C|^2\right],$$

then after applying $\hat{\mathcal{R}}(\alpha = 120°; z)$ to f we would obtain, using an almost identical procedure,

$$\begin{aligned}\hat{\mathcal{R}}f(\boldsymbol{r}_1, \boldsymbol{r}_2, \boldsymbol{r}_3) &= f\left(\hat{R}^{-1}\boldsymbol{r}_1, \hat{R}^{-1}\boldsymbol{r}_2, \hat{R}^{-1}\boldsymbol{r}_3\right) \\ &= \exp\left[-|\boldsymbol{r}_1-\boldsymbol{r}_B|^2\right] \cdot \exp\left[-|\boldsymbol{r}_2-\boldsymbol{r}_C|^2\right] \cdot \exp\left[-|\boldsymbol{r}_3-\boldsymbol{r}_A|^2\right],\end{aligned}$$

which represents a completely different function than the original $f(\boldsymbol{r}_1, \boldsymbol{r}_2, \boldsymbol{r}_3)$! Thus, $\hat{\mathcal{R}}$ does not represent any symmetry operation for $f(\boldsymbol{r}_1, \boldsymbol{r}_2, \boldsymbol{r}_3)$. If, however, we had taken a symmetrized function, e.g., $\tilde{f}(\boldsymbol{r}_1, \boldsymbol{r}_2, \boldsymbol{r}_3) = \sum_P \hat{P} f(\boldsymbol{r}_1, \boldsymbol{r}_2, \boldsymbol{r}_3)$, where $\hat{P}$ permutes the centres A, B, C, and the summation goes over all permutations, we would obtain an $\tilde{f}$ that would turn out to be symmetric with respect to $\hat{\mathcal{R}}(\alpha = 120°; z)$.

Example 10. ***Translation.*** Translation cannot be represented as a matrix transformation (C.1). It is, however, an *isometric operation*, i.e. preserves the distances among the points of the transformed object. This is sufficient for us. Let us enlarge the set of the allowed operations in 3D Euclidean space by isometry. Similarly, as in the case of rotations let us define a shift of the *function* $f(\boldsymbol{r})$. A shift of *function* $f(\boldsymbol{r})$ by vector $\boldsymbol{r}_1$ is such a transformation $\hat{\mathcal{R}}(\boldsymbol{r}_1)$ (in the Hilbert space) that the new function $\tilde{f}(\boldsymbol{r}) = f(\boldsymbol{r}-\boldsymbol{r}_1)$. As an example let us take function $f(\boldsymbol{r}) = \exp[-|\boldsymbol{r}-\boldsymbol{r}_0|^2]$ and let us shift it by vector $\boldsymbol{r}_1$. Translations obey the known relation (C.2):

isometric operation

$$\begin{aligned}\hat{\mathcal{R}}(\boldsymbol{r}_1)f(\boldsymbol{r}) &= f\left(\hat{R}^{-1}\boldsymbol{r}\right) = \exp\left[-\left|\hat{R}^{-1}\boldsymbol{r}-\boldsymbol{r}_0\right|^2\right] \\ &= \exp\left[-|\boldsymbol{r}-\boldsymbol{r}_1-\boldsymbol{r}_0|^2\right] = \exp\left[-\left|\boldsymbol{r}-(\boldsymbol{r}_1+\boldsymbol{r}_0)\right|^2\right] = f(\boldsymbol{r}-\boldsymbol{r}_1).\end{aligned}$$

---

[14] Note that, e.g., if one of the 1*s* orbitals had the opposite sign, function $f(\boldsymbol{r})$ would not have the symmetry of an equilateral triangle, although it would also be invariant with respect to *some* of the operations of an equilateral triangle.

Function $f(\boldsymbol{r})$ had been concentrated around point $\boldsymbol{r}_0$, while the new function $\hat{\mathcal{R}}(\boldsymbol{r}_1)f(\boldsymbol{r})$ is concentrated around the point indicated by vector $\boldsymbol{r}_1 + \boldsymbol{r}_0$, i.e. the function has been shifted by $\boldsymbol{r}_1$ (Fig. C.1.d). This transformation is (similar to case of rotations) unitary, because the scalar product between two functions f_1 and f_2 shifted by the same operation is preserved: $\langle f_1(\boldsymbol{r})|f_2(\boldsymbol{r})\rangle = \langle f_1(\boldsymbol{r}-\boldsymbol{r}_1)|f_2(\boldsymbol{r}-\boldsymbol{r}_1)\rangle$.

Symmetry group of the ammonia molecule

Imagine a model of the NH_3 molecule (rigid trigonal pyramid), Fig. C.2. A student sitting at the table plays with the model. We look at the model, then close our eyes for a second, and open them again. We see that the student smiles, but the coordinate system, the model and its position with respect to the coordinate system look exactly the same as before. Could the student have changed the position of the model? Yes, it is possible. For example, the student could rotate the model about the z axis (perpendicular to the table) by 120°, he might exchange two NH bonds in the model, he may also do nothing. Whatever the student could do is called a *symmetry operation*.

Let us make a list of all the symmetry operations allowed for the ammonia molecule (Table C.1). To this end, let us label the vertices of the triangle a, b, c and locate it in such a way that the centre of the triangle coincides with the origin of the coordinate system, and the y axis indicates vortex a.

Now let us check whether the operations given in Table C.1 form a group. Four conditions have to be satisfied. The first condition requires the existence of "multiplication" in the group, and that the product of any two elements gives an element of the group: $\hat{R}_i\hat{R}_j = \hat{R}_k$. The elements will be the symmetry operations of the equilateral triangle. The product $\hat{R}_i\hat{R}_j = \hat{R}_k$ means that operation $\hat{R}_k$ gives the same result as applying to the triangle operation $\hat{R}_j$ *first*, and *then* the result is

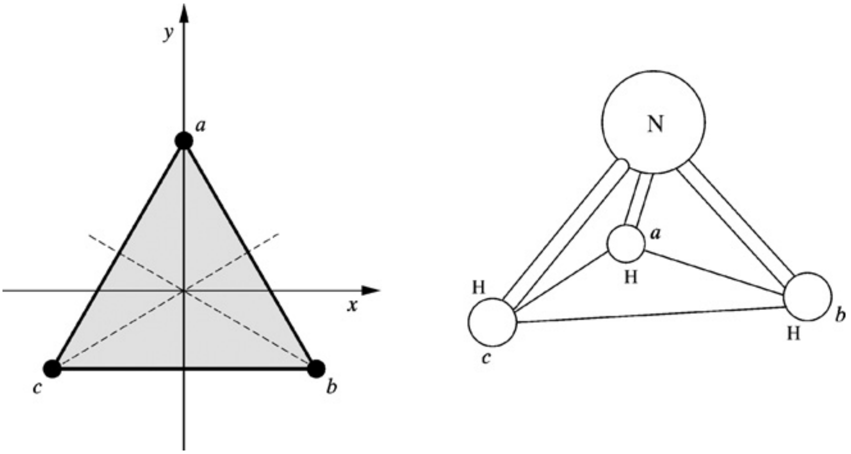


Fig. C.2. The equilateral triangle and the coordinate system. Positions a, b, c are occupied by hydrogen atoms, the nitrogen atom is (symmetrically) above the plane.

Table C.1. Symmetry operations of the ammonia molecule (the reflections pertain to the mirror planes perpendicular to the triangle, Fig. C.2, and go through the centre of the triangle)

Symbol	Description	Symbolic explanation
$\hat{E}$	do nothing	$\hat{E}\begin{bmatrix} a \\ c\ b \end{bmatrix} = \begin{bmatrix} a \\ c\ b \end{bmatrix}$
$\hat{A}$	reflection in the plane going through point *a* shown in Fig. C.2	$\hat{A}\begin{bmatrix} a \\ c\ b \end{bmatrix} = \begin{bmatrix} a \\ b\ c \end{bmatrix}$
$\hat{B}$	reflection in the plane going through point *b* shown in Fig. C.2	$\hat{B}\begin{bmatrix} a \\ c\ b \end{bmatrix} = \begin{bmatrix} c \\ a\ b \end{bmatrix}$
$\hat{C}$	reflection in the plane going through point *c* shown in Fig. C.2	$\hat{C}\begin{bmatrix} a \\ c\ b \end{bmatrix} = \begin{bmatrix} b \\ c\ a \end{bmatrix}$
$\hat{D}$	rotation by 120° (clockwise)	$\hat{D}\begin{bmatrix} a \\ c\ b \end{bmatrix} = \begin{bmatrix} c \\ b\ a \end{bmatrix}$
$\hat{F}$	rotation by −120° (counter-clockwise)	$\hat{F}\begin{bmatrix} a \\ c\ b \end{bmatrix} = \begin{bmatrix} b \\ a\ c \end{bmatrix}$

subject to operation $\hat{R}_i$. In this way the "multiplication table" C.2 can be obtained.

Further, using the table we may check whether the operation is associative. For example, we check whether $\hat{A}(\hat{B}\hat{C}) = (\hat{A}\hat{B})\hat{C}$. The left-hand side gives: $\hat{A}(\hat{B}\hat{C}) = \hat{A}\hat{D} = \hat{B}$. The right-hand side is: $(\hat{A}\hat{B})\hat{C} = \hat{D}\hat{C} = \hat{B}$. It agrees. It will agree for all the other entries in the table.

The unit operation is $\hat{E}$, as seen from the table, because multiplying by $\hat{E}$ does not change anything: $\hat{E}\hat{R}_i = \hat{R}_i\hat{E} = \hat{R}_i$. Also, using the table again, we can find the inverse element to any of the elements. Indeed, $\hat{E}^{-1} = \hat{E}$, because $\hat{E}$ times just $\hat{E}$ equals to $\hat{E}$. Further, $\hat{A}^{-1} = \hat{A}$, because $\hat{A}$ times $\hat{A}$ equals $\hat{E}$, etc., $\hat{B}^{-1} = \hat{B}$, $\hat{C}^{-1} = \hat{C}$, $\hat{D}^{-1} = \hat{F}$, $\hat{F}^{-1} = \hat{D}$.

Thus all the requirements are fulfilled and all these operations form a group of order $g = 6$. Note that in this group the operations do not necessarily commute, e.g., $\hat{C}\hat{D} = \hat{A}$, but $\hat{D}\hat{C} = \hat{B}$ (the group is not Abelian).

Table C.2. Group multiplication table

			second in the product					
		$\hat{R}_j$	$\hat{E}$	$\hat{A}$	$\hat{B}$	$\hat{C}$	$\hat{D}$	$\hat{F}$
first in the product	$\hat{R}_i$							
	$\hat{E}$		$\hat{E}$	$\hat{A}$	$\hat{B}$	$\hat{C}$	$\hat{D}$	$\hat{F}$
	$\hat{A}$		$\hat{A}$	$\hat{E}$	$\hat{D}$	$\hat{F}$	$\hat{B}$	$\hat{C}$
	$\hat{B}$		$\hat{B}$	$\hat{F}$	$\hat{E}$	$\hat{D}$	$\hat{C}$	$\hat{A}$
	$\hat{C}$		$\hat{C}$	$\hat{D}$	$\hat{F}$	$\hat{E}$	$\hat{A}$	$\hat{B}$
	$\hat{D}$		$\hat{D}$	$\hat{C}$	$\hat{A}$	$\hat{B}$	$\hat{F}$	$\hat{E}$
	$\hat{F}$		$\hat{F}$	$\hat{B}$	$\hat{C}$	$\hat{A}$	$\hat{E}$	$\hat{D}$

Classes

The group elements can be all divided into disjoint sets called classes. A class (to put it first in a simplified way) represents a set of operations that are similar, e.g., three reflection operations $\hat{A}$, $\hat{B}$ and $\hat{C}$ constitute one class, the rotations $\hat{D}$ and $\hat{F}$ form the second class, the third class is simply the element $\hat{E}$. Now, the precise definition.

> CLASS
> A class is a set of elements that are conjugate one to another. An element $\hat{R}_i$ is conjugate with $\hat{R}_j$ if we can find in group G such an element (let us denote it by $\hat{X}$) that $\hat{X}^{-1}\hat{R}_j\hat{X} = \hat{R}_i$.

Then, of course, element $\hat{R}_j$ is a conjugate to $\hat{R}_i$ as well. We check this by multiplying $\hat{R}_i$ from the left by $\hat{X} = \hat{Y}^{-1}$, and from the right by $\hat{X}^{-1} = \hat{Y}$ (which yields $\hat{Y}^{-1}\hat{R}_i\hat{Y} = \hat{X}\hat{R}_i\hat{X}^{-1} = \hat{X}\hat{X}^{-1}\hat{R}_j\hat{X}\hat{X}^{-1} = \hat{E}\hat{R}_j\hat{E} = \hat{R}_j$).

Let us have some practice using our table. We have $\hat{X}^{-1}\hat{E}\hat{X} = \hat{X}^{-1}\hat{X}\hat{E} = \hat{E}\hat{E} = \hat{E}$ for each $\hat{X} \in G$, i.e. $\hat{E}$ alone represents a class. Further, making $\hat{X}^{-1}\hat{A}\hat{X}$ for all possible $\hat{X}$ gives:

$$\begin{aligned}
\hat{E}^{-1}\hat{A}\hat{E} &= \hat{E}\hat{A}\hat{E} = \hat{A}\hat{E} = \hat{A},\\
\hat{A}^{-1}\hat{A}\hat{A} &= \hat{A}\hat{A}\hat{A} = \hat{E}\hat{A} = \hat{A},\\
\hat{B}^{-1}\hat{A}\hat{B} &= \hat{B}\hat{A}\hat{B} = \hat{F}\hat{B} = \hat{C},\\
\hat{C}^{-1}\hat{A}\hat{C} &= \hat{C}\hat{A}\hat{C} = \hat{D}\hat{C} = \hat{B},\\
\hat{D}^{-1}\hat{A}\hat{D} &= \hat{F}\hat{A}\hat{D} = \hat{B}\hat{D} = \hat{C},\\
\hat{F}^{-1}\hat{A}\hat{F} &= \hat{D}\hat{A}\hat{F} = \hat{B}\hat{F} = \hat{C}.
\end{aligned}$$

This means that $\hat{A}$ belongs to the same class as $\hat{B}$ and $\hat{C}$. Now we will create some conjugate elements to $\hat{D}$ and $\hat{F}$:

$$\begin{aligned}
\hat{A}^{-1}\hat{D}\hat{A} &= \hat{A}\hat{D}\hat{A} = \hat{B}\hat{A} = \hat{F},\\
\hat{B}^{-1}\hat{D}\hat{B} &= \hat{B}\hat{D}\hat{B} = \hat{C}\hat{B} = \hat{F},\\
\hat{C}^{-1}\hat{D}\hat{C} &= \hat{C}\hat{D}\hat{C} = \hat{A}\hat{C} = \hat{F},
\end{aligned}$$

etc. Thus $\hat{D}$ and $\hat{F}$ make a class. Therefore the group under consideration consists of the following classes: $\{\hat{E}\}\{\hat{A}, \hat{B}, \hat{C}\}\{\hat{D}, \hat{F}\}$.

> It is always like this: the *group is the sum of the disjoint classes*.

2 REPRESENTATIONS

A representation of the group is such a g-element sequence of square matrices (of the same dimension; a matrix is associated with each element of the group), that the matrices have a multiplication table consistent with the multiplication table of the group.

By consistency we mean the following. To each element of the group we assign a square matrix (of the same dimension for all elements). If the multiplication table for the group says that $\hat{R}_i\hat{R}_j = \hat{R}_k$, the matrix corresponding to $\hat{R}_i$ times the matrix corresponding to $\hat{R}_j$ is the matrix corresponding to $\hat{R}_k$. If this agrees for all $\hat{R}$, then we say that the matrices form a representation.[15]

We may create many group representations, see Table C.3.

The easiest thing is to see that Γ_1 satisfies the criterion of being a representation (the matrices have dimension 1, i.e. they are numbers). After looking for a while at Γ_2 we will say the same. Multiplying the corresponding matrices we will prove this for Γ_3 and Γ_4. For example, for Γ_3 the product of the matrices $\hat{B}$ and $\hat{C}$ gives the matrix corresponding to operation $\hat{D}$

$$\begin{bmatrix} \frac{1}{2} & -\frac{\sqrt{3}}{2} \\ -\frac{\sqrt{3}}{2} & -\frac{1}{2} \end{bmatrix}\begin{bmatrix} \frac{1}{2} & \frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & -\frac{1}{2} \end{bmatrix} = \begin{bmatrix} -\frac{1}{2} & \frac{\sqrt{3}}{2} \\ -\frac{\sqrt{3}}{2} & -\frac{1}{2} \end{bmatrix}.$$

If we had more patience, we could show this equally easily for the whole multiplication table of the group. Note that

there are many representations of a group.

Note also an interesting thing. Let us take a point with coordinates $(x, y, 0)$ and see what will happen to it when symmetry operations are applied (the coordinate system rests, while the point itself moves). The identity operation $\hat{E}$ leads to the following transformation matrix

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}$$

The results of the other operations are characterized by the following transformation matrices (you may check this step by step):

$$\hat{A}: \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \quad \hat{B}: \begin{bmatrix} \frac{1}{2} & -\frac{\sqrt{3}}{2} \\ -\frac{\sqrt{3}}{2} & -\frac{1}{2} \end{bmatrix} \quad \hat{C}: \begin{bmatrix} \frac{1}{2} & \frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & -\frac{1}{2} \end{bmatrix}$$

$$\hat{D}: \begin{bmatrix} -\frac{1}{2} & \frac{\sqrt{3}}{2} \\ -\frac{\sqrt{3}}{2} & -\frac{1}{2} \end{bmatrix} \quad \hat{F}: \begin{bmatrix} -\frac{1}{2} & -\frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & -\frac{1}{2} \end{bmatrix}.$$

---

[15] More formally: a representation is a *homomorphism of the group into the above set of matrices*.

Table C.3. Several representations of the equilateral triangle symmetry group

	Group elements					
Repr.	$\hat{E}$	$\hat{A}$	$\hat{B}$	$\hat{C}$	$\hat{D}$	$\hat{F}$
Γ_1	1	1	1	1	1	1
Γ_2	1	–1	–1	–1	1	1
Γ_3	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} \frac{1}{2} & -\frac{\sqrt{3}}{2} \\ -\frac{\sqrt{3}}{2} & -\frac{1}{2} \end{bmatrix}$	$\begin{bmatrix} \frac{1}{2} & \frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & -\frac{1}{2} \end{bmatrix}$	$\begin{bmatrix} -\frac{1}{2} & \frac{\sqrt{3}}{2} \\ -\frac{\sqrt{3}}{2} & -\frac{1}{2} \end{bmatrix}$	$\begin{bmatrix} -\frac{1}{2} & -\frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & -\frac{1}{2} \end{bmatrix}$
Γ_4	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} -1 & 0 & 0 \\ 0 & \frac{1}{2} & -\frac{\sqrt{3}}{2} \\ 0 & -\frac{\sqrt{3}}{2} & -\frac{1}{2} \end{bmatrix}$	$\begin{bmatrix} -1 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{\sqrt{3}}{2} \\ 0 & \frac{\sqrt{3}}{2} & -\frac{1}{2} \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & -\frac{1}{2} & \frac{\sqrt{3}}{2} \\ 0 & -\frac{\sqrt{3}}{2} & -\frac{1}{2} \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & -\frac{1}{2} & -\frac{\sqrt{3}}{2} \\ 0 & \frac{\sqrt{3}}{2} & -\frac{1}{2} \end{bmatrix}$

Note that the matrices obtained are identical to those for the representation Γ_3. Thus by transforming the coordinates of a point, we have generated a representation of the symmetry.

By transforming "anything" (coordinates of a point, vectors, functions) using the symmetry operations and collecting the results in the form of matrices, we always obtain a representation of the group.

Characters of representation

For any representation Γ, we may define vector $\boldsymbol{\chi}^{(\Gamma)}$ of dimension g, having as elements the traces of the representation matrices $\Gamma(\hat{R}_i)$

$$\mathrm{Tr}\,\Gamma = \sum_i \Gamma_{ii}, \tag{C.3}$$

$$\boldsymbol{\chi}^{(\Gamma)} \equiv \begin{bmatrix} \mathrm{Tr}\,\Gamma(\hat{R}_1) \\ \mathrm{Tr}\,\Gamma(\hat{R}_2) \\ \dots \\ \mathrm{Tr}\,\Gamma(\hat{R}_g) \end{bmatrix} \equiv \begin{bmatrix} \chi^{(\Gamma)}(\hat{R}_1) \\ \chi^{(\Gamma)}(\hat{R}_2) \\ \dots \\ \chi^{(\Gamma)}(\hat{R}_g) \end{bmatrix}. \tag{C.4}$$

The number $\chi^{(\Gamma)}(\hat{R}_i)$ is called the *character of representation* Γ that *corresponds to operation* $\hat{R}_i$. The characters of representations will play a most important role in the application of group theory to spectroscopy.

Irreducible representations

To explain what irreducible representation is, let us first define *reducible representations*. reducible representation

A representation is called *reducible* if its matrices can be transformed into what is called *block form* by using the transformation $\boldsymbol{P}^{-1}\Gamma(\hat{R}_i)\boldsymbol{P}$ for every matrix $\Gamma(\hat{R}_i)$, where $\boldsymbol{P}$ is a non-singular matrix.

In block form the non-zero elements can only be in the square blocks located on the diagonal, Fig. C.3. block form

If, using the same $\boldsymbol{P}$, we can transform each of the matrices $\Gamma(\hat{R}_i)$ and obtain the same block form, the representation is called *reducible.*

If we do not find such a matrix (because it does not exist), the representation is irreducible. If we carry out the transformation $\boldsymbol{P}^{-1}\Gamma(\hat{R}_i)\boldsymbol{P}$ (*similarity transformation*) for $i = 1, 2, \dots, g$ of a representation, the new matrices also form a representation Γ' called *equivalent* to Γ.

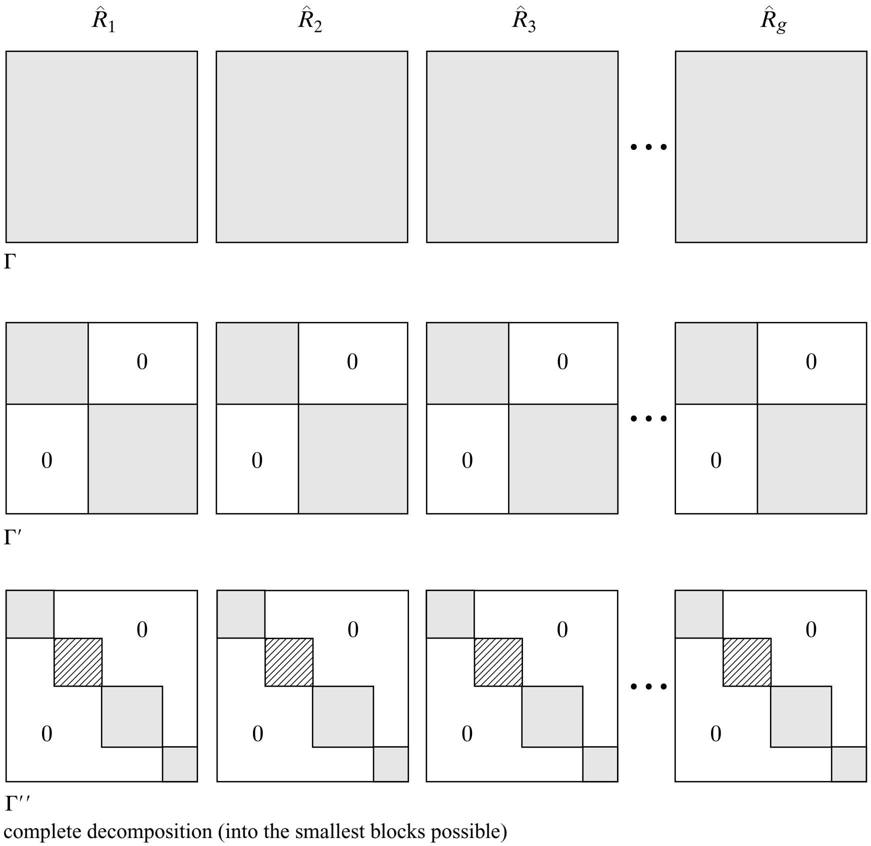


Fig. C.3. Reducible representation, block form and irreducible representation. In the first row the matrices $\Gamma(\hat{R}_i)$ are displayed which form a *reducible representation* (each matrix corresponds to the symmetry operation $\hat{R}_i$); the matrix elements are in general non-zero. The central row shows a representation Γ' equivalent to the first, i.e. related by a similarity transformation (through matrix $\boldsymbol{P}$); the new representation exhibits block form, i.e. in this particular case each matrix has two blocks of zeros, identical in all matrices. The last row shows an equivalent representation Γ'' that corresponds to the smallest square blocks (of non-zeros), i.e. the maximum number of blocks of identical form in all matrices. Not only Γ, Γ' and Γ'' are representations of the group, but also any sequence of individual blocks (as this shaded) is a representation. Thus, Γ'' is decomposed into the four irreducible representations.

This is easy to show. Indeed, group operations $\hat{R}_i$ and $\hat{R}_j$ correspond to matrices $\Gamma(\hat{R}_i)$ and $\Gamma(\hat{R}_j)$ in the original representation and to $\Gamma'(\hat{R}_i) = \boldsymbol{P}^{-1}\Gamma(\hat{R}_i)\boldsymbol{P}$ and $\Gamma'(\hat{R}_j) = \boldsymbol{P}^{-1}\Gamma(\hat{R}_j)\boldsymbol{P}$ in the equivalent representation (we will check in a moment whether this is indeed a representation). The product $\Gamma'(\hat{R}_i)\Gamma'(\hat{R}_j)$ equals to $\boldsymbol{P}^{-1}\Gamma(\hat{R}_i)\boldsymbol{P}\boldsymbol{P}^{-1}\Gamma(\hat{R}_j)\boldsymbol{P} = \boldsymbol{P}^{-1}\Gamma(\hat{R}_i)\Gamma(\hat{R}_j)\boldsymbol{P}$, i.e. to the matrix $\Gamma(\hat{R}_i)\Gamma(\hat{R}_j)$ transformed by *similarity transformation*, therefore everything goes with the same multiplication table. Thus matrices $\Gamma'(\hat{R}_i)$ also form a representation (Γ'). This means that we can create as many representations as we wish, it is sufficient to change

similarity transformation

matrix $\boldsymbol{P}$, and this is easy (since what we want is its singularity, i.e. the $\boldsymbol{P}^{-1}$ matrix has to exist).

The blocks are square matrices. It turns out that the set of the first blocks $\Gamma_1(\hat{R}_1), \Gamma_1(\hat{R}_2), \ldots, \Gamma_1(\hat{R}_g)$ (each block for one operation) is a representation, the set of the second blocks $\Gamma_2(\hat{R}_1), \Gamma_2(\hat{R}_2), \ldots, \Gamma_2(\hat{R}_g)$ forms a representation as well, etc. This is evident. It is sufficient to see what happens when we multiply two matrices in the same block form. The matrix product has the same block form and a particular block results from multiplication of the corresponding blocks of the matrices which are being multiplied.[16]

> In particular, maximum decomposition into blocks leads, of course, to blocks that are no longer decomposable, and therefore are irreducible representations.

irreducible representation

Properties of irreducible representations

For two irreducible representations α and β, the following *group orthogonality theorem* is satisfied:[17]

$$\sum_i [\Gamma^{(\alpha)}(\hat{R}_i)]_{mn}[\Gamma^{(\beta)}(\hat{R}_i)]^*_{m'n'} = \frac{g}{n_\alpha}\delta_{\alpha\beta}\delta_{mm'}\delta_{nn'}, \qquad \text{(C.5)}$$

where $\Gamma^{(\alpha)}(\hat{R})$ and $\Gamma^{(\beta)}(\hat{R})$ denote matrices that correspond to the group element $\hat{R}$ (m, n and m', n' determine the elements of the matrices), the summation goes over all the group elements, and n_α is the dimension of the irreducible representation α, i.e. the dimension of the matrices which form the representation. The symbol $*$ means complex conjugation.[18] We create two g-dimensional vec-

---

[16] Let us explain this by taking an example. We have two square matrices of dimension 4: $\boldsymbol{A}$ and $\boldsymbol{B}$, both having the block form:

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{A}_2 \end{bmatrix}, \quad \boldsymbol{B} = \begin{bmatrix} \boldsymbol{B}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{B}_2 \end{bmatrix} \quad \text{with}$$

$$\boldsymbol{A}_1 = \begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix}, \quad \boldsymbol{A}_2 = \begin{bmatrix} 2 & 2 \\ 2 & 3 \end{bmatrix}, \quad \boldsymbol{B}_1 = \begin{bmatrix} 1 & 3 \\ 3 & 2 \end{bmatrix}, \quad \boldsymbol{B}_2 = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}.$$

Let us check that $\boldsymbol{C} = \boldsymbol{AB}$ has *the same* block form

$$\boldsymbol{C} = \begin{bmatrix} \boldsymbol{C}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{C}_2 \end{bmatrix}$$

and that (which is particularly important for us) $\boldsymbol{C}_1 = \boldsymbol{A}_1\boldsymbol{B}_1$ and $\boldsymbol{C}_2 = \boldsymbol{A}_2\boldsymbol{B}_2$. Indeed, multiplying $\boldsymbol{AB}$ we have

$$\boldsymbol{C} = \begin{bmatrix} 6 & 11 & 0 & 0 \\ 7 & 7 & 0 & 0 \\ 0 & 0 & 6 & 6 \\ 0 & 0 & 7 & 8 \end{bmatrix}, \quad \text{i.e.} \begin{bmatrix} 6 & 11 \\ 7 & 7 \end{bmatrix} = \boldsymbol{C}_1, \quad \begin{bmatrix} 6 & 6 \\ 7 & 8 \end{bmatrix} = \boldsymbol{C}_2.$$

Hence, indeed $\boldsymbol{C}_1 = \boldsymbol{A}_1\boldsymbol{B}_1$ and $\boldsymbol{C}_2 = \boldsymbol{A}_2\boldsymbol{B}_2$.

[17] For the proof see H. Eyring, J. Walter, G.E. Kimball, "*Quantum Chemistry*", New York, Wiley (1944).

[18] It is important only for complex representations Γ.

tors: one composed of components $[\Gamma^{(\alpha)}(\hat{R}_i)]_{mn}$, the other from $[\Gamma^{(\beta)}(\hat{R}_i)]^*_{m'n'}$, $i = 1, 2, \dots, g$. Group orthogonality theorem says that

- if $\alpha \neq \beta$, the vectors are orthogonal,
- if $m \neq m'$ *or* $n \neq n'$, again the two vectors are orthogonal. The formula kills everything, except the two identical irreducible representations and we choose *the same* elements as the vector components.

Characters of irreducible representations

The most important consequence of the group orthogonality theorem is the equation:

$$\sum_i \chi^{(\alpha)}(\hat{R}_i)\chi^{(\beta)}(\hat{R}_i)^* = g\delta_{\alpha\beta}, \tag{C.6}$$

where $\chi^{(\alpha)}(\hat{R}_i)$ is a character of the *irreducible representation* α corresponding to symmetry operation $\hat{R}_i$. Eq. (C.6), in view of eq. (C.4), may be rewritten as a scalar product in a unitary space (Appendix B)

$$\langle \boldsymbol{\chi}^{(\beta)} | \boldsymbol{\chi}^{(\alpha)} \rangle = g\delta_{\alpha\beta}. \tag{C.7}$$

Eq. (C.7) can be obtained from the group orthogonality theorem after setting $m = n$ and $m' = n'$, and summing over m and m':

$$\begin{aligned}\langle \boldsymbol{\chi}^{(\beta)} | \boldsymbol{\chi}^{(\alpha)} \rangle &= \sum_i \sum_m \sum_{m'} [\Gamma^{(\alpha)}(\hat{R}_i)]_{mm} [\Gamma^{(\beta)}(\hat{R}_i)]^*_{m'm'} \\ &= \frac{g}{n_\alpha}\delta_{\alpha\beta} \sum_m \sum_{m'} (\delta_{mm'})^2 = \frac{g}{n_\alpha}\delta_{\alpha\beta} n_\alpha = g\delta_{\alpha\beta}.\end{aligned}$$

Decomposing reducible representations into irreducible ones

It is important that

equivalent representations have identical characters,

because the trace of a matrix is invariant with respect to any similarity transformation. Indeed, for two equivalent representations Γ and Γ', for any $\hat{R}_i$ we have $\Gamma'(\hat{R}_i) = \boldsymbol{P}^{-1}\Gamma(\hat{R}_i)\boldsymbol{P}$, which gives

$$\begin{aligned}\chi^{(\Gamma')}(\hat{R}_i) &= \sum_m \big(\boldsymbol{P}^{-1}\Gamma(\hat{R}_i)\boldsymbol{P}\big)_{mm} = \sum_{mkl} P^{-1}_{mk}\Gamma_{kl}P_{lm} = \sum_{kl}\Gamma_{kl}\sum_m P_{lm}P^{-1}_{mk} \\ &= \sum_{kl}\Gamma_{kl}\big(\boldsymbol{P}\boldsymbol{P}^{-1}\big)_{lk} = \sum_{kl}\Gamma_{kl}\delta_{lk} = \sum_k \Gamma_{kk} = \chi^{(\Gamma)}(\hat{R}_i).\end{aligned}$$

In particular, the character of a representation is the same as its block form (with the maximum number of blocks, which correspond to irreducible representations):

$$\chi(\hat{R}_i) = \sum_{\alpha} \mathrm{a}(\alpha)\chi^{(\alpha)}(\hat{R}_i), \tag{C.8}$$

or, in other words,

$$\boldsymbol{\chi} = \sum_{\alpha} \mathrm{a}(\alpha)\boldsymbol{\chi}^{(\alpha)}, \tag{C.9}$$

where $\mathrm{a}(\alpha)$ *is a natural number telling us how many times the irreducible representation* α *appears in block form*. The above formula comes from the very definition of the trace (the sum of the diagonal elements).

We will need another property of the characters. Namely,

> the characters corresponding to the elements of a class are equal.

Indeed, two elements of group $\hat{R}_i$ and $\hat{R}_j$ which belong to the same class are related to one another by relation $\hat{R}_i = X^{-1}\hat{R}_j X$, where X is an element of the group. The same multiplication table is valid for the representations (from the definition of the representation), thus

$$\Gamma(\hat{R}_i) = \Gamma(X^{-1})\Gamma(\hat{R}_j)\Gamma(X) = [\Gamma(X)]^{-1}\Gamma(\hat{R}_j)\Gamma(X). \tag{C.10}$$

This concludes the proof, because here the matrices $\Gamma(\hat{R}_i)$ and $\Gamma(\hat{R}_j)$ are related by a similarity transformation, and therefore have identical characters. From now on we can write $\chi(C)$ instead of $\chi(\hat{R})$, where C denotes the class to which operation $\hat{R}_i$ belongs.

Eq. (C.8) can be now modified appropriately. It can be rewritten as

$$\begin{aligned}\langle \boldsymbol{\chi}^{(\beta)} | \boldsymbol{\chi}^{(\alpha)} \rangle &= \sum_{C} n_C \chi^{\alpha}(C)\chi^{\beta}(C)^* = \sum_{C} [\sqrt{n_C}\chi^{(\alpha)}(C)][\sqrt{n_C}\chi^{(\beta)}(C)^*] \\ &= g\delta_{\alpha\beta},\end{aligned} \tag{C.11}$$

where C stands for the class, and n_C tells us how many operations belong to the class. This notation reminds us that the numbers $[\sqrt{n_C}\chi^{(\alpha)}(C)]$ for a fixed α and changing class C may be treated as the components of a vector (its dimension is equal to the number of classes) and that the vectors which correspond to different irreducible representations are orthogonal. The dimension of the vectors is equal to the number of classes, say, k. Since the number of orthogonal vectors, each of dimension k, cannot exceed k, then the number of the different irreducible representations is equal the number of classes.

In future applications it will be of key importance to find such a natural number $\mathrm{a}(\alpha)$ which tells us how many times the irreducible representation α is encountered in a reducible representation. The formula for $\mathrm{a}(\alpha)$ is the following

$$\mathrm{a}(\alpha) = \frac{1}{g}\sum_C n_C \chi(C)\chi^{(\alpha)}(C)^*. \tag{C.12}$$

The proof is simple. From the scalar product of both sides of eq. (C.9) with the vector $\boldsymbol{\chi}^{(\beta)}$ after using eq. (C.7) we obtain

$$\langle\boldsymbol{\chi}^{(\beta)}|\boldsymbol{\chi}\rangle = \sum_\alpha \mathrm{a}(\alpha)\langle\boldsymbol{\chi}^{(\beta)}|\boldsymbol{\chi}^{(\alpha)}\rangle = \sum_\alpha \mathrm{a}(\alpha) g\delta_{\alpha\beta} = \mathrm{a}(\beta)g \text{ or } \mathrm{a}(\alpha) = \frac{1}{g}\langle\boldsymbol{\chi}^{(\alpha)}|\boldsymbol{\chi}\rangle.$$

This is the formula sought, because the characters are the same for all operations of the same class.

Note that

to find $\mathrm{a}(\alpha)$ it is sufficient to know the *characters* of the representations, the representations themselves are not necessary.

Tables of characters of irreducible representations

Any textbook on the application of group theory in molecular spectroscopy contains tables of characters of irreducible representations which correspond to various symmetry groups of molecules.[19]

To apply group theory to a particular molecule, we first have to find the table of characters mentioned above. To this end:

- the Born–Oppenheimer approximation is used, therefore the positions of the nuclei are fixed in space (“geometry”),
- from the geometry, we make a list of all the symmetry operations which transform it into itself,
- we identify the corresponding symmetry group.[20]

To find the proper table, we may use the Schoenflies notation for the symmetry[21] (there are also some other notations):

$\hat{E}$ the symbol of the identity operation (i.e. do nothing);

---

[19] The tables are constructed by considering possible symmetries (symmetry groups), creating suitable matrix representations, using similarity transformations to find the irreducible representations, by summing the diagonal elements, we obtain the required character tables.

[20] This may be done by using a flow chart, e.g., given by P.W. Atkins, “*Physical Chemistry*”, sixth edition, Oxford University Press, Oxford, 1998.

[21] Artur Moritz Schoenflies (1853–1928), German mathematician, professor at the universities in Göttingen, Königsberg, Frankfurt am Main. Schoenflies proved (independently of J.S. Fiodorow and W. Barlow) the existence of the complete set of 230 space groups of crystals.

$\hat{C}_n$ rotation by angle $\frac{2\pi}{n}$ about the n-fold symmetry axis;
$\hat{C}_n^m$ rotation by $\frac{2\pi m}{n}$ about the n-fold symmetry axis;
$\hat{\sigma}_v$ reflection in the plane through the axis of the highest symmetry;
$\hat{\sigma}_h$ reflection in the plane perpendicular to the axis of the highest symmetry;
$\hat{\imath}$ inversion with respect to the centre of symmetry;
$\hat{S}_n$ rotation by angle $\frac{2\pi}{n}$ about the n-fold symmetry axis with subsequent reflection in the plane perpendicular to it;
$\hat{S}_n^m$ rotation by angle $\frac{2\pi m}{n}$ about the n-fold symmetry axis with subsequent reflection in the plane perpendicular to it.

The set of symmetry operations obtained forms the symmetry group. The symmetry groups also have their special symbols. The Schoenflies notation for the symmetry groups of some simple molecules is given in Table C.4 (their geometry corresponding to the energy minimum).

A molecule may be much more complicated, but often its symmetry is identical to that of a simple molecule (e.g., one of those reported in the table).

When we finally identify the table of characters suitable for the molecule under consideration, it is time to look at it carefully. For example, for the ammonia molecule we find the table of characters (Table C.5).

In the upper left corner the name of the group is displayed (C_{3v}). The symmetry operations are listed in the same row (in this case $\hat{E}$, $\hat{\sigma}_v$, $\hat{C}_3$).[22] The operations are collected in classes, and the number of such operations in the class is given: the identity operation ($\hat{E}$) forms the first class, the three reflection operations (hence $3\hat{\sigma}_v$, previously called $\hat{A}, \hat{B}, \hat{C}$) corresponding to the planes which contain the threefold symmetry axis, *two* rotation operations (hence, $2\hat{C}_3$, previ-

Table C.4. Examples of the symmetry group (for a few molecules in their ground-state optimum geometry)

Molecule	Group
H_2O	C_{2v}
NH_3	C_{3v}
CH_4	T_d
benzene	D_{6h}
naphthalene	D_{2h}

Table C.5. C_{3v} group. Table of characters

C_{3v}	$\hat{E}$	$3\hat{\sigma}_v$	$2\hat{C}_3$		
A_1	1	1	1	z	$x^2+y^2,\ z^2$
A_2	1	-1	1	R_z	
E	2	0	-1	$(x,y)(R_x,R_y)$	$(x^2-y^2,xy)(xz,yz)$

---

[22] The same symmetry operations as discussed on p. 911.

ously called $\hat{D}$ and $\hat{F}$) about the threefold symmetry axis (by 120° and by 240°, or $-120°$, and the rotation by 360° is identical to $\hat{E}$).

We have information about the irreducible representations in the second and subsequent rows, one row for each representation. The number of irreducible representations is equal to the number of classes (three in our case), i.e. the table of characters is square. On the left-hand side we have the symbol of the representation telling us about its dimension (if the symbol is A, the dimension is 1, if it is E the dimension is 2, if T then 3). Thus, unfortunately, the letter E plays a double role in the table:[23] as the identity operation $\hat{E}$ and as E – the symbol of an irreducible representation. In a given row (irreducible representation), the number below the symbol for class is the corresponding character. For the identity operation $\hat{E}$, the corresponding matrices are unit matrices, and the calculated character is therefore equal to the dimension of the irreducible representation.

The simplest representation possible is of great importance, all the characters equal 1 (in our case A_1). This will be called the fully symmetric representation.

Example 11. ***Decomposition of a reducible representation.*** Let us find how the reducible representation Γ_4 from p. 914 may be decomposed into irreducible representations. First we see from eq. (C.12) that characters rather than the representations themselves are required. The characters $\chi^{(\Gamma_4)}$ are calculated by summing up the diagonals of the matrix representations for the corresponding classes, $\chi^{(\Gamma_4)}$: 3 (class $\hat{E}$), -1 (class $\hat{\sigma}_v$), 0 (class $\hat{C}_3$). Let us first ask how many times (a_{A_1}) the irreducible representation A_1 is encountered in Γ_4. The characters of A_1 (Table C.5) are: $1, 1, 1$ for the corresponding classes. The number of the operations in the classes is respectively n_C: $1, 3, 2$. From (C.12) we find $a(A_1) = \frac{1}{6}(1\cdot 3\cdot 1 + 3\cdot(-1)\cdot 1 + 2\cdot 0\cdot 1) = 0$. Similarly we find $a(A_2) = \frac{1}{6}(1\cdot 3\cdot 1 + 3\cdot(-1)\cdot(-1) + 2\cdot 0\cdot 1) = 1$ and $a(E) = \frac{1}{6}(1\cdot 3\cdot 2 + 3\cdot(-1)\cdot 0 + 2\cdot 0\cdot(-1)) = 1$. Thus, we may write that $\Gamma_4 = A_2 + E$. This exercise will be of great help when the selection rules in spectroscopy are considered.

Projection operator on an irreducible representation

We will soon need information about whether a particular function exhibits certain symmetry properties in the system under consideration. We will need certain projection operators to do this.

$$\hat{P}^{(\alpha)} = \frac{n_\alpha}{g}\sum_i \chi^{(\alpha)*}(\hat{R}_i)\hat{R}_i \tag{C.13}$$

represents the projection operator which projects onto the space of such functions which transform according to the irreducible representation $\Gamma^{(\alpha)}$.

---

[23] This unfortunate traditional notation will not lead to trouble.

This means that either $\hat{P}^{(\alpha)}f$ *transforms according to the irreducible representation* $\Gamma^{(\alpha)}$ or we obtain zero. To be a projection operator, $\hat{P}^{(\alpha)}$ has to satisfy[24]

$$\hat{P}^{(\alpha)}\hat{P}^{(\beta)} = \delta_{\alpha\beta}\hat{P}^{(\alpha)}. \tag{C.14}$$

We can also prove that

$$\boxed{\sum_{\alpha}\hat{P}^{(\alpha)} = 1, \qquad \text{(C.15)}}$$

where the summation goes over all irreducible representations of the group.

---

[24] This means that two functions which transform according to *different* irreducible representations are orthogonal, and that the projection of an already projected function changes nothing. Here is the proof. After noting that $\hat{R}\hat{S} = \hat{Q}$, or $\hat{S} = \hat{R}^{-1}\hat{Q}$ we have

$$\begin{aligned}
\hat{P}^{(\alpha)}\hat{P}^{(\beta)} &= \frac{n_\alpha n_\beta}{g^2}\sum_{\hat{R},S}\chi^{(\alpha)*}(\hat{R})\chi^{(\beta)*}(S)\hat{R}\hat{S} \\
&= \frac{n_\alpha n_\beta}{g^2}\sum_{Q}\hat{Q}\sum_{\hat{R}}\chi^{(\alpha)*}(\hat{R})\chi^{(\beta)*}(\hat{R}^{-1}\hat{Q}).
\end{aligned}$$

Note, that

$$\chi^{(\beta)*}(\hat{R}^{-1}\hat{Q}) = \sum_{k}\Gamma^{(\beta)*}_{kk}(\hat{R}^{-1}\hat{Q}) = \sum_{k}\sum_{l}\Gamma^{(\beta)*}_{kl}(\hat{R}^{-1})\Gamma^{(\beta)*}_{lk}(\hat{Q}).$$

After inserting this result we have

$$\begin{aligned}
\hat{P}^{(\alpha)}\hat{P}^{(\beta)} &= \frac{n_\alpha n_\beta}{g^2}\sum_{Q}\hat{Q}\sum_{\hat{R}}\sum_{m}\Gamma^{(\alpha)*}_{mm}(\hat{R})\sum_{k}\sum_{l}\Gamma^{(\beta)*}_{kl}(\hat{R}^{-1})\Gamma^{(\beta)*}_{lk}(\hat{Q}) \\
&= \frac{n_\alpha n_\beta}{g^2}\sum_{Q}\hat{Q}\sum_{\hat{R}}\sum_{k,l,m}\Gamma^{(\alpha)*}_{mm}(\hat{R})\Gamma^{(\beta)}_{lk}(\hat{R})\Gamma^{(\beta)*}_{lk}(\hat{Q}) \\
&= \frac{n_\alpha n_\beta}{g^2}\sum_{Q}\hat{Q}\sum_{k,l,m}\Gamma^{(\beta)*}_{lk}(\hat{Q})\sum_{\hat{R}}\left[\Gamma^{(\alpha)*}_{mm}(\hat{R})\Gamma^{(\beta)}_{lk}(\hat{R})\right],
\end{aligned}$$

because from the unitary character of the representation matrices $\Gamma^{(\beta)}(\hat{R}^{-1})$ and $\Gamma^{(\beta)}(\hat{R})$ we have $\Gamma^{(\beta)*}_{kl}(\hat{R}^{-1}) = \Gamma^{(\beta)}_{lk}(\hat{R})$. From the group theorem of orthogonality (eq. (C.5)) we have

$$\begin{aligned}
\hat{P}^{(\alpha)}\hat{P}^{(\beta)} &= \frac{n_\alpha n_\beta}{g^2}\frac{g}{n_\alpha}\sum_{Q}\hat{Q}\sum_{k,l,m}\Gamma^{(\beta)*}_{lk}(\hat{Q})\delta_{ml}\delta_{mk}\delta_{\alpha\beta} \\
&= \delta_{\alpha\beta}\frac{n_\alpha}{g}\sum_{Q}\hat{Q}\sum_{m}\Gamma^{(\alpha)*}_{mm}(\hat{Q}) \\
&= \delta_{\alpha\beta}\frac{n_\alpha}{g}\sum_{Q}\chi^{(\alpha)*}(Q)\hat{Q} = \delta_{\alpha\beta}\hat{P}^{(\alpha)},
\end{aligned}$$

as we wished to show, eq. (C.14).

The transformation of a function according to irreducible representation

The right-side of a character table (like Table C.5) contains the symbols x, y, z, $(x^2 - y^2, xy)$, R_x, R_y, R_z. These symbols will be needed to establish the selection rules in spectroscopy (UV-VIS, IR, Raman). They pertain to the coordinate system (the z axis coincides with the axis of highest symmetry). Let us leave the symbols R_x, R_y, R_z for a while.

We have some polynomials in the rows of the table. *The polynomials transform according to the irreducible representation which corresponds to the row*.[25] If a polynomial (displayed in a row of the table of characters) is subject to the projection $\hat{P}^{(\alpha)}$, then:

- if α does not correspond to the row, we obtain 0;
- if α corresponds to the row, we obtain either the polynomial itself (if the irreducible representation has dimension 1), or, if the dimension of the irreducible representation is greater than 1, a linear combination of the polynomials given in the same row (in parentheses).

If function f transforms according to a one-dimensional irreducible representation, the function is an eigenfunction of all the symmetry operators $\hat{\mathcal{R}}$, with the corresponding eigenvalues $\chi^{(\alpha)}(\hat{R})$.

Let us come back to R_x, R_y, R_z. Imagine R_x, R_y, R_z as oriented circles perpendicular to a rotation axis (i.e. x, y or z) which symbolizes rotations about these axes. For example, operation $\hat{E}$ and the two rotations $\hat{C}_3$ leave the circle R_z unchanged, while operations $\hat{\sigma}_v$ change its orientation to the opposite one, hence R_z transforms according to the irreducible representation A_2. It turns out, that R_x and R_y transform into their linear combinations under the symmetry operations and therefore correspond to a two-dimensional irreducible representation (E).

3 GROUP THEORY AND QUANTUM MECHANICS

Representation basis

If we have two equivalent[26] nuclei in a molecule, this *always* results from a molecular symmetry, i.e. at least one symmetry operation exchanges the positions of these two nuclei. There is no reason at all that electrons should prefer one such nucleus rather than the other.[27] Let us focus on molecular orbitals calculated for a fully symmetric Fock operator.[28] Therefore,

---

[25] Please recall the definition of the symmetry operation given on p. 907: $\hat{\mathcal{R}}f(\boldsymbol{r}) = f(\boldsymbol{r})$, where $\hat{\mathcal{R}}f(\boldsymbol{r}) = f(\hat{R}^{-1}\boldsymbol{r})$.

[26] With respect to physical and chemical properties.

[27] This may be not true for non-stationary states. The reason is simple. Imagine a long polymer molecule with two equivalent atoms at its ends. If one is touched by the tip of a tunnelling microscope and one electron is transferred to the polymer, a non-stationary asymmetric electron state is created.

[28] Limiting ourselves to molecular orbitals is not essential.

each molecular orbital has to be such, that when it is squared the electron density is the same on the equivalent nuclei.

What will happen, however, to the *molecular orbital itself*? Squaring removes information about its sign. The signs of both atoms may be the same (symmetric orbital), but they may also be opposite[29] (antisymmetric orbital). For example, the bonding orbital for the hydrogen molecule is symmetric with respect to the reflection in the plane perpendicular to the internuclear axis[30] and passing through its centre, while the antibonding orbital is antisymmetric with respect to the operation.

We know how to apply symmetry operations on molecular orbitals (p. 908) and transform them to other functions.

Under such a symmetry operation the orbital either remains unchanged (like the bonding mentioned above), or changes sign (like the antibonding).

or, if the orbital level is degenerate, we may obtain a different function. This function corresponds to the same energy, because in applying any symmetry operation we only exchange equivalent nuclei, which are otherwise treated on an equal footing in the Hamiltonian.

---

[29] This pertains to non-degenerate orbital levels. For a degenerate level any linear combination of the eigenfunctions (associated with the same level) is also an eigenfunction as good as those which entered the linear combination. A symmetry operation acting on an orbital gives another orbital corresponding to the same energy. In such a case, the squares of both orbitals in general does not exhibit the symmetry of the molecule. However, we can find such a linear combination of both, the square preserves the symmetry.

[30] Let us see what it really means in a very *formal* way (it may help us in more complicated cases). The coordinate system is located in the middle of the internuclear distance (on the x axis, the internuclear distance equals $2A$). The bonding orbital $\varphi_1 = N_1(a+b)$ and the antibonding orbital $\varphi_2 = N_2(a-b)$, where N_1 and N_2 are the normalization constants, the $1s$ atomic orbitals have the following form

$$a \equiv \frac{1}{\sqrt{\pi}}\exp\bigl[-|\boldsymbol{r}-\boldsymbol{A}|\bigr] = \frac{1}{\sqrt{\pi}}\exp\Bigl[-\sqrt{(x-A)^2+y^2+z^2}\Bigr],$$

$$b \equiv \frac{1}{\sqrt{\pi}}\exp\bigl[-|\boldsymbol{r}+\boldsymbol{A}|\bigr] = \frac{1}{\sqrt{\pi}}\exp\Bigl[-\sqrt{(x+A)^2+y^2+z^2}\Bigr],$$

$$\boldsymbol{A} = (A,0,0).$$

The operator $\hat{\sigma}$ of the reflection in the plane $x=0$ corresponds to the following unitary transformation matrix of the coordinates $\boldsymbol{U} = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$. Therefore, the inverse matrix $\boldsymbol{U}^{-1} = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$, i.e. the transformation $\boldsymbol{U}^{-1}\boldsymbol{r}$ means $x \to -x$, $y \to y$, $z \to z$, which transforms $a \to b$ and $b \to a$. Hence $\hat{\sigma}(a+b) = (b+a) = (a+b)$, $\hat{\sigma}(a-b) = (b-a) = -(a-b)$.

In both cases the molecular orbital represents an eigenfunction of the symmetry operator with eigenvalues $+1$ and -1, respectively.

If we obtain another orbital (φ_2), then we may begin to play with it by applying all the symmetry operations. Some operations will lead to the same (new) orbital, sometimes with the opposite sign. After other operations we will obtain the old orbital φ_1, sometimes with the opposite sign, and sometimes these operations will lead to a *third* orbital φ_3. Then we apply the symmetry operations to the third orbital, etc. until the final set of orbitals is obtained which transform into themselves when subject to symmetry operations. The set of such linearly independent orbitals φ_i, $i = 1, \ldots, n$, may be treated as the basis set in a vector space.

All the results of the application of operation $\hat{\mathcal{R}}_i$ on the orbitals φ_i are collected in a transformation matrix $\boldsymbol{R}_i$:

$$\hat{\mathcal{R}}_i \boldsymbol{\varphi} = \boldsymbol{R}_i^T \boldsymbol{\varphi}, \quad \text{where } \boldsymbol{\varphi} = \begin{bmatrix} \varphi_1 \\ \ldots \\ \varphi_n \end{bmatrix}. \tag{C.16}$$

The matrices $\boldsymbol{R}_i$, $i = 1, 2, \ldots, g$, form the n-dimensional representation (in general reducible) of the symmetry group of the molecule.

Indeed, let us see what happens if we apply operation $\hat{\mathcal{T}} = \hat{\mathcal{R}}_1 \hat{\mathcal{R}}_2$ to the function φ_i:

$$\hat{\mathcal{T}} \varphi_i = (\hat{\mathcal{R}}_1 \hat{\mathcal{R}}_2) \varphi_i = \hat{\mathcal{R}}_1 \boldsymbol{R}_2^T \boldsymbol{\varphi} = \boldsymbol{R}_2^T \hat{\mathcal{R}}_1 \boldsymbol{\varphi} = \boldsymbol{R}_2^T \boldsymbol{R}_1^T \boldsymbol{\varphi} = (\boldsymbol{R}_1 \boldsymbol{R}_2)^T \boldsymbol{\varphi}.$$

This means that all the matrices $\boldsymbol{R}_i$ form a representation.

BASIS OF A REPRESENTATION
The set of linearly independent functions φ_i, which served to create the representation, forms the *basis of the representation*.

The basis need not have been composed of the orbitals, it could be expressions like x, y, z or $x^2, y^2, z^2, xy, xz, yz$ or any linearly independent functions, provided they transform into themselves under symmetry operations. We may begin from an atomic orbital, and after applying symmetry operations will soon obtain a basis set which contains this orbital and all the other equivalent orbitals.

Decomposition of a function into irreducible representation components

Let us take a function f belonging to a Hilbert space. Since (see eq. (C.15)) $\sum_\alpha \hat{P}^{(\alpha)} = 1$, where α goes over all the irreducible representations of the group, f can be written as the sum of its components f_α, each component (belonging to the corresponding subspace of the Hilbert space) transforming according to the irreducible representation α

$$f = 1 \cdot f = \sum_\alpha \hat{P}^{(\alpha)} f = \sum_\alpha f_\alpha. \tag{C.17}$$

In view of (C.14), components f_α and f_β are automatically orthogonal if $\alpha \neq \beta$.

Example 12. ***Decomposition of a function.*** Let us take three hydrogen atoms in the configuration of an equilateral triangle, and assume we are in the simplest version of the molecular orbitals in the LCAO MO approximation, i.e. the atomic basis set is composed of the three 1s orbitals a, b, c centred on the three nuclei. Let us check whether the following functions:

$$u_1 = a + b + c,$$

$$u_2 = b - c,$$

$$u_3 = a - c,$$

form the basis to a (reducible) representation. If symmetry operations are applied to a, b, c, they transform into each other (cf. Fig. C.2), and the results obtained are easily shown as linear combinations of the functions u_1, u_2, u_3 (with $\boldsymbol{R}_i^T$ as transformation matrices). For example (see Table C.1, p. 911), $\hat{A}u_1 = a + b + c = u_1$, $\hat{A}u_2 = -b + c = -u_2$, $\hat{A}u_3 = a - b = -u_2 + u_3$. Hence,

$$\boldsymbol{A}^T = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & -1 & 1 \end{bmatrix}. \tag{C.18}$$

In this way (see (C.16)) we obtain $\boldsymbol{R}_i$ as:

$$\boldsymbol{E} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \boldsymbol{A} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & -1 \\ 0 & 0 & 1 \end{bmatrix}, \quad \boldsymbol{B} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & -1 \end{bmatrix}, \tag{C.19}$$

$$\boldsymbol{C} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \quad \boldsymbol{D} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & -1 \end{bmatrix}, \quad \boldsymbol{F} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & -1 \\ 0 & 1 & 0 \end{bmatrix}. \tag{C.20}$$

Let us check that $\boldsymbol{DF} = \boldsymbol{E}$ and $\boldsymbol{AD} = \boldsymbol{B}$, i.e. exactly as for operations: $\hat{D}\hat{F} = \hat{E}$, $\hat{A}\hat{D} = \hat{B}$, and so on. Thus this is a representation, moreover, this is a representation *in a block form*, because u_1 always transforms within itself, while u_2 and u_3 mix between themselves. It can be shown that this mixing cannot be avoided by any choice of u. Hence, u_1 alone represents the basis of a one-dimensional irreducible representation (A_1 – this is seen from the characters corresponding to the first block 1×1), while u_2 and u_3 form the basis of a two-dimensional irreducible representation (E). Note that from the mathematical form of the functions u, it follows that u_2 and u_3 have to correspond to the same energy and this energy is different from that corresponding to u_1. The conclusion is that a, b, c form the basis for a reducible representation, while their linear combinations u_1 and $\{u_2, u_3\}$ form the basis sets of two irreducible representations: A_1 and E. Any function which is a linear combination of a, b, c can also be represented as a linear combination of u_1, u_2, u_3.

The same symmetry orbitals can be obtained using the projection operators (C.13). Let us take any of functions a, b, c (the result does not depend on this choice), e.g., function a. In view of the beautiful equilateral triangle, such a function is no doubt a deformed object, which does not take the trouble to make the three vertices of the triangle equivalent. Let us see whether such a function has any component which transforms according to the irreducible representation A_1. To this end, let us use the projection operator $\hat{P}^{(A_1)}$ (Table C.5 of characters on p. 921 and eq. (C.13)): $\hat{P}^{(A_1)}a = \frac{1}{6}(a+b+c+a+b+c) = \frac{1}{3}(a+b+c)$. There *is* thus a fully symmetric component[31] in a. Now, let us use the same orbital a to obtain: $\hat{P}^{(A_2)}a = \frac{1}{6}(a+b+c-a-b-c) = 0$. This means that a does not contain anything which transforms according to A_2. Now it is the turn of the irreducible representation E: $\hat{P}^{(E)}a = \frac{2}{6}(2a-b-c+0\cdot a+0\cdot b+0\cdot c) = \frac{1}{3}[2(a-c)-(b-c)]$. We have now obtained is a linear combination of u_2 and u_3.

If the projections were made for function b, we would obtain a trivial repetition[32] for the irreducible representations A_1 and A_2 and a non-trivial result for the irreducible representation E: $\hat{P}^{(E)}b = \frac{2}{6}(2b-a-c+0\cdot a+0\cdot b+0\cdot c) = \frac{1}{3}[2(b-c)-(a-c)]$, which is just another linear combination of u_2 and u_3. These two functions are therefore inseparable and form the basis for a two-dimensional *irreducible* representation.

DECOMPOSITION INTO IRREDUCIBLE REPRESENTATIONS
Any function that is a linear combination of the basis functions of a *reducible* representation can be decomposed into a linear combination of the basis functions of those *irreducible* representations which form the reducible representation.

Most important

MOST IMPORTANT…
The wavefunctions corresponding to an energy level

- *form the basis of an irreducible representation* of the symmetry group of the molecule, or in other words, transform according to this irreducible representation
- the dimension of the representation is equal to the degeneracy of the energy level.

This is how it should be, because if a symmetry operation acts on an eigenfunction of the Hamiltonian, we will have only two possible results: 1) we obtain the same function to the accuracy of the sign (in the case of a one-dimensional representation, by definition irreducible); 2) another function *corresponding to the*

---

[31] This sentence carries a simple message, that by mixing symmetric objects we may obtain an asymmetric object, e.g., the asymmetric function $a+2b$ can be represented by the linear combination u_1+u_2, both functions transforming according to an irreducible representation of the symmetry group.

[32] $\hat{P}^{(A_1)}b = \frac{1}{3}(a+b+c)$ and $\hat{P}^{(A_2)}b = 0$.

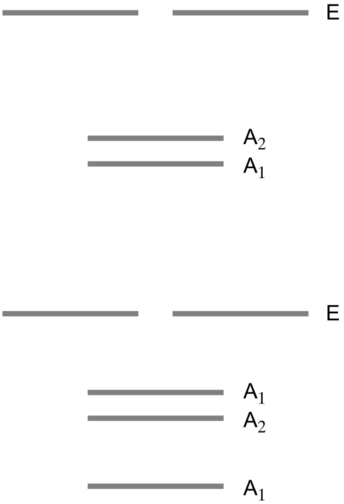


Fig. C.4. Each energy level corresponds to an irreducible representation of the symmetry group of the Hamiltonian. Its linearly independent eigenfunctions corresponding to a given level form the basis of the irreducible representation, or in other words, transform according to this representation. The number of the basis functions is equal to the degeneracy of the level.

same energy (because of the same physical situation). Acting on the function obtained and repeating the whole procedure, we will finally arrive at a set of n linearly independent functions which correspond to the same energy (the basis of a n-dimensional irreducible representation).

This means (Fig. C.4) that

> the energy levels may be labelled, each label corresponding to a single irreducible representation.

Eugene Wigner was the first who obtained this result. This will be of fundamental importance when the selection rules in spectroscopy will be considered.

We usually have plenty of energy levels, while the number of irreducible representations is small. Thus, in general there will be many levels with the same label. Group theory will never tell us either how many levels correspond to a particular irreducible representation, or to what energy they correspond.

4 INTEGRALS IMPORTANT IN SPECTROSCOPY

Direct product of the irreducible representations

We are quickly approaching the application of group theory to optical transitions in spectroscopy. The most important issue here will be a decision as to whether an integral is zero or non-zero. If the integral is zero, the transition is forbidden, if it is non-zero, it is allowed. To make such a decision we have to use what is known as the *direct product of irreducible representations*. Imagine basis functions $\{\varphi_i\}$ and $\{\psi_j\}$ which correspond to irreducible representations α and β of the symmetry group of a molecule. Let us make a set $\{\varphi_i\psi_j\}$ of all their possible products (i.e. the Cartesian product).

DIRECT PRODUCT
The products $\{\varphi_i\psi_j\}$ when subject to symmetry operations, lead (as usual) to a representation, and this representation is called the *direct product* $\Gamma^\alpha \times \Gamma^\beta$ *of the irreducible representations* Γ^α and Γ^β.

The functions $\{\varphi_i\psi_j\}$ form the basis set of a representation (reducible in general). The matrices of the representations we obtain as usual by applying symmetry operations (eq. (C.16)):

$$\begin{aligned}\hat{\mathcal{R}}\big[\varphi_i(\boldsymbol{r})\psi_j(\boldsymbol{r})\big] &= \varphi_i\big(\hat{R}^{-1}\boldsymbol{r}\big)\psi_j\big(\hat{R}^{-1}\boldsymbol{r}\big) = \sum_k \Gamma_{ki}^{(\alpha)}\varphi_k \sum_l \Gamma_{lj}^{(\beta)}\psi_l \\ &= \sum_{kl} \Gamma_{ki}^{(\alpha)}\Gamma_{lj}^{(\beta)}\varphi_k\psi_l = \sum_{kl} Z_{ij,kl}\varphi_k\psi_l,\end{aligned}$$

where $\Gamma_{ki}^{(\gamma)}$ are the matrix elements of the irreducible representation γ, $Z_{ij,kl} = \Gamma_{ki}^{(\alpha)}\Gamma_{lj}^{(\beta)}$. Of course,

the dimension of this representation is the product of the dimensions of the representations α and β, because this is the number of the functions $\varphi_k\psi_l$.

The characters $\chi^{(\alpha\times\beta)}$ of the representation can easily be obtained from the characters of the irreducible representations, we just have to multiply the latter:

$$\chi^{(\alpha\times\beta)}\big(\hat{R}\big) = \chi^{(\alpha)}\big(\hat{R}\big)\chi^{(\beta)}\big(\hat{R}\big). \tag{C.21}$$

Indeed, the formula is justified by:

$$\begin{aligned}\chi^{(\alpha\times\beta)}\big(\hat{R}\big) &= \sum_{kl} Z_{kl,kl} = \sum_{kl} \Gamma_{kk}^{(\alpha)}\Gamma_{ll}^{(\beta)} = \Big(\sum_k \Gamma_{kk}^{(\alpha)}\Big)\Big(\sum_l \Gamma_{ll}^{(\beta)}\Big) \\ &= \chi^{(\alpha)}\big(\hat{R}\big)\chi^{(\beta)}\big(\hat{R}\big).\end{aligned} \tag{C.22}$$

This rule can be naturally generalized for higher number of irreducible representations in the direct product (just multiply the characters of the irreducible representations). in a while we will have the product of three irreducible representations.

When is an integral bound to be zero?

Everyone studying this book should know how to calculate the integral

$$\int_{-1}^{+1} x\,\mathrm{d}x = \left[\frac{x^2}{2}\right]_{-1}^{+1} = \frac{1}{2} - \frac{1}{2} = 0.$$

Note, however, that we can tell what the value of the integral is without any calculation, just by looking at the integrand. Indeed, the integrand is odd with respect to the transformation $x \to -x$, i.e. the plot of the integral is an antisymmetric function with respect to the reflection in the plane perpendicular to x at $x = 0$. The integration limits are symmetric with respect to that point. An integral means the area under the plot, therefore what we gain for $x > 0$, we lose for $x < 0$ and the integral will be exactly zero.

> The force of group theory relies, even with a complicated integrands, on being able to tell immediately whether the integral *is* equal to zero. This allows us to predict whether an optical transition is allowed or forbidden.

We have to stress that these conclusions will be valid independent of the approximations used to calculate the molecular wave functions. The reason is that they follow from the symmetry, which is identical for exact and approximate wave functions.

The previous example can be generalized. Let us take the integral

$$\int f_\alpha f_\beta f_\gamma \ldots \mathrm{d}\tau, \tag{C.23}$$

where $f_\alpha, f_\beta, f_\gamma, \ldots$ *transform according to the irreducible representations* $\Gamma^{(\alpha)}, \Gamma^{(\beta)}, \Gamma^{(\gamma)}, \ldots$, of a symmetry group, and the integration is over the whole space.

> WHEN IS THE INTEGRAL EQUAL TO ZERO?
> If a representation (in general reducible), being the direct product of the irreducible representations $\Gamma^{(\alpha)}, \Gamma^{(\beta)}, \Gamma^{(\gamma)}, \ldots$ does not contain the fully symmetric representation (its all characters are equal to 1), the integral is equal to zero.

This is precisely our goal in this Appendix and this is why we have been working so hard with symmetry groups, operations, characters, etc. The essence of the theorem is very simple. The product $f_\alpha f_\beta f_\gamma \ldots$ transforms according to the (in general *reducible*) representation, which is the direct product of the irreducible representations $\Gamma^{(\alpha)}, \Gamma^{(\beta)}, \Gamma^{(\gamma)}, \ldots$. This means that according to eq. (C.17) the integrand $f_\alpha f_\beta f_\gamma \ldots$ can be represented as a linear combination of the basis functions of all the irreducible representations: $f_\alpha f_\beta f_\gamma \ldots = \sum_\mu g_\mu$, where g_μ transforms according to the irreducible representation $\Gamma^{(\mu)}$. Therefore, the integral (C.23) is the sum of the integrals

$$\int f_\alpha f_\beta f_\gamma \ldots \mathrm{d}\tau = \sum_\mu \int g_\mu \,\mathrm{d}\tau. \tag{C.24}$$

Let us take one such integral: $\int g_\mu \,\mathrm{d}\tau$. Note that the integration is over the whole space (i.e. the integration limits are symmetric). If the integrand g_μ were antisymmetric with respect to one or more symmetry operations, the integral would automatically be zero (the same argument as for $\int x \,\mathrm{d}x$). From this it follows that all integrals in the sum would be zero except the single one which contains the integrand transforming according to the fully symmetric representation.[33]

There are two (important to us) special cases of this theorem.

Two special cases

$\int f_\alpha f_\beta \, d\tau = \delta_{\alpha\beta} A$, i.e. in order for the integral not to vanish we have to have: $\Gamma^{(\alpha)} = \Gamma^{(\beta)}$.

The proof is very simple and relies on the fact that the characters of the fully symmetric irreducible representation are equal to 1. The number of times, a(A), the fully symmetric representation A, is present in the direct product $\Gamma^{(\alpha)} \times \Gamma^{(\beta)}$, we calculate from the formula:

$$\begin{aligned} \mathrm{a(A)} &= \frac{1}{g}\sum_i \chi^{(\alpha\times\beta)}(\hat{R}_i)\chi^{(\mathrm{A})}(\hat{R}_i)^* = \frac{1}{g}\sum_i \chi^{(\alpha\times\beta)}(\hat{R}_i) \\ &= \frac{1}{g}\sum_i \chi^{(\alpha)}(\hat{R}_i)\chi^{(\beta)}(\hat{R}_i)^* = \delta_{\alpha\beta}. \end{aligned} \tag{C.25}$$

This means that the fully symmetric representation is always present in $\Gamma^{(\alpha)} \times \Gamma^{(\alpha)}$ and therefore the integral does not vanish.[34]

Let us take the integral

$$\int f_\alpha f_\beta f_\gamma \, d\tau, \tag{C.26}$$

where $f_\alpha, f_\beta, f_\gamma$ transform according to the irreducible representations α, β, γ. For the integral not to vanish the direct product $\Gamma^{(\alpha)} \times \Gamma^{(\beta)}$ has to contain the representation $\Gamma^{(\gamma)}$.

This means that to have integral (C.26) not vanish, the function $f_\alpha f_\beta$ decomposes (eq. (C.17)) in such a way that there is a non-zero component belonging to $\Gamma^{(\gamma)}$. If this happens, according to the previous case, a component of the integrand will transform according to the fully symmetric representation, which will save the integral (C.26) from vanishing.

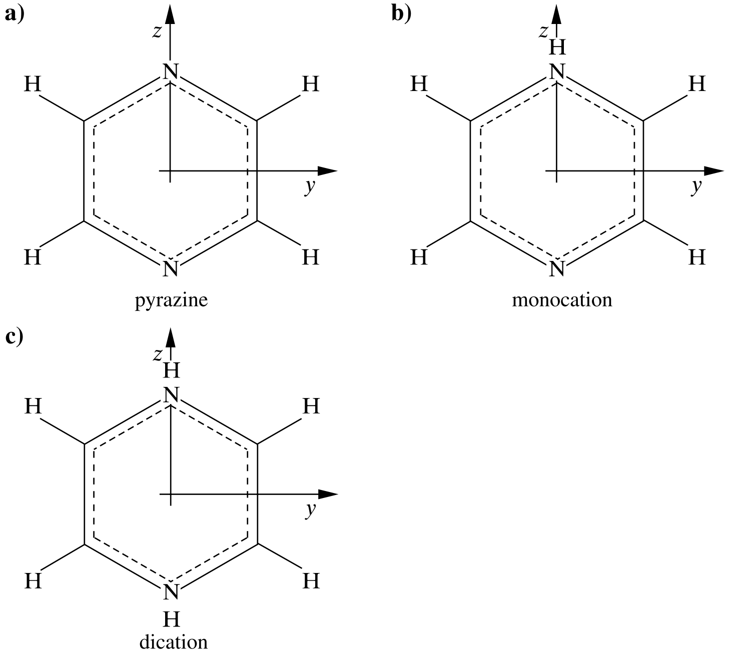
Selection rules for electronic transitions (UV-VIS)

The selection rules will be shown taking the example of pyrazine and its mono- and diprotonated ions (Fig. C.5).

---

[33] Only for the fully symmetric representation are all the characters equal to 1, and therefore the corresponding function does not change under symmetry operations.

[34] This is easy to understand. What transforms, according to $\Gamma^{(\alpha)} \times \Gamma^{(\alpha)}$, is the product of two (in general different) functions, each belonging to $\Gamma^{(\alpha)}$. It means that the function behaves in a very special way (typical for $\Gamma^{(\alpha)}$) under symmetry operations, e.g., changes sign under $\hat{R}_1$, while other operations leave it unchanged. If we have a product of two such functions, this means the product does not change at all under $\hat{R}_1$ (and, of course, the other operations), i.e. transforms according to the fully symmetric operation. This is why the fully symmetric representation is always present in $\Gamma^{(\alpha)} \times \Gamma^{(\alpha)}$.



pyrazine

monocation

dication

Fig. C.5. Pyrazine (a) and its mono- (b) and diprotonated (c) derivatives. The x axis is perpendicular to the ring plane, the y axis is in the ring plane perpendicular to the NN axis, and z is the NN axis.

A glimpse at the chemical formulae is sufficient to tell that the monocation of pyrazine has the same symmetry as H_2O, which corresponds to symmetry group C_{2v} (see Table C.4), while pyrazine and its diprotonated derivative have the symmetry identical with that of naphthalene, i.e. D_{2h}. Let us focus first on the last case.

Example 13. ***Pyrazine and its diprotonated derivative.*** Every book on group theory contains a table of characters of the symmetry group D_{2h} (Table C.6, x axis perpendicular to the plane of the molecule, z goes through the nitrogen atoms).

Table C.6. D_{2h} group table of characters

D_{2h}	$\hat{E}$	$\hat{C}_2(z)$	$\hat{C}_2(y)$	$\hat{C}_2(x)$	$\hat{\imath}$	$\hat{\sigma}(xy)$	$\hat{\sigma}(xz)$	$\hat{\sigma}(yz)$	
A_g	1	1	1	1	1	1	1	1	x^2, y^2, z^2
B_{1g}	1	1	−1	−1	1	1	−1	−1	R_z xy
B_{2g}	1	−1	1	−1	1	−1	1	−1	R_y xz
B_{3g}	1	−1	−1	1	1	−1	−1	1	R_x yz
A_u	1	1	1	1	−1	−1	−1	−1	
B_{1u}	1	1	−1	−1	−1	−1	1	1	z
B_{2u}	1	−1	1	−1	−1	1	−1	1	y
B_{3u}	1	−1	−1	1	−1	1	1	−1	x

From Table C.6 we see[35] that what we call irreducible representations are simply the *distinct rhythms* of pluses and minuses, which after squaring, give the fully symmetric behaviour. All the electronic states of pyrazine and its diprotonated derivative can be described by the irreducible representation labels: $A_g, B_{1g}, B_{2g}, B_{3g}, A_u, B_{1u}, B_{2u}, B_{3u}$.

We may ask: what are the selection rules for state-to-state optical transitions? Are all transitions allowed, or are some of them forbidden? From the theory of the electromagnetic field (cf. Chapters 7 and 12) it follows that the probability of the transition between states k and l is proportional to $|\mu_{kl}(x)|^2$ or to $|\mu_{kl}(y)|^2$ or to $|\mu_{kl}(z)|^2$, respectively,[36] depending on the electromagnetic wave polarization along axes x, y or z axes, with:

$$\begin{aligned}\mu_{kl}(x) &= \int \psi_k^* \hat{\mu}_x \psi_l \, \mathrm{d}\tau,\\ \mu_{kl}(y) &= \int \psi_k^* \hat{\mu}_y \psi_l \, \mathrm{d}\tau,\\ \mu_{kl}(z) &= \int \psi_k^* \hat{\mu}_z \psi_l \, \mathrm{d}\tau,\end{aligned} \tag{C.27}$$

where ψ stands for the electronic states k, and l [eq. (6.8) on p. 225], $\hat{\mu}_x, \hat{\mu}_y, \hat{\mu}_z$ are the operators of the molecular dipole moment components,[37] e.g.,

$$\hat{\mu}_z = \sum_i q_i z_i,$$

q_i is the electric charge of the particle (electron or nucleus) having its z component equal to z_i. Since we will decide, by using group theory, whether this integral[38] vanishes or not, the important thing is that μ_x transforms in exactly the same way as the coordinate x. The integrand $\psi_k^* \hat{\mu}_x \psi_l$ transforms as the direct product of the three irreducible representations: that of ψ_k, that of $\hat{\mu}_x$ and that of ψ_l.

Excitations from the ground-state

Suppose we

- have a molecule in its ground-state ψ_k (thus, belonging to the fully symmetric irreducible representation A_g),

---

[35] Note that all the irreducible representations of the symmetry group of the molecules under consideration are one-dimensional, hence their energy levels are non-degenerate.

[36] From the equality $|\mu_{kl}(x)|^2 = |\mu_{lk}(x)|^2$ and similarly for y and z, it follows that the optical excitation and the corresponding emission have the same probability.

[37] This may look alarming, because the operator depends on the choice of the coordinate system (cf. Appendix X). Do not worry, everything is all right. Even if the dipole moment depends on such a choice, any two choices give dipole moments differing by a *constant vector*. This vector being a constant can be shifted outside the integral and the integral itself will become zero, because ψ_k and ψ_l are orthogonal. Thus, to our delight, light absorption does not depend on the choice of the coordinate system. This is fine.

[38] The integration goes over all the electronic coordinates.

- immobilize the molecule in space (say, in a crystal),
- introduce the coordinate system in the way described above,
- irradiate the molecule with light polarized along the x axis

polarization x

and ask *which states will the molecule be excited to*. The direct product of A_g and the irreducible representation to which x belongs, decomposes into some irreducible representations. For the optical transition to be allowed, we have to find among them the irreducible representation to which ψ_l belongs (just recall that $\int f_\alpha f_\beta \, d\tau = \delta_{\alpha\beta} A$). Only then will the integrand contain something that has the chance to transform according to the fully symmetric representation. The x coordinate belongs to the representation B_{3u} (see Table C.6, last column). Therefore, let us see what represents the direct product $A_g \times B_{3u}$. We have eq. (C.12) for the integer $a(\alpha)$ that is the number of irreducible representations α in a given reducible representation. Let us calculate this number for the representation (in general reducible) being the direct product, and all the irreducible representations α. In this particular case the direct product is[39] $A_g \times B_{3u}$. We have

$$\begin{aligned} a(A_g) &= \frac{1}{8}[1\times 1 + 1\times(-1) + 1\times(-1) + 1\times 1 + 1\times(-1) + 1\times 1 \\ &\quad + 1\times 1 + 1\times(-1)] = 0, \\ a(B_{1g}) &= \frac{1}{8}[1\times 1 + 1\times(-1) + (-1)\times(-1) + (-1)\times 1 \\ &\quad + 1\times(-1) + 1\times 1 + (-1)\times 1 + (-1)\times(-1)], \\ &= 0 \end{aligned}$$

etc., all zeros, and finally

$$\begin{aligned} a(B_{3u}) = \frac{1}{8}[&1\times 1 + (-1)\times(-1) + (-1)\times(-1) + 1\times 1 + (-1)\times(-1) + 1\times 1 \\ &+ 1\times 1 + (-1)\times(-1)] = 1, \end{aligned}$$

exactly as we have expected. Thus, we can write[40]

$$A_g \times B_{3u} = B_{3u}.$$

---

[39] The characters of $A_g \times B_{3u}$ are as follows (in order of the symmetry operations in the table of characters):

$$1 \quad -1 \quad -1 \quad 1 \quad -1 \quad 1 \quad 1 \quad -1$$

i.e. they are identical to those of the (it turns out irreducible) representation B_{3u}. Such a product is really easy to form. In the table of characters one finger goes horizontally over the characters of A_g (they are all equal to 1), while the second finger moves similarly over the characters of B_{3u} and we multiply what the first finger shows by what the second finger shows. The result is the character of the direct product $A_g \times B_{3u}$, which in this case turns out to be exactly the character of B_{3u}. This is why we may expect that $a(\alpha)$ will all be zero except $a(B_{3u}) = 1$.

[40] We may say that the fully symmetric representation plays the role of unity in the multiplication of irreducible representations.

Now only those ψ_l are allowed in optical transitions (from the ground state A_g) that are labelled by B_{3u}, because only the direct product $B_{3u} \times B_{3u}$ may contain the fully symmetric irreducible representation A_g. Thus, the transitions $A_g \Rightarrow B_{3u}$ as well as $B_{3u} \Rightarrow A_g$ are allowed, if the light is polarized along x, i.e. perpendicular to the ring of the molecule.

polarization y

Now let us take light polarized along y, i.e. within the molecular plane, perpendicularly to the N–N line. This time we are interested in the irreducible representations that arise from $A_g \times B_{2u}$, because y transforms according to B_{2u}. Very similarly [by analyzing $a(\alpha)$] we find that

$$A_g \times B_{2u} = B_{2u}.$$

This means that the allowed states are now of the B_{2u} type.

polarization z

Similarly, for polarization along z (z belongs to B_{1u}), i.e. along the nitrogen-nitrogen direction, we have

$$A_g \times B_{1u} = B_{1u}.$$

Thus for polarization parallel to the NN axis of the molecule, absorption may occur from the ground state to any state of the B_{1u} type (and *vice versa*).

Nothing more can be said when relying solely on group theory. We will not get any information about the energies of the transitions, or about the corresponding intensities. To get this additional (and important) information we have work hard to solve the Schrödinger equation, rather than count on some easy profits obtained by the primitive multiplication of integers (as in group theory). To obtain the intensities, we have to calculate the transition moment integrals μ_{kl}. However, group theory, by excluding from the spectrum *many transitions* (forbidden ones), provides a lot of important information on the molecule. Table C.7 collects the calculated light frequencies[41] ($\bar{\nu}$ in wavenumbers, or cm^{-1}, $\nu = c\bar{\nu}$, where ν is the usual frequency), the *oscillator strengths* f_{kl} (in a.u.)

transition moment

$$f_{kl} = \frac{4\pi c}{3}\nu|\mu_{kl}|^2, \tag{C.28}$$

as well as the polarization of light for excitations from the electronic ground state for pyrazine and the pyrazine monocation. It is seen that the left-hand side of Table C.7 is consistent with the selection rules derived above. Indeed, a large f_{kl} only corresponds to those transitions from the ground state of the pyrazine that have been predicted as allowed (B_{1u}, B_{2u} and B_{3u}). The predicted polarization also agrees with the observed polarization.

oscillator strength

Excitations from an excited state

Calculations for absorption from the ground-state were particularly simple. Now let us see whether anything will be more complicated for the transitions from an excited state of the B_{2g} type of symmetry. We are going to calculate $a(\alpha)$ (for every α) for the following representations:

---

[41] J. Koput, unpublished results.

Table C.7. Wave numbers ($\bar{\nu}$), oscillator strengths (f_{kl}) and light polarization (in parentheses)

Pyrazine			Pyrazine monocation		
Excited state	$\bar{\nu}$	f_{kl}	Excited state	$\bar{\nu}$	f_{kl}
B_{3u}	28960	0.015(x)	B_1	27440	0.007(x)
B_{2u}	36890	0.194(y)	B_2	34130	0.280(y)
B_{2g}	38890	0.0	A_2	45100	0.0
A_u	41710	0.0	A_1	49720	0.126(z)
B_{1u}	49800	0.183(z)	B_1	57380	0.012(x)
B_{1g}	57070	0.0	A_2	57710	0.0
B_{1u}	57420	0.426(z)	A_1	58210	0.625(z)
A_u	60170	0.0	A_2	59830	0.0
B_{2g}	60970	0.0	B_2	60370	0.010(y)

for polarization along x: $B_{2g} \times B_{3u}$
for polarization along y: $B_{2g} \times B_{2u}$
for polarization along z: $B_{2g} \times B_{1u}$.

The characters of the representation $B_{2g} \times B_{3u}$ are the following (Table C.6, the first finger goes along B_{2g}, the second – along B_{3u}, etc.)

$$1 \quad -1 \quad -1 \quad -1 \quad -1 \quad 1 \quad 1$$

and are identical with the characters of B_{1u}. Hence, even without any calculation of $a(\alpha)$, we have $B_{2g} \times B_{3u} = B_{1u}$. Thus the transitions (for the polarization along x) are allowed only for the states labelled by B_{1u}, because otherwise there is no chance of obtaining a fully symmetric integrand. Similarly, by multiplying B_{2g} and B_{2u} we obtain the following characters of $B_{2g} \times B_{2u}$:

$$1 \quad 1 \quad 1 \quad 1 \quad -1 \quad -1 \quad -1 \quad -1$$

and these are identical to the characters of A_u, therefore $B_{2g} \times B_{2u} = A_u$. If the polarization of light is along y, the only excitations (or deexcitations) possible are for states belonging to A_u. Finally, for polarization along z, we find the characters of $B_{2g} \times B_{1u}$:

$$1 \quad -1 \quad -1 \quad 1 \quad -1 \quad 1 \quad 1 \quad -1$$

that turn out to be those of B_{3u}. This means that $B_{2g} \times B_{1u} = B_{3u}$ and that the transitions are possible only for states belonging to B_{3u}.

Example 14. ***Pyrazine monocation.*** As to the selection rules, nothing was said so far about the pyrazine monocation. We will be interested in excitations from the electronic ground state (as in Table C.7). The pyrazine monocation corresponds to symmetry group C_{2v} (Table C.8).

The ground state belongs to the fully symmetric irreducible representation A_1. Since (as before) we begin by excitations from the ground state, let us see which irreducible representations arise from $A_1 \times B_1$ (for the x polarization of light, see Table C.8, x transforms according to B_1), $A_1 \times B_2$ (for the y polarization)

Table C.8. C_{2v} group characters

C_{2v}	E	C_2	$\sigma_v(xz)$	$\sigma_v(yz)$		
A_1	1	1	1	1	z	x^2, y^2, z^2
A_2	1	1	−1	−1	R_z	xy
B_1	1	−1	1	−1	x, R_y	xz
B_2	1	−1	−1	1	y, R_x	yz

and $A_1 \times A_1$ (for the z polarization). We calculate the characters of $A_1 \times B_1$ by multiplying 1 by

$$1 \quad -1 \quad 1 \quad -1,$$

and checking in Table C.8 that these correspond to B_1 (it has to be like this, because the characters of A_1 are all equal to 1), i.e. $A_1 \times B_1 = B_1$. Similarly, even without immediate checking, we see that $A_1 \times B_2 = B_2$ and $A_1 \times A_1 = A_1$. In this way the following allowed transitions from the ground state (A_1) have been predicted:

for polarization along x: $A_1 \rightarrow B_1$
for polarization along y: $A_1 \rightarrow B_2$
for polarization along z: $A_1 \rightarrow A_1$.

This agrees with $f_{kl} \neq 0$ values of Table C.7.

Now we are able to compare the spectrum for pyrazine and for its monocation, Table C.7. Attaching a proton to the pyrazine (creating its monocation) does not look like something that would ruin the UV-VIS spectrum. We might expect that the frequencies of the bands, even their intensities should be somehow similar in both molecules. As we can see from the Table, the frequencies are similar indeed. For both molecules there are forbidden ($f_{kl} = 0$) and allowed ($f_{kl} \neq 0$) transitions. Note that what is allowed for the pyrazine is also allowed for its cation, the light polarization coincides, even the values of f_{kl} are similar (we have taken into account that the transition to B_{1u} in pyrazine with frequency 49800 cm^{-1} corresponds to the transition to A_1 in the monocation with frequency 49720 cm^{-1}). In the monocation there are some additional transitions allowed: to B_1 and to B_2. This is quite understandable, because the number of symmetry operations for the monocation is smaller, and the higher molecular symmetry the more numerous are forbidden transitions. If a molecule had no symmetry operations at all (except of course the identity symmetry), then all transitions would be allowed.

Thus, practically with zero effort, we find the selection rules in UV-VIS for any molecule we want.

Selection rules in IR and Raman spectra

The selection rules derived above pertain to electronic transitions, when the positions of the nuclei are fixed in space. Now a vibrational excitation of the molecule will be considered, while the electronic state is assumed to be unchanged. The vibrations of a molecule are related to its vibrational levels (each corresponding to an irreducible representation) and the corresponding vibrational wave functions,

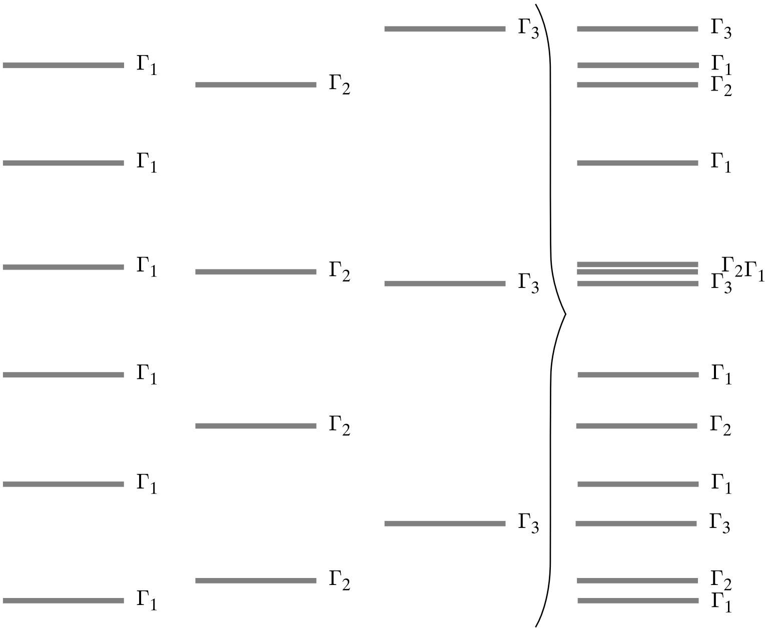


Fig. C.6. Small amplitude harmonic vibrations of a molecule (N atoms) are described by $3N - 6$ independent harmonic oscillators (normal modes). Each normal mode is characterized by an irreducible representation. A diagram shows the vibrational energy levels of three normal modes corresponding to the irreducible representations $\Gamma_1, \Gamma_2, \Gamma_3$. The modes have different frequencies, hence the interlevel separations are different for all of them (but equal for a given mode due to the harmonic potential). On the right-hand side all these levels are shown together.

and the IR spectrum results from transitions between such levels. Fig. C.6 shows the energy levels of three normal modes.

In the harmonic approximation the problem of small amplitude vibrations (Chapters 6 and 7) reduces to the $3N - 6$ normal modes (N is the number of atoms in the molecule). Each of the normal modes may be treated as an independent harmonic oscillator. A normal mode moves all the atoms with a certain frequency about their equilibrium positions in a concerted motion (the same phase). The relative deviations (i.e. the ratios of the amplitudes) of the vibrating atoms from equilibrium are characteristic for the mode, while the deviation itself is obtained from them by multiplication by the corresponding normal mode coordinate $Q \in (-\infty, \infty)$. The value $Q = 0$ corresponds to the equilibrium positions of all the atoms, Q and $-Q$ correspond to two opposite deviations of any atom from its equilibrium position.

Each normal mode belongs to an irreducible representation of the symmetry group of the molecule. What does it really mean? In any mode the displacements of the *equivalent* atoms from equilibrium have the same absolute value, although they may differ in sign.

We assume that small atomic deviations satisfy the symmetry requirements of the symmetry group of the molecule (valid for all atoms in equilibrium positions) and transform according to the irreducible representation, to which the normal mode belongs. Squaring the deviations destroys information about their signs, i.e. the absolute values of the deviations of the equivalent atoms are the same. This means that the squares of deviations transform according to the fully symmetric representation of the group.

To establish the vibrational selection rules, let us first define the vibrational states of $3N - 6$ harmonic oscillators (normal modes). The ground state of the system is no doubt the state in which *every* normal mode i is in its ground state, $\psi_{i,0}$. The ground-state wave function of the i-th normal mode reads as (p. 166)

$$\psi_{i,0} = N_0 \exp\left(-a_i Q_i^2\right), \tag{C.29}$$

where $a_i > 0$ is a constant, and Q_i is the normal mode coordinate. Whatever this normal mode is, the wave function contains the *square* of Q_i, i.e. the sign of the deviations of the equivalent atoms *is irrelevant*.

The squares of the deviations, and therefore function $\psi_{i,0}$ itself, transform independently of i.

Let us denote this fully symmetric irreducible representation by A_1. The wavefunction of the first *excited state* of a normal mode has the form (p. 166)

$$\psi_{i,1} = N_1 Q_i \exp\left(-a_i Q_i^2\right) \tag{C.30}$$

and we see that $\psi_{i,1}$ transforms exactly as the coordinate Q_i does, i.e. according to the irreducible representation to which the normal mode belongs (because Q_i^2 in the exponent and therefore the exponent itself both belong to the fully symmetric representation). In the harmonic approximation the total vibrational wavefunction of the system of $3N - 6$ normal (i.e. independent) oscillators can be written as:

$$\psi_0^{\mathrm{osc}} = \psi_{1,0}\psi_{2,0}\psi_{3,0}\ldots\psi_{3N-6,0}, \tag{C.31}$$

the zeros in the indices mean that all the modes are in their ground states. This means that ψ_0^{osc} transforms according to the representation being the direct product $A_1 \times A_1 \times A_1 \times \cdots \times A_1 = A_1$ (a banality, all the characters of A_1 are equal 1). Now let us focus on the excited states of the $3N - 6$ vibrational modes. The excited states may be quite complex, but the most important (and the simplest) are those with all the normal modes in their ground states, except a single mode that is in its first excited state. A transition from the many-oscillator ground state to such an excited state is called a *fundamental transition*. The intensities of the fundamental transitions are by at least one order of magnitude larger than the others. This is why we will focus on the selection rules for such transitions. Let us take one such

fundamental transition

singly excited state (with the first mode excited):

$$\psi_1^{\mathrm{osc}} = \psi_{1,1}\psi_{2,0}\psi_{3,0}\ldots\psi_{3N-6,0}. \tag{C.32}$$

The function $\psi_{1,1}$ corresponding to the first excited state transforms according to the irreducible representation Γ, to which the normal mode 1 belongs. Thus, ψ_1^{osc} transforms according to $\Gamma \times A_1 \times A_1 \times A_1 \times \cdots \times A_1 = \Gamma$, i.e. it belongs to the same irreducible representation as $\psi_{1,1}$. Of course, if the only excited mode were the i-th one, then the many-oscillator wavefunction would belong to the same irreducible representation as the wavefunction of the particular oscillator does. We will need this result later.

IR selection rules. Let us consider a molecule with a fixed position in a Cartesian coordinate system. To excite the molecule, IR light (because the separation of the vibrational levels corresponds to the infrared region) is used, which is polarized along the x axis. Theory of electromagnetism says the transition integral[42] decides the intensity of the absorption

$$\int \psi_0^{\mathrm{osc}} \hat{\mu}_x \psi_1^{\mathrm{osc}} \,\mathrm{d}\tau, \tag{C.33}$$

where $\hat{\mu}_x$ stands for the dipole moment component x. The selection rules establish which integrals of this kind will be zero for symmetry reasons. To this end we need information about the irreducible representations to which ψ_0^{osc}, $\hat{\mu}_x$, ψ_1^{osc} belong.[43] Since ψ_0^{osc} transforms according to A_1, for the integral to survive, the function ψ_1^{osc} has to belong to the same irreducible representation as $\hat{\mu}_x$ (and therefore x itself). We showed above that ψ_1^{osc} belongs to the same irreducible representation to which the normal mode 1 belongs. In other words, the rule is:

> SELECTION RULE IN IR
> the transition from the ground state is allowed for those normal modes that transform as x, where x is the direction of light polarization, and similarly for light polarization along y and z.

Raman selection rules. The physics of Raman spectra[44] is different: rather than direct absorption this is light scattering (in the UV-VIS region) on molecules. It turns out, that beside the light the source is emitting, we also detect quanta of energy lower or higher by $h\nu$, where ν is the vibrational frequency of the molecule.

---

[42] The integration goes over the coordinates of the nuclei.

[43] We are going to analyze the direct product of these three representations. If it contains the fully symmetric representation, the integral is not zero.

[44] Chandrasekhar Venkata Raman (1888–1970), Indian physicist, professor at the University of Calcutta and at the Indian Scientific Institute in Bangalore. In 1928 Raman discovered light scattering that is accompanied by a change of frequency (by frequency of the molecular vibrations). Raman received the Nobel prize in 1930 "*for his work on the scattering of light and for the discovery of the effect named after him*".

For the Raman scattering to be non-zero, at least one of the following integrals should be non-zero

$$\int \psi_0^{\mathrm{osc}} \hat{\alpha}_{qq'} \psi_1^{\mathrm{osc}} \, \mathrm{d}\tau, \tag{C.34}$$

where $\hat{\alpha}_{qq'}$ with $q, q' = x, y, z$ is a component of the polarizability tensor which transforms as one of the following (cf. eq. (12.40), p. 636): $qq' = x^2$, y^2, z^2, xy, xz, yz or their linear combinations (this information is available in the tables of characters). Identical reasoning leads to the conclusion that

> the normal mode excited in a fundamental transition has to belong to the same irreducible representation as the product qq'.

It remains to be seen to *which irreducible representations the normal modes belong*. The procedure consists of two stages.

Stage 1. First, the global Cartesian coordinate system is chosen, Fig. C.7. In this system we draw the equilibrium configuration of the molecule, with numbered atoms. A local Cartesian coordinate system is located on each atom with axes parallel to the axes of the global coordinate system. For each atom, we draw the arrows of its displacements along x, y and z, oriented towards the positive values (altogether $3N$ displacements), assuming that the displacements of equivalent atoms have to be the same. When symmetry operations are applied, these displacements transform into themselves[45] and therefore form a basis set of a (reducible) representation Γ of the symmetry group of the molecule (in its equilibrium position). This representation will be decomposed into the irreducible representations.

Stage 2. The reducible representation describes the genuine (internal) vibrations as well as the six apparent vibrations (three translations and three rotations). The apparent vibrations correspond to those irreducible representations that are associated to x, y, z (translations) and R_x, R_y, R_z (rotations). We know from the corresponding table of characters what the later ones are. Summing up: the reducible representation mentioned above has to be decomposed into irreducible representations. The decomposition yields $\Gamma = \mathrm{a}(\Gamma_1)\Gamma_1 + \mathrm{a}(\Gamma_2)\Gamma_2 + \mathrm{a}(\Gamma_3)\Gamma_3 \ldots$. From this decomposition we have to subtract (in order to eliminate the apparent vibrations) all the irreducible representations the x, y, z, R_x, R_y and R_z belong.

After these two stages we are left with the number of the irreducible representations which pertain to the genuine vibrations.[46] Only after this can we establish the vibrational selection rules according to the procedure used before. All this will be shown by a simple example of the carbonate anion CO_3^{2-} that in its equilibrium configuration corresponds to the D_{3h} symmetry group, Fig. C.7.

---

[45] For example, a displacement of an atom along x under a symmetry operation turns out to be a displacement of another equivalent atom.

[46] Rather internal motions. Note that some of these genuine vibrations may correspond to rotations of the functional groups in the molecule.

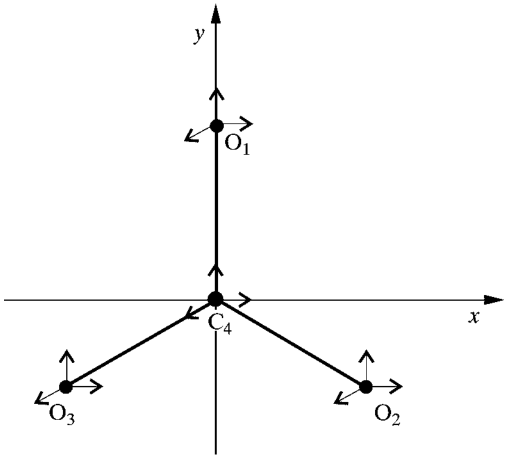


Fig. C.7. The carbonate anion CO_3^{2-}, the coordinate system used and the versors describing the displacements of the atoms.

Example 15. ***IR and Raman spectra of the carbonate anion.*** To decompose a reducible representation into irreducible ones, we do not need the reducible representation be given in full details. It is sufficient to know its characters (p. 920). These characters are easy to deduce by considering what happens to the displacement vectors along x_i, y_i, z_i (for atom i) under all the symmetry operations. What will greatly simplify our task is that *only the diagonal* elements of the matrices of the reducible representation contribute to the characters. How it looks in practice is shown in Table C.9.

Thus, the characters of the reducible representation have been found. To decompose the representation, we have to know the table of characters for the D_{3h} symmetry group, Table C.10.

Let us write (in the same order as in Table C.10) the characters of the reducible representation just found:

$$12 \quad 0 \quad -2 \quad 4 \quad -2 \quad 2.$$

Now, let us find (p. 920) how many times $[a(\alpha)]$ the irreducible representation α is present in the reducible representation (the sum over classes: number of operations in class × the calculated character × the character of irreducible representation):

$$a(A') = \frac{1}{12}[1 \times 12 \times 1 + 2 \times 0 \times 1 + 3 \times (-2) \times 1 + 1 \times 4 \times 1 + 2 \times (-2) \times 1 \\ + 3 \times 2 \times 1] = 1.$$

Similarly, we find (only needing to know how to multiply such numbers as $1, 2, 3$) that

$$a(A) = 1, \quad a(E') = 3, \quad a(A_1'') = 0, \quad a(A_2'') = 2, \quad a(E) = 1.$$

This means that the reducible representation in question decomposes into

$$\Gamma = A_1' + A_2' + 3E' + 2A_2'' + E''. \tag{C.35}$$

Table C.9.

Class	The character of the corresponding matrix
E	$\chi(E)=12$ *Justification*: each versor transforms into itself. Hence each diagonal element is equal to 1, and the number of them is equal to 3 times the number of atoms = 12
$2C_3$	$\chi(C_3)=0$ *Justification*: 0 from the oxygens, because they transform into other oxygens $+1(\text{from } z_4)+\cos 120°(\text{from } x_4)+\cos 120°(\text{from } y_4)=0$
$3C_2$	$\chi(C_2)=-2$ *Justification*: it is sufficient to consider only one of the operations of the class – others will have the same character. Let us take the rotation about the C_2 axis going through O_1 and C. Then the only unit vectors that transform into themselves (eventually changing sign – then the contribution to the character is −1) are those related to O_1 and C. We have $\chi(C_2)=-1(\text{from } z_4)+(-1)(\text{from } z_1)-1(\text{from } x_1)-1(\text{from } x_4)+1(\text{from } y_1)+1(\text{from } y_4)=-2$
σ_h	$\chi(\sigma_h)=4$ *Justification*: the contribution from each atom will be the same, i.e. χ will be equal to 4 times the contribution from a single atom, the latter equals: $-1(\text{from } z)+1(\text{from } x)+1(\text{from } y)=1$
$2S_3$	$\chi(S_3)=-2$ *Justification*: only C gives a contribution, which is equal to: $-1(\text{from } z_4)-\frac{1}{2}(\text{from } x_4)-\frac{1}{2}(\text{from } y_4)=-2$
$3\sigma_v$	$\chi(\sigma_v)=2$ *Justification*: Let us take only a single operation from the class, the one, which represents the reflection in the plane going through O_1 and C_4. Then the contributions to χ are the same for both atoms, and one gives: $-1(\text{from } x)+1(\text{from } z)+1(\text{from } y)=1$.

Table C.10. Characters of the irreducible representations of symmetry group D_{3h}

D_{3h}	E	$2C_3$	$3C_2$	σ_h	$2S_3$	$3\sigma_v$		
A_1'	1	1	1	1	1	1		x^2+y^2, z^2
A_2'	1	1	−1	1	1	−1	R_z	
E'	2	−1	0	2	−1	0	x, y	x^2-y^2, xy
A_1''	1	1	1	−1	−1	−1		
A_2''	1	1	−1	−1	−1	1	z	
E''	2	−1	0	−2	1	0	R_x, R_y	xz, yz

From the table of characters, we see that the apparent vibrations (see the irreducible representations corresponding to x, y, z, R_x, R_y, R_z) belong to A_2'', E', A_2', E''. After subtracting them from Γ, we obtain the irreducible representations that correspond to the genuine vibrations:

$$A_1',\ A_2'',\ 2E',$$

i.e. one vibration of symmetry A_1' (and a certain frequency ν_1), two vibrations (each doubly degenerate) of symmetry E' (they differ by frequency $\nu_3 \neq \nu_4$) and one vibration of A_2'' symmetry (corresponding to frequency ν_2).

SELECTION RULES FOR IR:
Therefore, we expect the following selection rules for the fundamental transitions in the IR spectrum for the CO_3^{2-} anion:

1. x and y belong to representation E', and therefore frequencies ν_3 and ν_4 are active in IR;
2. z belongs to representation A_2'', and therefore frequency ν_2 is active in IR.

SELECTION RULES FOR RAMAN SPECTRA
For the Raman spectra we expect the following selection rules. The vibrations with the frequency will be active:

1. ν_1, because $x^2 + y^2$ and z^2 belong to A_1';
2. ν_3 and ν_4, because $x^2 - y^2$ and xy belong to E',

while the vibration of the frequency ν_2 will be inactive in Raman spectroscopy, because none of the polarizability components (symbolized by x^2, y^2, etc.) belongs to A_2''.

The results are collected in Table C.11 (sign "+" = active vibration, sign "–" = inactive vibration, the polarization of the light is shown in parentheses).

As seen from Table C.11, in case of the carbonate anion the vibration ν_1 is inactive in IR, but active in Raman spectroscopy, while the opposite is true for ν_2. The vibrations with frequencies ν_3 and ν_4 are active both in IR and in Raman spectra.

EXCLUSION RULE
If the molecule under study has a centre of symmetry, the *exclusion rule* is valid, i.e. the vibrations that are active in IR are inactive in the Raman spectrum, and *vice versa*.

This follows from the fact that, in this case, x, y, z belong to different irreducible representations than $x^2,\ y^2,\ z^2,\ xy,\ xz,\ yz$. Indeed, the x, y, z are antisymmetric with respect to the inversion operation, whereas $x^2,\ y^2,\ z^2,\ xy,\ xz,\ yz$ or their combinations are symmetric with respect to inversion. This guarantees that they belong to *different* irreducible representations, therefore for a molecule with

Table C.11. Transitions in CO_3^{2-} active (+) and inactive (−) in IR and in Raman spectra

Representation	ν	IR (polarization)	Raman
A_1'	ν_1	−	+
A_2''	ν_2	+ (z)	−
E'	ν_3	+ (circular)	+
E'	ν_4	+ (circular)	+

a centre of inversion, the vibrations active in IR are inactive in Raman spectra and *vice versa*.

When do the conclusions drawn from group theory fail?

When deriving the selection rules the following assumptions have been made:

- the molecule is *isolated*,
- elements are represented by *the same isotope*,
- the molecule is in a *stationary state*,
- the vibrations have *small amplitudes*,
- the vibrations are *harmonic*,
- the electromagnetic field interacts with the molecule only through the *electric* field-molecule interaction,
- in the interaction of the molecule with the electromagnetic field only what are called the *dipole transitions* are involved.[47]

However, in practice the molecule is never isolated. In particular, the interactions it undergoes in the liquid or solid state are sufficiently strong to deform the molecule. As a result, we have to do (especially in a liquid) with a population of molecules, each in a different geometry, usually devoid of *any particular symmetry* (for a single molecule this means a non-stationary state), although the molecule is not far away from perfect symmetry ("broken symmetry").

Suppose for a while that the molecule under consideration *is* indeed isolated. In a substance we usually have several isotopomers, with different distributions of isotopes in the molecules. In most cases this also means broken symmetry. *Broken symmetry means that the selection rules are violated.*

> In practice, broken symmetry means that the selection rules cause only a small intensity of forbidden transitions with respect to allowed transitions.

When considering electronic transitions, we assumed that the molecule stays in its equilibrium geometry, often of high symmetry. This may be the most probable configuration,[48] but the vibrations and rotations deform it. An electronic excitation is fast, and usually takes place with a molecular geometry that differs slightly from the most probable and most symmetric one. This will cause a transition, forbidden for perfectly symmetric geometry, to have a non-negligible intensity.

Deriving the selection rules for IR and Raman spectra, we assumed that the equivalent atoms can differ only by the *sign* of the deviation from the equilibrium position, but its absolute value is the same. This is how it would be for a harmonic

---

[47] That is, the electric field of the electromagnetic wave within the molecule is assumed to be *homogeneous*. Then the only term in the Hamiltonian related to the light-molecule interaction is $-\hat{\boldsymbol{\mu}}\boldsymbol{\mathcal{E}}$, where $\hat{\boldsymbol{\mu}}$ stands for the dipole moment operator of the molecule and $\boldsymbol{\mathcal{E}}$ is the electric field intensity.

[48] The maximum of the ground-state probability density for the *harmonic* oscillator indeed corresponds to the equilibrium geometry. This is why the selection rules work at all (although in an approximate way).

oscillator. An anharmonicity therefore introduces another reason why a (harmonically) forbidden transition will have non-negligible intensity.

The electromagnetic field has an electric and magnetic component. The selection rules we have derived have not taken into account the presence of the magnetic field. Taking into account the magnetic field introduces some additional selection rules. Also, the wavelength of an electromagnetic wave in the UV-VIS region is of the order of thousands of Å, whereas the length of the molecule is usually of the order of a few Å. This means that the assumption that the electric field of the electromagnetic wave is homogeneous looks good, but in any case the field is not perfectly homogeneous. The deviations will be small, but non-zero. Taking this into account by including further terms besides $-\hat{\boldsymbol{\mu}}\boldsymbol{\mathcal{E}}$, we obtain the interaction of the electric field gradient with the quadrupole moment of the molecule, as well as further terms. This also weakens the selection rules.

Despite these complications, group theory allows for understanding the basic features of molecular spectra. It sometimes works even if the molecule being studied has no symmetry at all, because of a substituent that interferes. Some electronic or vibrational excitations are of a local spatial character and pertain to a portion of the molecule that is (nearly) symmetric. Due to this some optical transitions that are absolutely allowed, because the molecule as a whole does not have any symmetry,[49] *will still have a very low intensity.*

---

[49] But they would be forbidden if the portion in question represented a separate molecule and were symmetric.

D. A TWO-STATE MODEL

The Schrödinger equation $\hat{H}\psi = E\psi$ is usually solved by expanding the unknown wave function ψ in a series[1] of complete basis set $\{\phi_i\}_{i=1}^{N}$ of states ϕ_i, where N in principle equals ∞ (instead in practice we end up with a chosen large value of N). The expansion gives $\hat{H}\sum_j c_j\phi_j = E\sum_j c_j\phi_j$, or $\sum_j c_j(\hat{H}\phi_j - E\phi_j) = 0$. By multiplying this equation successively by ϕ_i^*, $i = 1, 2, \ldots, N$, and integrating we obtain a set of N linear equations for the unknown coefficients[2] c_i:

$$\sum_j c_j(H_{ij} - ES_{ij}) = 0,$$

where the Hamiltonian matrix elements $H_{ij} \equiv \langle\phi_i|\hat{H}\phi_j\rangle$, and the overlap integrals $S_{ij} \equiv \langle\phi_i|\phi_j\rangle$. The summation going to infinity makes impossible any simple insight into the physics of the problem. However, in many cases what matters most are only *two* states of comparable energies, while other states, being far away on the energy scale, do not count in practice (have negligible c_j). If indeed only two states were in the game (the two-state model), the situation could be analyzed in detail. The conclusions drawn are of great conceptual (and smaller numerical) importance.

For the sake of simplicity, in further analysis the functions ϕ_j will be assumed normalized and real.[3] Then, for $N = 2$ we have $H_{12} = \langle\phi_1|\hat{H}\phi_2\rangle = \langle\hat{H}\phi_1|\phi_2\rangle = \langle\phi_2|\hat{H}\phi_1\rangle = H_{21}$, and all H_{ij} are real numbers (in most practical applications $H_{12}, H_{11}, H_{22} \leqslant 0$). The overlap integral will be denoted by $S \equiv \langle\phi_1|\phi_2\rangle = \langle\phi_2|\phi_1\rangle$. After introducing the abbreviation $h \equiv H_{12}$ we have

$$c_1(H_{11} - E) + c_2(h - ES) = 0,$$
$$c_1(h - E) + c_2(H_{22} - ES) = 0.$$

A non-trivial solution of these secular equations exists only if the secular determinant satisfies

$$\begin{vmatrix} H_{11} - E & h - ES \\ h - ES & H_{22} - E \end{vmatrix} = 0.$$

---

[1] As a few examples just recall the CI, VB (Chapter 10), and MO (Chapter 8) methods.

[2] The same set of equations ("secular equations") is obtained after using the Ritz method (Chapter 5).

[3] This pertains to almost all applications. For complex functions the equations are only slightly more complicated.

After expanding the determinant, we obtain a quadratic equation for the unknown energy E:

$$(H_{11}-E)(H_{22}-E)-(h-ES)^2=0$$

with its two solutions[4]

$$E_\pm=\frac{1}{1-S^2}\left\{\frac{H_{11}+H_{22}}{2}-hS\right.$$
$$\left.\mp\sqrt{\left(\frac{H_{11}-H_{22}}{2}\right)^2+\left(h-S\sqrt{H_{11}H_{22}}\right)^2+2hS\left(\sqrt{H_{11}H_{22}}-\frac{H_{11}+H_{22}}{2}\right)}\right\}.$$

After inserting the above energies into the secular equations we obtain the following two sets of solutions c_1 and c_2:

$$\left(\frac{c_1}{c_2}\right)_\pm=\frac{1}{(h-H_{11}S)}\left\{\frac{H_{11}-H_{22}}{2}\pm\sqrt{\left(\frac{H_{11}-H_{22}}{2}\right)^2+(h-H_{11}S)(h-H_{22}S)}\right\}.$$

Using the abbreviations

$$\Delta=\frac{H_{11}-H_{22}}{2},$$

and $E_{ar}=(H_{11}+H_{22})/2$ for the arithmetic mean, as well as $E_{geom}=\sqrt{H_{11}H_{22}}$ for the geometric mean, we get a simpler formula for the energy

$$E_\pm=\frac{1}{1-S^2}\left\{E_{ar}-hS\mp\sqrt{\Delta^2+(h-SE_{geom})^2+2hS(E_{geom}-E_{ar})}\right\}.$$

Now, let us consider some important special cases.

Case I. $H_{11}=H_{22}$ and $S=0$ (ϕ_1 and ϕ_2 correspond to the same energy and do not overlap).

Then, $\Delta=0$, $E_{ar}=E_{geom}=H_{11}$ and we have

$$E_\pm=H_{11}\pm h,\qquad \left(\frac{c_1}{c_2}\right)_\pm=\pm 1.$$

For $h<0$ this means that E_+ corresponds to stabilization (with respect to ϕ_1 or ϕ_2 states), while E_- corresponds to destabilization (by the same value of $|h|$). The wave functions contain equal contributions of ϕ_1 and ϕ_2 and (after normalization) are

---

[4] It is most practical to use Mathematica coding:

```
Solve[(H11-EdS)*(H22-EdS)-(h-EdS*S)^2==0,EdS]
Solve[(H11-EdS)*(H22-EdS)-(h-EdS*S)^2==0
&&c1*(H11-EdS)+c2*(h-EdS*S)==0
&&c1*(h-EdS*S)+c2*(H22-EdS)==0,{c1,c2},{EdS}]
```

$$\psi_+ = \frac{1}{\sqrt{2}}(\phi_1 + \phi_2), \qquad \psi_- = \frac{1}{\sqrt{2}}(\phi_1 - \phi_2).$$

Case II. $H_{11} = H_{22}$ and $S \neq 0$ (ϕ_1 and ϕ_2 correspond to the same energy, but their overlap integral is non-zero).

Then,

$$E_\pm = \frac{H_{11} \pm h}{1 \pm S}, \qquad \left(\frac{c_1}{c_2}\right)_\pm = \pm 1.$$

Here also E_+ corresponds to stabilization, and E_- to destabilization (because of the denominator, this time the destabilization is larger than the stabilization). The wave functions have the same contributions of ϕ_1 and ϕ_2 and (after normalization) are equal to

$$\psi_+ = \frac{1}{\sqrt{2(1+S)}}(\phi_1 + \phi_2), \qquad \psi_- = \frac{1}{\sqrt{2(1-S)}}(\phi_1 - \phi_2).$$

Case III. $H_{11} \neq H_{22}$ and $S = 0$ (ϕ_1 and ϕ_2 correspond to different energies and the overlap integral equals zero).

This time

$$E_\pm = E_{\text{ar}} \mp \sqrt{\Delta^2 + h^2},$$

$$\left(\frac{c_1}{c_2}\right)_\pm = \frac{1}{h}\left(\Delta \pm \sqrt{\Delta^2 + h^2}\right). \tag{D.1}$$

Here also the state of E_+ means stabilization, while E_- corresponds to destabilization (both effects are equal).

Let us consider a limiting case when the mean energy in state ϕ_1 is much lower than that in ϕ_2 ($H_{11} \ll H_{22}$), and in addition $\frac{\Delta}{h} \gg 0$. For the state with energy E_+ we have $\frac{c_1}{c_2} \simeq \frac{2\Delta}{h}$, i.e. c_1 is very large, while c_2 is very small (this means that ψ_+ is very similar to ϕ_1). In state ψ_- the same ratio of coefficients equals $\frac{c_1}{c_2} \simeq 0$, which means a domination of ϕ_2.

Thus, if two states differ very much in their energies (or h is small, which means the overlap integral is also small), they do not change in practice (do not mix together).

This is why at the beginning of this appendix, we admitted only ϕ_1 and ϕ_2 of comparable energies.

E. DIRAC DELTA FUNCTION

Paul Dirac introduced some useful formal tools (for example his notation for integrals and operators, p. 19) including an object then unknown to mathematicians, which turned out to be very useful in physics. This is called the Dirac delta function $\delta(x)$. We may think of it as of a function[1]

- which is non-zero only very close to $x = 0$, where its value is $+\infty$;
- the surface under its plot is equal to 1, which is highlighted by a *symbolic* equation

$$\int_{-\infty}^{\infty} \delta(x)\, \mathrm{d}x = 1.$$

When we look at a straight thin reed protruding from a lake (the water level $=$ 0), then we have to do with something similar to the Dirac delta function. The only importance of the Dirac delta function lies in its specific behaviour, when integrating the product $f(x)\delta(x)$ and the integration *includes the point* $x = 0$, namely:

$$\int_a^b f(x)\delta(x)\, \mathrm{d}x = f(0). \tag{E.1}$$

This result is well understandable: the integral means the surface under the curve $f(x)\delta(x)$, but since $\delta(x)$ is so concentrated at $x = 0$, then it pays to take seriously only those $f(x)$ that are "extremely close" to $x = 0$. Over there $f(x)$ is equal to $f(0)$. The constant $f(0)$ can be taken out of the integral, which itself therefore has the form $\int_{-\infty}^{\infty} \delta(x)\, \mathrm{d}x = 1$. This is why we get the right hand side of the previous equation. Of course, $\delta(x - c)$ represents the same narrow peak, but at $x = c$, therefore, for $a \leqslant c \leqslant b$ we have

$$\boxed{\int_a^b f(x)\delta(x - c)\, \mathrm{d}x = f(c).} \tag{E.2}$$

1 APPROXIMATIONS TO $\delta(x)$

The Dirac delta function $\delta(x)$ can be approximated by many functions, which depend on a certain parameter and have the following properties:

---

[1] More precisely this is not a function, but what is called a distribution. The theory of distributions was developed by mathematicians only after Dirac.

- when the parameter tends to a limit, the values of the functions for x distant from 0 become smaller and smaller, while for x close to zero they get larger and larger (a peak close to $x = 0$);
- the integral of the function tends to (or is close to) 1 when the parameter approaches its limit value.

Here are several functions that approximate the Dirac delta function:

- a rectangular function centred at $x = 0$ with the surface of the rectangle equal to 1 ($a \to 0$):

$$f_1(x; a) = \begin{cases} \dfrac{1}{a} & \text{for } -\dfrac{a}{2} \leqslant x \leqslant \dfrac{a}{2}, \\ 0 & \text{for other;} \end{cases}$$

- a (normalized to 1) Gaussian function[2] ($a \to \infty$):

$$f_2(x; a) = \sqrt{\frac{a}{\pi}} e^{-ax^2};$$

- a function:

$$f_3(x; a) = \frac{1}{\pi} \lim \frac{\sin ax}{x} \quad \text{when } a \to \infty;$$

- the last function is (we will use this when considering the interaction of matter with radiation):[3]

$$f_4(x; a) = \frac{1}{\pi a} \lim \frac{\sin^2(ax)}{x^2} \quad \text{when } a \to \infty.$$

---

[2] Let us see how an approximation $f_2 = \sqrt{\frac{a}{\pi}} e^{-ax^2}$ does the job of the Dirac delta function when $a \to \infty$. Let us take a function $f(x) = (x-5)^2$ and consider the integral

$$\int_{-\infty}^{\infty} f(x) f_2(x)\, dx = \sqrt{\frac{a}{\pi}} \int_{-\infty}^{\infty} (x-5)^2 e^{-ax^2}\, dx = \sqrt{\frac{a}{\pi}} \left(\frac{1}{4a} \sqrt{\frac{\pi}{a}} + 0 + 25 \sqrt{\frac{\pi}{a}} \right) = \frac{1}{4a} + 25.$$

When $a \to \infty$, the value of the integral tends to $25 = f(0)$, as it has to be when the Dirac delta function is used instead of f_2.

[3] The function under the limit symbol may be treated as $A[\sin(ax)]^2$ with amplitude A *decaying* as $A = 1/x^2$, when $|x| \to \infty$. For small values of x, the $\sin(ax)$ changes as ax (as seen from its Taylor expansion), hence for small x the function changes as a^2. This means that when $a \to \infty$, there will be a dominant peak close to $x = 0$, although there will be some smaller side-peaks clustering around $x = 0$. The surface of the dominating peak may be approximated by a triangle of base $2\pi/a$ and height a^2, and we obtain its surface equal to πa, hence the "approximate normalization factor" $1/(\pi a)$ in f_4.

2 PROPERTIES OF $\delta(x)$

Function $\delta(cx)$

Let us see what $\delta(cx)$ is equal to:

$$\delta(cx) = \lim_{a\to\infty}\sqrt{\frac{a}{\pi}}\exp\left(-ac^2x^2\right) = \lim_{a\to\infty}\sqrt{\frac{ac^2}{\pi c^2}}\exp\left(-ac^2x^2\right)$$

$$= \frac{1}{|c|}\lim_{ac^2\to\infty}\sqrt{\frac{ac^2}{\pi}}\exp\left(-ac^2x^2\right) = \frac{1}{|c|}\delta(x).$$

Therefore,

$$\delta(cx) = \frac{1}{|c|}\delta(x). \tag{E.3}$$

Dirac δ in 3D

The 3D Dirac delta function is defined in the Cartesian coordinate system as

$$\delta(\boldsymbol{r}) = \delta(x)\delta(y)\delta(z),$$

where $\boldsymbol{r} = (x, y, z)$. Then, $\delta(\boldsymbol{r})$ denotes a peak of infinite height at $\boldsymbol{r} = \boldsymbol{0}$, and $\delta(\boldsymbol{r} - \boldsymbol{A})$ denotes an identical peak at the position shown by the vector $\boldsymbol{A}$ from the origin. Each of the peaks is normalized to 1, i.e. the integral over the whole 3D space is equal to 1. This means that the formula (E.1) is satisfied, but this time $x \in R^3$.

3 AN APPLICATION OF THE DIRAC DELTA FUNCTION

When may such a concept as the Dirac delta function be useful? Here is an example. Let us imagine that we have (in 3D space) two molecular charge distributions: $\rho_A(\boldsymbol{r})$ and $\rho_B(\boldsymbol{r})$. Each of the distributions consists of an electronic part and an nuclear part.

How can such charge distributions be represented mathematically? There is no problem in mathematical representation of the electronic parts, they are simply some functions of the position $\boldsymbol{r}$ in space: $-\rho_{\mathrm{el},A}(\boldsymbol{r})$ and $-\rho_{\mathrm{el},B}(\boldsymbol{r})$ for each molecule. The integrals of the corresponding electronic distributions yield, of course, $-N_A$ and $-N_B$ (in a.u.), or minus the number of the electrons (minus, because the electrons carry negative charge). How do we write the nuclear charge distribution as a function of $\boldsymbol{r}$? There is no way to do it without the Dirac delta function. With the function our task is simple:

$$\rho_{\mathrm{nucl},A}(\boldsymbol{r}) = \sum_{a\in A} Z_{A,a}\delta(\boldsymbol{r} - \boldsymbol{r}_a),$$

$$\rho_{\text{nucl},B}(\boldsymbol{r}) = \sum_{b\in B} Z_{B,b}\delta(\boldsymbol{r}-\boldsymbol{r}_b).$$

We put delta functions with the "intensities" equal to the nuclear charges in the nuclear positions . For neutral molecules $\int \rho_{\text{nucl},A}(\boldsymbol{r})\,\mathrm{d}^3\boldsymbol{r}$ and $\int \rho_{\text{nucl},B}(\boldsymbol{r})\,\mathrm{d}^3\boldsymbol{r}$ have to give $+N_A$ and $+N_B$, respectively. Indeed, we have

$$\int \rho_{\text{nucl},A}(\boldsymbol{r})\,\mathrm{d}^3\boldsymbol{r} = \sum_{a\in A} Z_{A,a}\int \delta(\boldsymbol{r}-\boldsymbol{r}_a)\,\mathrm{d}^3\boldsymbol{r} = \sum_{a\in A} Z_{A,a} = N_A,$$
$$\int \rho_{\text{nucl},B}(\boldsymbol{r})\,\mathrm{d}^3\boldsymbol{r} = \sum_{b\in B} Z_{B,b}\int \delta(\boldsymbol{r}-\boldsymbol{r}_b)\,\mathrm{d}^3\boldsymbol{r} = \sum_{b\in B} Z_{B,b} = N_B.$$

Thus the Dirac delta function enables us to write the total charge distributions and their interactions in an elegant way:

$$\rho_A(\boldsymbol{r}) = -\rho_{\text{el},A}(\boldsymbol{r}) + \rho_{\text{nucl},A}(\boldsymbol{r}),$$
$$\rho_B(\boldsymbol{r}) = -\rho_{\text{el},B}(\boldsymbol{r}) + \rho_{\text{nucl},B}(\boldsymbol{r}).$$

To demonstrate the difference, let us write the electrostatic interaction of the two charge distributions with and without the Dirac delta functions. The first gives the following expression

$$E_{\text{inter}} = \sum_{a\in A}\sum_{b\in B}\frac{Z_{A,a}Z_{B,b}}{|\boldsymbol{r}_a-\boldsymbol{r}_b|} - \sum_{a\in A}\int \rho_{\text{el},B}(\boldsymbol{r})\frac{Z_{A,a}}{|\boldsymbol{r}-\boldsymbol{r}_a|}\,\mathrm{d}^3\boldsymbol{r}$$
$$-\sum_{b\in B}\int \rho_{\text{el},A}(\boldsymbol{r})\frac{Z_{B,b}}{|\boldsymbol{r}-\boldsymbol{r}_b|}\,\mathrm{d}^3\boldsymbol{r} + \iint \frac{\rho_{\text{el},A}(\boldsymbol{r})\rho_{\text{el},B}(\boldsymbol{r}')}{|\boldsymbol{r}-\boldsymbol{r}'|}\,\mathrm{d}^3\boldsymbol{r}\,\mathrm{d}^3\boldsymbol{r}'.$$

The four terms mean the following interactions respectively: nuclei of A – nuclei of B, nuclei of A – electrons of B, electrons of A – nuclei of B, electrons of A – electrons of B. With the Dirac delta function the same expression reads:

$$E_{\text{inter}} = \int \frac{\rho_A(\boldsymbol{r})\rho_B(\boldsymbol{r}')}{|\boldsymbol{r}-\boldsymbol{r}'|}\,\mathrm{d}^3\boldsymbol{r}\,\mathrm{d}^3\boldsymbol{r}'.$$

The last expression comes from the definition of the Coulomb interaction and the definition of the integral.[4]

No matter how the charge distributions looks, whether they are diffuse (the electronic ones) or point-like (those of the nuclei), the formula is always the same.

---

[4] Of course, the two notations are equivalent, because inserting the total charge distributions into the last integral as well as using the properties of the Dirac delta function, gives the first expression for E_{inter}.

F. TRANSLATION VS MOMENTUM AND ROTATION VS ANGULAR MOMENTUM

It was shown in Chapter 2 that the Hamiltonian $\hat{H}$ commutes with any translation (p. 61) or rotation (p. 63) operator, denoted as $\hat{\mathcal{U}}$:

$$[\hat{H}, \hat{\mathcal{U}}] = 0. \tag{F.1}$$

1 THE FORM OF THE $\hat{\mathcal{U}}$ OPERATOR

Below it will be demonstrated for $\boldsymbol{\kappa}$, meaning first a translation vector, and then a rotation angle about an axis in 3D space, that operator $\hat{\mathcal{U}}$ is of the form

$$\hat{\mathcal{U}} = \exp\left(-\frac{i}{\hbar}\boldsymbol{\kappa} \cdot \hat{\boldsymbol{K}}\right), \tag{F.2}$$

where $\hat{\boldsymbol{K}}$ stands for a Hermitian operator (with x, y, z components) acting on functions of points in 3D Cartesian space.

Translation and momentum operators

Translation of a function by a vector $\Delta\boldsymbol{r}$ just represents function f in the coordinate system translated in the opposite direction, i.e. $f(\boldsymbol{r} - \Delta\boldsymbol{r})$, see Fig. 2.3 and p. 62. If vector $\Delta\boldsymbol{r}$ is *infinitesimally small, then, in order to establish the relation between* $f(\boldsymbol{r} - \Delta\boldsymbol{r})$ *and* $f(\boldsymbol{r})$, *it is of course sufficient to know the gradient of* f (neglecting, obviously, the quadratic and higher terms in the Taylor expansion):

$$f(\boldsymbol{r} - \Delta\boldsymbol{r}) = f(\boldsymbol{r}) - \Delta\boldsymbol{r} \cdot \nabla f = (1 - \Delta\boldsymbol{r} \cdot \nabla) f(\boldsymbol{r}). \tag{F.3}$$

We will compose a large translation of a function (by vector $\boldsymbol{T}$) from a number of small increments $\Delta\boldsymbol{r} = \frac{1}{N}\boldsymbol{T}$, where N is a veeery large natural number. Such a tiny translation will be repeated N times, thus recovering the translation of the function by $\boldsymbol{T}$. In order for the gradient formula to be exact, we have to ensure N tending to infinity. Recalling the definition $\exp(ax) = \lim_{N\to\infty}(1 + \frac{a}{x})^N$, we have:

$$\begin{aligned} \hat{\mathcal{U}}(\boldsymbol{T}) f(\boldsymbol{r}) = f(\boldsymbol{r} - \boldsymbol{T}) &= \lim_{N\to\infty}\left(1 - \frac{\boldsymbol{T}}{N}\nabla\right)^N f(r) \\ &= \exp(-\boldsymbol{T}\cdot\nabla) f = \exp\left(-\frac{i}{\hbar}\boldsymbol{T} \cdot \hat{\boldsymbol{p}}\right) f(\boldsymbol{r}), \end{aligned}$$

where $\hat{\boldsymbol{p}} = -i\hbar\nabla$ is the total momentum operator (see Chapter 1). Thus, for translations we have $\boldsymbol{\kappa} \equiv \boldsymbol{T}$ and $\hat{\boldsymbol{K}} \equiv \hat{\boldsymbol{p}}$.

Rotation and angular momentum operator

Imagine a function $f(\boldsymbol{r})$ of positions in 3D Cartesian space (think, e.g., about a probability density distribution centred somewhere in space). Now suppose the function is to be rotated about the z axis (the unit vector showing its direction is $\boldsymbol{e}$) by an angle α, so we have another function, let us denote it by $\hat{\mathcal{U}}(\alpha; \boldsymbol{e})f(\boldsymbol{r})$. What is the relation between $f(\boldsymbol{r})$ and $\hat{\mathcal{U}}(\alpha; \boldsymbol{e})f(\boldsymbol{r})$? This is what we want to establish. This relation corresponds to the opposite rotation (i.e. by the angle $-\alpha$, see Fig. 2.1 and p. 58) of the coordinate system:

$$\hat{\mathcal{U}}(\alpha; \boldsymbol{e})f(\boldsymbol{r}) = f\big(\boldsymbol{U}^{-1}\boldsymbol{r}\big) = f\big(\boldsymbol{U}(-\alpha; \boldsymbol{e})\boldsymbol{r}\big),$$

where $\boldsymbol{U}$ is a 3×3 orthogonal matrix. The new coordinates $x(\alpha), y(\alpha), z(\alpha)$ are expressed by the old coordinates x, y, z through[1]

$$\boldsymbol{r}' \equiv \begin{pmatrix} x(\alpha) \\ y(\alpha) \\ z(\alpha) \end{pmatrix} = \begin{pmatrix} \cos\alpha & \sin\alpha & 0 \\ -\sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}.$$

Therefore the rotated function $\hat{\mathcal{U}}(\alpha; \boldsymbol{e})f(\boldsymbol{r}) = f(x(\alpha), y(\alpha), z(\alpha))$. The function can be expanded in the Taylor series about $\alpha = 0$:

$$\begin{aligned} \hat{\mathcal{U}}(\alpha; \boldsymbol{e})f(\boldsymbol{r}) &= f\big(x(\alpha), y(\alpha), z(\alpha)\big) = f(x, y, z) + \alpha\left(\frac{\partial f}{\partial \alpha}\right)_{\alpha=0} + \cdots \\ &= f(x, y, z) + \alpha\left(\frac{\partial x(\alpha)}{\partial \alpha}\frac{\partial f}{\partial x} + \frac{\partial y(\alpha)}{\partial \alpha}\frac{\partial f}{\partial y} + \frac{\partial z(\alpha)}{\partial \alpha}\frac{\partial f}{\partial z}\right)_{\alpha=0} + \cdots \\ &= f(x, y, z) + \alpha\left[y\frac{\partial}{\partial x} - x\frac{\partial}{\partial y}\right]f + \cdots \end{aligned}$$

Now instead of the large rotation angle α, let us consider first an infinitesimally small rotation by angle $\varepsilon = \alpha/N$, where N is a huge natural number. In such a situation we retain only the first two terms in the previous equation:

$$\begin{aligned} \hat{\mathcal{U}}\left(\frac{\alpha}{N}; \boldsymbol{e}\right)f(\boldsymbol{r}) &= f(x, y, z) + \frac{\alpha}{N}\left[y\frac{\partial}{\partial x} - x\frac{\partial}{\partial y}\right]f(x, y, z) \\ &= \left(1 + \frac{\alpha}{N}\frac{i\hbar}{i\hbar}\left[y\frac{\partial}{\partial x} - x\frac{\partial}{\partial y}\right]\right)f = \left(1 + \frac{\alpha}{N}\frac{1}{i\hbar}\left[x\hat{p}_y - y\hat{p}_x\right]\right)f \\ &= \left(1 - \frac{\alpha}{N}\frac{i}{\hbar}\hat{J}_z\right)f. \end{aligned}$$

---

[1] A positive value of the rotation angle means an anticlockwise motion within the xy plane (x axis horizontal, y vertical, z axis pointing to us).

If such a rotation is repeated N times, we recover the rotation of the function by a (possibly large) angle α (the limit assures that ε is infinitesimally small):

$$\begin{aligned}\hat{\mathcal{U}}(\alpha;\boldsymbol{e})f(\boldsymbol{r}) &= \lim_{N\to\infty}\left[\hat{\mathcal{U}}\left(\frac{\alpha}{N};\boldsymbol{e}\right)\right]^N f(\boldsymbol{r}) = \lim_{N\to\infty}\left(1-\frac{\alpha}{N}\frac{i}{\hbar}\hat{J}_z\right)^N f(\boldsymbol{r})\\ &= \exp\left(-i\frac{\alpha}{\hbar}\hat{J}_z\right)f = \exp\left(-\frac{i}{\hbar}\alpha\boldsymbol{e}\cdot\hat{\boldsymbol{J}}\right)f.\end{aligned}$$

Thus for rotations $\hat{\mathcal{U}}(\alpha;\boldsymbol{e}) = \exp(-\frac{i}{\hbar}\alpha\boldsymbol{e}\cdot\hat{\boldsymbol{J}})$, and, therefore, we have $\boldsymbol{\kappa}\equiv\alpha\boldsymbol{e}$ and $\hat{\boldsymbol{K}}\equiv\hat{\boldsymbol{J}}$.

This means that, in particular for rotations about the x, y, z axes (with the corresponding unit vectors $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}$) we have, respectively

$$\left[\hat{\mathcal{U}}(\alpha;\boldsymbol{x}), \hat{J}_x\right] = 0, \tag{F.4}$$

$$\left[\hat{\mathcal{U}}(\alpha;\boldsymbol{y}), \hat{J}_y\right] = 0, \tag{F.5}$$

$$\left[\hat{\mathcal{U}}(\alpha;\boldsymbol{z}), \hat{J}_z\right] = 0. \tag{F.6}$$

Useful relation

The relation (F.1) means that for any translation or rotation

$$\hat{\mathcal{U}}\hat{H}\hat{\mathcal{U}}^{-1} = \hat{H}$$

and taking into account the general form of eq. (F.2) we have for any such transformation a series containing nested commutators (valid for any $\boldsymbol{\kappa}$)

$$\begin{aligned}\hat{H} = \hat{\mathcal{U}}\hat{H}\hat{\mathcal{U}}^{-1} &= \exp\left(-\frac{i}{\hbar}\boldsymbol{\kappa}\cdot\hat{\boldsymbol{K}}\right)\hat{H}\exp\left(\frac{i}{\hbar}\boldsymbol{\kappa}\cdot\hat{\boldsymbol{K}}\right)\\ &= \left(1-\frac{i}{\hbar}\boldsymbol{\kappa}\cdot\hat{\boldsymbol{K}}+\cdots\right)\hat{H}\left(1+\frac{i}{\hbar}\boldsymbol{\kappa}\cdot\hat{\boldsymbol{K}}+\cdots\right)\\ &= \hat{H} - \frac{i}{\hbar}\boldsymbol{\kappa}\cdot[\hat{\boldsymbol{K}},\hat{H}] - \frac{\kappa^2}{2\hbar^2}[[\hat{\boldsymbol{K}},\hat{H}],\hat{\boldsymbol{K}}] + \cdots,\end{aligned}$$

where each term in "$+\cdots$" contains $[\hat{\boldsymbol{K}},\hat{H}]$. This means that to satisfy the equation we necessarily have

$$\left[\hat{\boldsymbol{K}},\hat{H}\right] = \boldsymbol{0}. \tag{F.7}$$

2 THE HAMILTONIAN COMMUTES WITH THE TOTAL MOMENTUM OPERATOR

In particular this means $[\hat{\boldsymbol{p}},\hat{H}] = \boldsymbol{0}$, i.e.

$$\left[\hat{p}_\mu,\hat{H}\right] = 0 \tag{F.8}$$

for $\mu = x, y, z$. Of course, we also have $[\hat{p}_\mu, \hat{p}_\nu] = 0$ for $\mu, \nu = x, y, z$.

Since all these four operators mutually commute, the total wave function is simultaneously an eigenfunction of $\hat{H}$ and $\hat{p}_x$, $\hat{p}_y$, $\hat{p}_z$, i.e. the energy and the momentum of the centre of mass can both be measured (without making any error) in a space-fixed coordinate system (see Appendix I). From the definition, the momentum of the centre of mass is identical to the total momentum.[2]

3 THE HAMILTONIAN, $\hat{J}^2$ AND $\hat{J}_z$ DO COMMUTE

Eq. (F.7) for rotations means $[\hat{\boldsymbol{J}}, \hat{H}] = \boldsymbol{0}$, i.e. in particular

$$[\hat{J}_x, \hat{H}] = 0, \tag{F.9}$$

$$[\hat{J}_y, \hat{H}] = 0, \tag{F.10}$$

$$[\hat{J}_z, \hat{H}] = 0. \tag{F.11}$$

The components of the angular momentum operators satisfy the following commutation rules: [3]

$$\begin{aligned} [\hat{J}_x, \hat{J}_y] &= i\hbar \hat{J}_z, \\ [\hat{J}_y, \hat{J}_z] &= i\hbar \hat{J}_x, \\ [\hat{J}_z, \hat{J}_x] &= i\hbar \hat{J}_y. \end{aligned} \tag{F.12}$$

---

[2] Indeed the position vector of the centre of mass is defined as $\boldsymbol{R}_{\mathrm{CM}} = \frac{\sum_i m_i \boldsymbol{r}_i}{\sum_i m_i}$, and after differentiation with respect to time $(\sum_i m_i)\dot{\boldsymbol{R}}_{\mathrm{CM}} = \sum_i m_i \dot{\boldsymbol{r}}_i = \sum_i \boldsymbol{p}_i$. The right-hand side represents the momentum of all the particles (i.e. the total momentum), whereas the left is simply the momentum of the centre of mass.

[3] The commutation relations can be obtained by using the definitions of the operators involved directly: $\hat{J}_x = y\hat{p}_z - z\hat{p}_y$, etc. For example,

$$\begin{aligned} [\hat{J}_x, \hat{J}_y]f &= [(y\hat{p}_z - z\hat{p}_y)(z\hat{p}_x - x\hat{p}_z) - (z\hat{p}_x - x\hat{p}_z)(y\hat{p}_z - z\hat{p}_y)]f \\ &= [(y\hat{p}_z z\hat{p}_x - z\hat{p}_x y\hat{p}_z) - (y\hat{p}_z x\hat{p}_z - x\hat{p}_z y\hat{p}_z) \\ &\quad - (z\hat{p}_y z\hat{p}_x - z\hat{p}_x z\hat{p}_y) + (z\hat{p}_y x\hat{p}_z - x\hat{p}_z z\hat{p}_y)]f \\ &= (y\hat{p}_z z\hat{p}_x - z\hat{p}_x y\hat{p}_z)f - (yx\hat{p}_z\hat{p}_z - yx\hat{p}_z\hat{p}_z) \\ &\quad - (z^2\hat{p}_y\hat{p}_x - z^2\hat{p}_x\hat{p}_y) + (xz\hat{p}_y\hat{p}_z - x\hat{p}_z z\hat{p}_y) \\ &= (y\hat{p}_z z\hat{p}_x - yz\hat{p}_x\hat{p}_z)f - 0 - 0 + (xz\hat{p}_y\hat{p}_z - x\hat{p}_z z\hat{p}_y)f \\ &= (-i\hbar)^2\left[y\frac{\partial f}{\partial x} - x\frac{\partial f}{\partial y}\right] = i\hbar\hat{J}_z f. \end{aligned}$$

Eqs. (F.9)–(F.11) are not independent, e.g., eq. (F.11) can be derived from eqs. (F.9) and (F.10). Indeed,

$$\begin{aligned}
\big[\hat{J}_z, \hat{H}\big] &= \hat{J}_z\hat{H} - \hat{H}\hat{J}_z \\
&= \frac{1}{i\hbar}[\hat{J}_x, \hat{J}_y]\hat{H} - \frac{1}{i\hbar}\hat{H}[\hat{J}_x, \hat{J}_y] \\
&= \frac{1}{i\hbar}[\hat{J}_x, \hat{J}_y]\hat{H} - \frac{1}{i\hbar}[\hat{J}_x, \hat{J}_y]\hat{H} \\
&= 0.
\end{aligned}$$

Also, from eqs. (F.9), (F.10) and (F.11) it also follows that

$$\big[\hat{J}^2, \hat{H}\big] = 0, \tag{F.13}$$

because from Pythagoras' theorem $\hat{J}^2 = \hat{J}_x^2 + \hat{J}_y^2 + \hat{J}_z^2$.

Do $\hat{J}_x, \hat{J}_y, \hat{J}_z$ commute with $\hat{J}^2$? Let us check the commutator $[\hat{J}_z, \hat{J}^2]$:

$$\begin{aligned}
\big[\hat{J}_z, \hat{J}^2\big] &= \big[\hat{J}_z, \hat{J}_x^2 + \hat{J}_y^2 + \hat{L}_z^2\big] \\
&= \big[\hat{J}_z, \hat{J}_x^2 + \hat{J}_y^2\big] \\
&= \hat{J}_z\hat{J}_x^2 - \hat{J}_x^2\hat{J}_z + \hat{J}_z\hat{J}_y^2 - \hat{J}_y^2\hat{J}_z \\
&= \big(i\hbar\hat{J}_y + \hat{J}_x\hat{J}_z\big)\hat{J}_x - \hat{J}_x\big(-i\hbar\hat{J}_y + \hat{J}_z\hat{J}_x\big) + \big(-i\hbar\hat{J}_x + \hat{J}_y\hat{J}_z\big)\hat{J}_y \\
&\quad - \hat{J}_y\big(i\hbar\hat{J}_x + \hat{J}_z\hat{J}_y\big) \\
&= 0.
\end{aligned}$$

Thus,

$$\big[\hat{J}_z, \hat{J}^2\big] = 0, \tag{F.14}$$

and also by the argument of symmetry (the space is isotropic)

$$\big[\hat{J}_x, \hat{J}^2\big] = 0, \tag{F.15}$$

$$\big[\hat{J}_y, \hat{J}^2\big] = 0. \tag{F.16}$$

Now we need to determine the set of the operators that all mutually commute. Only then can all the physical quantities, to which the operators correspond, have definite values when measured. Also the wave function can be an eigenfunction of all of these operators and it can be labelled by quantum numbers, each corresponding to an eigenvalue of the operators in question. *We cannot choose, for these operators, the whole set of* $\hat{H}, \hat{J}_x, \hat{J}_y, \hat{J}_z, \hat{J}^2$, *because, as was shown above,* $\hat{J}_x, \hat{J}_y, \hat{J}_z$ *do not commute among themselves* (although they do with $\hat{H}$ and $\hat{J}^2$).

The only way is to choose as the set of operators *either* $\hat{H}, \hat{J}_z, \hat{J}^2$ *or* $\hat{H}, \hat{J}_x, \hat{J}^2$ *or* $\hat{H}, \hat{J}_y, \hat{J}^2$. Traditionally, we choose $\hat{H}, \hat{J}_z, \hat{J}^2$ as the set of mutually commuting operators (z is known as the *quantization axis*).

4 ROTATION AND TRANSLATION OPERATORS DO NOT COMMUTE

Now we may think about adding $\hat{p}_x$, $\hat{p}_y$, $\hat{p}_z$, to the above set of operators. The operators $\hat{H}$, $\hat{p}_x$, $\hat{p}_y$, $\hat{p}_z$, $\hat{J}^2$ and $\hat{J}_z$ *do not* represent a set of mutually commuting operators. The reason for this is that $[\hat{p}_\mu, \hat{J}_\nu] \neq 0$ for $\mu \neq \nu$, which is a consequence of the fact that, in general, rotation and translation operators do not commute as shown in Fig. F.1.

5 CONCLUSION

It is, therefore, impossible to make all the operators $\hat{H}$, $\hat{p}_x$, $\hat{p}_y$, $\hat{p}_z$, $\hat{J}^2$ and $\hat{J}_z$ commute in a *space fixed coordinate system*. What we are able to do, though, is to write the total wave function Ψ_{pN} in the space fixed coordinate system as a product of the plane wave $\exp(i\boldsymbol{p}_{\mathrm{CM}} \cdot \boldsymbol{R}_{\mathrm{CM}})$ depending on the centre-of-mass variables and on the wave function Ψ_{0N} depending on internal coordinates[4]

$$\Psi_{pN} = \Psi_{0N} \exp(i\boldsymbol{p}_{\mathrm{CM}} \cdot \boldsymbol{R}_{\mathrm{CM}}), \tag{F.17}$$

which is an eigenfunction of the total (i.e. centre-of-mass) momentum operators:

$$\hat{p}_x = \hat{p}_{CM,x}, \quad \hat{p}_y = \hat{p}_{CM,y}, \quad \hat{p}_z = \hat{p}_{CM,z}.$$

The function Ψ_{0N} is the total wave function written in the centre-of-mass coordinate system (a special body-fixed coordinate system, see Appendix I), in which the total angular momentum operators $\hat{J}^2$ and $\hat{J}_z$ are now defined. The three operators $\hat{H}, \hat{J}^2$ and $\hat{J}_z$ commute in any space-fixed or body-fixed coordinate system (including the centre-of-mass coordinate system), and therefore the corresponding physical quantities (energy and angular momentum) have exact values. In this particular coordinate system: $\hat{\boldsymbol{p}} = \hat{\boldsymbol{p}}_{\mathrm{CM}} = \boldsymbol{0}$. We may say, therefore, that

in the centre-of-mass coordinate system $\hat{H}$, $\hat{p}_x$, $\hat{p}_y$, $\hat{p}_z$, $\hat{J}^2$ and $\hat{J}_z$ all do commute.

---

[4] See Chapter 2 and Appendix I, where the total Hamiltonian is split into the sum of the centre-of-mass and internal coordinate Hamiltonians; N is the quantum number for the spectroscopic states.

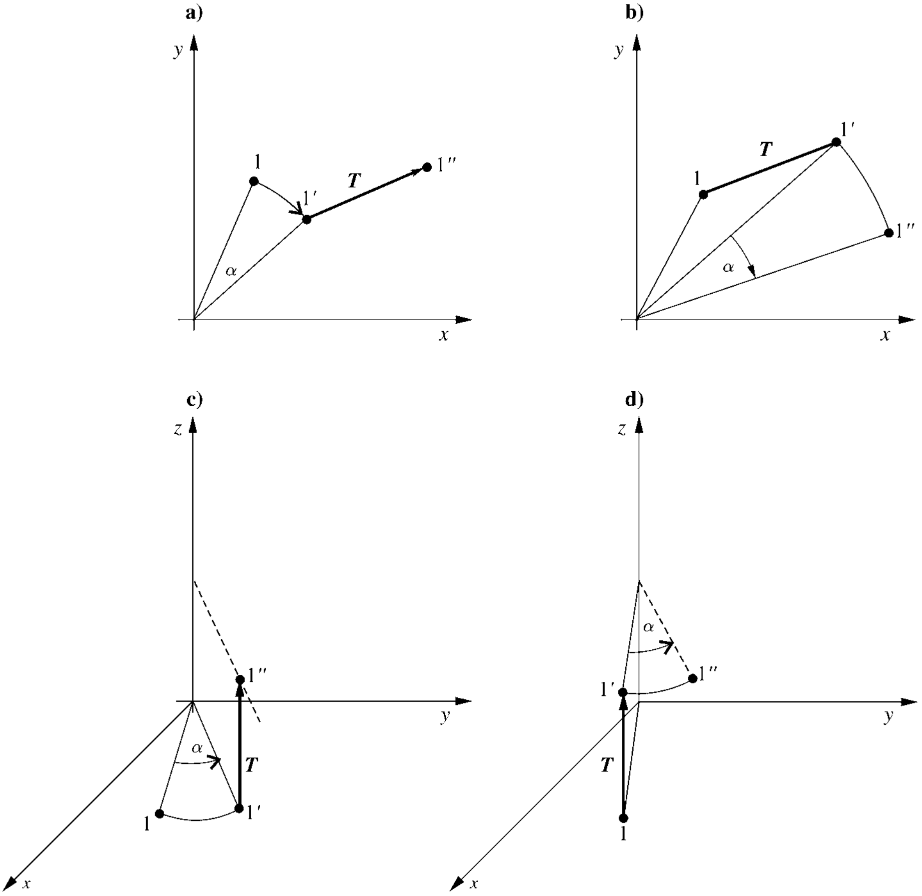


Fig. F.1. In general, translation $\hat{\mathcal{U}}(\boldsymbol{T})$ and rotation $\hat{\mathcal{U}}(\alpha; \boldsymbol{e})$ operators do not commute. The example shows what happens to a point belonging to the xy plane. (a) A rotation $\hat{\mathcal{U}}(\alpha; z)$ by angle α about the z axis takes place first and then a translation $\hat{\mathcal{U}}(\boldsymbol{T})$ by a vector $\boldsymbol{T}$ (restricted to the xy plane) is carried out. (b) The operations are applied in the reverse order. As we can see the results are different (two points $1''$ have different positions in Figs. (a) and (b)), i.e. the two operators do not commute: $\hat{\mathcal{U}}(\boldsymbol{T})\hat{\mathcal{U}}(\alpha; z) \neq \hat{\mathcal{U}}(\alpha; z)\hat{\mathcal{U}}(\boldsymbol{T})$. This after expanding $\hat{\mathcal{U}}(\boldsymbol{T}) = \exp[-\frac{i}{\hbar}(T_x \hat{p}_x + T_y \hat{p}_y)]$ and $\hat{\mathcal{U}}(\alpha; z) = \exp(-\frac{i}{\hbar}\alpha \hat{J}_z)$ in Taylor series, and taking into account that T_x, T_y, α are arbitrary numbers, leads to the conclusion that $[\hat{J}_z, \hat{p}_x] \neq 0$ and $[\hat{J}_z, \hat{p}_y] \neq 0$. Note, that *some* translations and rotations do commute, e.g., $[\hat{J}_z, \hat{p}_z] = [\hat{J}_x, \hat{p}_x] = [\hat{J}_y, \hat{p}_y] = 0$, because we see by inspection (c,d) that any translation by $\boldsymbol{T} = (0, 0, T_z)$ is independent of any rotation about the z axis, etc.

G. VECTOR AND SCALAR POTENTIALS

Maxwell equations

charge density

current

The electromagnetic field is described by two vector fields: the electric field intensity $\boldsymbol{\mathcal{E}}$ and the magnetic field intensity $\boldsymbol{H}$, both depending on position in space (Cartesian coordinates x, y, z) and time t. The vectors $\boldsymbol{\mathcal{E}}$ and $\boldsymbol{H}$ are determined by the electric charges and their currents. The charges are defined by the *charge density* function $\rho(x, y, z, t)$ such that $\rho(x, y, z, t)\,\mathrm{d}V$ at time t represents the charge in the infinitesimal volume $\mathrm{d}V$ that contains the point (x, y, z). The velocity of the charge in position x, y, z measured at time t represents the vector field $\boldsymbol{v}(x, y, z, t)$, while the *current* at point x, y, z measured at t is equal to $\boldsymbol{i}(x, y, z, t) = \rho(x, y, z, t)\boldsymbol{v}(x, y, z, t)$.

It turns out (as shown by James Maxwell), that $\boldsymbol{H}$, $\boldsymbol{\mathcal{E}}$, ρ and $\boldsymbol{i}$ are interrelated by the Maxwell equations (c stands for the speed of light)

$$\nabla \times \boldsymbol{\mathcal{E}} + \frac{1}{c}\frac{\partial \boldsymbol{H}}{\partial t} = \boldsymbol{0}, \tag{G.1}$$

$$\nabla \times \boldsymbol{H} - \frac{1}{c}\frac{\partial \boldsymbol{\mathcal{E}}}{\partial t} = \frac{4\pi}{c}\boldsymbol{i}, \tag{G.2}$$

$$\nabla \cdot \boldsymbol{\mathcal{E}} = 4\pi\rho, \tag{G.3}$$

$$\nabla \cdot \boldsymbol{H} = 0. \tag{G.4}$$

The Maxwell equations have an alternative notation, which involves two new quantities: the *scalar potential* ϕ and the *vector potential* $\boldsymbol{A}$ that replace $\boldsymbol{\mathcal{E}}$ and $\boldsymbol{H}$:

$$\boldsymbol{\mathcal{E}} = -\nabla\phi - \frac{1}{c}\frac{\partial \boldsymbol{A}}{\partial t}, \tag{G.5}$$

$$\boldsymbol{H} = \nabla \times \boldsymbol{A}. \tag{G.6}$$

After inserting $\boldsymbol{\mathcal{E}}$ from eq. (G.5) into eq. (G.1), we obtain its automatic satisfaction:

$$\begin{aligned}\nabla \times \boldsymbol{\mathcal{E}} + \frac{1}{c}\frac{\partial \boldsymbol{H}}{\partial t} &= \nabla \times \left(-\nabla\phi - \frac{1}{c}\frac{\partial \boldsymbol{A}}{\partial t}\right) + \frac{1}{c}\frac{\partial \boldsymbol{H}}{\partial t}\\ &= -\nabla \times \nabla\phi - \frac{1}{c}\frac{\partial \nabla \times \boldsymbol{A}}{\partial t} + \frac{1}{c}\frac{\partial \boldsymbol{H}}{\partial t} = \boldsymbol{0},\end{aligned}$$

because

$$\nabla \times \nabla\phi = \left[\frac{\partial}{\partial y}\left(\frac{\partial\phi}{\partial z}\right) - \frac{\partial}{\partial z}\left(\frac{\partial\phi}{\partial y}\right), \frac{\partial}{\partial z}\left(\frac{\partial\phi}{\partial x}\right) - \frac{\partial}{\partial x}\left(\frac{\partial\phi}{\partial z}\right), \frac{\partial}{\partial x}\left(\frac{\partial\phi}{\partial y}\right) - \frac{\partial}{\partial y}\left(\frac{\partial\phi}{\partial x}\right)\right]$$
$$= [0,0,0] = \mathbf{0} \tag{G.7}$$

and $\nabla \times \boldsymbol{A} = \boldsymbol{H}$.

Eq. (G.4) gives also automatically

$$\nabla \cdot (\nabla \times \boldsymbol{A}) = \frac{\partial}{\partial x}\left(\frac{\partial A_z}{\partial y} - \frac{\partial A_y}{\partial z}\right) + \frac{\partial}{\partial y}\left(\frac{\partial A_x}{\partial z} - \frac{\partial A_z}{\partial x}\right) + \frac{\partial}{\partial z}\left(\frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y}\right) = 0.$$

Eqs. (G.2) and (G.3) transform into

$$\nabla \times (\nabla \times \boldsymbol{A}) + \frac{1}{c}\frac{\partial \nabla\phi}{\partial t} + \frac{1}{c^2}\frac{\partial^2 \boldsymbol{A}}{\partial t^2} = \frac{4\pi}{c}\boldsymbol{i},$$
$$-\nabla \cdot (\nabla\phi) - \frac{1}{c}\frac{\partial \nabla \cdot \boldsymbol{A}}{\partial t} = 4\pi\rho,$$

which in view of the identity $\nabla \times (\nabla \times \boldsymbol{A}) = \nabla(\nabla \cdot \boldsymbol{A}) - \Delta\boldsymbol{A}$ and $\nabla \cdot (\nabla\phi) = \Delta\phi$, gives two additional Maxwell equations (besides eqs. (G.5) and (G.6))

$$\nabla\left(\nabla \cdot \boldsymbol{A} + \frac{1}{c}\frac{\partial\phi}{\partial t}\right) - \Delta\boldsymbol{A} + \frac{1}{c^2}\frac{\partial^2 \boldsymbol{A}}{\partial t^2} = \frac{4\pi}{c}\boldsymbol{i}, \tag{G.8}$$

$$\Delta\phi + \frac{1}{c}\nabla \cdot \frac{\partial \boldsymbol{A}}{\partial t} = -4\pi\rho. \tag{G.9}$$

To characterize the electromagnetic field we may use either $\boldsymbol{\mathcal{E}}$ and $\boldsymbol{H}$ or the two potentials, ϕ and $\boldsymbol{A}$.

Arbitrariness of the potentials ϕ and A

Potentials ϕ and $\boldsymbol{A}$ *are not defined uniquely, i.e. many different potentials lead to the same intensities of electric and magnetic fields*. If we made the following modifications in ϕ and $\boldsymbol{A}$:

$$\phi' = \phi - \frac{1}{c}\frac{\partial f}{\partial t}, \tag{G.10}$$
$$\boldsymbol{A}' = \boldsymbol{A} + \nabla f, \tag{G.11}$$

where f is an *arbitrary differentiable function* (of x, y, z, t), then ϕ' and $\boldsymbol{A}'$ lead to *the same* (see the footnote) $\boldsymbol{\mathcal{E}}$ and $\boldsymbol{H}$:

$$\boldsymbol{\mathcal{E}}' = -\nabla\phi' - \frac{1}{c}\frac{\partial \boldsymbol{A}'}{\partial t} = \left(-\nabla\phi + \frac{1}{c}\nabla\frac{\partial f}{\partial t}\right) - \frac{1}{c}\left(\frac{\partial \boldsymbol{A}}{\partial t} + \frac{\partial}{\partial t}(\nabla f)\right) = \boldsymbol{\mathcal{E}},$$

$$\boldsymbol{H}' = \nabla \times \boldsymbol{A}' = \nabla \times \boldsymbol{A} + \nabla \times \nabla f = \boldsymbol{H}.$$

Choice of potentials A and ϕ for a homogeneous magnetic field

From the second Maxwell equation (G.6), we see that, if the magnetic field $\boldsymbol{H}$ is time-independent, we get the time-independent $\boldsymbol{A}$. Profiting from the non-uniqueness of $\boldsymbol{A}$, we choose it in such a way as to satisfy (what is called the *Coulombic gauge*)[1]

Coulombic gauge

$$\nabla \cdot \boldsymbol{A} = 0, \tag{G.12}$$

which diminishes the arbitrariness, but does not remove it.

Let us take the example of an atom in a homogeneous magnetic field $\boldsymbol{H}$. Let us locate the origin of the coordinate system on the nucleus, the choice being quite natural for an atom, and let us construct the vector potential at position $\boldsymbol{r} = (x, y, z)$ as

$$\boldsymbol{A}(\boldsymbol{r}) = \frac{1}{2}[\boldsymbol{H} \times \boldsymbol{r}]. \tag{G.13}$$

As has been shown above, this is not a unique choice, there are an infinite number of them. All the choices are equivalent from the mathematical and physical point of view, they differ however by a peanut, the economy of computations. It appears that this choice of $\boldsymbol{A}$ is at least a logical one. The choice is also consistent with the Coulombic gauge (eq. (G.12)), because

$$\begin{aligned}\nabla \cdot \boldsymbol{A} &= \frac{1}{2}\nabla \cdot [\boldsymbol{H} \times \boldsymbol{r}] = \frac{1}{2}\nabla \cdot [\boldsymbol{H} \times \boldsymbol{r}]\\ &= \frac{1}{2}\nabla \cdot [H_y z - yH_z, H_z x - zH_x, H_x y - xH_y]\\ &= \frac{1}{2}\left[\frac{\partial}{\partial x}(H_y z - yH_z) + \frac{\partial}{\partial y}(H_z x - zH_x) + \frac{\partial}{\partial z}(H_x y - xH_y)\right] = 0,\end{aligned}$$

and also with the Maxwell equations (p. 962), because

$$\begin{aligned}&\nabla \times \boldsymbol{A}\\ &= \frac{1}{2}\nabla \times [\boldsymbol{H} \times \boldsymbol{r}] = \frac{1}{2}\nabla \cdot [\boldsymbol{H} \times \boldsymbol{r}]\\ &= \frac{1}{2}\nabla \times [H_y z - yH_z, H_z x - zH_x, H_x y - xH_y]\\ &= \frac{1}{2}\left[\begin{array}{c}\frac{\partial}{\partial y}(H_x y - xH_y) - \frac{\partial}{\partial z}(H_z x - zH_x), \frac{\partial}{\partial z}(H_y z - yH_z) - \frac{\partial}{\partial x}(H_x y - xH_y),\\ \frac{\partial}{\partial x}(H_z x - zH_x) - \frac{\partial}{\partial y}(H_y z - yH_z)\end{array}\right]\\ &= \boldsymbol{H}.\end{aligned}$$

Thus, this is the correct choice.

---

[1] The Coulombic gauge, even if only one of the possibilities, is almost exclusively used in molecular physics. The word "*gauge*" comes from the railway terminology referring to the different distances between the rails.

Practical importance of this choice

An example of possible choices of $\boldsymbol{A}$ is given in Fig. G.1.

If we shifted the vector potential origin far from the physical system under consideration (Fig. G.1.b), the values of $|\boldsymbol{A}|$ on all the particles of the system would be gigantic. $\boldsymbol{A}$ would be practically homogeneous *within* the atom or molecule. If we calculated $\nabla \times \boldsymbol{A} = \boldsymbol{H}$ on a particle of the system, then however horrifying it might look, we would obtain almost $\mathbf{0}$, because $\nabla \times \boldsymbol{A}$ means the differentiation of $\boldsymbol{A}$, and for a homogeneous field this yields zero. Thus we are going to study the system in a magnetic field, but the field disappeared! Very high accuracy would be needed to calculate $\nabla \times \boldsymbol{A}$ correctly as differences between two large numbers, which is always a risky business numerically due to the cancellation of accuracies. *It is therefore seen that the numerical results do depend critically on the choice of the origin of* $\boldsymbol{A}$ (arbitrary from the point of view of mathematics and physics). *It is always better to have the origin inside the system.*

Vector potential causes the wave function to change phase

The Schrödinger equation for a particle of mass m and charge q is

$$-\frac{\hbar^2}{2m}\Delta\Psi(\boldsymbol{r}) + V\Psi = E\Psi(\boldsymbol{r}),$$

where $V = q\phi$ with ϕ standing for the scalar electric potential.

The probability density of finding the particle at a given position depends on $|\Psi|$ rather than Ψ itself. This means that the wave function could be harmlessly multiplied by a phase factor $\Psi'(\boldsymbol{r}) = \Psi(\boldsymbol{r})\exp[-\frac{iq}{\hbar c}\chi(\boldsymbol{r})]$, where $\chi(\boldsymbol{r})$ could be any (smooth[2]) function of the particle's position $\boldsymbol{r}$. Then we have $|\Psi| = |\Psi'|$ at any $\boldsymbol{r}$. If $\Psi'(\boldsymbol{r})$ is as good as Ψ is, it would be nice if it kindly satisfied the Schrödinger equation like Ψ does, of course with the same eigenvalue

$$-\frac{\hbar^2}{2m}\Delta\Psi'(\boldsymbol{r}) + V\Psi'(\boldsymbol{r}) = E\Psi'(\boldsymbol{r}).$$

Let us see what profound consequences this has. The left-hand side of the last equation can be transformed as follows

$$\begin{aligned}
&-\frac{\hbar^2}{2m}\Delta\Psi'(\boldsymbol{r}) + V\Psi'(\boldsymbol{r})\\
&\quad= -\frac{\hbar^2}{2m}\left[\exp\left(-\frac{iq}{\hbar c}\chi\right)\Delta\Psi + \Psi\Delta\exp\left(-\frac{iq}{\hbar c}\chi\right) + 2(\nabla\Psi)\left(\nabla\exp\left(-\frac{iq}{\hbar c}\chi\right)\right)\right]\\
&\qquad+ V\exp\left(-\frac{iq}{\hbar c}\chi\right)\Psi\\
&\quad= -\frac{\hbar^2}{2m}\left[\exp\left(-\frac{iq}{\hbar c}\chi\right)\Delta\Psi + \Psi\nabla\left[\left(-\frac{iq}{\hbar c}\right)\exp\left(-\frac{iq}{\hbar c}\chi\right)\nabla\chi\right]\right.
\end{aligned}$$

---

[2] See Fig. 2.5.

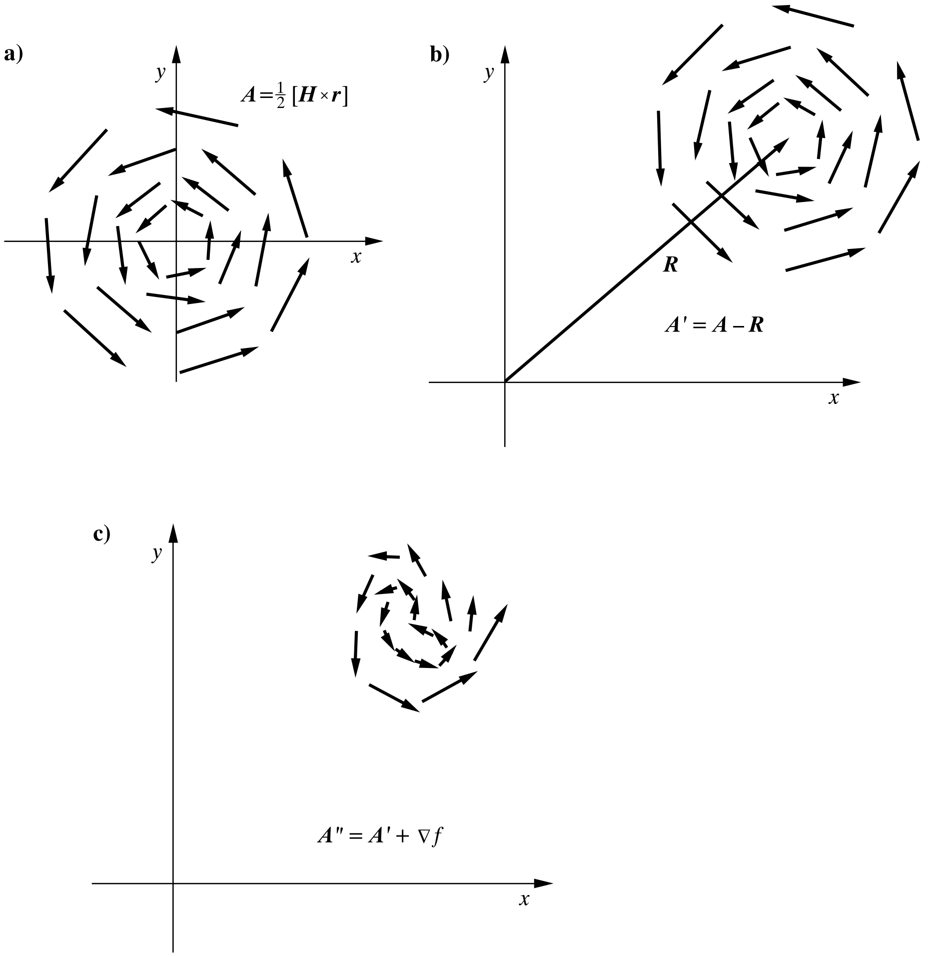


Fig. G.1. How do we understand the arbitrariness of the vector potential A? Figs. (a), (b), (c) represent schematically three physically equivalent vector potentials $\boldsymbol{A}$. Fig. (a) shows a section in the plane $z = 0$ (axis z protrudes towards the reader from the xy plane) of the vector field $\boldsymbol{A} = \frac{1}{2}(\boldsymbol{H} \times \boldsymbol{r})$ with $\boldsymbol{H} = (0, 0, H)$ and $H > 0$. We see that vectors $\boldsymbol{A}$ become longer and longer, when we leave the origin (where $\boldsymbol{A} = \mathbf{0}$), they "rotate" counter-clockwise. Such $\boldsymbol{A}$ therefore determines $\boldsymbol{H}$ directed perpendicularly to the page and oriented towards the reader. By the way, note that any *shift* of the potential obtained should give the same magnetic field perpendicular to the drawing, Fig. (b). This is what we get (Fig. (b)) after adding, according to eq. (G.11), the gradient of function $f = ax + by + c$ to potential $\boldsymbol{A}$, because $\boldsymbol{A} + \nabla f = \boldsymbol{A} + (\boldsymbol{i}a + \boldsymbol{j}b) = \boldsymbol{A} - \boldsymbol{R} = \boldsymbol{A}'$, where $\boldsymbol{R} = -(\boldsymbol{i}a + \boldsymbol{j}b) = \text{const}$. The transformation is only one of the possibilities. If we took an arbitrary smooth function $f(x, y)$, e.g., with many maxima, minima and saddle points (as in the mountains), we would deform Fig. (b) by expanding or shrinking it like a pancake. In this way we might obtain the situation shown in Fig. (c). All these situations a,b,c are physically indistinguishable (on condition that the scalar potential ϕ is changed appropriately).

$$
\begin{aligned}
&+2(\nabla\Psi)\bigg[\bigg(-\frac{iq}{\hbar c}\bigg)\exp\bigg(-\frac{iq}{\hbar c}\chi\bigg)\nabla\chi\bigg]\bigg]+V\exp\bigg(-\frac{iq}{\hbar c}\chi\bigg)\Psi\\
&=-\frac{\hbar^2}{2m}\bigg[\exp\bigg(-\frac{iq}{\hbar c}\chi\bigg)\Delta\Psi+\Psi\bigg(-\frac{iq}{\hbar c}\bigg)\bigg[\bigg(-\frac{iq}{\hbar c}\bigg)\exp\bigg(-\frac{iq}{\hbar c}\chi\bigg)(\nabla\chi)^2\\
&+\exp\bigg(-\frac{iq}{\hbar c}\chi\bigg)\Delta\chi\bigg]\bigg]-\frac{\hbar^2}{2m}2(\nabla\Psi)\bigg[\bigg(-\frac{iq}{\hbar c}\bigg)\exp\bigg(-\frac{iq}{\hbar c}\chi\bigg)\nabla\chi\bigg]\\
&+V\exp\bigg(-\frac{iq}{\hbar c}\chi\bigg)\Psi.
\end{aligned}
$$

Dividing the Schrödinger equation by $\exp(-\frac{iq}{\hbar c}\chi)$ we obtain

$$
\begin{aligned}
&-\frac{\hbar^2}{2m}\bigg[\Delta\Psi+\Psi\bigg(-\frac{iq}{\hbar c}\bigg)\bigg[\bigg(-\frac{iq}{\hbar c}\bigg)(\nabla\chi)^2+\Delta\chi\bigg]+2(\nabla\Psi)\bigg[\bigg(-\frac{iq}{\hbar c}\bigg)\nabla\chi\bigg]\bigg]+V\Psi\\
&=E\Psi(\boldsymbol{r}).
\end{aligned}
$$

Let us define a vector field $\boldsymbol{A}(\boldsymbol{r})$ using function $\chi(\boldsymbol{r})$

$$\boldsymbol{A}(\boldsymbol{r})=\nabla\chi(\boldsymbol{r}). \tag{G.14}$$

Hence, we have

$$-\frac{\hbar^2}{2m}\bigg[\Delta\Psi+\Psi\bigg(-\frac{iq}{\hbar c}\bigg)\bigg[\bigg(-\frac{iq}{\hbar c}\bigg)\boldsymbol{A}^2+\nabla\boldsymbol{A}\bigg]+2(\nabla\Psi)\bigg[\bigg(-\frac{iq}{\hbar c}\bigg)\boldsymbol{A}\bigg]\bigg]+V\Psi=E\Psi(\boldsymbol{r}),$$

and introducing the momentum operator $\hat{\boldsymbol{p}}=-i\hbar\nabla$ we obtain

$$\frac{1}{2m}\bigg[\hat{\boldsymbol{p}}^2\Psi+\Psi\bigg[\bigg(\frac{q}{c}\bigg)^2\boldsymbol{A}^2-\bigg(\frac{q}{c}\bigg)\hat{\boldsymbol{p}}\boldsymbol{A}\bigg]-2(\hat{\boldsymbol{p}}\Psi)\bigg(\frac{q}{c}\bigg)\boldsymbol{A}\bigg]+V\Psi=E\Psi(\boldsymbol{r}),$$

or finally

$$\frac{1}{2m}\bigg(\hat{\boldsymbol{p}}-\frac{q}{c}\boldsymbol{A}\bigg)^2\Psi+V\Psi=E\Psi, \tag{G.15}$$

which is the equation corresponding to the particle moving in electromagnetic field with vector potential $\boldsymbol{A}$, see p. 654.

Indeed, the last equation can be transformed in the following way

$$\frac{1}{2m}\bigg[\hat{\boldsymbol{p}}^2\Psi+\bigg(\frac{q}{c}\bigg)^2\boldsymbol{A}^2\Psi-\frac{q}{c}\hat{\boldsymbol{p}}(\boldsymbol{A}\Psi)-\frac{q}{c}\boldsymbol{A}\hat{\boldsymbol{p}}\Psi\bigg]+V\Psi=E\Psi,$$

which after using the equality[3] $\hat{\boldsymbol{p}}(\boldsymbol{A}\Psi)=\Psi\hat{\boldsymbol{p}}\boldsymbol{A}+\boldsymbol{A}\hat{\boldsymbol{p}}\Psi$ gives the expected result [eq. (G.15)].

---

[3] Remember that $\hat{\boldsymbol{p}}$ is proportional to the first derivative operator.

In conclusion, if a particle moves in a vector potential field $\boldsymbol{A}$ from $\boldsymbol{r}_0$ to $\boldsymbol{r}$, then its wave function changes the phase by δ

$$\delta = -\frac{q}{\hbar c} \int_{\boldsymbol{r}_0}^{\boldsymbol{r}} \boldsymbol{A}(\boldsymbol{r})\, \mathrm{d}\boldsymbol{r},$$

or, putting it in a different way: if the wave function undergoes a phase change, it means that the particle moves in the vector potential of an electromagnetic field.

The incredible Aharonov–Bohm effect

In a small area (say, in the centre of the Grand Place in Brussels, where we like to locate the origin of the coordinate system) there is a magnetic field flux corresponding to field intensity $\boldsymbol{H}$ directed along the z axis (perpendicular to the market place surface). Now let us imagine a particle of electric charge q enclosed in a 3D box (say, a cube) of small dimensions located at a *very long distance* from the origin, and therefore from the magnetic flux, say, in Lisbon. Therefore, the *magnetic field in the box is equal to zero.* Now we decide to travel with the box: Lisbon, Cairo, Ankara, StPetersburg, Stockholm, Paris, and back to Lisbon. Did the wave function of the particle in the box change during the journey?

Let us see. The magnetic field $\boldsymbol{H}$ is related to the vector potential $\boldsymbol{A}$ through the relation $\nabla \times \boldsymbol{A} = \boldsymbol{H}$. This means that the particle was subject to a huge vector potential field (see Fig. G.1) all the time, although the magnetic field was practically zero. Since the box is back to Lisbon, the phase acquired by the particle in the box[4] is an integral over the closed trajectory (loop)

$$\delta = -\frac{q}{\hbar c} \oint \boldsymbol{A}(\boldsymbol{r})\, \mathrm{d}\boldsymbol{r}.$$

However, from the Stokes equation, we can replace the integral by an integral over a surface enclosed by the loop

$$\delta = -\frac{q}{\hbar c} \oint \boldsymbol{A}(\boldsymbol{r})\, \mathrm{d}\boldsymbol{r} = -\frac{q}{\hbar c} \iint \nabla \times \boldsymbol{A}(\boldsymbol{r})\, \mathrm{d}\boldsymbol{S}.$$

This may be written as

$$\delta = -\frac{q}{\hbar c} \iint \boldsymbol{H}\, \mathrm{d}\boldsymbol{S} = -\frac{q}{\hbar c} \Phi,$$

where Φ is the magnetic flux (of the magnetic field $\boldsymbol{H}$) intersecting the loop surface, which contains, in particular, the famous market place of Brussels. Thus, despite the fact that the particle could not feel the magnetic field $\boldsymbol{H}$ (because it was zero in the box), its wave function underwent a change of phase, which is detectable experimentally (in interference experiments).

Does the pair of potentials $\boldsymbol{A}$ and ϕ contain the same information as $\boldsymbol{\mathcal{E}}$ and $\boldsymbol{H}$? The Aharonov–Bohm effect (see also p. 780) suggests that $\boldsymbol{A}$ and ϕ are more important.

---

[4] A non-zero δ requires a more general $\boldsymbol{A}$ than that satisfying eq. (G.14).

H. OPTIMAL WAVE FUNCTION FOR A HYDROGEN-LIKE ATOM

In several contexts we encounter the problem of the mean value of the Hamiltonian for a hydrogen-like atom (the a.u. are used throughout)

$$\hat{H} = -\frac{1}{2}\Delta - \frac{Z}{r}$$

with the normalized function

$$\Phi(r, \theta, \phi; c) = \sqrt{\frac{c^3}{\pi}} \exp(-cr),$$

where r, θ, ϕ are the spherical coordinates of the electron (the position of the nucleus is fixed at the origin).

Calculation of the mean value of the Hamiltonian, i.e. the mean value of the energy

$$\varepsilon(\Phi) = \langle \Phi | \hat{H} | \Phi \rangle$$

requires calculation of the mean value of the kinetic energy:

$$\bar{T} = \left\langle \Phi \middle| -\frac{1}{2}\Delta \middle| \Phi \right\rangle$$

and the mean value of the potential energy (Coulombic attraction of an electron by a nucleus of charge Z)

$$\bar{V} = -Z\left\langle \Phi \middle| \frac{1}{r} \middle| \Phi \right\rangle.$$

Therefore,

$$\varepsilon = \bar{T} + \bar{V}.$$

First, the Laplacian $\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$ may be expressed in spherical coordinates

$$\Delta = \frac{1}{r^2}\frac{\partial}{\partial r} r^2 \frac{\partial}{\partial r} + \frac{1}{r^2 \sin\theta}\frac{\partial}{\partial \theta}\sin\theta \frac{\partial}{\partial \theta} + \frac{1}{r^2 \sin^2\theta}\frac{\partial^2}{\partial \phi^2}, \tag{H.1}$$

and in view of the fact that Φ is spherically symmetric (it only depends on r)

$$\begin{aligned}\left\langle \Phi \middle| -\frac{1}{2}\Delta \middle| \Phi \right\rangle &= -\frac{1}{2}\left\langle \Phi \middle| \frac{1}{r^2}\frac{\partial}{\partial r} r^2 \frac{\partial}{\partial r} \middle| \Phi \right\rangle \\ &= -\frac{1}{2}\frac{c^3}{\pi}(-c)\left[\int_0^\infty r^2\left[\frac{2}{r}-c\right]\exp(-2cr)\,\mathrm{d}r \int_0^\pi \sin\theta\,\mathrm{d}\theta \int_0^{2\pi}\mathrm{d}\phi\right] \\ &= \frac{1}{2}c^4 4\left[\int_0^\infty \left[2r - cr^2\right]\exp(-2cr)\,\mathrm{d}r\right] \\ &= 2c^4\left[2\int_0^\infty r\exp(-2cr)\,\mathrm{d}r - c\int_0^\infty r^2\exp(-2cr)\,\mathrm{d}r\right] \\ &= 4c^4(2c)^{-2} - 2c^5 2(2c)^{-3} = c^2 - \frac{1}{2}c^2 = \frac{1}{2}c^2,\end{aligned}$$

where we have used (this formula is often exploited throughout the book)

$$\int_0^\infty r^n \exp(-\beta r)\,\mathrm{d}r = n!\beta^{-(n+1)}. \tag{H.2}$$

Similarly the second integral gives

$$\begin{aligned}-Z\left\langle \Phi \middle| \frac{1}{r} \middle| \Phi \right\rangle &= -Z\frac{c^3}{\pi}\left[\int_0^\infty r\exp(-2cr)\,\mathrm{d}r \int_0^\pi \sin\theta\,\mathrm{d}\theta \int_0^{2\pi}\mathrm{d}\phi\right] \\ &= -4Zc^3(2c)^{-2} = -Zc.\end{aligned}$$

Therefore, finally

$$\varepsilon = \frac{1}{2}c^2 - Zc. \tag{H.3}$$

We may want to use the variational method for finding the ground-state wave function. In this method we minimize the mean value of the Hamiltonian with respect to parameters in the variational function Φ. We may treat c as such a parameter. Hence, minimizing ε we force $\frac{\partial \varepsilon}{\partial c} = 0$, and therefore $c_{\mathrm{opt}} = Z$. Note that in this particular case:

- Such a value of c gives the *exact* ground-state of the hydrogen-like atom from the variational function.
- The ground-state energy computed with $c_{\mathrm{opt}} = Z$ gives $\varepsilon = \frac{1}{2}Z^2 - ZZ = -\frac{1}{2}Z^2$, which is the *exact* ground-state energy.
- virial theorem The quantity $-\frac{\bar{V}}{\bar{T}} = \frac{Zc}{\frac{1}{2}c^2} = 2\frac{Z}{c}$. For $c = c_{\mathrm{opt}} = Z$ we have what is called the *virial theorem*

$$-\frac{\bar{V}}{\bar{T}} = 2. \tag{H.4}$$

I. SPACE- AND BODY-FIXED COORDINATE SYSTEMS

Space-fixed coordinate system (SFCS)

A planetoid (or molecule) moves through empty space, we observe it from our (inertial[1]) space ship. To carry out observations of the planetoid (molecule), we have to install some equipment in our space ship and to fix a Cartesian coordinate system on it. This will enable us to describe the planetoid whatever happens to it. This is the Space-Fixed Coordinate System (SFCS), its orientation with respect to distant stars does not change in time.

If the molecule does not interact with anything, then with respect to the SFCS (see Chapter 2)

- its *total energy* remains invariant (because of the homogeneity of time),
- its *total momentum* remains invariant (because of the homogeneity of space),
- its *total angular momentum vector* remains invariant (because of the isotropy of space).

An observer on another space ship (also inertial) will see the same phenomena in exactly the same way,[2] the energy, momentum and angular momentum will also be invariant, but in general they will be different from what was measured in the first space ship.

Let us introduce the vectors $\boldsymbol{r}_i = (x_i, y_i, z_i)$ into the SFCS showing (from the origin of the coordinate system) the particles, from which our molecule is composed (i.e. the electrons and the nuclei), $i = 1, 2, \dots, N$. Then, using the SFCS, we write the Hamiltonian of the system, the operators of the mechanical quantities we are interested in, we calculate all the wave functions we need, compare with spectra measured in the SFCS, etc.

Body-fixed coordinate system (BFCS)

One day, however, we may feel that we do not like the SFCS, because to describe the molecule we use too many variables. Of course, this is not a sin, but only a

---

[1] No rotation. We will convince ourselves that our SFCS is inertial by measuring how a point-like mass moves (assumed to be non-interacting with the rest of the space ship). If it moves along a straight line with a constant velocity, the SFCS is inertial. In a *non-inertial* coordinate system the description of the physical phenomena in the molecule will look different.

[2] In the non-relativistic approximation. The Doppler effect, with the change in electromagnetic wave frequency due to the motion (even uniform) of the emitting object *is seen* in the experiment. The effect is of a relativistic character, i.e. vanishes, if we assume an infinite velocity of light.

waste of our time. Indeed, since in all inertial systems we have the same physics, we can separate the motion of the centre of mass[3] (the total mass $M = \sum_i m_i$). The centre of mass with position

$$\boldsymbol{R}_{\text{CM}} = \frac{\sum_i m_i \boldsymbol{r}_i}{M}$$

moves with a constant velocity along a straight line in the SFCS, which can easily be taken into account after the solution is obtained, and in most cases it is irrelevant. This is why we decide to introduce the Cartesian coordinates $(X_{\text{CM}}, Y_{\text{CM}}, Z_{\text{CM}}) = \boldsymbol{R}_{\text{CM}}$ in the hope that in future we will be able to get rid of them. Now we need to introduce a coordinate system (of the missing $3N - 3$ variables) located on the molecule, called the *body-fixed coordinate system* (BFCS). How to define this? Well, it should be a coordinate system that will define any configuration of the particles in the molecule unambiguously. There are a lot of such coordinate systems. Here you have some of the possibilities for the BFCS (*in all of them their axes are parallel to the corresponding axes of the SFCS*[4]). We may choose one of the following sets[5] of position vectors:

- $\boldsymbol{R}_{\text{CM}}$, then, we locate in the BFCS *on any of the particles* (say, the one indicated by vector $\boldsymbol{r}_1$), and the BFCS positions of the other particles are shown by: $\boldsymbol{r}'_i = \boldsymbol{r}_i - \boldsymbol{r}_1$ for $i = 2, 3, \ldots, N$.
- $\boldsymbol{R}_{\text{CM}}$, the vector $\boldsymbol{R} = \boldsymbol{r}_2 - \boldsymbol{r}_1$ *indicating particle* 2 *from particle* 1, and the remaining particles are shown by the vectors which begin in the centre of the section linking particles 1 and 2: $\boldsymbol{r}'_i = \boldsymbol{r}_i - \frac{(\boldsymbol{r}_1+\boldsymbol{r}_2)}{2}$ for $i = 3, 4, \ldots, N$.
- $\boldsymbol{R}_{\text{CM}}$, and all the vectors showing the particles (except particle 1): $\boldsymbol{r}'_i = \boldsymbol{r}_i - \boldsymbol{R}_{\text{CM}}$ for $i = 2, 3, \ldots, N$. the position vector of the particle 1 can be calculated from the coordinates already given.

Centre-of-mass separation

After writing the Hamiltonian $\hat{\mathcal{H}}$ in the SFCS, and introducing *any of the above choices* of coordinate system, we obtain $\hat{\mathcal{H}} = \hat{H}_{\text{CM}} + \hat{H}$, where

$$\hat{H}_{\text{CM}} = -\frac{\hbar^2}{2M}\Delta_{\text{CM}}$$

with $\Delta_{\text{CM}} = \frac{\partial^2}{\partial X_{\text{CM}}^2} + \frac{\partial^2}{\partial Y_{\text{CM}}^2} + \frac{\partial^2}{\partial Z_{\text{CM}}^2}$, and $\hat{H}$ that does not contain X_{CM}, Y_{CM}, Z_{CM}.

---

[3] The exact separation of the centre-of-mass motion in SFCS, as well as (not shown in this Appendix) the exact separation of rotation of the molecule have been shown in the paper by R.T. Pack, J.O. Hirschfelder, *J. Chem. Phys.* 49 (1968) 4009 for the first time.

[4] Only after introducing the axes of the coordinate system associated with the particles, and not with the SFCS, separation of rotation is possible.

[5] There are other possible choices.

At any of the choices the operator $\hat{H}$ is identical, but the mathematical formula for $\hat{H}$ will be different, because different coordinates are used.

Thus, the total Hamiltonian in the SFCS is

$$\hat{\mathcal{H}} = \hat{H}_{\mathrm{CM}}(X_{\mathrm{CM}}, Y_{\mathrm{CM}}, Z_{\mathrm{CM}}) + \hat{H}(\boldsymbol{r}),$$

where $\boldsymbol{r}$ *symbolizes*[6] all the other variables. The key result is that the two operators on the right *do depend on different variables*.

The goal of the above changes to the coordinate system was to show that the Schrödinger equation written in the SFCS, i.e. $\hat{\mathcal{H}}\Psi = \mathcal{E}\Psi$, splits into *two* Schrödinger equations ("separation of variables"):

- $\hat{H}_{\mathrm{CM}}\psi_{\mathrm{CM}} = E_{\mathrm{CM}}\psi_{\mathrm{CM}}$ describing the motion of a free "particle" of mass M and coordinates $X_{\mathrm{CM}}, Y_{\mathrm{CM}}, Z_{\mathrm{CM}}$ (the "centre-of-mass motion"), with $\psi_{\mathrm{CM}} = \exp(i\boldsymbol{p}_{\mathrm{CM}} \cdot \boldsymbol{R}_{\mathrm{CM}})$, where $\boldsymbol{p}_{\mathrm{CM}}$ stands for the total momentum of the system;
- $\hat{H}\psi = E\psi$, where

$$\mathcal{E} = E + E_{\mathrm{CM}},$$
$$\Psi(\boldsymbol{R}_{\mathrm{CM}}, \boldsymbol{r}) = \psi_{\mathrm{CM}}(\boldsymbol{R}_{\mathrm{CM}}) \cdot \psi(\boldsymbol{r}).$$

The proof is simple. Let us check that the product wave function satisfies the Schrödinger equation. The left-hand side is:

$$\begin{aligned}
\hat{\mathcal{H}}\big[\psi_{\mathrm{CM}}(\boldsymbol{R}_{\mathrm{CM}}) \cdot \psi(\boldsymbol{r})\big] &= \hat{H}_{\mathrm{CM}}\big[\psi_{\mathrm{CM}}(\boldsymbol{R}_{\mathrm{CM}}) \cdot \psi(\boldsymbol{r})\big] + \hat{H}\big[\psi_{\mathrm{CM}}(\boldsymbol{R}_{\mathrm{CM}}) \cdot \psi(\boldsymbol{r})\big] \\
&= \psi(\boldsymbol{r}) \cdot \hat{H}_{\mathrm{CM}}\psi_{\mathrm{CM}}(\boldsymbol{R}_{\mathrm{CM}}) + \psi_{\mathrm{CM}}(\boldsymbol{R}_{\mathrm{CM}}) \cdot \hat{H}\psi(\boldsymbol{r}) \\
&= \psi(\boldsymbol{r}) \cdot E_{\mathrm{CM}}\psi_{\mathrm{CM}}(\boldsymbol{R}_{\mathrm{CM}}) + \psi_{\mathrm{CM}}(\boldsymbol{R}_{\mathrm{CM}}) \cdot E\psi(\boldsymbol{r}) \\
&= (E + E_{\mathrm{CM}})\big[\psi_{\mathrm{CM}}(\boldsymbol{R}_{\mathrm{CM}}) \cdot \psi(\boldsymbol{r})\big]
\end{aligned}$$

and this equals the right side $\mathcal{E}\Psi$.

Example 1. ***Centre-of-mass separation for the first choice of the coordinates.*** We use the first choice of coordinates for the system of *two* particles. In the SFCS

$$\hat{\mathcal{H}} = -\frac{\hbar^2}{2m_1}\Delta_1 - \frac{\hbar^2}{2m_2}\Delta_2 + V.$$

The new coordinates are:

$$X_{\mathrm{CM}} = \frac{\sum_i m_i x_i}{M}, \qquad Y_{\mathrm{CM}} = \frac{\sum_i m_i y_i}{M}, \qquad Z_{\mathrm{CM}} = \frac{\sum_i m_i z_i}{M},$$
$$x = x_2 - x_1, \quad y = y_2 - y_1, \quad z = z_2 - z_1.$$

---

[6] For the sake of brevity.

Then,[7]

$$\begin{aligned}\frac{\partial}{\partial x_1} &= \frac{\partial X_{\mathrm{CM}}}{\partial x_1}\frac{\partial}{\partial X_{\mathrm{CM}}} + \frac{\partial Y_{\mathrm{CM}}}{\partial x_1}\frac{\partial}{\partial Y_{\mathrm{CM}}} + \frac{\partial Z_{\mathrm{CM}}}{\partial x_1}\frac{\partial}{\partial Z_{\mathrm{CM}}} + \frac{\partial x}{\partial x_1}\frac{\partial}{\partial x} + \frac{\partial y}{\partial x_1}\frac{\partial}{\partial y} + \frac{\partial z}{\partial x_1}\frac{\partial}{\partial z}\\ &= \frac{m_1}{M}\frac{\partial}{\partial X_{\mathrm{CM}}} + 0 + 0 - \frac{\partial}{\partial x} + 0 + 0 = \frac{m_1}{M}\frac{\partial}{\partial X_{\mathrm{CM}}} - \frac{\partial}{\partial x}\end{aligned}$$

and similarly for y_1 and z_1. Further,

$$\begin{aligned}\frac{\partial}{\partial x_2} &= \frac{\partial X_{\mathrm{CM}}}{\partial x_2}\frac{\partial}{\partial X_{\mathrm{CM}}} + \frac{\partial Y_{\mathrm{CM}}}{\partial x_2}\frac{\partial}{\partial Y_{\mathrm{CM}}} + \frac{\partial Z_{\mathrm{CM}}}{\partial x_2}\frac{\partial}{\partial Z_{\mathrm{CM}}} + \frac{\partial x}{\partial x_2}\frac{\partial}{\partial x} + \frac{\partial y}{\partial x_2}\frac{\partial}{\partial y} + \frac{\partial z}{\partial x_2}\frac{\partial}{\partial z}\\ &= \frac{m_2}{M}\frac{\partial}{\partial X_{\mathrm{CM}}} + 0 + 0 + \frac{\partial}{\partial x} + 0 + 0 = \frac{m_2}{M}\frac{\partial}{\partial X_{\mathrm{CM}}} + \frac{\partial}{\partial x}\end{aligned}$$

and similarly for y_2 and z_2.

Hence, the kinetic energy operator (after constructing the proper Laplacians from the operators above)

$$\begin{aligned}\hat{T} &= -\frac{\hbar^2}{2m_1}\Delta_1 - \frac{\hbar^2}{2m_2}\Delta_2\\ &= -\frac{\hbar^2}{2m_1}\left[\left(\frac{m_1}{M}\right)^2\frac{\partial^2}{\partial X_{\mathrm{CM}}^2} + \frac{\partial^2}{\partial x^2} - 2\frac{m_1}{M}\frac{\partial^2}{\partial X_{\mathrm{CM}}\partial x}\right] + (\text{similarly for } y \text{ and } z)\\ &\quad - \frac{\hbar^2}{2m_2}\left[\left(\frac{m_2}{M}\right)^2\frac{\partial^2}{\partial X_{\mathrm{CM}}^2} + \frac{\partial^2}{\partial x^2} + 2\frac{m_2}{M}\frac{\partial^2}{\partial X_{\mathrm{CM}}\partial x}\right] + (\text{similarly for } y \text{ and } z)\\ &= -\frac{\hbar^2}{2M}\Delta_{\mathrm{CM}} - \frac{\hbar^2}{2\mu}\Delta,\end{aligned}$$

where the reduced mass μ of the two particles: $\frac{1}{\mu} = \frac{1}{m_1} + \frac{1}{m_2}$, and $\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$.

Our derivation is over, and the operator $\hat{H}$ has been found. It turns out to be[8] (note, that the new coordinates also have to be introduced in the potential energy V) of the form

---

[7] According to the mathematical analysis we have to write the contributions of all the differential operators $\frac{\partial}{\partial u}$ of the new coordinates u multiplied by their "coupling constants" $\frac{\partial u}{\partial x_1}$ with the coordinate x_1.

[8] The kinetic energy operator has a quite interesting form. Particle 1 *rests* right at the origin of the BFCS ($x = 0$, $y = 0$, $z = 0$), and therefore its kinetic energy operator is absent in $\hat{H}$. There is the kinetic energy of particle 2, *but its mass is equal to* μ, not to m_2. The coordinates x, y, z (measured from the origin of the BFCS) correspond to particle 2. For example, for the hydrogen-like atom, if someone takes the nucleus as particle 1, and the electron as particle 2, then x, y, z show the *electron*

$$\hat{H} = -\frac{\hbar^2}{2\mu}\Delta + V.$$

Example 2. ***Centre-of-mass separation for the third choice of coordinates.*** Let us take the same two particles again, but this time use the third choice of coordinate system.

$$X_{CM} = \frac{\sum_i m_i x_i}{M}, \qquad Y_{CM} = \frac{\sum_i m_i y_i}{M}, \qquad Z_{CM} = \frac{\sum_i m_i z_i}{M},$$
$$x = x_2 - X_{CM}, \quad y = y_2 - Y_{CM}, \quad z = z_2 - Y_{CM}.$$

Then,

$$\frac{\partial}{\partial x_1} = \frac{\partial X_{CM}}{\partial x_1}\frac{\partial}{\partial X_{CM}} + \frac{\partial Y_{CM}}{\partial x_1}\frac{\partial}{\partial Y_{CM}} + \frac{\partial Z_{CM}}{\partial x_1}\frac{\partial}{\partial Z_{CM}} + \frac{\partial x}{\partial x_1}\frac{\partial}{\partial x} + \frac{\partial y}{\partial x_1}\frac{\partial}{\partial y} + \frac{\partial z}{\partial x_1}\frac{\partial}{\partial z}$$
$$= \frac{m_1}{M}\frac{\partial}{\partial X_{CM}} + 0 + 0 - \frac{m_1}{M}\frac{\partial}{\partial x} + 0 + 0 = \frac{m_1}{M}\left(\frac{\partial}{\partial X_{CM}} - \frac{\partial}{\partial x}\right)$$

and similarly for y_1 and z_1. Further,

$$\frac{\partial}{\partial x_2} = \frac{\partial X_{CM}}{\partial x_2}\frac{\partial}{\partial X_{CM}} + \frac{\partial Y_{CM}}{\partial x_2}\frac{\partial}{\partial Y_{CM}} + \frac{\partial Z_{CM}}{\partial x_2}\frac{\partial}{\partial Z_{CM}} + \frac{\partial x}{\partial x_2}\frac{\partial}{\partial x} + \frac{\partial y}{\partial x_2}\frac{\partial}{\partial y} + \frac{\partial z}{\partial x_2}\frac{\partial}{\partial z}$$
$$= \frac{m_2}{M}\frac{\partial}{\partial X_{CM}} + 0 + 0 + \left(1 - \frac{m_2}{M}\right)\frac{\partial}{\partial x} + 0 + 0 = \frac{m_2}{M}\frac{\partial}{\partial X_{CM}} + \left(1 - \frac{m_2}{M}\right)\frac{\partial}{\partial x}$$
$$= \frac{m_2}{M}\frac{\partial}{\partial X_{CM}} + \frac{m_1}{M}\frac{\partial}{\partial x}$$

and similarly for y_2 and z_2.

Thus, the kinetic energy operator takes the form (after inserting the squares of the corresponding operators)

---

from the Cartesian coordinate system BFCS located on the nucleus. The potential energy operator

$$V = -\frac{Ze^2}{\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}} = -\frac{Ze^2}{\sqrt{x^2 + y^2 + z^2}}$$

corresponds to the Coulombic interaction of the electron of charge $-e$ and the nucleus of charge Ze. After the separation of the centre of mass, we are left with equation $\hat{H}\psi = E\psi$. The electron of mass μ is described by the wave function ψ. In the ground state $\psi = \frac{1}{\sqrt{\pi}}e^{-\sqrt{x^2+y^2+z^2}}$. This a the description of the hydrogen-like atom *according to an observer sitting at the nucleus*.

If another observer puts his armchair (with the axes of the BFCS carved on it) at the *electron*, then he would see the hydrogen-like atom "according to the electron". Since in V there are squares of x, y, z, and in the kinetic energy operator there are the *second* derivatives with respect to x, y, z, we would obtain the same wave function as before: $\psi = \frac{1}{\sqrt{\pi}}e^{-\sqrt{x^2+y^2+z^2}}$, where the particle moving with respect to the electron is the *nucleus*, but with mass equal to μ, i.e. the same as before. By the way, this μ is almost equal to the mass of the *electron*.

Thus, the two descriptions mean the same.

$$\begin{aligned}
\hat{T} &= -\frac{\hbar^2}{2m_1}\Delta_1 - \frac{\hbar^2}{2m_2}\Delta_2\\
&= -\frac{\hbar^2}{2m_1}\left[\left(\frac{m_1}{M}\right)^2\left(\frac{\partial^2}{\partial X_{\mathrm{CM}}^2} + \frac{\partial^2}{\partial x^2} - 2\frac{\partial^2}{\partial X_{\mathrm{CM}}\partial x}\right)\right] + (\text{similarly for } y \text{ and } z)\\
&\quad - \frac{\hbar^2}{2m_2}\left[\left(\frac{m_2}{M}\right)^2\frac{\partial^2}{\partial X_{\mathrm{CM}}^2} + \left(\frac{m_1}{M}\right)^2\frac{\partial^2}{\partial x^2} + 2\frac{m_1m_2}{M^2}\frac{\partial^2}{\partial X_{\mathrm{CM}}\partial x}\right]\\
&\quad + (\text{similarly for } y \text{ and } z)\\
&= -\frac{\hbar^2}{2M}\Delta_{\mathrm{CM}} - \frac{\hbar^2}{2m_1}\left(\frac{m_1}{M}\right)^2\Delta_{xyz}\\
&\quad - \frac{\hbar^2}{2m_2}\left(\frac{m_1}{M}\right)^2\Delta_{xyz} - \frac{\hbar^2}{2m_1}\left(\frac{m_1}{M}\right)^2\left(-2\frac{\partial^2}{\partial X_{\mathrm{CM}}\partial x}\right) + \cdots\\
&\quad - \frac{\hbar^2}{2m_2}2\frac{m_1m_2}{M^2}\frac{\partial^2}{\partial X_{\mathrm{CM}}\partial x} + \cdots\\
&= -\frac{\hbar^2}{2M}\Delta_{\mathrm{CM}} - \frac{\hbar^2}{2m_1}\left(\frac{m_1}{M}\right)^2\Delta_{xyz} - \frac{\hbar^2}{2m_2}\left(\frac{m_1}{M}\right)^2\Delta_{xyz}\\
&= -\frac{\hbar^2}{2M}\Delta_{\mathrm{CM}} - \frac{\hbar^2}{2}\left(\frac{m_1}{m_2M}\right)\Delta_{xyz}.
\end{aligned}$$

It is seen that once again we have reached a situation allowing us to separate the motion of the centre of mass in the Schrödinger equation. This time, however, the *form* of the operator $\hat{H}$ is different (e.g., Δ_{xyz} has only formally the same form as Δ), only because the variables are different (the operator remains *the same*). Once again this is the kinetic energy of a point-like particle[9] with coordinates x, y, z (defined in *this* example) and mass equal to $\frac{m_2M}{m_1}$.

---

[9] Let us first denote the nucleus as particle 1 and the electron as particle 2. Then, $\boldsymbol{R}_{\mathrm{CM}}$ almost shows the position of the nucleus, and x, y, z are almost the coordinates of the electron measured from the nucleus, while $\frac{m_2M}{m_1}$ is almost equal to the mass of the electron. Thus we have a situation which resembles Example 1.

If the particles are chosen the other way (the electron is particle 1 and the nucleus is particle 2), the same physical situation looks completely different. The values of x, y, z are very close to 0, while the mass of the effective point-like particle becomes very large.

Note, that the new coordinates describe the potential energy in a more complex way. We need differences of the kind $x_2 - x_1$, to insert them into Pythagoras' formula for the distance. We have

$$x_1 = X_{\mathrm{CM}}\frac{m_1+m_2}{m_1} - \frac{m_2}{m_1}x_2 = X_{\mathrm{CM}}\frac{m_1+m_2}{m_1} - \frac{m_2}{m_1}(x + X_{\mathrm{CM}}) = X_{\mathrm{CM}} - \frac{m_2}{m_1}x,$$

$$x_1 - x_2 = X_{\mathrm{CM}} - \frac{m_2}{m_1}x - x - X_{\mathrm{CM}} = -x\left(1 + \frac{m_2}{m_1}\right).$$

This gives immediately (r stands for the electron-centre of mass distance): $V(\text{new}) = -\frac{Ze^2}{(1+\frac{m_2}{m_1})r}$.

J. ORTHOGONALIZATION

1 SCHMIDT ORTHOGONALIZATION

Two vectors

Imagine two vectors $\boldsymbol{u}$ and $\boldsymbol{v}$, each of length 1 (i.e. normalized), with the dot product $\langle \boldsymbol{u}|\boldsymbol{v}\rangle = a$. If $a = 0$, the two vectors are orthogonal. We are interested in the case $a \neq 0$. Can we make such linear combinations of $\boldsymbol{u}$ and $\boldsymbol{v}$, so that the new vectors, $\boldsymbol{u}'$ and $\boldsymbol{v}'$, will be orthogonal? We can do this in many ways, two of them are called the Schmidt orthogonalization:

Case I: $\boldsymbol{u}' = \boldsymbol{u}$, $\boldsymbol{v}' = \boldsymbol{v} - \boldsymbol{u}\langle \boldsymbol{u}|\boldsymbol{v}\rangle$,
Case II: $\boldsymbol{u}' = \boldsymbol{u} - \boldsymbol{v}\langle \boldsymbol{v}|\boldsymbol{u}\rangle$, $\boldsymbol{v}' = \boldsymbol{v}$.

It is seen that Schmidt orthogonalization is based on a very simple idea. In Case I the first vector is left unchanged, while from the second vector, we cut out its component along the first (Fig. J.1). In this way the two vectors are treated differently (hence, the two cases above).

In this book the vectors we orthogonalize will be Hilbert space vectors (see Appendix B), i.e. the normalized wave functions. In the case of two such vectors ϕ_1 and ϕ_2 having a dot product $\langle \phi_1|\phi_2\rangle$ we construct the new orthogonal wave

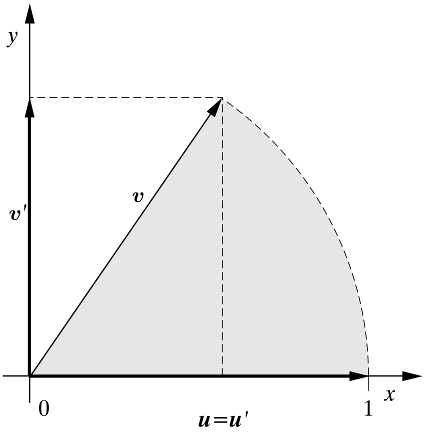


Fig. J.1. The Schmidt orthogonalization of the unit (i.e. normalized) vectors $\boldsymbol{u}$ and $\boldsymbol{v}$. The new vectors are $\boldsymbol{u}'$ and $\boldsymbol{v}'$. Vector $\boldsymbol{u}' \equiv \boldsymbol{u}$, while from vector $\boldsymbol{v}$ we subtract its component along $\boldsymbol{u}$. The new vectors are orthogonal.

functions $\psi_1 = \phi_1$, $\psi_2 = \phi_2 - \phi_1\langle\phi_1|\phi_2\rangle$ or $\psi_1 = \phi_1 - \phi_2\langle\phi_2|\phi_1\rangle$, $\psi_2 = \phi_2$ by analogy to the previous formulae.

More vectors

In case of many vectors the procedure is similar. First, we decide the order of the vectors to be orthogonalized. Then we begin the procedure by leaving the first vector unchanged. Then we continue, remembering that from a new vector we have to cut out all its components along the new vectors already found. Of course, the final set of vectors depends on the order chosen.

2 LÖWDIN SYMMETRIC ORTHOGONALIZATION

Imagine the normalized but non-orthogonal basis set wave functions collected as the components of the vector $\boldsymbol{\phi}$. By making proper linear combinations of the wave functions, we will get the orthogonal wave functions. The *symmetric* orthogonalization (as opposed to the Schmidt orthogonalization) treats all the wave functions on an equal footing. Instead of the old non-orthogonal basis set $\boldsymbol{\phi}$, we construct a new basis set $\boldsymbol{\phi}'$ by a linear transformation $\boldsymbol{\phi}' = \boldsymbol{S}^{-\frac{1}{2}}\boldsymbol{\phi}$, where $\boldsymbol{S}$ is the overlap matrix with the elements $S_{ij} = \langle\phi_i|\phi_j\rangle$, and the square matrix $\boldsymbol{S}^{-\frac{1}{2}}$, and its cousin $\boldsymbol{S}^{\frac{1}{2}}$, are defined in the following way. First, we diagonalize $\boldsymbol{S}$ using a unitary matrix $\boldsymbol{U}$, i.e. $\boldsymbol{U}^\dagger\boldsymbol{U} = \boldsymbol{U}\boldsymbol{U}^\dagger = \mathbf{1}$ (for real $\boldsymbol{S}$ the matrix $\boldsymbol{U}$ is orthogonal, $\boldsymbol{U}^T\boldsymbol{U} = \boldsymbol{U}\boldsymbol{U}^T = \mathbf{1}$),

$$\boldsymbol{S}_{\text{diag}} = \boldsymbol{U}^\dagger\boldsymbol{S}\boldsymbol{U}.$$

The eigenvalues of $\boldsymbol{S}$ are always positive, therefore the diagonal elements of $\boldsymbol{S}_{\text{diag}}$ can be replaced by their square roots, thus producing the matrix denoted by the *symbol* $\boldsymbol{S}^{\frac{1}{2}}_{\text{diag}}$. Using the latter matrix we define the matrices

$$\boldsymbol{S}^{\frac{1}{2}} = \boldsymbol{U}\boldsymbol{S}^{\frac{1}{2}}_{\text{diag}}\boldsymbol{U}^\dagger \quad \text{and} \quad \boldsymbol{S}^{-\frac{1}{2}} = \left(\boldsymbol{S}^{\frac{1}{2}}\right)^{-1} = \boldsymbol{U}\boldsymbol{S}^{-\frac{1}{2}}_{\text{diag}}\boldsymbol{U}^\dagger.$$

Their *symbols* correspond to their properties:

$$\boldsymbol{S}^{\frac{1}{2}}\boldsymbol{S}^{\frac{1}{2}} = \boldsymbol{U}\boldsymbol{S}^{\frac{1}{2}}_{\text{diag}}\boldsymbol{U}^\dagger\boldsymbol{U}\boldsymbol{S}^{\frac{1}{2}}_{\text{diag}}\boldsymbol{U}^\dagger = \boldsymbol{U}\boldsymbol{S}^{\frac{1}{2}}_{\text{diag}}\boldsymbol{S}^{\frac{1}{2}}_{\text{diag}}\boldsymbol{U}^\dagger = \boldsymbol{U}\boldsymbol{S}_{\text{diag}}\boldsymbol{U}^\dagger = \boldsymbol{S},$$

similarly $\boldsymbol{S}^{-\frac{1}{2}}\boldsymbol{S}^{-\frac{1}{2}} = \boldsymbol{S}^{-1}$. Also, a straightforward calculation gives[1] $\boldsymbol{S}^{-\frac{1}{2}}\boldsymbol{S}^{\frac{1}{2}} = \mathbf{1}$.

---

[1] The matrix $\boldsymbol{S}^{-\frac{1}{2}}$ is no longer a symbol anymore. Let us check whether the transformation $\boldsymbol{\phi}' = \boldsymbol{S}^{-\frac{1}{2}}\boldsymbol{\phi}$ indeed gives orthonormal wave functions (vectors). Remembering that $\boldsymbol{\phi}$ represents a vertical vector with components ϕ_i (being functions): $\int \boldsymbol{\phi}^*\boldsymbol{\phi}^T\,\mathrm{d}\tau = \boldsymbol{S}$, while $\int \boldsymbol{\phi}'^*\boldsymbol{\phi}'^T\,\mathrm{d}\tau = \int \boldsymbol{S}^{-\frac{1}{2}}\boldsymbol{\phi}^*\boldsymbol{\phi}^T\boldsymbol{S}^{-\frac{1}{2}}\,\mathrm{d}\tau = \mathbf{1}$. This is what we wanted to show.

An important feature of symmetric orthogonalization is[2] that among all possible orthogonalizations it ensures that

$$\sum_i \|\phi_i - \phi_i'\|^2 = \text{minimum}$$

where $\|\phi_i - \phi_i'\|^2 \equiv \langle \phi_i - \phi_i' | \phi_i - \phi_i' \rangle$. This means that

> the symmetrically orthogonalized functions ϕ_i' are the "least distant" from the original functions ϕ_i. Thus symmetric orthogonalization means a gentle pushing the directions of the vectors in order to get them to be orthogonal.

Example

Symmetric orthogonalization will be shown taking the example of two non-orthogonal vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ (instead of functions ϕ_1 and ϕ_2), each of length 1, with a dot product[3] $\langle \boldsymbol{u}|\boldsymbol{v}\rangle = a \neq 0$. We decide to consider vectors with real components, hence $a \in R$. First we have to construct matrix $\boldsymbol{S}^{-\frac{1}{2}}$. Here is how we arrive there. Matrix $\boldsymbol{S}$ is equal to $\boldsymbol{S} = \begin{pmatrix} 1 & a \\ a & 1 \end{pmatrix}$, and as we see it is symmetric. First, let us diagonalize $\boldsymbol{S}$. To achieve this, we apply the orthogonal transformation $\boldsymbol{U}^\dagger \boldsymbol{S} \boldsymbol{U}$ (thus, in this case $\boldsymbol{U}^\dagger = \boldsymbol{U}^T$), where (to ensure the orthogonality of the transformation matrix) we choose

$$\boldsymbol{U} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}, \quad \text{and therefore} \quad \boldsymbol{U}^\dagger = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

with angle θ to be specified. After the transformation we have:

$$\boldsymbol{U}^\dagger \boldsymbol{S} \boldsymbol{U} = \begin{pmatrix} 1 - a\sin 2\theta & a\cos 2\theta \\ a\cos 2\theta & 1 + a\sin 2\theta \end{pmatrix}.$$

We see that if we chose $\theta = 45°$, the matrix $\boldsymbol{U}^\dagger \boldsymbol{S} \boldsymbol{U}$ will be *diagonal*[4] (this is what we would like to have):

$$\boldsymbol{S}_{\text{diag}} = \begin{pmatrix} 1 - a & 0 \\ 0 & 1 + a \end{pmatrix}.$$

We then construct

$$\boldsymbol{S}_{\text{diag}}^{\frac{1}{2}} = \begin{pmatrix} \sqrt{1 - a} & 0 \\ 0 & \sqrt{1 + a} \end{pmatrix}.$$

---

[2] G.W. Pratt, S.P. Neustadter, *Phys. Rev*. 101 (1956) 1248.

[3] $-1 \leqslant a \leqslant 1$.

[4] In such a case the transformation matrix is

$$\boldsymbol{U} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}.$$

Next, we form[5]

$$\boldsymbol{S}^{\frac{1}{2}} = \boldsymbol{U}\boldsymbol{S}_{\text{diag}}^{\frac{1}{2}}\boldsymbol{U}^{\dagger} = \frac{1}{2}\begin{pmatrix} \sqrt{1-a}+\sqrt{1+a} & \sqrt{1+a}-\sqrt{1-a} \\ \sqrt{1+a}-\sqrt{1-a} & \sqrt{1-a}+\sqrt{1+a} \end{pmatrix}$$

and the matrix $\boldsymbol{S}^{-\frac{1}{2}}$ needed for the transformation is equal to

$$\boldsymbol{S}^{-\frac{1}{2}} = \boldsymbol{U}\boldsymbol{S}_{\text{diag}}^{-\frac{1}{2}}\boldsymbol{U}^{\dagger} = \boldsymbol{U}\begin{pmatrix} \frac{1}{\sqrt{1-a}} & 0 \\ 0 & \frac{1}{\sqrt{1+a}} \end{pmatrix}\boldsymbol{U}^{\dagger} = \frac{1}{2}\begin{pmatrix} \frac{1}{\sqrt{1-a}}+\frac{1}{\sqrt{1+a}} & \frac{1}{\sqrt{1+a}}-\frac{1}{\sqrt{1-a}} \\ \frac{1}{\sqrt{1+a}}-\frac{1}{\sqrt{1-a}} & \frac{1}{\sqrt{1-a}}+\frac{1}{\sqrt{1+a}} \end{pmatrix}.$$

Now we are ready to construct the orthogonalized vectors:[6]

$$\begin{pmatrix} \boldsymbol{u}' \\ \boldsymbol{v}' \end{pmatrix} = \frac{1}{2}\begin{pmatrix} \frac{1}{\sqrt{1-a}}+\frac{1}{\sqrt{1+a}} & \frac{1}{\sqrt{1+a}}-\frac{1}{\sqrt{1-a}} \\ \frac{1}{\sqrt{1+a}}-\frac{1}{\sqrt{1-a}} & \frac{1}{\sqrt{1-a}}+\frac{1}{\sqrt{1+a}} \end{pmatrix}\begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{v} \end{pmatrix},$$

$$\boldsymbol{u}' = C\boldsymbol{u} + c\boldsymbol{v},$$

$$\boldsymbol{v}' = c\boldsymbol{u} + C\boldsymbol{v}.$$

where the "large" coefficient

$$C = \frac{1}{2}\left(\frac{1}{\sqrt{1-a}} + \frac{1}{\sqrt{1+a}}\right),$$

and there is a "small" admixture

$$c = \frac{1}{2}\left(\frac{1}{\sqrt{1+a}} - \frac{1}{\sqrt{1-a}}\right).$$

As we can see the new (orthogonal) vectors are formed from the old ones (non-orthogonal) by an *identical* (hence the name "*symmetric orthogonalization*") admixture of the old vectors, i.e. the contribution of $\boldsymbol{u}$ and $\boldsymbol{v}$ in $\boldsymbol{u}'$ is the same as that of $\boldsymbol{v}$ and $\boldsymbol{u}$ in $\boldsymbol{v}'$.

> The new vectors are obtained by correcting the directions of the old ones, each by the same angle.

This is illustrated in Fig. J.2.

---

[5] They are symmetric matrices. For example,

$$(\boldsymbol{S}^{\frac{1}{2}})_{ij} = (\boldsymbol{U}\boldsymbol{S}_{\text{diag}}^{\frac{1}{2}}\boldsymbol{U}^{\dagger})_{ij} = \sum_k\sum_l U_{ik}(\boldsymbol{S}_{\text{diag}}^{\frac{1}{2}})_{kl}U_{jl} = \sum_k\sum_l U_{ik}(\boldsymbol{S}_{\text{diag}}^{\frac{1}{2}})_{kl}\delta_{kl}U_{jl}$$

$$= \sum_k U_{ik}(\boldsymbol{S}_{\text{diag}}^{\frac{1}{2}})_{kk}U_{jk} = (\boldsymbol{S}^{\frac{1}{2}})_{ji}.$$

[6] We see that if the vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ were already orthogonal, i.e. $a = 0$, then $\boldsymbol{u}' = \boldsymbol{u}$ and $\boldsymbol{v}' = \boldsymbol{v}$. Of course, we like this result.

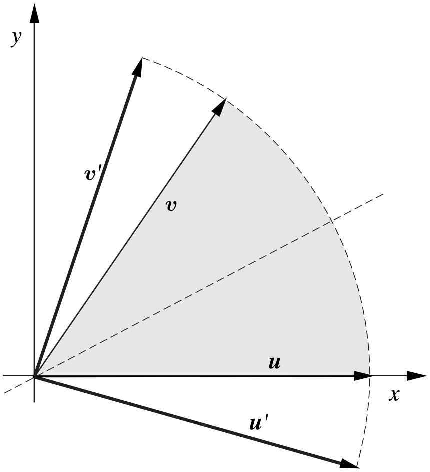


Fig. J.2. The symmetric (or Löwdin's) orthogonalization of the normalized vectors $\boldsymbol{u}$ and $\boldsymbol{v}$. The vectors are just pushed away by the same angle in such a way as to ensure $\boldsymbol{u}'$ and $\boldsymbol{v}'$ become orthogonal.

The new vectors automatically have length 1, the same as the starting vectors.

K. DIAGONALIZATION OF A MATRIX

In quantum chemistry we often encounter the following mathematical problem.

We have a Hermitian[1] matrix $\boldsymbol{A}$ (of dimension n), i.e. $\boldsymbol{A}^\dagger = \boldsymbol{A}$, and are interested in all numbers λ (called "eigenvalues"[2]) and the corresponding column vectors ("eigenvectors" of dimension n) $\boldsymbol{L}$, that satisfy the following equation

$$(\boldsymbol{A} - \lambda \mathbf{1})\boldsymbol{L} = \mathbf{0}, \tag{K.1}$$

where $\mathbf{1}$ is the unit matrix (of dimension n). There are n solutions to the last equation: n eigenvalues of λ and also n eigenvectors $\boldsymbol{L}$. Some eigenvalues λ may be equal (degeneracy), i.e. two or more linearly independent eigenvectors $\boldsymbol{L}$ correspond to a single eigenvalue λ. From (K.1) it is shown that any vector $\boldsymbol{L}$ is determined only to the accuracy of a multiplicative factor.[3] This is why, in future, there will be justification for normalizing them to unity.

In quantum chemistry the eigenvalue problem is solved in two ways: one is easy for $n \leqslant 2$, but more and more difficult for larger n, the second (using computers) treats all cases uniformly.

- The first way sees the eigenvalue equation as a set of linear homogeneous equations for the unknown components of vector $\boldsymbol{L}$. Then the condition for the non-trivial solution[4] to exist is: $\det(\boldsymbol{A} - \lambda\mathbf{1}) = 0$. This condition can be fulfilled only for some particular values of λ, which are to be found by expanding the determinant and solving the resulting n-th degree polynomial equation for λ. Then each solution λ is inserted into eq. (K.1) and the components of the corresponding vector $\boldsymbol{L}$ are found using any method applicable to linear equations. Thus, we end up with λ_k and $\boldsymbol{L}_k$ for $k = 1, 2, 3, \ldots, n$.
- The second way is based on diagonalization of $\boldsymbol{A}$.

First, let us show that *the same* λ's satisfy the eigenvalue equation (K.1), but with a much simpler matrix. To this end let us multiply (K.1) by (at the moment)

---

[1] In practice, matrix $\boldsymbol{A}$ is usually real, and therefore satisfies $(\boldsymbol{A}^T)^* = \boldsymbol{A}^T = \boldsymbol{A}$, i.e. $\boldsymbol{A}$ is symmetric.

[2] They are real.

[3] In other words, a unnormalized wave function still satisfies the Schrödinger equation, or an arbitrary amplitude can be assigned to any normal mode.

[4] The trivial one is obviously $\boldsymbol{L} = \mathbf{0}$, which is however unacceptable, since the wave function cannot vanish everywhere, or atoms have to vibrate, etc.

the *arbitrary* non-singular[5] square matrix[6] $\boldsymbol{B}$. We obtain the following chain of transformations $\boldsymbol{B}^{-1}(\boldsymbol{A}-\lambda\mathbf{1})\boldsymbol{L}=\boldsymbol{B}^{-1}(\boldsymbol{A}\boldsymbol{B}\boldsymbol{B}^{-1}-\lambda\mathbf{1})\boldsymbol{L}=(\boldsymbol{B}^{-1}\boldsymbol{A}\boldsymbol{B}-\lambda\mathbf{1})\boldsymbol{B}^{-1}\boldsymbol{L}=(\tilde{\boldsymbol{A}}-\lambda\mathbf{1})\tilde{\boldsymbol{L}}=\mathbf{0}$, where[7] $\tilde{\boldsymbol{A}}=\boldsymbol{B}^{-1}\boldsymbol{A}\boldsymbol{B}$, and $\tilde{\boldsymbol{L}}=\boldsymbol{B}^{-1}\boldsymbol{L}$. Thus, another matrix and other eigenvectors, but *the same* λ's! Now, let us choose such a special $\boldsymbol{B}$ so as to have the resulting equation as simple as possible, i.e. with a diagonal $\tilde{\boldsymbol{A}}$. Then we will know,[8] what the λ values have to be in order to satisfy the equation $(\tilde{\boldsymbol{A}}-\lambda\mathbf{1})\tilde{\boldsymbol{L}}=\mathbf{0}$.

Indeed, if $\tilde{\boldsymbol{A}}$ were diagonal, then

$$\det(\tilde{\boldsymbol{A}}-\lambda\mathbf{1})=\prod_{k=1}^{n}(\tilde{A}_{kk}-\lambda)=0,$$

which gives the solution $\lambda_k=\tilde{A}_{kk}$. Then, it is easy to find the corresponding vector $\tilde{\boldsymbol{L}}_k$. For example, $\tilde{\boldsymbol{L}}_1$ we find from equation $(\tilde{\boldsymbol{A}}-\lambda_1\mathbf{1})\tilde{\boldsymbol{L}}_1=\mathbf{0}$ in the following way:[9]

$$\begin{pmatrix} 0 & 0 & 0 & \dots & 0 \\ 0 & \tilde{A}_{22}-\lambda_1 & 0 & \dots & 0 \\ 0 & 0 & \tilde{A}_{33}-\lambda_1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & \tilde{A}_{nn}-\lambda_1 \end{pmatrix}\begin{pmatrix} \tilde{L}_{1,1} \\ \tilde{L}_{1,2} \\ \tilde{L}_{1,3} \\ \dots \\ \tilde{L}_{1,n} \end{pmatrix}=\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix},$$

which means that to get $\mathbf{0}$ on the right side, we have to have an *arbitrary* $\tilde{L}_{1,1}$, while the other $\tilde{L}_{1,j}=0$ for $j=2,3,\dots,n$.

To have the length of $\tilde{\boldsymbol{L}}_1$ equal to 1, it is sufficient to put $\tilde{L}_{1,1}=1$. Similarly, we easily find that the vectors $\tilde{\boldsymbol{L}}_k$ corresponding to λ_k simply represent the column vectors with all components equal to 0 except component k, which equals 1. We are interested in vectors $\boldsymbol{L}$, rather than $\tilde{\boldsymbol{L}}$. We get these vectors from $\boldsymbol{L}=\boldsymbol{B}\tilde{\boldsymbol{L}}$, and taking into account the form of $\tilde{\boldsymbol{L}}$, this means that $\boldsymbol{L}_k$ is nothing else but the k-th column of matrix $\boldsymbol{B}$. Since $\boldsymbol{B}$ is known, because this is precisely the matrix which led to the diagonalization, there is therefore no problem with $\boldsymbol{L}$:

> the columns of $\boldsymbol{B}$ represent the eigenvectors $\boldsymbol{L}$ of the equation $(\boldsymbol{A}-\lambda\mathbf{1})\boldsymbol{L}=\mathbf{0}$.

This is it.

---

[5] That is, its inverse matrix exists.

[6] To be found.

[7] Such a *unitary matrix* $\boldsymbol{B}$ (i.e. satisfying $(\boldsymbol{B}^T)^*=\boldsymbol{B}^{-1}$) can be found, that $\boldsymbol{B}^{-1}\boldsymbol{A}\boldsymbol{B}$ is *real and diagonal*. When (as is the case in most applications) we have to do with real matrices, then instead of unitary and Hermitian matrices, we have to do with orthogonal and symmetric matrices, respectively.

[8] Just by looking.

[9] The λ has been replaced by λ_1, because we are interested in getting $\tilde{\boldsymbol{L}}_1$.

L. SECULAR EQUATION $(H - \varepsilon S)c = 0$

A typical ε approach for solving an eigenvalue problem is its "algebraization", i.e. representation of the wave function as a linear combination of the known basis functions with the unknown coefficients. Then instead of searching for a function, we try to find the expansion coefficients $\boldsymbol{c}$ from the secular equation[1] $(\boldsymbol{H} - \varepsilon\boldsymbol{S})\boldsymbol{c} = 0$. Our goal is to reduce this task to the eigenvalue problem of a matrix. If the basis set used is orthonormal, the goal would be immediately achieved, because the secular equation would be reduced to $(\boldsymbol{H} - \varepsilon\boldsymbol{1})\boldsymbol{c} = 0$, i.e. the eigenvalue problem. However, in most cases the basis set used is not orthonormal. We may, however, orthonormalize the basis. We will achieve this using symmetric orthogonalization (see Appendix J, p. 977).

Instead of the old basis set (collected in the vector $\boldsymbol{\phi}$), in which the matrices $\boldsymbol{H}$ and $\boldsymbol{S}$ were calculated: $H_{ij} = \langle\phi_i|\hat{H}\phi_j\rangle$, $S_{ij} = \langle\phi_i|\phi_j\rangle$ we will use the orthogonal basis set $\boldsymbol{\phi}' = \boldsymbol{S}^{-\frac{1}{2}}\boldsymbol{\phi}$, where $\boldsymbol{S}^{-\frac{1}{2}}$ is calculated as described in Appendix J. Then we multiply the secular equation $(\boldsymbol{H} - \varepsilon\boldsymbol{S})\boldsymbol{c} = \boldsymbol{0}$ from the left by $\boldsymbol{S}^{-\frac{1}{2}}$ and make the following transformations

$$\begin{aligned}
\big(\boldsymbol{S}^{-\frac{1}{2}}\boldsymbol{H} - \varepsilon\boldsymbol{S}^{-\frac{1}{2}}\boldsymbol{S}\big)\boldsymbol{c} &= 0,\\
\big(\boldsymbol{S}^{-\frac{1}{2}}\boldsymbol{H}\boldsymbol{1} - \varepsilon\boldsymbol{S}^{-\frac{1}{2}}\boldsymbol{S}\big)\boldsymbol{c} &= 0,\\
\big(\boldsymbol{S}^{-\frac{1}{2}}\boldsymbol{H}\boldsymbol{S}^{-\frac{1}{2}}\boldsymbol{S}^{\frac{1}{2}} - \varepsilon\boldsymbol{S}^{-\frac{1}{2}}\boldsymbol{S}\big)\boldsymbol{c} &= 0,\\
\big(\boldsymbol{S}^{-\frac{1}{2}}\boldsymbol{H}\boldsymbol{S}^{-\frac{1}{2}}\boldsymbol{S}^{\frac{1}{2}} - \varepsilon\boldsymbol{S}^{\frac{1}{2}}\big)\boldsymbol{c} &= 0,\\
\big(\boldsymbol{S}^{-\frac{1}{2}}\boldsymbol{H}\boldsymbol{S}^{-\frac{1}{2}} - \varepsilon\boldsymbol{1}\big)\boldsymbol{S}^{\frac{1}{2}}\boldsymbol{c} &= 0,\\
\big(\tilde{\boldsymbol{H}} - \varepsilon\boldsymbol{1}\big)\tilde{\boldsymbol{c}} &= 0
\end{aligned}$$

with $\tilde{\boldsymbol{H}} = \boldsymbol{S}^{-\frac{1}{2}}\boldsymbol{H}\boldsymbol{S}^{-\frac{1}{2}}$ and $\tilde{\boldsymbol{c}} = \boldsymbol{S}^{\frac{1}{2}}\boldsymbol{c}$.

The new equation represents the eigenvalue problem, which we solve by diagonalization of $\tilde{\boldsymbol{H}}$ (Appendix K, p. 982). Thus,

> the equation $(\boldsymbol{H} - \varepsilon\boldsymbol{S})\boldsymbol{c} = \boldsymbol{0}$ is equivalent to the eigenvalue problem $(\tilde{\boldsymbol{H}} - \varepsilon\boldsymbol{1})\tilde{\boldsymbol{c}} = \boldsymbol{0}$. To obtain $\tilde{\boldsymbol{H}}$, we have to diagonalize $\boldsymbol{S}$ to calculate $\boldsymbol{S}^{\frac{1}{2}}$ and $\boldsymbol{S}^{-\frac{1}{2}}$.

---

[1] See Chapter 5.

Secular equation and normalization

If we used non-normalized basis functions in the Ritz method, this would not change the eigenvalues obtained from the secular equation. The only thing that would change are the eigenvectors. Indeed, imagine we have solved the secular equation for the normalized basis set functions: $(\boldsymbol{H} - \varepsilon \boldsymbol{S})\boldsymbol{c} = \mathbf{0}$. The eigenvalues ε have been obtained from the secular determinant $\det(\boldsymbol{H} - \varepsilon \boldsymbol{S}) = 0$. Now we wish to destroy the normalization and take new basis functions, which are the old basis set functions multiplied by some numbers, the i-th function by a_i. Then a new overlap integral and the corresponding matrix element of the Hamiltonian $\hat{H}$ would be $S'_{ij} = a_i a_j S_{ij}$, $H'_{ij} = a_i a_j H_{ij}$. The new secular determinant $\det(\boldsymbol{H}' - \varepsilon \boldsymbol{S}')$ may be expressed by the old secular determinant times a number.[2] This number is irrelevant, since what matters is that the determinant is equal to 0. Thus, whether in the secular equation we use the normalized basis set or not, the eigenvalues do not change. The eigenfunctions are also identical, although the eigenvectors $\boldsymbol{c}$ are different – they have to be, because they multiply different functions (which are proportional to each other).

If we ask whether the eigenvalues of the matrices $\boldsymbol{H}$ are $\boldsymbol{H}'$ identical, the answer would be: no.[3] However, in quantum chemistry we do not calculate the eigenvalues[4] of $\boldsymbol{H}$, but solve the secular equation $(\boldsymbol{H}' - \varepsilon \boldsymbol{S}')\boldsymbol{c} = \mathbf{0}$. If $\boldsymbol{H}'$ changes with respect to $\boldsymbol{H}$, there is a corresponding change of $\boldsymbol{S}'$ when compared to $\boldsymbol{S}$. This guarantees that the ε's do not change.

---

[2] We divide the new determinant by a_1, which means dividing the elements of the first row by a_1 and in this way removing from them a_1, both in $\boldsymbol{H}'$ and in $\boldsymbol{S}'$. Doing the same with a_2 and the second row, etc., and then repeating the procedure for columns (instead of rows), we finally get the old determinant times a number.

[3] This is evident, just think of diagonal matrices.

[4] Although we often say it this way.

M. SLATER–CONDON RULES

The Slater determinants represent something like the daily bread of quantum chemists. Our goal is to learn how to use the Slater determinants when they are involved in the calculation of the mean values or the matrix elements of some important operators. We will need this in the Hartree–Fock method, as well as in other important methods of quantum chemistry.

Only the final results of the derivations presented in this Appendix are the most important.

Antisymmetrization operator

The antisymmetrization operator is defined as

$$\hat{A} = \frac{1}{N!}\sum_P (-1)^p \hat{P}, \tag{M.1}$$

where $\hat{P}$ represents the permutation operator of N objects (in our case – electrons), while $(-1)^p$ stands for the parity of the permutation P, "even" ("odd") – if a given permutation P can be created from an even (odd) number p of transpositions (i.e. exchanges) of two elements.

The operator $\hat{A}$ has some nice features. The most important is that, when applied to *any* function, it produces either a function that is antisymmetric with respect to the permutations of N elements, or zero.[1] This means that $\hat{A}$ represents a sort of magic wand: whatever it touches it makes antisymmetric or causes it disappear! The antisymmetrizer is also idempotent, i.e. does not change any function that is already antisymmetric, which means $\hat{A}^2 = \hat{A}$.

Let us check that $\hat{A}$ is indeed idempotent. First we obtain:

$$\hat{A}^2 = (N!)^{-1}\sum_P (-1)^p \hat{P} (N!)^{-1}\sum_P (-1)^p \hat{P} = (N!)^{-2}\sum_{PP'} (-1)^{p+p'} \hat{P}\hat{P}'. \tag{M.2}$$

Of course $\hat{P}\hat{P}'$ represents a permutation opera tor,[2] which is then multiplied by its own parity $(-1)^{p+p'}$ and there is a sum over such permutations at a given fixed $\hat{P}'$.

---

[1] In the near future these elements will be identified with the electronic coordinates (one element will be represented by the space and spin coordinates of a single electron: x, y, z, σ).

[2] The permutations form the permutation group.

From "*Solid State and Molecular Theory*", Wiley, London, 1975 by John Slater on the permutation group: "*(...) It was at this point that Wigner, Hund, Heitler and Weyl entered the picture, with their "Grup-*

Independently of what $\hat{P}'$ is *we obtain the same result*[3] $N!$ times, and therefore:

$$\hat{A}^2 = (N!)^{-2} N! \sum_P (-1)^p \hat{P} = \hat{A}.$$

This is what we wanted to show.

The operator $\hat{A}$ is Hermitian. Since $\hat{P}$ represents a (permutational) symmetry operator, it therefore conserves the scalar product. This means that for the two vectors ψ_1 and ψ_2 of the Hilbert space we obtain[4]

$$\begin{aligned}&\langle \psi_1(1,2,\ldots,N) | \hat{A}\psi_2(1,2,\ldots,N)\rangle \\ &\quad = (N!)^{-1} \sum_P (-1)^p \langle \hat{P}^{-1}\psi_1(1,2,\ldots,N) | \psi_2(1,2,\ldots,N)\rangle.\end{aligned}$$

The summation over $\hat{P}$ can be replaced by the summation over $\hat{P}^{-1}$:

$$(N!)^{-1} \sum_{P^{-1}} (-1)^p \langle \hat{P}^{-1}\psi_1(1,2,\ldots,N) | \psi_2(1,2,\ldots,N)\rangle.$$

Since the parity p of the permutation $\hat{P}^{-1}$ is the same as that of $\hat{P}$, hence $(N!)^{-1}\sum_{P^{-1}}(-1)^p \hat{P}^{-1} = \hat{A}$, what shows that $\hat{A}$ is Hermitian: $\langle\psi_1|\hat{A}\psi_2\rangle = \langle\hat{A}\psi_1|\psi_2\rangle$, or[5]

$$\hat{A}^\dagger = \hat{A}. \tag{M.3}$$

Slater–Condon rules

The Slater–Condon rules serve to express the matrix elements involving the Slater determinants (which represent many-electron wave functions):

---

penpest": the pest of group theory, as certain disgruntled individuals who had never studied group theory in school described it. (…) The authors of the "Gruppenpest" wrote papers, which were incomprehensible to those like me who had not studied group theory (…). The practical consequences appeared to be negligible, but everyone felt that to be in the mainstream of quantum mechanics, we had to learn about it. (…) It was a frustrating experience, worthy of the name of a pest".

[3] Of course, $\hat{P}\hat{P}' = \hat{P}''$ has the parity $(-1)^{p+p'}$, because this is how such a permutation parity is to be calculated: first we make p transpositions to get $\hat{P}$, and next making p' transpositions we obtain the permutation $\hat{P}\hat{P}'$. Note that when keeping $\hat{P}'$ fixed and taking $\hat{P}$ from all possible permutations, we are running with $\hat{P}\hat{P}'$ over all possible permutations as well. This is because the complete set of permutations is obtained independently of what the starting permutation looks like, i.e. independently of $\hat{P}'$.

[4] The conservation of the scalar product $\langle\psi_1|\psi_2\rangle = \langle\hat{P}\psi_1|\hat{P}\psi_2\rangle$ means that the lengths of the vectors ψ_1 and $\hat{P}\psi_1$ are the same (similarly with ψ_2), and that the angle between the vectors is also conserved. If $\hat{P}$ is acting on ψ_2 alone, and ψ_1 does not change, the angle resulting from the scalar product $\langle\psi_1|\hat{P}\psi_2\rangle$ is of course different, because only one of the vectors (ψ_2) has been transformed (which means the rotation of a unit vector in the Hilbert space). *The same angle* would be obtained, if its partner ψ_1 were transformed in the *opposite* direction, i.e. when the operation $\hat{P}^{-1}\psi_1$ has been performed. Hence from the equality of the angles we have $\langle\psi_1|\hat{P}\psi_2\rangle = \langle\hat{P}^{-1}\psi_1|\psi_2\rangle$.

[5] $\hat{A}^\dagger$ stands for the *adjoint* operator with respect to $\hat{A}$, i.e. for arbitrary functions belonging to its domain we have $\langle\psi_1|\hat{A}\psi_2\rangle = \langle\hat{A}^\dagger\psi_1|\psi_2\rangle$. There is a subtle difference (ignored in the present book) among the self-adjoint ($\hat{A}^\dagger = \hat{A}$) and Hermitian operators in mathematical physics (they differ by definition of their domains).

$$\Psi = \frac{1}{\sqrt{N!}}\begin{vmatrix} \phi_1(1) & \phi_1(2) & \dots & \phi_1(N) \\ \phi_2(1) & \phi_2(2) & \dots & \phi_2(N) \\ \dots & \dots & \dots & \dots \\ \phi_N(1) & \phi_N(2) & \dots & \phi_N(N) \end{vmatrix}. \tag{M.4}$$

The normalized Slater determinant has the form: $\Psi = \sqrt{N!}\hat{A}(\phi_1\phi_2\cdots\phi_N)$, where $\phi_1\phi_2\cdots\phi_N$ represents the product $\phi_1(1)\phi_2(2)\cdots\phi_N(N)$, and therefore, the normalization constant before the determinant itself $\det[\phi_1(1)\phi_2(2)\cdots\phi_N(N)]$ is equal to $(N!)^{-1/2}$.

Quantum chemists love Slater determinants, because they are built of *one-electron* "bricks" ϕ_i called the spinorbitals (we assume them orthonormal) and because any Slater determinant is *automatically* antisymmetric with respect to the exchange of the coordinates of any two electrons (shown as arguments of ϕ_i's), the factor $\frac{1}{\sqrt{N!}}$ ensures the normalization. At the same time any Slater determinant automatically satisfies the Pauli exclusion principle, because any attempt to use the same spinorbitals results in two rows being equal, and in consequence, having $\Psi = 0$ everywhere.[6]

Using Slater determinants gives quantum chemists a kind of comfort, since all the integrals which appear when calculating the matrix elements of the Hamiltonian are relatively simple. The most complicated ones contain the coordinates of two electrons.

WHAT KIND OF OPERATORS WILL WE BE DEALING WITH?

1. The sum of one-electron operators $\hat{F} = \sum_i \hat{h}(i)$.
2. The sum of two-electron operators $\hat{G} = \sum_{i<j} \hat{g}(i,j)$.

In both cases the summation goes over all the electrons. Note that $\hat{h}$ has the identical form independent of the particular electron; the same pertains to $\hat{g}$.

The future meaning of the $\hat{F}$ and $\hat{G}$ operators is quite obvious, the first pertains to the non-interacting electrons (electronic kinetic energy with $\hat{h}(i) = -\frac{1}{2}\Delta_i$ or the interaction of the electrons with the nuclei), the second operator deals with the electronic repulsion, with $\hat{g}(i,j) = \frac{1}{r_{ij}}$.

WHAT ARE THE SLATER–CONDON RULES ALL ABOUT?

The Slater–Condon rules show how to express the matrix elements of many-electron operators $\hat{F}$ and $\hat{G}$ with the Slater determinants by *the matrix elements of the operators $\hat{h}$ and $\hat{g}$ calculated with orthonormal spinorbitals ϕ_i.*

---

[6] Which is a kind of catastrophe in theory: because our system *is somewhere* and can be found there with a certain non-zero probability.

The operators $\hat{F}$ and $\hat{G}$ are invariant with respect to any permutation of the electrons (Chapter 2). In other words, the formulae for $\hat{F}$ and $\hat{G}$ do not change before and after any relabelling of the electrons. This means that any permutation operator commutes with $\hat{F}$ and $\hat{G}$. Since $\hat{A}$ is a linear combination of such commuting operators, then $\hat{A}\hat{F} = \hat{F}\hat{A}$ and $\hat{A}\hat{G} = \hat{G}\hat{A}$.

A simple trick used in the proofs below

All the proofs given below are based on the same simple trick. First, the integral under consideration is transformed into the sum of the following terms $\langle\phi_1(1)\phi_2(2)\cdots\phi_N(N)|\hat{A}\hat{X}|\phi_1(1)\phi_2(2)\cdots\phi_N(N)\rangle$, where $\hat{X} = \hat{h}(i)$ or $\hat{g}(i,j)$. Then we recall that $\hat{A}$ is a linear combination of the permutation operators, and that in the integral $\langle\phi_1(1)\phi_2(2)\cdots\phi_N(N)|\hat{X}|\phi_{n_1}(1)\phi_{n_2}(2)\cdots\phi_{n_N}(N)\rangle$ only a few terms will survive.

- In the case $\hat{X} = \hat{h}(i)$ we obtain a product of one-electron integrals

  $$\begin{aligned}&\langle\phi_1(1)\phi_2(2)\cdots\phi_N(N)|\hat{X}|\phi_{n_1}(1)\phi_{n_2}(2)\cdots\phi_{n_N}(N)\rangle\\ &\quad=\langle\phi_1(1)|\phi_{n_1}(1)\rangle\langle\phi_2(2)|\phi_{n_2}(2)\rangle\cdots\langle\phi_i(i)|\hat{h}(i)|\phi_{n_i}(i)\rangle\cdots\langle\phi_N(N)|\phi_{n_N}(N)\rangle.\end{aligned}$$

  Since the spinorbitals are orthonormal, only one term will survive, the one which has $(n_1, n_2, \ldots, n_{i-1}, n_{i+1}, \ldots, n_N) = (1, 2, \ldots, i-1, i+1, \ldots, N)$. All the overlap integrals which appear there are equal to 1. Only one of the one-electron integrals will give something else: $\langle\phi_i(i)|\hat{h}(i)|\phi_{n_i}(i)\rangle$, but in this integral also we have to have $n_i = i$, because of the overlap integrals which force the matching of the indices mentioned above.
- In the case $\hat{X} = \hat{g}(i,j)$ we make the same transformations, but the rule for survival of the integrals pertains to the *two*-electron integral which involves the coordinates of the electrons i and j (not one-electron as before). Note that this time we will have some *pairs* of integrals which are going to survive, because the exchange of indices $ij \to ji$ also makes an integral survive.

I Slater–Condon rule

If ψ represents a normalized Slater determinant, then

$$\overline{F} = \langle\psi|\hat{F}|\psi\rangle = \sum_{i=1}^{N}\langle i|\hat{h}|i\rangle, \tag{M.5}$$

$$\overline{G} = \langle\psi|\hat{G}|\psi\rangle = \frac{1}{2}\sum_{i,j}\big(\langle ij|ij\rangle - \langle ij|ji\rangle\big), \tag{M.6}$$

where

$$\langle i|\hat{h}|r\rangle \equiv \sum_{\sigma_1}\int \phi_i^*(1)\hat{h}(1)\phi_r(1)\,\mathrm{d}V_1, \tag{M.7}$$

$$\langle ij|kl\rangle \equiv \sum_{\sigma_1}\sum_{\sigma_2}\iint \phi_i^*(1)\phi_j^*(2)g(1,2)\phi_k(1)\phi_l(2)\,\mathrm{d}V_1\,\mathrm{d}V_2, \tag{M.8}$$

where the summation pertains to two spin coordinates (for electrons 1 and 2).

Proof: **Operator $\hat{F}$.**

$$\overline{F} = \langle\psi|\hat{F}|\psi\rangle = N!\langle\hat{A}(\phi_1\phi_2\cdots\phi_N)|\hat{F}|\hat{A}(\phi_1\phi_2\cdots\phi_N)\rangle.$$

Using $\hat{A}\hat{F}=\hat{F}\hat{A}$, $\hat{A}^\dagger=\hat{A}$ and $\hat{A}^2=\hat{A}$ we get

$$\begin{aligned}\overline{F} &= N!\langle\phi_1\phi_2\cdots\phi_N|\hat{A}[(\hat{h}(1)\phi_1\phi_2\cdots\phi_N)+\cdots+(\phi_1\phi_2\cdots\hat{h}(N)\phi_N)]\rangle\\ &= \frac{N!}{N!}\langle\phi_1\phi_2\cdots\phi_N|[(\hat{h}(1)\phi_1\phi_2\cdots\phi_N)+\cdots+(\phi_1\phi_2\cdots\hat{h}(N)\phi_N)]\rangle,\end{aligned}$$

because what gives the non-zero contribution from the antisymmetrizer $\hat{A} = (N!)^{-1}(1 + \text{other permutations})$ is only the first term with the operator of multiplication by 1. Other terms disappear after any attempt at integration. As a result we have:

$$\overline{F} = \langle\phi_1|\hat{h}|\phi_1\rangle + \langle\phi_2|\hat{h}|\phi_2\rangle + \cdots + \langle\phi_N|\hat{h}|\phi_N\rangle = \sum_i h_{ii}, \tag{M.9}$$

which is what we wanted to show.

Operator $\hat{G}$. Now let us consider the expression for $\overline{G}$

$$\overline{G} = N!\langle\hat{A}(\phi_1\phi_2\cdots\phi_N)|\hat{G}|\hat{A}(\phi_1\phi_2\cdots\phi_N)\rangle,$$

where once again $N!$ comes from the normalization of ψ. Taking (as above) into account that $\hat{A}^\dagger=\hat{A}$, $\hat{A}^2=\hat{A}$, $\hat{G}\hat{A}=\hat{A}\hat{G}$, we get

$$\begin{aligned}\overline{G} &= N!\langle(\phi_1\phi_2\cdots\phi_N)|\hat{A}|[\hat{g}(1,2)\phi_1\phi_2\cdots\phi_N + \hat{g}(1,3)\phi_1\phi_2\cdots\phi_N+\cdots]\rangle\\ &= \langle\phi_1(1)\phi_2(2)|\hat{g}(1,2)|\phi_1(1)\phi_2(2)\rangle\\ &\quad -\langle\phi_1(1)\phi_2(2)|\hat{g}(1,2)|\phi_2(1)\phi_1(2)\rangle\\ &\quad +\langle\phi_1(1)\phi_3(3)|\hat{g}(1,3)|\phi_1(1)\phi_3(3)\rangle\\ &\quad -\langle\phi_1(1)\phi_3(3)|\hat{g}(1,3)|\phi_3(1)\phi_1(3)\rangle+\cdots\end{aligned} \tag{M.10}$$

This transformation needs some explanation. The factor $N!$ before the integral is annihilated by $1/N!$ coming from the antisymmetrizer. The remainder of the antisymmetrizer permutes the electrons in the ket $|[\hat{g}(1,2)\phi_1\phi_2\cdots\phi_N + \hat{g}(1,3)\phi_1\phi_2\cdots\phi_N+\cdots]\rangle$. In the first term [with $\hat{g}(1,2)$] the integrals with only those permutations of electrons $3,4,\ldots,N$ will survive which perfectly match the permutation $\phi_1(1)\phi_2(2)\cdots\phi_N(N)$, because otherwise the overlap integrals of the spinorbitals (over the coordinates of the electrons $2,3,\ldots,N$) will make them zero. This is why the first term will give rise to only *two* permutations which result in non-zero integrals: in the first two positions we will have $\phi_1(1)\phi_2(2)$, and in the other $\phi_1(2)\phi_2(1)$. Of course, they will differ by sign, and this is why we have the minus sign in the second surviving integral. Similar reasoning may be followed for the term with $\hat{g}(1,3)$, as well as for the other terms.

Thus, we have shown that

$$\overline{G} = \sum_{i<j} \big(\langle ij|ij\rangle - \langle ij|ji\rangle\big) = \frac{1}{2}\sum_{i,j} \big(\langle ij|ij\rangle - \langle ij|ji\rangle\big), \tag{M.11}$$

the factor $\frac{1}{2}$ takes care of the fact that there are only $\frac{N(N-1)}{2}$ interelectronic interactions $g(i,j)$ (the upper triangle of table $N \times N$). There is no restriction in the summation over $i, j = 1, 2, \dots, N$, because any attempt to take the "illegal" self-interaction (corresponding to $i = j$) gives zero, because of the identity of the Coulomb ($\langle ij|ij\rangle$) and exchange ($\langle ij|ji\rangle$) integrals. This is the formula we wanted to prove.

A special case: double occupation

The integrals in the expressions for $\overline{F}$ and G contain spinorbitals and the integration goes over the electronic space-and-spin coordinates. When the spinorbitals are expressed by the orbitals and the spin functions, we may perform the summation over the spin coordinates. The double occupation case is the most popular and the most important, when every orbital is used to form two spinorbitals[7]

$$\begin{aligned} \phi_1(1) &= \varphi_1(1)\alpha(1), \\ \phi_2(1) &= \varphi_1(1)\beta(1), \\ \phi_3(1) &= \varphi_2(1)\alpha(1), \\ \phi_4(1) &= \varphi_2(1)\beta(1), \\ &\dots \end{aligned} \tag{M.12}$$

or

$$\begin{aligned} \phi_{2i-1}(1) &= \varphi_i(1)\alpha(1), \\ \phi_{2i}(1) &= \varphi_i(1)\beta(1), \end{aligned} \tag{M.13}$$

$i = 1, 2, \dots, N/2$.

Thus, the one electron spinorbitals which represent the building blocks of the Slater determinant, are products of a spatial function (orbital φ), and one of the two simple functions of the spin coordinate σ (α or β functions, cf. p. 28).

---

[7] The functions below are written as if they were dependent on the coordinates of electron number 1. The reason is that we want to stress that they all are *one-electron functions.* Electron 1 serves here as an example (and when needed may be replaced by the other electron). The symbol "1" means $(x_1, y_1, z_1, \sigma_1)$ if it is an argument of a spinorbital, (x_1, y_1, z_1) if it corresponds to an orbital, and σ_1 if it corresponds to a spin function.

The first Slater–Condon rule (M.9) may be transformed as follows (for definition of the integrals see p. 334)

$$\overline{F}=\sum_{i=1}^{N}\langle i|\hat{h}|i\rangle=\sum_{i=1}^{\mathrm{MO}}\sum_{\sigma}\langle i\sigma|\hat{h}|i\sigma\rangle=2\sum_{i}^{\mathrm{MO}}(i|\hat{h}|i)\equiv 2\sum_{i}^{\mathrm{MO}}h_{ii} \tag{M.14}$$

where the summations denoted by MO go over the occupied *orbitals* (their number being $N/2$), the factor 2 results from the summation over σ, which gives the same result for the two values of σ (because of the double occupation of the orbitals).

Let us perform a similar operation with $\overline{G}$. The formula for $\overline{G}$ is composed of two parts

$$\overline{G}=\mathrm{I}-\mathrm{II}. \tag{M.15}$$

The first part reads as

$$\mathrm{I}=\frac{1}{2}\sum_{i}^{\mathrm{MO}}\sum_{\sigma_i}\sum_{j}^{\mathrm{MO}}\sum_{\sigma_j}\langle i\sigma_i, j\sigma_j|i\sigma_i, j\sigma_j\rangle$$

where $i\sigma_i,\dots$ etc. stands for the spinorbital composed of the orbital φ_i and a spin function that depends on σ_i. For any pair of values of σ_i, σ_j, the integral yields the same value (at a given pair of i, j) and therefore (cf. p. 334),

$$\mathrm{I}=\frac{1}{2}\sum_{i}^{\mathrm{MO}}\sum_{j}^{\mathrm{MO}}4(ij|ij)=2\sum_{i}^{\mathrm{MO}}\sum_{j}^{\mathrm{MO}}(ij|ij).$$

The fate of part II will be a little different:

$$\mathrm{II}=\frac{1}{2}\sum_{i}^{\mathrm{MO}}\sum_{\sigma_i}\sum_{j}^{\mathrm{MO}}\sum_{\sigma_j}\langle i\sigma_i, j\sigma_j|j\sigma_j, i\sigma_i\rangle=\frac{1}{2}\sum_{i}^{\mathrm{MO}}\sum_{j}^{\mathrm{MO}}2(ij|ji)=\sum_{i}^{\mathrm{MO}}\sum_{j}^{\mathrm{MO}}(ij|ji),$$

because this time the summation over σ_i and σ_j gives a non-zero result in half the cases when compared to the previous case. The pairs $(\sigma_i,\sigma_j)=(\frac{1}{2},\frac{1}{2}),(-\frac{1}{2},-\frac{1}{2})$ give a non-zero (and the same) result, while $(\frac{1}{2},-\frac{1}{2}),(-\frac{1}{2},\frac{1}{2})$ end up with zero (recall that, by convention, the electrons in the integral have the order 1 2 1 2). Finally the double occupation leads to

$$\overline{G}=\sum_{i,j}^{\mathrm{MO}}\big[2(ij|ij)-(ij|ji)\big]. \tag{M.16}$$

II Slater–Condon rule

Suppose we are interested in two matrix elements: $F_{12}\equiv\langle\psi_1|\hat{F}|\psi_2\rangle$ and $G_{12}\equiv\langle\psi_1|\hat{G}|\psi_2\rangle$ and the two Slater determinants ψ_1 and ψ_2 differ only in that spinorbital ϕ_i in ψ_1 has been replaced by ϕ_i' (*orthogonal to all other spinorbitals*) in ψ_2. Then the Slater–Condon rule states that

$$F_{12} = \langle i|\hat{h}|i'\rangle, \tag{M.17}$$

$$G_{12} = \sum_{j=1} \big(\langle ij|i'j\rangle - \langle ij|ji'\rangle\big). \tag{M.18}$$

***Proof*: Operator $\hat{F}$.** Using $\hat{F}\hat{A} = \hat{A}\hat{F}$, $\hat{A}^\dagger = \hat{A}$ and $\hat{A}^2 = \hat{A}$, we obtain $\hat{A}^\dagger\hat{F}\hat{A} = \hat{A}\hat{F}\hat{A} = \hat{A}\hat{A}\hat{F} = \hat{A}\hat{F}$ and therefore

$$F_{12} = N!\langle\phi_1\cdots\phi_i\cdots|\hat{A}\hat{F}|\phi_1\cdots\phi_i'\cdots\phi_N\rangle.$$

$$\begin{aligned}
F_{12} &= N!\langle\phi_1\phi_2\cdots\phi_i\cdots\phi_N|\hat{A}|\big[\hat{h}(1)\phi_1\cdots\phi_i'\cdots\phi_N \\
&\quad + \phi_1\hat{h}(2)\phi_2\cdots\phi_i'\cdots\phi_N + \cdots + \phi_1\cdots\phi_i'\cdots\hat{h}(N)\phi_N\big]\rangle \\
&= \sum_P (-1)^p\langle\phi_1\phi_2\cdots\phi_i\cdots\phi_N|\hat{P}\big[\hat{h}(1)\phi_1\cdots\phi_i'\cdots\phi_N \\
&\quad + \phi_1\hat{h}(2)\phi_2\cdots\phi_i'\cdots\phi_N + \cdots + \phi_1\cdots\phi_i'\cdots\hat{h}(N)\phi_N\big]\rangle.
\end{aligned}$$

Note first that the only integral to survive should involve ϕ_i and ϕ_i' in such a way that it leads to the one-electron integral $\langle\phi_i|\hat{h}|\phi_i'\rangle$. This however happens only if the i-th term in the square bracket intervenes [that with $\hat{h}(i)$]. Indeed, let us take an integral which is *not* like that ($i \neq 1$): $\langle\phi_1\phi_2\cdots\phi_i\cdots\phi_N|\hat{P}\hat{h}(1)\phi_1\phi_2\cdots\phi_i'\cdots\phi_N\rangle$. Whatever permutation $\hat{P}$ is, $\hat{h}$ will always go with ϕ_1, while ϕ_i' will therefore be *without* $\hat{h}$. When integrating over the electronic coordinates we obtain the product of one-electron integrals (for subsequent electrons), and in this product we always pinpoint the *overlap* integral of ϕ_i' multiplied by one of the spinorbitals $\phi_1, \phi_2, \ldots, \phi_N$. This integral (and therefore the whole product) is equal to 0, because ϕ_i' is orthogonal to *all* the spinorbitals. Identical reasoning can be given for $\hat{h}(2), \hat{h}(3), \ldots$, but not for $\hat{h}(i)$, and we obtain:

$$F_{12} = \sum_P (-1)^p\langle\phi_1\phi_2\cdots\phi_i\cdots\phi_N|\hat{P}\big[\phi_1\phi_2\cdots\hat{h}(i)\phi_i'\cdots\phi_N\big]\rangle.$$

The only integral to survive is that which corresponds to $\hat{P} = 1$, because in other cases the orthogonality of the spinorbitals will make the product of the one-electron integrals equal to zero. Thus, finally we prove that

$$F_{12} = \langle i|h|i'\rangle. \tag{M.19}$$

Operator $\hat{G}$. From $\hat{A}^\dagger = \hat{A}$, $\hat{A}\hat{G}\hat{A} = \hat{A}\hat{A}\hat{G} = \hat{A}\hat{G}$ we obtain the following transformation

$$\begin{aligned}
G_{12} &= N!\langle\hat{A}(\phi_1\phi_2\cdots\phi_N)|\hat{A}\hat{G}|\phi_1\cdots\phi_i'\cdots\phi_N\rangle\rangle \\
&= N!\langle\hat{A}(\phi_1\phi_2\cdots\phi_N)|\{\big[\hat{g}(1,2)|\phi_1\cdots\phi_i'\cdots\phi_N\rangle\big]
\end{aligned}$$

$$+\left[\hat{g}(1,3)\middle|\phi_1\cdots\phi_i'\cdots\phi_N\right\rangle\Big]+\cdots\Big\}\Big\rangle$$
$$=\frac{1}{2}\sum_{k,l}{}'\sum_P(-1)^p\big\langle\hat{P}(\phi_1\cdots\phi_i\cdots\phi_N)\big|\hat{g}(k,l)\big|\phi_1\cdots\phi_i'\cdots\phi_N\big\rangle.$$

The number of g terms is equal to the number of interelectronic interactions. The prime in the summation $k,l=1,2,\dots,N$ over interactions $\hat{g}(k,l)$ means that $k\neq l$ (we count the interactions twice, but the factor $\frac{1}{2}$ takes care of that). Note that, due to the orthogonality of the spinorbitals, for a given $\hat{g}(k,l)$ the integrals are all zero if $k\neq i$ and $l\neq i$. Thus, the integrals to survive have to have $k=i$ or $l=i$. Therefore (prime in the summation means the summation index i is to be excluded),

$$\begin{aligned}
G_{12}&=\frac{1}{2}\sum_{l}{}'\sum_P(-1)^p\big\langle\hat{P}(\phi_1\cdots\phi_i\cdots\phi_N)\big|\hat{g}(i,l)\big|\phi_1\cdots\phi_i'\cdots\phi_N\big\rangle\\
&\quad+\frac{1}{2}\sum_{k}{}'\sum_P(-1)^p\big\langle\hat{P}(\phi_1\cdots\phi_i\cdots\phi_N)\big|\hat{g}(k,i)\big|\phi_1\cdots\phi_i'\cdots\phi_N\big\rangle\\
&=\frac{1}{2}\sum_{l}{}'\big[\langle\phi_i\phi_l|\phi_i'\phi_l\rangle-\langle\phi_i\phi_l|\phi_l\phi_i'\rangle\big]+\frac{1}{2}\sum_{k}{}'\big[\langle\phi_i\phi_k|\phi_i'\phi_k\rangle-\langle\phi_i\phi_k|\phi_k\phi_i'\rangle\big]\\
&=\sum_{j}{}'\big[\langle\phi_i\phi_j|\phi_i'\phi_j\rangle-\langle\phi_i\phi_j|\phi_j\phi_i'\rangle\big],
\end{aligned}$$

because only those two-electron integrals will survive which involve both ϕ_i *and* ϕ_i', and the two other spinorbitals involved are bound to be identical (and have either the index k or l depending on whether $l=i$ or $k=i$). The difference in the square brackets results from two successful permutations $\hat{P}$, in which we have the order i,j or j,i (in the last term). Finally, for the sake of simplicity leaving only the indices for the spinorbitals, we obtain

$$G_{12}=\sum_{j(\neq i)}\big[\langle ij|i'j\rangle-\langle ij|ji'\rangle\big] \tag{M.20}$$

and after adding $0=\langle ii|i'i\rangle-\langle ii|ii'\rangle$ we have[8]

$$G_{12}=\sum_{j}\big\{\langle ij|i'j\rangle-\langle ij|ji'\rangle\big\}. \tag{M.21}$$

This is our result.

---

[8] With this formula, we may forget at once that the integration has been carried out over the coordinates of electrons i and j. It does not matter what the symbol of the coordinate is over which an integration is performed in a definite integral. When in the future, we have to declare which coordinates we are going to integrate over in $\langle ij|i'j\rangle$, it is absolutely safe to put any electrons. In the present book it will be electron 1 and electron 2.

III Slater–Condon rule

If ψ_1 and ψ_2 differ by two spinorbitals, say, in ψ_1 are ϕ_i and ϕ_s, and in ψ_2 we have ϕ_i' and ϕ_s' (normalized and orthogonal to themselves and to all other spinorbitals), i.e. ϕ_i' replaces ϕ_i while ϕ_s' replaces ϕ_s (all other spinorbitals are of the same order), then

$$F_{12} = 0, \tag{M.22}$$

$$G_{12} = \langle is|i's'\rangle - \langle is|s'i'\rangle, \tag{M.23}$$

***Proof*: Operator $\hat{F}$.**

$$\begin{aligned} F_{12} &= N!\langle(\phi_1\phi_2\cdots\phi_N)|\hat{A}\hat{F}(\phi_1'\phi_2'\cdots\phi_N')\rangle \\ &= N!\langle(\phi_1\phi_2\cdots\phi_N)|\hat{A}\{(\hat{h}(1)\phi_1'\phi_2'\cdots\phi_N') + (\phi_1'\hat{h}(2)\phi_2'\cdots\phi_N') + \cdots \\ &\quad + (\phi_1'\phi_2'\cdots\hat{h}(N)\phi_N')\}\rangle \\ &= 0, \end{aligned}$$

where the spinorbitals in ψ_2 have been additionally labelled by primes (to stress that they *may* differ from those of ψ_1). In each term there will be $N-1$ overlap integrals between spinorbitals and one integral involving $\hat{h}$. Therefore, there will *always* be at least one overlap integral involving *different* spinorbitals. This will produce zero.

Operator $\hat{G}$. There will be something surviving in G_{12}. Using the previous arguments, we have

$$\begin{aligned} G_{12} &= N!\langle(\phi_1\phi_2\cdots\phi_N)|\hat{A}(g(1,2)\phi_1'\phi_2'\cdots\phi_N') + (g(1,3)\phi_1'\phi_2'\cdots\phi_N') + \cdots\rangle \\ &= \langle\phi_1\phi_2|g(1,2)|\phi_1'\phi_2'\rangle - \langle\phi_1\phi_2|g(1,2)|\phi_2'\phi_1'\rangle \\ &\quad + \langle\phi_1\phi_3|g(1,3)|\phi_1'\phi_3'\rangle - \langle\phi_1\phi_3|g(1,3)|\phi_3'\phi_1'\rangle + \cdots \\ &= \langle\phi_1\phi_2|\phi_1'\phi_2'\rangle - \langle\phi_1\phi_2|\phi_2'\phi_1'\rangle \\ &\quad + \langle\phi_1\phi_3|\phi_1'\phi_3'\rangle - \langle\phi_1\phi_3|\phi_3'\phi_1'\rangle + \cdots . \end{aligned}$$

Note that $N!$ cancels $1/N!$ from the antisymmetrizer, and in the ket we have all possible permutations. A term to survive, it has to engage all four spinorbitals: i, i', s, s', otherwise the overlap integrals will kill it. Therefore, only two terms will survive and give

$$G_{12} = \langle is|i's'\rangle - \langle is|s'i'\rangle. \tag{M.24}$$

IV Slater–Condon rule

Using the above technique it is easy to show that if the Slater determinants ψ_1 and ψ_2 differ by *more than two* (orthogonal) spinorbitals, the matrix elements $F_{12} = 0$

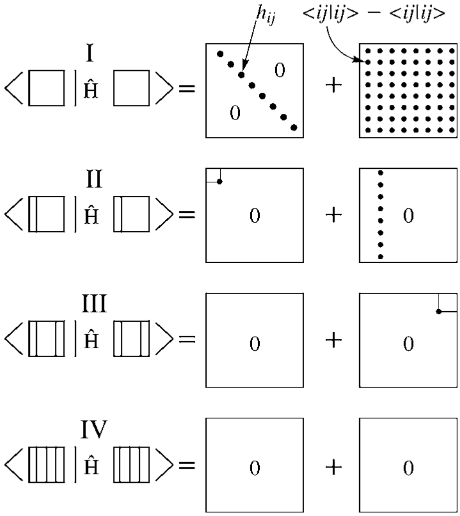


Fig. M.1. Four Slater–Condon rules (I, II, III, IV) for easy remembering. On the left side we see pictorial representations of matrix elements of the total Hamiltonian $\hat{H}$. The squares inside the brackets represent the Slater determinants. Vertical lines in bra stand for those spinorbitals, which are different in bra and in ket functions. On the right we have two square matrices collecting the h_{ij}'s and $\langle ij|ij\rangle - \langle ij|ji\rangle$ for $i, j = 1, \ldots, N$. The dots in the matrices symbolize non-zero elements.

and $G_{12} = 0$. This happens because operators $\hat{F}$ and $\hat{G}$ represent the sum of, at most, two-electron operators, which will involve at most four spinorbitals and there will always be an extra overlap integral over the orthogonal spinorbitals.[9]

The Slater–Condon rules are schematically depicted in Fig. M.1.

---

[9] If the operators were more than two-particle, the result would be different.

N. LAGRANGE MULTIPLIERS METHOD

Imagine a Cartesian coordinate system of $n+m$ dimensions with the axes labelled $x_1, x_2, \ldots, x_{n+m}$ and a function[1] $E(\boldsymbol{x})$, where $\boldsymbol{x} = (x_1, x_2, \ldots, x_{n+m})$. Suppose that we are interested in finding the lowest value of E, but only among such $\boldsymbol{x}$ that satisfy m conditions (*conditional extremum*):

conditional extremum

$$W_i(\boldsymbol{x}) = 0 \tag{N.1}$$

for $i = 1, 2, \ldots, m$. The constraints cause the number of *independent* variables to be n.

If we calculated the differential dE at point $\boldsymbol{x}_0$, which corresponds to an extremum of E, then we obtain 0:

$$0 = \sum_{j=1}^{n+m} \left(\frac{\partial E}{\partial x_j} \right)_0 \mathrm{d}x_j, \tag{N.2}$$

where the derivatives are calculated at the point of the extremum. The quantities dx_j stand for infinitesimally small increments. From (N.2) we cannot draw the conclusion that the $(\frac{\partial E}{\partial x_j})_0$ are equal to 0. This would be true if the increments dx_j were independent, but they are not. Indeed, we find the relations between them by making differentials of conditions W_i:

$$\sum_{j=1}^{n+m} \left(\frac{\partial W_i}{\partial x_j} \right)_0 \mathrm{d}x_j = 0 \tag{N.3}$$

for $i = 1, 2, \ldots, m$ (the derivatives are calculated for the extremum).

This means that the number of truly independent increments is only n. Let us try to exploit this. To this end let us multiply each equation (N.3) by a number ϵ_i (*Lagrange multiplier*), which will be fixed in a moment. Then, let us add together all the conditions (N.3), and subtract the result from eq. (N.2). We get

Joseph Louis de Lagrange (1736–1813), French mathematician of Italian origin, self-taught; professor at the Artillery School in Turin, then at the École Normale Supérieure in Paris. His main achievements are in variational calculus, mechanics, and also in number theory, algebra and mathematical analysis.

---

[1] Symbol E is chosen to suggest that, in our applications, the quantity will have the meaning of energy.

$$\sum_{j=1}^{n+m}\left[\left(\frac{\partial E}{\partial x_j}\right)_0-\sum_i \epsilon_i\left(\frac{\partial W_i}{\partial x_j}\right)_0\right]\mathrm{d}x_j=0,$$

where the summation extends over $n+m$ terms. The summation may be carried out in two steps. First, let us sum up the first n terms, and afterwards sum the other terms

$$\sum_{j=1}^{n}\left[\left(\frac{\partial E}{\partial x_j}\right)_0-\sum_i \epsilon_i\left(\frac{\partial W_i}{\partial x_j}\right)_0\right]\mathrm{d}x_j+\sum_{j=n+1}^{n+m}\left[\left(\frac{\partial E}{\partial x_j}\right)_0-\sum_i \epsilon_i\left(\frac{\partial W_i}{\partial x_j}\right)_0\right]\mathrm{d}x_j=0.$$

The multipliers ϵ_i have so far been treated as "undetermined". Well, we may force them to make each of the terms in the second summation equal zero[2]

$$\left(\frac{\partial E}{\partial x_j}\right)_0-\sum_i \epsilon_i\left(\frac{\partial W_i}{\partial x_j}\right)_0=0,\quad \text{for } j=n+1,\dots,n+m. \tag{N.4}$$

Hence, the first summation alone is 0

$$\sum_{j=1}^{n}\left[\left(\frac{\partial E}{\partial x_j}\right)_0-\sum_i \epsilon_i\left(\frac{\partial W_i}{\partial x_j}\right)_0\right]\mathrm{d}x_j=0,$$

which means that now we have only n increments $\mathrm{d}x_j$, and therefore they are *independent*. Since for any (small) $\mathrm{d}x_j$, the sum is always 0, the only reason for this could be that each parenthesis [ ] *individually equals zero*

$$\left(\frac{\partial E}{\partial x_j}\right)_0-\sum_i \epsilon_i\left(\frac{\partial W_i}{\partial x_j}\right)_0=0\quad \text{for } j=1,\dots,n.$$

Euler equation

This set of n equations (called the *Euler equations*) together with the m conditions (N.1) and m equations (N.4), gives a set of $n+2m$ equations with $n+2m$ unknowns (m epsilons and $n+m$ components x_i of the vector $\boldsymbol{x}_0$).

For a conditional extremum, the constraint $W_i(\boldsymbol{x})=0$ has to be satisfied for $i=1,2,\dots,m$ and

$$\left(\frac{\partial E}{\partial x_j}\right)_0-\sum_i \epsilon_i\left(\frac{\partial W_i}{\partial x_j}\right)_0=0\quad \text{for } j=1,\dots,n+m.$$

The x_i found from these equations determine the position $\boldsymbol{x}_0$ of the conditional extremum E.

Whether it is a minimum, a maximum or a saddle point, is decisive for the analysis of the matrix of the second derivative (Hessian). If its eigenvalues calculated at $\boldsymbol{x}_0$ are all positive (negative), it is a minimum[3] (maximum), in other cases it is a saddle point.

---

[2] This is possible if the determinant which is built of coefficients $(\frac{\partial W_i}{\partial x_j})_0$ is non-zero (this is what we have to assume). For example, if several conditions were identical, the determinant would be zero.

[3] In this way we find a minimum; no information is available as to whether is it global or local.

Example 1. ***Minimizing a paraboloid going along a straight line off centre.*** Let us take a paraboloid

$$E(x,y)=x^2+y^2.$$

This function has, of course, a minimum at $(0,0)$, but the minimum is of no interest to us. What we want to find is a minimum of E, but only *when* x and y satisfy some conditions. In our case there will only be one:

$$W=\frac{1}{2}x-\frac{3}{2}-y=0. \tag{N.5}$$

This means that we are interested in a minimum of E *along the straight line* $y=\frac{1}{2}x-\frac{3}{2}$.

The Lagrange multipliers method works as follows:

- We differentiate W and multiply by an unknown (Lagrange) multiplier ϵ thus getting: $\epsilon(\frac{1}{2}\,\mathrm{d}x-\mathrm{d}y)=0$.
- This result (i.e. 0) is subtracted[4] from $\mathrm{d}E=2x\,\mathrm{d}x+2y\,\mathrm{d}y=0$ and we obtain $\mathrm{d}E=2x\,\mathrm{d}x+2y\,\mathrm{d}y-\frac{1}{2}\varepsilon\,\mathrm{d}x+\varepsilon\,\mathrm{d}y=0$.
- In the last expression, the coefficients at $\mathrm{d}x$ and $\mathrm{d}y$ have to be equal to zero.[5] In this way we obtain two equations: $2x-\frac{1}{2}\epsilon=0$ and $2y+\epsilon=0$.
- The third equation needed is the constraint $y=\frac{1}{2}x-\frac{3}{2}$.
- The solution to these three equations gives a set of x,y,ϵ which corresponds to an extremum. We obtain: $x=\frac{3}{5}$, $y=-\frac{6}{5}$, $\epsilon=\frac{12}{5}$. Thus we have obtained, not only the position of the minimum ($x=\frac{3}{5}$, $y=-\frac{6}{5}$), but also the Lagrange multiplier ϵ. The minimum value of E, which has been encountered along the straight line $y=\frac{1}{2}x-\frac{3}{2}$ is equal to $E(\frac{3}{5},-\frac{6}{5})=(\frac{3}{5})^2+(-\frac{6}{5})^2=\frac{9+36}{25}=\frac{9}{5}$.

Example 2. ***Minimizing a paraboloid moving along a circle (off centre).*** Let us take the same paraboloid (N.5), but put another constraint

$$W=(x-1)^2+y^2-1=0. \tag{N.6}$$

This condition means that we want to go around a circle of radius 1, centred at $(1,0)$, and see at which point (x,y) we have the lowest value[6] of E. The example is chosen in such a way as to answer the question first *without any calculations*. Indeed, the circle goes through $(0,0)$, therefore, this point has to be found as the minimum. Beside this, we should find a maximum at $(2,0)$, because this is the point on the circle which is most distant from $(0,0)$.

Well, let us see whether the Lagrange multipliers method will give the same result.

After differentiation of W, multiplying it by the multiplier ϵ, subtracting the result from dE and rearranging the terms, we obtain

---

[4] Or added – no matter (in that case we get a different value of ϵ).

[5] This is only possible now.

[6] Or, in other words, we intersect the paraboloid with the cylindrical surface of radius 1 and the cylinder axis (parallel to the axis of symmetry of the paraboloid) is shifted to $(1,0)$.

$$\mathrm{d}E = \left[2x - \epsilon(2x-2)\right]\mathrm{d}x + 2y(1-\epsilon)\,\mathrm{d}y = 0,$$

which (after forcing the coefficients at dx and dy to be zero) gives a set of three equations

$$\begin{aligned} 2x - \epsilon(2x-2) &= 0,\\ 2y(1-\epsilon) &= 0,\\ (x-1)^2 + y^2 &= 1. \end{aligned}$$

Please check that this set has the following solutions: $(x, y, \epsilon) = (0, 0, 0)$ and $(x, y, \epsilon) = (2, 0, 2)$. The solution $(x, y) = (0, 0)$ corresponds to the minimum, while the solution $(x, y) = (2, 0)$ corresponds to the maximum.[7] This is what we expected to get.

Example 3. ***Minimizing the mean value of the harmonic oscillator Hamiltonian.*** This example is different: it will pertain to the extremum of a *functional*.[8] *We are often going to encounter this in the methods of quantum chemistry.* Let us take the energy functional

$$E[\phi] = \int_{-\infty}^{\infty} \mathrm{d}x\, \phi^* \hat{H} \phi \equiv \langle \phi | \hat{H} \phi \rangle,$$

where $\hat{H}$ stands for the harmonic oscillator Hamiltonian:

$$\hat{H} = -\frac{\hbar^2}{2m}\frac{\mathrm{d}^2}{\mathrm{d}x^2} + \frac{1}{2}kx^2.$$

If we were asked what function ϕ ensures the minimum value of $E[\phi]$, such a function could be found right away, it is $\phi = 0$. Yes, indeed, because the kinetic energy integral and the potential energy integral are positive numbers, except in the situation when $\phi = 0$, then the result is zero. Wait a minute! This is not what we thought of. We want ϕ to have a probabilistic interpretation, like any wave function, and therefore $\langle \phi | \phi \rangle = 1$, and not zero. Well, in such a case we want to minimize $E[\phi]$, *but forcing the normalization condition is satisfied all the time.* Therefore we search for the extremum of $E[\phi]$ *with condition* $W = \langle \phi | \phi \rangle - 1 = 0$. It is easy to foresee that what the method has to produce (if it is to be of any value) is the normalized ground-state wave function for the harmonic oscillator. How will the Lagrange multipliers method get such result?

The answer is on p. 198.

---

[7] The method does not give us information about the kind of extremum found.

[8] The argument of a functional is a function that produces the value of the functional (a number).

O. PENALTY FUNCTION METHOD

Very often we are interested in the minimization of a ("target") function,[1] i.e. in finding such values of variables, which ensure a minimum of the function when some constraints are satisfied. Just imagine hiking in the Smoky Mountains: we want to find the point of the lowest ground elevation provided that we hike along a straight line from, say, Gatlinburg to Cherokee.

Suppose the target function for minimization (which corresponds to the elevation of the ground in the Smoky Mountains region) is the function $f(x_1, x_2, \ldots, x_{n+m})$, but the variables x_i have to fulfil m equations ("constraints"):

$$\phi_i(x_1, x_2, \ldots, x_{n+m}) = 0 \quad \text{for } i = 1, 2, \ldots, m.$$

For such tasks we have at least three possibilities. The first is to eliminate m variables (by using the conditions) and express them by others. In this way the target function f takes into account all the constraints and depends only on n independent variables. Then the target function is to be minimized. The second possibility is to use the Lagrange multipliers method (see Appendix N). In both cases there is, however, the complication that the conditions to be satisfied might be quite complex and therefore solution of the corresponding equations may be difficult to achieve. An easier solution may be to choose a penalty method. The idea behind the penalty method is quite simple. Why go to the trouble of trying to satisfy the conditions $\phi_i = 0$, when we could propose the following: instead of function f let us minimize its modification

$$F = f + \sum_{i=1}^{m} K_i \phi_i^2,$$

where the penalty coefficients $K_i > 0$ are chosen to be large.[2] When minimizing F we *admit that the conditions* $\phi_i = 0$ *could be non-satisfied*, but any attempt to violate them introduces to F a positive contribution $\sum_{i=1}^{m} K_i \phi_i^2$. This means that, for minimization of F, it would always be better to explore such points in space (Fig. O.1) for which $\sum_{i=1}^{m} K_i \phi_i^2 = 0$. If the K's are large enough, the procedure will force the choice $\phi_i^2 = 0$, or $\phi_i = 0$ for $i = 1, 2, \ldots, m$, and this is what has to be satisfied during minimization.

Note that the task would be much more difficult if ϕ_i^2 had more than one minimum that corresponds to $\phi_i = 0$. This penalty method is worth keeping in our tool

---

[1] If we change the sign of the target function, the task is equivalent to maximization.

[2] This means a high penalty.

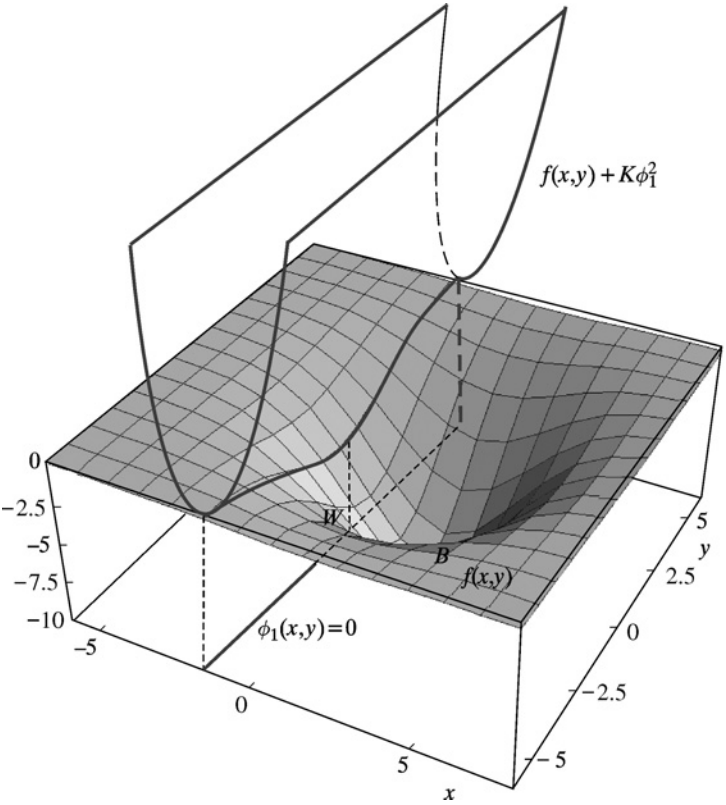


Fig. O.1. How does the penalty method work? We have to minimize $f(x, y)$, but under the condition that x and y satisfy the equation $\phi_1(x, y) = 0$ (black line). Function $f(x, y)$ exhibits a single minimum at point B, but this minimum is of no interest to us, because we are looking for a conditional minimum. To find it we minimize the sum $f(x, y) + K\phi_1^2$ with the penalty function $K\phi_1^2 \geqslant 0$ *allowing* any deviation from the black line $\phi_1(x, y) = 0$. However, going off this line does not pay, because this is precisely what switches the penalty on. As a result, at sufficiently large K we obtain the conditional minimum W. This is what the game is all about.

box, because it is general and easily applicable. For the method to work, it has to have a sufficiently large K. However, if K is too large, the numerical results might be of poor quality, since the procedure would first of all take care of the penalty, paying little attention to f. It is recommended that we take a few values of K and check whether the results depend on this.

As an example of the penalty function method, let us take the docking of two molecules. Our goal is to give such values of the atomic coordinates of both molecules as to ensure the contacts of some particular atoms of both molecules within some precise distance limits for the contacting atoms. The task sounds trivial, until we try to accomplish it in practice (especially for large molecules). The goal can be rather easily achieved when the penalty function method is used. We do the following. To the existing force field (i.e. an approximate electronic energy, Chapter 7) we simply add a penalty for not satisfying the desired contacts. For a single pair of the atoms (a contact) the penalty could be set as

$$K(r - r_0)^2,$$

where r stands for the distance of the atoms, and r_0 is the optimum (desired) contact distance. At a chosen starting geometry the atoms are far from achieving the optimum distance, and therefore the force field energy is supplemented by a large distance-dependent penalty. The energy is so high that the minimization procedure tries to remove the penalty and relax the system. Often this can be done in only one way: by docking the molecules in such a way as to achieve the proper contact distance.

P. MOLECULAR INTEGRALS WITH GAUSSIAN TYPE ORBITALS 1s

The normalized 1s spherically symmetric Gaussian Type Orbital (GTO) centred at the point shown by the vector $\boldsymbol{R}_p$ reads as

$$\chi_p \equiv \left(\frac{2\alpha_p}{\pi}\right)^{\frac{3}{4}} \exp\bigl(-\alpha_p|\boldsymbol{r}-\boldsymbol{R}_p|^2\bigr).$$

The molecular integrals usually involve, at most, four such orbitals: $\chi_p, \chi_q, \chi_r, \chi_s$, with corresponding centres $\boldsymbol{R}_p, \boldsymbol{R}_q, \boldsymbol{R}_r, \boldsymbol{R}_s$, and exponents $\alpha_p, \alpha_q, \alpha_r, \alpha_s$, respectively. Since any product of the 1s GTOs represents a (non-normalized) 1s GTO centred between the centres of the individual GTOs (see p. 359), let us denote the centre of $\chi_p\chi_q$ by $\boldsymbol{R}_k = \frac{\alpha_p\boldsymbol{R}_p+\alpha_q\boldsymbol{R}_q}{\alpha_p+\alpha_q}$, and the centre of $\chi_r\chi_s$ by $\boldsymbol{R}_l = \frac{\alpha_r\boldsymbol{R}_r+\alpha_s\boldsymbol{R}_s}{\alpha_r+\alpha_s}$. Then all the integrals needed are as follows:[1]

overlap integral:

$$S_{pq} = \langle\chi_p|\chi_q\rangle = \left(\frac{4\alpha_p\alpha_q}{(\alpha_p+\alpha_q)^2}\right)^{\frac{3}{4}} \exp\left(\frac{-\alpha_p\alpha_q}{\alpha_p+\alpha_q}|\boldsymbol{R}_p-\boldsymbol{R}_q|^2\right); \tag{P.1}$$

kinetic energy integral:

$$T_{pq} = \left\langle\chi_p\left|-\frac{1}{2}\Delta\right|\chi_q\right\rangle = \frac{\alpha_p\alpha_q}{\alpha_p+\alpha_q}\left(3-\frac{2\alpha_p\alpha_q}{\alpha_p+\alpha_q}|\boldsymbol{R}_p-\boldsymbol{R}_q|^2\right)S_{pq}; \tag{P.2}$$

nuclear attraction integral:[2]

$$V_{pq}^{\alpha} = \left\langle\chi_p\left|\frac{1}{|\boldsymbol{r}-\boldsymbol{R}_\alpha|}\right|\chi_q\right\rangle = 2\sqrt{\frac{\alpha_p+\alpha_q}{\pi}}F_0\bigl((\alpha_p+\alpha_q)|\boldsymbol{R}_\alpha-\boldsymbol{R}_k|^2\bigr)S_{pq}; \tag{P.3}$$

electron repulsion integral:

$$\begin{aligned}(pr|qs) &= (\chi_p\chi_r|\chi_q\chi_s) = \int \chi_p(1)^*\chi_q(1)\frac{1}{r_{12}}\chi_r^*(2)\chi_s(2)\,\mathrm{d}v_1\,\mathrm{d}v_2\\ &= \frac{2}{\sqrt{\pi}}\frac{\sqrt{\alpha_p+\alpha_q}\sqrt{\alpha_r+\alpha_s}}{\sqrt{\alpha_p+\alpha_q+\alpha_r+\alpha_s}}F_0\left(\frac{(\alpha_p+\alpha_q)(\alpha_r+\alpha_s)}{\alpha_p+\alpha_q+\alpha_r+\alpha_s}|\boldsymbol{R}_k-\boldsymbol{R}_l|^2\right)S_{pq}S_{rs}\end{aligned} \tag{P.4}$$

---

[1] S.F. Boys, *Proc. Roy. Soc. (London)* A200 (1950) 542.

[2] In order to interpret this integral (in a.u.) as the Coulombic attraction of the electronic charge $\chi_p^*(1)\chi_q(1)$ by a nucleus (of charge Z, located at $\boldsymbol{R}_\alpha$) we have to multiply the integral by $-Z$.

with F_0 defined as[3]

$$F_0(t) = \frac{1}{\sqrt{t}} \int_0^{\sqrt{t}} \exp\left(-u^2\right) \mathrm{d}u. \tag{P.5}$$

Note that for an atom (all the centres coincide) we have $t = 0$ and $F_0(0) = 1$.

Do these formulae work?

The formulae look quite complex. If they are correct, they have to work in several simple situations. For example, if the electronic distribution $\chi_p^*(1)\chi_q(1)$ centred at $\boldsymbol{R}_k$ is far away from the nucleus, then we have to obtain the Coulombic interaction of the charge of $\chi_p^*(1)\chi_q(1)$ and the nucleus. The total charge of the electron cloud $\chi_p^*(1)\chi_q(1)$ is obviously equal to S_{pq}, and therefore $\frac{S_{pq}}{|\boldsymbol{R}_\alpha - \boldsymbol{R}_k|}$ should be a very good estimation of the nuclear attraction integral, right?

What we need is the asymptotic form of $F_0(t)$, if $t \to \infty$. This can be deduced from our formula for $F_0(t)$. The integrand is concentrated close to $t = 0$. For $t \to \infty$, the contributions to the integral become negligible and the integral itself can be replaced by $\int_0^\infty \exp(-u^2)\,\mathrm{d}u = \sqrt{\pi}/2$. This gives $[F_0(t)]_{\text{asympt.}} = \frac{\sqrt{\pi}}{2\sqrt{t}}$ and

$$\begin{aligned}
\left(V_{pq}^\alpha\right)_{\text{asympt.}} &= 2\sqrt{\frac{\alpha_p + \alpha_q}{\pi}} F_0\left((\alpha_p + \alpha_q)|\boldsymbol{R}_\alpha - \boldsymbol{R}_k|^2\right) S_{pq} \\
&= 2\sqrt{\frac{\alpha_p + \alpha_q}{\pi}} \frac{\sqrt{\pi}}{2\sqrt{(\alpha_p + \alpha_q)|\boldsymbol{R}_\alpha - \boldsymbol{R}_k|^2}} S_{pq} = \frac{S_{pq}}{|\boldsymbol{R}_\alpha - \boldsymbol{R}_k|},
\end{aligned}$$

exactly as we expected. If $\chi_p = \chi_q$, then $S_{pq} = 1$ and we simply get the Coulombic law for the unit charges. It works.

Similarly, if in the electronic repulsion integral $\chi_p = \chi_q$, $\chi_r = \chi_s$ and the distance $|\boldsymbol{R}_k - \boldsymbol{R}_l| = R$ is large, then we should get the Coulombic law for the two point-like unit charges at distance R. Let us see. Asymptotically

$$\begin{aligned}
(pr|qs)_{\text{asympt.}} &= \frac{2}{\sqrt{\pi}} \frac{\sqrt{\alpha_p + \alpha_q}\sqrt{\alpha_r + \alpha_s}}{\sqrt{\alpha_p + \alpha_q + \alpha_r + \alpha_s}} F_0\left(\frac{(\alpha_p + \alpha_q)(\alpha_r + \alpha_s)}{\alpha_p + \alpha_q + \alpha_r + \alpha_s}|\boldsymbol{R}_k - \boldsymbol{R}_l|^2\right) \\
&= \frac{2}{\sqrt{\pi}} \frac{\sqrt{\alpha_p + \alpha_q}\sqrt{\alpha_r + \alpha_s}}{\sqrt{\alpha_p + \alpha_q + \alpha_r + \alpha_s}} \frac{\sqrt{\pi}}{2\sqrt{\frac{(\alpha_p+\alpha_q)(\alpha_r+\alpha_s)}{\alpha_p+\alpha_q+\alpha_r+\alpha_s}|\boldsymbol{R}_k - \boldsymbol{R}_l|^2}} = \frac{1}{R},
\end{aligned}$$

which is exactly what we should obtain.

---

[3] The values of $F_0(t)$ are reported in L.J. Schaad, G.O. Morrell, *J. Chem. Phys.* 54 (1971) 1965.

Q. SINGLET AND TRIPLET STATES FOR TWO ELECTRONS

An angular momentum is a vector, and this pertains also to spin angular momenta (see Chapter 1). The spin angular momentum of a certain number of elementary particles is the sum of their spin vectors. To obtain the total spin vector, we therefore have to add the x components of the spins of the particles, similarly for the y and z components, and to construct the total vector from them. Then we might be interested in the corresponding spin operators. These operators will be created using Pauli matrices.[1]

In this way we find immediately that, for a single particle, the following identity holds

$$\hat{S}^2 = \hat{S}_x^2 + \hat{S}_y^2 + \hat{S}_z^2 = \hat{S}_z^2 + \hat{S}_+\hat{S}_- - \hbar\hat{S}_z, \tag{Q.1}$$

lowering and raising operators

where $\hat{S}^2 \equiv \hat{\boldsymbol{S}}^2$ and $\hat{S}_+$ and $\hat{S}_-$ are the *lowering* and *raising* operators, respectively,

$$\hat{S}_+ = \hat{S}_x + i\hat{S}_y, \tag{Q.2}$$

$$\hat{S}_- = \hat{S}_x - i\hat{S}_y, \tag{Q.3}$$

which satisfy the useful relations justifying their names:

$$\hat{S}_+\alpha = 0, \qquad \hat{S}_+\beta = \hbar\alpha,$$
$$\hat{S}_-\alpha = \hbar\beta, \qquad \hat{S}_-\beta = 0.$$

For any stationary state the wave function is an eigenfunction of the square of the total spin operator and of the z-component of the total spin operator. The one and two-electron cases are the only ones for which the total wave function is the *product* of space and spin parts.

The maximum projection of the electron spin on the z axis is equal to $\frac{1}{2}$ (in a.u.). Hence, the maximum projection for the total spin of two electrons is equal to 1. This means that in this case only two spin states are possible: the *singlet state* corresponding to $S = 0$ and the *triplet state* with $S = 1$ (see Postulate V). In the singlet state the two electronic spins are opposite ("pairing of electrons"), while in the triplet state the spin vectors are "parallel" (cf. Fig. 1.11 in Chapter 1). As always, the possible projection of the total spin takes one of the values: $M_S =$

---

[1] See Postulate VI in Chapter 1.

$-S, -S+1, \ldots, +S$, i.e. $M_S = 0$ for the singlet state and $M_S = -1, 0, +1$ for the triplet state.

Now it will be shown that the two-electron spin function $\alpha(1)\beta(2) - \alpha(2)\beta(1)$ ensures the singlet state. First, let us construct the square of the total spin of the two electrons:

$$\boldsymbol{S}^2 = (\boldsymbol{s}_1 + \boldsymbol{s}_2)^2 = s_1^2 + s_2^2 + 2\boldsymbol{s}_1\boldsymbol{s}_2.$$

Thus to create operator $\hat{\boldsymbol{S}}^2$ we need operators $\hat{s}_1^2$ and $\hat{s}_2^2$, which will be expressed by the lowering and raising operators according to eq. (Q.1), and the scalar product $\hat{s}_1\hat{s}_2$ expressed as the sum of the products of the corresponding components x, y and z (we know how they act, see Postulate V in Chapter 1). If $\hat{S}^2$ acts on $\alpha(1)\beta(2)$, after five lines of derivation we obtain

$$\hat{S}^2\big[\alpha(1)\beta(2)\big] = \hbar^2\big[\alpha(1)\beta(2) + \alpha(2)\beta(1)\big]$$

similarly

$$\hat{S}^2\big[\alpha(2)\beta(1)\big] = \hbar^2\big[\alpha(1)\beta(2) + \alpha(2)\beta(1)\big].$$

Now we will use this result to calculate

$$\hat{S}^2\big[\alpha(1)\beta(2) - \alpha(2)\beta(1)\big] \quad \text{and} \quad \hat{S}^2\big[\alpha(1)\beta(2) + \alpha(2)\beta(1)\big].$$

We have

$$\begin{aligned}
&\hat{S}^2\big[\alpha(1)\beta(2) - \alpha(2)\beta(1)\big] \\
&\quad = 0 \times \big[\alpha(1)\beta(2) - \alpha(2)\beta(1)\big] \\
&\quad \equiv S(S+1)\hbar^2\big[\alpha(1)\beta(2) - \alpha(2)\beta(1)\big],
\end{aligned}$$

where $S = 0$ (singlet) and

$$\begin{aligned}
&\hat{S}^2\big[\alpha(1)\beta(2) + \alpha(2)\beta(1)\big] = 2, \\
&\hbar^2\big[\alpha(1)\beta(2) + \alpha(2)\beta(1)\big] \equiv S(S+1)\hbar^2\big[\alpha(1)\beta(2) + \alpha(2)\beta(1)\big],
\end{aligned}$$

where $S = 1$ (triplet).

If operator $\hat{S}_z = \hat{s}_{1z} + \hat{s}_{2z}$ acts on $[\alpha(1)\beta(2) - \alpha(2)\beta(1)]$, we obtain

$$0 \times \big[\alpha(1)\beta(2) - \alpha(2)\beta(1)\big].$$

This means that, in the singlet state, the projection of the spin on the z axis is equal to 0. This is what we expect from a singlet state function.

On the other hand, if $\hat{S}_z = \hat{s}_{1z} + \hat{s}_{2z}$ acts on $[\alpha(1)\beta(2) + \alpha(2)\beta(1)]$, we have

$$0 \times \big[\alpha(1)\beta(2) + \alpha(2)\beta(1)\big],$$

i.e. the function $[\alpha(1)\beta(2)+\alpha(2)\beta(1)]$ is such a triplet function which corresponds to the zero projection of the total spin. A similarly simple calculation for the spin functions $\alpha(1)\alpha(2)$ and $\beta(1)\beta(2)$ gives the eigenvalue $S_z = \hbar$ and $S_z = -\hbar$, respectively. Therefore, after normalization[2] finally

$\frac{1}{\sqrt{2}}[\alpha(1)\beta(2) - \alpha(2)\beta(1)]$ is a singlet function, while: $\frac{1}{\sqrt{2}}[\alpha(1)\beta(2) + \alpha(2)\beta(1)]$, $\alpha(1)\alpha(2)$ and $\beta(1)\beta(2)$ represent three triplet functions.

---

[2] For example let us check the normalization of the singlet function $\frac{1}{\sqrt{2}}[\alpha(1)\beta(2) - \alpha(2)\beta(1)]$:

$$\sum_{\sigma_1}\sum_{\sigma_2}\left\{\frac{1}{\sqrt{2}}\left[\alpha(1)\beta(2)-\alpha(2)\beta(1)\right]\right\}^2$$

$$=\sum_{\sigma_1}\sum_{\sigma_2}\frac{1}{2}\{[\alpha(1)]^2[\beta(2)]^2+[\alpha(2)]^2[\beta(1)]^2-2[\alpha(2)\beta(2)][\alpha(1)\beta(1)]\}$$

$$=\frac{1}{2}\left\{\sum_{\sigma_1}[\alpha(1)]^2\sum_{\sigma_2}[\beta(2)]^2+\sum_{\sigma_2}[\alpha(2)]^2\sum_{\sigma_1}[\beta(1)]^2-2\sum_{\sigma_2}[\alpha(2)\beta(2)]\sum_{\sigma_1}[\alpha(1)\beta(1)]\right\}$$

$$=\frac{1}{2}\{1\cdot 1+1\cdot 1-2\cdot 0\cdot 0\}=1.$$

R. THE HYDROGEN MOLECULAR ION IN THE SIMPLEST ATOMIC BASIS SET

Consider the quantum mechanical description of the hydrogen molecular ion in its simplest version. Let us use molecular orbital theory with the atomic basis set composed of only two Slater Type Orbitals (STO): $1s_a$ and $1s_b$ centred on the nuclei a and b. The mean value of the Hamiltonian calculated with the bonding $(+)$ and antibonding $(-)$ orbital (see Chapter 8 and Appendix D) reads as

$$E_\pm = \frac{H_{aa} \pm H_{ab}}{1 \pm S},$$

where the Hamiltonian (in a.u.)[1] $\hat{H} = -\frac{1}{2}\Delta - \frac{1}{r_a} - \frac{1}{r_b} + \frac{1}{R}$ and S stands for the overlap integral of the two atomic orbitals. Thus we have

$$\begin{aligned} E_\pm &= \frac{1}{R} + \frac{H_{aa} \pm H_{ab}}{1 \pm S} = \frac{1}{R} + \frac{\left(-\frac{1}{2}\Delta - \frac{1}{r_a} - \frac{1}{r_b}\right)_{aa} \pm \left(-\frac{1}{2}\Delta - \frac{1}{r_a} - \frac{1}{r_b}\right)_{ab}}{1 \pm S} \\ &= \frac{1}{R} + \frac{E_H + V_{aa,b} \pm E_H S \pm V_{ab,b}}{1 \pm S} = E_H + \frac{1}{R} + \frac{V_{aa,b} \pm V_{ab,b}}{1 \pm S}, \end{aligned}$$

where E_H means the energy of the hydrogen atom, while the nuclear attraction integrals are

$$V_{aa,b} = -\left(a\left|\frac{1}{r_b}\right|a\right), \qquad V_{ab,b} = -\left(a\left|\frac{1}{r_b}\right|b\right).$$

The energy $E_\pm$ is a function of the internuclear distance R, which is hidden in the dependence of the integrals on R. We want to have this function explicitly. To this end we have to compute the integrals S, $V_{aa,b}$ and $V_{ab,b}$. We use the elliptic coordinates (Fig. R.1):

$$\mu = \frac{r_a + r_b}{R}, \qquad \nu = \frac{r_a - r_b}{R}, \qquad \phi = \arctan\left(\frac{y}{x}\right).$$

The volume element in the elliptic coordinates is $\mathrm{d}V = R^3/8(\mu^2 - \nu^2)\,\mathrm{d}\mu\,\mathrm{d}\nu\,\mathrm{d}\phi$, where $1 \leqslant \mu < \infty$, $-1 \leqslant \nu \leqslant 1, 0 \leqslant \phi \leqslant 2\pi$.

---

[1] See Fig. R.1 for explaining symbols.

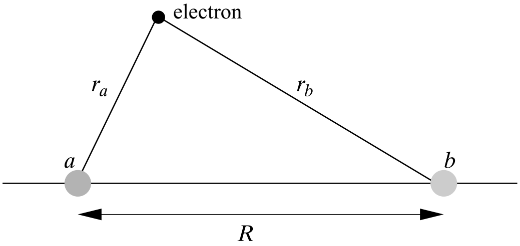


Fig. R.1. The elliptic coordinates $\mu = \frac{r_a+r_b}{R}$, $\nu = \frac{r_a-r_b}{R}$ built using distances r_a and r_b from the two foci (where the nuclei are, their distance is R) of the ellipse. The angle ϕ measures the rotation of the plane defined by ab and the corresponding electron about the ab axis.

We will need two auxiliary integrals:

$$A_n(\sigma,\alpha) = \int_\sigma^\infty \mu^n \exp(-\alpha x)\,\mathrm{d}x = \exp(-\alpha\sigma)\sum_{k=0}^{n} \frac{n!}{(n-k)!}\frac{\sigma^{n-k}}{\alpha^{k+1}},$$

$$B_n(\alpha) = \int_{-1}^{+1} x^n \exp(-\alpha x)\,\mathrm{d}x = A_n(-1,\alpha) - A_n(1,\alpha).$$

The integrals $A_n(\sigma,\alpha)$ satisfy the following recurrence relation:

$$A_n(\sigma,\alpha) = \sigma^n A_0(\sigma,\alpha) + \frac{n}{\alpha}A_{n-1}(\sigma,\alpha),$$

$$A_0(\sigma,\alpha) = \frac{1}{\alpha}\exp(-\sigma\alpha).$$

These are some simple auxiliary integrals (we will need them in a moment):

$$A_1(\sigma,\alpha) = \sigma\frac{1}{\alpha}\exp(-\sigma\alpha) + \frac{1}{\alpha}\frac{1}{\alpha}\exp(-\sigma\alpha) = \frac{1}{\alpha}\left(\sigma+\frac{1}{\alpha}\right)\exp(-\sigma\alpha),$$

$$\begin{aligned} A_2(\sigma,\alpha) &= \sigma^2\frac{1}{\alpha}\exp(-\sigma\alpha) + \frac{2}{\alpha}\left(\frac{1}{\alpha}\left(\sigma+\frac{1}{\alpha}\right)\exp(-\sigma\alpha)\right) \\ &= \frac{1}{\alpha}\exp(-\sigma\alpha)\left[\sigma^2 + \frac{2}{\alpha}\left(\sigma+\frac{1}{\alpha}\right)\right], \end{aligned}$$

$$B_0(\alpha) = \frac{1}{\alpha}\exp(\alpha) - \frac{1}{\alpha}\exp(-\alpha) = \frac{1}{\alpha}\big[\exp(\alpha)-\exp(-\alpha)\big],$$

$$\begin{aligned} B_1(\alpha) &= \frac{1}{\alpha}\left(-1+\frac{1}{\alpha}\right)\exp(\alpha) - \frac{1}{\alpha}\left(1+\frac{1}{\alpha}\right)\exp(-\alpha) \\ &= \frac{1}{\alpha}\left[\left(\frac{1}{\alpha}-1\right)\exp(\alpha) - \left(\frac{1}{\alpha}+1\right)\exp(-\alpha)\right]. \end{aligned}$$

Thus, the overlap integral S is calculated in the following way

$$S = \frac{R^3}{8\pi}\int_1^\infty \mathrm{d}\mu\,\exp(-R\mu)\int_{-1}^{+1}\mathrm{d}\nu\,(\mu^2-\nu^2)\int_0^{2\pi}\mathrm{d}\phi$$

$$
\begin{aligned}
&= \frac{R^3}{2}\left[\int_1^{\infty} \mathrm{d}\mu\, \mu^2 \exp(-R\mu) - \frac{1}{3}\int_1^{\infty} \mathrm{d}\mu \exp(-R\mu)\right] \\
&= \frac{R^3}{2}\left[A_2(1,\alpha) - \frac{1}{3}A_0(1,\alpha)\right] \\
&= \frac{R^3}{2}\left[\frac{1}{R}\exp(-R)\left(1+\frac{2}{R}+\frac{2}{R^2}\right) - \frac{1}{3}\frac{1}{R}\exp(-R)\right] \\
&= \exp(-R)\left(\frac{R^2}{3}+R+1\right).
\end{aligned}
$$

Thus we have explicit dependence on R. The formula for S satisfies correctly the limiting cases: $\lim_{R\to\infty} S(R) = 0$ and $\lim_{R\to 0} S(R) = 1$ (normalization of the $1s$ orbital). Besides

$$
\begin{aligned}
\frac{\mathrm{d}S}{\mathrm{d}R} &= -\exp(-R)\left(\frac{R^2}{3}+R+1\right) + \exp(-R)\left(\frac{2}{3}R+1\right) \\
&= -\exp(-R)\left(\frac{R^2+R}{3}\right) < 0,
\end{aligned}
$$

i.e. the overlap integral of the $1s$ functions decreases from 1 to 0, if $R \to \infty$ (see Fig. R.2.a).

We see that for small R the function S decreases gently, while for larger R it decreases fast.[2]

Using the elliptic coordinates and the formulae for the integrals $A_n(\sigma,\alpha)$ and $B_n(\alpha)$ we obtain

$$
\begin{aligned}
-V_{aa,b} &= \left(a\left|\frac{1}{r_b}\right|a\right) = \frac{1}{\pi}\int \exp(-2r_a)\frac{1}{r_b}\,\mathrm{d}\tau \\
&= \frac{R^3}{8\pi}\frac{2}{R}\int_1^{\infty} \mathrm{d}\mu\, \exp\bigl[-R(\mu+\nu)\bigr]\int_{-1}^{+1} \mathrm{d}\nu\, \frac{(\mu^2-\nu^2)}{\mu-\nu}\int_0^{2\pi} \mathrm{d}\phi \\
&= \frac{R^2}{4\pi}2\pi\int_1^{\infty} \mathrm{d}\mu \int_{-1}^{+1} \mathrm{d}\nu\, \exp(-R\mu)\exp(-R\nu)(\mu+\nu) \\
&= \frac{R^2}{2}\left[\int_1^{\infty} \mathrm{d}\mu\, \mu \exp(-R\mu)\int_{-1}^{+1} \mathrm{d}\nu\, \exp(-R\nu)\right. \\
&\quad \left. + \int_1^{\infty} \mathrm{d}\mu\, \exp(-R\mu)\int_{-1}^{+1} \mathrm{d}\nu\, \nu \exp(-R\nu)\right] \\
&= \frac{R^2}{2}\bigl[A_1(1,R)B_0(R) + A_0(1,R)B_1(R)\bigr] = \frac{1}{R} - \exp(-2R)\left(1+\frac{1}{R}\right).
\end{aligned}
$$

---

[2] Just to get an idea: at $R = 5$ a.u. (quite typical for van der Waals complexes) the value of the overlap integral is of the order of 0.1.

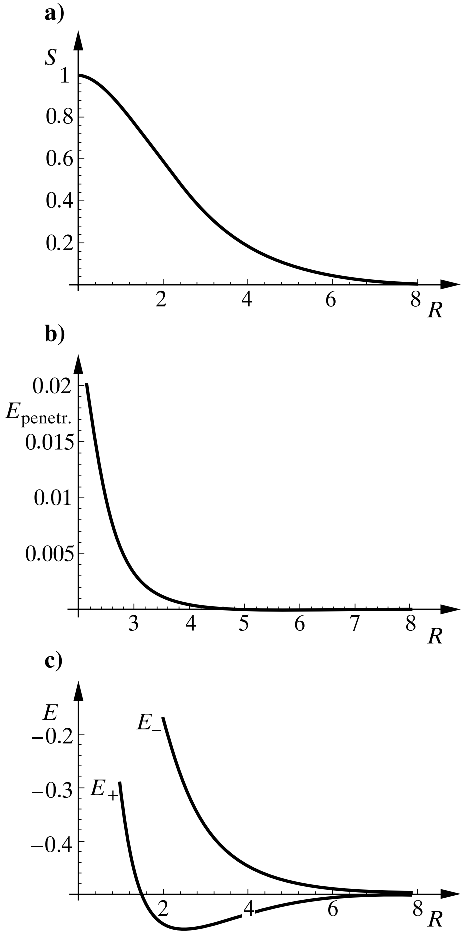


Fig. R.2. The hydrogen molecule in the simplest basis set of two 1*s* Slater type orbitals (STO). (a) The overlap integral S as a function of the internuclear distance R. (b) The penetration energy represents the difference between the electron–proton interaction calculated assuming the electronic charge distribution and the same energy calculated with the point charges (the electron is located on nucleus a). (c) The energies E_+ and E_- of the bonding (lower curve) and of the antibonding (upper curve) orbitals. Energies and distances in a.u.

This is an interesting result. The integral $-V_{aa,b}$ means $(a|\frac{1}{r_b}|a)$, which at large R should give the Coulombic interaction of the two unit point charges, i.e. $\frac{1}{R}$. This is what we have as the first term. The second term: $E_{\text{penetr}} = -\exp(-2R)(1+\frac{1}{R})$ represents what is known as *penetration energy* resulting from the non-point-like character of one of the interacting charges.[3]

penetration energy

From Fig. R.2.b we see that the penetration energy vanishes much faster that the overlap integral. No wonder it vanishes as $\exp(-2R)$, while the overlap integral vanishes only as $\exp(-R)$.

It is seen that

> the diffuse charges interact more weakly.

---

[3] The electron cloud of electronic density a^2.

On the one hand diffuse charges offer the chance to be close in space (this increases the interaction), on the other hand some charges become more distant. The second effect prevails and therefore the penetration energy makes the Coulombic interaction weaker.

What will happen if $R \to 0$?

Let us expand the exponential function in the Taylor series. We obtain

$$\begin{aligned}\lim_{R\to 0}\left[V_{aa,b}(R)\right] &= -\lim_{R\to 0}\left[\frac{1}{R} - \left[1 - 2R + \frac{1}{2}R^2 + \cdots\right]\left(1 + \frac{1}{R}\right)\right] \\ &= -\lim_{R\to 0}\left(\frac{1}{R} - 1 + 2R - \frac{1}{2}R^2 - \frac{1}{R} + 2 + \frac{1}{2}R + \cdots\right) = -1.\end{aligned}$$

This is exactly what we get for the hydrogen atom when calculating:

$$\begin{aligned}V_{aa,a} &= -\int \mathrm{d}v\,\frac{1}{r}(1s)^2 = -\frac{1}{\pi}\int \exp(-2r)\frac{1}{r}r^2 \sin\theta\,\mathrm{d}r\,\mathrm{d}\theta\,\mathrm{d}\phi \\ &= -4\int_0^\infty r\exp(-2r)\,\mathrm{d}r = -4\times 2^{-2} = -1.\end{aligned}$$

Thus everything is all right.

Similarly we calculate

$$\begin{aligned}-V_{ab,b} &= \left(a\left|\frac{1}{r_b}\right|b\right) = \frac{1}{\pi}\int \exp\bigl(-(r_a + r_b)\bigr)\frac{1}{r_b}\,\mathrm{d}v \\ &= \frac{1}{\pi}\frac{2}{R}\int \exp(-R\mu)\frac{1}{(\mu - \nu)}\frac{R^3}{8}\bigl(\mu^2 - \nu^2\bigr)\,\mathrm{d}\mu\,\mathrm{d}\nu\,\mathrm{d}\phi \\ &= \frac{R^2}{2}\int_1^\infty\int_{-1}^{+1}\mathrm{d}\mu\,\mathrm{d}\nu\bigl[\mu\exp(-R\mu) + \nu\exp(-R\mu)\bigr] \\ &= \frac{R^2}{2}2A_1(1,R) + 0 = (1+R)\exp(-R).\end{aligned}$$

If $R \to \infty$, then $-V_{ab,b} \to 0$, which is the correct behaviour. Do we get $V_{aa,a} = -1$, if $R \to 0$? Again, let us expand the exponential function:

$$\begin{aligned}V_{aa,a} &= -\lim_{R\to 0}(1+R)\exp(-R) = -\lim_{R\to 0}(1+R)\left(1 - R + \frac{R^2}{2} + \cdots\right) \\ &= -\lim_{R\to 0}\left[1 + R - R - R^2 + \frac{R^2}{2} + \cdots\right] = -1.\end{aligned}$$

This is what we expected.

Bonding and antibonding orbital energy

If we insert the results into the formula for the energy of the bonding and antibonding orbitals, we obtain the most important formulae for the problem being considered:

$$
\begin{aligned}
E_{\pm} &= E_H + \frac{1}{R} + \frac{V_{aa,b} \pm V_{ab,b}}{1 \pm S} \\
&= E_H + \frac{1}{R} + \frac{-\frac{1}{R} + \exp(-2R)\left(1 + \frac{1}{R}\right) \pm (-1 - R)\exp(-R)}{1 \pm \left[\exp(-R)\left(\frac{R^2}{3} + R + 1\right)\right]}.
\end{aligned}
$$

The plots of $E_{\pm}$ are shown in Fig. R.2.c. It is seen that in the quite primitive LCAO MO approximation, the bonding energy is lower than the energy of the hydrogen atom E_H for all sufficiently large R (a single minimum). The energy of the antibonding orbital is higher than E_H for all R (no minimum). This simple theory predicts the position of the energy minimum for the ground state as $R_e = 2.5$ a.u., while the experimental value is equal[4] ca. 2.0 a.u.

---

[4] These two quantities are not directly comparable, because the experimental value does not correspond exactly to the position of the minimum (because of anharmonicity).

S. POPULATION ANALYSIS

On p. 569 the electronic density ρ is defined. If the wave function is a Slater determinant (p. 332) and assuming the double occupancy of orbitals φ_i, we have (see (11.5)):

$$\rho(\boldsymbol{r}) = 2\left[\left|\varphi_1(\boldsymbol{r})\right|^2 + \left|\varphi_2(\boldsymbol{r})\right|^2 + \cdots + \left|\varphi_{\frac{N}{2}}(\boldsymbol{r})\right|^2\right]. \tag{S.1}$$

The density distribution ρ may be viewed as a cloud carrying a charge $-Ne$ and eq. (S.1) says that the cloud is composed of individual clouds of molecular orbitals, each carrying two electrons. On the other hand in the LCAO approximation any molecular orbital is represented by the sum of atomic orbitals. If we insert the LCAO expansion into ρ, then ρ becomes the sum of the contributions, each being the product of two atomic orbitals. There is a temptation to go even further and to divide ρ somehow *into the contributions of particular atoms*, calculate the charge corresponding to such contributions and locate the (point) charge on the nucleus.[1] We might say therefore, what the "electron population" residing on the particular atoms is (hence the name: population analysis).

Mulliken population analysis

Such tricks are of course possible, and one of them is called Mulliken population analysis. From (S.1), after using the LCAO expansion $\varphi_i = \sum_r c_{ri}\chi_r$, we have ($S_{rs}$ stands for the overlap integrals between the atomic orbitals r and s, and c are the corresponding LCAO coefficients)

$$\begin{aligned} N &= \int \rho(\boldsymbol{r})\,\mathrm{d}V = 2\sum_{i=1}^{N/2}\int \left|\varphi_i(\boldsymbol{r})\right|^2 \mathrm{d}V = \sum_i\sum_{rs} 2c_{ri}^* c_{si} S_{rs} \\ &= \sum_{rs} P_{rs}S_{rs} = \mathrm{Tr}(\boldsymbol{PS}), \end{aligned} \tag{S.2}$$

where $\boldsymbol{P}$ is called the charge and bond-order matrix

$$P_{sr} = \sum_i 2c_{ri}^* c_{si}. \tag{S.3}$$

The summation over r and s may be carried out, being careful from which atom A the particular atomic orbital comes (we assume that the AO's are centred on the nuclei). We get an equivalent formula (A ad B denote atoms):

---

[1] This number need not be an integer.

$$N = \sum_{A} \sum_{r \in A} \sum_{B} \sum_{s \in B} P_{rs} S_{rs}.$$

Afterwards we may choose the following partitionings:

Atomic partitioning:

$$N = \sum_{A} q_A, \quad q_A = \sum_{r \in A} \left(\sum_{B} \sum_{s \in B} P_{rs} S_{rs} \right),$$

Mulliken charges

where q are called the *Mulliken charges*. They are often calculated in practical applications and serve to provide information on how much of the electronic density ρ is concentrated on atom A. Such a quantity is of interest because it may be directly linked to the reactivity of atom A, often identified with its ability to be attacked by nucleophilic or electrophilic agents.[2] Also, if we measure the dipole moment, we would like to know why this moment is large in a molecule. By performing Mulliken analysis, we might be able to identify those atoms that are responsible for this. This might be of value when interpreting experimental data.

Atomic and bond partitioning: The summation may also be performed in a slightly different way

$$N = \sum_{A} \sum_{r,s \in A} P_{rs} S_{rs} + \sum_{A<B} 2 \sum_{r \in A} \sum_{s \in B} P_{rs} S_{rs} = \sum_{A} \bar{q}_A + \sum_{A<B} \bar{q}_{AB}.$$

The first term represents the contributions $\bar{q}_A$ of the atoms, the second pertains to the atomic pairs $\bar{q}_{AB}$.

The latter populations are large and positive for those pairs of atoms for which chemists assign chemical bonds.

The bond population $\bar{q}_{AB}$ may be treated as a measure of whether in the $A-B$ atomic interaction, bonding or antibonding character prevails.[3] If, for two atoms, $\bar{q}_{AB} < 0$, we may say that they are not bound by any chemical bond, if $\bar{q}_{AB}$ is large, then we may treat it as an indication that these two atoms are bound by a chemical bond or bonds.

---

[2] We have to remember that, besides electrons, this atom has a nucleus. This has to be taken into account when calculating the atomic net charge.

[3] P_{rs} is the sum (over the occupied orbitals) of the products of the LCAO coefficients of two atoms in each of the occupied molecular orbitals. The equal signs of these coefficients (with $S_{rs} > 0$) means a *bonding* interaction (recall Chapter 8 and Appendix R on p. 1009) and such a contribution increases P_{rs}. The opposite signs of the coefficients (with $S_{rs} > 0$) corresponds to the antibonding interactions and in such a case the corresponding contribution decreases P_{rs}. If $S_{rs} < 0$, then the words "bonding" and "antibonding" above have to be exchanged, but the effect remains the same. This means that the product $P_{rs}S_{rs}$ in all cases correctly controls the bonding ($P_{rs}S_{rs} > 0$) or antibonding ($P_{rs}S_{rs} < 0$) effects.

Example 1. ***The hydrogen molecule.*** Let us take the simplest example. First, let us consider the electronic ground-state in the simplest molecular orbital approximation, i.e. two electrons are described by the normalized orbital in the form (a, b denote the $1s$ atomic orbitals centred on the corresponding nuclei; note that this is the famous bonding orbital)

$$\varphi_1 = N_1(a+b),$$

where $N_1 = (2+2S)^{-\frac{1}{2}}$, and $S \equiv (a|b)$. Then,

$$P_{sr} = \sum_i 2c_{ri}^* c_{si} = 2c_{r1}^* c_{s1} = (1+S)^{-1},$$

independent of the indices r and s. Of course,

$$\boldsymbol{S} = \begin{pmatrix} 1 & S \\ S & 1 \end{pmatrix}, \quad \text{and therefore} \quad \boldsymbol{PS} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}.$$

Thus, $\mathrm{Tr}\,(\boldsymbol{PS}) = 2 =$ the number of electrons $= P_{11}S_{11} + P_{22}S_{22} + 2P_{12}S_{12} = q_A + q_B + q_{AB}$, with $q_A = q_B = (1+S)^{-1}$, and $q_{AB} = \frac{2S}{1+S} > 0$. Thus we immediately see that the HH bond *has an electronic population greater than zero*, i.e. the atom–atom interaction is *bonding*.

Let us now consider H_2 with two electrons occupying the normalized orbital of a different character[4] $\varphi_2 = N_2(a-b)$, with $N_2 = (2-2S)^{-\frac{1}{2}}$, then

$$P_{sr} = \sum_i 2c_{ri}^* c_{si} = 2c_{r2}^* c_{s2} = (1-S)^{-1}$$

for $(r,s) = (1,1)$ and $(r,s) = (2,2)$ while $P_{rs} = -(1-S)^{-1}$ for $(r,s) = (1,2)$ and $(r,s) = (2,1)$.

Now, let us calculate

$$\boldsymbol{PS} = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$$

and $\mathrm{Tr}(\boldsymbol{PS}) = 2 =$ the number of electrons $= P_{11}S_{11} + P_{22}S_{22} + 2P_{12}S_{12} = q_A + q_B + q_{AB}$, but now $q_A = q_B = (1-S)^{-1}$ and $q_{AB} = -\frac{2S}{1-S} < 0$. Thus, a glance at q_{AB} tells us that this time the atoms are interacting in an antibonding way.

A similar analysis for polyatomic molecules gives more subtle and more interesting results.

Other population analyses

Partitioning of the electron cloud of N electrons according to Mulliken population analysis represents only one of possible choices. For a positively definite matrix[5] $\boldsymbol{S}$ (and the overlap matrix is always positively definite) we may introduce the powers

---

[4] We do not want to suggest anything, but this orbital is notorious for antibonding character.

[5] I.e. all the eigenvalues positive.

of the matrix[6] $\boldsymbol{S}^x$, where x is an arbitrary real number (in a way shown in Appendix J on p. 977), and we have $\boldsymbol{S}^{1-x}\boldsymbol{S}^x = \boldsymbol{S}$. Then we may write[7]

$$N = \mathrm{Tr}(\boldsymbol{PS}) = \mathrm{Tr}\big(\boldsymbol{S}^x\boldsymbol{PS}^{1-x}\big). \tag{S.4}$$

Löwdin population analysis

Now, we may take any x and for this value construct the corresponding partition of N electronic charges into atoms. If $x = 0$ or 1, then we have a Mulliken population analysis, if $x = \frac{1}{2}$ then we have what is called the *Löwdin population analysis*, etc.

Multipole representation

Imagine a charge distribution $\rho(\boldsymbol{r})$. Let us choose a Cartesian coordinate system. We may calculate the Cartesian moments of the distribution: $\int \rho(\boldsymbol{r})\,\mathrm{d}V$, i.e. the total charge, then $\int x\rho(\boldsymbol{r})\,\mathrm{d}V$, $\int y\rho(\boldsymbol{r})\,\mathrm{d}V$, $\int z\rho(\boldsymbol{r})\,\mathrm{d}V$, i.e. the components of the dipole moment, $\int x^2\rho(\boldsymbol{r})\,\mathrm{d}V$, $\int y^2\rho(\boldsymbol{r})\,\mathrm{d}V$, $\int z^2\rho(\boldsymbol{r})\,\mathrm{d}V$, $\int xy\rho(\boldsymbol{r})\,\mathrm{d}V$, $\int xz\rho(\boldsymbol{r})\,\mathrm{d}V$, $\int yz\rho(\boldsymbol{r})\,\mathrm{d}V$ – the components of the quadrupole moment, etc. The moments mean a complete description of $\rho(\boldsymbol{r})$ as concerns its interaction with another (distant) charge distribution. The higher the powers of x, y, z (i.e. the higher the moment) the more important distant parts of $\rho(\boldsymbol{r})$ are. If $\rho(\boldsymbol{r})$ extends to infinity (and for atoms and molecules it does), higher order moments tend to infinity. This means trouble when the consecutive interactions of the multipole moments are calculated (multipole expansion, Appendix X) and indeed, the multipole expansion "explodes", i.e. diverges.[8] This would not happen if the interacting charge distributions could be enclosed in two spheres.

There is also another problem: where to locate the origin of the coordinate system, with respect to which the moments are calculated? The answer is: *anywhere*. Wherever such an origin is located it is OK from the point of view of mathematics. However, such choices may differ enormously from the practical point of view. For example, let us imagine a spherically symmetric charge distribution. If the origin is located at its centre (as "most people would do"), then we have a quite simple description of $\rho(\boldsymbol{r})$ by using the moments, namely, the only non-zero moment is the charge, i.e. $\int \rho(\boldsymbol{r})\,\mathrm{d}V$. If, however, the origin is located off centre, all the moments would be non-zero. They are all needed to calculate accurately the interaction of the charge distribution (with anything). As we can see, it is definitely better to locate the origin at the centre of $\rho(\boldsymbol{r})$.

Well, and what if the charge distribution $\rho(\boldsymbol{r})$ were divided into segments and each segment represented by a set of multipoles? It would be all right, although, in view of the above example, it would be better to locate the corresponding origins at the centre of the segments. It is clear that, in particular, it would be OK if the segments were very small, e.g., the cloud was cut into tiny cubes and we consider

---

[6] They are symmetric matrices as well.

[7] We easily check that $\mathrm{Tr}(\boldsymbol{ABC}) = \mathrm{Tr}(\boldsymbol{CAB})$. Indeed, $\mathrm{Tr}(\boldsymbol{ABC}) = \sum_{i,k,l} A_{ik}B_{kl}C_{li}$, while $\mathrm{Tr}(\boldsymbol{CAB}) = \sum_{i,k,l} C_{ik}A_{kl}B_{li}$. Changing summation indices $k \to i$, $l \to k$, $i \to l$ in the last formula, we obtain $\mathrm{Tr}(\boldsymbol{ABC})$.

[8] Although the first terms (i.e. before the "explosion") may give accurate results.

every cube's content as a separate cloud.[9] But, well…, what are the multipoles for? Indeed, it would be sufficient to take only the charges of the cubes, because they approximate the original charge distribution. In this situation higher multipoles would certainly be irrelevant! Thus we have two extreme cases:

- a single origin and an infinite number of multipoles,
- or an infinite number of centres and monopoles (charges) only.

We see that when the origins are located on atoms, we have an intermediary situation and it might be sufficient to have a few multipoles per atom.[10] This is what the concept of what is called the *cumulative multipole moments* is all about (CAMM[11]). Besides the isotropic atomic charges $q_a = M_a^{(000)}$ calculated in an arbitrary population analysis, we have, in addition, higher multipoles $M_a^{(klm)}$ (atomic dipoles, quadrupoles, octupoles, etc.) representing the anisotropy of the atomic charge distribution (i.e. they describe the deviations of the atomic charge distributions from the spherical)

cumulative multipole moments

$$M_a^{(klm)} = Z_a x_a^k y_a^l z_a^m - \sum_{r\in a}\sum_{s} D_{sr}\left(r\middle|x^k y^l z^m\middle|s\right)$$
$$-\sum_{\substack{k'\leqslant k}}\sum_{l'\leqslant l}\sum_{\substack{m'\leqslant m,\\ (k',l',m')\neq(k,l,m)}} \binom{k}{k'}\binom{l}{l'}\binom{m}{m'}$$
$$\times\, x_a^{k-k'} y_a^{l-l'} z_a^{m-m'} \cdot M_a^{k'l'm'},$$

where $M_a^{(klm)}$ is the multipole moment of the "klm" order with respect to the Cartesian coordinates x, y, z located on atom a, $M_a^{(000)}$ stands for the atomic charge, e.g., from the Mulliken population analysis, Z_a is the nuclear charge of the atom a, $(r|x^k y^l z^m|s)$ stands for the one-electron integral of the corresponding multipole moment, and $D_{sr}\chi_r^*\chi_s$ represents the electronic density contribution related to AO's: χ_s and χ_r and calculated by any method (LCAO MO SCF, CI, MP2, DFT, etc.). We may also use multipole moments expressed by spherical harmonic functions as proposed by Stone.[12]

---

[9] The clouds might eventually overlap.

[10] If the clouds overlapped, the description of each centre by an infinite number of multipoles would lead to a redundancy ("overcompleteness"). I do not know of any trouble of that kind, but in my opinion trouble would come if the number of origins were large. This is in full analogy with the overcompleteness of the LCAO expansion. These two examples differ by a secondary feature: in the LCAO, instead of moments, we have the s, p, d, … orbitals, i.e. some moments multiplied by exponential functions.

[11] W.A. Sokalski and R. Poirier, *Chem. Phys. Lett.* 98 (1983) 86; W.A. Sokalski, A. Sawaryn, *J. Chem. Phys.* 87 (1987) 526.

[12] A.J. Stone, *Chem. Phys. Lett.* 83 (1981) 233; A.J. Stone, M. Alderton, *Mol. Phys.* 56 (1985) 1047.

T. THE DIPOLE MOMENT OF A LONE ELECTRON PAIR

The electronic lone pairs play an important role in intermolecular interactions. In particular, a lone pair protruding in space towards its partner has a large dipole moment,[1] which may interact electrostatically with its partner's multipole moments (see Appendix X, p. 1038). Let us see how the dipole moment depends on the atom to which it belongs and on the type of hybridization.

Suppose the electronic lone pair is described by the normalized hybrid

$$h = \frac{1}{\sqrt{1+\lambda^2}}\big[(2s) + \lambda(2p_x)\big],$$

with the normalized $2s$ and $2p_x$ atomic orbitals. The coefficient λ may change from $-\infty$ to $+\infty$ giving a different degree of hybridization. Fig. T.1 shows for comparison two series of the hybrids: for carbon and fluorine atoms. If $\lambda = 0$, we have the pure $2s$ orbital, if $\lambda = \pm\infty$ we obtain the pure $\pm 2p_x$ orbital.

The dipole moment of a single electron described by h is calculated[2] as ($N = \frac{1}{\sqrt{1+\lambda^2}}$):

$$\begin{aligned}
\mu_x &= \langle h|{-x}|h\rangle = -N^2\big[\langle 2s|x|2s\rangle + \lambda^2\langle 2p_x|x|2p_x\rangle + 2\lambda\langle 2s|x|2p_x\rangle\big],\\
\mu_y &= \mu_z = 0,
\end{aligned}$$

where x stands for the x coordinate of the electron.

The first two integrals equal zero, because the integrand represents an odd function[3] with respect to the reflection in the plane $x = 0$. As a result

$$\mu_x = -N^2 2\lambda\langle 2s|x|2p_x\rangle.$$

We will limit ourselves to $\lambda \geqslant 0$, which means we are considering hybrids protruding to the right-hand side[4] as in Fig. T.1, and since $\langle 2s|x|2p_x\rangle > 0$, then $\mu_x \leqslant 0$. The negative sign stresses the fact that a negative electron is displaced to the right-hand side (positive x).

---

[1] Calculated with respect to the nucleus; a large dipole moment means here a large length of the dipole moment vector.

[2] Atomic units have been used throughout, and therefore μ is expressed in a.u.

[3] Please recall that the orbital $2p_x$ represents a spherically symmetric factor multiplied by x.

[4] The hybrids with $\lambda < 0$ differ only by protruding to the left-hand side.

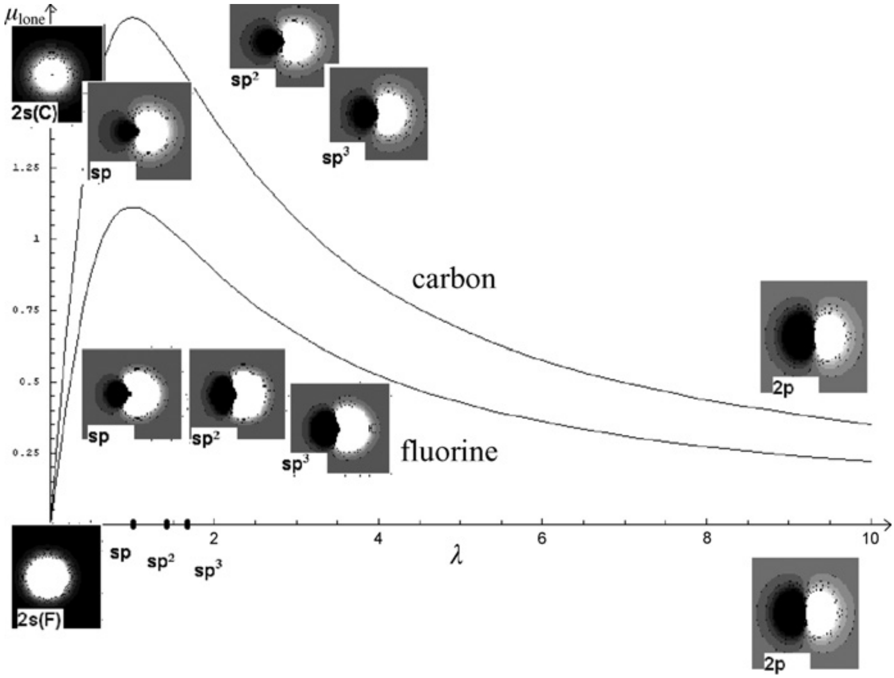


Fig. T.1. The length of the dipole moment vector μ_{lone} (in a.u.) as a function of the mixing parameter λ for carbon (upper curve) and fluorine (lower curve) atoms. The figure shows the shape of different hybrids $h = \frac{1}{\sqrt{1+\lambda^2}}[(2s) + \lambda(2p_x)]$ which correspond to various mixing of the $2s$ and $2p_x$ carbon Slater orbitals (with exponential factor $\zeta = 1.625$) and fluorine orbitals ($\zeta = 2.60$); from the left: $\lambda = 0$, $\lambda = 1$ (sp), $\lambda = 1.41$ (sp^2), $\lambda = 1.73$ (sp^3), $\lambda = 1000$. All the hybrids are shown in square windows of 10 a.u. The fluorine orbitals are more compact due to the larger charge of the nucleus. A hybrid orbital which corresponds to $\lambda < 0$ looks exactly like that with $\lambda' = -\lambda$, except it is reflected with respect to the yz plane. The maximum dipole moment corresponds to the sp hybridization.

To calculate $\langle 2s|x|2p_x\rangle$ *we need to specify the atomic orbitals* $2s$ *and* $2p$. As the $2s$ and $2p$ atomic orbitals, let us take Slater type orbitals:

$$2s = N' r \exp(-\zeta r),$$
$$2p_x = N'' x \exp(-\zeta r),$$

where the exponential factor ζ (the same for both orbitals) is calculated using simple rules for building the Slater orbitals, see p. 355.

Using the integral

$$\int_0^\infty x^n \exp(-\alpha x)\,\mathrm{d}x = n!\alpha^{-(n+1)},$$

we obtain the normalization constants $N' = \zeta^2\sqrt{\frac{\zeta}{3\pi}}$ and $N'' = \zeta^2\sqrt{\frac{\zeta}{\pi}}$. The contribution of two electrons ("lone electron pair") to the dipole moment is, therefore, equal to

$$
\begin{aligned}
\mu_{\text{lone}} = 2\mu_x &= -N^2|\lambda|(2s|xp_x) = -2N^2N'N''(2\lambda)\int r x^2 \exp(-2\zeta r)\,\mathrm{d}v \\
&= -2N^2N'N''2\lambda \int r^3 x^2 \exp(-2\zeta r)\sin\theta\,\mathrm{d}r\,\mathrm{d}\theta\,\mathrm{d}\phi \\
&= -2N^2N'N''2\lambda \int_0^\infty \mathrm{d}r\, r^5 \exp(-2\zeta r)\int_0^\pi \sin^3\theta\,\mathrm{d}\theta \int_0^{2\pi}\cos^2\phi\,\mathrm{d}\phi \\
&= -2N^2N'N''2\lambda \frac{5!}{(2\zeta)^6}\frac{4}{3}\pi = -\frac{4\lambda}{(1+\lambda^2)}\zeta^2\sqrt{\frac{\zeta}{3\pi}}\zeta^2\sqrt{\frac{\zeta}{\pi}}\frac{5!}{(2\zeta)^6}\frac{4}{3}\pi \\
&= -\frac{\lambda}{1+\lambda^2}\frac{10}{\zeta\sqrt{3}}.
\end{aligned}
$$

THE DIPOLE MOMENT OF A LONE PAIR $\mu_{\text{lone}} = -\frac{\lambda}{1+\lambda^2}\frac{10}{\zeta\sqrt{3}}$.

The dipole moment at $\lambda = 0$, i.e. for the pure $2s$ orbital, is equal to 0, for $\lambda = \infty$, i.e. for the pure $2p_x$ orbital it is also equal 0. It is interesting to see for which hybridization the length of dipole moment has a maximum. We easily find

$$
\frac{\partial|\mu_{\text{lone}}|}{\partial\lambda} = \frac{10}{\zeta\sqrt{3}}\frac{(1+\lambda^2)-2\lambda^2}{(1+\lambda^2)} = 0,
$$

which gives $\lambda = \pm 1$, independently of ζ.

Thus

the maximum dipole moment is at the $1:1$ mixing of $2s$ and $2p$, i.e. for digonal hybridization (for any element), Fig. T.1.

From Table T.1 it is seen that the dipole moment of a lone pair strongly depends on the chemical element,[5] and to a lesser extent on hybridization.

Table T.1. The length of the dipole moments μ_{lone} (a.u.) corresponding to doubly occupied hybrid atomic orbitals. The orbital exponents of $2s$ and $2p$ STO's are identical and calculated using the rules given by Slater: $\zeta_C = 1.625$, $\zeta_N = 1.95$, $\zeta_O = 2.275$, $\zeta_F = 2.60$

Atom	Digonal $\lambda = 1$	Trigonal $\lambda = \sqrt{2}$	Tetrahedral $\lambda = \sqrt{3}$
C	1.776	1.675	1.538
N	1.480	1.396	1.282
O	1.269	1.196	1.099
F	1.110	1.047	0.962

---

[5] From the practical point of view, it is probably most important to compare nitrogen and oxygen lone pairs. Thus, coordination of a cation by amines should correspond to a stronger interaction than that by hydroxyl groups.

U. SECOND QUANTIZATION

When we work with a basis set composed of Slater determinants we are usually confronted with a large number of matrix elements involving one- and two-electron operators. The Slater–Condon rules (Appendix M) are doing the job to express these matrix elements by the one-electron and two-electron integrals. However, we may introduce an even easier tool called second quantization, which is equivalent to the Slater–Condon rules.

The vacuum state

In the second quantization formalism we introduce a reference state for the system under study, which is a Slater determinant (usually the Hartree–Fock wave function) composed of N orthonormal spinorbitals, where N is the number of electrons. This function will be denoted in short by Φ_0 or in a more detailed way by $\Phi^N(n_1, n_2, \ldots, n_\infty)$. The latter notation means that we have to do with a normalized N electron Slater determinant, and in parenthesis we give the occupancy list ($n_i = 0, 1$) for the infinite number of orthonormal spinorbitals considered in the basis set and listed one by one in the parenthesis. This simply means that some spinorbitals are present in the determinant (they have $n_i = 1$), while others are absent[1] ($n_i = 0$). Hence, $\sum_i n_i = N$. The reference state is often called the *vacuum state*. The subscript 0 in Φ_0 means that we are going to consider a single-determinant approximation to the *ground state*. Besides the reference state, some normalized Slater determinants of the excited states will be considered, with other occupancies, *including those corresponding to the number of electrons which differs from N*.

The creation and annihilation of electrons

Let us make quite a strange move, and consider operators that change the number of electrons in the system. To this end, let us define the creation operator[2] $\hat{k}^\dagger$ of the electron going to occupy spinorbital k and the annihilation operator $\hat{k}$ of an electron leaving spinorbital k:

---

[1] For example, the symbol $\Phi^2(001000100000\ldots)$ means a normalized Slater determinant of dimension 2, containing the spinorbitals 3 and 7. The symbol $\Phi^2(001000\ldots)$ is a nonsense, because the number of "ones" has to be equal to 2, etc.

[2] The domain of the operators represents the space spanned by the Slater determinants built of spinorbitals.

Richard Feynman, in one of his books, says jokingly that he could not understand the very sense of the operators. If we annihilate or create an electron, then what about the system's electroneutrality? Happily enough, these operators will always act in creator–annihilator pairs.

CREATION AND ANNIHILATION OPERATORS

$$\hat{k}^{\dagger}\Phi^{N}(\ldots n_k \ldots) = \theta_k (1-n_k)\Phi^{N+1}(\ldots 1_k, \ldots),$$
$$\hat{k}\Phi^{N}(\ldots n_k \ldots) = \theta_k n_k \Phi^{N-1}(\ldots 0_k, \ldots),$$

where $\theta_k = (-1)^{\sum_{j<k} n_j}$.

The symbol 1_k means that the spinorbital k is present in the Slater determinant, while 0_k means that this spinorbital is empty, i.e. is not present in the Slater determinant. The factors $(1-n_k)$ and n_k ensure an important property of these operators, namely that

any attempt at *creating* an electron on an already *occupied spinorbital* gives zero, similarly any attempt at *annihilating an empty spinorbital* also gives zero.

It can be easily shown,[3] that (as the symbol suggests) $\hat{k}^{\dagger}$ is simply the adjoint operator with respect to $\hat{k}$.

The above operators have the following properties that make them equivalent to the Slater–Condon rules:

ANTICOMMUTATION RULES

$$[\hat{k}, \hat{l}]_+ = 0,$$
$$[\hat{k}^{\dagger}, \hat{l}^{\dagger}]_+ = 0,$$
$$[\hat{k}^{\dagger}, \hat{l}]_+ = \delta_{kl},$$

anticommutator

where the symbol $[\hat{A}, \hat{B}]_+ = \hat{A}\hat{B} + \hat{B}\hat{A}$ is called the *anticommutator*.[4] It is simpler than the Slater–Condon rules, isn't it? Let us check the rule $[\hat{k}^{\dagger}, \hat{l}]_+ = \delta_{kl}$. We have to check how it works for all possible occupancies of the spinorbitals k and l, (n_k, n_l): $(0,0)$, $(0,1)$, $(1,0)$ and $(1,1)$.

Case: $(n_k, n_l) = (0,0)$

$$[\hat{k}^{\dagger}, \hat{l}]_+ \Phi^{N}(\ldots 0_k \ldots 0_l \ldots) = [\hat{k}^{\dagger}\hat{l} + \hat{l}\hat{k}^{\dagger}]\Phi^{N}(\ldots 0_k \ldots 0_l \ldots)$$
$$= \hat{k}^{\dagger}\hat{l}\Phi^{N}(\ldots 0_k \ldots 0_l \ldots) + \hat{l}\hat{k}^{\dagger}\Phi^{N}(\ldots 0_k \ldots 0_l \ldots)$$

---

[3] *Proof*. Let us take two Slater determinants $\Phi_a = \Phi^{N+1}(\ldots 1_k \ldots)$ and $\Phi_b = \Phi^{N}(\ldots 0_k \ldots)$, in both of them the occupancies of all other spinorbitals are identical. Let us write the normalization condition for Φ_b in the following way: $1 = \langle \Phi_b | \theta_k \hat{k}\Phi_a \rangle = \theta_k \langle \Phi_b | \hat{k}\Phi_a \rangle = \theta_k \langle \hat{k}^{\#}\Phi_b | \Phi_a \rangle$, where as $\hat{k}^{\#}$ has been denoted the operator adjoint to $\hat{k}$, θ_k appeared in order to compensate for ($\theta_k^2 = 1$) the θ_k produced by the annihilator. On the other hand, from the normalization condition of Φ_a we see that $1 = \langle \Phi_a | \Phi_a \rangle = \theta_k \langle \hat{k}^{\dagger}\Phi_b | \Phi_a \rangle$. Hence, $\theta_k \langle \hat{k}^{\#}\Phi_b | \Phi_a \rangle = \theta_k \langle \hat{k}^{\dagger}\Phi_b | \Phi_a \rangle$ or $\hat{k}^{\#} = \hat{k}^{\dagger}$, This is what we wanted to show.

[4] The above formulae are valid under the (common) assumption that the spinorbitals are orthonormal. If this assumption is not true, only the last anticommutator changes to the form $[\hat{k}^{\dagger}, \hat{l}]_+ = S_{kl}$, where S_{kl} stands for the overlap integral of spinorbitals k and l.

$$
\begin{aligned}
&= 0 + \hat{l}\theta_k \Phi^{N+1}(\ldots 1_k \ldots 0_l \ldots) \\
&= \theta_k \hat{l} \Phi^{N+1}(\ldots 1_k \ldots 0_l \ldots) \\
&= \theta_k \delta_{kl} \theta_k \Phi^N(\ldots 0_k \ldots) \\
&= \delta_{kl} \Phi^N(\ldots 0_k \ldots).
\end{aligned}
$$

So far so good.

Case: $(n_k, n_l) = (0, 1)$

$$
\begin{aligned}
\left[\hat{k}^\dagger, \hat{l}\right]_+ \Phi^N(\ldots 0_k \ldots 1_l \ldots) &= \left[\hat{k}^\dagger \hat{l} + \hat{l}\hat{k}^\dagger\right] \Phi^N(\ldots 0_k \ldots 1_l \ldots) \\
&= \hat{k}^\dagger \hat{l} \Phi^N(\ldots 0_k \ldots 1_l \ldots) + \hat{l}\hat{k}^\dagger \Phi^N(\ldots 0_k \ldots 1_l \ldots) \\
&= \theta_k \theta_l \Phi^N(\ldots 1_k \ldots 0_l \ldots) - \theta_k \theta_l \Phi^N(\ldots 1_k \ldots 0_l \ldots) \\
&= \delta_{kl} \Phi^N(\ldots 0_k \ldots 1_l \ldots).
\end{aligned}
$$

This is what we expected.[5]

Case: $(n_k, n_l) = (1, 0)$

$$
\begin{aligned}
\left[\hat{k}^\dagger, \hat{l}\right]_+ \Phi^N(\ldots 1_k \ldots 0_l \ldots) &= \left[\hat{k}^\dagger \hat{l} + \hat{l}\hat{k}^\dagger\right] \Phi^N(\ldots 1_k \ldots 0_l \ldots) \\
&= \hat{k}^\dagger \hat{l} \Phi^N(\ldots 1_k \ldots 0_l \ldots) + \hat{l}\hat{k}^\dagger \Phi^N(\ldots 1_k \ldots 0_l \ldots) \\
&= (0 + 0) \Phi^N(\ldots 1_k \ldots 0_l \ldots) \\
&= \delta_{kl} \Phi^N(\ldots 1_k \ldots 0_l \ldots).
\end{aligned}
$$

This is OK.

Case: $(n_k, n_l) = (1, 1)$

$$
\begin{aligned}
\left[\hat{k}^\dagger, \hat{l}\right]_+ \Phi^N(\ldots 1_k \ldots 1_l \ldots) &= \left[\hat{k}^\dagger \hat{l} + \hat{l}\hat{k}^\dagger\right] \Phi^N(\ldots 1_k \ldots 1_l \ldots) \\
&= \hat{k}^\dagger \hat{l} \Phi^N(\ldots 1_k \ldots 1_l \ldots) + \hat{l}\hat{k}^\dagger \Phi^N(\ldots 1_k \ldots 1_l \ldots) \\
&= \hat{k}^\dagger \hat{l} \Phi^N(\ldots 1_k \ldots 1_l \ldots) + 0 \\
&= \theta_k^2 \delta_{kl} \Phi^N(\ldots 1_k \ldots 1_l \ldots) \\
&= \delta_{kl} \Phi^N(\ldots 1_k \ldots).
\end{aligned}
$$

The formula has been proved.

Operators in the second quantization

Creation and annihilation operators may be used to represent one- and two-electron operators.[6] The resulting matrix elements with Slater determinants *correspond exactly to the Slater–Condon rules* (see Appendix M, p. 986).

---

[5] What decided is the change of sign (due to θ_k) when the order of the operators has changed.

[6] The original operator and its representation in the language of the second quantization are not identical in practical applications. The second ones can act only on the Slater determinants or their combinations. Since we are going to work with the creation and annihilation operators in only those methods which use Slater determinants (CI, MC SCF, etc.), the difference is irrelevant.

One-electron operators

The operator $\hat{F}=\sum_i \hat{h}(i)$ is the sum of the one-electron operators[7] $\hat{h}(i)$ acting on functions of the coordinates of electron i.

The I Slater–Condon rule says (see Appendix M), that for the Slater determinant ψ built of the spinorbitals ϕ_i, the matrix element $\langle\psi|\hat{F}\psi\rangle=\sum_i h_{ii}$, where $h_{ij}=\langle\phi_i|\hat{h}\phi_j\rangle$.

In the second quantization

$$\hat{F}=\sum_{ij}^{\infty} h_{ij}\hat{i}^{\dagger}\hat{j}.$$

Interestingly, the summation extends to infinity, and therefore the operator is independent of the number of electrons in the system.

Let us check whether the formula is correct. Let us insert $\hat{F}=\sum_{ij} h_{ij}\hat{i}^{\dagger}\hat{j}$ into $\langle\psi|\hat{F}\psi\rangle$. We have

$$\langle\psi|\hat{F}\psi\rangle=\left\langle\psi\middle|\sum_{ij} h_{ij}\hat{i}^{\dagger}\hat{j}\psi\right\rangle=\sum_{ij} h_{ij}\langle\psi|\hat{i}^{\dagger}\hat{j}\psi\rangle=\sum_{ij} h_{ij}\delta_{ij}=\sum_i h_{ii}.$$

This is correct.

What about the II Slater–Condon rule (the Slater determinants ψ_1 and ψ_2 differ by a single spinorbital: the spinorbital i in ψ_1 is replaced by the spinorbital i' in ψ_2)? We have

$$\langle\psi_1|\hat{F}\psi_2\rangle=\sum_{ij} h_{ij}\langle\psi_1|\hat{i}^{\dagger}\hat{j}\psi_2\rangle.$$

The Slater determinants that differ by one spinorbital produce an overlap integral equal to zero,[8] therefore $\langle\psi_1|\hat{F}\psi_2\rangle=h_{ii'}$. Thus, the operator in the form $\hat{F}=\sum_{ij} h_{ij}\hat{i}^{\dagger}\hat{j}$ ensures equivalence with all the Slater–Condon rules.

Two-electron operators

Similarly, we may use the creation and annihilation operators to represent the two-electron operators $\hat{G}=\frac{1}{2}\sum_{ij}' \hat{g}(i,j)$. In most cases $\hat{g}(i,j)=\frac{1}{r_{ij}}$ and $\hat{G}$ has the form:

$$\hat{G}=\frac{1}{2}\sum_{ij}{}' \frac{1}{r_{ij}}=\frac{1}{2}\sum_{ijkl}^{\infty}\langle ij|kl\rangle\hat{j}^{\dagger}\hat{i}^{\dagger}\hat{k}\hat{l}.$$

---

[7] Most often this will be the kinetic energy operator, the nuclear attraction operator, the interaction with the external field or the multipole moment.

[8] It is evident, that if in this situation the Slater determinants ψ_1 and ψ_2 differed by more than a single spinorbital, we would get zero (III and IV Slater–Condon rule).

Here also the summation extends to infinity and the operator is independent of the number of electrons in the system.

The proof of the I Slater–Condon rule relies on the following chain of equalities

$$\begin{aligned}\langle\psi|\hat{G}\psi\rangle &= \frac{1}{2}\sum_{ijkl}\langle ij|kl\rangle\langle\psi|\hat{\jmath}^{\dagger}\hat{\imath}^{\dagger}\hat{k}\hat{l}\psi\rangle = \frac{1}{2}\sum_{ijkl}\langle ij|kl\rangle\langle\hat{\imath}\hat{\jmath}\psi|\hat{k}\hat{l}\psi\rangle\\ &= \frac{1}{2}\sum_{ijkl}\langle ij|kl\rangle(\delta_{ik}\delta_{jl}-\delta_{il}\delta_{jk}) = \frac{1}{2}\sum_{ij}\big(\langle ij|ij\rangle-\langle ij|ji\rangle\big),\end{aligned}$$

because the overlap integral $\langle\hat{\imath}\hat{\jmath}\psi|\hat{k}\hat{l}\psi\rangle$ of the two Slater determinants $\hat{\imath}\hat{\jmath}\psi$ and $\hat{k}\hat{l}\psi$ is non-zero in the two cases only: either if $i=k$, $j=l$ or if $i=l$, $j=k$ (then the sign has to change). This is what we get from the Slater–Condon rules.

For the II Slater–Condon rule we have (instead of the spinorbital i in ψ_1 we have the spinorbital i' in ψ_2):

$$\langle\psi_1|\hat{G}\psi_2\rangle = \frac{1}{2}\sum_{Ijkl}\langle Ij|kl\rangle\langle\psi_1|\hat{\jmath}^{\dagger}\hat{I}^{\dagger}\hat{k}\hat{l}\psi_2\rangle = \frac{1}{2}\sum_{Ijkl}\langle Ij|kl\rangle\langle\hat{I}\hat{\jmath}\psi_1|\hat{k}\hat{l}\psi_2\rangle, \tag{U.1}$$

where the summation index I has been introduced in order not to mix with spinorbital i. In the overlap integral $\langle\hat{I}\hat{\jmath}\psi_1|\hat{k}\hat{l}\psi_2\rangle$ the sets of the spinorbitals in the Slater determinant $\hat{I}\hat{\jmath}\psi_1$ and in the Slater determinant $\hat{k}\hat{l}\psi_2$ have to be *identical*, otherwise the integral will equal zero. However, in ψ_1 and ψ_2 we already have a difference of one spinorbital. Thus, first we have to get rid of these spinorbitals (i and i'). For the integral to survive[9] we have to have at least one of the following conditions satisfied:

- $I=i$ and $k=i'$ (and then $j=l$),
- $j=i$ and $k=i'$ (and then $I=l$),
- $I=i$ and $l=i'$ (and then $j=k$),
- $j=i$ and $l=i'$ (and then $I=k$).

This means that, taking into account the above cases in eq. (U.1), we obtain

$$\begin{aligned}\langle\psi_1|\hat{G}\psi_2\rangle &= \frac{1}{2}\sum_{j}\langle ij|i'j\rangle\langle\hat{\imath}\hat{\jmath}\psi_1|\hat{\imath}'\hat{\jmath}\psi_2\rangle + \frac{1}{2}\sum_{l}\langle li|i'l\rangle\langle\hat{l}\hat{\imath}\psi_1|\hat{\imath}'\hat{l}\psi_2\rangle\\ &\quad+\frac{1}{2}\sum_{j}\langle ij|ji'\rangle\langle\hat{\imath}\hat{\jmath}\psi_1|\hat{\jmath}\hat{\imath}'\psi_2\rangle + \frac{1}{2}\sum_{k}\langle ki|ki'\rangle\langle\hat{k}\hat{\imath}\psi_1|\hat{k}\hat{\imath}'\psi_2\rangle\\ &= \frac{1}{2}\sum_{j}\langle ij|i'j\rangle - \frac{1}{2}\sum_{l}\langle li|i'l\rangle - \frac{1}{2}\sum_{j}\langle ij|ji'\rangle + \frac{1}{2}\sum_{k}\langle ki|ki'\rangle\\ &= \frac{1}{2}\sum_{j}\langle ij|i'j\rangle - \frac{1}{2}\sum_{j}\langle ji|i'j\rangle - \frac{1}{2}\sum_{j}\langle ij|ji'\rangle + \frac{1}{2}\sum_{j}\langle ji|ji'\rangle\end{aligned}$$

---

[9] This is a necessary, but not a sufficient condition.

$$= \frac{1}{2}\sum_j \langle ij|i'j\rangle - \frac{1}{2}\sum_j \langle ij|ji'\rangle - \frac{1}{2}\sum_j \langle ij|ji'\rangle + \frac{1}{2}\sum_j \langle ij|i'j\rangle$$

$$= \sum_j \langle ij|i'j\rangle - \sum_j \langle ij|ji'\rangle,$$

where the coordinates of electrons 1 and 2 have been exchanged in the two sums. Notice that the overlap integrals

$$\langle \hat{i}\hat{j}\psi_1|\hat{i}'\hat{j}\psi_2\rangle = \langle \hat{k}\hat{i}\psi_1|\hat{k}\hat{i}'\psi_2\rangle = 1,$$

because the Slater determinants $\hat{i}\psi_1$ and $\hat{i}'\psi_2$ are identical. Also, from the anticommutation rules

$$\langle \hat{l}\hat{i}\psi_1|\hat{i}'\hat{l}\psi_2\rangle = \langle \hat{i}\hat{j}\psi_1|\hat{j}\hat{i}'\psi_2\rangle = -1.$$

Thus the II Slater–Condon rule has been correctly reproduced:

$$\langle \psi_1|\hat{G}\psi_2\rangle = \sum_j \left[\langle ij|i'j\rangle - \langle ij|ji'\rangle\right].$$

We may conclude that the definition of the creation and annihilation operators and the simple anticommutation relations are equivalent to the Slater–Condon rules. This opens up the space spanned by the Slater determinants for us, i.e. all the integrals involving Slater determinants can be easily transformed into one- and two-electron integrals involving spinorbitals.

V. THE HYDROGEN ATOM IN THE ELECTRIC FIELD – VARIATIONAL APPROACH

Polarization of an atom or molecule can be calculated by using the finite field method described on p. 639. Let us apply this method to the hydrogen atom. Its polarizability was already calculated using a simple version of perturbation theory (p. 636). This time we will use the variational method.

The Hamiltonian for the isolated hydrogen atom (within the Born–Oppenheimer approximation) reads as

$$\hat{H}^{(0)} = -\frac{1}{2}\Delta_e - \frac{1}{r},$$

where the first term is the electronic kinetic energy operator, and the second its Coulomb interaction energy with the nucleus (proton–electron distance is denoted by r). The atom is in a homogeneous electric field $\boldsymbol{\mathcal{E}} = (0, 0, \mathcal{E})$ with $\mathcal{E} > 0$ and as in perturbation theory (p. 636), the total Hamiltonian has the form

$$\hat{H} = \hat{H}^{(0)} + V$$

with $V = z\mathcal{E}$, where z denotes the coordinate of the electron and the proton is at the origin (the derivation of the formula is given on p. 636, the exchange of z to x does not matter).

The variational wave function ψ is proposed in the form

$$\psi = \chi_1 + c\chi_2, \tag{V.1}$$

where $\chi_1 = \frac{1}{\sqrt{\pi}} \exp(-r)$ is the 1s orbital of the hydrogen atom (ground state) and χ_2 is the normalized[1] p-type orbital

$$\chi_2 = Nz\exp(-\zeta r).$$

---

[1] N can be easily calculated from the normalization condition

$$1 = N^2 \int \left[z\exp(-\zeta r)\right]^2 \mathrm{d}V = N^2 \int_0^\infty \mathrm{d}r\, r^4 \exp(-2\zeta r) \int_0^\pi \mathrm{d}\theta \sin\theta \cos^2\theta \int_0^{2\pi} \mathrm{d}\phi$$
$$= N^2 4!\,(2\zeta)^{-5}\,\frac{2}{3} 2\pi = N^2 \frac{\pi}{\zeta^5}.$$

This gives $N = \sqrt{\frac{\zeta^5}{\pi}}$.

There are two variational parameters c and ζ. Let us assume for a while that we have fixed the value of ζ, so the only variational parameter is c. The wave function ψ is a linear combination of two expansion functions ("two-state model"): χ_1 and χ_2. Therefore, optimal energy follows from the Ritz method, according to case III of Appendix D on p. 948:

$$E = E_{\text{ar}} \pm \sqrt{\Delta^2 + h^2}, \tag{V.2}$$

where arithmetic mean energy $E_{\text{ar}} \equiv \frac{H_{11}+H_{22}}{2}$, while $\Delta \equiv \frac{H_{11}-H_{22}}{2}$ and $h \equiv H_{12} = H_{21}$ with

$$H_{ij} \equiv \langle \chi_i | \hat{H} \chi_j \rangle = \langle \chi_i | \hat{H}^{(0)} \chi_j \rangle + \langle \chi_i | V \chi_j \rangle.$$

Let us calculate all the ingredients of the energy given by (V.2).

First, let us note that $H_{11} \equiv \langle \chi_1 | \hat{H}^{(0)} \chi_1 \rangle = -\frac{1}{2}$ a.u., since χ_1 is the ground state of the isolated hydrogen atom (p. 178), and $V_{11} = \langle \chi_1 | V \chi_1 \rangle = 0$, because the integrand is antisymmetric with respect to $z \to -z$.

Now let us calculate $H_{22} = H_{22}^{(0)} + V_{22}$. Note that $V_{22} = 0$, for the same reason as V_{11}. We have

$$H_{22}^{(0)} = -\frac{1}{2} \langle \chi_2 | \Delta_e \chi_2 \rangle - \left\langle \chi_2 \middle| \frac{1}{r} \chi_2 \right\rangle.$$

The second integral is

$$\begin{aligned}\left\langle \chi_2 \middle| \frac{1}{r} \chi_2 \right\rangle &= N^2 \int_0^\infty \mathrm{d}r\, r^3 \exp(-2\zeta r) \int_0^\pi \mathrm{d}\theta \sin\theta \cos^2\theta \int_0^{2\pi} \mathrm{d}\phi \\ &= \frac{\zeta^5}{\pi} \cdot 3!(2\zeta)^{-4} \cdot \frac{2}{3} \cdot 2\pi = \frac{1}{2}\zeta,\end{aligned}$$

where the dots separate the values of the corresponding integrals.[2] In Appendix R, the reader will find the main ingredients needed to calculate the first integral of $H_{22}^{(0)}$:

$$\begin{aligned}\langle \chi_2 | \Delta_e \chi_2 \rangle &= N^2 \left\langle r\cos\theta \exp(-\zeta r) \middle| \left[\frac{1}{r^2} \frac{\partial}{\partial r} r^2 \frac{\partial}{\partial r} + \frac{1}{r^2 \sin\theta} \frac{\partial}{\partial \theta} \sin\theta \frac{\partial}{\partial \theta} \right.\right. \\ &\qquad \left.\left. + \frac{1}{r^2 \sin^2\theta} \frac{\partial^2}{\partial \phi^2} \right] r\cos\theta \exp(-\zeta r) \right\rangle \\ &= N^2 \left[\begin{array}{c} \langle r\cos\theta \exp(-\zeta r) | \cos\theta \frac{1}{r^2} \frac{\partial}{\partial r} [r^2 \exp(-\zeta r) - \zeta r^3 \exp(-\zeta r) \rangle + \\ \langle r\cos\theta \exp(-\zeta r) | \frac{(-2\cos\theta)}{r^2} r \exp(-\zeta r) \rangle + 0 \end{array} \right] \\ &= N^2 \left[\left\langle r\cos\theta \exp(-\zeta r) \middle| \cos\theta \left[\frac{2}{r} - \zeta - 3\zeta + \zeta^2 r \right] \exp(-\zeta r) \right\rangle \right.\end{aligned}$$

---

[2] Note that, in spherical coordinates, the volume element $\mathrm{d}V = r^2 \sin\theta\, \mathrm{d}r\, \mathrm{d}\theta\, \mathrm{d}\phi$. In derivations of this Appendix (and not only) we often use the equality $\int_0^\infty \mathrm{d}x\, x^n \exp(-\alpha r) = n! \alpha^{-(n+1)}$.

$$
\begin{aligned}
&+\left\langle r\cos\theta\exp(-\zeta r)\middle|\frac{(-2\cos\theta)}{r}\exp(-\zeta r)\right\rangle\Bigg]\\
&=\frac{\zeta^5}{\pi}\left(\frac{2}{3}\cdot 2\pi\right)\left[\frac{2\cdot 2}{(2\zeta)^3}-\frac{4\zeta\cdot 3!}{(2\zeta)^4}+\frac{\zeta^2\cdot 4!}{(2\zeta)^5}-\frac{2\cdot 2!}{(2\zeta)^3}\right]\\
&=-\zeta^2.
\end{aligned}
$$

Thus, we obtain $H_{22}=\frac{1}{2}\zeta^2-\frac{1}{2}\zeta$. This formula looks good, since for $\chi_2=2p_z$, i.e. for $\zeta=\frac{1}{2}$ we get correctly (see p. 178) $H_{22}=E_{2p}=-\frac{1}{8}$ a.u., the energy of orbital $2p$.

Let us turn to the non-diagonal matrix element of the Hamiltonian: $H_{12}=H_{12}^{(0)}+V_{12}$. Note, that $H_{12}^{(0)}=0$, because χ_1 is an eigenfunction of $\hat{H}^{(0)}$ and $\langle\chi_1|\chi_2\rangle=0$. Thus,

$$
\begin{aligned}
h&=N\mathcal{E}\left\langle r\cos\theta\exp(-\zeta)\middle|r\cos\theta\frac{1}{\sqrt{\pi}}\exp(-r)\right\rangle\\
&=N\mathcal{E}\frac{1}{\sqrt{\pi}}\int_0^\infty \mathrm{d}r\, r^4\exp\bigl[-(\zeta+1)r\bigr]\int_0^\pi \mathrm{d}\theta\,\sin\theta\cos^2\theta\int_0^{2\pi}\mathrm{d}\phi\\
&=\mathcal{E}\frac{\sqrt{\zeta^5}}{\pi}\cdot 4!(\zeta+1)^{-5}\cdot\frac{2}{3}\cdot 2\pi=32\frac{\sqrt{\zeta^5}}{(\zeta+1)^5}\mathcal{E}.
\end{aligned}
$$

Now we can write eq. (V.2) as a function of ζ:

$$
E=\frac{1}{4}\bigl(\zeta^2-\zeta-1\bigr)-\sqrt{\frac{1}{16}\bigl(\zeta^2-\zeta+1\bigr)^2+\zeta^5\left(\frac{2}{\zeta+1}\right)^{10}\mathcal{E}^2}.\tag{V.3}
$$

We would like to expand this expression in a power series of $\mathcal{E}$ to highlight the coefficient at $\mathcal{E}^2$, because this coefficient is related to the polarizability. The expansion gives (in a.u.)

$$
E\approx\frac{1}{4}\bigl(\zeta^2-\zeta-1\bigr)-\frac{1}{4}\bigl(\zeta^2-\zeta+1\bigr)-\frac{1}{2}\alpha_{zz}\mathcal{E}^2+\cdots=-\frac{1}{2}-\frac{1}{2}\alpha_{zz}\mathcal{E}^2+\cdots,
$$

where, according to eq. (12.24), the polarizability (in a.u.) reads as

$$
\alpha_{zz}=4\cdot\frac{\zeta^5}{|\zeta^2-\zeta+1|}\left(\frac{2}{\zeta+1}\right)^{10}.\tag{V.4}
$$

Several numerical values of α_{zz} calculated using (V.3) and (V.4), are given on p. 639. They may be compared with the exact result $\alpha_{zz}=4.5$ a.u.

W. NMR SHIELDING AND COUPLING CONSTANTS – DERIVATION

This section is for those who do not fully believe the author, and want to check whether the final formulae for the shielding and coupling constants in nuclear magnetic resonance are indeed valid (Chapter 12).

1 SHIELDING CONSTANTS

Let us begin with eq. (12.87).

Applying vector identities

We are going to apply some vector identities[1] in the operators $\hat{B}_3, \hat{B}_4, \hat{B}_5$. The first identity is $\boldsymbol{u} \cdot (\boldsymbol{v} \times \boldsymbol{w}) = \boldsymbol{v} \cdot (\boldsymbol{w} \times \boldsymbol{u}) = \boldsymbol{w} \cdot (\boldsymbol{u} \times \boldsymbol{v})$, which simply means three equivalent ways of calculating the volume of a parallelepiped (cf. p. 437). This identity applied to $\hat{B}_3$ and $\hat{B}_4$ gives

$$\hat{B}_3 = \frac{e}{mc} \sum_A \sum_j \gamma_A \frac{\boldsymbol{I}_A \cdot \hat{\boldsymbol{L}}_{Aj}}{r_{Aj}^3}, \tag{W.1}$$

$$\hat{B}_4 = \frac{e}{2mc} \sum_j \boldsymbol{H} \cdot \hat{\boldsymbol{L}}_{0j}. \tag{W.2}$$

Let us transform the term $\hat{B}_5$ by using the following identity $(\boldsymbol{u} \times \boldsymbol{v}) \cdot (\boldsymbol{w} \times \boldsymbol{s}) = (\boldsymbol{u} \cdot \boldsymbol{w})(\boldsymbol{v} \cdot \boldsymbol{s}) - (\boldsymbol{v} \cdot \boldsymbol{w})(\boldsymbol{u} \cdot \boldsymbol{s})$:

$$\begin{aligned}
\hat{B}_5 &= \frac{e^2}{2mc^2} \sum_A \sum_j \gamma_A (\boldsymbol{H} \times \boldsymbol{r}_{0j}) \cdot \frac{\boldsymbol{I}_A \times \boldsymbol{r}_{Aj}}{r_{Aj}^3} \\
&= \frac{e^2}{2mc^2} \sum_A \sum_j \gamma_A \big[(\boldsymbol{H} \cdot \boldsymbol{I}_A)(\boldsymbol{r}_{0j} \cdot \boldsymbol{r}_{Aj}) - (\boldsymbol{r}_{0j} \cdot \boldsymbol{I}_A)(\boldsymbol{H} \cdot \boldsymbol{r}_{Aj})\big] \cdot \frac{1}{r_{Aj}^3}.
\end{aligned}$$

Putting things together

Now we are all set to put all this baroque furniture into its place, i.e. into eq. (12.87) for ΔE

---

[1] The reader may easily check each of them.

$$\Delta E = \sum_A \Delta E_A, \tag{W.3}$$

where ΔE_A stands for the contribution of nucleus A:

$$\begin{aligned}
\Delta E_A = {} & -\gamma_A \langle \psi_0^{(0)} | (\boldsymbol{I}_A \cdot \boldsymbol{H}) \psi_0^{(0)} \rangle \\
& + \frac{e^2}{2mc^2} \gamma_A \left\langle \psi_0^{(0)} \middle| \sum_j \left[(\boldsymbol{H} \cdot \boldsymbol{I}_A)(\boldsymbol{r}_{0j} \cdot \boldsymbol{r}_{Aj}) - (\boldsymbol{r}_{0j} \cdot \boldsymbol{I}_A)(\boldsymbol{H} \cdot \boldsymbol{r}_{Aj}) \right] \cdot \frac{1}{r_{Aj}^3} \psi_0^{(0)} \right\rangle \\
& + \frac{e^2}{2m^2c^2} \gamma_A \left[\left\langle \psi_0^{(0)} \middle| \left(\sum_j \frac{\boldsymbol{I}_A \cdot \hat{\boldsymbol{L}}_{Aj}}{r_{Aj}^3} \right) \hat{R}_0 \left(\sum_j \boldsymbol{H} \cdot \hat{\boldsymbol{L}}_{0j} \right) \psi_0^{(0)} \right\rangle \right. \\
& \left. + \left\langle \psi_0^{(0)} \middle| \left(\sum_j \boldsymbol{H} \cdot \hat{\boldsymbol{L}}_{0j} \right) \hat{R}_0 \left(\sum_j \frac{\boldsymbol{I}_A \cdot \hat{\boldsymbol{L}}_{Aj}}{r_{Aj}^3} \right) \psi_0^{(0)} \right\rangle \right].
\end{aligned}$$

Averaging over rotations

The expression for ΔE_A represents a bilinear form with respect to the components of vectors $\boldsymbol{I}_A$ and $\boldsymbol{H}$

$$\Delta E_A = \boldsymbol{I}_A^T \boldsymbol{C}_A \boldsymbol{H},$$

where $\boldsymbol{C}_A$ stands for a square matrix[2] of dimension 3, and $\boldsymbol{I}_A$ and $\boldsymbol{H}$ are vertical three-component vectors.

A contribution to the energy such as ΔE_A cannot depend on our choice of coordinate system axes x, y, z, i.e. on the components of $\boldsymbol{I}_A$ and $\boldsymbol{H}$. We will obtain the same energy if we rotate the axes (orthogonal transformation) in such a way as to diagonalize $\boldsymbol{C}_A$. The resulting diagonalized matrix $\boldsymbol{C}_{A,\mathrm{diag}}$ has three eigenvalues (composing the diagonal) corresponding to the new axes x', y', z'. *The very essence of averaging is that none of these axes are to be privileged in any sense.* This is achieved by constructing the averaged matrix

$$\begin{aligned}
& \frac{1}{3} \left[(\boldsymbol{C}_{A,\mathrm{diag}})_{x'x'} + (\boldsymbol{C}_{A,\mathrm{diag}})_{y'y'} + (\boldsymbol{C}_{A,\mathrm{diag}})_{z'z'} \right] \\
& \quad = \left(\bar{\boldsymbol{C}}_{A,\mathrm{diag}} \right)_{x'x'} = \left(\bar{\boldsymbol{C}}_{A,\mathrm{diag}} \right)_{y'y'} = \left(\bar{\boldsymbol{C}}_{A,\mathrm{diag}} \right)_{z'z'} \equiv C_A
\end{aligned}$$

where $(\bar{\boldsymbol{C}}_{A,\mathrm{diag}})_{qq'} = \delta_{qq'} C_A$ for $q, q' = x', y', z'$. Note that since the transformation was orthogonal (i.e. the trace of the matrix is preserved), the number C_A may also be obtained from the original matrix $\boldsymbol{C}_A$

$$\begin{aligned}
C_A &= \frac{1}{3} \left[(\boldsymbol{C}_{A,\mathrm{diag}})_{x'x'} + (\boldsymbol{C}_{A,\mathrm{diag}})_{y'y'} + (\boldsymbol{C}_{A,\mathrm{diag}})_{z'z'} \right] \\
&= \frac{1}{3} [\boldsymbol{C}_{A,xx} + \boldsymbol{C}_{A,yy} + \boldsymbol{C}_{A,zz}].
\end{aligned} \tag{W.4}$$

---

[2] We could write its elements from equation for ΔE_A, but their general form will turn out to be not necessary.

Then the averaged energy ΔE becomes (note the resulting dot product)

$$\Delta\bar{E} = \sum_A \boldsymbol{I}_A^T \bar{\boldsymbol{C}}_{A,\mathrm{diag}} \boldsymbol{H} = \sum_A C_A (\boldsymbol{I}_A \cdot \boldsymbol{H}).$$

Thus we obtain the sum of energy contributions over the nuclei, each contribution with its own coefficient averaged over rotations[3]

$$\begin{aligned}\Delta\bar{E} = {} & -\sum_A \gamma_A \boldsymbol{I}_A \cdot \boldsymbol{H} \biggl\{ 1 - \frac{e^2}{2mc^2} \biggl\langle \psi_0^{(0)} \biggl| \sum_j \frac{2}{3} (\boldsymbol{r}_{0j} \cdot \boldsymbol{r}_{Aj}) \frac{1}{r_{Aj}^3} \psi_0^{(0)} \biggr\rangle \\ & - \frac{e^2}{2m^2c^2} \frac{1}{3} \biggl\langle \psi_0^{(0)} \biggl| \biggl[\biggl(\sum_j \frac{\hat{\boldsymbol{L}}_{Aj}}{r_{Aj}^3} \biggr) \hat{R}_0 \biggl(\sum_j \hat{\boldsymbol{L}}_{0j} \biggr) + \biggl(\sum_j \hat{\boldsymbol{L}}_{0j} \biggr) \hat{R}_0 \biggl(\sum_j \frac{\hat{\boldsymbol{L}}_{Aj}}{r_{Aj}^3} \biggr) \biggr] \psi_0^{(0)} \biggr\rangle \biggr\},\end{aligned} \tag{W.5}$$

with the matrix elements

$$(\hat{\boldsymbol{U}})_{kl} = \langle \psi_k^{(0)} | \hat{\boldsymbol{U}} \psi_l^{(0)} \rangle$$

of the corresponding operators $\hat{\boldsymbol{U}} = (\hat{U}_x, \hat{U}_y, \hat{U}_z)$.

Finally, after comparing the formula with eq. (12.80), we obtain the shielding constant for nucleus A (the change of sign in the second part of the formula comes from the change in the denominator) given in eq. (12.88).

---

[3] Indeed, making $C_A = \frac{1}{3}[\boldsymbol{C}_{A,xx} + \boldsymbol{C}_{A,yy} + \boldsymbol{C}_{A,zz}]$ for the terms of eq. (W.3) we have the following contributions (term by term):

- $-\gamma_A \frac{1}{3}[1+1+1] = -\gamma_A$;
- $$\begin{aligned}& \frac{e^2}{2mc^2}\gamma_A \frac{1}{3}\biggl[\biggl\langle \psi_0^{(0)} \biggl| \sum_j \boldsymbol{r}_{0j} \cdot \boldsymbol{r}_{Aj} \frac{1}{r_{Aj}^3} \psi_0^{(0)} \biggr\rangle + \biggl\langle \psi_0^{(0)} \biggl| \sum_j \boldsymbol{r}_{0j} \cdot \boldsymbol{r}_{Aj} \frac{1}{r_{Aj}^3} \psi_0^{(0)} \biggr\rangle \\ & + \biggl\langle \psi_0^{(0)} \biggl| \sum_j \boldsymbol{r}_{0j} \cdot \boldsymbol{r}_{Aj} \frac{1}{r_{Aj}^3} \psi_0^{(0)} \biggr\rangle \biggr] = \frac{e^2}{2mc^2}\gamma_A \biggl\langle \psi_0^{(0)} \biggl| \sum_j \boldsymbol{r}_{0j} \cdot \boldsymbol{r}_{Aj} \frac{1}{r_{Aj}^3} \psi_0^{(0)} \biggr\rangle;\end{aligned}$$
- $$\begin{aligned}& -\frac{e^2}{2mc^2}\gamma_A \biggl\langle \psi_0^{(0)} \biggl| \sum_j \frac{1}{3}[x_{0j}x_{Aj} + y_{0j}y_{Aj} + z_{0j}z_{Aj}] \frac{1}{r_{Aj}^3} \psi_0^{(0)} \biggr\rangle \\ & = -\frac{e^2}{2mc^2}\gamma_A \biggl\langle \psi_0^{(0)} \biggl| \sum_j \frac{1}{3} \boldsymbol{r}_{0j} \cdot \boldsymbol{r}_{Aj} \frac{1}{r_{Aj}^3} \psi_0^{(0)} \biggr\rangle + \frac{1}{3}\frac{e^2}{2m^2c^2}\gamma_A {\sum_k}' \frac{1}{E_0^{(0)} - E_k^{(0)}} \\ & \quad \times \biggl[\biggl\langle \psi_0^{(0)} \biggl| \biggl(\sum_j \frac{\hat{L}_{Ajx}}{r_{Aj}^3} \biggr) \psi_k^{(0)} \biggr\rangle \biggl\langle \psi_k^{(0)} \biggl| \sum_j \hat{L}_{0jx} \psi_0^{(0)} \biggr\rangle + \text{similarly } y, z + \text{cc} \biggr] \\ & = \frac{1}{3}\frac{e^2}{2m^2c^2}\gamma_A {\sum_k}' \frac{1}{E_0^{(0)} - E_k^{(0)}} \times \frac{1}{3}\biggl[\biggl\langle \psi_0^{(0)} \biggl| \biggl(\sum_j \frac{\hat{\boldsymbol{L}}_{Aj}}{r_{Aj}^3} \biggr) \psi_k^{(0)} \biggr\rangle \biggl\langle \psi_k^{(0)} \biggl| \sum_j \hat{\boldsymbol{L}}_{0j} \psi_0^{(0)} \biggr\rangle + \text{cc} \biggr],\end{aligned}$$

where cc means the "complex conjugate" counterpart. This reproduces eq. (W.5).

2 COUPLING CONSTANTS

Averaging over rotations

In each contribution on p. 670 there is a double summation over the nuclear spins, which, after averaging over rotations (as for the shielding constant) gives rise to an energy dependence of the kind

$$\sum_{A<B} \gamma_A \gamma_B K_{AB} (\hat{\boldsymbol{I}}_A \cdot \hat{\boldsymbol{I}}_B),$$

which is required in the NMR Hamiltonian. Now, let us take the terms E_{DSO}, E_{PSO}, E_{SD}, E_{FC} and average them over rotations producing $\bar{E}_{\mathrm{DSO}}$, $\bar{E}_{\mathrm{PSO}}$, $\bar{E}_{\mathrm{SD}}$, $\bar{E}_{\mathrm{FC}}$:

- $$\begin{aligned}\bar{E}_{\mathrm{DSO}} = {} & \frac{e^2}{2mc^2} \sum_{A,B} \sum_j \gamma_A \gamma_B \boldsymbol{I}_A \cdot \boldsymbol{I}_B \left\langle \psi_0^{(0)} \middle| \frac{\boldsymbol{r}_{Aj} \cdot \boldsymbol{r}_{Bj}}{r_{Aj}^3 r_{Bj}^3} \psi_0^{(0)} \right\rangle \\ & - \frac{e^2}{2mc^2} \sum_{A,B} \sum_j \gamma_A \gamma_B \frac{1}{3} \boldsymbol{I}_A \cdot \boldsymbol{I}_B \\ & \times \left\{ \left\langle \psi_0^{(0)} \middle| \frac{x_{Aj} x_{Bj}}{r_{Aj}^3 r_{Bj}^3} \psi_0^{(0)} \right\rangle + \left\langle \psi_0^{(0)} \middle| \frac{y_{Aj} y_{Bj}}{r_{Aj}^3 r_{Bj}^3} \psi_0^{(0)} \right\rangle + \left\langle \psi_0^{(0)} \middle| \frac{z_{Aj} z_{Bj}}{r_{Aj}^3 r_{Bj}^3} \psi_0^{(0)} \right\rangle \right\},\end{aligned}$$

because the first part of the formula does not need any averaging (it is already in the appropriate form), the second part is averaged according to (W.4). Therefore,

$$\bar{E}_{\mathrm{DSO}} = \frac{e^2}{3mc^2} \sum_{A,B} \sum_j \gamma_A \gamma_B \boldsymbol{I}_A \cdot \boldsymbol{I}_B \left\langle \psi_0^{(0)} \middle| \frac{\boldsymbol{r}_{Aj} \cdot \boldsymbol{r}_{Bj}}{r_{Aj}^3 r_{Bj}^3} \psi_0^{(0)} \right\rangle.$$

- $$\begin{aligned}\bar{E}_{\mathrm{PSO}} & = \langle \psi_0^{(0)} | \hat{B}_3 \hat{R}_0 \hat{B}_3 \psi_0^{(0)} \rangle_{\mathrm{aver}} \\ & = \left(\frac{i\hbar e}{mc}\right)^2 \sum_{A,B} \sum_{j,l} \gamma_A \gamma_B \left\langle \psi_0^{(0)} \middle| \nabla_j \cdot \frac{\boldsymbol{I}_A \times \boldsymbol{r}_{Aj}}{r_{Aj}^3} \hat{R}_0 \nabla_l \cdot \frac{\boldsymbol{I}_B \times \boldsymbol{r}_{Bl}}{r_{Bl}^3} \psi_0^{(0)} \right\rangle_{\mathrm{aver}} \\ & = \left(\frac{i\hbar e}{mc}\right)^2 \sum_{A,B} \sum_{j,l} \gamma_A \gamma_B \left\langle \psi_0^{(0)} \middle| \nabla_j \cdot \frac{\boldsymbol{r}_{Aj} \times \boldsymbol{I}_A}{r_{Aj}^3} \hat{R}_0 \nabla_l \cdot \frac{\boldsymbol{r}_{Bl} \times \boldsymbol{I}_B}{r_{Bl}} \psi_0^{(0)} \right\rangle_{\mathrm{aver}} \\ & = -\left(\frac{\hbar e}{mc}\right)^2 \sum_{A,B} \sum_{j,l} \gamma_A \gamma_B \\ & \quad \times \left\langle \psi_0^{(0)} \middle| \boldsymbol{I}_A \cdot \left(\nabla_j \times \frac{\boldsymbol{r}_{Aj}}{r_{Aj}^3}\right) \hat{R}_0 \boldsymbol{I}_B \cdot \left(\nabla_l \times \frac{\boldsymbol{r}_{Bl}}{r_{Bl}}\right) \psi_0^{(0)} \right\rangle_{\mathrm{aver}},\end{aligned}$$

where the subscript "aver" means the averaging of eq. (W.4) and the identity $\boldsymbol{A} \cdot (\boldsymbol{B} \times \boldsymbol{C}) = (\boldsymbol{A} \times \boldsymbol{B}) \cdot \boldsymbol{C}$ has been used. We have the following chain of equalities (involving[4] the electronic momenta $\hat{\boldsymbol{p}}_j$ and angular momenta $\boldsymbol{L}_{Aj}$ with respect to the nucleus A, where j means electron number j)

$$\left(\frac{i\hbar e}{mc}\right)^2 \sum_{A,B}\sum_{j,l}\gamma_A\gamma_B\left\langle\psi_0^{(0)}\middle|\boldsymbol{I}_A\cdot\frac{1}{i\hbar}(\boldsymbol{r}_{Aj}\times\hat{\boldsymbol{p}}_j)\hat{R}_0\boldsymbol{I}_B\cdot\frac{1}{i\hbar}(\boldsymbol{r}_{Bl}\times\hat{\boldsymbol{p}}_l)\psi_0^{(0)}\right\rangle_{\text{aver}}$$

$$=\left(\frac{e}{mc}\right)^2 \sum_{A,B}\sum_{j,l}\gamma_A\gamma_B\langle\psi_0^{(0)}|\boldsymbol{I}_A\cdot(\boldsymbol{r}_{Aj}\times\hat{\boldsymbol{p}}_j)\hat{R}_0\boldsymbol{I}_B\cdot(\boldsymbol{r}_{Bl}\times\hat{\boldsymbol{p}}_l)\psi_0^{(0)}\rangle_{\text{aver}}$$

$$=\left(\frac{e}{mc}\right)^2 \sum_{A,B}\sum_{j,l}\gamma_A\gamma_B\langle\psi_0^{(0)}|\boldsymbol{I}_A\cdot\hat{\boldsymbol{L}}_{Aj}\hat{R}_0\boldsymbol{I}_B\cdot\hat{\boldsymbol{L}}_{Bl}\psi_0^{(0)}\rangle_{\text{aver}}$$

$$=\left(\frac{e}{mc}\right)^2 \sum_{A,B}\sum_{j,l}\gamma_A\gamma_B\boldsymbol{I}_A\cdot\boldsymbol{I}_B\frac{1}{3}\big\{\langle\psi_0^{(0)}|\hat{\boldsymbol{L}}_{Aj,x}\hat{R}_0\hat{\boldsymbol{L}}_{Bl,x}\psi_0^{(0)}\rangle$$

$$+\langle\psi_0^{(0)}|\hat{\boldsymbol{L}}_{Aj,y}\hat{R}_0\hat{\boldsymbol{L}}_{Bl,y}\psi_0^{(0)}\rangle+\langle\psi_0^{(0)}|\hat{\boldsymbol{L}}_{Aj,z}\hat{R}_0\hat{\boldsymbol{L}}_{Bl,z}\psi_0^{(0)}\rangle\big\}.$$

Thus, finally

$$\bar{E}_{\text{PSO}}=\frac{1}{3}\left(\frac{e}{mc}\right)^2\sum_{A,B}\sum_{j,l}\gamma_A\gamma_B\boldsymbol{I}_A\cdot\boldsymbol{I}_B\langle\psi_0^{(0)}|\hat{\boldsymbol{L}}_{Aj}\hat{R}_0\hat{\boldsymbol{L}}_{Bl}\psi_0^{(0)}\rangle.$$

- $\bar{E}_{\text{SD}}=\langle\psi_0^{(0)}|\hat{B}_6\hat{R}_0\hat{B}_6\psi_0^{(0)}\rangle_{\text{aver}}$

---

[4] Let us have a closer look at the operator $(\nabla_j\times\frac{\boldsymbol{r}_{Aj}}{r_{Aj}^3})$ acting on a function (it is necessary to remember that ∇_j in $\nabla_j\times\frac{\boldsymbol{r}_{Aj}}{r_{Aj}^3}$ is not just acting on the components of $\frac{\boldsymbol{r}_{Aj}}{r_{Aj}^3}$ alone, but in fact on $\frac{\boldsymbol{r}_{Aj}}{r_{Aj}^3}$ times a wave function) f: Let us see:

$$\left(\nabla_j\times\frac{\boldsymbol{r}_{Aj}}{r_{Aj}^3}\right)f=\boldsymbol{i}\left(\nabla_j\times\frac{\boldsymbol{r}_{Aj}}{r_{Aj}^3}\right)_x f+\boldsymbol{j}\left(\nabla_j\times\frac{\boldsymbol{r}_{Aj}}{r_{Aj}^3}\right)_y f+\boldsymbol{k}\left(\nabla_j\times\frac{\boldsymbol{r}_{Aj}}{r_{Aj}^3}\right)_z f$$

$$=\boldsymbol{i}\left(\frac{\partial}{\partial y_j}\frac{z_{Aj}}{r_{Aj}^3}-\frac{\partial}{\partial z_j}\frac{y_{Aj}}{r_{Aj}^3}\right)_x f+\text{similarly with } y \text{ and } z$$

$$=\boldsymbol{i}\left(-3\frac{y_{Aj}z_{Aj}}{r_{Aj}^4}+\frac{z_{Aj}}{r_{Aj}^3}\frac{\partial}{\partial y_j}+3\frac{y_{Aj}z_{Aj}}{r_{Aj}^4}-\frac{y_{Aj}}{r_{Aj}^3}\frac{\partial}{\partial z_j}\right)_x f+\text{similarly with } y \text{ and } z$$

$$=\boldsymbol{i}\left(\frac{z_{Aj}}{r_{Aj}^3}\frac{\partial}{\partial y_j}-\frac{y_{Aj}}{r_{Aj}^3}\frac{\partial}{\partial z_j}\right)_x f+\text{similarly with } y \text{ and } z$$

$$=\boldsymbol{i}\left(\frac{z_{Aj}}{r_{Aj}^3}\frac{\partial}{\partial y_j}-\frac{y_{Aj}}{r_{Aj}^3}\frac{\partial}{\partial z_j}\right)_x f+\text{similarly with } y \text{ and } z$$

$$=-\frac{1}{i\hbar}(-\boldsymbol{r}_{Aj}\times\hat{\boldsymbol{p}}_j)f=\frac{1}{i\hbar}(\boldsymbol{r}_{Aj}\times\hat{\boldsymbol{p}}_j)f.$$

$$= \gamma_{\text{el}}^2 \sum_{j,l=1}^{N} \sum_{A,B} \gamma_A \gamma_B \Big\langle \psi_0^{(0)} \Big| \Big[\frac{\hat{\boldsymbol{s}}_j \cdot \boldsymbol{I}_A}{r_{Aj}^3} - 3 \frac{(\hat{\boldsymbol{s}}_j \cdot \boldsymbol{r}_{Aj})(\boldsymbol{I}_A \cdot \boldsymbol{r}_{Aj})}{r_{Aj}^5} \Big]$$

$$\times \hat{R}_0 \Big[\frac{\hat{\boldsymbol{s}}_l \cdot \boldsymbol{I}_B}{r_{Bl}^3} - 3 \frac{(\hat{\boldsymbol{s}}_l \cdot \boldsymbol{r}_{Bl})(\boldsymbol{I}_B \cdot \boldsymbol{r}_{Bl})}{r_{Bl}^5} \Big] \psi_0^{(0)} \Big\rangle_{\text{aver}}$$

$$= \gamma_{\text{el}}^2 \sum_{j,l=1}^{N} \sum_{A,B} \gamma_A \gamma_B \boldsymbol{I}_A \cdot \boldsymbol{I}_B \frac{1}{3} \Big\{ \Big\langle \psi_0^{(0)} \Big| \Big[\frac{\hat{\boldsymbol{s}}_{j,x}}{r_{Aj}^3} - 3 \frac{(\hat{\boldsymbol{s}}_j \cdot \boldsymbol{r}_{Aj}) x_{Aj}}{r_{Aj}^5} \Big]$$

$$\times \hat{R}_0 \Big[\frac{\hat{\boldsymbol{s}}_{l,x}}{r_{Bl}^3} - 3 \frac{(\hat{\boldsymbol{s}}_l \cdot \boldsymbol{r}_{Bl})(x_{Bl})}{r_{Bl}^5} \Big] \psi_0^{(0)} \Big\rangle$$

$$+ \Big\langle \psi_0^{(0)} \Big| \Big[\frac{\hat{\boldsymbol{s}}_{j,y}}{r_{Aj}^3} - 3 \frac{(\hat{\boldsymbol{s}}_j \cdot \boldsymbol{r}_{Aj}) y_{Aj}}{r_{Aj}^5} \Big] \hat{R}_0 \Big[\frac{\hat{\boldsymbol{s}}_{l,y}}{r_{Bl}^3} - 3 \frac{(\hat{\boldsymbol{s}}_l \cdot \boldsymbol{r}_{Bl})(y_{Bl})}{r_{Bl}^5} \Big] \psi_0^{(0)} \Big\rangle$$

$$+ \Big\langle \psi_0^{(0)} \Big| \Big[\frac{\hat{\boldsymbol{s}}_{j,z}}{r_{Aj}^3} - 3 \frac{(\hat{\boldsymbol{s}}_j \cdot \boldsymbol{r}_{Aj}) z_{Aj}}{r_{Aj}^5} \Big] \hat{R}_0 \Big[\frac{\hat{\boldsymbol{s}}_{l,z}}{r_{Bl}^3} - 3 \frac{(\hat{\boldsymbol{s}}_l \cdot \boldsymbol{r}_{Bl})(z_{Bl})}{r_{Bl}^5} \Big] \psi_0^{(0)} \Big\rangle \Big\}.$$

Therefore,

$$\bar{E}_{\text{SD}} = \frac{1}{3} \gamma_{\text{el}}^2 \sum_{j,l=1}^{N} \sum_{A,B} \gamma_A \gamma_B \boldsymbol{I}_A \cdot \boldsymbol{I}_B$$

$$\times \Big\langle \psi_0^{(0)} \Big| \Big[\frac{\hat{\boldsymbol{s}}_j}{r_{Aj}^3} - 3 \frac{(\hat{\boldsymbol{s}}_j \cdot \boldsymbol{r}_{Aj}) \boldsymbol{r}_{Aj}}{r_{Aj}^5} \Big] \hat{R}_0 \Big[\frac{\hat{\boldsymbol{s}}_l}{r_{Bl}^3} - 3 \frac{(\hat{\boldsymbol{s}}_l \cdot \boldsymbol{r}_{Bl})(\boldsymbol{r}_{Bl})}{r_{Bl}^5} \Big] \psi_0^{(0)} \Big\rangle.$$

- $\bar{E}_{\text{FC}} = \langle \psi_0^{(0)} | \hat{B}_7 \hat{R}_0 \hat{B}_7 \psi_0^{(0)} \rangle$

$$= \gamma_{\text{el}}^2 \sum_{j,l=1} \sum_{A,B} \gamma_A \gamma_B \langle \psi_0^{(0)} | \delta(\boldsymbol{r}_{Aj}) \hat{\boldsymbol{s}}_j \cdot \boldsymbol{I}_A \hat{R}_0 \delta(\boldsymbol{r}_{Bl}) \hat{\boldsymbol{s}}_l \cdot \boldsymbol{I}_B \psi_0^{(0)} \rangle_{\text{aver}}$$

$$= \gamma_{\text{el}}^2 \sum_{j,l=1} \sum_{A,B} \gamma_A \gamma_B \boldsymbol{I}_A \cdot \boldsymbol{I}_B \frac{1}{3} \{ \langle \psi_0^{(0)} | \delta(\boldsymbol{r}_{Aj}) \hat{\boldsymbol{s}}_{j,x} \hat{R}_0 \delta(\boldsymbol{r}_{Bl}) \hat{\boldsymbol{s}}_{l,x} \psi_0^{(0)} \rangle$$

$$+ \langle \psi_0^{(0)} | \delta(\boldsymbol{r}_{Aj}) \hat{\boldsymbol{s}}_{j,y} \hat{R}_0 \delta(\boldsymbol{r}_{Bl}) \hat{\boldsymbol{s}}_{l,y} \psi_0^{(0)} \rangle$$

$$+ \langle \psi_0^{(0)} | \delta(\boldsymbol{r}_{Aj}) \hat{\boldsymbol{s}}_{j,z} \hat{R}_0 \delta(\boldsymbol{r}_{Bl}) \hat{\boldsymbol{s}}_{l,z} \psi_0^{(0)} \rangle \}.$$

Hence,

$$\bar{E}_{\text{FC}} = \frac{1}{3} \left(\frac{8\pi}{3} \right)^2 \gamma_{\text{el}}^2 \sum_{j,l=1} \sum_{A,B} \gamma_A \gamma_B \boldsymbol{I}_A \cdot \boldsymbol{I}_B \langle \psi_0^{(0)} | \delta(\boldsymbol{r}_{Aj}) \hat{\boldsymbol{s}}_j \hat{R}_0 \delta(\boldsymbol{r}_{Bl}) \hat{\boldsymbol{s}}_l \psi_0^{(0)} \rangle.$$

The results mean that the coupling constants J are just as reported on p. 671.

X. MULTIPOLE EXPANSION

What is the multipole expansion for?

In the perturbational theory of intermolecular interactions (Chapter 13) the perturbation operator (V) plays an important role. The operator contains all the Coulombic charge–charge interactions, where one of the point charges belongs to subsystem A, the second to B. Therefore, according to the assumption behind the perturbational approach (large intermolecular distance) there is a *guarantee* that both charges are distant in space. For example, for two interacting hydrogen atoms (electron 1 at the nucleus a, electron 2 at nucleus b, a.u. are used)

$$V = -\frac{1}{r_{a2}} + \frac{1}{r_{12}} - \frac{1}{r_{b1}} + \frac{1}{R}, \tag{X.1}$$

where R stands for the internuclear distance. A short inspection convinces us that the mean value of the operator $-\frac{1}{r_{a2}} + \frac{1}{r_{12}}$, with the wave function[1] $\psi_{A,n_1}(1)\psi_{B,n_2}(2)$, would give something close to zero, because both distances in the denominators are almost equal to each other, Fig. X.1.a. The same can be said of the two other terms of V. This is why, the situation is similar (see Chapter 13) to weighing the captain's hat, which we criticized so harshly in the supermolecular approach to supermolecular forces, see Fig. 13.4.

> What could we do to prevent a loss of accuracy? This is precisely the goal of the multipole expansion for each of the operators $\frac{1}{r_{ij}}$.

Coordinate system

What is the multipole expansion really? We will explain this in a moment. Let us begin quietly with introducing two Cartesian coordinate systems: one on molecule A, the second on molecule B (Fig. X.1.b).

This can be done in several ways. Let us begin by choosing the origins of the coordinate systems. How do we choose them? Is it irrelevant? It turns out that the choice is important. Let us stop the problem here and come back to it later on. Just as a signal, let me communicate the conclusion: the origins should be chosen in the neighbourhood of the centres of mass (charges) of the interacting molecules. Let

---

[1] $\psi_{A,n_1}(1)$ means an excited state (n_1 is the corresponding quantum number) of atom A, $\psi_{B,n_2}(2)$ similarly for atom B. Note that electron 1 is always close to nucleus a, electron 2 close to nucleus b, while A and B are far distant.

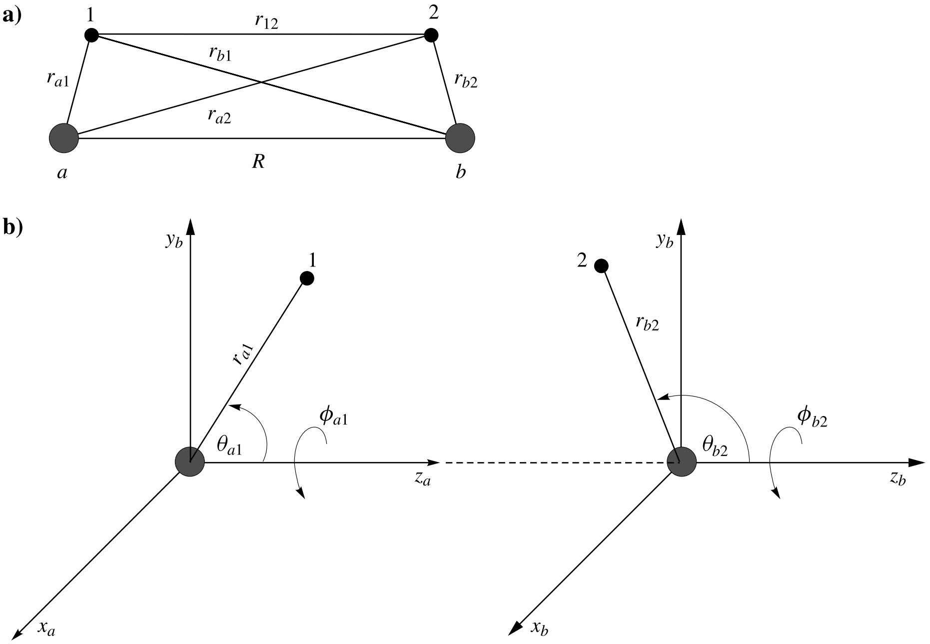


Fig. X.1. The coordinate system used in the multipole expansion. (a) Interparticle distances. The large black dots denote the origins of the two Cartesian coordinate systems, labelled a and b, respectively. We assume particle 1 always resides close to a, particle 2 always close to b. The figure gives a notation related to the distances considered. (b) Two Cartesian coordinate systems (and their polar counterparts): one associated with the centre a, the second one with centre b (the x and y axes are parallel in both systems, the z axes are collinear). Note that the two coordinate systems are not on the same footing: the z axis of a points towards b, while the coordinate system b does not point to a. Sometimes in the literature we introduce an alternative coordinate system with "equal footing" by changing $z_b \to -z_b$ (then the two coordinate systems point to each other), but this leads to different "handedness" ("right-" or "left-handed") of the systems and subsequently to complications for chiral molecules. Let us stick to the "non-equivalent choice".

us introduce the axes by taking the z axes (z_a and z_b) collinear pointing in the same direction, axes x_a and x_b as well as y_a and y_b, pairwise parallel.

The multipole series and the multipole operators of a particle

With such a coordinate system the Coulomb interaction of particles 1 and 2 (with charges q_1 and q_2) can be expanded using the following approximation[2]

$$\frac{q_1 q_2}{r_{12}} \cong \sum_{k=0}^{n_k} \sum_{l=0}^{n_l} \sum_{m=-s}^{m=+s} A_{kl|m|} R^{-(k+l+1)} \hat{M}_a^{(k,m)}(1)^* \hat{M}_b^{(l,m)}(2), \qquad \text{(X.2)}$$

---

[2] It represents an approximation because it is not valid for $R < |\boldsymbol{r}_{a1} - \boldsymbol{r}_{b2}|$, and this may happen in real systems (the electron clouds extend to infinity), also because n_k, n_l are finite instead of equal to ∞.

where the coefficient

$$A_{kl|m|} = (-1)^{l+m} \frac{(k+l)!}{(k+|m|)!(l+|m|)!} \tag{X.3}$$

whereas

MULTIPOLE MOMENT OPERATORS

$\hat{M}_a^{(k,m)}(1)$ and $\hat{M}_b^{(l,m)}(2)$ represent, respectively, the m-th components of the 2^k-pole and 2^l-pole of particle 1 in the coordinate system on a and of particle 2 in the coordinate system on b:

$$\hat{M}_a^{(k,m)}(1) = q_1 r_{a1}^k P_k^{|m|}(\cos\theta_{a1})\exp(im\phi_{a1}), \tag{X.4}$$

$$\hat{M}_b^{(l,m)}(2) = q_2 r_{b2}^l P_l^{|m|}(\cos\theta_{b2})\exp(im\phi_{b2}), \tag{X.5}$$

with r, θ, ϕ standing for the spherical coordinates of a particle (in coordinate system a or b, Fig. X.1.b), the associated Legendre polynomials $P_k^{|m|}$ with $|m| \leqslant k$ are defined as (cf. p. 176)

$$P_k^{|m|}(x) = \frac{1}{2^k k!}\left(1-x^2\right)^{|m|/2} \frac{\mathrm{d}^{k+|m|}}{\mathrm{d}x^{k+|m|}}\left(x^2-1\right)^k, \tag{X.6}$$

n_k and n_l in principle have to be equal to ∞, but in practice take finite integer values, s is the lower of the summation indices k, l.

Maybe an additional remark would be useful concerning the nomenclature: any multipole may be called a 2^k-pole (however strange this name looks), because this "multi" means the number 2^k. If we know how to make powers of two, and in addition have some contact with the world of the ancient Greeks and Romans, we will know how to compose the names of the successive multipoles: $2^0 = 1$, hence *mono*pole; $2^1 = 2$, hence *di*pole, $2^2 = 4$, hence, *quadru*pole, etc. The names, however, are of no importance. The formulae for the multipoles are important.

Multipole moment operators for many particles

A while ago a definition of the multipole moments of a single point-like charged particle was introduced. However, the multipole moments will be calculated in future, practically always for a molecule. Then,

THE TOTAL MULTIPOLE MOMENT OPERATOR

The total multipole moment operator represents the sum of the same operators for the individual particles (of course, all them have to be calculated in the same coordinate system): $\hat{M}_a^{(k,m)}(A) = \sum_{i\in A} \hat{M}_a^{(k,m)}(i)$.

The first thing we have to stress about multipole moments is that, in principle, they depend on the choice of the coordinate system (Fig. X.2).

This will soon be seen when inspecting the formulae for multipole moments.

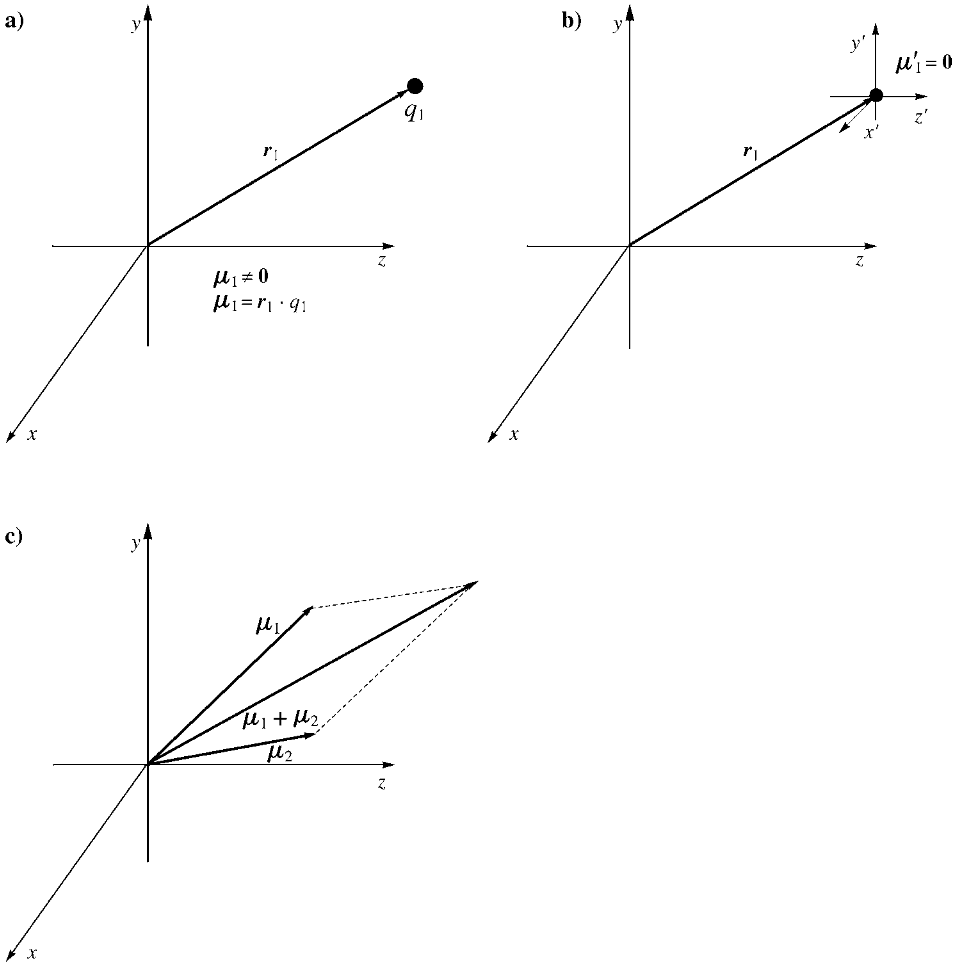


Fig. X.2. The multipole moments (or, simply multipoles) in general depend on the choice of coordinate system. (a) The dipole moment of a point-like particle with charge q_1 is equal to $\boldsymbol{\mu}_1$. (b) The dipole moment of the same particle in a coordinate system with the origin on the particle. Here we obtain $\boldsymbol{\mu}'_1 = 0$. (c) The dipole moment of two particles represents the sum of the dipole moments of the individual particles (in a common coordinate system).

Examples

Let us take a few examples for particle 1 in the coordinate system a (for the sake of simplicity we skip the indices). The case with $k = 0$ is obviously the simplest one, and we should always begin with the simplest things. If $k = 0$, then (because of $P_k^{|m|}$) $m = 0$, and the monopole therefore has a single component $M^{(00)}$

$$\hat{M}^{(0,0)} = qr^0 P_0^0(\cos\theta)\exp(\mathrm{i}0\phi) = q. \tag{X.7}$$

Table X.1. Multipole moments $\hat{M}^{(k,m)}$ divided by q

m k	0	± 1	± 2	± 3
0 charge	1	–	–	–
1 dipole	z	$x+\mathrm{i}y$ $x-\mathrm{i}y$	–	–
2 quadrupole	$\frac{1}{2}(3z^2-r^2)$	$3z(x+\mathrm{i}y)$ $3z(x-\mathrm{i}y)$	$3(x+\mathrm{i}y)^2$ $3(x-\mathrm{i}y)^2$	– –
3 octupole	$\frac{1}{2}(5z^3-3zr^2)$	$\frac{3}{2}(x+\mathrm{i}y)(5z^2-r^2)$ $\frac{3}{2}(x-\mathrm{i}y)(5z^2-r^2)$	$15z(x+\mathrm{i}y)^2$ $15z(x-\mathrm{i}y)^2$	$15(x+\mathrm{i}y)^3$ $15(x-\mathrm{i}y)^3$

Hence,

MONOPOLE
The monopole for a particle simply means its charge.

Let us go to $k=1$, i.e. to the dipole moment. Since $m=-1,0,+1$, the dipole moment has three components. First, let us consider $\hat{M}^{(1,0)}$

$$\hat{M}^{(1,0)} = qr^1 P_1^0(\cos\theta)\exp(\mathrm{i}0\phi) = qr\cos\theta = qz. \tag{X.8}$$

DIPOLE MOMENT OPERATOR
Thus the z-component of the dipole moment operator of a single particle is equal to qz. The other components are:

$$\begin{aligned} M^{(1,1)} &= qr^1 P_1^1(\cos\theta)\exp(\mathrm{i}\phi) = qr\sin\theta(\cos\phi + \mathrm{i}\sin\phi) \\ &= q(x+\mathrm{i}y), \\ M^{(1,-1)} &= qr^1 P_1^1(\cos\theta)\exp(-\mathrm{i}\phi) = qr\sin\theta(\cos\phi - \mathrm{i}\sin\phi) \\ &= q(x-\mathrm{i}y). \end{aligned}$$

After a careful (but a little boring) derivation, we arrive at Table X.1 (up to the octupole). Just to make the table simpler, every multipole moment of the particle has been divided by q.

Thus the operator of the 2^k-pole moment of a charged particle simply represents a k-th degree polynomial of x, y, z.

The multipoles depend on the coordinate system chosen

Evidently any multipole moment value (except the monopole) depends on my imagination because I am free to choose any coordinate system I want and, e.g., the z coordinate of the particle in such a system will also depend on me! It turns out that if we calculate the multipole moments, then

> the lowest non-vanishing multipole moment does not depend on the coordinate system translation, the other moments in general do depend on it.

This is not peculiar for the moments defined by eqs. (X.4) or (X.5), but represents a property of every term of the form $x^n y^l z^m$. Indeed, $k = n + l + m$ tells us that we have to do with a 2^k-pole. Let us shift the origin of the coordinate system by the vector $\boldsymbol{L}$. Then the $x^n y^l z^m$ moment calculated in the new coordinate system, i.e. $x'^n y'^l z'^m$ is equal to

$$\begin{aligned}(x')^n (y')^l (z')^m &= (x + L_x)^n (y + L_y)^l (z + L_z)^m \\ &= x^n y^l z^m + \text{a linear combination of } \textit{lower multipole moments}. \qquad \text{(X.9)}\end{aligned}$$

If, for some reason, all the lower moments are equal to zero, this would mean the invariance of the moment of choice of the coordinate system.

Let us take, e.g., the system ZnCl^+. In the first approximation, the system may be approximated by *two point-like charges* Zn^{++} and Cl^-. Let us locate these charges on the z axis in such a way that Zn^{++} has the coordinate $z = 0$, and Cl^- $z = 5$. Now we would like to calculate the z component of the dipole moment:[3] $M^{(1,0)} = \mu_z = q_1 z_1 + q_2 z_2 = (+2)0 + (-1)5 = -5$. What if we had chosen another coordinate system? Let us check what would happen if the origin of the coordinate system were shifted towards the positive z by 10 units. In such a case the ions have the coordinates $z_1' = -10$, and $z_2' = -5$, and, as the z component of the dipole moment we obtain

$$M^{(1,0)'} = \mu_z' = q_1 z_1' + q_2 z_2' = (+2)(-10) + (-1)(-5) = -15. \qquad \text{(X.10)}$$

Thus, the dipole moment depends on the choice of the coordinate system. However, the monopole of the system is equal to $(+2) + (-1) = +1$ and this number will not change with any shift of the coordinate system. Therefore,

> the dipole moment of a molecular ion depends on us, through arbitrary choice of the coordinate system.

Interaction energy of non-point like multipoles

In our chemical understanding of intermolecular interactions, multipole–multipole (mainly dipole–dipole, as for interactions in, e.g., water) interactions play an important role. The dipolar molecules have non-zero dimensions and therefore they

---

[3] Since we have to do with point charges, the calculation of the multipole moments reduces simply to inserting the values of the coordinates of the corresponding charges into the multipole operator.

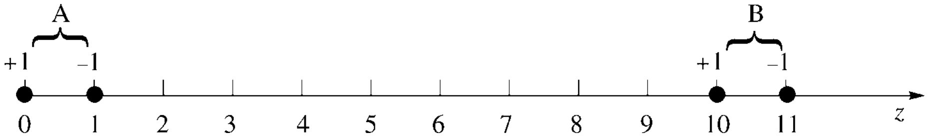


Fig. X.3. The interaction of *non-pointlike* dipoles also contains interactions of higher multipoles.

represent something other than point-like dipoles. Let us clarify this by taking the simple example of two dipolar systems located on the z axis (Fig. X.3): the system A consists of the two charges $+1$ at $z = 0$ and -1 at $z = 1$, while system B also has two charges $+1$ with $z = 10$ and -1 with $z = 11$.

The first idea is that we have to do with the interaction of two dipoles and that's all there is to it. Let us check whether everything is OK. The checking is very easy, because what really interacts are the charges, no dipoles whatsoever. Thus the exact interaction of systems A and B is $(+1)(+1)/10 + (+1)(-1)/11 + (-1)(+1)/9 + (-1)(-1)/10 = 2/10 - 1/11 - 1/9 = -0.0020202$. What would give such a dipole–dipole interaction? Such a task immediately poses the question of how such an interaction is to be calculated.

The first advantage of the multipole expansion is that it produces the formulae for the multipole–multipole interactions.

We have the dipole–dipole term in the form $R^{-3}(\mu_{ax}\mu_{bx} + \mu_{ay}\mu_{by} - 2\mu_{az}\mu_{bz}) = -2R^{-3}\mu_{az}\mu_{bz}$, because the x and y components of our dipole moments are equal zero. Since A and B are neutral, it is absolutely irrelevant which coordinate system is to be chosen to calculate the dipole moment components. Therefore let us use the global coordinate system, in which the positions of the charges have been specified. Thus, $\mu_{az} = (+1)\cdot 0 + (-1)\cdot 1 = -1$ and $\mu_{bz} = (+1)\cdot 10 + (-1)\cdot 11 = -1$.

What is R?

Now, we are encountering a serious problem (which we *always* encounter in the multipole expansion), what is R? We are forced to choose the two local coordinate systems in A and B. We *arbitrarily* decide here to locate these origins in the middle of each dipolar system, and therefore $R = 10$. It looks like a reasonable choice, and as will be shown later on, it really is. We are all set to calculate the dipole–dipole interaction: $-2\cdot 10^{-3}(-1)(-1) = -0.0020000$. Close! The exact calculated interaction energy is -0.0020202. Where is the rest? Is there any error in our dipole–dipole interaction formula? We simply forgot that our dipolar systems represent not only the dipole moments, but also have non-zero octupole moments (the quadrupoles are equal zero) and non-zero higher odd-order multipoles, and we did not take them into account. If somebody calculated all the interactions of such multipoles, we would recover the correct interaction energy with any desired accuracy. How come, however, that such a simple dipolar system also has a non-zero octupole moment? The answer is simple: it is because the dipole is not

Table X.2. Are the multipole moments zero or non-zero?

	Li^+	HCl	H_2	CH_4	HCl^+
monopole $k=0$	q	0	0	0	q
dipole $k=1$	0	μ	0	0	μ
quadrupole $k=2$	0	Q	Q	0	Q
octupole $k=3$	0	Oct	0	Oct	Oct

point-like.[4] The conclusion from this story is that the reader has to pay attention to whether we have to deal with point-like or non-point-like multipole moments.

Just as a little exercise, Table X.2 shows which multipole moments are zero and which are non-zero for a few simple chemical systems. All this follows from the symmetry of their nuclear framework in the electronic ground state.

Properties of the multipole expansion

When performing multipole expansions, at least three simple questions arise:

a) How do we truncate the expansion, i.e. how do we choose the values of n_k and n_l in eq. (X.2)?
b) Since the multipole moments depend, in general, on the coordinate system chosen, what sort of miracle makes the multipole expansion *of the energy*, independent of the coordinate system?
c) When does the multipole expansion make sense, i.e. when does it converge?

Truncating the multipole expansion and its coordinate system dependence

It turns out that questions a and b are closely related to each other. When n_k and n_l are finite and non-zero,[5] then, however horrifying it might be, the result of the multipole expansion is in general coordinate-dependent. If, however, n_k and n_l satisfy $n_k + n_l = \text{const}$, we may shift both coordinate systems (the same translation for both) however we like, and the interaction energy calculated remains invariant.[6] Such a recipe for n_k and n_l corresponds to taking all the terms with a given power of R^{-1}.

> In other words, if we take all the terms with a given R^{-m} dependence, the result does not depend on the same translations of both coordinate systems.

---

[4] Just think about a multipole component of the form qz^n calculated with respect to the centre of each subsystem.

[5] Zero would introduce large errors in most applications.

[6] L.Z. Stolarczyk, L. Piela, *Int. J. Quantum Chem.* 15 (1979) 701.

This means that to maintain the invariance of the energy with respect to equal translations of both coordinate systems, we have to calculate all terms satisfying $n_k + n_l = n_{\max}$ in the multipole expansion. If, e.g., $n_{\max} = 2$, we have to calculate the term proportional to R^{-1} or the charge–charge interaction (it will be invariant), proportional to R^{-2} or charge–dipole and dipole–charge terms (their sum is also invariant), proportional to R^{-3} or charge–quadrupole, quadrupole–charge and dipole–dipole (their sum is invariant as well).

Let us imagine scientists calculating the interaction energy of two molecules. As will be shown later, in their multipole expansion they will have the charges of both interacting *molecules*, their dipole moments, their quadrupole moments, etc. Our scientists are systematic fellows, and therefore I bet they will begin by calculating the multipole moments for each molecule, up to a certain maximum multipole moment (say, the quadrupole; the calculations become more and more involved, which makes their decision easier). Then they will be ready to calculate all the individual multipole–multipole interaction contributions. They will make a table of such interactions (rows: the multipole moments of A; columns: the multipole moments of B) and calculate all the entries in their table. Then many of their colleagues would sum *all* the entries of the table in order not to waste their time. This will be a mistake. The scientists might not suspect that, due to this procedure, their result depends on the choice of coordinate system, which is always embarrassing. However, our scientists will do something else. They will sum the entries corresponding to: charge–charge, charge–dipole, dipole–charge, charge–quadrupole, quadrupole–charge, dipole–dipole and they will throw the other entries into the waste paper basket. Having made this decision, the scientists will gain a lot: their interaction energy will not depend on how they translated the a and b coordinate systems.

Now, we will illustrate this by a simple formulae and see how it works in practice. We have said before that it is decisive to take the complete set of terms with the given dependence on R^{-1}. Otherwise horrible things happen. Let us take such a complete set of terms with $k + l = 2$. We will see how nicely they behave upon the translation of the coordinate system, and how nasty the behaviour of individual terms is. Let us begin with the charge–dipole term. The term in the multipole expansion corresponds to $k = 0$ and $l = 2$:

$$(-1)^2\frac{2!}{2!R^3}\hat{M}^{(00)}(1)^*\hat{M}^{(20)}(2) = q_1q_2R^{-3}\frac{1}{2}\left(3z_2^2 - r_2^2\right).$$

The next term $(k = 1, l = 1)$ has three contributions coming from the summation over m:

$$(-1)\frac{2!}{1!1!R^3}\hat{M}^{(10)}(1)^*\hat{M}^{(10)}(2) + (-1)^2\frac{2!}{2!2!R^3}\hat{M}^{(11)}(1)^*\hat{M}^{(11)}(2)$$
$$+ (-1)^0\frac{2!}{2!2!R^3}\hat{M}^{(1-1)}(1)^*\hat{M}^{(1-1)}(2) = q_1q_2R^{-3}\left[(x_1x_2 + y_1y_2) - 2z_1z_2\right].$$

The third term $(k = 2, l = 0)$:

$$(-1)^2\frac{2!}{2!R^3}\hat{M}^{(20)}(1)^*\hat{M}^{(00)}(2)=q_1q_2R^{-3}\frac{1}{2}\left(3z_1^2-r_1^2\right).$$

Note that each of the calculated terms depends separately on the translation along the z axis of the origins of the interacting objects. Indeed, by taking $z+T$ instead of z we obtain: for the first term

$$\begin{aligned}&q_1q_2R^{-3}\left[\frac{1}{2}\left(3(z_2+T)^2-x_2^2-y_2^2-(z_2+T)^2\right)\right]\\&\quad=q_1q_2R^{-3}\left[\frac{1}{2}\left(3z_2^2-r_2^2\right)+\frac{1}{2}\left(6Tz_2+3T^2-2Tz_2-T^2\right)\right],\end{aligned}$$

for the second term

$$\begin{aligned}&q_1q_2R^{-3}\left[(x_1x_2+y_1y_2)-2(z_1+T)(z_2+T)\right]\\&\quad=q_1q_2R^{-3}\left[(x_1x_2+y_1y_2)-2z_1z_2\right]+R^{-3}\left[-2Tz_1-2Tz_2-2T^2\right],\end{aligned}$$

for the third term

$$\begin{aligned}&q_1q_2R^{-3}\frac{1}{2}\left(3(z_1+T)^2-x_1^2-y_1^2-(z_1+T)^2\right)\\&\quad=q_1q_2R^{-3}\left[\frac{1}{2}\left(3z_1^2-r_1^2\right)+\frac{1}{2}\left(6Tz_1+3T^2-2Tz_1-T^2\right)\right].\end{aligned}$$

If someone still has the illusion that the coordinate system dependence is negligible, this is about the right time to change their opinion. Evidently, each term depends on what we chose as T, and T can be *anything*! If I were really malicious, I would obtain a monstrous dependence on T.

Now, let us add all the individual terms together to form the complete set for $k+l=2$:

$$\begin{aligned}&q_1q_2\left\{R^{-3}\left[\frac{1}{2}\left(3z_2-r_2^2\right)+\left(2Tz_2+T^2\right)\right]+R^{-3}\left[(x_1x_2+y_1y_2)-2z_1z_2\right]\right.\\&\qquad\left.+R^{-3}\left[-2Tz_1-2Tz_2-2T^2\right]+R^{-3}\left[\frac{1}{2}\left(3z_1-r_1^2\right)+\left(2Tz_1+T^2\right)\right]\right\}\\&\quad=q_1q_2R^{-3}\left\{\frac{1}{2}\left(3z_2-r_2^2\right)+\left[(x_1x_2+y_1y_2)-2z_1z_2\right]+\frac{1}{2}\left(3z_1-r_1^2\right)\right\}.\end{aligned}$$

The dependence on T has disappeared as if touched by a magic wand.[7] *The complete set does not depend on* T! This is what I wanted to show.

---

[7] We may also prove that equal but arbitrary rotations of both coordinate systems about the z axis also lead to a similar invariance of interaction energy.

Convergence of the multipole expansion

I owe the reader an explanation about the convergence of the multipole expansion (point c, Fig. X.4). Well,

we may demonstrate that the multipole expansion convergence depends on how the molecules are located in space with respect to one another. The convergence criterion reads

$$|\boldsymbol{r}_{b2} - \boldsymbol{r}_{a1}| < R, \tag{X.11}$$

where $\boldsymbol{r}_{a1}$ denotes the vector pointing the particle 1 from its coordinate system origin, similarly for vector $\boldsymbol{r}_{b2}$.

The readers will easily be convinced if they draw two spheres that are tangent to each other (this is the most dangerous situation) and then consider possible $\boldsymbol{r}_{a1}$ and $\boldsymbol{r}_{b2}$ vectors. Whatever the $\boldsymbol{r}_{a1}$ and $\boldsymbol{r}_{b2}$ vectors are, our criterion will be fulfilled. The criterion is, however, even more general than to allow two non-overlapping

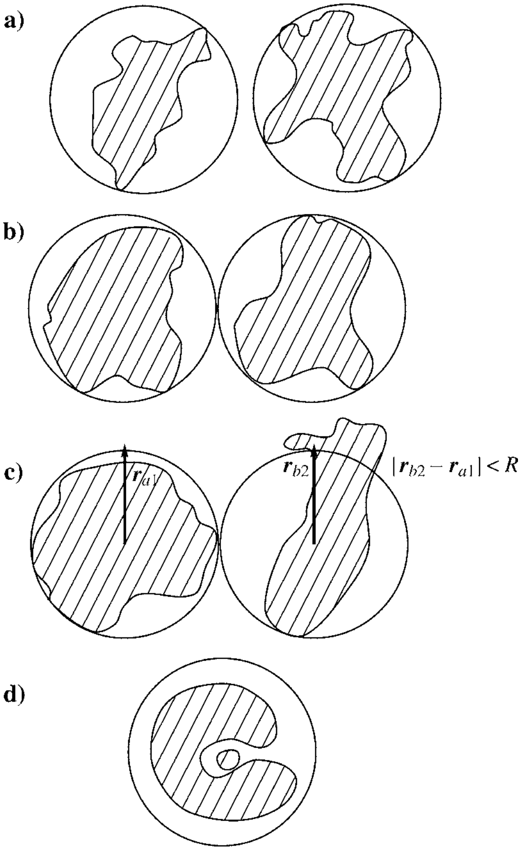


Fig. X.4. Convergence of the multipole expansion. The expansion converges in situations (a–c), diverges in (d).

spheres. It is easy to find locations of the two particles that are outside the spheres, and yet the convergence criterion is fulfilled. For example, let us take two tangent spheres with radii ρ_1 and ρ_2 (their centres are on the x axis) as well as vectors $\boldsymbol{r}_{a1} = (0, \rho_1, 0)$ and $\boldsymbol{r}_{b2} = (0, u, 0)$, where $u = \rho_1 + R/10$ and $u > \rho_2$. Then, $|\boldsymbol{r}_{b2} - \boldsymbol{r}_{a1}| = R/10 < R$, i.e. the convergence criterion is satisfied, despite the fact that particle 2 is outside its sphere.

For our purposes it is sufficient to remember that

> when the two particles are in their non-overlapping spheres, the multipole expansion converges.

Can we make such an assumption? Our goal is the application of the multipole expansion in the case of intermolecular interactions. Are we able to enclose both molecules in two non-overlapping spheres? Sometimes certainly not, e.g., if a small molecule A is to be docked in the cavity of a large molecule B. This is a very interesting case (Fig. X.4.d), but what we have most often in quantum chemistry are two *distant* molecules. Is everything all right then? Apparently the molecules can be enclosed in the spheres, but if we recall that the electronic density extends to infinity (although it decays very fast), we feel a little scared. *Almost* the whole density distribution could be enclosed in such spheres, but outside the spheres there is also something. It turns out that this very fact causes

> the multipole expansion for the interaction energy of such diffused charge distributions to diverge, i.e. if we go to very high terms we will get infinity.

However strange it might look, in mathematics we are also able to extract very useful information from divergent series, if they converge *asymptotically*, see p. 210. This is precisely the situation when multipole expansion is applied to the diffuse charge distributions that such molecules have. This is why the multipole expansion is useful.[8] It also has the important advantage of being physically appealing, because thanks to it we may interpret interaction energy in terms of the properties of the individual interacting molecules (their charges, dipole, quadrupole, etc. moments).

---

[8] If the calculations were feasible to a high degree of accuracy, the multipole expansion might be of small importance.

Y. PAULI DEFORMATION

Two molecules, when non-interacting are independent and the wave function of the total system might be taken as a *product* of the wave functions for the individual molecules. When the same two molecules are interacting, any product-like function represents only an approximation, sometimes a very poor approximation,[1] because according to a postulate of quantum mechanics, the wave function has to be antisymmetric with respect to the exchange of electronic labels, while the product does not fulfil this. More exactly, the approximate wave function has to belong to the irreducible representation of the symmetry group of the Hamiltonian (see Appendix C, p. 903), to which the ground state wave function belongs. This means first of all that the Pauli exclusion principle is to be satisfied.

PAULI DEFORMATION
The product-like wave function has to be made antisymmetric. This causes some changes in the electronic charge distribution (electronic density), which will be called the Pauli deformation.

The Pauli deformation may be viewed as a mechanical distortion of both interacting molecules due to mutual pushing. The reason why two rubber balls deform when pushed against each other is the same: the electrons of one ball cannot occupy the same space as the electrons (with the same spin coordinates) of the second ball. The most dramatic deformation takes place close to the contact area of these balls.

The norm of the difference of $\varphi^{(0)}$ and $\psi^{(0)}$ represents a very stringent measure of the difference between two functions: any deviation gives a contribution to the measure. We would like to know, how the electronic density has changed, where the electrons flow from, and where they go to. The electron density ρ (a function of position in space) is defined as the sum of densities ρ_i of the particular electrons:

$$\rho(x,y,z)=\sum_{i=1}^{N}\rho_i(x,y,z),$$

$$\rho_i(x_i,y_i,z_i)=\sum_{\sigma_i=-\frac{1}{2}}^{+\frac{1}{2}}\int\frac{\mathrm{d}\tau}{\mathrm{d}\tau_i}|\psi|^2, \tag{Y.1}$$

---

[1] For example, when the intermolecular distance is short, the molecules push each other and deform (maybe strongly), and the product-like function is certainly inadequate.

where $d\tau = d\tau_1 d\tau_2 \cdots d\tau_N$, and therefore the integration goes over the coordinates (space and spin) of all the electrons except electron i. In addition, there is also a summation over the spin coordinate of electron "i", simply because we are not interested in its value. As seen, the integral of $\rho(x, y, z)$ over x, y, z is equal to N, therefore $\rho(x, y, z)$ represents an electron cloud carrying N electrons, as defined in eq. (11.1) on p. 569. We make the two molecules approach without changing their charge distribution (the system is described by the electron density corresponding to the wave function $\psi = \varphi^{(0)}$), and then we allow the Pauli exclusion principle to operate to ensure the proper symmetry of the wave function (the system is therefore described by a new wave function $\psi = \psi^{(0)}$) by applying a suitable projection operator. What happens to the electronic density? Will it change or not?

Let us see what happens when we make two hydrogen atoms approach and then two helium atoms.

H_2 case

In the case of two hydrogen atoms[2]

$$\varphi^{(0)} = 1s_a(1)\alpha(1)1s_b(2)\beta(2) \equiv a(1)\alpha(1)b(2)\beta(2),$$

where we have used the abbreviation $1s_a(1) \equiv a$ and $1s_b(1) \equiv b$. After inserting $\psi = \varphi^{(0)}$ into (Y.1), integration over space and summation over spin coordinates gives

$$\rho^{(0)} = \rho_1(x, y, z) + \rho_2(x, y, z),$$

where

$$\begin{aligned}\rho_1(x, y, z) &= \sum_{\sigma_1=-\frac{1}{2}}^{+\frac{1}{2}} \int \frac{d\tau}{d\tau_1} \left|a(1)\alpha(1)b(2)\beta(2)\right|^2 \\ &= \sum_{\sigma_1=-\frac{1}{2}}^{+\frac{1}{2}} \int d\tau_2 \left|a(1)\alpha(1)b(2)\beta(2)\right|^2 = a^2.\end{aligned}$$

Similarly,

$$\begin{aligned}\rho_2(x, y, z) &= \sum_{\sigma_2=-\frac{1}{2}}^{+\frac{1}{2}} \int \frac{d\tau}{d\tau_2} \left|a(1)\alpha(1)b(2)\beta(2)\right|^2 \\ &= \sum_{\sigma_2=-\frac{1}{2}}^{+\frac{1}{2}} \int d\tau_1 \left|a(1)\alpha(1)b(2)\beta(2)\right|^2 = b^2.\end{aligned}$$

---

[2] We arbitrarily assign the spin function α to electron 1 and the spin function β to electron 2. We might have done this in the opposite way, but it does not change anything.

Thus finally $\rho^{(0)} = a^2 + b^2$. This density is normalized to 2 – as it has to be, because the electron cloud $\rho(x, y, z)$ carries two electrons. Now, let us do the same for the wave function $\psi^{(0)} = N\hat{\mathcal{A}}\varphi^{(0)}$, where $\hat{\mathcal{A}}$ stands for the idempotent projection operator (13.23), and the normalization constant $N = \frac{2}{\sqrt{1+S^2}}$ with $S = (a|b)$, all quantities described in Chapter 13 on the symmetry adapted perturbation theory:

$$\begin{aligned}
\rho(x, y, z) &= \rho_1(x, y, z) + \rho_2(x, y, z),\\
\rho_1(x, y, z) &= \sum_{\sigma_1=\pm\frac{1}{2}} \int \mathrm{d}\tau_2 \, |\psi^{(0)}|^2\\
&= N^2\frac{1}{8}\int \mathrm{d}V_2 \big[a(1)b(2) + a(2)b(1)\big]^2 \sum_{\sigma_1}\sum_{\sigma_2}\frac{1}{2}\big[\alpha(1)\beta(2) - \alpha(2)\beta(1)\big]^2\\
&= N^2\frac{1}{8}\int \mathrm{d}V_2 \big[a(1)b(2) + a(2)b(1)\big]^2\\
&= N^2\frac{1}{8}\big(a^2 + b^2 + 2abS\big)\\
&= \frac{1}{2(1+S^2)}\big(a^2 + b^2 + 2abS\big),\\
\rho_2(x, y, z) &= \rho_1(x, y, z).
\end{aligned}$$

As seen, the density $\rho_1(x, y, z)$ is normalized to 1 – this is what we get after integration over dV_1. A similar calculation for ρ_2 would give the same result, because $|\psi^{(0)}|^2$ is symmetric with respect to the exchange of electrons[3] 1 and 2. Therefore, the *change* in the electron density due to the proper symmetry projection (including the Pauli exclusion principle) is:

$$\rho - \rho^{(0)} = \frac{a^2 + b^2 + 2abS}{1+S^2} - \big(a^2 + b^2\big) = \frac{2S}{1+S^2}ab - \frac{S^2}{1+S^2}a^2 - \frac{S^2}{1+S^2}b^2. \quad \text{(Y.2)}$$

Thus, it turns out that as a result of the Pauli exclusion principle (i.e. of the antisymmetrization of the wave function) an electron density $a^2S^2/(1+S^2)$ flows from atom a, a similar thing happens to atom b, where the electronic density decreases by $b^2S^2/(1+S^2)$. Both these electronic clouds go to the bond region – we find them as an electron cloud $2abS/(1+S^2)$ with a maximum in the middle of the bond, and of course, the integral of $\rho - \rho^{(0)}$ is equal to zero (Fig. Y.1.a).

Thus,

> in the hydrogen molecule the Pauli exclusion principle caused the two atoms to stick together (the two electrons increase their probability to be in the region between the two nuclei).

---

[3] This was not the case for $\varphi^{(0)}$.

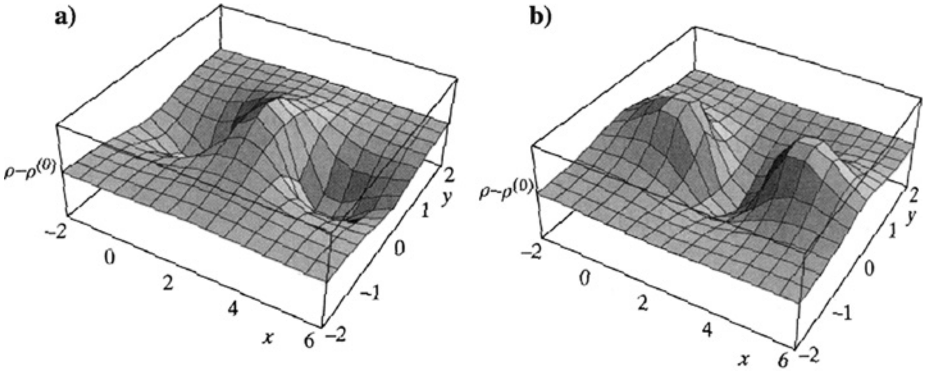


Fig. Y.1. Comparison of the Pauli deformation for two hydrogen atoms and for two helium atoms. (a) Two hydrogen atoms. Visualization of $\rho - \rho^{(0)}$ calculated in the plane containing the nuclei ("the net result is zero"). One of the protons is located at the origin, the other has coordinates $(0, R, 0)$, with $R = 4$ a.u. For this distance the overlap integral (see Appendix R, p. 1009) $S = (1 + R + \frac{R^2}{3})\exp(-R)$ is 0.189. As we can see, the electron density has flown from the nuclei to the bond. (b) Two helium atoms. The only difference with respect to (a) is that two electrons have been added. The visualization of $\rho - \rho^{(0)}$ reveals a *completely different pattern.* This time the electron density has been removed from the bond region and increased in the region of the nuclei.

This is what the Pauli exclusion principle dictates. Besides this we have, of course, all the physical interactions (electron repulsion, attraction with the nuclei) and the kinetic energy, but none of these effects has yet been taken into account.[4] Fig. Y.1(a) shows only the deformation that results from forcing the proper symmetry in the wave function.

He_2 case

Let us see what happens if we make similar calculations for two helium atoms. To compare the future result with the H_2 case, let us keep everything the same (the internuclear distance R, the atomic orbitals, the overlap integral S, etc.), except that the number of electrons changes from two to four. This time the calculation will be a little bit more tedious, because four-electron wave functions are more complicated than two-electron functions. For example, the function $\varphi^{(0)}$ this time is the product of the two Slater determinants – one for atom a, the other for atom b:

---

[4] Indeed, all these effects have been ignored, because we neither calculated the energy, nor used the Hamiltonian. However, the very fact that we write: $\varphi^{(0)} = a(1)\alpha(1)b(2)\beta(2)$, where a and b stand for the properly centred 1s orbitals, means that the electron–nucleus interaction has been *implicitly* taken into account (this is why the 1s orbital appears). Similarly, when we project the product-like function and obtain $\psi^{(0)}$ proportional to $[a(1)b(2) + a(2)b(1)][\alpha(1)\beta(2) - \alpha(2)\beta(1)]$, then besides the above mentioned electron–nucleus interactions (manifested by the 1s orbitals) we obtain an interesting effect: when one electron is on nucleus a, the second electron runs to nucleus b. It looks as if they have repelled each other. This is, however, at the level of the *mathematical formula of the function* ("function design"), as if the function has already been quite well designed for the future, and takes into account the physical interactions.

$$\begin{aligned}\varphi^{(0)} &= N\begin{vmatrix} a\alpha(1) & a\alpha(2) \\ a\beta(1) & a\beta(2)\end{vmatrix}\begin{vmatrix} b\alpha(3) & b\alpha(4) \\ b\beta(3) & b\beta(4)\end{vmatrix} \\ &= N'a(1)a(2)b(3)b(4)\left[\frac{1}{\sqrt{2}}\big[\alpha(1)\beta(2)-\alpha(2)\beta(1)\big]\right] \\ &\quad\times\left[\frac{1}{\sqrt{2}}\big[\alpha(3)\beta(4)-\alpha(4)\beta(3)\big]\right], \end{aligned} \tag{Y.3}$$

where the normalization constant $N'=1$ (easy to verify: just square the function and integrate). We obtain directly from the definition[5]

$$\rho^{(0)}=\rho_1+\rho_2+\rho_3+\rho_4=2a^2+2b^2,$$

which, after integration, gives four electrons, as should be. The function $\varphi^{(0)}$ is “illegal”, because it does not fulfil the Pauli exclusion principle, e.g., the exchange of electrons 1 and 3 does not lead to a change of the sign of the wave function.

Now let us focus on $\psi^{(0)}$. Please note that $\varphi^{(0)}$ of eq. (Y.3) may be written alternatively as:

$$\varphi^{(0)}=N\begin{vmatrix} a\alpha(1) & a\alpha(2) & 0 & 0 \\ a\beta(1) & a\beta(2) & 0 & 0 \\ 0 & 0 & b\alpha(3) & b\alpha(4) \\ 0 & 0 & b\beta(3) & b\beta(4)\end{vmatrix},$$

where N is a normalization constant.

Antisymmetrization of $\varphi^{(0)}$, in which electrons 1 and 2 occupy orbital a, and electrons 3 and 4 occupy orbital b, is equivalent to completing the Slater determinant[6] in such a way as to allow for the exchange of electrons between the subsystems:

$$\begin{aligned}\psi^{(0)} &= N\frac{1}{2}(1+I)\hat{A}\varphi^{(0)} = N\frac{1}{2}(1+I)\begin{vmatrix} a\alpha(1) & a\alpha(2) & a\alpha(3) & a\alpha(4) \\ a\beta(1) & a\beta(2) & a\beta(3) & a\beta(4) \\ b\alpha(1) & b\alpha(2) & b\alpha(3) & b\alpha(4) \\ b\beta(1) & b\beta(2) & b\beta(3) & b\beta(4)\end{vmatrix} \\ &= N\begin{vmatrix} a\alpha(1) & a\alpha(2) & a\alpha(3) & a\alpha(4) \\ a\beta(1) & a\beta(2) & a\beta(3) & a\beta(4) \\ b\alpha(1) & b\alpha(2) & b\alpha(3) & b\alpha(4) \\ b\beta(1) & b\beta(2) & b\beta(3) & b\beta(4)\end{vmatrix},\end{aligned}$$

where, according to (13.23), $\hat{A}$ stands for the idempotent antisymmetrization operator, and $\frac{1}{2}(1+I)$ represents an idempotent symmetrization operator acting on the nuclear coordinates. The last equality follows from the fact that this particular

---

[5] This may also be calculated in your head (note that the spin functions in the square brackets are normalized).

[6] The Slater determinant containing linearly independent spinorbitals *guarantees* the antisymmetry.

Slater determinant is already symmetric with respect to the exchange of nuclei,[7] which is equivalent to $a \leftrightarrow b$.

Any determinant is invariant with respect to the addition of any linear combination of rows (columns) to a given row (column). For reasons that will become clear soon, let us make a series of such operations. First, let us add the third row to the first one, then multiply the third row by 2 (any multiplication is harmless for the determinant, because at the end it will be normalized) and subtract the first row from the third one. Then let us perform a similar series of operations on rows 2 and 4 (instead of 1 and 3), and at the end let us multiply rows 1 and 3 by $\frac{1}{\sqrt{2(1+S)}}$, and rows 2 and 4 by $\frac{1}{\sqrt{2(1-S)}}$. The result of these operations is the Slater determinant with the doubly occupied bonding molecular orbital $\sigma = \frac{1}{\sqrt{2(1+S)}}(a+b)$ and the doubly occupied antibonding molecular orbital $\sigma^* = \frac{1}{\sqrt{2(1-S)}}(a-b)$

$$\psi^{(0)} = \frac{1}{\sqrt{4!}} \begin{vmatrix} \sigma\alpha(1) & \sigma\alpha(2) & \sigma\alpha(3) & \sigma\alpha(4) \\ \sigma\beta(1) & \sigma\beta(2) & \sigma\beta(3) & \sigma\beta(4) \\ \sigma^*\alpha(1) & \sigma^*\alpha(2) & \sigma^*\alpha(3) & \sigma^*\alpha(4) \\ \sigma^*\beta(1) & \sigma^*\beta(2) & \sigma^*\beta(3) & \sigma^*\beta(4) \end{vmatrix}.$$

All the spinorbitals involved are orthonormal (in contrast to what was in the original determinant) and the corresponding electronic density is easy to write – it is the sum of squares of the molecular orbitals multiplied by their occupancies (cf. p. 1015):

$$\rho(x,y,z) = 2\sigma^2 + 2(\sigma^*)^2.$$

Now let us calculate the Pauli deformation

$$\begin{aligned} \rho - \rho^{(0)} &= \frac{a^2+b^2+2ab}{1+S} + \frac{a^2+b^2-2ab}{1-S} - 2\big(a^2+b^2\big) \\ &= -\frac{4S}{1-S^2}ab + \frac{2S^2}{1-S^2}a^2 + \frac{2S^2}{1-S^2}b^2. \end{aligned} \tag{Y.4}$$

Integration of the difference gives zero, as should be. Note that the formula is similar to that which we obtained for the hydrogen molecule, but this time the electron flow is completely different (Fig. Y.1.b).

In the case of He_2 the Pauli exclusion principle makes the electron density decrease in the region between the nuclei and increase close to the nuclei. In the case of the hydrogen molecule, the two atoms stuck together, while the two helium atoms deform as if they were rubber balls squeezed together (Pauli deformation).

---

[7] This corresponds to the exchange of rows in the determinant: the first with the third, and the second with the fourth. A single exchange changes the sign of the determinant, therefore the two exchanges leave the determinant invariant.

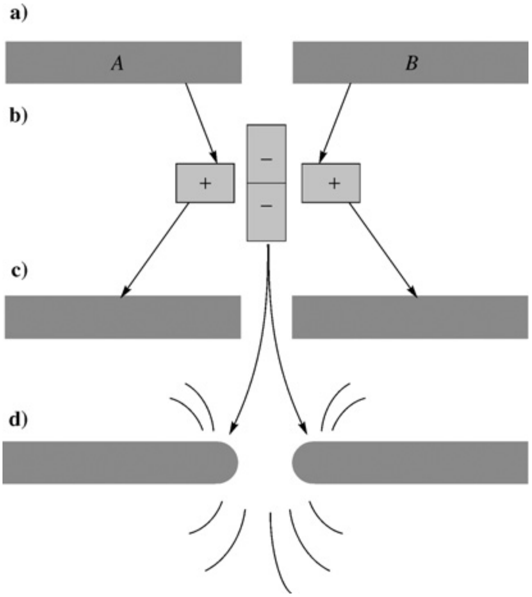


Fig. Y.2. The locality of the Pauli deformation (diagram). (a) Two polymeric chains A and B (with electronic densities in the form of the elongated rectangles corresponding to the isolated molecules A and B) approach one another (b) the Pauli deformation consists of the two density gains (the rectangles with +) and a single electron loss (the rectangles with −). Let us assume that the surfaces of the rectangles are equal to the corresponding integrals of the charge distributions $-4S/(1-S^2)ab$ in the contact region, $2S^2/(1-S^2)a^2$ on molecule A and $2S^2/(1-S^2)b^2$ on polymer B – this is why the electron density loss has a rectangle twice as large as any of the electron density gains (c) a partial Pauli deformation: the density gain $2S^2/(1-S^2)a^2$ for molecule A has been added to the initial density distribution, and similarly for molecule B (the rectangles became larger, but *locally the corresponding increase is small*). (d) In order to represent the total Pauli deformation from the result obtained at point c we subtracted the density distribution $4S/(1-S^2)ab$ which is located in the contact region. As a result *the Pauli deformation, when viewed locally, is large only in the contact region*.

The *only* thing that has been changed with respect to the hydrogen molecule is the increase in the number of electrons from two to four (we have kept the orbital exponents equal to 1 and the internuclear distance equal to 4 a.u. unchanged). This change results in a qualitative difference in the Pauli deformation.

Two large molecules

For two helium atoms, the Pauli deformation means decreasing the electron density in the region between the nuclei and a corresponding increase in the density on the nuclei. This looks dangerous! What if, instead of two helium atoms, we have two closed-shell long molecules A and B that touch each other with their terminal parts? *Would the Pauli deformation be local, or would it extend over the whole system?* Maybe the distant parts of the molecules would deform as much as the contact regions?

The answer may be deduced from eq. (Y.4). The formula suggests that the electronic density change pertains to the whole system. When the formula was derived, we concentrated on two helium atoms. However, *nothing* would change in the derivation if we had in mind a doubly occupied molecular orbital a that extends over the whole polymer A and a similar orbital b that extends over B. In such a case the formula (Y.4) would be identical. The formula says: the three deformation contributions cancel if we integrate them over the total space.[8] The first deformation means a density deficiency (minus sign), the other two mean density gains (plus sign). The first of these contributions is *certainly located close to the contact region* of A and B. The two others (of the same magnitude) have a spatial form such that a^2 and b^2 (i.e. extend over the whole polymer chains A and B), but are scaled by the factor $2S^2/(1-S^2)$. Since the contributions cancel in space (when integrated), this means that *the density gain extends over the polymeric molecules and, therefore, locally is very small; the larger the interacting molecules the smaller the local change*. The situation is therefore similar to an inflatable balloon pressed with your finger. We have a large deformation at the contact region , what corresponds to $-\frac{4S}{1-S^2}ab$, but in fact the *whole* balloon deforms. Because this deformation has to extend over the whole balloon, the local deformation on the other side of the toy is extremely small. Therefore, common sense has arrived at a quantum mechanical explanation.[9]

> This means that the Pauli deformation has a *local character*: it takes place almost exclusively in the region of contact between both molecules.

Two final remarks

- The Pauli deformation, treated as a spatial charge density distribution has a region with positive charge (some electron density flowed from there) and negative charge (where the electron density has increased). The Pauli charge distribution participates in the Coulombic interactions within the system. If such an interaction is represented by a multipole–multipole interaction, the Pauli deformation has no monopole, or charge. In general, the other multipole moments of the Pauli deformation are non-zero. In particular, the Pauli deformation multipoles resulting from the exchange interaction of molecules A and B may interact with the electric multipoles of molecule C, thus contributing to the three-body effect.
- If the two systems A and B approach each other in such a way that $S=0$, the Pauli deformation is zero. $S=0$ might occur, e.g., if the two molecules approach along the nodal surfaces of the frontier molecular orbitals.

---

[8] But of course at a given point they do not cancel in general.

[9] Good for both of them.

Z. ACCEPTOR–DONOR STRUCTURE CONTRIBUTIONS IN THE MO CONFIGURATION

In Chapter 14 the Slater determinants were constructed in three different ways using:

- molecular orbitals (MO picture),
- acceptor and donor orbitals (AD picture),
- atomic orbitals (VB picture).

Then, the problem appeared of how to express one picture by another, in particular this was of importance for expressing the MO picture as an AD. More specifically, we are interested in calculating the contribution of an acceptor–donor structure[1] in the Slater determinant written in the MO formalism, where the molecular orbitals are expressed by the donor (n) and acceptor (χ and χ^*) orbitals in the following way

$$\begin{aligned}\varphi_1 &= a_1 n + b_1\chi - c_1\chi^*,\\ \varphi_2 &= a_2 n - b_2\chi - c_2\chi^*,\\ \varphi_3 &= -a_3 n + b_3\chi - c_3\chi^*.\end{aligned} \tag{Z.1}$$

We assume that $\{\varphi_i\}$ form an orthonormal set. For simplicity, it is also assumed that in the first approximation the orbitals $\{n, \chi, \chi^*\}$ are also orthonormal. Then we may write that a Slater determinant in the MO picture (denoted by X_i) represents a linear combination of the Slater determinants (Y_j) containing exclusively donor and acceptor orbitals:

$$X_i = \sum_j c_i(Y_j)Y_j,$$

where the coefficient $c_i(Y_k) = \langle Y_k|X_i\rangle$ at the Slater determinant Y_k is the contribution of the acceptor–donor structure Y_k in X_i.

In Chapter 14 three particular cases are highlighted, and they will be derived below. We will use the antisymmetrizer

$$\hat{A} = \frac{1}{N!}\sum_P (-1)^p\hat{P}$$

introduced in Chapter 10 ($\hat{P}$ is the permutation operator, and p is its parity).

---

[1] That is, of a Slater determinant built of acceptor and donor orbitals.

Case $c_0(DA)$

The $c_0(DA)$ coefficient means the contribution of the structure $n^2\chi^2$, i.e.

$$\Psi(DA) = (4!)^{-\frac{1}{2}} \det[n\bar{n}\chi\bar{\chi}] = (4!)^{\frac{1}{2}} \hat{A}[n\bar{n}\chi\bar{\chi}]$$

in the ground-state Slater determinant

$$\Psi_0 = (4!)^{-\frac{1}{2}} \det[\varphi_1\bar{\varphi}_1\varphi_2\bar{\varphi}_2] = (4!)^{\frac{1}{2}} \hat{A}[\varphi_1\bar{\varphi}_1\varphi_2\bar{\varphi}_2].$$

We have to calculate

$$\begin{aligned}
c_0(DA) &= \langle Y_k|X_i\rangle = \langle\Psi(DA)|\Psi_0\rangle \\
&= 4!\langle\hat{A}[n\bar{n}\chi\bar{\chi}]|\hat{A}[\varphi_1\bar{\varphi}_1\varphi_2\bar{\varphi}_2]\rangle \\
&= 4!\langle[n\bar{n}\chi\bar{\chi}]|\hat{A}^2[\varphi_1\bar{\varphi}_1\varphi_2\bar{\varphi}_2]\rangle \\
&= 4!\langle[n\bar{n}\chi\bar{\chi}]|\hat{A}[\varphi_1\bar{\varphi}_1\varphi_2\bar{\varphi}_2]\rangle \\
&= 4!\langle\big[n(1)\bar{n}(2)\chi(3)\bar{\chi}(4)\big]|\hat{A}\big[\varphi_1(1)\bar{\varphi}_1(2)\varphi_2(3)\bar{\varphi}_2(4)\big]\rangle,
\end{aligned}$$

where we have used $\hat{A}$ as Hermitian and idempotent. Next, we have to write all the 24 permutations $[\varphi_1(1)\bar{\varphi}_1(2)\varphi_2(3)\bar{\varphi}_2(4)]$ (taking into account their parity) and then perform integration over the coordinates of all the four electrons (together with summation over the spin variables):

$$\begin{aligned}
c_0(DA) = \int &\mathrm{d}\tau_1\,\mathrm{d}\tau_2\,\mathrm{d}\tau_3\,\mathrm{d}\tau_4\,\big[n(1)\bar{n}(2)\chi(3)\bar{\chi}(4)\big]^* \\
&\times \sum_P (-1)^p P\big[\varphi_1(1)\bar{\varphi}_1(2)\varphi_2(3)\bar{\varphi}_2(4)\big].
\end{aligned}$$

The integral to survive *has* to have perfect matching of the spin functions between $[n(1)\bar{n}(2)\chi(3)\bar{\chi}(4)]$ and $\hat{P}[\varphi_1(1)\bar{\varphi}_1(2)\varphi_2(3)\bar{\varphi}_2(4)]$. This makes 20 of these permutations vanish. Only four integrals will survive:

$$\begin{aligned}
&c_0(DA) \\
&= \int \mathrm{d}\tau_1\,\mathrm{d}\tau_2\,\mathrm{d}\tau_3\,\mathrm{d}\tau_4\,\big[n(1)\bar{n}(2)\chi(3)\bar{\chi}(4)\big]^*\big[\varphi_1(1)\bar{\varphi}_1(2)\varphi_2(3)\bar{\varphi}_2(4)\big] \\
&\quad - \int \mathrm{d}\tau_1\,\mathrm{d}\tau_2\,\mathrm{d}\tau_3\,\mathrm{d}\tau_4\,\big[n(1)\bar{n}(2)\chi(3)\bar{\chi}(4)\big]^*\big[\varphi_1(1)\bar{\varphi}_1(4)\varphi_2(3)\bar{\varphi}_2(2)\big] \\
&\quad - \int \mathrm{d}\tau_1\,\mathrm{d}\tau_2\,\mathrm{d}\tau_3\,\mathrm{d}\tau_4\,\big[n(1)\bar{n}(2)\chi(3)\bar{\chi}(4)\big]^*\big[\varphi_1(3)\bar{\varphi}_1(2)\varphi_2(1)\bar{\varphi}_2(4)\big] \\
&\quad + \int \mathrm{d}\tau_1\,\mathrm{d}\tau_2\,\mathrm{d}\tau_3\,\mathrm{d}\tau_4\,\big[n(1)\bar{n}(2)\chi(3)\bar{\chi}(4)\big]^*\big[\varphi_1(3)\bar{\varphi}_1(4)\varphi_2(1)\bar{\varphi}_2(2)\big]
\end{aligned}$$

$$
\begin{aligned}
&= \int \mathrm{d}\tau_1\, n(1)^*\varphi_1(1) \int \mathrm{d}\tau_2\, \bar{n}(2)^*\bar{\varphi}_1(2) \int \mathrm{d}\tau_3\, \chi(3)^*\varphi_2(3) \int \mathrm{d}\tau_4\, \bar{\chi}(4)^*\bar{\varphi}_2(4)\\
&\quad - \int \mathrm{d}\tau_1\, n(1)^*\varphi_1(1) \int \mathrm{d}\tau_2\, \bar{n}(2)^*\bar{\varphi}_2(2) \int \mathrm{d}\tau_3\, \chi(3)^*\varphi_2(3) \int \mathrm{d}\tau_4\, \bar{\chi}(4)^*\bar{\varphi}_1(4)\\
&\quad - \int \mathrm{d}\tau_1\, n(1)^*\varphi_2(1) \int \mathrm{d}\tau_2\, \bar{n}(2)^*\bar{\varphi}_1(2) \int \mathrm{d}\tau_3\, \chi(3)^*\varphi_1(3) \int \mathrm{d}\tau_4\, \bar{\chi}(4)^*\bar{\varphi}_2(4)\\
&\quad + \int \mathrm{d}\tau_1\, n(1)^*\varphi_2(1) \int \mathrm{d}\tau_2\, \bar{n}(2)^*\bar{\varphi}_2(2) \int \mathrm{d}\tau_3\, \chi(3)^*\varphi_1(3) \int \mathrm{d}\tau_4\, \bar{\chi}(4)^*\bar{\varphi}_1(4)\\
&= (a_1)^2(-b_2)^2 - a_1a_2(-b_2)b_1 - a_2a_1b_1(-b_2) + (a_2)^2(b_1)^2\\
&= (a_1)^2(b_2)^2 + a_1a_2b_2b_1 + a_2a_1b_1b_2 + (a_2)^2(b_1)^2\\
&= a_1b_2(a_1b_2 + a_2b_1) + a_2b_1(a_1b_2 + a_2b_1)\\
&= (a_1b_2 + a_2b_1)^2 = \begin{vmatrix} a_1 & a_2 \\ b_1 & -b_2 \end{vmatrix}^2 .
\end{aligned}
$$

Hence,

$$c_0(DA) = \begin{vmatrix} a_1 & a_2 \\ b_1 & -b_2 \end{vmatrix}^2$$

which agrees with the formula on p. 805.

Case $c_2(DA)$

The $c_2(DA)$ represents the contribution of the structure $\Psi(DA) = (4!)^{\frac{1}{2}}\hat{A}[n\bar{n}\chi\bar{\chi}]$ in the Slater determinant corresponding to the double excitation $\Psi_{2d} = (4!)^{\frac{1}{2}}\hat{A}[\varphi_1\bar{\varphi}_1\varphi_3\bar{\varphi}_3]$. We are interested in the integral

$$
\begin{aligned}
c_2(DA) &= \langle \Psi(DA) | \Psi_{2d} \rangle\\
&= 4! \big\langle \big[n(1)\bar{n}(2)\chi(3)\bar{\chi}(4)\big] \big| \hat{A}\big[\varphi_1(1)\bar{\varphi}_1(2)\varphi_3(3)\bar{\varphi}_3(4)\big]\big\rangle .
\end{aligned}
$$

This case is very similar to the previous one, the only difference is the substitution $\varphi_2 \to \varphi_3$. Therefore, everything goes the same way as before, but this time we obtain:

$$
\begin{aligned}
&c_2(DA)\\
&= \int \mathrm{d}\tau_1\, n(1)^*\varphi_1(1) \int \mathrm{d}\tau_2\, \bar{n}(2)^*\bar{\varphi}_1(2) \int \mathrm{d}\tau_3\, \chi(3)^*\varphi_3(3) \int \mathrm{d}\tau_4\, \bar{\chi}(4)^*\bar{\varphi}_3(4)\\
&\quad - \int \mathrm{d}\tau_1\, n(1)^*\varphi_1(1) \int \mathrm{d}\tau_2\, \bar{n}(2)^*\bar{\varphi}_3(2) \int \mathrm{d}\tau_3\, \chi(3)^*\varphi_3(3) \int \mathrm{d}\tau_4\, \bar{\chi}(4)^*\bar{\varphi}_1(4)\\
&\quad - \int \mathrm{d}\tau_1\, n(1)^*\varphi_3(1) \int \mathrm{d}\tau_2\, \bar{n}(2)^*\bar{\varphi}_1(2) \int \mathrm{d}\tau_3\, \chi(3)^*\varphi_1(3) \int \mathrm{d}\tau_4\, \bar{\chi}(4)^*\bar{\varphi}_3(4)
\end{aligned}
$$

$$+\int d\tau_1\, n(1)^*\varphi_3(1)\int d\tau_2\, \bar{n}(2)^*\bar{\varphi}_3(2)\int d\tau_3\, \chi(3)^*\varphi_1(3)\int d\tau_4\, \bar{\chi}(4)^*\bar{\varphi}_1(4),$$

or

$$\begin{aligned} c_2(DA) &= (a_1)^2(b_3)^2 - a_1(-a_3)b_3b_1 - (-a_3)a_1b_1b_3 + (-a_3)^2(b_1)^2 \\ &= (a_1)^2(b_3)^2 + a_1a_3b_3b_1 + a_3a_1b_1b_3 + (a_3)^2(b_1)^2 = (a_1b_3 + a_3b_1)^2 \\ &= \begin{vmatrix} a_1 & b_1 \\ -a_3 & b_3 \end{vmatrix}^2. \end{aligned}$$

We have

$$c_2(DA) = \begin{vmatrix} a_1 & b_1 \\ -a_3 & b_3 \end{vmatrix}^2$$

which also agrees with the result used on p. 806.

Case $c_3(DA)$

This time we have to calculate the contribution of $\Psi(DA) = (4!)^{\frac{1}{2}}\hat{A}[n\bar{n}\chi\bar{\chi}]$ in the Slater determinant $\Psi_{3d} = (4!)^{\frac{1}{2}}\hat{A}[\varphi_2\bar{\varphi}_2\varphi_3\bar{\varphi}_3]$, therefore

$$\begin{aligned} c_2(DA) &= \langle\Psi(DA)|\Psi_{3d}\rangle \\ &= 4!\langle[n(1)\bar{n}(2)\chi(3)\bar{\chi}(4)]|\hat{A}[\varphi_2(1)\bar{\varphi}_2(2)\varphi_3(3)\bar{\varphi}_3(4)]\rangle. \end{aligned}$$

This is a similar case to the previous one, but we have to exchange $\varphi_1 \rightarrow \varphi_2$. We obtain:

$$\begin{aligned} &c_3(DA) \\ &= \int d\tau_1\, n(1)^*\varphi_2(1)\int d\tau_2\, \bar{n}(2)^*\bar{\varphi}_2(2)\int d\tau_3\, \chi(3)^*\varphi_3(3)\int d\tau_4\, \bar{\chi}(4)^*\bar{\varphi}_3(4) \\ &\quad - \int d\tau_1\, n(1)^*\varphi_2(1)\int d\tau_2\, \bar{n}(2)^*\bar{\varphi}_3(2)\int d\tau_3\, \chi(3)^*\varphi_3(3)\int d\tau_4\, \bar{\chi}(4)^*\bar{\varphi}_2(4) \\ &\quad - \int d\tau_1\, n(1)^*\varphi_3(1)\int d\tau_2\, \bar{n}(2)^*\bar{\varphi}_2(2)\int d\tau_3\, \chi(3)^*\varphi_2(3)\int d\tau_4\, \bar{\chi}(4)^*\bar{\varphi}_3(4) \\ &\quad + \int d\tau_1\, n(1)^*\varphi_3(1)\int d\tau_2\, \bar{n}(2)^*\bar{\varphi}_3(2)\int d\tau_3\, \chi(3)^*\varphi_2(3)\int d\tau_4\, \bar{\chi}(4)^*\bar{\varphi}_2(4), \end{aligned}$$

or

$$\begin{aligned} c_3(DA) &= (a_2)^2(b_3)^2 - a_2(-a_3)b_3(-b_2) - (-a_3)a_2(-b_2)b_3 + (-a_3)^2(-b_2)^2 \\ &= (a_2)^2(b_3)^2 - a_2a_3b_3b_2 - a_3a_2b_2b_3 + (a_3)^2(b_2)^2 \\ &= a_2b_3[a_2b_3 - a_3b_2] - a_3b_2[a_2b_3 - a_3b_2] \\ &= (a_2b_3 - a_3b_2)^2 = \begin{vmatrix} a_2 & -b_2 \\ -a_3 & b_3 \end{vmatrix}^2. \end{aligned}$$

Finally,

$$c_3(DA) = \begin{vmatrix} a_2 & -b_2 \\ -a_3 & b_3 \end{vmatrix}^2$$

and again agreement with the formula on p. 806 is obtained.

Table Z.1. Units of physical quantities

Quantity	Unit	Symbol	Value
light velocity		c	$299792.458\ \frac{\text{km}}{\text{s}}$
Planck constant		h	$6.6260755 \cdot 10^{-34}\ \text{J} \cdot \text{s}$
mass	electron rest mass	m_0	$9.1093897 \cdot 10^{-31}\ \text{kg}$
charge	element. charge = a.u. of charge	e	$1.60217733 \cdot 10^{-19}\ \text{C}$
action	$\frac{h}{2\pi}$	$\hbar$	$1.05457266 \cdot 10^{-34}\ \text{J} \cdot \text{s}$
length	bohr = a.u. of length	a_0	$5.29177249 \cdot 10^{-11}\ \text{m}$
energy	hartree = a.u. of energy	E_h	$4.3597482 \cdot 10^{-18}\ \text{J}$
time	a.u. of time	$\frac{\hbar}{E_h}$	$2.418884 \cdot 10^{-17}\ \text{s}$
velocity	a.u. of velocity	$\frac{a_0 E_h}{\hbar}$	$2.187691 \cdot 10^{6}\ \frac{\text{m}}{\text{s}}$
momentum	a.u. of momentum	$\frac{\hbar}{a_0}$	$1.992853 \cdot 10^{-24}\ \frac{\text{kg}\,\text{m}}{\text{s}}$
electr. dipole moment	a.u. of electr. dipole	ea_0	$8.478358 \cdot 10^{-30}\ \text{C} \cdot \text{m}$ (2.5415 D)
magn. dipole	Bohr magneton	$\frac{e\hbar}{2m_0c}$	$0.92731 \cdot 10^{-20}\ \frac{\text{erg}}{\text{gauss}}$
polarizability		$\frac{e^2 a_0^2}{E_h}$	$1.648778 \cdot 10^{-41}\ \frac{\text{C}^2\text{m}^2}{\text{J}}$
electric field		$\frac{E_h}{ea_0}$	$5.142208 \cdot 10^{11}\ \frac{\text{V}}{\text{m}}$
Boltzm. constant		k_B	$1.380658 \cdot 10^{-23}\ \frac{\text{J}}{\text{K}}$
Avogadro constant		N_A	$6.0221367 \cdot 10^{23}\ \text{mol}^{-1}$

Table Z.2. Conversion coefficients

	a.u.	erg	eV	$\frac{\text{kcal}}{\text{mole}}$	1 cm^{-1}	1 Hz	1 K
1 a.u.	1	$4.35916 \cdot 10^{-11}$	27.2097	627.709	$2.194746 \cdot 10^{5}$	$6.579695 \cdot 10^{15}$	$3.15780 \cdot 10^{5}$
1 erg	$2.29402 \cdot 10^{10}$	1	$6.24197 \cdot 10^{11}$	$1.43998 \cdot 10^{13}$	$5.03480 \cdot 10^{15}$	$1.50940 \cdot 10^{26}$	$7.2441 \cdot 10^{15}$
1 eV	$3.67516 \cdot 10^{-2}$	$1.60206 \cdot 10^{-12}$	1	23.0693	$8.06604 \cdot 10^{3}$	$2.41814 \cdot 10^{14}$	$1.16054 \cdot 10^{4}$
1 $\frac{\text{kcal}}{\text{mol}}$	$1.59310 \cdot 10^{-3}$	$6.9446 \cdot 10^{-14}$	$4.33477 \cdot 10^{-2}$	1	$3.49644 \cdot 10^{2}$	$1.048209 \cdot 10^{13}$	$5.0307 \cdot 10^{2}$
1 cm^{-1}	$4.556336 \cdot 10^{-6}$	$1.98618 \cdot 10^{-16}$	$1.23977 \cdot 10^{-4}$	$2.86005 \cdot 10^{-3}$	1	$2.997930 \cdot 10^{10}$	1.43880
1 Hz	$1.519827 \cdot 10^{-16}$	$6.62517 \cdot 10^{-27}$	$4.13541 \cdot 10^{-15}$	$9.54009 \cdot 10^{-14}$	$3.335635 \cdot 10^{-11}$	1	$4.7993 \cdot 10^{-11}$
1 K	$3.16676 \cdot 10^{-6}$	$1.38044 \cdot 10^{-16}$	$8.6167 \cdot 10^{-5}$	$1.98780 \cdot 10^{-3}$	0.69502	$2.08363 \cdot 10^{10}$	1

This page intentionally left blank

NAME INDEX

SUBJECT INDEX