Progress in Mathematical Physics
63

Bertrand Duplantier
Editor

Time

Poincaré Seminar 2010

Birkhäuser

Birkhäuser

Progress in Mathematical Physics
Volume 63

For further volumes:
http://www.birkhauser-science.com/series/4813

Bertrand Duplantier
Editor

Time

Poincaré Seminar 2010

Birkhäuser

Editor
Bertrand Duplantier
CEA Saclay
Service de Physique Théorique
Gif-sur-Yvette Cedex
France

ISBN 978-3-0348-0358-8 ISBN 978-3-0348-0359-5 (eBook)
DOI 10.1007/978-3-0348-0359-5
Springer Basel Heidelberg New York Dordrecht London

Library of Congress Control Number: 2013932223

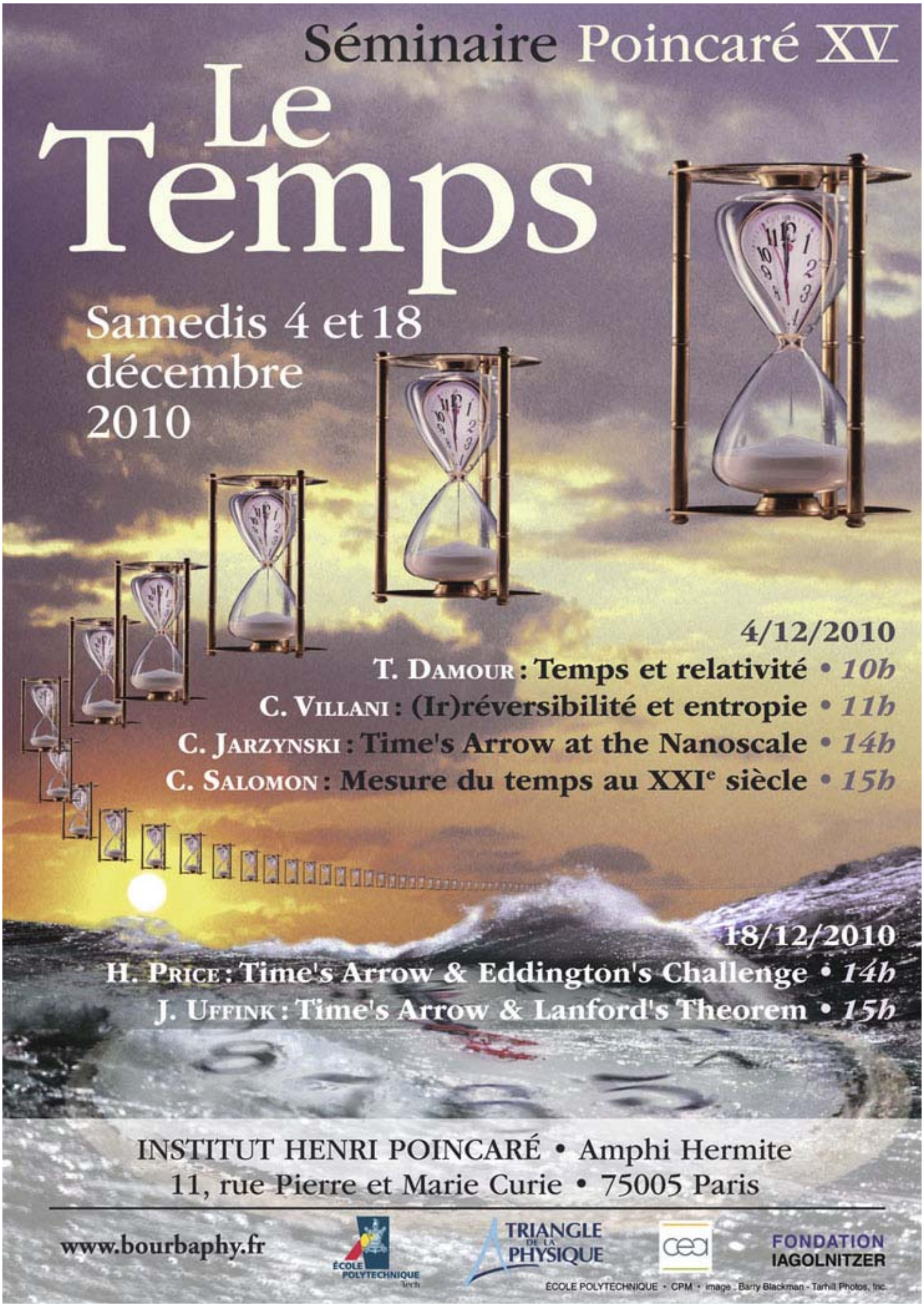
Mathematics Subject Classification (2000): 37A35, 37A60, 37N20, 81V80, 82-02, 82C05, 82C35, 82C40, 83F05

Cover design: deblik, Berlin

Printed on acid-free paper

Springer Basel is part of Springer Science+Business Media (www.birkhauser-science.com)

Contents


Séminaire Poincaré XV
Le Temps
Samedis 4 et 18 décembre 2010
4/12/2010
T. DAMOUR : Temps et relativité • 10h
C. VILLANI : (Ir)réversibilité et entropie • 11h
C. JARZYNSKI : Time's Arrow at the Nanoscale • 14h
C. SALOMON : Mesure du temps au XXI^e siècle • 15h
18/12/2010
H. PRICE : Time's Arrow & Eddington's Challenge • 14h
J. UFFINK : Time's Arrow & Lanford's Theorem • 15h
INSTITUT HENRI POINCARÉ • Amphi Hermite
11, rue Pierre et Marie Curie • 75005 Paris
www.bourbaphy.fr
ÉCOLE POLYTECHNIQUE
TRIANGLE DE LA PHYSIQUE
cea
FONDATION IAGOLNITZER
ÉCOLE POLYTECHNIQUE • CPM • image : Barry Blackman - Tarhill Photos, Inc.

Foreword

This book is the eleventh in a series of Proceedings for the *Séminaire Poincaré*, which is directed towards a broad audience of physicists, mathematicians, and philosophers of science.

The goal of this Seminar is to provide up-to-date information about general topics of great interest in physics. Both the theoretical and experimental aspects of the topic are covered, generally with some historical background. Inspired by the *Nicolas Bourbaki Seminar* in mathematics, hence nicknamed *"Bourbaphy"*, the Poincaré Seminar is held twice a year at the Institut Henri Poincaré in Paris, with written contributions prepared in advance. Particular care is devoted to the pedagogical nature of the presentations, so as to fulfill the goal of being accessible to a large audience of scientists.

This new volume of the Poincaré Seminar Series, *"Time"* (*"Temps"*), corresponds to the fifteenth such seminar, held on December 4 and 18, 2010. It presents an interdisciplinary view of the concept of time, which poses some of the most challenging questions in science, and to the human mind in its quest for an understanding of the universe. This volume describes recent developments related to the mathematical, physical, experimental, and philosophical facets of this fascinating concept. Its title could actually be **'Time's arrow'**, a phrase which seems to have been first coined by Sir Arthur Eddington in *The Nature of the Physical World* (1928), in a **challenge to physics** recalled by H. Price in his contribution below:

> "*Time's Arrow.* The great thing about time is that it goes on. But this is an aspect of it which the physicist sometimes seems inclined to neglect. In the four-dimensional world ... the events past and future lie spread out before us as in a map. The events are there in their proper spatial and temporal relation; but there is no indication that they undergo what has been described as "the formality of taking place" and the question of their doing or undoing does not arise. We see in the map the path from past to future or from future to past; but there is no signboard to indicate that it is a one-way street. Something must be added to the geometrical conceptions comprised in Minkowski's world before it becomes a complete picture of the world as we know it."

The first survey, by Thibault Damour, titled *"Time and Relativity"*, offers a broad description of the manifold fundamental physical issues at play with time, thereby serving, in effect, as an introductory chapter to the book. It recalls the evolution of Boltzmann's ideas about the physical origin of the Second Law, and the possibility that the "flow of time" is an emergent phenomenon, locally induced by the entropy time-gradient. The changes of perspective implied by Special and General Relativity are then described, before focusing on the deep question of whether relativistic gravity (including black holes), primordial cosmology, and last, but not least, quantum mechanics, are at the root of the strong time-asymmetry of our Universe.

The second article, *"(Ir)reversibility and Entropy"*, by the 2010 Fields medalist CÉDRIC VILLANI, is a masterpiece whose original French version has also been retained here for the vigor of its style, in addition to its English translation. It addresses in exquisite detail, within classical mechanics, the foundational issues associated with Time's arrow, entropy, (pre- and post-collisional) chaos, and approach to equilibrium, as seen through the lenses of Boltzmann's and Vlasov's kinetic equations. Noted work, partly done in collaboration with L. DESVILLETTES and C. MOUHOT, which led to the Fields award, is reported here. For Boltzmann's equation, it addresses in particular the Cercignani conjecture, the instability of the hydrodynamic approximation, and the fine evolution of entropy. In the case of Vlasov's equation, whose evolution, in contrast to that of Boltzmann's, is *isentropic*, the essentials of the existence proof of non-linear Landau damping, and its deep physical meaning, are described. A final section, titled "Paradoxes lost", summarizes the subtle present-day explanations to many outstanding historical issues concerning (ir)reversibility. It is a must for any reader interested in the foundations of a (classical) statistical mechanics perspective on Nature.

In the third contribution, *"Equalities and Inequalities: Irreversibility and the Second Law of Thermodynamics at the Nanoscale"*, CHRISTOPHER JARZYNSKI offers a beautifully concise but complete description of recent fundamental advances in the "thermodynamics of small systems", at the scale precisely relevant to biological physics. These recently derived theoretical predictions, which pertain to fluctuations of work and entropy production in systems far from thermal equilibrium, go by the general name of *"fluctuation theorems"*, the most famous ones being the *non-equilibrium work relation*, a.k.a. *Jarzynski's equality*, the *Crooks time-symmetry relation*, the *transient* fluctuation theorem of Evans and Searles, and the *steady-state* fluctuation theorem of Gallavotti and Cohen. These results allow one to rewrite thermodynamic inequalities as *equalities*, and reveal that nonequilibrium fluctuations encode equilibrium information, with practical applications in computational thermodynamics and in the analysis of single-molecule manipulation experiments. They also have far-reaching scientific and philosophical consequences: the ability of thermodynamics to set the direction of Time's arrow can now be **quantified**.

In *"Time's Measurement in the XXIst Century"*, CHRISTOPHE SALOMON describes its amazing improvement in precision over the last 4 centuries, modern atomic clocks having gained 13 orders of magnitude with respect to the Huygens pendulum. He explains the principles of atomic and fountain clocks; the latest *optical* ones with 10^{-17} accuracy use the frequency division due to 2005 Nobel laureates T.W. Haensch and J.L. Hall, and allow for a precision of 1 picosecond per day, i.e., 5 seconds over the age of the Universe! At NIST in 2010, general relativity effects could thus be detected at a distance of 33 centimeters! The author then describes the design of the new cold atom space clock PHARAO, in which ultra-slow Cesium atoms produced by laser cooling will yield an atomic resonance 5 to 10 times narrower than in a fountain and about 10^4 times narrower than in a commercial Cesium clock. This clock will be a core element of the European space mission ACES (Atomic Clock Ensemble in Space), to be installed in 2015 onboard the International Space Station. The comparison of this microgravity high-stability time scale in space to those on the ground will allow a test of Einstein's gravitational shift with a record precision of 2×10^{-6}. Repeated frequency comparisons between space and ground clocks that operate with different atoms will enable a test of the stability over time of the fundamental constants of physics, an issue first raised by Paul A.M. Dirac in 1937. The ultra-stable clocks of the ACES mission will also allow for a 'relativistic geodesy' of the Earth.

The following contribution, written by HUW PRICE, a leading philosopher of science, aims to clarify the difficult and subtle logical issues arising from the puzzle of the time-asymmetry of our universe, as reflected in particular in thermodynamics. In *"Time's Arrow and Eddington's Challenge"*, he expounds the latter challenge, taken from *The Nature of the Physical World*:

> "But is he [i.e., the physicist] ready to forgo that knowledge of the going on of time ... , and content himself with the time inferred from sense-impressions which is emaciated of all dynamic quality? No doubt some will reply that they are content; to these I would say – Then *show your good faith by reversing the dynamic quality of time* (which you may freely do if it has no importance in Nature), and, just for a change, give us a picture of the universe passing from the more random to the less random state ... If you are an astronomer, tell how waves of light hurry in from the depths of space and condense on to stars; how the complex solar system unwinds itself into the evenness of a nebula. ... If you genuinely believe that a contra-evolutionary theory is just as true and as significant as an evolutionary theory, *surely it is time that a protest should be made against the entirely one-sided version currently taught.*" (H.P.'s emphasis.)

The author's aim is to provide a logical, disambiguating but subtle philosophical guide to the puzzle of the time-asymmetry of thermodynamic phenomena and the time-symmetry of the underlying microphysics. While many approaches to the thermodynamic asymmetry look for a dynamical explanation of the Second Law – a dynamical cause or factor, responsible for entropy increase (like in Boltzmann's *H*-theorem, based on the assumption of molecular chaos), he insists on the distinct puzzle of the low entropy past boundary condition (to be completed by Boltzmann's appeal to high entropy statistical macrostates). The puzzle of the thermodynamic arrow thus becomes a puzzle for cosmology in the past, as already in Eddington's view:

> "We are thus driven to admit anti-chance; and apparently the best thing we can do with it is to sweep it up into a heap at the beginning of time."

Price then recommends a kind of healthy scepticism about the universality of the Second Law of thermodynamics, quoting Burbury (1904) who denied any "right to supplement it by a large draft of the scientific imagination", in contrast to Eddington's view that it holds "the supreme position among the laws of Nature"! At the end of his article, Price returns to some of Eddington's other ideas – the "going on" or the "passage" of time (the "All is flux" of Heraclitus of Ephesus) and the role of consciousness. He sides with Einstein, Boltzmann and Weyl to advocate a static 'Block Universe' picture (thus echoing Parmenides of Elea). Nevertheless, he argues, Eddington's challenge should be taken up in cosmology, and perhaps in microphysics, in the hope of vindicating Boltzmann's 'Copernican' atemporal perspective. In this context, while the Boltzmann-Schütz hypothesis of low-entropy fluctuations defining local Time's arrows can be critically analyzed, a multiverse approach would seem to restore an overall time-symmetry. More gems are thus to be found in Eddington, the 'cosmological thinker', as in this final note:

> "[I]t is practically certain that a universe containing mathematical physicists will at any assigned date be in the state of maximum disorganization which is not inconsistent with the existence of such creatures."

Last, but not least, a short *poème en prose* by Catherine de Mitry, titled *"Image of Time's Irreversibility"* (*"Image de l'irréversibilité du Temps"*), offers a poetical ending to this volume.

This book, by the depth of the topics covered in the subject of 'Time', should be of broad interest to mathematicians, physicists and philosophers of science. We further hope that the continued publication of this series of Proceedings will serve the scientific community, at both the professional and graduate levels. We thank the Commissariat à l'Énergie Atomique et aux Énergies Alternatives (Division des Sciences de la Matière), the Daniel Iagolnitzer Foundation, the Triangle de la Physique Foundation and the École polytechnique for sponsoring this Seminar. Special thanks are due to Chantal Delongeas for the preparation of the manuscript.

Bertrand Duplantier
Institut de Physique Théorique
Saclay, CEA, France, July 2012

Time, 1–17

Poincaré Seminar 2010

Time and Relativity

Thibault Damour

Abstract. We discuss the interplay between the apparent fundamental irreversibility of Time ("Second Law") and Einstein's views about Space, Time and Matter. A particular attention is given to Boltzmann's 1897 entropic (and anthropic) fluctuation argument, leading to the idea that the "flow of time" is an emergent illusory phenomenon. The basic issue raised by Boltzmann is still the focus of active discussions in modern cosmology, that we briefly review.

1. Introduction

Time has many facets, related to most fields of human endeavour, and to many separate fields of science:

- Metaphysics: e.g., Heraclitus' *"Panta rei"* ("everything flows"); Zeno's arrow, Leibniz's relational time, Kant's ideality of time,. . .
- Spirituality: samsara, maya, sunyata, brahmanda (cosmic egg), orphism, bereishit (book of genesis), eternal return through ekpyrosis, death and resurrection, eternity,. . .
- Psychology: awareness, flow of consciousness,. . .
- Literature: e.g., from Virgil's "fugit irreparabile tempus" to Proust's *In Search of Lost Time.*
- Music: rhythm, tempo, frequency,. . .
- Historical studies: from Herodotus to Fernand Braudel.
- Technology: from Sun dials to LED watches.
- Biology: circadian cycles, aging, programmed cell-death, evolution of species, mitochondrial DNA mutation rate,. . .
- Sociology: working hours, summer time,. . .
- Probability Theory: Bayesian inference, stochastic differential equations, Markov processes, Kolmogorov-Chaitin complexity,. . .
- Astronomy: day, month, year, celestial mechanics, the origin of the solar system, chaos,. . .
- Metrology: atomic clocks, lasers, frequency comparisons,. . .

- Thermodynamics: irreversibility, the Second Law ($dS/dt \geq 0$),...
- Statistical Physics: Boltzmann's equation, Boltzmann's H-theorem, fluctuation-dissipation, Onsager's relations,...
- Chemistry: chaotic chemical reactions, Belousov-Zhabotinsky, self-organisation,...
- Hydrodynamics: Navier-Stokes, viscosity,...
- Information Theory: from Brillouin and Szilard to Shannon, Landauer and Bennett.
- Electromagnetism: retarded potentials versus advanced ones, radiation, the Einstein-Ritz debate, Wheeler-Feynman,...
- Classical Dynamics: Liouville's theorem, periodic systems, quasi-periodic motions, Poincaré recurrences, Lyapunov exponents, chaos, strange attractors,...
- Geology: ...
- Paleontology: ...
- Archeology: ...
- Special Relativity: time dilation, twin paradox, light-cone, Poincaré-Minkowski spacetime geometry,...
- General Relativity: gravitational redshift, GPS, warped spacetime, black holes, worm-holes, time travel, closed time-like curves,...
- Astrophysics: Doppler effect, gravitational redshift, pulsar timing,...
- Cosmology: big bang, expansion of the universe, big crunch, spacelike singularities, inflation, eternal inflation,...
- Quantum Theory: "collapse of the wave function", the measurement issue, the time-energy uncertainty relation, the Zeno effect,...
- Nuclear Physics: nuclear decay, radioactive isotope dating,...
- Atomic Physics: stationary states, quantum transitions, lifetime of unstable states, Ramsey transitions,...
- Quantum Field Theory: Stückelberg-Feynman propagators, Wick rotation, CPT,...
- Quantum Gravity: spacetime foam, (de-)emergence of space time at spacelike singularities, gauge-gravity duality, holography,...

This (certainly incomplete) list illustrates the all pervading significance of the concept of Time. The present contribution will focus only on a few aspects of Time, namely those relating its apparent fundamental irreversibility ("Second Law") to Einstein's revolutionary ideas about Space, Time and Matter, and their import in current developments in physics and cosmology. Before coming to grips with these issues, we shall set the stage by recalling the "common conception of Time" (which was enshrined in Newton's *Principia*), as well as the ground-breaking ideas introduced by Boltzmann in 1897. Our treatment will be rather brief and superficial. The interested reader is referred to the books of Paul Davies [1], Brian Greene [2], Alex Vilenkin [3] and Sean Carroll [4] for more complete discussions, and references to the huge literature on Time.

2. The common conception of Time

In an often quoted sentence of his *Confessions*, Saint Augustine wrote: "What is time? If no one asks me, I know. If I wish to explain it to one that asketh, I know not." However, when pressed to answer the question "what is time?", it seems likely that the most common answer would roughly be that Time is something exterior to the material universe around us, that "passes", or "flows", thereby creating our perception of reality as a "now", as well as inexorably dragging this perception from the past to the future. I do not know for how long this conception of time has been commonly held by human beings (nor do I know whether other animals share it). Among the ancient Greeks, it seems that, with the important exception of Parmenides and his school, the Heraclitean view of a "flow of time" (common to us and the universe) was considered as the standard one. Jumping to more recent times, it seems that mechanical clocks appeared in European convents near the end of the thirteenth century. [They were used to indicate the passing of time to the monks, whose daily prayer and work schedules had to be strictly regulated.]

Later, mechanical clocks became part of the everyday life of ordinary citizens, through the construction of clock towers, notably on cathedrals. Their presence in the city contributed to imposing the conception of a universal time, before the basic scientific advances of the seventeenth century. We have in mind here Galileo (who noticed the isochronism of small pendulum oscillations, and introduced time in the dynamical description of reality), Huyghens (isochronism of a cycloidal pendulum), and Newton. Let us recall how Newton describes his conceptions of time in the scholium to the Definitions at the beginning of his monumental Philosophiae Naturalis Principia Mathematica (1687) [5]:

> "Hitherto I have laid down the definitions of such words as are less known, and explained the sense in which I would have them to be understood in the following discourse. I do not define time, space, place, and motion, as being well known to all. Only I must observe, that the common people conceive those quantities under no other notions but from the relation they bear to sensible objects. And thence arise certain prejudices, for the removing of which it will be convenient to distinguish them into absolute and relative, true and apparent, mathematical and common.
>
> I. Absolute, true, and mathematical time, of itself, and from its own nature, flows equably without relation to anything external, and by another name is called duration: relative, apparent, and common time is some sensible and external (whether accurate or unequable) measure of duration by the means of motion, which is commonly used instead of true time; such as an hour, a month, a year.
>
> II. Absolute space, in its own nature, without relation to anything external, remains always similar and immovable. Relative space is

> some movable dimension or measure of the absolute spaces, which our senses determine by its position to bodies; [...]".

The "Newtonian" concept of absolute time was developed at a time when longcase clocks were used in private homes, and when people started to carry pocket watches. [It seems that Blaise Pascal (1623–1662) was the first to attach his pocket watch to his wrist.] Since that time the conjunction of the tick-tock of public or individual clocks and watches, and of the successful development of the Newtonian description of reality (from celestial mechanics to industrial devices) has "hammered" the common conception of time, recalled above, deeply into the minds of most people.

3. Boltzmann and the first time revolution

The common (and Newtonian) concept of time underwent a first revolution at the end of the nineteenth century, through the work of Boltzmann on the second principle of thermodynamics.

Let us recall that the Second Law of thermodynamics states that the entropy, S, of an isolated system can only increase with time: $dS/dt \geq 0$. This law formalizes, in particular, the many irreversibilities that one observes everyday. E.g., the fact that we see ice cubes melting in a glass of hot water, but we never see a glass of tepid water separating into ice cubes and hot water. Clearly, the Second Law also underlies the fact that we have memories of the past but not of the future, essentially because one needs to have at hand low-entropy reservoirs either to record information on "blank slates" or to "erase" already recorded information (as was discussed by Brillouin, Szilard, Landauer and Bennet; see references in [4]).

We recall that Boltzmann thought he had succeeded (in 1872) in deriving the irreversible increase of the entropy of an isolated mechanical system, $dS/dt \geq 0$, from an innocent-looking assumption about the number of collisions in a gas (the so-called "Stosszahlansatz"). [See [6], p. 88.] However, several scientists raised objections to the "proof" of the "H-theorem" of Boltzmann. [We recall that Boltzmann discussed the evolution of the quantity $H \equiv \int f \ln f \, dq \, dp$ which is the *negative* of the (Boltzmann) entropy S of a gas, described by its one-particle phase-space distribution $f(q,p)$.] First, soon after Boltzmann published his "theorem", Lord Kelvin, Maxwell, Loschmidt and others pointed out that the time-symmetry of the underlying (Newtonian) dynamics of colliding atoms made it impossible to derive, as a mathematical theorem, a time-dissymmetric result such as $dS/dt \geq 0$. This led Boltzmann to stating that the Second Law had only a *statistical* validity, though an overwhelmingly probable one. Twenty years later ($\sim$ 1896), Zermelo raised a new objection based on the recurrence theorem of Poincaré (1890). Indeed, the fact that an isolated system having a *compact* (or, at least, *finite measure*) phase space will recur infinitely many times to a state very close to its initial state (however "improbable" it may be) seems to undermine the existence of a molecular basis of the Second Law. This objection (or, more precisely, a second, related objection of Zermelo concerning the choice of initial state) led Boltzmann to proposing radically new ways of thinking about the physical origin of the Second Law. First, Boltzmann acknowledges that the only way to explain, through considerations of the (reversible) dynamics of molecules, the deeply irreversible nature of the Second Law, is by means of an *assumption* about the state of the universe. He calls this assumption, *assumption* A, and introduces it as follows (Boltzmann 1897, see p. 238 in [6]):

> "The second law will be explained mechanically by means of assumption A (which is of course unprovable) that the universe, considered as a mechanical system – or at least a very large part of it which surrounds us – started from a very improbable state, and is still in an improbable state."

Actually, this rather convoluted sentence means that Boltzmann is here hesitating between two different assumptions. Later in his text, he will clarify their differences, and refer to them as "two kinds of pictures". For clarity, let us give them two different names, say:

- assumption A^{global}: the *entire* universe started from a very improbable state, and is still in an improbable state;
- assumption A^{local}: only the (large but) *local patch* of the universe that surrounds us finds itself at present in a very improbable state.

The meaning of assumption A^{global} is clear. On the other hand, to understand the meaning of assumption A^{local}, it is worth to quote the sentences in which Boltzmann explains the second possible "picture" for understanding the origin of the Second Law:

> "However, one may suppose that the eons during which this[1] improbable state lasts, and the distance from here to Sirius, are minute compared to the age and size of the universe. There must then be in the universe, which is in thermal equilibrium as a whole and therefore dead, here and there relatively small regions of the size of our galaxy (which we call worlds), which during the relatively short time of eons deviate significantly from thermal equilibrium. Among these worlds the state probability increases as often as it decreases. For the universe as a whole the two directions of time are indistinguishable, just as in space there is no up and down. However, just as at a certain place on the earth's surface we can call "down" the direction towards the centre of the earth, so a living being that finds itself in such a world at a certain period of time can define the time direction as going from the less probable to more probable states (the former will be the "past" and the latter the "future") and by virtue of this definition he will find that this small region,

[1]The text of Boltzmann is somewhat confusing as he is here grammatically referring to the "very improbable state" (of assumption A^{global}) in which "the entire universe finds itself at present". We think, however, that he has in mind only a *local* version of assumption A.

isolated from the rest of the universe, is "initially" always in an improbable state. This viewpoint seems to me to be the only way in which one can understand the validity of the second law and the heat death of each individual world without invoking an unidirectional change of the entire universe from a definite initial state to a final state. The objection that it is uneconomical and hence senseless to imagine such a large part of the universe as being dead in order to explain why a small part is living – this objection I consider invalid. I remember only too well a person who absolutely refused to believe that the sun could be 20 million miles from the earth, on the grounds that it is inconceivable that there could be so much space filled only with aether and so little with life."

The visionary nature of this remarkable text has only been appreciated rather recently, especially in the context of modern cosmology. In modern parlance, Boltzmann's vision consists of:

- viewing our visible universe as a spatially and temporally localized entropy fluctuation within an infinite (or much larger) universe;
- appealing (as emphasized in [7]) to a form of the Anthropic Principle: while, globally, the universe is in a "heat death" state, life can exist only in regions where a large enough fluctuation of the entropy away from its maximum takes place;
- considering the "flow of time" as an *emergent illusory phenomenon*, locally induced by the local (spacetime) value of the entropy time-gradient.

I find the latter point the most revolutionary one from the conceptual point of view. It is not clear how many readers of Boltzmann fully realized that if indeed there exist, in the total spacetime[2], *antichronal regions*[3], i.e., local spacetime regions where sentient beings experience time as flowing in the opposite direction than us, this clearly means that the common conception of time as "flowing" externally to the entire universe is incorrect, and that the "flow of time" is a mere (biolo-psycholo-gical) illusion.

The "anthropic-fluctuation" scenario suggested by Boltzmann has been discussed, and rejected, by several physicists. Landau and Lifshitz, in the first English edition (1938) of their *Statistical Physics* volume [8] write:

"Boltzmann attempted to remove this contradiction by his "fluctuation hypothesis". He suggested that in the relatively small part of the universe observed by us, chance fluctuations from the statistical equilibrium of the whole universe are taking place, or, in other words, the impression that the universe does not obey statistical laws is due to our part of the universe being in the course of an enormous fluctuation.

---

[2]Let us emphasize that there is no real anachronism in phrasing Boltzmann's view in terms of *spacetime*. Indeed, for instance, the influential book of H.G. Wells, *The Time Machine* (whose first chapter contains a vivid explanation of Time as a fourth dimension, additional to the three dimensions of Space) was published in 1895, i.e., two years before Boltzmann wrote his text.

[3]i.e., temporal analogues of *antipodal* regions on the Earth.

> The fact that it is possible to observe such a colossal fluctuation (over a volume exceeding 10^{75} c.c.) Boltzmann explained by supposing that just such a fluctuation is a necessary condition for the existence of the observer (a condition favouring biological development of organisms, for instance). This argument is, however, quite false, since there would be an enormously greater probability of a smaller fluctuation for which there existed for instance only a single observer, without the myriads of stars prepared for him, and in any case it would be sufficient for the possibility of observing the universe to have this deviation from equilibrium in a volume of only 10^{55} c.c. (containing the sun and nearest stars). In this connexion we should remark that the probability of fluctuations is so small that it is in general not possible to observe any appreciable fluctuations at all."

A similar argument was presented (in 1963) by Feynman in his *Lectures on Physics* [9]: "Therefore, from the hypothesis that the world is a fluctuation, all of the predictions are that if we look at a part of the world we have never seen before, we will find it mixed up, and not like the piece we looked at. If our order were due to a fluctuation, we would not expect order anywhere but where we have just noticed it." (because the probability of a fluctuation is proportional to $\exp(-S)$, so that "minimal" fluctuations corresponding to the minimal anthropic-compatible local decrease of entropy are a priori much more probable). Then, concerning the origin of the Second Law Feynman concludes that the "one-wayness" displayed by the evolution of any local (isolated) thermodynamical system "cannot be completely understood until the mystery of the beginnings of the history of the universe are reduced still further from speculation to scientific understanding."

On their side, Landau and Lifshitz offered, in the second English edition (1959) of [8], more precise suggestions about the ultimate origin of the Second Law. On the one hand, they point out that:

> "The answer is to be sought in the general theory of relativity. The point is that when we consider large regions of the system, the gravitational fields which they contain begin to become important. According to the general theory of relativity, the latter represent simply changes in the space time metric which is described by the metric tensor g_{ik}. In the study of the statistical properties of bodies, the metrical properties of space time can, in a certain sense, be regarded as the "external conditions" in which these bodies are situated. The assumption that after a long enough interval of time a closed system must eventually reach a state of equilibrium depends obviously on the external conditions remaining constant. But the metric tensor g_{ik} is, generally speaking, a function not only of the co-ordinates but of the time as well, so that the "external conditions" are by no means constant. It is important to note with this that the gravitational field cannot itself be counted as part of the closed system because in that case the conservation laws, which, as

we have seen, are the very foundation of statistics, would become simply identities. As a result of this, in the general theory of relativity the universe as a whole must be regarded not as a closed system, but as one which is in a variable gravitational field. In this case the application of the law of increase of entropy does not imply the necessity of statistical equilibrium."

On the other hand, they raise doubts about the possibility of deriving the Second Law from any (intrinsically time symmetric) classical theory, and they suggest that the time-dissymmetry present in Quantum Mechanics (when adopting the Copenhagen interpretation of measurements) might be related to the Second Law.

4. Einstein, Special Relativity and Time

Textbook presentations of Special Relativity often fail to convey the revolutionary nature, with respect to the "common conception of time", of the seminal paper of Einstein in June 1905. It is true that many of the equations, and mathematical considerations, of this paper were also contained[4] in a 1904 paper of Lorentz, and in two papers of Poincaré submitted in June and July 1905. It is also true that the central informational core of a physical theory is defined by its fundamental equations, and that for some theories (notably Quantum Mechanics) the fundamental equations were discovered before their physical interpretation. However, in the case of Special Relativity, the egregious merit of Einstein was, apart from his new mathematical results and his new physical predictions (notably about the comparison of the readings of clocks which have moved with respect to each other) the *conceptual* breakthrough that the rescaled "local time" variable t' of Lorentz was "purely and simply, the time", as experienced by a moving observer. This new conceptualization of time implied a deep upheaval of the common conception of time. Max Planck immediately realized this and said, later, that Einstein's breakthrough exceeded in audacity everything that had been accomplished so far in speculative science, and that the idea of non-Euclidean geometries was, by comparison, mere "child's play".

The paradigm of the special relativistic upheaval of the usual concept of time is the *twin paradox*. Let us emphasize that this striking example of time dilation proves that *time travel (towards the future) is possible*. As a gedanken experiment (if we neglect practicalities such as the technology needed for reaching velocities comparable to the velocity of light, the cost of the fuel and the capacity of the traveller to sustain high accelerations), it shows that a sentient being can jump, "within a minute" (of his experienced time) arbitrarily far in the future, say sixty million years ahead, and see, and be part of, what (will) happen then on Earth. This is a clear way of realizing that the future "already exists" (as we can experience it "in a minute"). No wonder that many people, attached to the

[4]It is probable that Einstein knew neither the 1904 paper of Lorentz, nor the June 1905 short paper of Poincaré. For historical discussions and references to the original papers, see, e.g., the 2005 Poincaré seminar on Einstein [10] and the book [11].

usual idea of an external flow of time, refused to believe that the travelling twin will come back younger than his sedentary brother. This was notably the case of Bergson whose philosophy was based on a phenomenological intuition of time ("la durée"), experienced in its eternal flow as an "immediate datum of consciousness". Bergson characterized his view of time as follows [12]:

> "Common sense believes in a unique time, the same for all beings and for all things [...]. Each of us feels themselves to experience duration [...] there is no reason, we think, that our duration is not as well the duration of all things."

Today, many experiments have confirmed the reality of time dilation (see the contribution of Christophe Salomon to this seminar). In spite of this, the special relativistic revolution in the concept of time has had little effect on "common-sense". In view of the fact that Copernicus' *De Revolutionibus* appeared in 1543, and that the new world view that this book pioneered started affecting "common sense" only a couple of centuries later, maybe we should not (yet) worry about the little effect that Einstein's 1905 insight has had on the man in the street.

5. General Relativity and Time

General Relativity opened the door to an even deeper upheaval of the common concept of time. However, most popular treatments of science have a tendency, when speaking of General Relativity (GR), and especially when describing relativistic cosmological models (Inflation, Big Bang,...), to use a language which suggests that GR reintroduces the notion of *temporal flow*, which Special Relativity had abolished. Far from it. The spacetime of GR is just a "timeless" as the special relativistic one. The Big Bang should not be referred to as the "birth" of the universe, or its "creation" *ex nihilo*, but as one of the possible "boundaries" of a strongly deformed (timeless) spacetime block.

Far from reintroducing the notion of temporal flow, the infinite variety of possible Einsteinian cosmological models furnish some striking examples of *conceivable "worlds"* where the unreality of this flow becomes palpable. For example, one can imagine a spacetime containing both big bangs (i.e., "lower" boundaries) and big crunches ("upper" boundaries), and such that the privileged "arrow of time" defined by the gradient of entropy in the vicinity of these various spacetime boundaries is, for each boundary, directed towards the interior of the spacetime (as it is for the boundary of our spacetime that is conventionally called "the Big Bang"). The simplest such spacetime, one with one big bang and one big crunch was suggested by Gold [13] as a model of our universe, and as an illustration of a conceivable correlation between the expansion of the universe and the increase of entropy. Hawking thought for a while that this time-symmetric model (featuring a reversal of the time-arrow around the stage of maximum size of the universe) might come out naturally from his Euclidean approach to quantum gravity [14]. However, Page [15] argued against this conclusion.

As already mentioned, we are considering that the "thermodynamic arrow of time", i.e., the direction of time with respect to which entropy grows, is what determines the sensation of "the passage of time", through the irreversibility of the process of memorization in the neuronal structures which give rise to the phenomenon of consciousness. In this view (which only assumes some minimal form of "psycho-physical parallelism") the "flow of time" is illusory, i.e., does not correspond to any "real" passage of time, while the "arrow of time" does correspond to a "real" structure of spacetime, namely a certain "stratification" of spacetime by hypersurfaces of varying entropy. [Note that this stratification is "static", and does not correspond to the common idea of a "stratum of the present" which would "move" towards the future, like a projector successively illuminating the various "entropy strata" of spacetime.]

Another example of a relativistic cosmos which puts into question the usual notion of temporal flow is the one introduced in 1949 by the famous mathematician Kurt Gödel [16]. Gödel's cosmos does not admit a "stratification" by global spacelike hypersurfaces. Locally, this spacetime admits a Lorentzian structure, i.e., it contains a regular field of lightcones separating timelike from spacelike directions. Near each point, one can therefore define pieces of spacelike hypersurfaces, and use them to distinguish the "upper" parts of the lightcones (the "future-directed" timelike vectors) from their "lower" parts (the "past-directed" ones). However, such a construction cannot be done globally because Gödel has shown that there exist "closed time-like curves" (CTCs), i.e., worldlines, representing the history of observers living in this cosmos, which close in on themselves like circles. In other words, in Gödel's spacetime it is possible to *travel into the past.* Gödel even showed that given any "starting" point P in spacetime (e.g., "here and now" for you, reader of those lines), and any wished "arrival" point Q (e.g., Mount Golgotha, on a certain Friday of April A.D. 33), one can travel (along an initially future-directed time-like path) from P to Q in a finite time (which can be, in principle, as short as wished). As far as we know, the structure of our cosmological spacetime does not include the feature of Gödel's one that leads to CTCs (namely the existence of a "rotation field" that can progressively tip the lightcones so as to reverse their orientation). However, the point of Gödel was not to claim that our universe is similar to his model but was to give a *conceivable* cosmos (solution of Einstein's field equations) in which the usual notion of universal time-flow becomes meaningless. The mere possibility of having such a solution[5] of Einstein's equations shows that, in General Relativity, the "external flow of time" can only be an "illusion", which depends on some particular structure of the spacetime we "live in". As is well known, time travels can lead to paradoxical situations, but none of these paradoxes constitute a proof of non-existence. We should keep in mind, as an analogy, that the "twin paradox" has often been used as a proof of the inconsistency of

[5]It was later found that many other solutions of Einstein's theory of General Relativity can lead to time travel and CTCs: e.g., an overcritical rotating Kerr solution, solutions containing wormholes,...

the special relativistic time-dilation. We know, however, that it corresponds to a real effect, and that the "paradox" was just due to conceptual conservatism. For a detailed discussion of time travel's classical and quantum physics see [17].

6. Relativistic Gravity and the Second Law

In the last two Sections we have been mainly trying to show that, at the conceptual level, both the Special and the General Theories of Relativity suggest that one should open one's mind and stop being formatted by the traditional (and deeply ingrained) idea that Time exists as an entity outside the material world around us, and "drags" the "common now of the universe" as it "passes". In the words of Einstein:

> "For us, physicists in the soul[6] the distinction between past, present and future is only a stubbornly persistent illusion."

In the following, we shall take for granted the idea that, as Einstein wrote once to his friend Michele Besso, "subjective time with its "now" [does] not have any objective significance", i.e., that it does not correspond to a "unique time, the same for all beings and for all things", as Bergson described the "common sense" idea of time. On the other hand, we shall take for granted that the subjective experience we have, as human beings, of the "flow of time" is ultimately rooted in the Second Law of thermodynamics, i.e., in the (objective) fact that, as said Boltzmann, "the universe, considered as a mechanical system – or at least a very large part of it which surrounds us – started from a very improbable state, and is still in an improbable state."

Within this view, the basic question that needs to be addressed is: what is the physical origin of the massive time-dissymmetry embodied in the very special past (and present) state of the entire visible universe? This issue will be the main topic of the rest of this lecture. Let us start by noting that several different sectors of physics (or of the world around us) exhibit important time dissymmetries:

1. Thermodynamics: the Second Law;
2. Electrodynamics: retarded-potential radiation;[7]
3. Expansion of the Universe;
4. Irreversible behaviour of black holes (see below);
5. Quantum Mechanics: irreversibility in the Copenhagen interpretation of measurements.

[6] The German expression used by Einstein is "gläubige Physiker", which is often translated as "believing physicists". Nevertheless, all the philosophical context of Einstein's thought shows that one must not understand the word "believing" in the sense of a traditional religious belief, but rather in the sense of a deep belief in the rationality of the universe. Because of this, it seems to us more appropriate to translate "gläubige Physiker" by "physicists in the soul", or by "convinced physicists".

[7] Note that the observations of binary pulsars have also shown that gravitational radiation is emitted via *retarded* potentials, rather than advanced ones.

Our view is that the facts 2, and the interpretation 5, have the same origin as 1, namely a very special state in the past (and today). We shall therefore focus on the points 3 and 4, i.e., on the question whether relativistic gravity (and cosmology) are related with the origin of the Second Law.

We saw above that Landau and Lifshitz suggested (starting in 1959) that relativistic cosmology might indeed be closely related with the Second Law. I am not sure who was the first to suggest such a connection. Though Friedmann [18] was the first to introduce time-dependence in cosmology, and to suggest a phoenix-type cyclic universe, undergoing successive bounces, I am not aware of his discussing the issue of the Second Law within such a model. This was specifically discussed by Tolman in 1932 [19]. Let us quote one of his main conclusions:

> "The main purpose of this article has been a further examination of the bearings of relativistic thermodynamics on the well-known problem of the entropy of the universe as a whole. The work has again illustrated the necessity of using relativistic rather than classical thermodynamics in treating this problem, and has demonstrated that the framework of general relativity at least provides a class of conceivable models of the universe which would undergo a continued series of expansions and contractions without being brought to rest by the irreversible processes which accompany these changes. The findings of relativistic thermodynamics thus stand in sharp contrast to the familiar conclusion of the classical thermodynamics that the continued occurrence of irreversible processes would lead to an ultimate condition of maximum entropy and minimum free energy where change would cease."

The point of Tolman is interesting but does not really address the issue of the origin of the Second Law.

A few months before the paper of Tolman (which was submitted on November 13, 1931) the issue of time asymmetry (and "time's arrow") in cosmology was discussed by Eddington and by Lemaître. See the lecture of Huw Price for a discussion of Eddington's ideas. Here, I will only consider the ideas of Lemaître, focussing on his remarkable Letter to Nature, [20]. This Letter is very short and is worth quoting in its entirety:

> "**The Beginning of the World from the Point of View of Quantum Theory**
>
> Sir Arthur Eddington[8] states that, philosophically, the notion of a beginning of the present order of Nature is repugnant to him. I would rather be inclined to think that the present state of quantum theory suggests a beginning of the world very different from the present order of Nature. Thermodynamical principles from the point of view of quantum theory may be stated as follows: (1) Energy of constant total amount is distributed in discrete quanta. (2) The number of distinct

[8] *Nature*, Mar. 21, p. 447.

quanta is ever increasing. If we go back in the course of time we must find fewer and fewer quanta, until we find all the energy of the universe packed in a few or even in a unique quantum.

Now, in atomic processes, the notions of space and time are no more than statistical notions; they fade out when applied to individual phenomena involving but a small number of quanta. If the world has begun with a single quantum, the notions of space and time would altogether fail to have any meaning at the beginning; they would only begin to have a sensible meaning when the original quantum had been divided into a sufficient number of quanta. If this suggestion is correct, the beginning of the world happened a little before the beginning of space and time. I think that such a beginning of the world is far enough from the present order of Nature to be not at all repugnant.

It may be difficult to follow up the idea in detail as we are not yet able to count the quantum packets in every case. For example, it may be that an atomic nucleus must be counted as a unique quantum, the atomic number acting as a kind of quantum number. If the future development of quantum theory happens to turn in that direction, we could conceive the beginning of the universe in the form of a unique atom, the atomic weight of which is the total mass of the universe. This highly unstable atom would divide in smaller and smaller atoms by a kind of super-radio-active process. Some remnant of this process might, according to Sir James Jean's idea, foster the heat of the stars until our low atomic number atoms allowed life to be possible.

Clearly the initial quantum could not conceal in itself the whole course of evolution; but, according to the principle of indeterminacy, that is not necessary. Our world is now understood to be a world where something really happens; the whole story of the world need not have been written down in the first quantum like a song on the disc of a phonograph. The whole matter of the world must have been present at the beginning, but the story it has to tell may be written step by step."

Though the formulation of Lemaître is somewhat unclear and imprecise, it carries a deep vision of a possible interlocking between: (1) the Second Law; (2) Quantum Mechanics; and (3) the emergence of the universe (and of space and time) from "a single quantum" [that he later explicitly connected to his work on expanding cosmological models, with a cosmological constant, under the name of "the primeval atom" ("l'atome primitif")]. Besides the vision, what can be retained today of the suggestions of Lemaître is unclear. However, the issues discussed by Lemaître have been hotly discussed ever since. The already mentioned works of Gold, of Hawking and of Page, being examples of such discussions.

Let us note that both Tolman and Lemaître (especially in his subsequent, more detailed papers) were mainly discussing the issue of the increase of the entropy of the material content of the universe, taking the Einsteinian spacetime essentially as an external, self-consistent time-dependent background. The issue of the "entropy" to be attributed to this external gravitational field was (apparently) not considered. This issue got a decisive impetus from the work of Christodoulou [21], Christodoulou and Ruffini [22], and Hawking [23] on the *irreversible* aspects of the physics of black holes. These authors discovered that, during the interaction of one or several black holes with external particles or fields, a certain quantity (the square irreducible mass, or the area) could only increase. The later work of Bekenstein [24] and Hawking [25] led to attribute to any black hole the entropy (with $c = 1$)

$$S_{\mathrm{BH}} = \frac{A}{4\,G\hbar} \tag{1}$$

where A is the area of the horizon.

The statistical physics meaning of Eq. (1) is still rather mysterious (in spite of remarkable results in string theory), but there is no doubt that it is telling us something deep about a three-way link between quantum theory, general relativity and thermodynamics.

7. Primordial cosmology and the Second Law

To end this brief survey, let us mention some of the recent attempts at connecting the Second Law with primordial cosmology. Several possibilities have been suggested (references on the recent works alluded below can be easily obtained from the web, or from the books quoted at the beginning of this lecture).

- The "chaotic inflation paradigm" (Linde) [or, alternatively, the eternal inflation paradigm (Vilenkin, Linde)] argues that our entire visible universe developed from a roughly homogeneous Planck-scale patch of a "random" universe. The inflationary mechanism, together with the a posteriori condition of looking only at large (inflated) patches, seems to naturally introduce a dissymmetry, explaining why the post-inflationary universe "starts" in a rather low-entropy state (compared, say, to the present one).
- The "special boundary paradigms" wish to add to the dynamical laws of nature, an additional prescription to select the global state of the universe. For instance, R. Penrose suggests to impose the vanishing of the Weyl curvature at "initial" spacetime singularities. Another example, is the "no-boundary" proposal of Hartle and Hawking which tries to restrict the quantum amplitude of the universe by generalizing the Euclidean-time characterization of "ground state" wavefunctions in quantum field theory.
- The "quantum tunnelling paradigms" wish to describe our universe as the result of a quantum tunnelling from some previous state. [Note that this is reminiscent of the "super-radioactive process" contemplated by Lemaître.] The "tunnelling from nothing" scenario of Vilenkin is similar to the Euclidean-time description of pair creation. Many scenarios explored various possible tunnellings between different "vacua" of some underlying theory (Garriga

and Vilenkin, Dyson, Kleban and Susskind, Albrecht and Sorbo, Carroll and Chen, ...).

All these studies have usefully stretched our imagination about the possible origin of our world. However, it is not clear that any of them provides a satisfactory answer to the basic question of the origin of the Second Law. For instance, the chaotic inflation scenario looks a priori quite appealing. It uses the "ironing" effect of inflation to stretch a small, inhomogeneous patch into a huge, nearly homogeneous space. Moreover, as the inflationary behaviour is a dynamical attractor, some authors have argued that "most" initial states will be inflated and thereby ironed out (Belinsky, Khalatnikov, Grishchuk and Zeldovich 1985, Kofman, Linde and Mukhanov 2002). But, other authors (Khalfin 1989, Hollands and Wald 2002) have argued that "very "special" initial conditions are nevertheless needed in order to enter an era of inflation". Basically, their argument is similar to the old objections of Kelvin, Maxwell, Loschmidt and Zermelo to Boltzmann: the time-reversibility of the underlying (general relativistic) dynamics, and the invariance of some Liouville measure imply that *any* present state of the universe (as inhomogeneous as wished), must come from *some* initial state. Therefore the latter initial state was *not* ironed out by inflation, which shows that only a special class of initial states can be ironed out by inflation. In spite of the apparent strength of this argument, some specific aspects of gravity, and of the interplay between gravity and quantum mechanics, make it problematic. We have here in mind three issues:

- Contrary to what happens in usual dynamical systems, or for usual (non gravitational) fields, the Liouville measure of spatially compact, finite-energy systems *does not have a finite integral* when one includes the gravitational degrees of freedom. [This comes from a famous *minus sign* associated to the conformal mode in gravity.] Because of this, we cannot use the Liouville measure (or its various possible quotients) to estimate the likelihood of some state.
- Some authors (most notably R. Penrose) estimate the probability of the initial state by assuming that the irreversible behaviour linked to gravitational clumping in our universe can be quantified in terms of some "entropy" S_g linked to the gravitational field; and they use the Bekenstein-Hawking Black Hole entropy (1) to estimate S_g now. However, this type of estimate is not justified by our (rather incomplete) knowledge of the thermodynamics of self-gravitating systems.
- Quantum gravity considerations suggest that inhomogeneous modes having Planck-scale wavelengths ($\lambda \sim \ell_P \equiv \sqrt{\hbar G}$) must be (effectively) excluded from the Hilbert space of physical quantum states. This has two types of effects: (1) it introduces a high-frequency cut-off and thereby allows inflation to iron out *all* the initial states with $\lambda \gtrsim \ell_P$; (2) it effectively introduces a violation of the conservation of the number of states (which is the quantum version of Liouville's theorem) during the expansion.

In conclusion, we see that the basic issue raised by Boltzmann long ago is still with us. What is new is that we now think that its answer (if any) lies at the interplay between relativistic gravity and quantum mechanics. This reminds us of the suggestion of Landau and Lifshitz, except that most people interested in this issue do not interpret quantum mechanics in the Copenhagen way, but rather in the Everett's one [26].

References

[1] Davies, Paul: *About Time, Einstein's Unfinished Revolution* (Touchstone, Simon & Schuster, New York, 1995); see also his earlier (more technical) book: *The Physics of Time Asymmetry* (Survey University Press, London, 1974).

[2] Greene, Brian: *The Fabric of the Cosmos, Space, Time and the Texture of Reality* (Albert A. Knopf, New York, 2004).

[3] Vilenkin, Alex: *Many Worlds in One, The Search for Other Universes* (Hill and Wang, New York, 2006).

[4] Carroll, Sean: *From Eternity to Here, The Quest for the Ultimate Theory of Time* (Dutton, New York, 2010).

[5] Newton, I.: *The Mathematical Principles of Natural Philosophy*, trans. A. Motte (University of California Press, Berkeley, 1934), p. 6.

[6] Brush, S.G.: *Kinetic Theory: Vol.* 2 – *Irreversible processes* (Pergamon, Oxford, 1966).

[7] Barrow, J.D., and Tipler, F.J.: *The Anthropic Cosmological Principle* (Oxford Univ. Press, Oxford, 1988).

[8] Landau, L., and Lifshitz, E.: *Statistical Physics*, both (Clarendon Press, Oxford, 1938) and (Pergamon Press, London, 1959), see §8, i.e., the end of Chapter 1.

[9] Feynman, R.P., Leighton, R.B., and Sands, M.: *The Feynman Lectures on Physics*, (Addison-Wesley, Reading, Massachusetts, 1963), Volume I, pages 46-8 and 46-9.

[10] Damour, T., Darrigol, O., Duplantier, B., and Rivasseau, V.: *Einstein* 1905–2005*, Poincaré Seminar* 2005 (Birkhäuser, Basel, 2006).

[11] Damour, T.: *Si Einstein m'était conté* (Le Cherche Midi, Paris, 2005); English translation: *Once Upon Einstein* (A.K. Peters, Wellesley, 2006), translated by E. Novak.

[12] Bergson, H.: *Durée et Simultanéité. À propos de la théorie d'Einstein* (Félix Alcan, Paris, 1922).

[13] Gold, T.: American Journal of Physics **30**, 403 (1962).

[14] Hawking, S.W.: Phys. Rev. D **32**, 2489 (1985).

[15] Page, D.N.: Phys. Rev. D **32**, 2496 (1985).

[16] Gödel, K.: Reviews of Modern Physics **21**, 447 (1949).

[17] Deutsch, D.: *The Fabric of Reality* (Allen Lane, New York, 1997).

[18] For a useful selection of the cosmological works (translated in French) of Friedmann and Lemaître, together with a detailed presentation by J.P. Luminet, see A. Friedmann and G. Lemaître, *Essais de Cosmologie*, preceded by *L'Invention du Big Bang*

by J.P. Luminet, texts selected and translated by J.P. Luminet and A. Grib (Le Seuil, Paris, 1997).

[19] Tolman, R.C.: Phys. Rev. **39**, 320 (1932).

[20] Lemaître, G.: Nature **127**, 706 (9 May 1931).

[21] Christodoulou, D.: Phys. Rev. Lett. **25**, 1596 (1970).

[22] Christodoulou, D., and Ruffini, R.: Phys. Rev. D **4**, 3552 (1971).

[23] Hawking, S.W.: Phys. Rev. Lett. **26**, 1344 (1971).

[24] Bekenstein, J.: Phys. Rev. D **7**, 2333 (1973).

[25] Hawking, S.W.: Comm. Math. Phys. **43**, 199 (1975).

[26] Everett, H.: Reviews Mod. Phys. **29**, 454 (1957).

Thibault Damour
Institut des Hautes Études Scientifiques
Le Bois-Marie
35, route de Chartres
F-91440 Bures-sur-Yvette, France

Time, 19–79

Poincaré Seminar 2010

(Ir)reversibility and Entropy

Cédric Villani

Abstract. In this text, I evoke the issue of time in classical statistical physics, insisting on the problem of time's arrow and irreversibility. Boltzmann's entropy and the H Theorem play a key part, as well as the damping without information loss imagined by Landau. Some paradoxes of various fame are analyzed from the physical and mathematical points of view.

La cosa più meravigliosa è la felicità del momento
L. Ferré

Time's arrow is part of our daily life and we experience it every day: broken mirrors do not come back together, human beings do not rejuvenate and rings grow unceasingly in tree trunks. In sum, time always flows in the same direction! Nonetheless, the fundamental laws of classical physics do not favor any time direction and conform to a rigorous symmetry between past and future. It is possible, as discussed in the article by T. Damour in this same volume, that irreversibility is inscribed in other physical laws, for example on the side of general relativity or quantum mechanics. Since Boltzmann, **statistical physics** has advanced another explanation: time's arrow translates a constant flow of less likely events toward more likely events. Before continuing with this interpretation, which constitutes the guiding principle of the whole exposition, I note that the flow of time is not necessarily based on a single explanation.

At first glance, Boltzmann's suggestion seems preposterous: it is not because an event is *probable* that it is actually achieved, but time's arrow seems inexorable and seems not to tolerate any exception. The answer to this objection lies in a catchphrase: **separation of scales**. If the fundamental laws of physics are exercised on the microscopic, particulate (atoms, molecules, ...) level, phenomena that we can sense or measure involve a considerable number of particles. The effect of this number is even greater when it enters combinatoric computations: if N, the number of atoms participating in an experiment, is of order 10^{10}, this is already considerable, but $N!$ or 2^N are supernaturally large, invincible numbers.

The innumerable debates between physicists that have been pursued for more than a century, and that are still pursued today, give witness to the subtlety and depth of Maxwell's and Boltzmann's arguments, banners of a small scientific revolution that was accomplished in the 1860s and 1870s, and which saw the birth of the fundamentals of the modern kinetic theory of gases, the universal concept of statistical entropy and the notion of macroscopic irreversibility. In truth, the arguments are so subtle that Maxwell and Boltzmann themselves sometimes went astray, hesitating on certain interpretations, alternating naive errors with profound concepts; the greatest scientists at the end of the nineteenth century, e.g., Poincaré and Lord Kelvin, were not to be left behind. We find an overview of these delays in the book by Damour already mentioned; for my part, I am content to present a "decanted" version of Boltzmann's theory. At the end of the text I shall evoke the way in which Landau shattered Boltzmann's paradigm, discovering an apparent irreversibility where there seemed not to be any and opening up a new mine of mathematical problems.

In retracing the history of the statistical interpretation of time's arrow, I shall have occasion to make a voyage to the heart of profound problems that have agitated mathematicians and physicists for more than a century.

The notation used in this exposition are generally classical; I denote $\mathbb{N} = \{1, 2, 3, \ldots\}$ and log = natural logarithm.

1. Newton's inaccessible realm

I shall adopt here a purely classical description of our physical universe, in accordance with the laws enacted by Newton: the ambient space is Euclidean, time is absolute and acceleration is equal to the product of the mass by the resultant of the forces.

In the case of the description of a gas, these hypotheses are questionable: according to E.G.D. Cohen, the quantum fluctuations are not negligible on the mesoscopic level. The probabilistic nature of quantum mechanics is still debated; we nevertheless accept that the resulting increased uncertainty due to taking these uncertainties into account can but arrange our affairs, at least qualitatively, and we thus concentrate on the classical and deterministic models, "à la Newton".

1.1. The solid sphere model

In order to fix the ideas, we consider a system of ideal spherical particles bouncing off one another: let there be N particles in a box Λ. We let $X_i(t)$ denote the position at time t of the center of the ith particle. The rules of motion are stated as follows:

- We suppose that initially the particles are well separated ($i \neq j \implies |X_i - X_j| > 2r$) and separated from the walls ($d(X_i, \partial\Lambda) > r$ for all i).
- While these separation conditions are satisfied, the movement is uniformly rectilinear: $\ddot{X}_i(t) = 0$ for each i, where we denote $\ddot{X} = d^2X/dt^2$, the acceleration of X.

- When two particles meet, their velocities change abruptly according to Descartes' laws: if $|X_i(t) - X_j(t)| = 2r$, then

$$\begin{cases} \dot{X}_i(t^+) = \dot{X}_i(t^-) - 2\left\langle \dot{X}_i(t^-) - \dot{X}_j(t^-), n_{ij} \right\rangle n_{ij}, \\ \dot{X}_j(t^+) = \dot{X}_j(t^-) - 2\left\langle \dot{X}_j(t^-) - \dot{X}_i(t^-), n_{ji} \right\rangle n_{ji}, \end{cases}$$

  where $n_{ij} = (X_i - X_j)/|X_i - X_j|$ denotes the unit vector joining the centers of the colliding balls.
- When a particle encounters the boundary, its velocity also changes: if $|X_i - x| = r$ with $x \in \partial\Lambda$, then

$$\dot{X}_i(t^+) = \dot{X}_i(t^-) - 2\left\langle \dot{X}_i(t), n(x) \right\rangle n(x),$$

  where $n(x)$ is the exterior normal to Λ at x, supposed well defined.

These rules are not sufficient for completely determining the dynamics: we cannot exclude *a priori* the possibility of triple collisions, simultaneous collisions between particles and the boundary, or again an infinity of collisions occurring in a finite time. However, such events are of probability zero if the initial conditions are drawn at random with respect to Lebesgue measure (or Liouville measure) in phase space [40, appendix 4.A]; we thus neglect these eventualities. The dynamic thus defined, as simple as it may be, can then be considered as a caricature of our complex universe if the number N of particles is very large. Studied for more than a century, this caricature has still not yielded all its secrets; far from that.

1.2. Other Newtonian models

Beginning with the emblematic model of hard spheres, we can define a certain number of more or less complex variants:

- replace dimension 3 by an arbitrary dimension $d \geq 2$ (dimension 1 is likely pathological);
- replace the boundary condition (elastic rebound) by a more complex law [40, chapter 8];
- or, instead, eliminate the boundaries, always delicate, by setting the system in the whole space $\mathbb{R}^d$ (but we may then add that the number of particles must then be infinite so as keep a nonzero global mean density) or in a torus of side L, $\mathbb{T}^d_L = \mathbb{R}^d/(L\mathbb{Z}^d)$, which will be my choice of preference in the sequel;
- replace the contact interaction of hard spheres by another interaction between point particles, e.g., associated with an interaction potential between two bodies: $\phi(x - y) =$ potential exerted at point x by a material point situated at y.

Among the notable interaction potentials in dimension 3 we mention (within a multiplicative constant):

– the **Coulomb** potential: $\phi(x - y) = 1/|x - y|$;
– the **Newtonian** potential: $\phi(x - y) = -1/|x - y|$;
– the **Maxwellian** potential: $\phi(x - y) = 1/|x - y|^4$.

The Maxwellian interaction was artificially introduced by Maxwell and Boltzmann in the context of the statistical study of gases; it leads to important simplifications in certain formulas. There exists a taxonomy of other potentials (Lennard-Jones, Manev…). The hard spheres correspond to the limiting case of a potential that equals 0 for $|x-y|>r$ and $+\infty$ for $|x-y|<2r$.

Suppose, more generally, that the interaction takes place on a scale of order r and with an intensity a. We end up with a **system of point particles with interaction potential**

$$\ddot{X}_i(t) = -a \sum_{j\neq i} \nabla\phi\left(\frac{X_i - X_j}{r}\right), \tag{1}$$

for each $i \in \{1,\dots,N\}$; we thus suppose that $X_i \in \mathbb{T}^d_L$. Here again, the dynamic is well defined except for a set of exceptional initial conditions and it is associated with a **Newtonian flow** $\mathcal{N}_t$, which maps the configuration at time s to the configuration at time $s+t$ ($t \in \mathbb{R}$ can be positive or negative).

1.3. Distribution functions

Even if one accepts the Newtonian model (1), it remains *inaccessible* to us: first because we cannot perceive the individual particles (too small), and because their number N is large. By well-designed experiments, we can measure the pressure exerted on a small surface, the temperature about a point, the mean density, etc. None of these quantities is expressed directly in terms of the X_i, but rather in terms of averages

$$\frac{1}{N}\sum_i \chi(X_i, \dot{X}_i), \tag{2}$$

where χ is a scalar function.

It may seem an idle distinction: in concentrating χ near the particle i, we retrieve the missing information. But quite clearly this is impossible: in practice χ is of *macroscopic* variation, e.g., of the order of the size of the box. Besides, the information contained in the averages (2) does not distinguish particles, so that we have to replace the vector of the $(X_i, \dot{X}_i)$ by the **empirical measure**

$$\hat{\mu}^N_t = \frac{1}{N}\sum_{i=1}^N \delta_{(X_i(t), \dot{X}_i(t))}. \tag{3}$$

The terminology "empirical" is well chosen: it is the measure that is observed by means (without intending a pun) of measurements.

To resume: our knowledge of the particle system is achieved only through the behavior of the empirical measure in a *weak topology* that models the macroscopic limitation of our experiments – laboratory experiments as well as sensory perceptions.

Frequently, on our own scale, the empirical measure appears continuous:

$$\hat{\mu}^N_t(dx\,dv) \simeq f(t,x,v)\,dx\,dv.$$

We often use the notation $f(t, \cdot) = f_t$. The density f is the **kinetic distribution** of the gas. The study of this distribution constitutes the kinetic theory of gases; the founder of this science is undoubtedly D. Bernoulli (around 1738), and the most famous contributors to it are Maxwell and Boltzmann. A brief history of kinetic theory can be found in [40, Chapter 1] and in the references there included.

We continue with the study of the Newtonian system. We can imagine that certain experiments allow for simultaneous measurement of the parameters of various particles, thus giving access to correlations between particles. This leads us to define, for example,

$$\hat{\mu}_t^{2;N}(dx_1\,dv_1\,dx_2\,dv_2) = \frac{1}{N(N-1)}\sum_{i\neq j}\delta_{(X_{i_1}(t),\dot{X}_{i_2}(t),X_{i_2}(t),\dot{X}_{i_2}(t))},$$

or more generally

$$\hat{\mu}_t^{k;N}(dx_1\,dv_1\,\ldots\,dx_k\,dv_k) = \frac{(N-k-1)!}{N!}\sum_{(i_1,\ldots,i_k)\text{ distinct}}\delta_{(X_{i_1}(t),\dot{X}_{i_1}(t),\ldots,X_{i_k}(t),\dot{X}_{i_k}(t))}.$$

The corresponding approximations are **distribution functions in k particles**:

$$\hat{\mu}_t^{k;N}(dx_1\,dv_1\,\ldots\,dx_k\,dv_k) \simeq f^{(k)}(t,x_1,v_1,\ldots,x_k,v_k).$$

Evidently, by continuing up until $k = N$, we find a measure $\hat{\mu}^{N;N}(dx_1\,\ldots\,dv_N)$ concentrated at the vector of particle positions and velocities (the mean over all permutations of the particles). But in practice we never go to $k = N$: k remains very small (going to 3 would already be a feat), whereas N is huge.

1.4. Microscopic randomness

In spite of the determinism of the Newtonian model, hypotheses of a probabilistic nature on the initial data have already been made, by supposing that they are not configured to end up in some unusual catastrophe such as a triple collision. We can now generalize this approach by considering a probability distribution on the set of initial positions and velocities:

$$\mu_0^N(dx_1\,dv_1\,\ldots\,dx_N\,dv_N),$$

which is called a **microscopic probability measure**. In the sequel we will use the abbreviated notation

$$dx^N\,dv^N := dx_1\,dv_1\,\ldots\,dx_N\,dv_N.$$

It is natural to choose μ_0^N symmetric, i.e., invariant under coordinate permutations. The data μ_0^N replace the measure $\hat{\mu}_0^{N;N}$ and generalize it, giving rise to a flow of measures, obtained by the action of the flow:

$$\mu_t^N = (\mathcal{N}_t)_\#\mu_0^N,$$

and the marginals

$$\mu_t^{k;N} = \int_{(x_1,v_1,\ldots,x_k,v_k)} \mu_t^N.$$

If the sense of the empirical measure is transparent (it is the "true" particle density), that of the microscopic probability measure is less evident. Let us assume that the initial state has been prepared by a great combination of circumstances about which we know little: we can only make suppositions and guesses. Thus μ_0^N is a probability measure on the set of possible initial configurations. A physical statement involving μ_0^N will, however, scarcely make sense if we use the precise form of this distribution (we cannot verify it, since we do not observe μ_0^N); but it will make good sense if a μ_0^N-almost certain property is stated, or indeed with μ_0^N-probability of 0.99 or more.

Likewise, the form of $\mu_t^{1;N}$ has scarcely any physical meaning. But if there is a phenomenon of concentration of measure due to the hugeness of N, then it may be hoped that

$$\mu_0^N \left[\operatorname{dist}\left(\hat{\mu}_t^N,\ f_t(x,v)\,dx\,dv\right) \geq r\right] \leq \alpha(N,r),$$

where dist is a well-chosen distance on the space of measures and $\alpha(N,r) \to 0$ when $r \to \infty$, all the faster that N is large (for example $\alpha(N,r) = e^{-c\,N\,r}$). We will then have

$$\begin{aligned}
\operatorname{dist}\left(\mu_t^{1;N}, f_t(x,v)\,dx\,dv\right) &= \operatorname{dist}\left(\int \hat{\mu}_t^N\,d\mu_t^N,\ f_t(x,v)\,dx\,dv\right)\\
&\leq \int \operatorname{dist}\left(\hat{\mu}_t^N, f_t\,dx\,dv\right)\\
&\leq \int_0^\infty \alpha(N,r)\,dr =: \eta(N).
\end{aligned}$$

If $\eta(N) \to 0$ when $N \to \infty$ it follows that, with very high probability, $\mu_t^{1;N}$ is an excellent approximation to $f(t,x,v)\,dx\,dv$, which itself is a good approximation to $\hat{\mu}_t^N$.

1.5. Micromegas

In this section I shall introduce two very different statistical descriptions: the macroscopic description $f(t,x,v)\,dx\,dv$ and the microscopic probabilities $\mu_t^N(dx^N\,dv^N)$. Of course, the quantity of information contained in μ^N is considerably more important than that contained in the macroscopic distribution: the latter informs us about the state of a typical particle, whereas a draw following the distribution μ_t^N informs us about the state of *all* particles. Think that if we have 10^{20} degrees of freedom, we will have to integrate 99999999999999999999 of them. For handling such vertiginous dimensions, we will require a fundamental concept: entropy.

2. The entropic world

The concept and the name entropy were introduced by Clausius in 1865 as part of the theory – then under construction – of thermodynamics. A few years later Boltzmann (certainly influenced by the statistical ideas put forward by Laplace, Quetelet and others) revolutionized the concept by giving it a statistical interpretation based on atomic theory. In addition to this section, the reader can consult, e.g., Balian [9, 10] about the notion of entropy in physical statistics.

2.1. Boltzmann's formula

Let a physical system be given, which we suppose is completely described by its microscopic state $z \in \mathcal{Z}$. Experimentally we only gain access to a partial description of that state, say $\pi(z) \in \mathcal{Y}$, where $\mathcal{Y}$ is a space of macroscopic states. I will not give precise hypotheses on the spaces $\mathcal{Z}$ and $\mathcal{Y}$, but with the introduction of measure theory we will implicitly assume that these are "Polish" (separable complete metric) spaces.

How can we estimate the amount of information that is lost when we summarize the microscopic information by the macroscopic? Assuming that $\mathcal{Y}$ and $\mathcal{Z}$ are denumerable, it is natural so suppose that the uncertainty associated with a state $y \in \mathcal{Y}$ is a function of the cardinality of the pre-image, i.e., $\#\pi^{-1}(y)$.

If we carry out two independent measures of two different systems, we are tempted to say that the uncertainties are additive. Now, with obvious notation, $\#\pi^{-1}(y_1, y_2) = (\#\pi_1^{-1}(y_1))\,(\#\pi_2^{-1}(y_2))$. To pass from this multiplicative operation to an addition, let us take a multiple of the logarithm. We thus end up with Boltzmann's celebrated formula, engraved on his tombstone in the Central Cemetery in Vienna:

$$S = k \log W, \tag{4}$$

where $W = \#\pi^{-1}(y)$ is the number of microscopic states compatible with the observed macroscopic state y and k is the so-called Boltzmann constant.[1]

In numerous cases, the space $\mathcal{Z}$ of microscopic configurations is continuous, and in applying Boltzmann's formula it is customary to replace the counting measure by a privileged measure: for example by Liouville measure if we are interested in a Hamiltonian system. Thus W in (4) can be the *volume* of microscopic states that are compatible with the macroscopic state y.

If the space $\mathcal{Y}$ of macroscopic configurations is likewise continuous, then this notion of volume must be handled cautiously: the fiber $\pi^{-1}(y)$ is typically of volume zero and thus of scarce interest. One is tempted to postulate, for a given topology,

$$S(y) = \text{f.p.}_{\varepsilon\to 0} \log \left|\pi^{-1}(B_\varepsilon(y))\right|,$$

where $B_\varepsilon(y)$ is the ball of radius ε centered at y and f.p. denotes the finite part, meaning that we excise the divergence in ε, if indeed it has a universal behavior.

[1]Even if this formula accurately reflects Boltzmann's thoughts, it was Planck who first wrote it in this particular form, around 1900.

If this last point is not at all evident, the universality is nonetheless verified in the particular case that interests us where the microscopic state $\mathcal{Z}$ is the space of configurations of N particles, i.e., $\mathcal{Y}^N$, and where we begin by taking the limit $N \to \infty$. In this limit, as we will see, the mean entropy per molecule tends to a finite value and we can subsequently take the limit $\varepsilon \to 0$, which corresponds to an arbitrarily precise *macroscopic* measure. The result is, within a sign, nothing other than Boltzmann's famous H *function*.

2.2. The entropy function H

Let us apply the preceding considerations to a macroscopic space made up of k different states: a macroscopic state is thus a vector $(f_1, \ldots, f_k)$ of frequencies with, of course, $f_1 + \cdots + f_k = 1$. It is supposed that the measure is absolute (no error) and that $Nf_j = N_j$ is entire for all j. The number of microscopic states associated with this macroscopic state then equals

$$W = \frac{N!}{N_1! \ldots N_k!}.$$

(If N_j positions are prepared in the jth state and if we number the positions from 1 to N, then there are $N!$ ways of arranging the N balls in the N positions and it is subsequently impossible to distinguish between permutations on the interior of any single box.)

According to Stirling's formula, when $N \to \infty$ we have $\log N! = N \log N - N + \log \sqrt{2\pi N} + o(1)$. It follows easily that

$$\begin{aligned} \frac{1}{N} \log W &= -\sum_i \frac{N_i}{N} \log \frac{N_i}{N} + O\left(\frac{k \log N}{N}\right) \\ &= -\sum f_i \log f_i + o(1). \end{aligned}$$

We note that we can also arrive at the same result without using Stirling's formula, thanks to the so-called method of types [41, Section 12.4].

If now we increase the number of experiments, we can formally make k tend to ∞, while making sure that k remains small compared to N. Let us suppose that we have at our disposal a reference measure ν on the macroscopic space $\mathcal{Y}$, and that we can separate this space into "cells" of volume (measure) $\delta > 0$, corresponding to the different states. When $\delta \to 0$, if the system has a statistical distribution $f(y)$ with respect to the measure ν, we can reasonably think that $f_i \simeq \delta\, f(y_i)$, where y_i is a representative point of cell number i. But then

$$\sum_i f_i \log \frac{f_i}{\delta} \simeq \delta \sum_i f(y_i) \log f(y_i) \simeq \int f \log f \, d\nu,$$

where the last approximation comes from the second sum being a Riemann sum of the integral.

We have ended up with **Boltzmann's H function**: being given a reference measure ν on a space $\mathcal{Y}$ and a probability measure μ on $\mathcal{Y}$,

$$H_\nu(\mu) = \int f \log f \, d\nu, \qquad f = \frac{d\mu}{d\nu}. \tag{5}$$

If ν is a probability measure, or more generally a measure of finite mass, it is easy to extend this formula to all probabilities μ by setting $H_\nu(\mu) = +\infty$ if μ is not absolutely continuous with respect to ν. If ν is a measure of infinite mass, more precautions must be taken; we could require at the very least the finiteness of $\int f\,(\log f)_- \, d\nu$.

We then note that if the macroscopic space $\mathcal{Y}$ bears a measure ν, then the microscopic space $\mathcal{Z} = \mathcal{Y}^N$ bears a natural measure $\nu^{\otimes N}$.

We are now ready to state the precise mathematical version of the formula for the function H: given a family $\{\varphi_j\}_{j\in\mathbb{N}}$ of bounded and uniformly continuous functions, then

$$\lim_{k\to\infty} \lim_{\varepsilon\to 0} \lim_{N\to\infty} \frac{1}{N} \log \nu^{\otimes N}\left[\left\{(y_1,\ldots,y_N) \in \mathcal{Y}^N; \quad \forall j \in \{1,\ldots,k\}, \right.\right.$$
$$\left.\left. \left| \int \varphi_j \, d\mu - \frac{1}{N}\sum_i \varphi_j(y_i) \right| \le \varepsilon \right\}\right] = -H_\nu(\mu). \tag{7}$$

We thus interpret N as the number of particles; the φ_j as a sequence of observables for which we measure the average value; and ε as the precision of the measurements. This formula summarizes in a concise manner the essential information contained in the function H.

If ν is a probability measure, statement (6) is known as **Sanov's theorem** [43] and is a leading result in the theory of large deviations. Before giving the interpretation of (6) in this theory, note that once we know how to treat the case where ν is a probability measure we easily deduce the case where ν is a measure of finite mass; however, I have no knowledge of any rigorous discussion in the case where ν is of infinite mass, even though we may expect that the result remains true.

2.3. Large deviations

Let ν be a probability measure and suppose that we independently draw random variables y_i according to ν. The empirical measure $\hat{\mu} = N^{-1}\sum \delta_{y_j}$ is then a random measure, almost certainly convergent to ν as $N \to \infty$ (it is Varadarajan's theorem, also called the fundamental law of statistics [49]). Of course it is possible that appearances deceive and that we think we are observing a measure μ distinct from ν. This probability decreases exponentially with N and is roughly proportional to $\exp(-N\,H_\nu(\mu))$; in other words, the Boltzmann entropy dictates the rarity of conditions that lead to the "unexpected" observation μ.

2.4. Information

Information theory was born in 1948 with the remarkable treatise of Shannon and Weaver [94] on the "theory of communication" which is now a pillar for the whole industry of information transmission.

In Shannon's theory, somewhat disembodied for its reproduction and impassionate discussion, the quantity of information carried by the decoding of a random signal is defined as a function of the reciprocal of the probability of the signal (which is rare and precious). Using the logarithm allows having the additivity property, and Shannon's formula for the mean quantity gained in the course of decoding is obtained: $\mathbb{E}\log(1/p(Y))$, where p is the law of Y. This of course gives Boltzmann's formula again!

2.5. Entropies on all floors

Entropy is not an intrinsic concept; it depends on the observer and the degree of knowledge that can be acquired through experiments and measures. The notion of entropy will consequently vary with the degree of precision of the description.

Boltzmann's entropy, as has been seen, informs us of the rarity of the kinetic distribution function $f(x,v)$ and the quantity of microscopic information remaining to be discovered once f is known.

If to the contrary we are given the microscopic state of all the microscopic particles, no hidden information remains and thus no more entropy. But if we are given a probability μ^N on the microscopic configurations, then the concept of entropy again has meaning: the entropy will be lower when the probability μ^N is concentrated and informative in itself. We thus find ourselves with a notion of **microscopic entropy**, $S_N = -H_N$,

$$H_N = \frac{1}{N}\int f^N \log f^N \, dx^N \, dv^N,$$

which is typically conserved by the Newtonian dynamic in consequence of Liouville's theorem. We can verify that

$$H_N \geq H(\mu^{1;N}),$$

with equality when μ^N is a tensor product and there are thus no correlations between particles. The idea is that the state of the microscopic particles is easier to obtain by multiparticle measurements than particle by particle – unless of course when the particles are independent!

In the other direction, we can also be given a *less precise* distribution than the kinetic distribution: this typically concerns a hydrodynamic description, which involves only the density field $\rho(x)$, the temperature $T(x)$ and mean velocity $u(x)$. The passage from the kinetic formalism to hydrodynamic formalism is accomplished by simple formulas:

$$\rho(x) = \int f(x,v)\,dv; \qquad u(x) = \rho(x)^{-1} \int f(x,v)\,v\,dv;$$
$$T(x) = \frac{1}{d\,\rho(x)} \int f(x,v)\,\frac{|v-u(x)|^2}{2}\,dv.$$

With this description is associated a notion of **hydrodynamic entropy**:

$$S_h = -\int \rho \log \frac{\rho}{T^{d/2}}.$$

This information is always lower than kinetic information. We have, finally, a hierarchy: first microscopic information at the low level, then "mesoscopic" information from the Boltzmann distribution function, finally "macroscopic information" contributed by the hydrodynamic description. The relative proportions of these different entropies constitute excellent means for appraising the physical state of the systems considered.

2.6. The universality of entropy

Initially introduced within the context of the kinetic theory of gases, entropy is an abstract and evolving mathematical concept, which plays an important role in numerous areas of physics, but also in branches of mathematics having nothing to do with physics, such as information theory and other sciences.

Some mathematical implications of the concept are reviewed in my survey *H-Theorem and beyond: Boltzmann's entropy in today's mathematics* [106].

3. Order and chaos

Intuitively, a microscopic system is ordered if all its particles are arranged in a coordinated, *correlated* way. On the other hand, it is chaotic if the particles, doing just as they please, act entirely independently from one another. Let us reformulate this idea: a distribution of particles is chaotic if each of the particles is oblivious to all the others, in the sense that a gain of information obtained for a given particle brings no gain in information about any other particle. This simple notion, key to Boltzmann's equation, presents some important subtleties that we will briefly mention.

3.1. Microscopic chaos

To say that random particles that are oblivious to each other is equivalent to saying that their joint law is tensorial. Of course, even if the particles are unaware of each other initially, they will enter into interaction right away and the independence property will be destroyed. In the case of hard spheres, the situation is still worse: the particles are obliged to consider each another since the spheres cannot interpenetrate. Their independence is thus to be understood asymptotically when the number of particles becomes very large; and experiments seeking to measure the degree of independence will involve but a finite number of particles. This leads naturally to the definition that follows.

Let $\mathcal{Y}$ be a macroscopic space and, for each N, let μ^N be a probability measure, assumed symmetric (invariant under coordinate permutations). We say that $\left(\mu^N\right)$ is chaotic if there exists a probability μ such that $\mu^N \simeq \mu^{\otimes N}$ in the sense of the weak topology of product measures. Explicitly, this means that for each $k \in \mathbb{N}$ and for all choices of the continuous functions $\varphi_1, \ldots, \varphi_k$ bounded on $\mathcal{Y}$, we have

$$\int_{\mathcal{Y}^N} \varphi_1(y_1) \ldots \varphi_k(y_k)\, \mu^N(dy_1 \, \ldots \, dy_N) \xrightarrow[N\to\infty]{} \left(\int \varphi_1 \, d\mu\right) \ldots \left(\int \varphi_k \, d\mu\right). \quad (7)$$

Of course, the definition can be quantified by introducing an adequate notion of distance, permitting us to measure the gap between μ^N and $\mu^{\otimes N}$. We can then say that a distribution μ^N is more or less chaotic. We again emphasize: what matters is the independence of a small number k of particles taken from among a large number N.

It can be shown (see the argument in [99]) that it is equivalent to impose property (7) for all $k \in \mathbb{N}$, or simply for $k = 2$. Thus chaos means precisely that 2 *particles drawn randomly from among* N *are asymptotically independent when* $N \to \infty$. The proof proceeds by observing the connections between chaos and empirical measure.

3.2. Chaos and empirical measure

By the law of large numbers, chaos automatically implies an asymptotic determinism: with very high probability, the empirical measure approaches the statistical distribution of an arbitrary particle when the total number of particles becomes gigantic.

It turns out that, conversely, *correlations accommodate very badly a macroscopic prescription of density.* Before giving a precise statement, we will illustrate this concept in a simple context. Consider a box with two compartments, in which we distribute a very large number N of *indistinguishable* balls. A highly correlated state would be a one in which all the particles occupy the same compartment: if I draw two balls at random, the state of first ball informs me completely about the state of the second. But of course, from the moment when the respective numbers of balls in the compartments are fixed and both nonzero, such a state of correlation is impossible. In fact, if the particles are indistinguishable, when two are drawn at random, the only information gotten is obtained by exploiting the fact that they are distinct, so that knowledge of the state of the first particle reduces slightly the number of possibilities for the state of the second. Thus, if the first particle occupies state 1, then the chances of finding the second particle in state 1 or 2 respectively are not $f_1 = N_1/N$ and $f_2 = N_2/N$, but $f_1' = (N_1 - 1)/(N - 1)$ and $f_2 = N_2/(N - 1)$. The joint distribution of a pair of particles is thus very close to the product law.

By developing the preceding argument, we arrive at an elementary but conceptually profound general result, whose proof can be found in Sznitman's course [99] (see also [40, p. 91]): *microscopic chaos is equivalent to the determinism of the empirical measure.* More precisely, the following statements are equivalent:

(i) $\left(\mu^N\right)$ is μ-chaotic;

(ii) the empirical measure $\widehat{\mu}^N$ associated with μ^N converges in law toward the deterministic measure μ.

By "empirical measure $\widehat{\mu}^N$ associated with μ^N" we understand the measure of the image of μ^N under the mapping $(y_1, \ldots, y_N) \longmapsto N^{-1} \sum \delta_{y_i}$, which is a measure of random probability. Convergence in law means that, for each continuous bounded function Φ on the space of probability measures, we have

$$\int \Phi\left(\frac{1}{N}\sum \delta_{y_i}\right) \mu^N(dy_1 \ldots dy_N) \xrightarrow[N\to\infty]{} \Phi(\mu).$$

In informal language, given a statistical quantity involving $\widehat{\mu}^N$, we can obtain an excellent approximation for large N by replacing, in the expression for this statistic, $\widehat{\mu}^N$ by μ.

The notion of chaos thus presented is weak and susceptible to numerous variants; the general idea being that μ^N must be close to $\mu^{\otimes N}$. The stronger concept of **entropic chaos** was introduced by Ben Arous and Zeitouni [13]: there $H_{\mu^{\otimes N}}(\mu^N) = o(N)$ is imposed. A related notion was developed by Carlen, Carvallo, Le Roux, Loss and Villani [32] in the case where the microscopic space is not a tensor product, but rather a sphere of large dimension; the measure $\mu^{\otimes N}$ is replaced by the restriction of the product measure to the sphere. Numerous other variants remain to be discovered.

3.3. The reign of chaos

In Boltzmann's theory, it is postulated that *chaos is the rule*: when a system is prepared, it is *a priori* in a chaotic state. Here are some possible arguments:

- if we can act on the macroscopic configuration, we will not have access to the microscopic structure and it is very difficult to impose correlations;
- the laws that underlie the microscopic variations are unknown to us and we may suppose that they involve a large number of factors destructive to correlations;
- the macroscopic measure observed in practice seems always well determined and not random;
- if we fix the macroscopic distribution, the entropy of a chaotic microscopic distribution is larger than the entropy of a nonchaotic microscopic distribution.

Let us explain the last argument. If we are given a probability μ on $\mathcal{Y}$, then the product probability $\mu^{\otimes N}$ is the maximum entropy among all the symmetric probabilities μ^N on $\mathcal{Y}^N$ having μ as marginal. In view of the large numbers N in play, this represents a phenomenally larger number of possibilities.

The microscopic measure μ_0^N can be considered as an object of Bayesian nature, an *a priori* probability on the space of possible observations. This choice, in general arbitrary, is made here in a canonical manner by maximization of the entropy: in some way we choose the distribution that leaves the most possibilities open and makes the observation the most likely. We thus join the scientific approach of maximum likelihood, which has proved its robustness and effectiveness – while skipping the traditional quarrel between frequentists and Bayesians!

The problem of the propagation of chaos consists of showing that our chaos hypothesis, made on the initial data (it is not entirely clear how), is propagated by the microscopic dynamic (which is well defined). The propagation of chaos is essential for two reasons: first, it shows that independence is asymptotically preserved, providing statistical information about the microscopic dynamic; secondly, it guarantees that *the statistical measure remains deterministic*, which allows hope for the possibility of a **macroscopic equation** governing the evolution of this empirical measure or its approximation.

3.4. Evolution of entropy

A recurrent theme in the study of dynamical systems, at least since Poincaré, is the search for invariant measures; the best-known example is Liouville measure for Hamiltonian systems. This measure possesses the remarkable additional property of tensorizing itself.

Suppose that we have a microscopic dynamic on $\mathcal{Y}^N$ and a measure ν on the space $\mathcal{Y}$ such that $\nu^{\otimes N}$ is an invariant measure for the microscopic dynamic; or more generally that there exists a ν-chaotic invariant measure on $\mathcal{Y}^N$. What happens with the preservation of microscopic volume in the limit $N \to \infty$?

A simple consequence of preservation of volume is conservation of macroscopic information $H_{\nu^{\otimes N}}(\mu_t^N)$, where μ_t^N is the image measure of μ_0^N through the microscopic evolution. In fact, since μ_t^N is preserved by the flow (by definition) and $\nu^{\otimes N}$ likewise, the density $f^N(t, y_1, \ldots, y_N)$ is constant along the trajectories of the system, and it follows that $\int f^N \log f^N \, d\nu^{\otimes N}$ is likewise constant.

Matters are more subtle for macroscopic information. Of course, if the various particles evolve independently from one another, the measure μ_t^N remains factored for all time, and we easily deduce that the macroscopic entropy remains constant. In general, the particles interact with one another, which destroys independence; however if there is propagation of chaos in a sufficiently strong sense, the independence is restored as $N \to \infty$, and we consequently have determinism for the empirical measure. So all the typical configurations for the microscopic initial measure μ_0^N give way, after a time t, to an empirical measure $\hat{\mu}_t^N \simeq \mu_t$, where μ_t is well determined. But it is possible that *other* microscopic configurations are compatible with the state μ_t, configurations that have not been obtained by evolution from typical initial configurations.

In other words: if we have a propagation of macroscopic determinism between the initial time and the time t, and if the microscopic dynamic preserves the reference measure produced, then we expect that the volume of the admissible microscopic states does not decrease between time 0 and time t. Keeping in mind the definition of entropy, we would have $e^{NS(t)} \geq e^{NS(0)}$, where $S(t)$ is the value of the entropy at time t. We thus expect that the entropy does not decrease over the course of the temporal evolution:

$$S(t) \geq S(0).$$

But then why not reverse the argument and say that chaos at time t implies chaos at time 0, by reversibility of the microscopic dynamic? This argument is in general inadmissible unless an exact notion of the chaos propagated is specified. The initial data prepared "at random" with just one kinetic distribution constraint, is supposed chaotic in a less strong sense; this depends on the microscopic evolution.

The notion of scale of interaction plays an important role here. Certain interactions take place on a macroscopic scale, other on a microscopic scale, which is to say that all or part of the interaction law is coded in parameters that are invisible on the macroscopic level. In this last case, the notion of chaos conducive to the propagation of the dynamic risks not being visible on the macroscopic scale and we can expect a degradation of the notion of chaos.

From there, the discussion must involve the details of the dynamic, and our worst troubles begin.

4. Chaotic equations

After the introduction of entropy and chaos, we can return to the Newtonian systems of Section 1, for which the phase space is composed of positions and velocities. A **kinetic equation** is an evolution equation bearing on the distribution $f(t, x, v)$; the important role of the velocity variable v justifies the terminology *kinetic*. By extension, in the case where there are external degrees of freedom (orientation of molecules for example), by extension we still speak of kinetic equations.

As descendants of Boltzmann, we pose the problem of deducing the macroscopic evolution starting from the underling microscopic model. This problem is in general of considerable difficulty. The fundamental equations are those of Vlasov, Boltzmann, Landau and Balescu–Lenard, published respectively in 1938, 1867, 1936 and 1960 (the more or less logical order of presentation of these equations does not entirely follow the order in which they were discovered…).

4.1. Vlasov's equation

Also called the Boltzmann equation without collisions, Vlasov's equation [112] is a mean field equation in the sense that all particles interact with one another (so each particle feels the mean contribution of the others). To deduce it from Newtonian dynamics, we begin by translating Newton's equation (1) as an equation in the empirical measure; for this we write the evolution equation of an arbitrary observable:

$$\begin{aligned}
&\frac{d}{dt}\frac{1}{N}\sum\nolimits_i \varphi(X_i(t),\dot{X}_i(t)) \\
&= \frac{1}{N}\sum\nolimits_i \left[\nabla_x\varphi(X_i,\dot{X}_i)\cdot\dot{X}_i + \nabla_v\varphi(X_i,\dot{X}_i)\cdot\ddot{X}_i\right] \\
&= \frac{1}{N}\sum\nolimits_i \left[\nabla_x\varphi(X_i,\dot{X}_i)\cdot\dot{X}_i + \nabla_v\varphi(X_i,\dot{X}_i)\cdot\left(a\sum\nolimits_j F(X_i-X_j)\right)\right].
\end{aligned}$$

This can be rewritten

$$\frac{\partial\widehat{\mu}^N}{\partial t} + \ v\cdot\nabla_x\ \hat{\mu}_t^N\ +\ aN\ \left(F*\hat{\mu}_t^N\right)\cdot\nabla_v\ \hat{\mu}_t^N\ = 0. \tag{8}$$

If now we suppose that $aN\simeq 1$ and we make the approximation

$$\hat{\mu}_t^N(dx\,dv)\simeq f(t,x,v)\,dx\,dv,$$

we obtain Vlasov's equation

$$\frac{\partial f}{\partial t} + \ v\cdot\nabla_x\ f\ +\ \left(\ F*_x\ \int f\,dv\right)\cdot\nabla_v\ f = 0. \tag{9}$$

We note well that $\hat{\mu}_t^N$ in (8) is a weak solution of Vlasov's equation, so that the passage to the limit is conceptually very simple: it is simply a stability result for the Vlasov equation.

Quite clearly I have gone a bit far, for this equation is nonlinear. If $\hat{\mu}\simeq f$ in the sense of the weak topology of measures, then $F*\hat{\mu}$ converges to $F*\int f\,dv$ in a topology determined by the regularity of F, and if this topology is weaker than uniform convergence, nothing guarantees that $(F*\hat{\mu})\hat{\mu}\simeq(F*f)f$.

If F is in fact bounded and uniformly continuous, then the above argument can be made rigorous. If F is furthermore L-Lipschitz, then we can do better and establish a stability estimate in weak topology: if (μ_t) and (μ_t'') are two weak solutions of Vlasov's equation, then

$$W_1\left(\mu_t,\mu_t'\right)\le e^{2\max(1,L)|t|}\,W_1\left(\mu_0,\mu_0'\right),$$

where W_1 is the Wasserstein distance of order 1,

$$W_1(\mu,\nu) = \sup\ :\left\{\int\varphi\,d\mu - \int\varphi\,d\nu;\quad \varphi\text{ 1-Lipschitz}\right\}.$$

Estimates of this sort are found in [95, Chapter 5] and date back to the 1970s (Dobrushin [48], Braun and Hepp [24], Neunzert [87]). Large deviation estimates can also be established as in [20].

However, for **singular interactions**, the problem of the Vlasov limit remains open, except for a result of Jabin and Hauray [64], which essentially assumes that

(a) $F(x-y) = O(|x-y|^{-s})$ with $0<s<1$; and

(b) the particles are initially well separated in phase space, so that

$$\inf_{j\neq i}\ \left(|X_i(0)-X_j(0)| + \left|\dot{X}_i(0)-\dot{V}_j(0)\right|\right)\ge\frac{c}{N^{\frac{1}{2d}}}.$$

Neither of these conditions is satisfied: the first lacks the Coulomb case of singularity order, while the second excludes the case of random data. However, it remains the sole result available at this time... To go further, it would be nice to have a sufficiently strong notion of chaos so as to be able to control the number of pairs (i, j) such that $|X_i(t) - X_j(t)|$ is small. In the absence of such controls, Vlasov's equation for singular interactions remains an act of faith.

This act of faith is very effective since the Vlasov–Poisson model, in which $F = -\nabla W$, where W a fundamental solution of $\pm\Delta$, is the universally accepted classic model in plasma physics [42, 71] as well as in astronomy [15]. In the first instance a particle is an electron, in the second a star! The only difference lies in the sign: repulsive interaction for electrons, attractive for stars. We should not be astonished to see stars considered in this way as microscopic objects: they are effectively so on the scale of a galaxy (which can tally 10^{12} stars...).

The theory of the Vlasov–Poisson equation itself remains incomplete. We can distinguish presently two principal theories, both developed in the entire space. That of Pfaffelmoser, simplified by Schaeffer and exposited for example in [51], supposes that the initial data f_i is C^1 with compact support; later this unsatisfactory compactness assumption was removed by Horst [61] by an improvement of the Pfaffelmoser–Schaeffer method. The concurrent theory is that Lions–Perthame, reviewed in [23]. Pfaffelmoser's theory has been adapted in spatially periodic context (see [12] or modify [61]), which is not the case for the Lions–Perthame theory.

4.2. Boltzmann's equation

Vlasov's equation loses its relevance when the interactions have a short range. A typical example is that of rarefied gas, for which the dominant interactions are binary and are uniquely produced in the course of "collisions" between particles.

Boltzmann's equation was established by Maxwell [80] and Boltzmann [21, 22]; it describes a situation where the interactions are of short range and where each particle undergoes $O(1)$ impacts per unit of time. Much more subtle than the situation of Vlasov's mean field, the Boltzmann situation is nonetheless simpler than the hydrodynamic one where the particles undergo a very large number of collisions per unit of time.

We start by establishing the equation informally. The movement of a particle occurs with alternation of rectilinear trajectories and collisions, during the course of which its velocity changes so abruptly that we can consider the event as instantaneous and localized in space. We first consider the emblematic case of hard spheres of radius r: a collision occurs when two particles, with respective positions x and y and with respective velocities v and w, are found in a configuration where $|x-y| = 2r$ and $(w-v)\cdot(y-x) < 0$. We then speak of a *precollisional configuration.* We let $\omega = (y - x)/|x - y|$.

We now come to the central point in all Boltzmann's argument: *when two particles encounter each other, with very high probability they will* (*almost*) *not be correlated*: think of two people who encounter each other for the first time. We can consequently apply the hypothesis of molecular chaos to such particles, and we find that the probability of an encounter between these particles is proportional to

$$f^{2;N}(t,x,v,x+2r\omega,w) \simeq f^{1;N}(t,x,v)\, f^{1;N}(t,x+2r\omega,w) \\ \simeq f^{1;N}(t,x,v)\, f^{1;N}(t,x,w),$$

provided thus that $(w-v)\cdot\omega < 0$. We likewise need to take into account the relative velocities in order to evaluate the influence of the particles of velocity w on the particles of velocity v : the probability of encountering a particle of velocity w in a unit of time is proportional to the product of $|v-w|$ by the effective section (in dimension 3 this is the apparent area of the particles, or πr^2) and by the cosine of the angle between $v-w$ and ω (the extreme case is where $v-w$ is orthogonal to ω, which is to say that the two particles but graze each other, clearly an event of probability zero. Each of these collisions removes a particle of velocity v, and we thus have a negative term, the *loss term*, proportional to

$$-\iint f(t,x,v)\, f(t,x,v_*)\, |(v-v_*)\cdot\omega|\, dv_*\, d\omega.$$

The velocities after the collision are easily calculated:

$$v' = v - (v-v_*)\cdot\omega\,\omega; \quad v'_* = v_* + (v-v_*)\cdot\omega\,\omega \tag{10}$$

These velocities do not matter for the final analysis.

However, we also need to take account of all the particles of velocity v that have been created by collisions between particles of arbitrary velocities. By microscopic reversibility, these velocities are of the form (v',v'_*), and our problem is to take account of all the possible pairs (v',v'_*), which in this problem of computing the *gain term* are the *pre-collisional* velocities. We thus again apply the hypothesis of pre-collisional chaos and obtain finally the expression of the **Boltzmann equation for solid spheres**:

$$\frac{\partial f}{\partial t} + v\cdot\nabla_x f = Q(f,f), \tag{11}$$

where

$$Q(f,f)(t,x,v) \\ = B\int_{\mathbb{R}^3}\int_{S^2_-} |(v-v_*)\cdot\omega|\, \Big(f(t,x,v')\, f(t,x,v'_*) - f(t,x,v)\, f(t,x,v_*)\Big)\, dv_*\, d\omega,$$

Here S^2_- denotes the pre-collisional configurations $\omega\cdot(v-v_*) < 0$, and B is a constant. By using the change of variable $\omega \to -\omega$ we can symmetrize this expression and arrive at the final expression (after changing the value of B)

$$Q(f,f)(t,x,v) \\ = B\int_{\mathbb{R}^3}\int_{S^2} |(v-v_*)\cdot\omega|\, \Big(f(t,x,v')\, f(t,x,v'_*) - f(t,x,v)\, f(t,x,v_*)\Big)\, dv_*\, d\omega. \tag{12}$$

The operator (12) is the **Boltzmann collision operator for hard spheres**. The problem now consists of justifying this approximation.

To do this, in the 1960s Grad proposed a precise mathematical limit: have r tend toward 0 and at the same time N toward infinity, so that $Nr^2 \to 1$, which is to say that the total effective section remains constant. Thus a given particle, moving among all the others, will typically encounter a finite number of them in a unit of time. One next starts with a microscopic probability density $f_0^N(x^N, v^N)\, dx^N\, dv^N$, which is allowed to evolve by the Newtonian flow $\mathcal{N}_t$, and one attempts to show that the first marginal $f^{1;N}(t, x, v)$ (obtained by integrating all the variables except the first position variable and the first velocity variable) converges in the limit to a solution of the Boltzmann equation.

The Boltzmann-Grad limit is also often called the **low density limit** [40, p. 60]: in fact, if we start from the Newtonian dynamic and fix the particle size, we will dilate the spatial scale by a factor $1/\sqrt{N}$ and the density will be of the order $N/N^{3/2} = N^{-1/2}$.

At the beginning of the 1970s, Cercgnani [37] showed that Grad's program could be completed if one proved a number of *plausible* estimates; shortly thereafter, independently, Lanford [69] sketched the proof of the desired result.

Lanford's theorem is perhaps the single most important mathematical result in kinetic theory. In this theorem, we are given microscopic densities f_0^N such that for each k the densities $f_0^{k;N}$ of the k particle marginals are continuous, satisfy Gaussian bounds at large velocities and converge uniformly outside the collisional configurations (those where the positions of two distinct particles coincide) to their limit $f_0^{\otimes k}$. The conclusion is that there exists a time $t_* > 0$ such that $f_t^{k;N}$ converges almost everywhere to $f_t^{\otimes k}$, where f_t is a solutions of Boltzmann's equation, for all time $t \in [0, t_*]$.

Lanford's estimates were later rewritten by Spohn [95] and by Illner and Pulvirenti [61, 62] who replaced the hypothesis of small time by a smallness hypothesis on the initial data, permitting Boltzmann's equation to be treated as a perturbation of free transport. These results are reviewed in [40, 90, 95].

The technique used by Lanford and his successors goes through the **BBGKY hierarchy** (Bogoliubov–Born–Green–Kirkwood–Yvon), the method by which the evolution of the marginal for a particle $f^{1;N}$ is expressed as a function of the marginal for two particles $f^{2;N}$; the evolution of a two-particle marginal $f^{2;N}$ as a function of a three-particle marginal $f^{3;N}$, and so forth. This procedure is especially uneconomical (in the preceding heuristic argument, we only use $f^{1;N}$ and $f^{2;N}$, but there is no known alternative).

Each of the equations of the hierarchy is then solved via Duhamel's formula, applying successively the free transport and collision operators, and by summing over all the possible collisional history. The solution at time t is thus formally expressed, as with an exponential operator, as a function of the initial data and we can apply the chaos hypothesis on (f_0^N).

We then pass to the limit $N \to \infty$ in each of the equations, after having verified that we can neglect pathological "recollisions", where a particle again encounters a particle that it had already encountered beforehand, and which is thus not unknown to it. This point is subtle: in [40, appendix 4.C] a dynamic that is *a priori* simpler than that of solid spheres, due to Uchiyama, is discussed, with only four velocities in the plane, for which the recollision configurations cannot be neglected, and the kinetic limit does not exist.

It remains to identify the result with the series of tensor products of the solution to Boltzmann's equation and conclude by using a uniqueness result.

Spohn [95, Section 4.6] shows that one can give more precise information on the microscopic distribution of the particles: on the small scale, this follows a homogeneous **Poisson law** in phase space. This is consistent with the intuitive idea of molecular chaos.

Lanford's theorem settled a controversy that had lasted since Boltzmann himself; but it leaves numerous questions in suspense. In the first place, it is limited to a small time interval (on which only about 1/6 of the particles have had time to collide... but the conceptual impact of the theorem is nonetheless important). The variant of Illner and Pulvirenti lifts this restriction of small time, but the proof does not lend itself to a bounded geometry. As for lifting the smallness restriction, at the moment it is but a distant dream.

Next, to this day the theorem has only been proved for a system of solid spheres; long-range interactions are not covered. Cercignani [36] notes that the limit of Boltzmann-Grad for such interactions poses subtle problems, even from the formal viewpoint.

Finally, the most frustrating thing is that Lanford avoided discussion of **pre-collisional chaos**, the notion that particles that are about to collide are not correlated. This notion is very subtle, because just after the collision, correlations have inevitably taken place. In other words, we have *pre-collisional chaos, but not post-collisional.*

What does pre-collisional chaos mean exactly? For the moment we do not have a precise definition. It is certainly a stronger notion than chaos in the usual sense; it involves too a de-correlation hypothesis that is seen on a set of codimension 1, i.e., configurations leading to collisions. We would infer that it is a notion of chaos where we have replaced the weak topology by a uniform topology; but that cannot be so simple, since chaos in a uniform topology also implies post-collisional chaos, which is incompatible with pre-collisional chaos! In fact, the continuity of the two-particle marginal along a collision would imply

$$\begin{aligned} f(t,x,v)\,f(t,x,v_*) &\simeq f^{(2;N)}(t,x,v,x+2r\omega,v_*) \\ &= f^{(2;N)}(t,x,v',x+2r\omega,v'_*) \simeq f^{(1;N)}(t,x,v')f^{(1;N)}(t,x,v'_*). \end{aligned}$$

Passing to the limit we would have

$$f(t,x,v')\,f(t,x,v'_*) = f(t,x,v)\,f(t,x,v_*),$$

and as we will see in Section 5.3 this implies that f is Gaussian in the velocity variable, which is of course false in general. Another argument for showing that post-collisional chaos must be incompatible with pre-collisional chaos consists of noting that if we have post-collisional chaos, the reasoning leading to the Boltzmann equation can be used again by expressing two-particle probabilities in terms of post-collisional probabilities... and we then obtain Boltzmann's equation in reverse:

$$\frac{\partial f}{\partial t} + v \cdot \nabla_x f = -Q(f, f).$$

As has been mentioned, Lanford's proof applies only to solid spheres; but Boltzmann's equation is used for a much larger range of interactions. The general form of the equation, say in dimension d, is the same as in (11):

$$\frac{\partial f}{\partial t} + v \cdot \nabla_x f = Q(f, f), \tag{13}$$

but now

$$Q(f, f) = \int_{\mathbb{R}^d} \int_{S^{d-1}} (f' f'_* - f f_*) \, \tilde{B}(v - v_*, \omega) \, dv_* d\omega \tag{14}$$

where $\tilde{B}(v - v_*, \omega)$ depends only on $|v - v_*|$ and $|(v - v_*) \cdot \omega|$. There exist several representations of this integral operator (see [103]); it is often convenient to change variables by introducing another angle, $\sigma = (v' - v'_*)/|v - v_*|$, so that the formulas (10) must be replaced by

$$v' = \frac{v + v_*}{2} + \frac{|v - v_*|}{2} \sigma, \qquad v'_* = \frac{v + v_*}{2} - \frac{|v - v_*|}{2} \sigma. \tag{15}$$

We must then replace the collision kernel $\tilde{B}$ by B so that

$$B \, d\sigma = \tilde{B} \, d\omega.$$

Explicitly, we find

$$\frac{1}{2} \tilde{B}(z, \omega) = \left| 2 \, \frac{z}{|z|} \cdot \omega \right|^{d-2} B(z, \sigma).$$

The precise form of B (or, in an equivalent way, of $\tilde{B}$) is obtained by a classical scattering computation that goes back to Maxwell and which can be found in [38]: for an impact parameter $p \geq 0$ and a relative velocity $z \in \mathbb{R}^3$, the deviation angle θ equals

$$\theta(p, z) = \pi - 2p \int_{s_0}^{+\infty} \frac{ds/s^2}{\sqrt{1 - \frac{p^2}{s^2} - 4\frac{\phi(s)}{|z|^2}}} = \pi - 2 \int_0^{\frac{p}{s_0}} \frac{du}{\sqrt{1 - u^2 - \frac{4}{|z|^2} \phi\left(\frac{p}{u}\right)}},$$

where s_0 is the positive root of

$$1 - \frac{p^2}{s_0^2} - 4\frac{\phi(s_0)}{|z|^2} = 0.$$

So B is implicitly defined by

$$B\left(|z|, \cos\theta\right) = \frac{p}{\sin\theta} \frac{dp}{d\theta} |z|. \tag{16}$$

We write either $B(|z|, \cos\theta)$ or $B(z, \sigma)$, it being understood that the deviation angle θ is the angle formed by the vectors $v - v_*$ and $v' - v'_*$.

When $\phi(r) = 1/r$, we recover Rutherford's formula for the Coulomb deviation:

$$B\left(|v - v_*|, \cos\theta\right) = \frac{1}{|v - v_*|^3 \sin^4(\theta/2)}.$$

When $\phi(r) = 1/r^{s-1}$, $s > 2$, the collision kernel is not computed explicitly, but is can be shown that (always in dimension 3)

$$B\left(|v - v_*|, \cos\theta\right) = b(\cos\theta)\,|v - v_*|^\gamma, \qquad \gamma = \frac{s-5}{s-1}. \tag{17}$$

Furthermore, the function b, defined implicitly, is locally smooth with a *non integrable angular singularity* when $\theta \to 0$:

$$\sin\theta\, b(\cos\theta) \sim K\theta^{-1-\nu}, \qquad \nu = \frac{2}{s-1}. \tag{18}$$

This singularity corresponds to collisions with large impact parameter p, where there is scant deflection. It is inevitable once the forces are of infinite range: in fact

$$\int_0^\pi B\left(|z|, \cos\theta\right) \sin\theta\, d\theta = |z| \int_0^\pi p\,\frac{dp}{d\theta}\, d\theta = |z| \int_0^{p_{\max}} p\, dp = \frac{|z|\, p_{\max}^2}{2}. \tag{19}$$

In the particular case $s = 5$, the collision kernel depends no longer on the relative velocity, but only on the *deviation angle*: we speak of Maxwellian molecules. By extension, we say that $B(v - v_*, \sigma)$ is a Maxwellian collision kernel if it depends only on the angle between $v - v_*$ and σ. The Maxwellian molecules are above all a phenomenological model, even if the interaction between a charged ion and a neutral particle in a plasma is regulated by such a law [42, Vol. 1, p. 149]. The potentials in $1/r^{s-1}$ for $s > 5$ are called hard potentials, for $s < 5$ soft potentials. Often the angular singularity $b(\cos\theta)$ is truncated to small values of θ.

The Boltzmann equation is important in modeling rarefied gases, as explained in [39]. Nonetheless, because of its eventful history and its conceptual content, as well as the impact of Boltzmann's treatise [22], this equation has exerted a fascination that goes far beyond its usefulness. The first mathematical works dedicated to it are those of Carleman[2] [26, 27], followed by Grad [57]. Besides the article by Lanford [69] already mentioned, a result that has had a great impact is the weak stability theorem of DiPerna–Lions [47]. The equation is well understood in the spatially homogeneous setting for hard potentials with angular truncation, see, e.g., [84]; and in the setting close to equilibrium, see, e.g., [60]. We refer to the reference treatises [38, 40, 103] for a number of other results.

4.3. Landau's equation

Boltzmann's collisional integral loses its meaning for Coulomb interactions because of the extremely slow decrease of the Coulomb potential. The grazing collisions, with large impact parameter, then become dominant.

[2] The monography [27] was incomplete at the time of Carleman's passing away, and was completed by Carleson.

In 1936, Landau [67] established, using formal arguments, an asymptotic of Boltzmann's kernel in this setting. Letting λ_D be the shielding distance (below which the Coulomb potential is no longer visible because of the global neutrality of the plasma), and r_0 the typical collision distance (distance of two particles whose interaction energy is comparable to the molecular excitation energy), the parameter $\Lambda = 2\lambda_D/r_0$ is the plasma parameter, and in the limit $\Lambda \to \infty$ (justified for "classical" plasmas), the Boltzmann operator can be formally replaced by a diffusive operator called Landau's operator:

$$Q_B(f,f) \simeq \frac{\log \Lambda}{2\pi\Lambda}\; Q_L(f,f), \tag{20}$$

$$Q_L(f,f) = \nabla_v \cdot \left(\int_{\mathbb{R}^3} a(v-v_*)\; \left[f(v_*)\,\nabla_v\, f(v) \;-\; f(v)\,\nabla_v\, f(v_*)\right] dv_* \right), \tag{21}$$

$$a(v-v_*) = \frac{L}{|v-v_*|}\; \Pi_{(v-v_*)^\perp}, \tag{22}$$

where L is a constant and $\Pi_{z^\perp}$ denotes orthogonal projection onto $z^\perp$.

The Landau approximation is now well understood mathematically in the context of a limit called **grazing collision asymptotics**; [3] can be consulted for a detailed discussion of this problem.

The Landau operator, both diffusive and integral, presents a remarkable structure. It is easily generalized to arbitrary dimensions $d \geq 2$, and the coefficient $L/|z|$ can be changed to $L\,|z|^{\gamma+2}$, where γ is the exponent appearing in (17). The models of hard potential type with $\gamma > 0$ have been completely studied in the spatially homogeneous case [45]; but it is definitely the case $\gamma = -3$ in dimension 3 that is physically interesting. In this case we only know how to prove the existence of weak solutions in the spatially homogeneous case (by adapting [1, Section 7] and the existence of strong solutions for perturbations of equilibrium [59]. This situation is entirely unsatisfactory.

4.4. The Balescu-Lenard equation

In 1960, Balescu [7] directly establishes a kinetic equation that describes the Coulomb interactions in a plasma; he thus recovers an equation published in another form by Bogoliubov [19] and simplified by Lenard. The reference [96] can be consulted for information on the genesis of the equation, and [8] for a synthetic presentation. The collision kernel in this equation takes the same form as (21), the difference is in the expression of the matrix $a(v-v_*)$, which now depends both on v and ∇f:

$$a_{BL}\big(v, v-v_*, \nabla f\big) = B \int_{|k|\leq K_0} \frac{k\otimes k}{|k|^4}\; \frac{\delta_{k\cdot(v-v_*)}}{\big|\epsilon(k, k\cdot v, \nabla f)\big|^2}\, dk, \tag{23}$$

$$\epsilon\big(k, k\cdot v, \nabla f\big) = 1 + \frac{b}{|k|^2} \int_{\mathbb{R}^3} \frac{k\cdot\nabla f(u)}{k\cdot(v-u) - i\,0}\, du.$$

This equation can also be obtained beginning with the study of long duration fluctuations in Vlasov's equation [71, Section 51].

The Balescu–Lenard equation is scarcely used because of its complexity. Under reasonable hypotheses, the Landau equation provides a good approximation [8, 70]. The procedure is adaptable interactions other than the Coulomb interaction, but in contrast with the limit of grazing collisions, it still provides the expression (21), the only change being in the coefficient L of (22), which is proportional to

$$\int_{\mathbb{R}^3} |k|\,|\hat{W}(k)|^2\,dk,$$

where W is the interaction potential. This equation is briefly reviewed in [95, Chapter 6].

The mathematical theory of the Balescu–Lenard equation is wide open, both with regard to establishing it and to studying its qualitative properties; one of the rare rigorous papers on the subject is the linearized study of R. Strain [96]. Even though little used, the Balescu–Lenard equation is nonetheless the most respected of the collisional models in plasmas and it is an intermediary that allows justification for using the Landau collision operator to represent long duration fluctuations in particle systems; its theory represents a formidable challenge.

5. Boltzmann's theorem *H*

In this section we will start with Boltzmann's equation and examine several of its most striking properties. Much more information can be found in my long review article [103].

5.1. Modification of observables by collisions

The statistical properties of a gas are manifested, in the kinetic model, by the evolution of observables $\iint f(t,x,v)\,\varphi(x,v)\,dx\,dv$. Still assuming conditions with periodic limits and all the required regularity, we may write

$$\begin{aligned}\frac{d}{dt}\iint f\,\varphi\,dx\,dv &= \iint (\partial_t f)\,\varphi\,dx\,dv \qquad (24)\\ &= -\iint v\cdot\nabla_x f\,\varphi\,dx\,dv + \iint Q(f,f)\,\varphi\,dx\,dv\\ &= \iint v\cdot\nabla_x\varphi\,f\,dx\,dv + \iiiint \tilde{B}\,\left(f'f'_* - ff_*\right)\,\varphi\,dx\,dv\,dv_*\,d\omega,\end{aligned}$$

where we are still using the notation $f' = f(t,x,v')$, etc.

In the term with the integral in $f'f'_*$ we now make the pre-postcollisional change of variables $(v,v_*) \longrightarrow (v',v'_*)$, for all $\omega \in S^{d-1}$. This change of variable is unitary (Jacobian determinant equal to 1) and preserves $\tilde{B}$ (its properties are traces of the microreversibility). After having renamed the variables, we obtain

$$\frac{d}{dt}\iint f\,\varphi\,dx\,dv = \iint v\cdot\nabla_x\varphi\,f\,dx\,dv + \iiiint \tilde{B}\,f\,f_*\,\left(\varphi\prime - \varphi\right)dv\,dv_*\,d\omega\,dx. \quad (25)$$

This is, incidentally, the form in which Maxwell wrote Boltzmann's equation from 1867 on... We deduce from (25) that $\iint f\,dx\,dv$ is constant (fortunately!!), and we get an important quantity, the effective momentum transfer cross section $M(v-v_*)$ defined by

$$M(v-v_*)\,(v-v_*) = \int \tilde{B}(v-v_*,\omega)\,(v'-v)\,d\omega.$$

Even when $\tilde{B}$ is a divergent integral, the quantity M may be finite, expressing the fact that the collisions modify the velocities in a statistically reasonable way. Readers may refer to [2, 3, 103] for more details on the treatment of grazing singularities of $\tilde{B}$.

Boltzmann would improve Maxwell's procedure by making better use of the symmetries of the equation. First, by making the pre-postcollisional change of variables in the whole second term of (24) we obtain

$$\iiiint \tilde{B}\,(f'f'_* - f\,f_*)\ \varphi\,dv\,dv_*\,d\omega\,dx = -\iiiint \tilde{B}\,(f'f'_* - f\,f_*)\ \varphi'\,dv\,dv_*\,d\omega\,dx. \tag{26}$$

Instead of exchanging the pre- and postcollisional configurations, we may exchange the particles together: $(v, v_*) \longmapsto (v_*, v)$, which also clearly has a unitary Jacobian determinant. This gives us two new forms from (26):

$$\iiiint \tilde{B}\,(f'f'_* - f\,f_*)\ \varphi_*\,dv\,dv_*\,d\omega\,dx = -\iiiint \tilde{B}\,(f'f'_* - f\,f_*)\ \varphi'_*\,dv\,dv_*\,d\omega\,dx. \tag{27}$$

By combining the four forms appearing in (26) and (27), we obtain

$$\begin{aligned}\frac{d}{dt}\iint f\,\varphi\,dv\,dx = &\iint f\,(v\cdot\nabla_x\varphi)\ dx\,dv \qquad (28)\\ &-\frac{1}{4}\iiiint \tilde{B}\ (f'f'_* - ff_*)\ (\varphi' + \varphi'_* - \varphi - \varphi_*)\ dx\,dv\,dv_*\,d\omega.\end{aligned}$$

As a consequence of (28), we note in the first place that $\iint f\,\varphi$ is preserved if φ satisfies the functional equation

$$\varphi(v') + \varphi(v'_*) = \varphi(v) + \varphi(v_*) \tag{29}$$

for each choice of velocities v, v_* and of the parameter ω. Such functions are called **collision invariants** and reduce, under extremely weak hypotheses, to just linear combinations of the functions

$$1, \quad v_j\ (1 \le j \le d), \quad \frac{|v|^2}{2}.$$

Readers may consult [40] in this regard. This is again natural: it is the macroscopic reflection of conservation of mass, the amount of motion and kinetic energy during microscopic interactions.

5.2. Theorem *H*

We now come to the discovery that will put Boltzmann among the greatest names in physics. We choose $\varphi = \log f$ and assume all the regularity needed for carrying through the calculations; in particular

$$\iint f\,v \cdot \nabla_x(\log f)\,dv\,dx = \iint v \cdot \nabla_x\,(f \log f - f)\;dv\,dx = 0.$$

Identity (28) thus becomes, taking into account the additive properties of the logarithm,

$$\frac{d}{dt}\iint f\,\log f\,dx\,dv = -\frac{1}{4}\iint\iint \tilde{B}\;(f\prime f\prime_* - f f_*)\;(\log f\prime f\prime_* - \log f f_*)\,. \qquad (30)$$

The logarithm function being increasing, the above expression is always nonpositive! Moreover, knowing that $\tilde{B}$ vanishes only on a set of measure zero, we see that the expression (30) is strictly negative whenever $f' f'_*$ is not equal to $f f_*$ almost everywhere, which is true for generic distributions. Thus, modulo the rigorous justification of the integrations by parts and a change of variables, we have proved that, *in Boltzmann's model, the entropy increases with time.*

The impact of this result is crucial. First, the heuristic microscopic reasoning of Section 3.4 has been replaced by a simple argument that leads directly to the limit equation. Next, even if it is a manifestation of the **second law of thermodynamics**, the increase in the entropy in Boltzmann's model is deduced by *logical reasoning* and not by a postulate (a law) which one accepts or not. Finally, of course, in doing so, Boltzmann displayed an **arrow of time** associated with his equation.

Not only is this macroscopic irreversibility not contradictory with microscopic reversibility, but it is in fact intimately linked to it: as has already been explained, it is the conservation of microscopic volume in phase space that guarantees the nondecrease of entropy. For the rest, as L. Carleson was already surprised to discover in 1979 while examining simplified models of Boltzmann's equation [35], it is precisely when the parameters of the dynamics are adjusted in such a way to achieve microscopic reversibility, that the H theorem holds. The phenomenon is well known in the context of the physics of granular media [105]: there the microscopic dynamic is dissipative (nonreversible), including a loss of energy due to friction, and the macroscopic dynamic does not satisfy Theorem H!

From the informational point of view, the increase in entropy means that the system always runs toward macroscopic states that are more and more probable. This probabilistic idea is exacerbated by the formidable power of the combinatorics: we suppose for example that we are considering a gas with $N \simeq 10^{16}$ particles (which is roughly what we find in 1 mm^3 of gas under ordinary conditions!), and between time $t = t_1$ and time $t = t_2$ the entropy increases only by 10^{-5}. The volume of microscopic possibilities is then multiplied by $e^{N\,[S(t_2)-S(t_1)]} = e^{10^{11}} \gg 10^{10^{10}}$. This phenomenal factor far exceeds the number of protons in the universe (10^{100}?) or the number of 1000-page books that could be written by combining all the alphabetic characters of all the languages in the world…

The intuitive interpretation of Theorem H is thus rather eloquent: the high entropy states occupy, at the microscopic level, a place so monstrously larger than the states of low entropy, that the microscopic system goes to them automatically. As we have seen, the logical reasoning justifying this scenario is complex and indirect, involving the propagation of chaos and macroscopic determinism – and to this day only a small portion of the program has been rigorously achieved.

5.3. Vanishing of entropy production

The increase in entropy admits a complement that is no less profound, frequently stated as a second part of Theorem H: the *characterization of cases of equality*, i.e., states for which the production of entropy vanishes.

We have seen in (30) that the entropy production equals

$$\int \mathrm{PE}\,(f(x,\,\cdot))\,dx, \tag{31}$$

where PE is the functional of "local production of entropy", acting on the kinetic distributions $f = f(v)$:

$$\mathrm{PE}\,(f) = \iiint \tilde{B}(v - v_*, \omega)\ (f(v')f(v'_*) - f(v)f(v_*))\ \log \frac{f(v')f(v'_*)}{f(v)f(v_*)}\, dv\, dv_*\, d\omega. \tag{32}$$

For all reasonable models, we have $\tilde{B}(z, \omega) > 0$ almost everywhere, and it follows that the entropy production vanishes only for a distribution satisfying the functional equation

$$f(v')f(v'_*) = f(v)f(v_*) \tag{33}$$

for (almost) all v, v_*, ω. By taking the logarithm in (33) we recover equation (29), which shows that f must be the exponential of a collision invariant. In view of the form of the latter, and taking into account the integrability constraint of f, we obtain $f(v) = e^{a+b\cdot v+c|v|^2/2}$, which can be rewritten

$$f(v) = \rho\,\frac{1}{(2\pi T)^{d/2}}\,\exp\left(-\frac{|v-u|^2}{2T}\right), \tag{34}$$

where $\rho \geq 0$, $T > 0$ and $u \in \mathbb{R}^d$ are constants. It is therefore a particular Gaussian, with covariance matrix proportional to the identity.

Maxwell already noticed that (34) makes Boltzmann's collision operator vanish: $Q(f, f) = 0$. Such a distribution is called **Maxwellian** in his honor. However, it is Boltzmann who first gave a convincing argument that the distributions (34) are *the only* solutions of the equation $\mathrm{PE}\,(f) = 0$, and consequently the only solutions of $Q(f, f) = 0$. Let's honor him by sketching a variant of his original proof.

We begin by averaging (33) over all angles $\sigma = (v' - v'_*)/|v - v_*|$; the left side $|S^{d-1}|^{-1} \int f' f'_*\, d\sigma$ is then the mean of the function $\sigma \to f(c + r\sigma)\, f(c - r\sigma)$, where $c = (v + v_*)/2$ and $r = |v - v_*|$. This thus depends only on c and r or, in an equivalent way, only on $m = v + v_*$ and $e = (|v|^2 + |v_*|^2)/2$, respectively the amount of motion and the total energy involved in a collision. After passing to the logarithm, we find for $\varphi = \log f$ the identity

$$\varphi(v) + \varphi(v_*) = G(m, e). \tag{35}$$

The operator $\nabla_v - \nabla_{v_*}$, applied to the left-hand side of (35), yields $\nabla\varphi(v) - \nabla\varphi(v_*)$. When we apply the same operator to the right-hand side, the contribution of m disappears, and the contribution of e is collinear with $v - v_*$. We thus conclude that $F = \nabla\varphi$ satisfies

$$F(v) - F(v_*) \text{ is collinear with } v - v_* \text{ for all } v, v_*$$

and it is easy to deduce that $F(v)$ is an affine transformation, whence the conclusion. (Here is a crude method for showing the affine character of F, assuming regularity: we start by writing a Taylor expansion and noting that the Jacobian matrix of F is a multiple of the identity at each point, say $\partial_i F^j(v) z_i = \lambda(v)\, z_j$; then by differentiating with respect to v_k we deduce that $\partial_{ik} F^j = 0$ if $i \neq j$, and it follows that all the coefficients $\partial_i F^j$ cancel, after which we easily see that DF is a multiple of the identity.)

As a consequence of (31) and (34), the distributions $f(x, v)$ that cancel the production of Boltzmann entropy are precisely the distributions of the form

$$f(x, v) = \rho(x)\ \frac{1}{(2\pi T(x))^{d/2}} \exp\left(-\frac{|v - u(x)|^2}{2T(x)}\right). \tag{36}$$

They are called **local Maxwellians** or else **hydrodynamic states**. In accordance with the kinetic description, these states are characterized by a considerable reduction in complexity, since they depend on but three fields: the density field ρ, the field of macroscopic velocities u and the temperature field T. These are the fields that enter into the hydrodynamic equations, whence the above terminology.

This discovery establishes a bridge between the kinetic and hydrodynamic descriptions: in a process where collisions are very numerous (weak Knudsen number), the finiteness of entropy production forces the dynamic to be concentrated near distributions that makes the entropy production vanish. This remark makes way for a vast program of hydrodynamic approximation of Boltzmann's equation, to which Hilbert alludes in his Sixth Problem. Readers can consult [54, 55, 93]. If the Boltzmann equation can be approached both by compressible and incompressible equations, we should note that it does not lead to the whole range of hydrodynamic equations, but only to those for perfect gases, i.e., those that conform to a law where pressure is proportional to ρT.

6. Entropic convergence: forced march to oblivion

If Maxwell discovered the importance of Gaussian velocity profiles, he did not, as Boltzmann remarks, prove that these profiles are actually induced by the dynamic. Boltzmann wanted to complete this program, and for this recover the Maxwellian profiles not only as equilibrium distributions, but also as limits of the kinetic equation asymptotically as time becomes large ($t \to \infty$). This conceptual leap aimed at basing equilibrium statistical mechanics on its nonequilibrium counterpart – usually much more delicate – is still topical in innumerable contexts.

I have written a good bit on this topic and readers can consult the survey article [103, Chapter 3], the course [108], the research article [46] or the research memoir [109]. In the sequel, in order to fix the ideas, I will suppose that the position variable lives in the torus $\mathbb{T}^d$.

6.1. Global Maxwellian

We have already encountered local Maxwellians that make the collision operator vanish. In order to make the operator $v \cdot \nabla_x$ also vanish, it is natural to look for Maxwellians whose parameters ρ, u, T are homogeneous, constants independent of position. A single set of these parameters is compatible with the laws of conservation of mass, momentum and energy. The distribution thus obtained is called **global Maxwellian**:

$$M_{\rho u T} = \rho \, \frac{1}{(2\pi T)^{d/2}} \exp\left(-\frac{|v-u|^2}{2T}\right).$$

Without loss of generality, even with a change in Galilean reference or physical scale, we may suppose that $\rho = 1$, $u = 0$ and $T = 1$, and we will denote the corresponding distribution by M.

This distribution is thus an equilibrium for Boltzmann's equation. Moreover, it is easy to verify that it is the distribution that maximizes entropy under the constraints of fixed mass, linear momentum and total energy. This selection criterion foreshadows the classical theory of equilibrium statistical mechanics and Gibbs' famous canonical ensembles (Gibbs measure).

6.2. The entropic argument

Boltzmann now uses Theorem H to give a more solid justification to the global Maxwellian: he notes that

- entropy increases strictly unless it is in a hydrodynamic state
- the global Maxwellian, stationary, is the only hydrodynamic solution of Boltzmann's equation.

The image that emerges is that the entropy will continue to increase as much as possible since the distribution never remains "stuck" on a hydrodynamic solution; the entropy will end up approaching the maximal entropy of the global Maxwellian, and convergence results.

In this regard we can make two remarks: the first is that the Lebesgue measure, which we have taken as the reference measure in Boltzmann's entropy, may be replaced by the Maxwellian: in fact

$$H(f) - H(M) = \iint f \log \frac{f}{M} \, dv \, dx = H_M(f),$$

where we have used the fact that $\log M$ is a collision invariant. The second remark is that the difference in entropies allows us to quantify the difference between the Gaussian and equilibrium, for example by virtue of the Csiszàr–Kullback–Pinsker inequality: $H_M(f) \geq \|f - M\|_{L^1}^2 / 2\|M\|_{L^1}$.

Boltzmann's reasoning is essentially correct and it is not difficult to transform it into a rigorous argument by showing that sufficiently regular solutions of the Boltzmann equation approach Maxwellian equilibrium. In the context of spatially homogeneous solutions, T. Carleman formalized this reasoning in 1932 [26].

However, Boltzmann did not have the means for making his argument qualitative; it would be necessary to wait almost a century before anyone dared to pose the problem of the speed of convergence toward Gaussian equilibrium, especially pertinent since the range of validity of Boltzmann's equation is not eternal and is limited in time by phenomena such as the Poincaré recurrence theorem.

6.3. The probabilistic approach of Mark Kac

At the beginning of the 1950s, Kac [66] attempted to understand convergence toward equilibrium for the Boltzmann equation and began by simplifying the model. Kac ignores positions, simplifies the collision geometry extravagantly and invents a stochastic model *where randomness is present in the interaction*: whenever two particles interact, one draws at random the parameters describing the collision. Positions being absent, the particles all interact, each with all the others, and thus a "mean field" model is produced. This simplified probabilistic model is for Kac an opportunity for formalizing mathematically the notion of propagation of chaos in the mean field equations, which would prove so fertile and would be taken up later by Sznitman [98] and many others.

Suspicious of Boltzmann's equation, Kac wants to explain the convergence by microscopic probabilistic reasoning on the level of an N-particle system; he attempts to obtain spectral gap estimates that are uniform in N. His approach seems naive nowadays in that it underestimates the difficulty of treating dimension N; nonetheless, the problem of determining the optimal spectral gap, resolved a half-century later, has proved to be very interesting [31, 65, 79]. For this subject readers can equally well consult [104, Section 6] and [32], where there is interest in the entropic version of this "microscopic" program.[3]

In 1966 McKean [81] resumed Kac's work and drew a parallel with the problems of the central limit theorem. He introduced the tools of information theory to the subject, in particular Fisher information [41], which measures the difficulty in estimating a parameter such as the velocity of the particles. The program would be completed by Tanaka [100], who discovered new contracting distances, and would culminate with the work of Carlen, Carvalho, Gebetta, Lu, Toscani on "the central limit theorem for the Boltzmann equation" [29, 30, 33, 34]. This theory encompasses basic convergence theorems based on the combinatorics of interactions between particles and tools from the study of central limit theorem (weak distances...), as well as counterexamples demonstrating extremely slow convergence to equilibrium.

This stochastic program allows us to dispense with Theorem H; in fact it has also permitted the updating of several Lyapunov functionals: Tanaka's contracting

[3]This program culminates in a recent manuscript by Mischler and Mouhot.

distance (of optimal transport), Fisher's information. Nonetheless, from a technical point of view, the whole theory remains essentially confined to Maxwellian interactions (in which $\tilde{B}(v - v_*, \omega)$ depends only on the angle between $v - v_*$ and ω) and to spatially homogeneous gases. Chapter 4 of [103] is dedicated to particular properties, highly elegant moreover, of these interactions.

For gaining generality and for studying inhomogeneous situations or non-Maxwellian interactions, the only robust approach known to this day is based on the H Theorem.

6.4. Cercignani's conjecture

Boltzmann's H Theorem is general and relevant, so that it is natural to look for its quantitative refinements. At the beginning of the 1980s, C. Cercignani asked if one could *estimate from below* the production of local entropy as a function of the "non-Gaussianity" of the kinetic distribution, ideally by a multiple of the information $H_M(f)$. It was not until a decade later that Carlen and Carvalho [28], Desvillettes [44], without answering Cercignani's question, could nonetheless present quantitative lower bounds for the production of entropy.

A more precise answer to this problem is obtained in my articles [101] and [103] (the first in collaboration with G. Toscani). Without loss of generality we suppose that $\int f\,dv = 1$, $\int f\,v\,dv = 0$, $\int f\,|v|^2\,dv = d$; the general case being deducible by change of scale or reference frame. We begin by mentioning a surprisingly simple example, taken from [104], which is applied in a nonphysical situation: if $B(v - v_*, \sigma) \geq K\,(1 + |v - v_*|^2)$, then

$$\mathrm{PE}\,(f) \geq \left(K_B \,\frac{|S^{d-1}|}{8}\;\frac{d-1}{1+2d}\right)\; T_f^*\;\; H_M(f), \tag{37}$$

where

$$T_f^* = \inf_{e \in S^{d-1}} \int f(v)\,(v \cdot e)^2\,dv.$$

The quantity T_f^* quantifies the nonconcentration of f near a hyperplane; it is estimated from below with any information on entropy or regularity, even automatically for radially symmetric distributions. The conciseness of the result masks a surprising proof technique whereby f is regularized by an auxiliary diffusion semigroup; under the effect of this this semigroup, the variation in the production of Boltzmann entropy is essentially estimated by the production of Landau entropy, which in turn is estimated in terms of Fisher information before integrating along the semigroup; see [104] or [108] for the details.

The hypothesis of quadratic growth in the relative velocity is not physically realistic; it is nonetheless optimal in the sense that there exist counterexamples [16] for kernels with growth $|v - v_*|^\gamma$, for each $\gamma < 2$. One can then work on inequality (37) in order to derive from it a weaker underestimate that applies to realistically effective sections, such as the model of solid spheres; the principal difficulty lies in controlling the quantity of entropy production induced by the small relative velocities ($|v - v_*| \leq \delta$). The logarithms make this control delicate; see [104] for the details. In the end, for each $\varepsilon > 0$ we obtain the inequality

$$\mathrm{PE}\,(f) \geq K_\varepsilon(f) \quad \Big[H(f) - H(M^f)\Big]^{1+\varepsilon}, \tag{38}$$

where M^f is the Maxwellian associated with f, i.e., that with parameters ρ, u, T corresponding to the density, mean velocity and temperature of f. The constant $K_\varepsilon(f)$ depends only on ε, on the C^r regularity of f for r sufficiently large, on a moment $\int f|v|^s\,dv$ for s sufficiently large, and on a lower bound $f \geq K\,e^{-A|v|^q}$. The question as to whether these hypotheses can be relaxed remains open.

6.5. Conditional convergence

Inequality (38) concerns a function $f = f(v)$ but does not include spatial dependence; this is inevitable since the variable x does not enter into the study of the global production of entropy. Of course, (38) immediately implies (modulo the regularity bounds) convergence in $O(t^{-\infty})$ for the spatially homogeneous equation, i.e., the distance between the distribution and equilibrium tends to 0 faster than t^{-k} for each k; yet this inequality does not resolve the inhomogeneous problem. The obstacle to be overcome is the *degeneracy of entropy production on hydrodynamic states.* A key to the study over long time of Boltzmann's equation thus consists in showing that not too much time is spent in a hydrodynamic, or approximately hydrodynamic, state. To avoid this trap, we can only depend on the transport, represented by the operator $v \cdot \nabla_x$. Grad [56] had understood in 1965, in a moreover rather obscure paper: "the question is whether the deviation from a local Maxwellian, which is fed by molecular streaming in the presence of spatial inhomogeneity, is sufficiently strong to ultimately wipe out the inhomogeneity" (...) "a valid proof of the approach to equilibrium in a spatially varying problem requires just the opposite of the procedure that is followed in a proof of the H-Theorem, viz., to show that the distribution function does not approach too closely to a local Maxwellian."

In 2000s, Desvillettes and I [46] rediscovered this principle formulated by Grad and we established a version of the **instability of hydrodynamic approximation**: if the system becomes, at a given moment, close to being hydrodynamic without being in equilibrium, then transport phenomena cause it to leave this hydrodynamic state. This is quantified, under the hypothesis of strong regularity, by studying second variations of the square of the norm, $\|f - M^f\|^2$, between $f = f(t,x,v)$ and the associated *local* Maxwellian

$$M^f(t,x,v) = \rho(t,x)\,\frac{1}{(2\pi T(t,x))^{d/2}}\exp\left(-\frac{|v-u(t,x)|^2}{2T(t,x)}\right),$$

$$\rho(t,x) = \int f(t,x,v)\,dv, \qquad u(t,x) = \frac{1}{\rho(t,x)}\int f(t,x,v)\,v\,dv,$$

$$T(t,x) = \frac{1}{d\,\rho(t,x)}\int f(t,x,v)|v-u(t,x)|^2\,dv.$$

In some way M^f is the best possible approximation of f by a hydrodynamic state, and the study of the variations $\|f - M^f\|$ allows us to verify that f cannot long remain close to a hydrodynamic state.

By adjoining (in an especially technical way, with the help of numerous functional inequalities) the quantitative H Theorem with the instability of hydrodynamic approximation, we end up with **conditional convergence**: a solution of the Boltzmann equation satisfying uniform regularity bounds converges toward equilibrium in $O(t^{-\infty})$. This result is constructive in the sense that the time constants involved depend only on the regularity bounds, on the form of the interaction and on the boundary conditions. The convergence resides in a system of inequalities that simultaneously involve the entropy and the distance to the hydrodynamic states. For example, one of them is written

$$\frac{d^2}{dt^2}\|f-M^f\|_{L^2}^2 \geq K\int |\nabla T|^2\,dx - \frac{C}{\delta^{1-\varepsilon}}\left(\|f-M^f\|_{L^2}^2\right)^{1-\varepsilon} - \delta[H(f)-H(M)]. \quad (39)$$

In order to understand the contribution of such an inequality, we suppose that f becomes hydrodynamic at some moment: then $f = M^f$ and (38) is useless. But if the temperature is inhomogeneous and if δ in (39) is very small, then we are left with $(d^2/dt^2)\|f - M^f\|_{L^2}^2 \geq$ const., which certainly keeps f from remaining close to M^f for very long. Once f has exited the hydrodynamic approximation, we can reapply (38), and so forth. This reasoning only works when the temperature is inhomogeneous, but we can find other inequalities involving macroscopic velocity gradients and the density. We thus pass from a "passive" argument to an "active" argument, where the increase in entropy is forced by differential inequalities rather than by the identification of a limit.

We end this section with several commentaries on the hypotheses. The regularity theory of the Boltzmann equation allows reduction of the general bounds to very particular bounds, e.g., it is known that the kinetic distribution is automatically minorized by a multiple of $e^{-|v|^q}$ if, for example, the equation is set on the torus and the solution is regular. It is also known that bounds on the moments of low order allow having bounds on arbitrarily high moments, etc. But regularity in the general context remains a celebrated open problem. The conditional convergence result shows that it is the final obstacle separating us from quantitative estimates of convergence to equilibrium; it likewise unifies the already known results on convergence: both the case of spatially homogeneous distributions and that of distributions close to equilibrium are situations in which we have an almost complete regularity theory.

In studying convergence toward equilibrium for the Boltzmann equation, we observe a subtle interaction between the collision operator (nonlinear, degeneratively dissipative) and the transport operator (linear, conservative). Neither of the two, taken separately, would be sufficient for inducing convergence, but the combination of the two succeeds. This situation arises rather frequently and recalls the hypoellipticity problem in the theory of partial differential equations. By analogy, the *hypocoercivity* problem is the study of the convergence properties for potentially degenerate equations.

A somewhat systematic study of these situations, both for linear and nonlinear equations, was made in my memoir [107]. The general strategy consists of constructing Lyapunov functionals adapted to the dynamic, while adjoining by a natural functional (like entropy) a well-chosen term of lesser order. A case study is the "$A^*A + B$ theorem", inspired by Hörmander's sum of squares theorem, which gives sufficient conditions on the commutators between operators A and B, with B antisymmetric, for the evolution $e^{-t(A^*A+B)}$ to be hypocoercive. In the simpler variant, one of these conditions reminiscent of Hörmander's Lie algebra condition, is the coercivity of $A^*A + [A, B]^*[A, B]$.

Hypercoercivity theory has now taken on a life of its own and there are already a number of striking results; it continues to expand, especially in the nonlinear context. This is true as well both for kinetic theory, as in the paper [58] that will be mentioned in the next section, and outside of kinetic theory, as in the paper by Liverani and Olla on the hydrodynamic limits of certain particle systems [73].

In a nonlinear context, the principal result remains [109, theorem 51]; this general statement allows for simplification of the proof of the conditional convergence theorem for the Boltzmann equation, and includes new interactions and limit conditions. See [109, part III] for more details.

6.6. Linearized system

The rate of convergence to equilibrium can be determined by a linearized study. We begin by flushing out a classical logical mistake: the linearized study can in no case be a substitute for the nonlinear study, since linearization is only valid beginning from where the distribution is very close to equilibrium.

The linearized study of convergence requires overcoming three principal difficulties:

- quantitatively estimating the spectral gap for the linearized collision operator;
- performing a spectral study of the linearization in a space appropriate for the nonlinear problem, so as to achieve a "connection" between the nonlinear study and the linearized study;
- take into account the hydrodynamic degeneracy from a hypercoercive perspective: in fact, the linearized equation is just as degenerate as the nonlinear equation.

All of these difficulties have been resolved in the last decade by C. Mouhot and his collaborators Baranger, Gualdani and Mischler [11, 58, 82], at least in the emblematic case of solid spheres. Thus the recent article [58] establishes a conditional convergence result with exponential rate $O(e^{-\lambda t})$ instead of $O(t^{-\infty})$, and the rate λ is estimated in a constructive manner.

Exponential convergence is not a universal characteristic of Boltzmann's equation: we do not expect it for hard potentials or even for the moderately soft. To fix our ideas, let us suppose that the collision kernel behaves like $|v - v_*|^\gamma\, b(\cos\theta)$. In the case where $b(\cos\theta)\,\sin^{d-2}\theta$ is integrable (often by angular truncation for the grazing collisions), the linearized collision operator only admits a spectral gap for $\gamma \geq 0$. An abundance of grazing collisions permits extending this condition, as Mouhot and Strain [83] showed: if $b(\cos\theta)\,\sin^{d-2}\theta \simeq \theta^{-(1+\nu)}$ for $\theta \to 0$ (important grazing collisions), then the linearized collision operator only allows a spectral gap for $\gamma + \nu \geq 0$. The regularity theory is presently under development for such equations (work of Gressman–Strain, Alexandre–Morimoto–Ukai–Xu–Yang), and we can wager that within a few years the linearized theory will cover all these cases.

For too soft potentials (or for the Landau's model of Coulomb collisions), there is no spectral gap and the best result we can expect is fractional exponential convergence $O(e^{-\lambda t^\beta})$, $0 < \beta < 1$. Such estimates can be found in the paper of Guo and Strain [60].

6.7. Qualitative evolution of entropy

A recurrent theme in this whole section is the degeneracy related to hydrodynamic states, which disturbs the convergence to equilibrium. In the beginning years of this century, Desvillettes and I suggested that this degeneracy is reflected in oscillations in the production of entropy. Never previously observed, these oscillations have been identified in very precise numerical simulations by F. Filbet. Below I reproduce a striking curve, obtained with Boltzmann's equation in a one-dimensional periodic geometry.

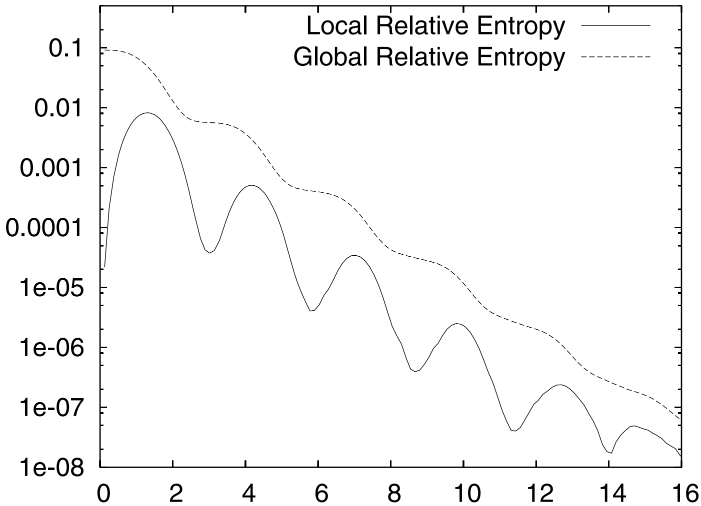


FIGURE 1. Logarithmic evolution of the kinetic function H and of the hydrodynamic function H for the Boltzmann equation in a periodic box.

In this diagram, the logarithm of the function H has been drawn as a function of time; the global rectilinear decrease thus corresponds to an exponential convergence toward the equilibrium state. The kinetic information has also been separated in to hydrodynamic information and “purely kinetic” information:

$$\int f \log \frac{f}{M} = \left(\int \rho \log \frac{\rho}{T^{d/2}} \right) + \int f \log \frac{f}{M^f};$$

the second quantity (purely kinetic information) is the curve that is seen just below the curve of the function H. When the two curves are distant from each other, the distribution is almost hydrodynamic; when they are close, the distribution is almost homogeneous. Starting from the hydrodynamic distribution, it deviates immediately, in conformance with the instability principle for the hydrodynamic approximation. One subsequently clearly sees oscillations between rather hydrodynamic states, associated with a slowing down in entropy production, and the more homogeneous states; these oscillations are important, given the logarithmic nature of the diagram. Filbet, Mouhot and Pareschi [50] present other curves and attempt to explain the oscillation frequency in a certain asymptotic process.

Here the Boltzmann equation nicely reveals its double nature, relevant for both transfer of information via collisions and fluid mechanics via the transport operator. It is often the marriage between the two aspects that proves delicate.

The relative importance of transport and collisions can be modulated by the boundary conditions; in the periodic context it comes down to the size of the box. A large box will permit important spatial variations, giving the hydrodynamic effects free rein, as in the above simulation. Nonetheless, we clearly see that even in this case, and contrary to an idea well ingrained even with specialists, the asymptotic regime is not hydrodynamic, in the sense that the ratio between hydrodynamic entropy and total kinetic entropy does not increase significantly as time passes, oscillating rather between minimum and maximum values.

We can ask ourselves what happens in a rather small box. Such a simulation is presented below.

The conclusion that we can draw from this figure is precisely opposite to our intuition, according to which the hydrodynamic effects dominate in the long run: quite the contrary, starting from a hydrodynamic situation, we quickly arrive at a situation that is almost homogeneous (visually we have the impression that at time $\simeq 0.7$ the hydrodynamic information represents scarcely more that 1% of the total information!). The inhomogeneous effects then resume their rights (at time $t = 1$ the information is divided into parts of the same order), after which it becomes resolutely homogeneous. In this example, the homogenization has proceeded faster than convergence to equilibrium. We’ll return to this figure, which has caused some perplexity, in Section 8.

6.8. Two nonconventional problems

I end this section by mentioning two peculiar problems linked to time’s arrow in Boltzmann’s equation that are perhaps just curiosities. The first is the classification

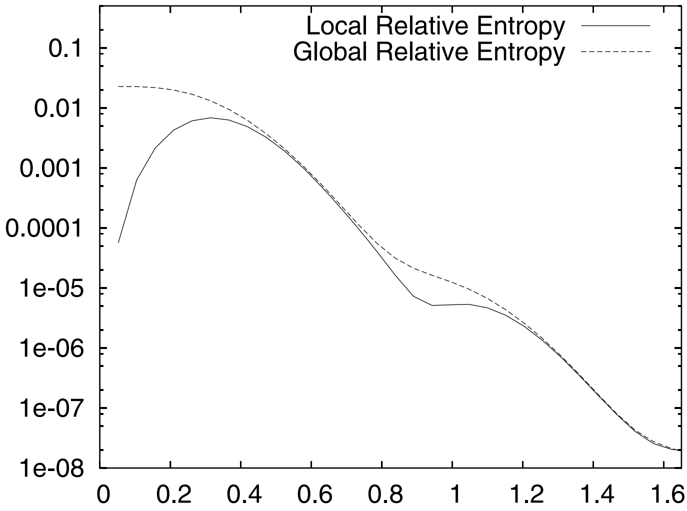


FIGURE 2. The same thing in a smaller box.

of eternal solutions of Boltzmann's equation: I tried to show in my doctoral thesis that, at least for the spatially homogeneous Boltzmann equation with Maxwellian molecules, there do not exist eternal solutions with finite energy. The second would be to instead look for self-similar solutions with finite energy that do not converge to Maxwellian equilibrium. For the first problem, [111] can be consulted for partial results; the conjecture is still viable, and Bobylev and Cercignani [18] have been able to show that there does not exist any eternal solution having finite moments of all orders. As to the second problem, it has been resolved by the same authors [17] using Fourier transform techniques.

7. Isentropic relaxation: living with ones memories

We now consider Vlasov's equation with interaction potential W:

$$\frac{\partial f}{\partial t} + v \cdot \nabla_x f - \left(\nabla W * \int f \, dv \right) \cdot \nabla_v f = 0. \tag{40}$$

Unlike Boltzmann's equation, equation (40) does not impose time's arrow and remains unchanged under the action of time reversal. The constancy of entropy corresponds to a preservation of microscopic information. The solution of Vlasov's equation at time t theoretically permits reconstructing the initial condition without loss of precision, simply by solving Vlasov's equation after having reversed the velocities.

Additionally, whereas Boltzmann's equation allows but a very small number of equilibria (the Maxwellians determined by the conservation laws), Vlasov's equation allows a considerable number of them. For example, *all* the homogeneous distributions $f^0 = f^0(v)$ are stationary. There exist yet many other stationary distributions, for example the family of Bernstein–Greene–Kruskal waves [14]. For all these reasons, there is nothing *a priori* that would lead us to conjecture a well-determined behavior over the long term and there is no indication at all of time's arrow. However, in 1946, L. Landau, released several years earlier from the soviet communist prisons where his frank speech had led him, suggested a very specific long term behavior for Vlasov's equation. It is based on an analysis of the linearized equation near a homogeneous equilibrium. Landau's prediction provoked a shock and a conceptual change which still today raises lively discussions [92]; in its sequel it began to be suspected that convergence toward equilibrium is not necessarily tied to an increase in entropy. This section is devoted to a survey of the question of isentropic convergence, while emphasizing the perturbation regime, which is the only one for which there are sound elements. More details can be found in my course [110].

7.1. Linearized analysis

We study Vlasov's equation near a homogeneous equilibrium $f^0(v)$. If we set $f(t,x,v) = f^0(v) + h(t,x,v)$, the equation becomes

$$\frac{\partial h}{\partial t} + v \cdot \nabla_x h + F[h] \cdot \nabla_v f^0 + F[h] \cdot \nabla_v h = 0, \tag{41}$$

where

$$F[h](t,x,) = -\iint \nabla W(x-y)\, h(t,y,v)\, dv$$

is the force induced by the distribution h.

By neglecting the quadratic term $F[h] \cdot \nabla_v h$ in (41), we obtain the **linearized Vlasov equation** near a homogeneous equilibrium:

$$\frac{\partial h}{\partial t} + v \cdot \nabla_x h + F[h] \cdot \nabla_v f^0 = 0. \tag{42}$$

Before examining (42), we consider the case without interaction ($W = 0$), i.e., the **free transport** $\partial_t f + v \cdot \nabla_x f = 0$. This equation is solved in $\mathbb{T}^d_x \times \mathbb{R}^d_v$ by $f(t,x,v) = f_i(x - vt, v)$, where f_i is the initial distribution. We change to Fourier variables by putting

$$\tilde{g}(k,\eta) = \iint g(x,v)\, e^{-2i\pi k\cdot x}\, e^{-2i\pi\eta\cdot v}\, dx\, dv;$$

the free transport solution is then written

$$\tilde{f}(t,k,\eta) = \tilde{f}_i(k, \eta + kt). \tag{43}$$

When $k \neq 0$, this expression tends to 0 when $t \to \infty$, with rate determined by the regularity of f_i in the velocity variable (Riemann–Lebesgue principle). All these nonzero spatial modes thus relax toward 0; it is the **homogenizing** action of free transport.

Equation (42) is not so easily solved; nonetheless, if we put $\rho(t,x) = \int h(t,x,v)\,dv$, we then discover that the various modes $\hat{\rho}(t,k)$ all satisfy independent equations for distinct values of k. This remarkable *decoupling* property for the modes is the foundation for Landau's analysis. For each k we have a Volterra equation for the kth mode:

$$\hat{\rho}(t,k) = \tilde{f}_i(k,kt) + \int_0^t K^0(k,t-\tau)\,\hat{\rho}(\tau,k)\,d\tau,$$

where

$$K^0(t,k) = -4\pi^2 \hat{W}(k)\tilde{f}^0(kt)\,|k|^2\,t.$$

The stability of Volterra equations is a classical problem. If u satisfies $u(t) = S(t) + \int_0^t K(t-\tau)\,u(\tau)\,d\tau$, then the rate of decrease of u is dictated by the worse of two rates: the rate of decrease of S of course, and on the other hand the width of the largest band $\{0 \le \Re\xi \le \Lambda\}$ that does not intersect any solution of the equation $K^L = 1$, where K^L is the Laplace transform of K. If $\Lambda > 0$, we thus have exponential stability for the linearization.

Adapted to our context, this result leads to the **Penrose stability criterion**, for which a multidimensional version will be stated. For each $k \in \mathbb{Z}^d$, we define $f_k^0 : \mathbb{R} \to \mathbb{R}_+$ by

$$f_k^0(r) = \int_{k^\perp} f^0\left(\frac{k}{|k|}\,r + z\right) dz;$$

in short, f_k^0 is the marginal of f^0 in the kth direction. Penrose's criterion [88] requires that for each $k \in \mathbb{Z}^d$,

$$\forall \omega \in \mathbb{R}, \qquad (f_k^0)'(\omega) = 0 \implies \hat{W}(k) \int \frac{(f_k^0)'(v)}{v-\omega}\,dv < 1.$$

If this criterion (essentially optimal) is satisfied, then there is exponential stability for the linearization: the force decreases exponentially fast, as do all inhomogeneities of the spatial density $\int h\,dv$.

The Penrose stability criterion is satisfied in numerous situations: in the case of a Coulomb interaction when the marginals of f^0 increase to the left of 0, decrease to the right (in other words, if $(f_k^0)'(z)/z < 0$ for $z \neq 0$); in particular if f^0 is a decreasing function of $|v|$, a Gaussian for example. Again in the Coulomb case, in dimension 3 or more, the criterion is verified if f^0 is isotropic. In the case of Newtonian attraction, things are more complex: for example, for a Gaussian distribution, the stability depends on the mass and the temperature of the distribution. This reflects the celebrated Jeans instability, according to which the Vlasov equation is linearly unstable for lengths greater than

$$L_J = \sqrt{\frac{\pi T}{\mathcal{G}\rho^0}},$$

where G is the constant of universal gravitation, ρ^0 the mass of the distribution f^0 and T its temperature. It is this instability which is responsible for the tendency of massive bodies to regroup themselves in "clusters" (galaxies, clusters of galaxies, etc.).

In summary, the linearized Vlasov equation about a stable homogeneous equilibrium (in the sense of Penrose) predicts an exponential dampening of the force, in an apparently irreversible manner. This discovery brought back the problematic of time's arrow in the theory of Vlasov's equation.

The study of the linearized Vlasov equation can be found in advanced treatises on plasma physics, like [71]; however, the treatment there is systematically obscured by the use of contour integrals in the complex plane, which arise from the inversion of the Laplace transform. This has been avoided in the presentation of [85, Section 3], based on the simple Fourier transform; or in the short version [110].

7.2. Nonlinear Landau dampening

The linearization effected by Landau perhaps is not an innocent operation, and for half a century doubts have been expressed on its validity. In 1960, Backus [6] remarked that replacing $\nabla_v(f^0 + h)$ by $\nabla_v f^0$ in the force term would be conceptually simple if $\nabla_v h$ remained small throughout all time; but if we replace h by the solution of the linearized equation, we see that its velocity gradient grows linearly in time, becoming arbitrarily large. This, suggests Backus, "destroys the validity of the linear theory". Backus's argument is questionable because $\nabla_v h$ is multiplied by $F[h]$ which one expects to see decrease exponentially; nevertheless heuristic considerations [86] suggest the failure of the linear approximation at the end of time $O(1/\sqrt{\delta})$, where δ is the size of the initial perturbation. Thereafter the curve (drawn by F. Filbet) represents the logarithm of the quotient of the energy computed using the nonlinear equation and that obtained from the linear equation, for different values of the perturbation amplitude; it is clearly seen that even for δ small, we end up in a process where the nonlinear effects cannot be neglected.

There are other reasons for distrusting the linearization. First, the eliminated term, $F[h] \cdot \nabla_v h$, is of higher degree in terms of derivatives of h with respect to velocity. Next, the linearization eliminates conservation of entropy, and favors the particular state f^0, which voids the discussion of reversibility.

In 1997, Isichenko [63] muddied the waters by arguing that convergence toward equilibrium cannot in general be more rapid than $O(1/t)$ for the nonlinear equation. This conclusion seemed to be contradicted by Caglioti and Maffei [25], who constructed exponentially damped solutions of the nonlinear equation. Numerical simulations (see below) are not very reliable over very long time and there is felt need for theorem.

In 2009, Mouhot and I established such a result [85]. If the interaction potential W is not too singular, in the sense that $\hat{W}(k) = O(1/|k|^2)$ (this hypothesis allows just Coulomb and Newton interactions!), and if f^0 is an analytic homogeneous equilibrium satisfying Penrose's stability condition, then there is nonlinear dynamic stability: starting with initial data f_i, analytic and such that $\|f_i - f^0\| = O(\delta)$ when δ is very small, we have decrease of the force in $O(e^{-2\pi\lambda|t|})$

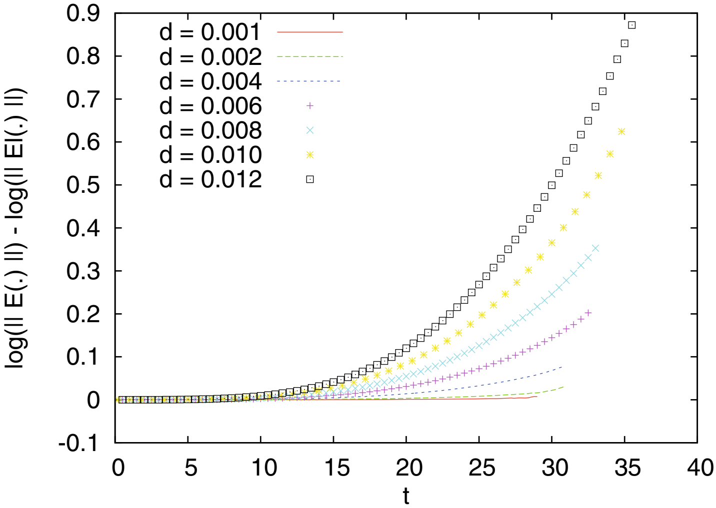


FIGURE 3. For a Vlasov evolution, the logarithmic ratio between the norms of the energy following the nonlinear equation to that following the linear equation, for different perturbation amplitudes.

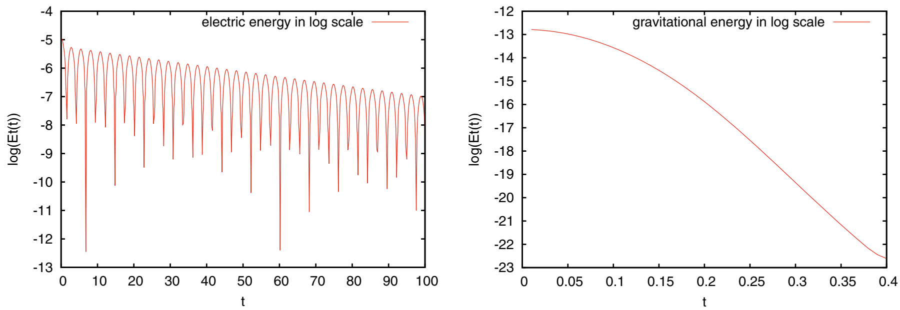


FIGURE 4. Evolution of the norm of the force field, for electrostatic interactions (left) and gravitational interactions (right). In the electrostatic case, the rapid oscillations are called Langmuir waves.

for all $\lambda < \min(\lambda_0, \lambda_i, \lambda_L)$, where λ_0 is the width of the band of complex analyticity of f^0 about $\mathbb{R}^d_v$, λ_i is the width of the band of complex analyticity of f_i in the variable v, and λ_L is the rate of the Landau convergence. In brief, linear damping implies nonlinear damping, with an arbitrarily small loss in rate of convergence.

The theorem also establishes the weak convergence of $f(t, \cdot)$ to an asymptotically homogeneous state $f_\infty(v)$. More precisely, the equation being invariant under time reversal, there is an asymptotic profile $f_{+\infty}$ for $t \to +\infty$, and another profile $f_{-\infty}$ for $t \to -\infty$. If Vlasov's equation is viewed as a dynamical system, there is then a remarkable behavior: the homoclinic/heteroclinic trajectories are so numerous that they fill an entire neighborhood of f^0 in analytic topology.

The nonlinear damping of Vlasov's equation is based on confinement and mixing. Containment is indispensable: it is known that Landau damping does not take place in all space, even for the linearized equation [52, 53]; in our case it is automatic because the phase space is the torus. Mixing takes place because of the differential velocity phenomenon: particles with different velocities move with different velocities in phase space; here it is almost a tautology. An example of a nonmixing system is the harmonic oscillator, where the trajectories borne by variables with different action move with constant angular velocity. Some of the other ingredients underlying the nonlinear study are:

- a reinterpretation of the problem in terms of regularity: instead of showing that there is damping, it is shown that $f(t, x, v)$ is "as regular" as the free transport solution, uniformly in time;
- "deflection" estimates: a particle placed in an exponentially decreasing force field follows a free transport asymptotic trajectory in a sense that can be quantified precisely;
- the stabilizing role of retarded response, *in echoes*, of the plasma: when one of the modes of the plasma is perturbed, the reaction of the other modes is not instantaneous, but follows with a slight retardation, because the effect of the modes is compensated outside of some instants of resonance;
- a Newton scheme that takes advantage of the fact that the linearized Vlasov equation is in some way completely integrable; the speed of convergence of this scheme compensates for the loss of decay that accompanies the solution of the linearization.

All these ingredients are described in more detail in [110]. The special place of the Newton scheme and of the complete integrability form an unexpected bridge with KAM (Kolomogorov–Arnold–Moser) theory. In some way the Vlasov nonlinear Vlasov equation, in the perturbative process, inherits some of the good properties of the completely integrable linearized Vlasov equation.

From the physical point of view, information goes toward the small kinetic scales: the oscillations of the distribution function are amplified when time becomes large, and become invisible. Lynden-Bell [75, 76] clearly understood this and used a striking formula for explaining: "a [galactic] system whose density has achieved a steady state will have information about its birth still stored in the peculiar velocities of its stars."

These oscillations, clearly visible in the figures below, are both a nuisance from the technical point of view and the fundamental physical mechanism that produces the impression of irreversibility. We note the difference with the mechanism called *radiation*, in which the energy is emitted on macroscopic scale and goes off to infinity: here to the contrary the energy literally vanishes into thin air…

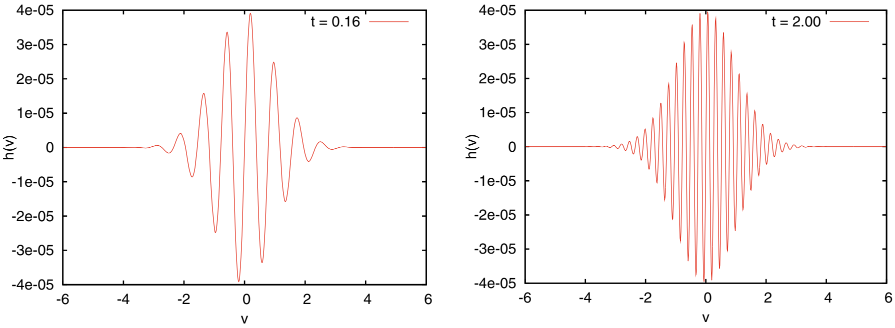


FIGURE 5. A section of the distribution function (in relation to a homogenous equilibrium) for gravitational Landau damping, at two different times.

7.3. Gliding regularity

The nonlinear damping theorem is based on a recent reinterpretation, in terms of regularity, that deserves some comments. We begin by talking about the cascade associated with free transport, represented on the diagram below:

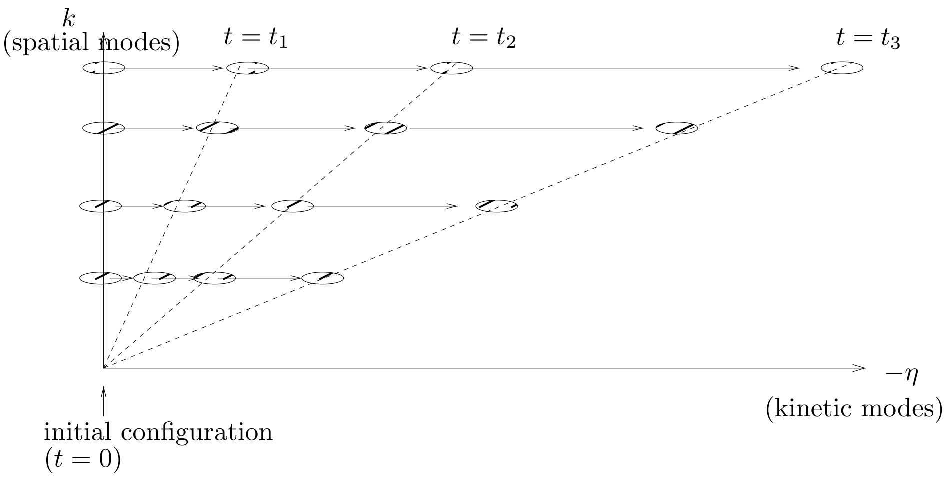


FIGURE 6. Evolution of energy in the space of frequencies along free transport or of a perturbation of this latter, the marks indicating the localization of energy in phase space.

This image, which is derived from formula (43), shows that the frequencies that matter vary over time: there is an overall movement toward high kinetic frequencies, and this movement is all the faster than the frequency is high. More precisely, the spatial mode of frequency k oscillates in the velocity variable with period of order $O(1/|k|t)$. The challenge of Landau damping is to show that this cascade, although distorted, is globally preserved by the effect of the interactions that *couple* the different modes.

These strong oscillations preclude any hope of obtaining bounds that are uniform in time, e.g., analytically regular in the usual sense. A key idea in [85] consists of concentrating on the Fourier modes that matter for the free transport solution, and thus to follow the cascade over the course of time. This concept is called **gliding regularity** and comes with a degradation of the regularity bounds in velocity, but simultaneously with an improvement of regularity in position, once velocity averages have been formed. Our interpretation of Landau damping is thus a transfer of regularity away from the variable v and toward the variable x, the regularity of the force improving, which implies that its amplitude dies off.

The analytic norm used in [85] is a bit complex: it has good algebraic properties that allow following the errors obtained by composition, it adapts well to the geometry of the problem, and follows free transport for measuring the gliding regularity:

$$\|f\|_{\mathcal{Z}_\tau^{\lambda,(\mu,\gamma);p}} = \sum_{k\in\mathbb{Z}^d}\sum_{n\in\mathbb{N}^d} \frac{\lambda^n}{n!}\, e^{2\pi\mu|k|}\ (1+|k|)^\gamma\ \left\|(\nabla_v + 2i\pi\tau k)^n\, \hat{f}(k,v)\right\|_{L^p(dv)} \quad (44)$$

(here $\hat{f}$ denotes the Fourier transform in the position variable, not in velocity). The exponent λ quantifies analytic regularity in velocity, the exponents μ and γ (by default $\gamma = 0$) quantify the regularity in position, and the parameter τ is to be taken as a time lag. Readers are referred to [85] for a study of the remarkable properties of this type of norm, and also for comparing results for more naive norms for which the nonlinear damping theorem can be stated.

The principal result of [85] consists in proving a uniform bound of type

$$\|f(t,\cdot) - f^0\|_{\mathcal{Z}_t^{\lambda,\mu;1}} = O(\delta).$$

This bound implies Landau damping, yet contains much more information: e.g., it shows that the higher spatial frequencies relax more quickly; it also implies nonlinear orbital stability under the Penrose condition, a problem that until now has resisted all the classical methods.

7.4. Nonlinear echoes and critical regularity

The celebrated plasma wave echo experiment [77, 78] describes the interaction of two waves generated by distinct spatial perturbations. If a first perturbation is sent at the initial time with a frequency k, there ensue oscillations with kinetic frequency $|k|t$, oscillations that do not attenuate over time but rather become more and more frenzied. If now at time τ a second perturbation with frequency ℓ is made to intervene, then oscillations with kinetic frequency $|\ell|(t-\tau)$ are generated. The two oscillation trains will be invisible to each other, due to averaging, except when they have the same kinetic frequency; this is produced in a time t such that

$kt + \ell(t - \tau) = 0$, or

$$t = \frac{\ell\tau}{k+\ell}; \tag{45}$$

where it is understood that k and ℓ are collinear and opposite in direction, with $|\ell| > |k|$. In a certain sense, in the long time asymptotic, the reaction to the second perturbation τ is achieved at a time u that is strictly greater than t. This delay is critical for explaining the stability of the nonlinear Vlasov equation. To get an idea of this gain, compare the inequality $u(t) \leq A + \int_0^t \tau\, u(\tau)\, d\tau$, which implies for u a growth essentially in $O(e^{t^2})$, to the inequality $u(t) \leq A + t\, u(t/2)$, which implies a very slow growth in $O(t^{\log t})$.

As a caricature of the estimates for the Vlasov–Poisson equation the family of inequalities

$$\varphi_k(t) \leq a(kt) + \frac{c\,t}{k^2}\,\varphi_{k+1}\left(\frac{kt}{k+1}\right)$$

can be proposed. Here $\varphi_k(t)$ represents roughly the norm of the kth mode of the spatial density at time t; $a(kt)$ represents the effect of the source (we ignore the linear term represented by a Volterra equation), the coefficient t translates the fact that the coupling occurs through the velocity gradient of f, and that the gradient grows linearly with time; $1/k^2$ is the Fourier transform of the interaction potential; we note in this regard that the interaction between modes is even more dangerous than the potential is singular; we keep only the interaction between the kth and the $(k+1)$-st mode; finally, the argument of the $(k+1)$-st mode is not t but $kt/(k+1)$, which represents a slight retardation with respect to t, as in the echoes formula. An explicit solution shows that

$$\varphi_k(t) \lesssim a(kt)\, \exp((ckt)^{1/3}).$$

These estimates can be adapted to the original Vlasov–Poisson equation; we then find, in the solution of the linearized equation about a nonstationary solution, a loss of regularity/decay that is fractional exponential. Under good assumption (as strong as the Penrose condition in the gravitational case, even stronger in the Coulomb case) we find essentially $\exp((kt)^{1/3})$; in the more general case the growth remains like a fractional exponential in kt. As it remains sub-exponential, this loss of regularity can be compensated by the exponential decay coming from the linear problem.

The loss of regularity depends essentially on the interaction, whereas the linear gain depends foremost on the regularity of the data: exponential for analytic data, polynomial for C^r data, fractional exponential for Gevrey data. The preceding discussion thus suggests that it is possible to extend the nonlinear damping theorem to Gevrey data. E.g., in the gravitational case, the critical exponent $1/3$ corresponds to a critical regularity Gevrey-3. We recall that a function is called Gevrey-ν if its successive derivatives do not grow faster than $O(n!^{\nu})$. Even if losing arbitrary little over ν, it is equivalent to requiring that its Fourier transform decay as a fractional exponential $\exp(c\,|\xi|^{1/\nu})$.

7.5. Speculations

The nonlinear Landau damping theorem opens a large number of questions. First, its extensions to geometries other than $\mathbb{T}^d$ is a real challenge, because then we lose the magical Fourier transform. The extension of inhomogeneous equilibria is still a distant dream; in fact, the linear stability of Bernstein–Greene–Kruskal waves is still not known!

Next, we have seen that it is known how to deal with damping under Gevrey regularity; but that the extension to lower regularities such as C^r regularity is an open problem. We have already emphasized the parallel with KAM theory, in which we know how to treat problems of class C^r; but in KAM the loss of regularity is only polynomial, and here it is much more severe. Certain variants of the KAM problem lead to a fractional exponential loss of regularity, and then it is likewise an open problem to work with regularity lower than Gevrey. In the immediate future, the only progress in C^r regularity suggested by [85] is the possibility of proving damping on time scales much larger than the nonlinear scale ($O(1/\delta)$ instead of $O(1/\sqrt{\delta})$, see [85, Section 13]); this development seems to depend on an original conjecture concerning the optimal constant occurring in certain interpolation inequalities). In Section 8 we shall again discuss a strategy permitting us to conceptually bypass this limitation of very high regularity.

Whatever the optimal regularity, it is not possible to obtain a Landau damping in the space with natural energy associated with the physical conservation laws. In fact, Lin and Zeng [72] show that nonlinear Landau damping is false if there are strictly less than two derivatives, in an appropriate sense.

Finally, even if Landau damping is but a perturbative phenomenon, it should be noted that its conceptual importance remains considerable because, at the present moment, it is the only little island that we are succeeding in exploring in the ocean of open problems related to isentropic relaxation. By its discovery, Landau raised awareness that physical systems may relax without there being any irreversibility and increase in entropy. In the 1960s, Lynden-Bell [75, 76] invoked this conceptual advance for solving the galactic relaxation problem, which appears in an approximately quasi-stationary state, whereas the relaxation times associated with the galactic Vlasov equation are vastly greater than the age of the universe. Subsequently, the **violent relaxation** principle – relaxation of the force field over certain times characteristic of the dynamic – has been generally accepted by astrophysicists, without there being any theoretical explanation to promote it. We have here a major scientific challenge.

8. Weak dissipation

Between the Boltzmann model that gives preference to collisions and that of Vlasov, which completely neglects them, we find a particularly interesting compromise in the Landau (or Fokker–Planck–Landau) model, weakly dissipative:

$$\frac{\partial f}{\partial t} + v \cdot \nabla_x f + F[f] \cdot \nabla_v f = \varepsilon\, Q_L(f, f), \tag{46}$$

where Q_L is the Landau operator (21).

In classical plasma physics, the coefficient ε equals $(\log \Lambda)/(2\pi\Lambda)$, where Λ is the plasma parameter, ordinarily very large (between 10^2 and 10^{30}). In a particle approach, the coefficient ε is a variation with respect to the limit of the mean field, proportional to $\log N/N$. The irreversible entropic effects modeled by the collision operator are only significantly apparent over large time $O(1/\varepsilon)$. Besides, regularizing effects are sensed instantly, even when they are weak. Interest in the study is thus multiple:

- it is a more realistic physical model than the "pure" Vlasov equation without collisions;
- it permits quantification, as a function of the small parameter ε, of the relative velocities of the homogenization (Landau damping) and entropic convergence phenomena;
- it permits bypassing the obstacle of Gevrey regularity that confronts the study of the noncollisional model.

Everything remains to be done here and I will merely sketch a long-term program.

8.1. A plausible scenario

Starting with a perturbation of homogeneous equilibrium with very rapid velocity decay, we should, in the course of temporal evolution per (46), remain close to a homogeneous regime; this is in the spirit of results of Arkeryd, Esposito and Pulvirenti [5] on the weakly inhomogeneous Boltzmann equation. In the homogeneous context, the operator of the right-hand side undoubtedly has the same regularization properties as a Laplacian in velocity, at least locally (the regularization properties become very weak at large velocities, but a very strong velocity decay is imposed). Assuming that this remains true in a weakly inhomogeneous context, we are left with a hypoelliptic equation that will regularize in all the variables, surely more quickly in velocity than in the position variable.

The hypoelliptic regularization in the Gevrey class has been but little studied, but using dimensional arguments we might think that in this context there is regularization in the Gevrey-$1/\alpha$ class, with velocity $O(\exp((\varepsilon t)^{-\alpha/(2-\alpha)}))$ in v, and $O(\exp((\varepsilon t)^{-3\alpha/(2-3\alpha)}))$ in x.

From another direction, in the Gevrey-$1/\alpha$ class, for $\alpha > 1/3$ we must have decay toward the homogeneous regime like $O(\exp -t^\alpha)$.

By combining the two effects we obtain homogenization on a $O(\varepsilon^{-\zeta})$ time scale, with $\zeta < 1$, which is a more rapid rate than the rate of increase of the entropy in $O(\varepsilon^{-1})$.

Balancing accounts, the coefficient ζ we might hope for is disappointing, of order 8/9. Among the steps used, the weakest link seems to be Gevrey regularization in x, which is extremely costly and perhaps not optimal since this regularity is not necessary in linear analysis. This motivates the development of a version of the nonlinear damping theorem in low regularity in x. If this regularity occurs, the coefficient becomes much better, of order 1/6...

8.2. Reexamining simulations

With this interpretation, we can now return to figure 6: using a small spatial box reinforces the effect of the operator $v \cdot \nabla_x$ at the expense of the collision operator, so that we are in a weakly dissipative process (the force field is zero). Then over long time homogenization happens more quickly than entropic relaxation. This does not explain everything, for two reasons: first, in this figure the initial condition is strongly (and not weakly) inhomogeneous; then the Boltzmann operator does not regularize. Nonetheless we may well want to believe that it is the homogenization by Landau damping that primarily manifests itself in this picture, before the collisions do their work in increasing entropy. (How to describe the temporary departure from the homogeneous process seems a mystery.)

9. Metastatistics

Here I use the word "metastatistics" to talk about statistics on the distribution function, which itself has a statistical content. This section will be short because we have scarcely more than speculations on the matter.

The Hewitt–Savage theorem, a reincarnation of the Krein–Milman theorem, describes the symmetric probability measures in a large number of variables as convex combinations of chaotic measures:

$$\mu^\infty = \int_{P(\mathcal{Y})} \mu^{\otimes\infty}\, \Pi(d\mu),$$

where Π is a probability measure on $P(\mathcal{Y})$, the space of probability measures on the macroscopic space. In brief, a microscopic uncertainty may be decomposed on two levels: besides the chaos with fixed macroscopic profile, there is the uncertainty about the macroscopic profile, which is to say the choice of profile μ that occurs with probability measure Π.

Now is there a natural probability measure Π on the space of admissible profiles? Ideally, such a measure will be **invariant under the dynamic**. In the context of the Boltzmann equation, the question really is not posed: only trivial measures, borne by Maxwellian equilibria, remain in contention. However, in the context of Vlasov's equation, the construction of nontrivial invariant measures is a fascinating problem. Such measures reflect the Hamiltonian nature of Vlasov's equation, studied for simplified interactions by Ambrosio and Gangbo [4].

A rather serious candidate for the status of invariant measure is Sturm's **entropic measure** [97], issuing from optimal transport theory, formally of the form $P = e^{-\beta H_\nu}$. Its complex definition has until now impeded success in proving its invariance. It should not be very difficult to modify the construction by appending an energy term. Sturm's measure is defined on a compact space, and there are perhaps subtleties in extending it to a kinetic context where the velocity space is not bounded. But the worst difficulty comes no doubt from the singularity of the typical measures: it is expected that $\mathbb{P}$-almost every measure is totally foreign to Lebesgue measure, and is supported by a set of codimension 1. This seems to close the door to every statistical study of damping based on regularity, and increases the mystery.[4]

Robert [91] and others have attempted to build a statistical theory of Vlasov's equation, starting from the notion of entropy, trying to predict the likely asymptotic state of dynamic evolution. The theory has gained some success, but it remains controversial. Furthermore, since the asymptotic state is obtained by a weak limit, the question arises of knowing whether an equality or inequality should be imposed on the constraints involving nonlinear functional density. For this topic readers can consult [102].

Then, this theory does not take into account the underlying evolution equation, postulating a certain universality with respect to the interaction. Isichenko [63] has remarked that the long-term asymptotic state, if it exists, must depend on fine details of the initial distribution and of the interaction, whereas the measures constructed by statistical theory only depend on invariants: energy, entropy, or other functionals of the form $\iint A(f)\,dx\,dv$. This objection has found substance with the counterexamples constructed in [85, Section 14], which show that the transformation $f(x,v) \to f(x,-v)$ can modify the final asymptotic state, while it preserves all the known invariants of the dynamic. The objection is perhaps surmountable, because these counterexamples are constructed in analytic regularity, i.e., in a class that must be invisible to a statistical treatment; but these counterexamples show the subtlety of the problem, and reinforce the feeling of difficulty in the construction of invariant measures.

10. Paradoxes lost

In this last section I will review a series of more or less famous paradoxes involving time's arrow and kinetic equations, and present their commonly accepted resolutions. A certain number of them involve infinity, a classic source of paradoxes such as "Hilbert's hotel" with an infinite number of rooms, where it is always possible to find a place for a new arrival even if the hotel is already full. On our scale, this paradox reflects our incapacity to account for the appearance or disappearance of a particle in relation to the gigantic number that make up our universe. The limit $N \to \infty$ (or the asymptotic $N \gg 1$, if like Boltzmann one prefers to avoid manipulating infinities) being the basis for statistical mechanics, it is not surprising that this paradox should arise.

In all the sequel, when I mention positive or negative time, or pre-collisional or post-collisional configuration, I am referring to the absolute microscopic time of Newton's equations.

[4] According to a personal communication by Mouhot, there are clues that Sturm's measure may be too singular to do the job.

10.1. The Poincaré–Zermelo paradox

In 1895 Poincaré [89] cast doubt on Boltzmann's theory because it seemed to contradict the fundamental properties of dynamical systems. A little later Zermelo [113] developed this point and noted that the inexorable increase in entropy prohibited the return of the system to the initial state, which is however predicted by the recurrence theorem (within an arbitrarily small error).

The same objection can be applied to the Landau damping problem: if the distribution tends to a homogeneous equilibrium, it will never return close to its initial state.

From the mathematical point of view, this reasoning clearly does not apply, since the Boltzmann equation involves an infinite number of degrees of freedom; it is only for a fixed number of particles that the recurrence theorem applies. From the physical point of view, the answer is a bit more subtle. On the one hand, the recurrence time diverges when the number N of particles tends to infinity, and this divergence is likely monstrously rapid! For a system of macroscopic size, albeit small, the recurrence theorem simply never applies, for it involves times well greater than the age of the universe. On the other hand, the validity of the Boltzmann equation is not eternal: for N fixed, the quality of the approximation will degrade with time, because chaos (simple or pre-collisional) is preserved only approximately. By the time that Poincaré recurrence takes place, the Boltzmann equation will have long ceased to be valid!![5]

10.2. Microscopic conservation of the volume

Poincaré's recurrence theorem is based on conservation of the volume in microscopic phase space (preservation of Liouville measure). The entropy is a direct function of the volume of microscopic admissible states; how can it increase if the volume of microscopic states is constant?

The answer to this question may seem surprising: it can be argued that the increase of entropy does not occur *despite* the conservation of microscopic volume, but *because* of this conservation; more precisely, it is what keeps entropy from decreasing. In fact, let us start at the initial time from all the typical configurations associated with a distribution f_i. After a time t, these typical configurations have evolved and are now associated with a distribution f_t, the transition from f_i to f_t being governed by the Boltzmann equation. The typical configurations associated with f_t are thus at least as numerous as the typical configurations associated with f_i, which clearly means that entropy cannot decrease.

In a microscopic irreversible model, we will typically have a contraction of microscopic phase space, linked to a dissipative phenomenon. The preceding argument does not apply then, and one can imagine that the entropy decreases, at least for certain initial data. It is indeed what happens, for example, in models of granular gases undergoing inelastic collisions.

[5] In real life, I think it likely that the validity of the Boltzmann equation is longer, because of slight non-Newtonian randomness, like quantum perturbations, which "renew" the equation; but this does not invalidate the reasoning.

10.3. Spontaneous appearance of time's arrow

How, starting from a microscopic equation that does not favor any time direction, can the Boltzmann equation predict an inexorable evolution toward positive times?

The answer is simple: there is not any inexorable evolution toward positive times, and the double direction of time is preserved. There is simply a particular choice of the initial data (instant of preparation of the experiment), that has fixed a particular time, say $t = 0$. Starting from there, one has a double arrow of time; entropy increases for positive times, and decreases for negative times.

10.4. Loschmidt's paradox

Loschmidt's paradox [74] formalizes the apparent contradiction that exists of a reversible microscopic dynamic and an irreversible evolution of the entropy. Let us suppose that we start from a given initial configuration and that at time t we stop the gas and reverse the velocities of all the particles. This operation does not change the entropy, and starting from this new initial data we can let the dynamic act anew. By microscopic reversibility, at the end of time $2t$ we will have returned to the point of departure; but the entropy will not have ceased to increase, whence the contradiction.

This paradox can be resolved in several ways, all of which come down to the same finding: *the degradation of the notion of chaos* between the initial time and time $t > 0$. On the mathematical level it is only known how to prove the weak convergence of $\mu_t^{1;N}$ to $f(t, \cdot)$ as $N \to \infty$, whereas the convergence is supposed uniform at the initial time. In fact, it is conjectured that the data (μ_t^N) satisfies the property (still to be defined. . .) of *pre-collisional chaos*, whereas the initial data is supposed to satisfy a complete chaos property. When the velocities are reversed, the hypothesis of pre-collisional chaos is transformed into post-collisional chaos, and the relevant equation is no longer the Boltzmann equation, but the "reverse" Boltzmann equation, in which a negative sign is placed before the collision operator. Entropy then increases toward negative times and no longer toward positive times, and all contradiction disappears.

To state matters in a more informal way: at the initial time the particles are all strangers to one another. After a time t, the particles that have just collided know each other still, while those which are about to collide do not know each other: the particles have a memory of the past and not of the future. When the velocities are reversed, the particles have a memory of the future and not of the past, and time begins to flow backwards!

Legend has it that Boltzmann, confronted with the velocity reversal paradox, responded: "Go ahead, reverse them!" Behind the jest is hidden a profound observation: reversal of velocities is an operation that is inaccessible to us because it requires microscopic knowledge of the system; and the notion of entropy emerges precisely from what we can only act upon macroscopically. Beginning in the 1950s, spin echo experiments allowed us to see the paradox from another angle [10].

10.5. Nonuniversal validity of Boltzmann's equation

This paradox is a variant of the preceding. Having understood that Boltzmann's equation does not apply after reversal of velocities, we will exploit this fact to put Herr Boltzmann in default. We redo the preceding experiment and choose as initial data the distribution obtained after reversal of velocities at time t. We let time act, and the relevant equation certainly is not Boltzmann's equation.

This paradox effectively shows that there are microscopic configurations that do not lead to Boltzmann's equation. Nevertheless, and it is thus that Boltzmann argued, these configurations are rare: precisely, they cause the appearance of correlations between pre-collisional velocities. This is not rarer than correlations between post-collisional velocities, but it is rarer than not having correlations at all! The Boltzmann equation is approximately true if we depart from a typical configuration, which is to say drawn according to a "strongly chaotic" law, but it does not hold for all initial configurations. Once these grand principles are stated, the quantitative work remains to be done.

10.6. Boltzmann's arbitrary procedure

To establish the Boltzmann equation, the encounter probabilities of particles are expressed in terms of the pre-collisional probabilities, which are arbitrary. If instead post-conditional probabilities had been used, a different equation would have been obtained, with a negative sign before the collision operator! Why then have confidence in Boltzmann?

The answer is still the same, of course, and depends on which side of the origin one is placed: for positive times, these are pre-collisional probabilities that are almost factored, whereas for negative times, these are post-collisional probabilities.

10.7. Maxwell's demon

Maxwell conceived a thought experiment in which a malicious demon positions himself in a box with two compartments and adroitly manipulates a small valve so that there is a flow of balls going from the right compartment toward the left but not the other way. The system thus evolves toward increased order, and the entropy decreases.

Of course, this cannot be considered an objection to the law of increasing entropy, and the experiment is intended to make us think: first, the demon should be part of the model and himself subjected to reversible mechanical laws, taking into account the energy needed for recognizing that a particle is approaching and for evaluating its velocity, for the mental work done, etc. If a complete account is made, we will find again, for sure, that the second law of thermodynamics is not violated.[6]

We note in this regard that recently experiments with Maxwell's demon have been realized with granular gases: as I myself saw with stupefaction in a film

[6] Maxwell's Demon has been the object of many discussions, in particular by Smoluchowski, Szilard, Gabor, Brillouin, Landauer and Bradbury; it has also inspired novelists like Pynchon. A recent paper by Binder and Danchin suggests to look for such concepts in the heart of living mechanisms.

made of an experiment, initially there is a receptacle with two vertically separated compartments and an opening above that allows communication, the two compartments are filled with inelastic particles in approximately equal number, the whole thing is agitated automatically, and little by little one of the compartments is emptied in favor of the other. An underlying principle is that in the fuller compartment the abundance of collisions results in cooling by dissipation of energy; and the particles jump less high, rendering it more and more difficult for them to leave the full compartment. We find again on this occasion the principle – already mentioned – according to which a dissipative (irreversible) dynamic does not necessarily lead to an increase of entropy, but to the opposite.

10.8. Convergence and reversibility

This paradox is a variant of the Loschmidt paradox; it applies both to the theme of Boltzmann entropic relaxation and to nonlinear Landau damping: how can there be convergence when $t \to +\infty$ if there is reversibility of the dynamic?? The answer is of childish simplicity: there is also convergence when $t \to -\infty$. For Vlasov, this was accomplished with the same equation, and we thus have a phenomenon of generalized homoclinic/heteroclinic. For Boltzmann, the equation changes according as to whether times are considered which are prior or subsequent to the chaotic data.

10.9. Stability and reversibility

This paradox is more subtle and applies to nonlinear Landau damping: asymptotic stability and reversibility of the dynamic automatically imply an instability, which seems contradictory.

We detail the argument. If we have stability in time $t \to +\infty$, let $f_\infty(v)$ be an asymptotically stable profile, which we assume to be even. We take a solution $\bar{f}(t, x, v)$, inhomogeneous, which converges toward $f_\infty(v)$. We then choose as initial data $f(T, x, -v)$ with T very large; we thus have data arbitrarily close to $f_\infty(-v) = f_\infty(v)$, and which brings us back after time T to the data $\bar{f}(0, x, v)$, rather removed from $f_\infty(v)$. In other words, the distribution f_∞ is *unstable*. How is this compatible with stability??

The answer, as explained, e.g., in [25], lies in the topology: in the theorem of asymptotic stability (nonlinear Landau damping), the convergence over large time occurs in the sense of the weak topology, with frenetic oscillations in the velocity distribution, which is compensated locally. When we say that a distribution f^0 is stable, that means that if we depart close to f^0 in the sense of the strong topology (e.g., analytic or Gevrey), then we remain close to f^0 in the sense of the weak topology. The asymptotic stability combined with the reversibility thus imply *instability in the sense of the weak topology*, which is perfectly compatible with stability in the sense of the strong topology.

10.10. Conservative relaxation

This problem is of a rather general nature. Vlasov's equation comes with a preservation of the amount of microscopic uncertainty (conservation of entropy). Moreover the distribution at time $t > 0$ allows reconstructing exactly the distribution at time $t = 0$: it suffices to solve Vlasov's equation after reversal of the velocities. We can say that Vlasov's equation forgets nothing; but convergence consists precisely in forgetting the episodes of the dynamic evolution!

The answer again lies in the weak convergence and the oscillations. Information will be lodged in these oscillations, information which is invisible because in practice we never measure the complete distribution function, but averages of this distribution function (recall the quote of Lynden-Bell reproduced at the end of Section 7.2). Every observable will converge toward its limit value, and there will be a "forgetfulness". The force field, obtained as mean of the kinetic distribution, converges toward 0 without this being contradictory to preservation of information: the information leaves the spatial variables so as to go into the kinetic variables. In particular, the spatial entropy $\int \rho \log \rho$ (where $\rho = \int f\, dv$) tends toward 0, whereas the total kinetic entropy $\int f \log f$ is conserved (but does not converge! Information is conserved for all time, but because of the weak convergence there is a loss of information in the passage to the limit $t \to \infty$).

Similarly, in nonlinear Landau damping, the energy of interaction – which is $\int W(x-y)\, \rho(x)\, \rho(y)\, dx\, dy$ – tends toward 0, and it is converted into kinetic energy (which can increase or decrease as a function of the interaction).

10.11. The echo experiment

In this famous experiment [77, 78] a plasma, prepared in a state of equilibrium, is excited at the initial time by a spatial frequency impulse k. At the end of a time τ, after relaxation of the plasma, it is excited anew by a spatial frequency ℓ, collinear and in the direction opposite to k, and of greater amplitude. We then wait and observe spontaneous response from the electric field of the plasma, called **echo**, which is produced with spatial frequency $k + \ell$ and around the time $t_e = (|\ell|/|k + \ell|)\tau$.

This experiment shows that the kinetic distribution of the plasma has kept track of past impulses: even if the force field has died off to the point of becoming negligible, the kinetic oscillations of the distribution remain present and evolve over the course of time. The first impulse subsists in the form of very rapid oscillations of period $(|k|t)^{-1}$, the second in the form of oscillations of period $(|\ell|(t-\tau))^{-1}$. A calculation, found, e.g., in [110, Section 7.3] shows that the distribution continues to oscillate rapidly in velocity, and the associated force remains negligible, up until the two trains of oscillations compensate almost exactly, which is manifested by an echo.

References

[1] Alexandre, R., Desvillettes, L., Villani, C. and Wennberg, B.: Entropy dissipation and long-range interactions. Arch. Rat. Mech. Anal. **152**, 327–355 (2000).

[2] Alexandre, R. and Villani, C.: On the Boltzmann equation for long-range interaction and the Landau approximation in plasma physics. Comm. Pure Appl. Math. **55**, 1, 30–70 (2002).

[3] Alexandre, R. and Villani, C.: On the Landau approximation in plasma physics. Ann. Inst. H. Poincaré Anal. Non Linéaire **21**, 1, 61–95 (2004).

[4] Ambrosio, L. and Gangbo, W.: Hamiltonian ODE's in the Wasserstein space of probability measures. Comm. Pure Appl. Math. **51**, 18–53 (2007).

[5] Arkeryd, L., Esposito, R. and Pulvirenti, M.: The Boltzmann equation for weakly inhomogeneous data. Comm. Math. Phys. **111**, 3, 393–407 (1987).

[6] Backus, G.: Linearized plasma oscillations in arbitrary electron distributions. J. Math. Phys. **1**, 178–191, 559 (1960).

[7] Balescu, R.: Irreversible processes in ionized gases. Phys. Fluids **3**, 52–63 (1960).

[8] Balescu, R. *Statistical Mechanics of Charged Particles.* Wiley-Interscience, New York, 1963.

[9] Balian, R.: Entropy, a protean concept. Poincaré Seminar 2, 119–145 (2003). Birkhäuser Verlag, Basel.

[10] Balian, R.: Information in statistical physics. Studies in Hist. and Phil. of Mod. Phys. **36**, 323–353 (2005).

[11] Baranger, C. and Mouhot, C.: Explicit spectral gap estimates for the linearized Boltzmann and Landau operators with hard potentials. Rev. Matem. Iberoam. **21**, 819–841 (2005).

[12] Batt, J. and Rein, G.: Global classical solutions of the periodic Vlasov–Poisson system in three dimensions. C. R. Acad. Sci. Paris Sér. I Math. **313**, 6, 411–416 (1991).

[13] Ben Arous, G. and Zeitouni, O.: Increasing propagation of chaos for mean field models. Ann. Inst. H. Poincaré Probab. Statist. **35**, 1, 85–102 (1999).

[14] Bernstein, I.B., Greene, J.M. and Kruskal, M.D.: Exact nonlinear plasma oscillations. Phys. Rev. **108**, 3, 546–550 (1957).

[15] Binney, J. and Tremaine, S.: *Galactic Dynamics*, 2nd ed. Princeton Series in Astrophysics. Princeton University Press, 2008.

[16] Bobylev, A.V. and Cercignani, C.: On the rate of entropy production for the Boltzmann equation. J. Statist. Phys. **94**, 3–4, 603–618 (1999).

[17] Bobylev, A.V. and Cercignani, C.: Exact eternal solutions of the Boltzmann equation. J. Statist. Phys. **106**, 5–6, 1019–1038 (2002).

[18] Bobylev, A.V. and Cercignani, C.: Self-similar solutions of the Boltzmann equation and their applications. J. Statist. Phys. **106**, 5–6, 1039–1071 (2002).

[19] Bogoliubov, N.N.: Problems of a dynamical theory in statistical physics. Studies in Statistical Mechanics 1, 1–118 (1962). Translation from the 1946 Russian original.

[20] Bolley, F., Guillin, A. and Villani, C.: Quantitative concentration inequalities for empirical measures on non-compact spaces. Probab. Theory Related Fields **137**, 3–4, 541–593 (2007).

[21] Boltzmann, L.: Weitere Studien über das Wärmegleichgewicht unter Gasmolekülen. *Sitzungsberichte Akad. Wiss.* 66 (1872), 275–370. Traduction : Further studies on the thermal equilibrium of gas molecules, in Kinetic Theory **2**, 88–174, Éd. S.G. Brush, Pergamon, Oxford (1966).

[22] Boltzmann, L.: *Lectures on gas theory.* University of California Press, Berkeley, 1964. English translation by Stephen G. Brush. Reprint of the 1896–1898 edition. Dover Publications, 1995.

[23] Bouchut, F.: Introduction à la théorie mathématique des équations cinétiques. In Session "L'Etat de la Recherche" de la SMF (1998). "Kinetic equations and asymptotic theory", F. Bouchut, F. Golse, M. Pulvirenti, coll. "Series in Appl. Math.", Gauthier-Villars, 2000.

[24] Braun, W. and Hepp, K.: The Vlasov dynamics and its fluctuations in the $1/N$ limit of interacting classical particles. Commun. Math. Phys. **56**, 125–146 (1977).

[25] Caglioti, E. and Maffei, C.: Time asymptotics for solutions of Vlasov–Poisson equation in a circle. J. Statist. Phys. **92**, 1–2, 301–323 (1998).

[26] Carleman, T.: Sur la théorie de l'équation intégrodifférentielle de Boltzmann. Acta Math. **60**, 369–424 (1932).

[27] Carleman, T.: *Problèmes Mathématiques dans la Théorie Cinétique des Gaz.* Almqvist & Wiksell, 1957.

[28] Carlen, E.A. and Carvalho, M.C.: Strict entropy production bounds and stability of the rate of convergence to equilibrium for the Boltzmann equation. J. Statist. Phys. **67**, 3–4, 575–608 (1992).

[29] Carlen, E.A., Carvalho, M.C. and Gabetta, E.: Central limit theorem for Maxwellian molecules and truncation of the Wild expansion. Comm. Pure Appl. Math. **53**, 370–397 (2000).

[30] Carlen, E.A., Carvalho, M.C. and Gabetta, E.: On the relation between rates of relaxation and convergence of Wild sums for solutions of the Kac equation. J. Funct. Anal. **220**, 362–387 (2005).

[31] Carlen, E.A., Carvalho, M.C. and Loss, M.: Determination of the spectral gap for Kac's master equation and related stochastic evolution. Acta Math. **191**, 1–54 (2003).

[32] Carlen, E.A., Carvalho, M.C., Le Roux, J., Loss, M. and Villani, C. Entropy and chaos in the Kac model. *Kinet. Relat. Models* **3**, 1, 85–122 (2010).

[33] Carlen, E.A., Gabetta, E. and Toscani, G.: Propagation of smoothness and the rate of exponential convergence to equilibrium for a spatially homogeneous Maxwellian gas. Comm. Math. Phys. **199**, 3, 521–546 (1999).

[34] Carlen, E.A. and Lu, M.: Fast and slow convergence to equilibrium Maxwellian molecules via Wild sums J. Stat. Phys. **112**, 1–2, 59–134 (2003).

[35] Carleson, L.: Some analytic problems related to statistical mechanics. In Euclidean Harmonic Analysis, J.J. Benedetto, éd. (Univ. of Maryland, 1979), Lect. Notes Math. 779, 5–45.

[36] Cercignani, C.: On the Boltzmann equation with cutoff potentials. Phys. Fluids **10**, 2097 (1967).

[37] Cercignani, C.: On the Boltzmann equation for rigid spheres. Transport Theory Statist. Phys. **2**, 3, 211–225 (1972).

[38] Cercignani, C.: *The Boltzmann equation and its applications.* Springer-Verlag, New York, 1988.

[39] Cercignani, C.: *Rarefied gas dynamics.* Cambridge University Press, Cambridge, 2000. From basic concepts to actual calculations.

[40] Cercignani, C., Illner, R. and Pulvirenti, M.: *The Mathematical Theory of Dilute Gases.* Springer, New York, 1994.

[41] Cover, T.M. and Thomas, J.A.: *Elements of information theory.* John Wiley & Sons Inc., New York, 1991. A Wiley-Interscience Publication.

[42] Delcroix, J.-L. and Bers, A.: *Physique des plasmas* (2 Vol.). InterEditions/CNRS Éditions, 1994.

[43] Dembo, A. and Zeitouni, O.: *Large deviations techniques and applications*, 2nd ed. Springer-Verlag, New York, 1998.

[44] Desvillettes, L.: Entropy dissipation rate and convergence in kinetic equations. Comm. Math. Phys. **123**, 4, 687–702 (1989).

[45] Desvillettes, L. and Villani, C.: On the spatially homogeneous Landau equation for hard potentials. Part I: existence, uniqueness and smoothness. Comm. Partial Differential Equations **25**, 1–2, 179–259 (2000).

[46] Desvillettes, L. and Villani, C.: On the trend to global equilibrium for spatially inhomogeneous kinetic systems: the Boltzmann equation. Invent. Math. **159**, 2, 245–316 (2005).

[47] DiPerna, R. and Lions, P.-L.: On the Cauchy problem for the Boltzmann equation: Global existence and weak stability. Ann. of Math. (2) **130**, 312–366 (1989).

[48] Dobrušin, R.L.: Vlasov equations. Funktsional. Anal. i Prilozhen. **13**, 2, 48–58, 96 (1979).

[49] Dudley, R.M.: *Real analysis and probability*, vol. 74 of Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge, 2002. Revised edition of the 1989 original.

[50] Filbet, F., Mouhot, C. and Pareschi, L.: Solving the Boltzmann equation in $N \log_2 N$. SIAM J. Sci. Comput. **28**, 3, 1029–1053 (2006).

[51] Glassey, R.T.: *The Cauchy problem in kinetic theory.* Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1996.

[52] Glassey, R. and Schaeffer, J.: Time decay for solutions to the linearized Vlasov equation. Transport Theory Statist. Phys. **23**, 4, 411–453 (1994).

[53] Glassey, R. and Schaeffer, J.: On time decay rates in Landau damping. Comm. Partial Differential Equations **20**, 3–4, 647–676 (1995).

[54] Golse, F.: From kinetic to macroscopic models. In Session "L'Etat de la Recherche" de la SMF (1998), "Kinetic equations and asymptotic theory", F. Bouchut, F. Golse, M. Pulvirenti, coll. "Series in Appl. Math.", Gauthier-Villars, 2000.

[55] Golse, F. and Saint-Raymond, L.: The Navier–Stokes limit of the Boltzmann equation for bounded collision kernels. Invent. Math. **155**, 1, 81–161 (2004).

[56] Grad, H.: On Boltzmann's H-theorem. J. Soc. Indust. Appl. Math. **13**, 1, 259–277 (1965).

[57] Grad, H.: Principles of the kinetic theory of gases. In Flügge's Handbuch der Physik, vol. XII. Springer-Verlag, 1958, pp. 205–294.

[58] Gualdani, M.P., Mischler, S. and Mouhot, C.: Factorization for non-symmetric operators and exponential H-theorem. Preprint, available online at http ://hal.archives-ouvertes.fr/ccsd-00495786

[59] Guo, Y.: The Landau equation in a periodic box. Commun. Math. Phys. **231**, 391–434 (2002).

[60] Guo, Y. and Strain, R.M.: Exponential decay for soft potentials near Maxwellian. Arch. Ration. Mech. Anal. **187**, 2, 287–339 (2008).

[61] Horst, E.: On the asymptotic growth of the solutions of the Vlasov–Poisson system. Math. Methods Appl. Sci. **16**, 75–86 (1993).

[62] Illner, R. and Pulvirenti, M.: Global validity of the Boltzmann equation for a two-dimensional rare gas in vacuum. Comm. Math. Phys. **105**, 2, 189–203 (1986). "Erratum and improved result", Comm. Math. Phys. 121, 1, 143–146 (1989).

[63] Illner, R. and Pulvirenti, M.: A derivation of the BBGKY-hierarchy for hard sphere particle systems. Transport Theory Statist. Phys. **16**, 7, 997–1012 (1987).

[64] Isichenko, M.: Nonlinear Landau damping in collisionless plasma and inviscid fluid. Phys. Rev. Lett. **78**, 12, 2369–2372 (1997).

[65] Jabin, P.-E. and Hauray, M.: N-particles approximation of the Vlasov equations with singular potential. Arch. Ration. Mech. Anal. **183**, 3, 489–524 (2007).

[66] Janvresse, E.: Spectral gap for Kac's model of Boltzmann equation. Ann. Probab. **29**, 1, 288–304 (2001).

[67] Kac, M.: Foundations of kinetic theory. In Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955, vol. III (Berkeley and Los Angeles, 1956), University of California Press, pp. 171–197.

[68] Landau, L.D.: Die kinetische Gleichung für den Fall Coulombscher Wechselwirkung. *Phys. Z. Sowjet* 10 (1936), 154. English translation: The transport equation in the case of Coulomb interactions, in Collected papers of L.D. Landau, edited and with an introduction by D. ter Haar, Pergamon Press, 1965, pp. 163–170.

[69] Landau, L.D.: On the vibration of the electronic plasma. J. Phys. USSR **10**, 25 (1946). English translation in JETP 16, 574. Reprinted in Collected papers of L.D. Landau, edited with an introduction by D. ter Haar, Pergamon Press, 1965, pp. 445–460; and in Men of Physics: L.D. Landau, Vol. 2, Pergamon Press, D. ter Haar, ed. (1965).

[70] Lanford, O.E.: Time evolution of large classical systems. In Dynamical systems, theory and applications (Rencontres, Battelle Res. Inst., Seattle, Wash., 1974). Springer, Berlin, 1975, pp. 1–111. Lecture Notes in Phys., Vol. 38.

[71] Lenard, A.: On Bogoliubov's kinetic equation for a spatially homogeneous plasma. Ann. Physics **10**, 390–400 (1960).

[72] Lifshitz, E.M. and Pitaevskiĭ, L.P.: *Course of theoretical physics* [*"Landau-Lifshits"*]. *Vol.* 10. Pergamon Press, Oxford, 1981. Translated from the Russian by J.B. Sykes and R.N. Franklin.

[73] Lin, Z. and Zeng, C.: BGK waves and nonlinear Landau damping. Preprint, 2010.

[74] Liverani, C. and Olla, S.: Toward the Fourier law for a weakly interacting anharmonic crystal. Preprint, 2010.

[75] Loschmidt, J.: Über den Zustand des Wärmegleichgewichtes eines Systems von Körpern mit Rücksicht auf die Schwerkraft. Wien. Ber. **73**, 128 (1876).

[76] Lynden-Bell, D.: The stability and vibrations of a gas of stars. Mon. Not. R. astr. Soc. **124**, 4, 279–296 (1962).

[77] Lynden-Bell, D.: Statistical mechanics of violent relaxation in stellar systems. Mon. Not. R. astr. Soc. **136**, 101–121 (1967).

[78] Malmberg, J. and Wharton, C.: Collisionless damping of electrostatic plasma waves. Phys. Rev. Lett. **13**, 6, 184–186 (1964).

[79] Malmberg, J., Wharton, C., Gould, R. and O'Neil, T.: Plasma wave echo experiment. Phys. Rev. Letters **20**, 3, 95–97 (1968).

[80] Maslen, D.: The eigenvalues of Kac's master equation. Math. Zeit. **243**, 291–331 (2003).

[81] Maxwell, J.C.: On the dynamical theory of gases. Philos. Trans. Roy. Soc. London Ser. A **157**, 49–88 (1867).

[82] McKean, H.J.: Speed of approach to equilibrium for Kac's caricature of a Maxwellian gas. Arch. Rational Mech. Anal. **21**, 343–367 (1966).

[83] Mouhot, C.: Rate of convergence to equilibrium for the spatially homogeneous Boltzmann equation with hard potentials. Comm. Math. Phys. **261**, 629–672 (2006).

[84] Mouhot, C. and Strain, R.M.: Spectral gap and coercivity estimates for linearized Boltzmann collision operators without angular cutoff. J. Math. Pures Appl. **87**, 5, 515–535 (2007).

[85] Mouhot, C. and Villani, C.: Regularity theory for the spatially homogeneous Boltzmann equation with cut-off. Arch. Rational Mech. Anal. **173**, 2, 169–212 (2004).

[86] Mouhot, C. and Villani, C.: On Landau damping. To appear in Acta Math. Available on line at arXiv:0904.2760

[87] O'Neil, T.: Collisionless damping of nonlinear plasma oscillations. Phys. Fluids **8**, 12, 2255–2262 (1965).

[88] Neunzert, H.: An introduction to the nonlinear Boltzmann–Vlasov equation. In Kinetic theories and the Boltzmann equation, C. Cercignani, ed., vol. 1048 of Lecture Notes in Math., Springer, Berlin, Heidelberg, 1984, pp. 60–110.

[89] Penrose, O.: Electrostatic instability of a non-Maxwellian plasma. Phys. Fluids **3**, 258–265 (1960).

[90] Poincaré, H.: Le mécanisme et l'expérience. Revue de Métaphysique et de Morale **I**, 534–537 (1893).

[91] Pulvirenti, M.: From particle to transport equations. In Session "L'État de la Recherche" de la SMF (1998). "Kinetic equations and asymptotic theory", F. Bouchut, F. Golse, M. Pulvirenti, coll. "Series in Appl. Math.", Gauthier-Villars, 2000.

[92] Robert, R.: Statistical mechanics and hydrodynamical turbulence. In Proceedings of the International Congress of Mathematicians, Vol. 1, 2 (Zürich, 1994) (Basel, 1995), Birkhäuser, pp. 1523–1531.

[93] Ryutov, D.D.: Landau damping: half a century with the great discovery. Plasma Phys. Control. Fusion **41**, A1–A12 (1999).

[94] Saint-Raymond, L.: *Hydrodynamic limits of the Boltzmann equation.* Lectures at SISSA, Trieste (2006), Springer Monograph, to appear.

[95] Shannon, C.E. and Weaver, W.: *The mathematical theory of communication.* University of Illinois Press, Urbana, Illinois, 1949.

[96] Spohn, H.: *Large scale dynamics of interacting particles.* Texts and Monographs in Physics. Springer-Verlag, Berlin, 1991.

[97] Strain, R.M.: On the linearized Balescu–Lenard equation. Comm. Partial Differential Equations **32**, 1551–1586 (2007).

[98] Sturm, K.-Th.: Entropic measure on multidimensional spaces. In Stochastic analysis, random fields and applications VI, Dalang, R., Dozzi, M., and Russo, F., eds. Progress in Probability, Birkhäuser.

[99] Sznitman, A.-S.: Equations de type de Boltzmann, spatialement homogènes. Z. Wahrsch. Verw. Gebiete **66**, 559–562 (1984).

[100] Sznitman, A.-S.: *Topics in propagation of chaos.* In École d'Été de Probabilités de Saint-Flour XIX (1989), Lect. Notes in Math. 1464 (1991), Springer, Berlin, pp. 165–251.

[101] Tanaka, H.: An inequality for a functional of probability distributions and its application to Kac's one-dimensional model of a Maxwellian gas. Z. Wahrscheinlichkeitstheorie und Verw. Gebiete **27**, 47–52 (1973).

[102] Toscani, G. and Villani, C.: Sharp entropy dissipation bounds and explicit rate of trend to equilibrium for the spatially homogeneous Boltzmann equation. Comm. Math. Phys. **203**, 3, 667–706 (1999).

[103] Turkington, B.: Statistical equilibrium measures and coherent states in two-dimensional turbulence. Comm. Pure Appl. Math. **52**, 7, 781–809 (1999).

[104] Villani, C.: *A review of mathematical topics in collisional kinetic theory.* In Handbook of Mathematical Fluid Dynamics I, 71–305, S. Friedlander et D. Serre, éds, North-Holland, Amsterdam, 2002.

[105] Villani, C. Cercignani's conjecture is sometimes true and always almost true. Commun. Math. Phys. **234**, 455–490 (2003).

[106] Villani, C.: Mathematics of granular materials. J. Stat. Phys. **124**, no. 2–4, 781–822 (2006).

[107] Villani, C. H-Theorem and beyond: Boltzmann's entropy in today's mathematics. Conference proceedings "Boltzmann's legacy", Erwin-Schrödinger Institute (Vienna, July 2007).

[108] Villani, C.: Hypocoercive diffusion operators. Text of my lecture International Congress of Mathematicians (Madrid, August 2006).

[109] Villani, C.: Entropy production and convergence to equilibrium. Notes pour une série de cours à l'Institut Henri Poincaré, Paris (automne 2001). In Entropy methods for the Boltzmann equation, 1–70, Lecture Notes in Math., Vol. 1916, Springer, Berlin, 2008.

[110] Villani, C.: *Hypocoercivity.* Mem. Amer. Math. Soc. **202**, no. 950 (2009).

[111] Villani, C.: *Landau damping.* Lecture notes, CEMRACS 2010. Available at www.math.univ-lyon1.fr/~villani

[112] Villani, C.: Is there any backward solution of the Boltzmann equation? Unpublished work, available at www.math.univ-lyon1.fr/~villani

[113] Vlasov, A.A.: On the oscillation properties of an electron gas. Zh. Èksper. Teoret. Fiz. **8**, 291–318 (1938).

[114] Zermelo, E.: Über einen Satz der Dynamik und die mechanische Wärmetheorie. Annalen der Physik **54**, 485–494 (1896).

Translated by Lester Senechal

Cédric Villani
Université de Lyon
& Institut Henri Poincaré
11 rue Pierre et Marie Curie
F-75231 Paris Cedex 05, FRANCE

Time, 81–143

Poincaré Seminar 2010

(Ir)réversibilité et entropie

Cédric Villani

Résumé. Dans ce texte j'évoque la question du temps en physique statistique classique, en insistant sur le problème de l'écoulement du temps et de l'irréversibilité. L'entropie de Boltzmann et le Théorème H occupent une place de choix, ainsi que l'amortissement sans perte d'information deviné par Landau. Certains paradoxes plus ou moins célèbres sont analysés des points de vue physique et mathématique.

La cosa più meravigliosa è la felicità del momento
L. Ferré

La **flèche du temps** fait partie de notre expérience sensible et nous en faisons l'expérience chaque jour : les miroirs brisés ne se recollent pas, les êtres humains ne rajeunissent pas, et les cernes croissent sans cesse dans les troncs des arbres… En somme, le temps s'écoule toujours dans le même sens ! Pourtant, les lois fondamentales de la physique classique ne privilégient aucune direction du temps et obéissent à une rigoureuse symétrie entre passé et futur. Il est possible, comme discuté dans l'article de T. Damour dans ce même volume, que l'irréversibilité soit inscrite dans d'autres lois de la physique, par exemple du côté de la relativité générale ou de la mécanique quantique. Depuis Boltzmann, la **physique statistique** avance une autre explication : la flèche du temps traduit un flot constant des événements moins probables vers les événements plus probables. Avant de continuer sur cette interprétation qui constitue le fil directeur de tout l'exposé, je noterai que l'écoulement du temps n'est pas forcément basé sur une explication unique.

Au premier examen, la suggestion de Boltzmann semble saugrenue : ce n'est pas parce qu'un événement est *probable* qu'il va se réaliser *effectivement*, or la flèche du temps semble inexorable et ne tolérer aucune exception. La réponse à cette objection tient dans un slogan : la **séparation d'échelles**. Si les lois fondamentales de la physique s'exercent au niveau microscopique, particulaire (atomes, molécules…), les phénomènes que nous pouvons sentir ou mesurer mettent en jeu un nombre considérable de particules. L'effet de ce nombre est d'autant plus grand qu'il intervient dans des calculs combinatoires : si N, le nombre d'atomes participant à une expérience, est de l'ordre de 10^{10}, c'est déjà considérable, mais $N!$ ou 2^N sont des nombres surnaturellement grands, invincibles.

Les innombrables débats entre physiciens qui se sont ensuivis pendant plus d'un siècle, et se poursuivent encore aujourd'hui, témoignent de la subtilité et de la profondeur des arguments de Maxwell et Boltzmann, porte-drapeaux d'une petite révolution scientifique qui s'accomplit dans les années 1860 et 1870, et qui vit naître les fondements de la théorie cinétique des gaz moderne, le concept universel d'entropie statistique, et la notion d'irréversibilité macroscopique. À dire vrai, les arguments étaient si subtils que Maxwell et Boltzmann s'y sont eux-mêmes parfois perdus, hésitant sur certaines interprétations, alternant les erreurs naïves avec les concepts profonds ; les plus grands scientifiques de la fin du dix-neuvième siècle, comme Poincaré ou Lord Kelvin, n'ont pas été en reste. On trouvera un aperçu de ces atermoiements dans le texte de Damour déjà cité ; pour ma part je me contenterai de présenter une version "décantée" de la théorie de Boltzmann. On évoquera à la fin de ce texte la façon dont Landau fit voler en éclat le paradigme de Boltzmann, découvrant une apparente irréversibilité là où il ne semblait pas y en avoir, et ouvrant une nouvelle mine de problèmes mathématiques.

En retraçant l'histoire de l'interprétation statistique de la flèche du temps, nous aurons l'occasion d'effectuer un voyage au cœur de problèmes profonds qui depuis plus d'un siècle agitent mathématiciens et physiciens.

Les notations utilisées dans cet exposé sont dans l'ensemble classiques ; je noterai $\mathbb{N} = \{1, 2, 3, \ldots\}$ et log = logarithme népérien.

1. Le royaume inaccessible de Newton

On va adopter ici une description purement classique de notre univers physique, selon les lois édictées par Newton : l'espace ambient est euclidien, le temps absolu, et l'accélération est égale au produit de la masse par la résultante des forces.

Dans le cas de la description d'un gaz, ces hypothèses sont discutables : d'après E.G.D. Cohen, les fluctuations quantiques ne sont pas négligeables au niveau mésoscopique. La nature probabiliste de la mécanique quantique est toujours débattue ; admettons cependant que l'incertitude accrue qui résulterait de la prise en compte de ces fluctuations ne puisse qu'arranger nos affaires, au moins qualitativement, et concentrons-nous donc sur des modèles classiques et déterministes, "à la Newton".

1.1. Le modèle des sphères dures

Pour fixer les idées, considérons un système de particules sphériques idéales rebondissant les unes sur les autres : soient N particules dans une boîte Λ, on désigne par $X_i(t)$ la position au temps t du centre de la particule numérotée i. Les règles du mouvement s'énoncent comme suit :

- On suppose qu'initialement les particules sont bien séparées ($i \neq j \Longrightarrow |X_i - X_j| > 2r$) et séparées de la paroi ($d(X_i, \partial\Lambda) > r$ pour tout i).
- Tant que ces conditions de séparation sont satisfaites, le mouvement est rectiligne uniforme : $\ddot{X}_i(t) = 0$ pour tout i, où l'on note $\ddot{X} = d^2X/dt^2$ l'accélération de X.
- Quand deux particules se rencontrent, leurs vitesses changent brutalement selon les lois de Descartes : si $|X_i(t) - X_j(t)| = 2r$, alors

$$\begin{cases} \dot{X}_i(t^+) = \dot{X}_i(t^-) - 2\Big\langle \dot{X}_i(t^-) - \dot{X}_j(t^-), n_{ij} \Big\rangle n_{ij}, \\ \dot{X}_j(t^+) = \dot{X}_j(t^-) - 2\Big\langle \dot{X}_j(t^-) - \dot{X}_i(t^-), n_{ji} \Big\rangle n_{ji}, \end{cases}$$

  où l'on note $n_{ij} = (X_i - X_j)/|X_i - X_j|$ le vecteur unitaire joignant les centres des boules en collision.
- Quand une particule rencontre le bord, sa vitesse change aussi : si $|X_i - x| = r$ avec $x \in \partial\Lambda$, alors

$$\dot{X}_i(t^+) = \dot{X}_i(t^-) - 2\big\langle \dot{X}_i(t), n(x) \big\rangle\, n(x),$$

  où $n(x)$ est la normale extérieure à Λ en x, supposée bien définie.

Ces règles ne sont pas suffisantes pour définir complètement la dynamique : on ne peut a priori exclure les possibilités de collision triple, collisions simultanées entre particules et le bord, ou encore d'une infinité de collisions se produisant en temps fini. Cependant ces événements sont de probabilité nulle si les conditions initiales sont tirées au hasard selon la mesure de Lebesgue (ou mesure de Liouville) dans l'espace des phases [40, Appendice 4.A] ; on négligera donc ces éventualités. La dynamique ainsi définie, toute simple qu'elle soit, peut alors être considérée comme une caricature de notre univers complexe si le nombre N de particules est très grand. Étudiée depuis plus d'un siècle, cette caricature n'a pas encore livré tous ses secrets, il s'en faut de beaucoup.

1.2. Autres modèles newtoniens

À partir du modèle emblématique des sphères dures, on peut définir un certain nombre de variantes plus ou moins complexes :

- remplacer la dimension 3 par une dimension $d \geq 2$ arbitraire (la dimension 1 étant probablement pathologique) ;
- remplacer la condition aux limites (rebond élastique sur les parois) par une loi plus complexe [40, Chapitre 8] ;
- ou au contraire éliminer les bords, toujours délicats, en posant le système dans l'espace entier $\mathbb{R}^d$ (mais on peut alors argumenter que le nombre de particules devrait être infini pour conserver une densité moyenne globale non nulle) ou dans un tore de côté L, $\mathbb{T}^d_L = \mathbb{R}^d/(L\mathbb{Z}^d)$, ce qui sera mon choix préférentiel dans la suite ;
- remplacer l'interaction de contact des sphères dures par une autre interaction entre particules pontuelles, par exemple associée à un potentiel d'interaction

à deux corps $\phi(x-y)$ = potentiel exercé au point x par un point matériel situé en y.

Parmi les potentiels d'interaction notables, on mentionne, en dimension 3, à constante multiplicative près :

- le potentiel **coulombien** : $\phi(x-y) = 1/|x-y|$;
- le potentiel **newtonien** : $\phi(x-y) = -1/|x-y|$;
- le potentiel **maxwellien** : $\phi(x-y) = 1/|x-y|^4$.

L'interaction maxwellienne a été introduite artificiellement par Maxwell et Boltzmann dans le cadre de l'étude statistique des gaz ; elle mène à d'importantes simplifications dans certaines formules. Il existe une zoologie d'autres potentiels (Lennard-Jones, Manev ...). Les sphères dures correspondent au cas limite d'un potentiel qui vaudrait 0 pour $|x-y| > r$ et $+\infty$ pour $|x-y| < 2r$.

Supposant plus généralement que l'interaction se fait sur une échelle d'ordre r et avec l'intensité a, on arrive à un **système de particules ponctuelles avec potentiel d'interaction** :

$$\ddot{X}_i(t) = -a \sum_{j \neq i} \nabla\phi\left(\frac{X_i - X_j}{r}\right), \tag{1}$$

pour tout $i \in \{1, \ldots, N\}$; on suppose donc $X_i \in \mathbb{T}^d_L$. Ici encore, la dynamique est bien définie en-dehors d'un ensemble de conditions initiales exceptionnelles, et associée à un **flot newtonien** : $\mathcal{N}_t$ qui à la configuration au temps s associe la configuration au temps $s+t$ ($t \in \mathbb{R}$ peut être positif ou négatif).

1.3. Fonctions de distribution

Même si l'on accepte le modèle newtonien (1), il nous reste *inaccessible* : d'abord parce que nous ne percevons pas les particules individuellement (trop petites), et parce que leur nombre N est considérable. Avec des expériences bien choisies, on peut mesurer la pression exercée sur une petite surface, la température autour d'un point, la densité moyenne, etc. Toutes ces quantités ne s'expriment pas directement en fonction des X_i, mais plutôt en fonction de *quantités moyennes*

$$\frac{1}{N} \sum_i \chi(X_i, \dot{X}_i), \tag{2}$$

où χ est une fonction scalaire.

La distinction peut sembler oiseuse : en particularisant χ, en le concentrant près de la particule i, on retrouve l'information manquante. Mais bien évidemment, cela est impossible : en pratique χ est à variation *macroscopique*, par exemple de l'ordre de la taille de la boîte. En outre, l'information contenue dans les moyennes (2) ne distingue pas les particules, de sorte que l'on a remplacé le vecteur des $(X_i, \dot{X}_i)$ par la **mesure empirique**

$$\hat{\mu}^N_t = \frac{1}{N} \sum_{i=1}^N \delta_{(X_i(t), \dot{X}_i(t))}. \tag{3}$$

La terminologie "empirique" est bien choisie : c'est la mesure que l'on observe au moyen (sans jeu de mots) de mesures.

Pour résumer : notre connaissance du système de particules s'effectue uniquement à travers le comportement de la mesure empirique dans une *topologie faible* qui modélise la limitation macroscopique de nos expériences – aussi bien expériences de laboratoire qu'expériences sensibles.

Très souvent, à notre échelle, la mesure empirique apparaît continue :

$$\hat{\mu}_t^N(dx\,dv) \simeq f(t,x,v)\,dx\,dv.$$

On notera souvent $f(t,\cdot) = f_t$. La densité f est la **distribution cinétique** du gaz. L'étude de cette distribution constitue la théorie cinétique des gaz ; le fondateur de cette science est sans doute D. Bernoulli (vers 1738), et les plus fameux contributeurs en sont Maxwell et Boltzmann. On trouvera une brève histoire de la théorie cinétique dans [40, Chapitre 1] et les références incluses.

Continuons l'étude du système newtonien. On peut imaginer que certaines expériences permettent des *mesures simultanées* des paramètres de plusieurs particules, donnant ainsi accès à des corrélations entre particules. Ceci nous mène à définir par exemple

$$\hat{\mu}_t^{2;N}(dx_1\,dv_1\,dx_2\,dv_2) = \frac{1}{N(N-1)}\sum_{i\neq j}\delta_{(X_{i_1}(t),\dot{X}_{i_2}(t),X_{i_2}(t),\dot{X}_{i_2}(t))},$$

ou plus généralement

$$\hat{\mu}_t^{k;N}(dx_1\,dv_1\,\ldots\,dx_k\,dv_k) = \frac{(N-k-1)!}{N!}\sum_{(i_1,\ldots,i_k)\,\text{distincts}}\delta_{(X_{i_1}(t),\dot{X}_{i_1}(t),\ldots,X_{i_k}(t),\dot{X}_{i_k}(t))}.$$

Les approximations correspondantes sont les **fonctions de distribution à k particules** :

$$\hat{\mu}_t^{k;N}(dx_1\,dv_1\,\ldots\,dx_k\,dv_k) \simeq f^{(k)}(t,x_1,v_1,\ldots,x_k,v_k).$$

Évidemment, en continuant jusqu'à $k = N$, on trouve une mesure

$$\hat{\mu}^{N;N}(dx_1\,\ldots\,dv_N)$$

concentrée sur le vecteur des positions et vitesses des particules (moyenné sur tous les choix de permutations des particules). Mais en pratique on ne va jamais jusqu'à $k = N$: k reste très petit (aller jusqu'à 3 serait déjà un exploit), alors que N est gigantesque.

1.4. Aléatoire microscopique

Malgré le déterminisme du modèle newtonien, on a déjà fait des hypothèses de nature probabiliste sur les données initiales, en supposant qu'elles ne sont pas configurées de sorte à aboutir à une catastrophe rare telle qu'une collision triple.

On peut maintenant généraliser cette approche en considérant une distribution de probabilité sur l'ensemble des positions et vitesses initiales :

$$\mu_0^N(dx_1\,dv_1\,\ldots\,dx_N\,dv_N),$$

que l'on appellera **mesure de probabilité microscopique**. Dans la suite on notera pour abréger

$$dx^N\,dv^N := dx_1\,dv_1\,\ldots\,dx_N\,dv_N.$$

Il est naturel de choisir μ_0^N symétrique, c'est-à-dire invariante par permutation des coordonnées. La donnée μ_0^N remplace la mesure $\hat{\mu}_0^{N;N}$ et la généralise. Elle donne lieu à un flot de mesures, obtenu par l'action du flot :

$$\mu_t^N = (\mathcal{N}_t)_\# \mu_0^N,$$

et des marginales

$$\mu_t^{k;N} = \int_{(x_1,v_1,\ldots,x_k,v_k)} \mu_t^N.$$

Si le sens de la mesure empirique est transparent (c'est la “vraie” densité de particules), celui de la mesure de probabilité microscopique est moins évident. Imaginons que l'état initial a été préparé par un grand concours de circonstances dont nous ne savons pas grand chose : on peut seulement faire des suppositions et des devinettes. Ainsi μ_0^N est une mesure de probabilité sur l'ensemble des possibles configurations initiales. Un énoncé physique faisant intervenir μ_0^N n'aura donc guère de sens s'il utilise la forme précise de cette distribution (nous ne pourrons le vérifier, car nous n'observons pas μ_0^N) ; mais il en aura un s'il énonce une propriété μ_0^N-presque sûre, ou bien sûre avec μ_0^N-probabilité 0.99 ou davantage.

De même, la forme de $\mu_t^{1;N}$ n'a guère de sens physique. Mais si l'on a un phénomène de *concentration de la mesure* dû au gigantisme de N, alors on peut espérer que

$$\mu_0^N\Big[\mathrm{dist}\,\big(\hat{\mu}_t^N,\ f_t(x,v)\,dx\,dv\big) \geq r\Big] \leq \alpha(N,r),$$

où dist est une distance bien choisie sur l'espace des mesures et $\alpha(N,r) \to 0$ quand $r \to \infty$, d'autant plus vite que N est grand (par exemple $\alpha(N,r) = e^{-c\,N\,r}$). On aura alors

$$\begin{aligned}
\mathrm{dist}\,\big(\mu_t^{1;N}, f_t(x,v)\,dx\,dv\big) &= \mathrm{dist}\,\left(\int \hat{\mu}_t^N\,d\mu_t^N,\ f_t(x,v)\,dx\,dv\right)\\
&\leq \int \mathrm{dist}\,(\hat{\mu}_t^N, f_t\,dx\,dv)\\
&\leq \int_0^\infty \alpha(N,r)\,dr =: \eta(N).
\end{aligned}$$

Si $\eta(N) \to 0$ quand $N \to \infty$, il s'ensuivra que, avec très forte probabilité, $\mu_t^{1;N}$ est une excellente approximation de $f(t,x,v)\,dx\,dv$, qui lui-même est une bonne approximation de $\hat{\mu}_t^N$.

1.5. Micromégas

Dans cette section on a introduit deux descriptions statistiques très différentes : la distribution macroscopique $f(t,x,v)\,dx\,dv$, et les probabilités microscopiques $\mu_t^N(dx^N\,dv^N)$. Bien sûr, la quantité d'information contenue dans μ^N est considérablement plus importante que celle qui est contenue dans la distribution macroscopique : cette dernière nous renseigne sur l'état d'une particule typique, alors qu'un tirage selon la distribution μ_t^N nous renseigne sur l'état de *toutes* les particules. Pensons que si l'on a 10^{20} degrés de liberté, il s'agit d'intégrer sur 99999999999999999999 d'entre eux. Pour manipuler des dimensions aussi vertigineuses, nous aurons besoin d'un concept fondamental : l'entropie.

2. Le monde entropique

Le concept et le nom d'entropie ont été introduits par Clausius en 1865 comme une partie de la théorie – alors en construction – de la thermodynamique. Quelques années plus tard, Boltzmann (certainement influencé par les idées statistiques développées par Laplace, Quetelet et d'autres) révolutionnait le concept en lui donnant une interprétation statistique basée sur la théorie atomique. En plus de cette section, on pourra consulter par exemple Balian [9, 10] sur la notion d'entropie en physique statistique.

2.1. Formule de Boltzmann

Soit un système physique, que l'on suppose complètement décrit par son état microscopique $z \in \mathcal{Z}$. L'expérience ne nous donne en général accès qu'à une description partielle de l'état, disons $\pi(z) \in \mathcal{Y}$, où $\mathcal{Y}$ est un espace d'états macroscopiques. Je ne donnerai pas d'hypothèses précises sur les espaces $\mathcal{Z}$ et $\mathcal{Y}$, mais dès que l'on introduira de la théorie de la mesure on supposera implicitement que ce sont des espaces polonais (métriques séparables complets).

Comment estimer la quantité d'information perdue quand on résume l'information microscopique par l'information macroscopique ? Supposons $\mathcal{Y}$ et $\mathcal{Z}$ dénombrables, alors il est naturel de penser que l'incertitude associée à un état $y \in \mathcal{Y}$ est fonction du cardinal de la pré-image, soit $\#\pi^{-1}(y)$.

Maintenant si l'on effectue deux mesures indépendantes de deux systèmes différents, on a envie de dire que les incertitudes s'ajoutent. Or, avec des notations évidentes, $\#\pi^{-1}(y_1, y_2) = (\#\pi_1^{-1}(y_1))\,(\#\pi_2^{-1}(y_2))$. Pour passer de cette opération multiplicative à une addition, on prendra un multiple du logarithme. On aboutit ainsi à la célèbre formule de Boltzmann, gravée sur sa tombe du Cimetière central de Vienne :

$$S = k\,\log W, \tag{4}$$

où $W = \#\pi^{-1}(y)$ est le nombre d'états microscopiques compatibles avec l'état macroscopique y observée, et k est la constante de Boltzmann. [1]

1. Même si la formule (4) exprime clairement la pensée de Boltzmann, c'est Planck qui le premier l'écrivit sous cette forme particulière, vers 1900.

Dans de nombreux cas, l'espace $\mathcal{Z}$ des configurations microscopiques est continu, et pour appliquer la formule de Boltzmann il convient de remplacer la mesure de comptage par une mesure privilégiée : par exemple la mesure de Liouville si l'on s'intéresse à un système hamiltonien. Ainsi W dans (4) pourra être *volume* d'états microscopiques compatibles avec l'état macroscopique y.

Si l'espace $\mathcal{Y}$ des configurations macroscopiques est également continu, alors cette notion de volume doit être maniée avec prudence : la fibre $\pi^{-1}(y)$ est typiquement de volume nul, ce qui n'a guère d'intérêt. On peut tenter de poser, pour une topologie donnée,

$$S(y) = \text{p.f.}_{\varepsilon \to 0} \log \big| \pi^{-1}(B_\varepsilon(y)) \big|,$$

où $B_\varepsilon(y)$ est la boule de rayon ε centrée en y et p.f. désigne la partie finie, c'est à dire que l'on retranche la divergence en ε, si tant est qu'elle ait un comportement universel.

Si ce dernier point n'a en général rien d'évident, l'universalité est cependant vérifiée dans le cas particulier qui nous intéresse où l'état microscopique $\mathcal{Z}$ est l'espace des configurations de N particules, soit $\mathcal{Y}^N$, et où l'on commence par prendre *la limite* $N \to \infty$. Dans cette limite, comme on va le voir, l'entropie moyenne par particule tend vers une valeur finie, et l'on pourra ensuite prendre la limite $\varepsilon \to 0$, qui correspond à une mesure *macroscopique* arbitrairement précise. Le résultat n'est autre que la fameuse *fonction H* de Boltzmann, à un signe près.

2.2. De l'entropie à la fonction *H*

Appliquons les considérations précédentes avec un espace macroscopique constitué de k états différents : un état macroscopique est donc un vecteur $(f_1, \ldots, f_k)$ de fréquences, avec bien sûr $f_1 + \cdots + f_k = 1$. On suppose que la mesure est absolue (pas d'erreur) et que $Nf_j = N_j$ est entier pour tout j. Le nombre d'états microscopiques associés à cet état macroscopique vaut alors

$$W = \frac{N!}{N_1! \ldots N_k!}.$$

(Si l'on prépare N_j emplacements dans l'état j et que l'on numérote tous les emplacements de 1 à N, alors on a $N!$ façons de ranger les N boules dans les N emplacements, et ensuite on est incapable de distinguer entre permutations à l'intérieur d'une même boîte.)

D'après la formule de Stirling, quand $N \to \infty$ on a $\log N! = N \log N - N + \log \sqrt{2\pi N} + o(1)$. Il s'ensuit facilement que

$$\begin{aligned} \frac{1}{N} \log W &= -\sum_i \frac{N_i}{N} \log \frac{N_i}{N} + O\left(\frac{k \log N}{N}\right) \\ &= -\sum f_i \log f_i + o(1). \end{aligned}$$

On note que l'on peut aussi aboutir au même résultat sans utiliser la formule de Stirling, grâce à la méthode dite des types [41, section 12.4].

Si maintenant on augmente le nombre d'expériences, on pourra formellement faire tendre $k \to \infty$, tout en tenant compte de ce que k reste très inférieur à N. Imaginons que l'on dispose d'une mesure de référence ν sur l'espace macroscopique $\mathcal{Y}$, et que l'on puisse l'on puisse séparer cet espace $\mathcal{Y}$ en "cellules" de volume (mesure) $\delta > 0$, correspondant aux différents états. Quand $\delta \to 0$, si le système a une distribution statistique $f(y)$ par rapport à la mesure ν, on peut raisonnablement penser que $f_i \simeq \delta\, f(y_i)$ où y_i est un point représentatif de la cellule numérotée i. Mais alors

$$\sum_i f_i \log \frac{f_i}{\delta} \simeq \delta \sum_i f(y_i) \log f(y_i) \simeq \int f \log f \, d\nu,$$

où la dernière approximation provient d'un argument de somme de Riemann.

Nous avons ainsi abouti à la **fonction H de Boltzmann** : étant donnée une mesure de référence ν sur un espace $\mathcal{Y}$, et μ une mesure de probabilité sur $\mathcal{Y}$,

$$H_\nu(\mu) = \int f \log f \, d\nu, \qquad f = \frac{d\mu}{d\nu}. \tag{5}$$

Si ν est une mesure de probabilité, ou plus généralement une mesure de masse finie, il est facile d'étendre cette formule à toutes les probabilités μ, en posant $H_\nu(\mu) = +\infty$ si μ n'est pas absolument continue par rapport à ν. Si ν est une mesure de masse infinie, plus de précautions sont de mise, on pourra requérir au minimum la finitude de $\int f\,(\log f)_-\, d\nu$.

On note ensuite que si l'espace macroscopique $\mathcal{Y}$ porte une mesure ν, alors l'espace microscopique $\mathcal{Z} = \mathcal{Y}^N$ porte une mesure naturelle $\nu^{\otimes N}$.

Nous sommes maintenant mûrs pour énoncer la traduction mathématique précise de la formule de la fonction H : on se donne $\{\varphi_j\}_{j\in\mathbb{N}}$ une famille dense de fonctions bornées et uniformément continues, alors

$$\lim_{k\to\infty} \lim_{\varepsilon\to 0} \lim_{N\to\infty} \frac{1}{N} \log \nu^{\otimes N}\Bigg[\bigg\{(y_1,\ldots,y_N) \in \mathcal{Y}^N; \quad \forall j \in \{1,\ldots,k\}, \\ \left|\int \varphi_j\, d\mu - \frac{1}{N}\sum_i \varphi_j(y_i)\right| \le \varepsilon\bigg\}\Bigg] = -H_\nu(\mu). \tag{6}$$

On interprètera donc N comme le nombre de particules ; les φ_j comme une suite d'observables dont on mesure la valeur moyenne ; et ε comme la précision des mesures. Cette formule résume de manière concise l'information essentielle contenue dans la fonction H.

Si ν est une mesure de probabilité, l'énoncé (6) est connu sous le nom de **théorème de Sanov** [43] et c'est un résultat phare de la théorie des grandes déviations. Avant de donner l'interprétation de (6) dans cette théorie, je note qu'une fois que l'on sait traiter le cas où ν est une mesure de probabilité on en déduit facilement le cas où ν est une mesure de masse finie ; en revanche je n'ai pas connaissance d'une discussion rigoureuse dans le cas où ν est de masse infinie, même si on s'attend à ce que le résultat reste vrai.

2.3. Grandes déviations

Soit ν une mesure de probabilité, et supposons que l'on tire des variables aléatoires y_j selon ν, de manière indépendante.

La mesure empirique $\hat{\mu} = N^{-1} \sum \delta_{y_j}$ est alors une mesure aléatoire, convergeant presque sûrement vers ν quand $N \to \infty$ (c'est le théorème de Varadarajan, aussi appelé loi fondamentale de la statistique [49]). Il se peut bien sûr que les apparences soient trompeuses, et que l'on croie observer une mesure μ distincte de ν. Cette probabilité d'observation décroît exponentiellement avec le coefficient N, et elle est en gros proportionnelle à $\exp(-N\, H_\nu(\mu))$; autrement dit, l'entropie de Boltzmann dicte la rareté des conditions qui mènent à l'observation "inattendue" μ.

2.4. Information

La théorie de l'information est née en 1948 avec le remarquable traité de Shannon et Weaver [95] sur la "théorie de la communication". C'est maintenant un pilier de toute l'industrie de la transmission de l'information.

Dans la théorie de Shannon, quelque peu désincarnée pour pouvoir être reproduite et abordée froidement, on définit la quantité d'information apportée par le déchiffrage d'un signal aléatoire, en fonction de l'inverse de la probabilité de ce signal (ce qui est rare est précieux). L'usage du logarithme permet d'avoir la propriété d'additivité, et l'on obtient la formule de Shannon pour la quantité moyenne d'information gagnée au cours du déchiffrage : $\mathbb{E} \log(1/p(Y))$, où p est la loi de Y. Ceci redonne bien sûr la formule de Boltzmann !

2.5. Entropies à tous les étages

L'entropie n'est pas un concept intrinsèque, elle dépend de l'observateur et du degré de connaissance qu'il peut acquérir par des expériences et mesures. Il est commode de penser que l'entropie est un concept *anthropique*. Une conséquence est que la notion d'entropie va varier avec le degré de précision de la description.

L'entropie de Boltzmann, comme on l'a vu, nous informe sur la rareté de la fonction de distribution cinétique $f(x, v)$, et la quantité d'information microscopique qui reste à découvrir une fois que l'on connaît f.

Si au contraire on se donne l'état microscopique de toutes les particules microscopiques, il n'y a plus d'information cachée et donc plus d'entropie. Mais si l'on se donne une probabilité sur les configurations microscopiques, μ^N, alors le concept d'entropie prend à nouveau son sens : l'entropie sera d'autant plus basse que la probabilité μ^N sera concentrée et informative en elle-même. On se retrouve alors avec une notion d'**entropie microscopique** $S_N = -H_N$,

$$H_N = \frac{1}{N} \int f^N \log f^N \, dx^N \, dv^N,$$

qui est typiquement *conservée* par la dynamique newtonienne, comme conséquence du théorème de Liouville. On peut vérifier que

$$H_N \geq H(\mu^{1;N}),$$

avec égalité quand μ^N est un produit tensoriel et qu'il n'y a donc pas de corrélations entre particules. L'idée est que l'état des particules microscopiques est plus facile à obtenir par mesures multiparticules, que particule par particule – sauf bien sûr si les particules sont indépendantes !

Dans l'autre sens, on peut aussi se donner une distribution *moins précise* que la distribution cinétique : il s'agit typiquement d'une description hydrodynamique, qui fait seulement intervenir des champs de densité $\rho(x)$, température $T(x)$ et vitesse moyenne $u(x)$. Le passage du formalisme cinétique au formalisme hydrodynamique se fait par les formules simples

$$\rho(x) = \int f(x,v)\,dv; \qquad u(x) = \rho(x)^{-1} \int f(x,v)\,v\,dv;$$
$$T(x) = \frac{1}{d\,\rho(x)} \int f(x,v)\,\frac{|v-u(x)|^2}{2}\,dv.$$

À cette description est associée une notion d'**entropie hydrodynamique** :

$$S_h = -\int \rho \log \frac{\rho}{T^{d/2}}.$$

Cette information est toujours plus basse que l'information cinétique. Nous avons finalement une hiérarchie : d'abord l'information microscopique de bas niveau, puis l'information “mésoscopique” de la fonction de distribution à la Boltzmann, enfin l'information “macroscopique” apportée par la description hydrodynamique. Les proportions relatives de ces différentes entropies constituent d'excellents moyens d'apprécier l'état physique des systèmes considérés.

2.6. Universalité de l'entropie

Initialement introduite dans le cadre de la théorie cinétique des gaz, l'entropie est un concept mathématique abstrait et protéiforme, qui joue un rôle important dans de nombreux domaines de la physique, mais aussi dans des branches des mathématiques qui n'ont rien à voir avec la physique, ainsi qu'en informatique et dans d'autres sciences.

Certaines retombées mathématiques de la notion introduite par Boltzmann sont passées en revue dans mon texte de synthèse, *H-Theorem and beyond : Boltzmann's entropy in today's mathematics* [107].

3. Ordre et chaos

Intuitivement, un système microscopique est ordonné si toutes les particules sont agencées de manière coordonnée, *corrélée*. Au contraire, il est chaotique si les particules, n'en faisant qu'à leur tête, agissent toutes indépendamment les unes des autres. Reformulons cette idée : une distribution de particules est chaotique si les particules s'ignorent toutes les unes les autres, au sens où un gain d'information obtenu sur une particule donnée n'apporte aucun gain d'information sur une autre particule. Cette notion simple, clé de l'équation de Boltzmann, présente des subtilités importantes que nous allons brièvement évoquer.

3.1. Chaos microscopique

Il est équivalent de dire de particules aléatoires qu'elles s'ignorent, ou que leur loi jointe est un produit tensoriel. Bien sûr, même si des particules s'ignorent au temps initial, elles vont tout de suite entrer en interaction, et la propriété d'indépendance sera détruite. Dans le cas du système des sphères dures, la situation est encore pire : les particules sont obligées de tenir compte les unes des autres puisque les sphères ne peuvent s'interpénétrer. L'indépendance est donc à comprendre en un sens asymptotique quand le nombre de particules devient très grand ; et les expériences visant à mesurer le degré d'indépendance ne feront intervenir qu'un nombre fini de particules. Ceci mène naturellement à la définition qui suit.

Soit un espace macroscopique $\mathcal{Y}$ et pour tout N une mesure de probabilité μ^N sur $\mathcal{Y}^N$, supposée symétrique (invariante par permutation des coordonnées). On dit que (μ^N) est **chaotique** s'il existe une probabilité μ telle que $\mu^N \simeq \mu^{\otimes N}$ au sens de la topologie faible des mesures produits. Explicitement, cela signifie que pour tout $k \in \mathbb{N}$ et pour toutes fonctions $\varphi_1, \ldots, \varphi_k$ continues bornées sur $\mathcal{Y}$,

$$\int_{\mathcal{Y}^N} \varphi_1(y_1) \ldots \varphi_k(y_k)\, \mu^N(dy_1 \,\ldots\, dy_N) \xrightarrow[N\to\infty]{} \left(\int \varphi_1\, d\mu\right) \ldots \left(\int \varphi_k\, d\mu\right). \tag{7}$$

Bien sûr, la définition peut être quantifiée, en introduisant une notion de distsance adéquate permettant de mesurer l'écart entre μ^N et $\mu^{\otimes N}$. On pourra ainsi dire qu'une distribution μ^N est plus ou moins chaotique. On répète encore une fois : ce qui compte c'est l'indépendance d'un petit nombre k de particules prises parmi un très grand nombre N.

On montre (voir l'argument dans [100]) qu'il est équivalent d'imposer la propriété (7) pour tout $k \in \mathbb{N}$, ou simplement pour $k = 2$. Ainsi, le chaos signifie exactement que *2 particules tirées au hasard parmi N sont asymptotiquement indépendantes quand $N \to \infty$*. La démonstration passe par l'observation des liens entre chaos et mesure empirique.

3.2. Chaos et mesure empirique

Par la loi des grands nombres, le chaos impliquera automatiquement un déterminisme asymptotique : avec très forte probabilité, la mesure empirique s'approchera de la distribution statistique d'une particule arbitraire quand le nombre total de particules deviendra gigantesque.

Il se trouve que réciproquement *les corrélations s'accommodent très mal d'une prescription de densité macroscopique.* Avant de donner un énoncé précis, on va illustrer ce concept dans un cadre simplissime. Soient une boîte à deux compartiments, dans laquelle on range un très grand nombre N de boules *indistinguables.* Un état très corrélé serait un état dans lequel toutes les particules occuperaient le même compartiment : si je tire deux boules au hasard, l'état de la première boule m'informe complètement sur l'état de la seconde. Mais bien sûr, à partir du moment où les quantités respectives de boules dans les compartiments sont fixées et toutes deux non nulles, un tel état de corrélation est impossible. En fait, si les particules sont indistinguables, quand on en tire deux au hasard, la seule information que l'on gagne est obtenue en exploitant le fait qu'elles sont distinctes, de sorte que la connaissance de l'état de la première particule réduit légérement le nombre de possibilités pour l'état de la seconde. Ainsi, si la première particule occupe l'état 1, alors les chances de trouver la seconde particule dans l'état 1 ou 2 respectivement ne sont pas $f_1 = N_1/N$ et $f_2 = N_2/N$, mais $f_1' = (N_1 - 1)/(N - 1)$ et $f_2 = N_2/(N - 1)$. La distribution jointe d'un couple de particules est donc très proche de la loi produit.

En développant l'argument précédent, on arrive à un résultat général élémentaire mais conceptuellement profond, dont on trouvera la démonstration dans le cours de Sznitman [100] (voir aussi [40, p. 91]) : *le chaos microscopique équivaut au déterminisme de la mesure empirique.* Plus précisément, les énoncés suivants sont équivalents :

(i) (μ^N) est μ-chaotique ;
(ii) la mesure empirique $\hat{\mu}^N$ associée à μ^N converge en loi vers la mesure déterministe μ.

Par “mesure empirique $\hat{\mu}^N$ associée à μ^N” on entend la mesure image de μ^N par l'application $(y_1, \dots, y_N) \longmapsto N^{-1} \sum \delta_{y_i}$; c'est une mesure de probabilité aléatoire. La convergence en loi signifie que pour toute fonction continue bornée Φ sur l'espace des mesures de probabilité, on a

$$\int \Phi\left(\frac{1}{N}\sum \delta_{y_i}\right) \mu^N(dy_1 \dots dy_N) \xrightarrow[N\to\infty]{} \Phi(\mu).$$

En langage informel, étant donnée une grandeur statistique faisant intervenir $\hat{\mu}^N$, on peut en donner une excellente approximation pour N grand en remplaçant, dans l'expression de cette statistique, $\hat{\mu}^N$ par μ.

La notion de chaos ainsi présentée est une notion faible, susceptible de nombreuses variantes ; l'idée générale étant que μ^N devrait être proche de $\mu^{\otimes N}$. La notion plus forte de **chaos entropique** est introduite par Ben Arous et Zeitouni [13] : on y impose $H_{\mu^{\otimes N}}(\mu^N) = o(N)$. Une notion proche est développée par Carlen, Carvalho, Le Roux, Loss et Villani [32] dans le cas où l'espace microscopique n'est pas un produit tensoriel, mais plutôt une sphère de grande dimension ; on remplace alors la mesure $\mu^{\otimes N}$ par la mesure produit *restreinte* à la sphère. De nombreuses autres variantes sont encore à découvrir.

3.3. Le règne du chaos

Dans la théorie de Boltzmann, on postule que *le chaos est la règle* : quand on prépare un système, il est a priori dans un état chaotique. Voici quelques arguments possibles :

- si nous pouvons agir sur la configuration macroscopique, nous n'avons pas accès à la structure microscopique et il est très difficile d'imposer des corrélations ;
- les lois sous-jacentes aux variations microscopiques nous sont inconnues et on peut supposer qu'elles font intervenir un très grand nombre de facteurs détruisant les corrélations ;
- la mesure macroscopique observée en pratique semble toujours bien déterminée et pas aléatoire ;
- si l'on fixe la distribution macroscopique, l'entropie d'une distribution microscopique chaotique est plus grande que l'entropie d'une distribution microscopique non chaotique.

Explicitons le dernier argument. Si l'on se donne une probabilité μ sur $\mathcal{Y}$, alors la probabilité produit $\mu^{\otimes N}$ est d'entropie maximale parmi toutes les probabilités symétriques μ^N sur $\mathcal{Y}^N$ ayant μ pour marginale. Au vu des grands nombres N mis en jeu, cela représente une quantité de possibilités phénoménalement plus grande.

La mesure microscopique μ_0^N peut être considérée comme un objet de nature bayésienne, une probabilité a priori sur l'espace des observations possibles. Ce choix, en général arbitraire, est ici fait d'une manière canonique, par maximisation de l'entropie : en quelque sorte, on choisit la distribution qui laisse le plus de possibilités ouvertes, et rend l'observation la plus probable. On rejoint ainsi la démarche scientifique du maximum de vraisemblance, qui a prouvé sa robustesse et son efficacité – et l'on dépasse la traditionnelle querelle des fréquentistes contre les bayésiens !

Le problème de la **propagation du chaos** consiste à montrer que cette hypothèse de chaos, faite sur la donnée initiale (préparée on ne sait trop comment), est propagée par la dynamique microscopique (qui elle est bien définie). La propagation du chaos est vitale à deux titres : d'abord elle montre que l'indépendance est asymptotiquement préservée, fournissant une information statistique sur la dynamique microscopique ; ensuite, elle garantit que *la mesure empirique reste déterministe*, ce qui laisse espérer la possibilité d'une **équation macroscopique** gouvernant l'évolution de cette mesure empirique ou de son approximation.

3.4. Évolution de l'entropie

Un thème récurrent dans l'étude qualitative des systèmes dynamiques, au moins depuis Poincaré, est la recherche de mesures invariantes ; l'exemple le plus connu est la mesure de Liouville pour les systèmes hamiltoniens. Cette mesure possède en outre la propriété remarquable de se tensoriser.

Supposons que l'on dispose d'une dynamique microscopique sur $\mathcal{Y}^N$, et d'une mesure ν sur l'espace $\mathcal{Y}$ telle que $\nu^{\otimes N}$ est une mesure invariante pour la dynamique microscopique ; ou plus généralement telle qu'il existe une mesure invariante ν-chaotique sur $\mathcal{Y}^N$. Que devient dans la limite $N \to \infty$ la préservation du volume microscopique ?

Une conséquence simple de cette préservation du volume est la *conservation de l'information microscopique* $H_{\nu^{\otimes N}}(\mu_t^N)$, où μ_t^N est la mesure image de μ_0^N

par l'évolution microscopique. En effet, comme μ_t^N est préservée par le flot (par définition) et $\nu^{\otimes N}$ également, la densité $f^N(t, y_1, \ldots, y_N)$ est constante le long des trajectoires du système, et il s'ensuite que $\int f^N \log f^N \, d\nu^{\otimes N}$ est également constante.

Les choses sont plus subtiles pour l'information macroscopique. Bien sûr, si les différentes particules évoluent indépendamment les unes des autres, la mesure μ_t^N reste factorisée pour tous les temps, et on en déduit facilement que l'entropie macroscopique reste constante. En général les particules interagissent les unes avec les autres, ce qui détruit l'indépendance ; cependant s'il y a propagation du chaos en un sens suffisamment fort, l'indépendance est restaurée pour $N \to \infty$, et en conséquence on a toujours déterminisme de la mesure empirique. Toutes les configurations typiques pour la mesure microscopique initiale μ_0^N donnent lieu, après un temps t, à une mesure empirique $\hat{\mu}_t^N \simeq \mu_t$, où μ_t est bien déterminée. *Mais* il est possible que *d'autres* configurations microscopiques soient compatibles avec l'état μ_t, configurations qui n'ont pas été obtenues par évolution des configurations typiques initiales.

En d'autres termes : si l'on a une propagation du déterminisme macroscopique entre le temps initial et le temps t, et que la dynamique microscopique préserve la mesure de référence produit, alors on s'attend à ce que le volume d'états microscopiques admissibles ne diminue pas entre le temps 0 et le temps t. Gardant en tête la définition de l'entropie, on aurait $e^{NS(t)} \geq e^{NS(0)}$, où $S(t)$ est la valeur de l'entropie au temps t. On s'attend donc à ce que l'entropie *ne diminue pas* au cours de l'évolution temporelle :

$$S(t) \geq S(0).$$

Mais alors, pourquoi ne pas renverser l'argument et dire que le chaos au temps t implique le chaos au temps 0, par réversibilité de la dynamique microscopique ? Cet argument est en général irrecevable tant que l'on ne précise pas la notion exacte de chaos qui est propagée. La donnée initiale, préparée "au hasard" avec juste une contrainte de distribution cinétique, est supposée chaotique en un sens très fort ; mais peut-être que le système au temps t est chaotique en un sens moins fort ; ceci dépend de l'évolution microscopique.

La notion d'échelle d'interaction joue ici un rôle important. Certaines interactions s'effectuent à échelle macroscopique, d'autres à échelle microscopique, c'est-à-dire que tout ou partie de la loi d'interaction est codé dans des paramètres qui sont invisibles au niveau macroscopique. Dans ce dernier cas, la notion de chaos propice à la propagation de la dynamique risque de ne pas être visible à l'échelle macroscopique, et on peut s'attendre à une dégradation de la notion de chaos.

À partir de là, la discussion doit faire intervenir les détails de la dynamique, et nos pires ennuis commencent.

4. Équations cinétiques

Après l'introduction de l'entropie et du chaos, nous pouvons revenir aux systèmes newtoniens de la section 1, dont l'espace des phases est fait de positions et vitesses. Une **équation cinétique** est une équation d'évolution portant sur la fonction de distribution $f(t, x, v)$; le rôle important de la variable de vitesse v justifie la terminologie *cinétique*. Par extension, dans le cas où il y a des degrés de liberté externes (orientation de molécules par exemple), on parle encore d'équations cinétiques par extension.

Dans la lignée de Boltzmann, on se pose le problème de déduire une évolution macroscopique à partir du modèle microscopique sous-jacent. Ce problème est en général d'une difficulté considérable. Les équations fondamentales sont celles de Vlasov, Boltzmann, Landau et Balescu-Lenard, publiées respectivement en 1938, 1867, 1936 et 1960. (L'ordre plus ou moins logique de présentation de ces équations ne suit pas l'ordre dans lequel elles ont été découvertes...)

4.1. Équation de Vlasov

Aussi appelée équation de Boltzmann sans collisions, l'équation de Vlasov [113] est une équation de champ moyen, au sens où toutes les particules interagissent les unes avec les autres (de sorte que chaque particule ressent la contribution moyenne des autres). Pour la déduire de la dynamique newtonienne, commençons par traduire l'équation de Newton (1) sous forme d'équation sur la mesure empirique ; pour cela on écrit l'équation d'évolution d'un observable arbitraire :

$$\begin{aligned}\frac{d}{dt}\frac{1}{N}\sum_i \varphi(X_i(t), \dot{X}_i(t)) \\ &= \frac{1}{N}\sum_i \Big[\nabla_x\varphi(X_i, \dot{X}_i)\cdot \dot{X}_i + \nabla_v\varphi(X_i, \dot{X}_i)\cdot \ddot{X}_i\Big] \\ &= \frac{1}{N}\sum_i \Big[\nabla_x\varphi(X_i, \dot{X}_i)\cdot \dot{X}_i + \nabla_v\varphi(X_i, \dot{X}_i)\cdot \Big(a\sum_j F(X_i - X_j)\Big)\Big].\end{aligned}$$

Ceci peut se réécrire

$$\frac{\partial \hat{\mu}^N}{\partial t} + v\cdot\nabla_x\hat{\mu}_t^N + aN\big(F * \hat{\mu}_t^N\big)\cdot\nabla_v\hat{\mu}_t^N = 0. \tag{8}$$

Si maintenant on suppose que $aN \simeq 1$, et que l'on réalise l'approximation

$$\hat{\mu}_t^N(dx\,dv) \simeq f(t, x, v)\,dx\,dv,$$

on obtient l'équation de Vlasov

$$\frac{\partial f}{\partial t} + v\cdot\nabla_x f + \left(F *_x \int f\,dv\right)\cdot\nabla_v f = 0. \tag{9}$$

Notons bien que $\hat{\mu}_t^N$ dans (8) est une solution faible de l'équation de Vlasov, de sorte que le passage à la limite est conceptuellement très simple : il s'agit simplement d'un résultat de stabilité de l'équation de Vlasov.

Bien évidemment, je suis allé un peu vite en besogne car cette équation est non linéaire. Si $\hat{\mu} \simeq f$ au sens de la topologie faible des mesures, alors $F * \hat{\mu}$

converge vers $F * \int f\,dv$ dans une topologie déterminée par la régularité de F, et si cette topologie est plus faible que la convergence uniforme, rien ne garantit que $(F * \hat{\mu})\hat{\mu} \simeq (F * f)f$.

Si F est effectivement bornée et uniformément continue, alors l'argument ci-dessus peut être rendu rigoureux. Si F est en outre L-lipschitzienne, alors on peut faire mieux et établir une estimation de stabilité en topologie faible : si (μ_t) et (μ'_t) sont deux solutions faibles de l'équation de Vlasov, alors

$$W_1\big(\mu_t, \mu'_t\big) \leq e^{2\max(1,L)|t|}\, W_1\big(\mu_0, \mu'_0\big),$$

où W_1 est la distance de Wasserstein d'ordre 1,

$$W_1(\mu,\nu) = \sup\ : \left\{ \int \varphi\, d\mu - \int \varphi\, d\nu; \quad \varphi \text{ 1-lipschitzienne} \right\}.$$

Des estimations de ce style se trouvent dans [96, Chapitre 5], et remontent aux années 1970 (Dobrushin [48], Braun et Hepp [24], Neunzert [88]). On peut aussi établir des estimations de grande déviation comme dans [20].

En revanche, pour des **interactions singulières**, le problème de la limite de Vlasov reste ouvert, à l'exception d'un résultat de Jabin et Hauray [65], qui suppose essentiellement que (a) $F(x-y) = O(|x-y|^{-s})$ avec $0 < s < 1$; et (b) les particules sont initialement bien séparées dans l'espace des phases, de sorte que

$$\inf_{j\neq i} \Big(\big|X_i(0) - X_j(0)\big| + \big|\dot{X}_i(0) - \dot{V}_j(0)\big| \Big) \geq \frac{c}{N^{\frac{1}{2d}}}.$$

Aucune de ces conditions n'est satisfaisante : la première manque le cas coulombien d'un ordre de singularité, tandis que la seconde exclut le cas de données aléatoires. Cependant il s'agit bien du seul résultat disponible à ce jour... Pour aller plus loin, on souhaiterait propager une notion de chaos suffisamment forte pour contrôler le nombre de paires (i,j) telles que $|X_i(t) - X_j(t)|$ est petit. En l'absence de tels contrôles, l'équation de Vlasov pour des interactions singulières reste un acte de foi.

Cet acte de foi est très efficace puisque le modèle de Vlasov–Poisson, dans lequel $F = -\nabla W$, W solution fondamentale de $\pm\Delta$, est le modèle classique universellement accepté aussi bien en physique des plasmas [42, 72] qu'en astrophysique [15]. Dans le premier cas, une particule est un électron, dans le second c'est une étoile ! La seule différence réside dans le signe : interaction répulsive pour les électrons, attractive pour les étoiles. Il ne faut pas s'étonner de voir les étoiles considérées ainsi comme des objets microscopiques : elles le sont effectivement à l'échelle d'une galaxie (qui peut compter 10^{12} étoiles...).

La théorie de l'équation de Vlasov–Poisson elle-même reste incomplète. On distingue actuellement deux théories principales, toutes deux développées dans l'espace entier. Celle de Pfaffelmoser, simplifiée par Schaeffer et exposée par exemple dans [51], suppose que la donnée initiale f_i est C^1 à support compact ; cette hypothèse insatisfaisante de compacité a été ensuite éliminée par Horst [61]. La théorie concurrente est celle de Lions–Perthame, passée en revue dans [23]. La théorie de Pfaffelmoser a été adaptée dans un cadre périodique en espace (voir [12] ou modifier [61]), ce qui n'est pas le cas de la théorie de Lions–Perthame.

4.2. Équation de Boltzmann

L'équation de Vlasov perd sa pertinence quand les interactions ont une courte portée. Un exemple typique est celui des gaz raréfiés, pour lesquels les interactions dominantes sont binaires et se produisent uniquement au cours de "collisions" entre particules.

L'équation de Boltzmann a été établie par Maxwell [81] et Boltzmann [21, 22] ; elle décrit un régime où les interactions sont à courte portée et où chaque particule subit $O(1)$ choc par unité de temps. Beaucoup plus subtil que le régime de champ moyen de Vlasov, le régime de Boltzmann est cependant plus simple que le régime hydrodynamique où les particules subissent un très grand nombre de collisions par unité de temps.

Établissons pour commencer l'équation informellement. Le mouvement d'une particule est fait d'une alternance de trajectoires rectilignes et de chocs au cours desquels sa vitesse variera si brusquement que l'on peut considérer cet événement comme instantané et localisé en espace. Examinons d'abord le cas emblématique des sphères dures de rayon r : un choc se produit quand deux particules de positions respectives x et y, de vitesses respectives v et w, se retrouvent dans la configuration où $|x-y| = 2r$ et $(w-v)\cdot(y-x) < 0$. On parle alors de *configuration précollisionnelle.* On notera $\omega = (y-x)/|x-y|$.

Ici vient maintenant le point central de toute l'argumentation de Boltzmann : *quand deux particules se rencontrent, avec très forte probabilité elles ne sont* (*presque*) *pas corrélées* : penser à deux individus qui se rencontrent pour la première fois. On peut en conséquence appliquer l'hypothèse de chaos moléculaire à de telles particules, et on trouve que la probabilité d'une rencontre entre ces particules est proportionnelle à

$$\begin{aligned} f^{2;N}(t,x,v,x+2r\omega,w) &\simeq f^{1;N}(t,x,v)\, f^{1;N}(t,x+2r\omega,w) \\ &\simeq f^{1;N}(t,x,v)\, f^{1;N}(t,x,w), \end{aligned}$$

à condition donc que $(w-v)\cdot\omega < 0$. Il faut également tenir compte des vitesses relatives pour évaluer l'influence des particules de vitesse w sur les particules de vitesse v : la probabilité de rencontrer une particule de vitesse w en une unité de temps est proportionnelle au produit de $|v-w|$ par la section efficace (en dimension 3 il s'agit de l'aire apparente des particules, soit πr^2) et par le cosinus de l'angle entre $v-w$ et ω (le cas extrême est celui où $v-w$ est orthogonal à ω, ce qui veut dire que les deux particules se frôlent, évidemment un événement de probabilité nulle. Chacune de ces collisions fait disparaître une particule de vitesse v, et on a donc un terme négatif, *terme de perte*, proportionnel à

$$-\iint f(t,x,v)\, f(t,x,v_*)\, |(v-v_*)\cdot\omega|\, dv_*\, d\omega.$$

Les vitesses après le choc se calculent facilement :

$$v' = v - (v - v_*) \cdot \omega\, \omega; \quad v'_* = v_* + (v - v_*) \cdot \omega\, \omega. \tag{10}$$

Ces vitesses nous importent peu dans le bilan.

Cependant, il faut aussi tenir compte de toutes les particules de vitesse v qui ont été créées par collision entre particules de vitesses arbitraires. Par réversibilité microscopique, ces vitesses sont de la forme (v', v'_*), et notre problème est de tenir compte de tous les couples possibles (v', v'_*), qui dans ce problème de calcul du *terme de gain* sont des vitesses *pré-collisionnelles*. On applique donc à nouveau l'hypothèse de chaos pré-collisionnel, et on obtient finalement l'expression de l'**équation de Boltzmann pour les sphères dures** :

$$\frac{\partial f}{\partial t} + v \cdot \nabla_x f = Q(f, f), \tag{11}$$

$$Q(f,f)(t,x,v) = B \int_{\mathbb{R}^3} \int_{S^2_-} \big|(v - v_*) \cdot \omega\big| \Big(f(t,x,v')\, f(t,x,v'_*) - f(t,x,v)\, f(t,x,v_*) \Big)\, dv_*\, d\omega,$$

où S^2_- désigne les configurations pré-collisionnelles $\omega \cdot (v - v_*) < 0$, et B est une constante. En utilisant le changement de variable $\omega \to -\omega$ on peut symétriser cette expression, et arriver à l'expression finale

$$Q(f,f)(t,x,v) = B \int_{\mathbb{R}^3} \int_{S^2} \big|(v - v_*) \cdot \omega\big| \Big(f(t,x,v')\, f(t,x,v'_*) - f(t,x,v)\, f(t,x,v_*) \Big)\, dv_*\, d\omega. \tag{12}$$

L'opérateur (12) est l'**opérateur de collision de Boltzmann pour les sphères dures**. Le problème consiste maintenant à justifier cette approximation.

Pour ce faire, dans les années 1960, Grad proposa une limite mathématique précise : faire tendre r vers 0 et en même temps N vers l'infini, de sorte que $Nr^2 \to 1$, c'est-à-dire que la section efficace totale reste constante. Ainsi une particule donnée, se déplaçant parmi toutes les autres, en rencontrera typiquement un nombre fini par unité de temps. On part ensuite d'une densité de probabilité microscopique $f_0^N(x^N, v^N)\, dx^N\, dv^N$, que l'on fait évoluer par le flot newtonien $\mathcal{N}_t$, et l'on cherche à montrer que la première marginale $f^{1;N}(t, x, v)$ (obtenue en intégrant toutes les variables sauf la première variable de position et la première variable de vitesse) converge à la limite vers une solution de l'équation de Boltzmann.

La limite de Boltzmann–Grad est aussi souvent appelée **limite de faible densité** [40, p. 60] : en effet, si l'on part de la dynamique newtonienne et que l'on fixe la taille des particules, on va dilater l'échelle spatiale d'un facteur $1/\sqrt{N}$, et la densité sera d'ordre $N/N^{3/2} = N^{-1/2}$.

Au début des années 1970, Cercignani [37] montrait que le programme de Grad pouvait être mené à bien si l'on prouvait un certain nombre d'estimations *plausibles* ; peu de temps après, indépendamment, Lanford [70] esquissait la preuve du résultat souhaité.

Le **théorème de Lanford** est peut-être le plus important résultat mathématique de la théorie cinétique. Dans ce théorème, on se donne des densités microscopiques f_0^N, telles que pour tout k les densités $f_0^{k;N}$ des marginales à k particules sont continues, satisfont des bornes gaussiennes aux grandes vitesses, et convergent uniformément en-dehors des configurations collisionnelles (celles où les positions de deux particules distinctes coïncident) vers leur limite $f_0^{\otimes k}$. La conclusion est qu'il existe un temps $t_* > 0$ tel que $f_t^{k;N}$ converge *presque partout* vers $f_t^{\otimes k}$, où f_t est solution de l'équation de Boltzmann, pour tout temps $t \in [0, t_*]$.

Les estimations de Lanford ont été plus tard réécrites par Spohn [96] et par Illner et Pulvirenti [62, 63] qui remplacèrent l'hypothèse de temps petit par une hypothèse de petitesse sur la donnée initiale, permettant de traiter l'équation de Boltzmann comme une perturbation du transport libre. Ces résultats sont passés en revue dans [40, 91, 96].

La technique employée par Lanford et ses successeurs passe par la **hiérarchie BBGKY** (Bogoljubov–Born–Green–Kirkwood–Yvon), méthode par laquelle on exprime l'évolution de la marginale à une particule $f^{1;N}$ en fonction de la marginale à deux particules $f^{2;N}$; l'évolution de la marginale à deux particules $f^{2;N}$ en fonction de la marginale à trois particules $f^{3;N}$, et ainsi de suite. Ce procédé est particulièrement peu économique (dans l'argument heuristique précédent, on a seulement utilisé $f^{1;N}$ et $f^{2;N}$, mais on ne lui connaît pas d'alternative.

On résout ensuite chacune des équations de la hiérarchie via la formule de Duhamel, en appliquant successivement les opérateurs de transport libre et de collision, et en sommant sur tous les histoires collisionnelles possibles. On exprime ainsi formellement, comme avec un opérateur exponentiel, la solution au temps t en fonction de la donnée initiale, et on peut appliquer l'hypothèse de chaos sur (f_0^N).

Puis on passe à la limite $N \to \infty$ dans chacune des équations, après avoir vérifié que l'on peut négliger des situations "pathologiques" de "recollision" où une particule rencontre à nouveau une particule qu'elle a déjà rencontrée une première fois, et qui ne lui est donc pas inconnue. Ce point est subtil : on discute dans [40, Appendice 4.C] une dynamique a priori plus simple que celle des sphères dures, due à Uchiyama, avec seulement quatre vitesses dans le plan, pour laquelle les configurations recollisionnelles ne peuvent être négligées, et la limite cinétique n'existe pas.

Il reste à identifier le résultat avec la suite des produits tensoriels de la solution de l'équation de Boltzmann, et à conclure en utilisant un résultat d'unicité.

Spohn [96, section 4.6] montre que l'on peut donner des informations plus précises sur la répartition microscopique des particules : à petite échelle, celle-ci suit une **loi de Poisson** homogène dans l'espace des phases. Ceci est cohérent avec l'idée intuitive du chaos moléculaire.

Le théorème de Lanford tranche une controverse qui durait depuis Boltzmann lui-même ; mais il laisse de nombreuses questions en suspens. En premier lieu, il est limité à un intervalle de temps petit (sur lequel seulement 1/6 environ des particules ont eu le temps d'entrer en collision... mais l'impact conceptuel du théorème n'en est pas moins important). La variante d'Illner et Pulvirenti lève cette restriction de temps petit, mais la preuve ne s'adapte pas à une géométrie bornée. Quant à lever la restriction de petitesse, c'est pour l'instant un rêve très lointain.

Ensuite, le théorème n'est à ce jour démontré que pour un système de sphères dures ; les interactions à longue portée ne sont pas couvertes. Cercignani [36] note que la limite de Boltzmann–Grad pour de telles interactions pose des problèmes subtils, même d'un point de vue formel.

Enfin, le plus frustrant est peut-être que Lanford évite la discussion du **chaos pré-collisionnel**, cette notion selon laquelle des particules qui sont sur le point de se rencontrer ne sont pas corrélées. Cette notion est très subtile ! Car juste après le choc, des corrélations auront fatalement eu lieu. En d'autre termes, nous avons *chaos pré-collisionnel, mais pas post-collisionnel.*

Que signifie exactement le chaos pré-collisionnel ? Nous n'en avons pas pour l'instant de définition précise. C'est certainement une notion plus forte que le chaos au sens habituel ; elle fait intervenir en plus une hypothèse de décorrélation qui se voit sur un ensemble de codimension 1, à savoir les configurations menant à des chocs. On aimerait en déduire que c'est une notion de chaos où l'on a remplacé la topologie faible par une topologie uniforme ; mais cela ne peut être aussi simple, car le chaos en topologie uniforme impliquerait aussi le chaos post-collisionnel, et ce dernier est incompatible avec le chaos pré-collisionnel ! En effet, la continuité de la marginale à deux particules le long d'un choc impliquerait

$$\begin{aligned} f(t,x,v)\,f(t,x,v_*) &\simeq f^{(2;N)}(t,x,v,x+2r\omega,v_*) \\ &= f^{(2;N)}(t,x,v',x+2r\omega,v'_*) \simeq f^{(1;N)}(t,x,v')\,f^{(1;N)}(t,x,v'_*). \end{aligned}$$

Passant à la limite on en déduirait

$$f(t,x,v')\,f(t,x,v'_*) = f(t,x,v)\,f(t,x,v_*),$$

et comme on le verra dans la section 5.3 ceci implique que f est gaussienne dans la variable de vitesse, en général faux bien sûr ! Un autre raisonnement pour montrer que le chaos post-collisionnel doit être incompatible avec le chaos pré-collisionnel consiste à remarquer que si l'on a chaos post-collisionnel, le raisonnement menant à l'équation de Boltzmann peut être effectué à nouveau en exprimant les probabilités à deux particules en fonction des probabilités post-collisionnelles... et on obtient alors l'équation de Boltzmann renversée :

$$\frac{\partial f}{\partial t} + v\cdot\nabla_x f = -Q(f,f).$$

Comme on l'a mentionné, la preuve de Lanford ne s'applique qu'aux sphères dures ; cependant l'équation de Boltzmann est utilisée pour une gamme bien plus large d'interactions. La forme générale de l'équation, disons en dimension d, est la même qu'en (11) :

$$\frac{\partial f}{\partial t} + v \cdot \nabla_x f = Q(f,f), \tag{13}$$

mais maintenant

$$Q(f,f) = \int_{\mathbb{R}^d} \int_{S^{d-1}} \big(f'f'_* - ff_*\big)\, \tilde{B}(v-v_*,\omega)\, dv_*\, d\omega \tag{14}$$

où $\tilde{B}(v-v_*,\omega)$ ne dépend que de $|v-v_*|$ et $|(v-v_*)\cdot\omega|$. Il existe plusieurs représentations de cet opérateur intégral (voir [104]) ; il est souvent commode de changer de variables en introduisant un autre angle, $\sigma = (v'-v'_*)/|v-v_*|$, de sorte que les formules (10) doivent être remplacées par

$$v' = \frac{v+v_*}{2} + \frac{|v-v_*|}{2}\,\sigma, \qquad v'_* = \frac{v+v_*}{2} - \frac{|v-v_*|}{2}\,\sigma. \tag{15}$$

On doit alors changer le noyau de collision $\tilde{B}$ par B tel que

$$B\, d\sigma = \tilde{B}\, d\omega.$$

Explicitement, on trouve

$$\frac{1}{2}\tilde{B}(z,\omega) = \left|2\,\frac{z}{|z|}\cdot\omega\right|^{d-2} B(z,\sigma).$$

La forme précise de B (ou, de manière équivalente, de $\tilde{B}$) est obtenue par un calcul de scattering classique qui remonte à Maxwell, et que l'on peut trouver dans [38] : pour un paramètre d'impact $p \geq 0$ et une vitesse relative $z \in \mathbb{R}^3$, l'angle de déviation θ vaut

$$\theta(p,z) = \pi - 2p\int_{s_0}^{+\infty} \frac{ds/s^2}{\sqrt{1-\frac{p^2}{s^2}-4\frac{\phi(s)}{|z|^2}}} = \pi - 2\int_0^{\frac{p}{s_0}} \frac{du}{\sqrt{1-u^2-\frac{4}{|z|^2}\phi\left(\frac{p}{u}\right)}},$$

où s_0 est la racine positive de

$$1 - \frac{p^2}{s_0^2} - 4\frac{\phi(s_0)}{|z|^2} = 0.$$

Alors B est défini implicitement par

$$B\left(|z|,\cos\theta\right) = \frac{p}{\sin\theta}\frac{dp}{d\theta}|z|. \tag{16}$$

On notera indifféremment $B(|z|,\cos\theta)$ ou $B(z,\sigma)$, étant entendu que l'angle de déviation θ est l'angle formé par les vecteurs $v-v_*$ et $v'-v'_*$.

Quand $\phi(r) = 1/r$, on retrouve la formule de Rutherford pour la déviation coulombienne :

$$B\big(|v-v_*|,\cos\theta\big) = \frac{1}{|v-v_*|^3\sin^4(\theta/2)}.$$

Quand $\phi(r) = 1/r^{s-1}$, $s>2$, le noyau de collision ne se calcule pas explicitement, mais on peut montrer que (toujours en dimension 3)

$$B\big(|v-v_*|,\cos\theta\big) = b(\cos\theta)\,|v-v_*|^\gamma, \qquad \gamma = \frac{s-5}{s-1}. \tag{17}$$

En outre, la fonction b, définie implicitement, est localement lisse avec une *singularité angulaire non intégrable* quand $\theta \to 0$:

$$\sin\theta\, b(\cos\theta) \sim K\theta^{-1-\nu}, \qquad \nu = \frac{2}{s-1}. \tag{18}$$

Cette singularité correspond aux collisions de grand paramètre d'impact p, qui ne sont presque pas déviées. Elle est inévitable dès lors que les forces sont de rayon infini : en effet

$$\int_0^\pi B\big(|z|, \cos\theta\big) \sin\theta\, d\theta = |z| \int_0^\pi p\, \frac{dp}{d\theta}\, d\theta = |z| \int_0^{p_{\max}} p\, dp = \frac{|z|\, p_{\max}^2}{2}. \tag{19}$$

Dans le cas particulier $s = 5$, le noyau de collision ne dépend plus de la vitesse relative, mais seulement de l'*angle de déviation* : on parle de moélcules maxwelliennes. Par extension, on dit que $B(v - v_*, \sigma)$ est un noyau de collision maxwellien s'il ne dépend que de l'angle entre $v - v_*$ et σ. Les molécules maxwelliennes sont avant tout un modèle phénoménologique, même si l'interaction entre un ion chargé et une particule neutre dans un plasma est régie par une telle loi [42, Vol.1, p. 149]. Les potentiels en $1/r^{s-1}$ pour $s > 5$ sont appelés potentiels durs, pour $s < 5$ potentiels mous. On tronque souvent la singularité angulaire $b(\cos\theta)$ aux petites valeurs de θ.

L'équation de Boltzmann est importante dans la modélisation des gaz raréfiés, comme expliqué dans [39]. Cependant, du fait de son histoire mouvementée et de son contenu conceptuel, ainsi que de l'impact du traité de Boltzmann [22], cette équation a exercé une fascination qui va bien au-delà de son utilité. Les premiers travaux mathématiques qui lui étaient consacrés ont été ceux de Carleman[2] [26] [27], suivis de Grad [57]. Outre l'article de Lanford [70] déjà cité, un résultat qui a eu un grand retentissement est le théorème de stabilité faible de DiPerna–Lions [47]. L'équation est bien comprise dans le régime spatialement homogène pour les potentiels durs avec troncature angulaire, voir par exemple [85] ; et dans le régime proche de l'équilibre, voir par exemple [60]. On renvoie aux traités de référence [38, 40, 104] pour quantité d'autres résultats.

4.3. Équation de Landau

L'intégrale collisionnelle de Boltzmann perd son sens pour des interactions coulombiennes, du fait de la décroissance extrêmement lente du potentiel coulombien. Les collisions rasantes, de grand paramètre d'impact, deviennent alors dominantes.

En 1936, Landau [68] établissait par des arguments formels une asymptotique du noyau de Boltzmann dans ce régime. Soient λ_D la distance d'écrantage (en-dessous de laquelle le potentiel coulombien n'est plus visible du fait de la neutralité globale du plasma), et r_0 la distance typique des collisions (distance de deux particules dont l'énergie d'interaction est comparable à l'énergie d'agitation moléculaire) ; le paramètre $\Lambda = 2\lambda_D/r_0$ est le **paramètre plasma**, et dans la limite

2. La monographie [27] était inachevée au moment de la disparition de Carleman, et c'est Carleson qui l'a complétée.

$\Lambda \to \infty$ (justifiée pour les plasmas dits classiques), on peut formellement remplacer l'opérateur de Boltzmann avec potentiel coulombien écranté par un opérateur diffusif appelé opérateur de Landau :

$$Q_B(f,f) \simeq \frac{\log \Lambda}{2\pi\Lambda} Q_L(f,f), \tag{20}$$

$$Q_L(f,f) = \nabla_v \cdot \left(\int_{\mathbb{R}^3} a(v-v_*) \left[f(v_*) \nabla_v f(v) - f(v) \nabla_v f(v_*) \right] dv_* \right); \tag{21}$$

$$a(v-v_*) = \frac{L}{|v-v_*|} \Pi_{(v-v_*)^\perp} \tag{22}$$

où L est une constante et $\Pi_{z^\perp}$ est le projecteur orthogonal sur $z^\perp$.

L'approximation de Landau est maintenant bien comprise mathématiquement, dans le cadre d'une limite appelée **asymptotique des collisions rasantes** ; on pourra se reporter à [3] pour une discussion détaillée de ce problème.

L'opérateur de Landau (21), à la fois diffusif et intégral, présente une structure remarquable. On le généralise facilement à des dimensions $d \geq 2$ arbitraires, et on peut changer le coefficient $L/|z|$ en $L\,|z|^{\gamma+2}$, où γ est la puissance apparaissant dans (17). Les modèles de type potentiel durs avec $\gamma > 0$ ont été complétement étudiés dans le cas spatialement homogène [45] ; mais c'est bien le cas $\gamma = -3$ en dimension 3 qui est physiquement intéressant. Dans ce cas on sait seulement démontrer l'existence de solutions faibles dans le cas spatialement homogène (en adaptant [1, section 7]) et l'existence de solutions fortes pour des perturbations de l'équilibre [59]. Cette situation est tout à fait insatisfaisante.

4.4. Équation de Balescu-Lenard

En 1960, Balescu [7] établit directement une équation cinétique qui décrit les interactions coulombiennes dans un plasma ; il retrouve ainsi une équation publiée sous une autre forme par Bogoljubov [19] et simplifiée par Lenard. On pourra consulter [97] pour plus d'informations sur la genèse de l'équation, et [8] pour une présentation synthétique. Le noyau de collision dans cette équation prend la même forme que (21), la différence est dans l'expression de la matrice $a(v-v_*)$, qui maintenant dépend aussi de v et ∇f :

$$a_{BL}\big(v, v-v_*, \nabla f\big) = B \int_{|k| \leq K_0} \frac{k \otimes k}{|k|^4} \frac{\delta_{k\cdot(v-v_*)}}{\big|\epsilon(k, k\cdot v, \nabla f)\big|^2} \, dk, \tag{23}$$

$$\epsilon\big(k, k \cdot v, \nabla f\big) = 1 + \frac{b}{|k|^2} \int_{\mathbb{R}^3} \frac{k \cdot \nabla f(u)}{k \cdot (v-u) - i\,0} \, du.$$

Cette équation peut aussi être obtenue à partir de l'étude des fluctuations de l'équation de Vlasov en temps grand [72, section 51].

L'équation de Balescu–Lenard n'est presque pas employée du fait de sa complexité. Sous des hypothèses raisonnables, l'équation de Landau en constitue une bonne approximation [8, 71]. La procédure s'adapte à d'autres interactions que l'interaction coulombienne, mais en contraste avec la limite des collisions rasantes, elle fournit toujours l'expression (21), le seul changement étant dans le coefficient L de (22) qui est proportionnel à

$$\int_{\mathbb{R}^3} |k|\,|\hat{W}(k)|^2\,dk,$$

où W est le potentiel d'interaction. Cette équation est brièvement passée en revue dans [96, Chapitre 6].

La théorie mathématique de l'équation de Balescu–Lenard est grande ouverte, aussi bien pour ce qui est de l'établir que d'étudier ses propriétés qualitatives ; un des rares travaux rigoureux sur le sujet est l'étude linéarisée de R. Strain [97]. Bien que peu utilisée, l'équation de Balescu–Lenard n'en est pas moins le plus respecté des modèles collisionnels dans les plasmas, et c'est un intermédiaire qui permet de justifier l'usage de l'opérateur de collision de Landau pour représenter les fluctuations en temps grand dans les systèmes de particules ; sa théorie représente un formidable défi.

5. Théorème *H* de Boltzmann

Dans cette section nous allons partir de l'équation de Boltzmann et examiner quelques-unes de ses propriétés les plus marquantes. On trouvera beaucoup plus d'informations dans mon long article de revue [104].

5.1. Modification des observables par les collisions

Les propriétés statistiques du gaz se manifestent, dans le modèle cinétique, par l'évolution des **observables** $\iint f(t,x,v)\,\varphi(x,v)\,dx\,dv$. Supposant toujours des conditions aux limites périodiques, et toute la régularité nécessaire, on peut écrire

$$\begin{aligned}\frac{d}{dt}\iint f\,\varphi\,dx\,dv &= \iint (\partial_t f)\,\varphi\,dx\,dv \\ &= -\iint v\cdot\nabla_x f\,\varphi\,dx\,dv + \iint Q(f,f)\,\varphi\,dx\,dv \\ &= \iint v\cdot\nabla_x\varphi\,f\,dx\,dv + \iiiint \tilde{B}\,\big(f'f'_* - f f_*\big)\,\varphi\,dx\,dv\,dv_*\,d\omega,\end{aligned} \tag{24}$$

où l'on note toujours $f' = f(t,x,v')$, etc.

On effectue maintenant dans le terme intégral en $f'f'_*$ le changement de variables pré-postcollisionnel $(v,v_*) \longrightarrow (v',v'_*)$, pour tout $\omega \in S^{d-1}$. Ce changement de variable est unitaire (déterminant jacobien égal à 1), et préserve $\tilde{B}$ (ces propriétés sont des traces de la microréversibilité). Après avoir renommé les variables, on obtient

$$\frac{d}{dt}\iint f\,\varphi\,dx\,dv = \iint v\cdot\nabla_x\varphi\,f\,dx\,dv + \iiiint \tilde{B}\,f\,f_*\,(\varphi' - \varphi)\,dv\,dv_*\,d\omega\,dx. \tag{25}$$

C'est d'ailleurs sous cette forme que Maxwell avait écrit l'équation de Boltzmann dès 1867... On déduit de (25) que $\iint f\,dx\,dv$ est constante (heureusement !!), et on en tire aussi une quantité importante, la *section efficace de transfert de quantité de mouvement*, $M(v-v_*)$, définie par

$$M(v-v_*)\,(v-v_*) = \int \tilde{B}(v-v_*,\omega)\,(v'-v)\,d\omega.$$

Même quand $\tilde{B}$ est d'intégrale divergente, la quantité M doit être finie, exprimant le fait que les collisions modifient les vitesses de manière statistiquement raisonnable. On pourra se reporter à [2, 3, 104] pour plus de détails sur le traitement des singularités rasantes de $\tilde{B}$.

Boltzmann va améliorer le procédé de Maxwell en tirant mieux partie des symétries de l'équation. D'abord, en effectuant le changement de variable pré-postcollisionnel dans tout le second membre de (24), on obtient

$$\iiiint \tilde{B}\,(f'f'_* - f\,f_*)\,\varphi\,dv\,dv_*\,d\omega\,dx = -\iiiint \tilde{B}\,(f'f'_* - f\,f_*)\,\varphi'\,dv\,dv_*\,d\omega\,dx. \tag{26}$$

Au lieu d'échanger les configurations précollisionnelles et postcollisionnelles, on peut maintenant échanger les particules entre elles : $(v,v_*) \longmapsto (v_*,v)$, ce qui est aussi bien évidemment de déterminant jacobien unité. Ceci nous donne deux nouvelles formes de (26) :

$$\iiiint \tilde{B}\,(f'f'_* - f\,f_*)\,\varphi_*\,dv\,dv_*\,d\omega\,dx = -\iiiint \tilde{B}\,(f'f'_* - f\,f_*)\,\varphi'_*\,dv\,dv_*\,d\omega\,dx. \tag{27}$$

En combinant les quatre formes apparaissant dans (26) et (27), on obtient

$$\begin{aligned} \frac{d}{dt}\iint f\,\varphi\,dv\,dx = &\iint f\big(v\cdot\nabla_x\varphi\big)\,dx\,dv \\ &- \frac{1}{4}\iiiint \tilde{B}\,\big(f'f'_* - ff_*\big)\,\big(\varphi' + \varphi'_* - \varphi - \varphi_*\big)\,dx\,dv\,dv_*\,d\omega. \end{aligned} \tag{28}$$

Comme conséquence de (28), on note en premier lieu que $\iint f\,\varphi$ est conservé si φ vérifie l'équation fonctionnelle

$$\varphi(v') + \varphi(v'_*) = \varphi(v) + \varphi(v_*) \tag{29}$$

pour tous choix des vitesses v, v_* et du paramètre ω. De telles fonctions sont appelées **invariants de collision**, et se réduisent, sous des hypothèses extrêmement faibles, aux seules combinaisons linéaires des fonctions

$$1, \quad v_j\ (1 \le j \le d), \quad \frac{|v|^2}{2}.$$

On pourra consulter [40] à ce sujet. Ceci est encore une fois naturel : il s'agit du reflet macroscopique de la conservation de masse, quantité de mouvement et énergie cinétique au cours des interactions microscopiques.

5.2. Théorème H

Nous arrivons maintenant à une découverte qui installera Boltzmann parmi les plus grands noms de la physique. Choisissons $\varphi = \log f$, et supposons toute la régularité souhaitée pour mener les calculs ; en particulier,

$$\iint f\, v \cdot \nabla_x (\log f)\, dv\, dx = \iint v \cdot \nabla_x \big(f \log f - f\big)\, dv\, dx = 0.$$

L'identité (28) devient ainsi, en tenant compte des propriétés additives du logarithme,

$$\frac{d}{dt} \iint f \, \log f \, dx\, dv = -\frac{1}{4} \iint \iint B \left(f' f'_* - f f_*\right) \Big(\log f' f'_* - \log f f_*\Big). \qquad (30)$$

La fonction logarithme étant croissante, l'expression ci-dessus est toujours négative ! En outre, sachant que B ne s'annule que sur un ensemble de mesure nulle, on voit que l'expression (30) est strictement négative dès que $f' f'_*$ n'est pas égal à $f f_*$ presque partout, ce qui est vrai pour des distributions génériques. Ainsi, modulo la justification rigoureuse des intégrations par parties et changement de variables, nous avons prouvé que *dans le modèle de Boltzmann, l'entropie augmente avec le temps.*

L'impact de ce résultat est capital. D'abord, le raisonnement microscopique heuristique de la section 3.4 a été remplacé par un argument simple qui porte directement sur l'équation limite. Ensuite, même s'il s'agit d'une manifestation de la **seconde loi de la thermodynamique**, la croissance de l'entropie dans le modèle de Boltzmann est déduite d'un *raisonnement logique* et non d'un postulat (une loi) qui ne se discute pas. Enfin, bien sûr, ce faisant, Boltzmann a mis en évidence une **flèche du temps** associée à son équation.

Non seulement cette irréversibilité macroscopique n'est pas contradictoire avec la réversibilité microscopique, mais elle lui est en fait intimement liée : comme on l'a déjà expliqué, c'est la préservation du volume microscopique dans l'espace des phases qui garantit la non-décroissance de l'entropie. Au reste, comme s'en étonnait déjà L. Carleson en 1979 en examinant des modèles simplifiés de l'équation de Boltzmann [35], c'est bien quand les paramètres sont ajustés de manière à avoir la réversibilité microscopique, que l'on obtient du même coup le théorème H. Le phénomène est bien connu dans le contexte de la physique des milieux granulaires [106] : la dynamique microscopique y est dissipative (non réversible), incluant une déperdition d'énergie due aux frottements, et la dynamique macroscopique ne vérifie pas le Théorème H !

Du point de vue informationnel, l'augmentation de l'entropie signifie que le système se dirige toujours vers des états macroscopiques de plus en plus probables. Cette notion de probabilité est exacerbée par la formidable puissance de la combinatoire : supposons par exemple que l'on considère un gaz de $N \simeq 10^{16}$ particules (c'est en gros ce que l'on trouve dans 1mm^3 de gaz aux conditions habituelles !), et qu'entre le temps $t = t_1$ et le temps $t = t_2$ l'entropie augmente de seulement 10^{-5}. Le volume des possibilités microscopiques est alors multiplié par

$e^{N\,[S(t_2)-S(t_1)]} = e^{10^{11}} \gg 10^{10^{10}}$. Ce facteur phénoménal dépasse de très loin le nombre de protons dans l'univers (10^{100} ?) ou le nombre de livres de 1000 pages que l'on peut écrire en combinant tous les caractères alphabétiques de toutes les langues du monde...

L'interprétation intuitive du Théorème H est donc assez parlante : les états de haute entropie occupent au niveau microscopique une place tellement monstrueusement plus grande que les états de basse entropie, que le système microscopique va automatiquement les visiter. Comme on l'a vu, le raisonnement logique justifiant ce scénario est complexe et indirect, passant par la propagation du chaos et le déterminisme macroscopique – et à ce jour seule une petite partie du programme a été réalisée rigoureusement.

5.3. Annulation de la production d'entropie

L'augmentation de l'entropie admet un complément non moins profond, que l'on énonce souvent en deuxième partie du Théorème H : la *caractérisation des cas d'égalité*, c'est à dire des états pour lesquels la production d'entropie s'annule.

On a vu dans (30) que la production d'entropie est égale à

$$\int \mathrm{PE}\,(f(x,\,\cdot\,))\,dx, \tag{31}$$

où PE est la fonctionnelle de "production locale d'entropie", agissant sur des distributions cinétiques $f = f(v)$:

$$\mathrm{PE}\,(f) = \iiint \tilde{B}(v-v_*,\omega)\,\big(f(v')f(v'_*) - f(v)f(v_*)\big)\,\log\frac{f(v')f(v'_*)}{f(v)f(v_*)}\,dv\,dv_*\,d\omega. \tag{32}$$

Pour tous les modèles raisonnables, on a $\tilde{B}(z,\omega) > 0$ presque partout, et il s'ensuit que la production d'entropie ne s'annule que pour une distribution vérifiant l'équation fonctionnelle

$$f(v')f(v'_*) = f(v)f(v_*) \tag{33}$$

pour (presque) tous v, v_*, ω. En prenant le logarithme dans (33) on retrouve l'équation (29), ce qui montre que f doit être l'exponentielle d'un invariant de collision. Au vu de la forme de ces derniers, et tenant compte de la contrainte d'intégrabilité de f, on obtient $f(v) = e^{a+b\cdot v+c|v|^2/2}$, que l'on peut réécrire

$$f(v) = \rho\,\frac{1}{(2\pi T)^{d/2}}\exp\left(-\frac{|v-u|^2}{2T}\right), \tag{34}$$

où $\rho \geq 0$, $T > 0$ et $u \in \mathbb{R}^d$ sont des constantes. Il s'agit donc d'une gaussienne particulière, avec matrice de covariance proportionnelle à l'identité.

Maxwell avait déjà remarqué que (34) annule l'opérateur de collision de Boltzmann : $Q(f,f) = 0$. En son honneur, une telle distribution est appelée **maxwellienne**. Cependant, c'est Boltzmann qui donna le premier un argument convainquant selon lequel les distributions (34) sont *les seules* solutions de l'équation $\mathrm{PE}\,(f) = 0$, et en conséquence les seules solutions de $Q(f,f) = 0$. Rendons-lui hommage en esquissant une variante de sa preuve originale.

Commençons par moyenner (33) sur tous les angles $\sigma = (v' - v'_*)/|v - v_*|$; le membre de gauche $|S^{d-1}|^{-1} \int f' f'_* \, d\sigma$ est alors la moyenne de la fonction $\sigma \to f(c + r\sigma) \, f(c - r\sigma)$ où $c = (v + v_*)/2$ et $r = |v - v_*|$. Cette quantité ne dépend donc que de c et r, ou de manière équivalente de $m = v + v_*$ et $e = (|v|^2 + |v_*|^2)/2$, respectivement la quantité de mouvement et l'énergie totales impliquées dans une collision. Après passage au logarithme, on trouve pour $\varphi = \log f$ l'identité

$$\varphi(v) + \varphi(v_*) = G(m, e). \tag{35}$$

L'opérateur $\nabla_v - \nabla_{v_*}$, appliqué au membre de gauche de (35), donne $\nabla\varphi(v) - \nabla\varphi(v_*)$. Quand on applique le même opérateur au membre de droite, la contribution de m disparaît, et la contribution de e est colinéaire à $v - v_*$, on conclut donc que $F = \nabla\varphi$ vérifie

$$F(v) - F(v_*) \text{ est colinéaire à } v - v_* \text{ pour tous } v, v_*$$

et il est facile d'en déduire que $F(v)$ est une application affine, d'où la conclusion. (Une méthode brutale pour montrer le caractère affine de F, supposant la régularité : on commence par noter, en écrivant une formule de Taylor, que la matrice jacobienne de F est en tout point un multiple de l'identité, soit $\partial_i F^j(v) z_i = \lambda(v) \, z_j$; ensuite en dérivant par rapport à v_k on déduit que $\partial_{ik} F^j = 0$ si $i \neq j$, et il s'ensuit que tous les coefficients $\partial_i F^j$ s'annulent, après quoi on voit facilement que DF est un multiple de l'identité.)

En conséquence de (31) et (34), les distributions $f(x, v)$ qui annulent la production d'entropie de Boltzmann sont exactement les distributions de la forme

$$f(x, v) = \rho(x) \, \frac{1}{(2\pi T(x))^{d/2}} \exp\left(-\frac{|v - u(x)|^2}{2T(x)}\right). \tag{36}$$

On les appelle **maxwelliennes locales** ou encore **états hydrodynamiques**. Par rapport à la description cinétique, ces états sont caractérisés par une réduction considérable de la complexité, puisqu'ils ne dépendent que de trois champs : le champ de densité ρ, le champ de vitesses macroscopiques u, et le champ de température T. Ce sont ces champs qui interviennent dans les équations hydrodynamiques, d'où la terminologie ci-dessus.

Cette découverte jette un pont entre la description cinétique et la description hydrodynamique : dans un régime où les collisions sont très nombreuses (faible nombre de Knudsen), la finitude de la production d'entropie force la dynamique à se concentrer près des distributions qui annulent cette production. Cette remarque donne lieu à un vaste programme d'approximation hydrodynamique de l'équation de Boltzmann, auquel Hilbert fait allusion dans son Sixième Problème. Sur ce sujet on pourra consulter [54, 55, 94]. Si l'équation de Boltzmann peut être approchée aussi bien par des équations compressibles qu'incompressibles, il convient de noter qu'elle ne mène pas à toute la gamme des équations hydrodynamiques, mais seulement à celles des gaz parfaits, c'est à dire suivant une loi de pression proportionnelle à ρT.

6. Convergence entropique : oubli à marche forcée

Si Maxwell a découvert l'importance des profils de vitesses gaussiens en théorie cinétique, il n'a pas, comme le remarque Boltzmann, démontré que ces profils sont effectivement induits par la dynamique. Boltzmann souhaite achever ce programme, et pour cela retrouver les profils maxwelliens non seulement comme distributions d'équilibre, mais aussi comme limites de l'équation cinétique dans l'asymptotique où le temps devient grand ($t \to +\infty$). Ce saut conceptuel visant à fonder la mécanique statistique d'équilibre sur sa contrepartie hors équilibre, d'habitude beaucoup plus délicate, est toujours d'actualité dans d'innombrables contextes.

J'ai beaucoup écrit sur ce thème et on pourra consulter l'article de revue [104, Chapitre 3], le cours [109], l'article de recherche [46] ou le mémoire de recherche [110]. Dans la suite, pour fixer les idées, je supposerai que la variable de position vit dans le tore $\mathbb{T}^d$.

6.1. Maxwellienne globale

Nous avons déjà rencontré les maxwelliennes locales qui annulent l'opérateur de collision. Pour annuler également l'opérateur de transport $v \cdot \nabla_x$, il est naturel de chercher des maxwelliennes dont les paramètres ρ, u, T sont homogènes, des constantes indépendantes de la position. Un seul jeu de ces paramètres est compatible avec les lois de conservation de masse, de quantité de mouvement et d'énergie. La distribution ainsi obtenue est appelée **maxwellienne globale**

$$M_{\rho u T} = \rho \, \frac{1}{(2\pi T)^{d/2}} \exp\left(-\frac{|v-u|^2}{2T}\right).$$

Sans perte de généralité, quitte à changer de repère galiléen ou d'échelle physique, on poura supposer que $\rho = 1$, $u = 0$ et $T = 1$ et on notera M la distribution correspondante.

Cette distribution est donc un équilibre pour l'équation de Boltzmann. En outre, il est facile de vérifier que c'est la distribution qui maximise l'entropie sous les contraintes de masse, quantité de mouvement et énergie totales fixées. Ce critère de sélection préfigure la théorie classique de la mécanique statistique d'équilibre et les fameux ensembles canoniques de Gibbs.

6.2. Argument entropique

Boltzmann utilise maintenant le Théorème H pour donner une justification plus solide à la maxwellienne globale : il note que

- l'entropie augmente strictement, sauf si elle est dans un état hydrodynamique
- la maxwellienne globale, stationnaire, est la seule solution hydrodynamique de l'équation de Boltzmann.

L'image qui se dessine est que l'entropie va continuer à augmenter autant que possible, puisque la distribution ne restera jamais "bloquée" sur une solution hydrodynamique ; l'entropie finira par approcher l'entropie maximale de la maxwellienne globale, et la convergence en résultera.

À ce sujet, on peut faire deux remarques : la première est que la mesure de Lebesgue, que nous avons prise comme mesure de référence dans l'entropie de Boltzmann, peut être remplacée par la maxwellienne : en effet,

$$H(f) - H(M) = \iint f \log \frac{f}{M}\, dv\, dx = H_M(f),$$

où l'on a utilisé le fait que $\log M$ est un invariant de collision. La seconde remarque est que la différence d'entropies permet de quantifier l'écart à la gaussienne d'équilibre, en vertu par exemple de l'inégalité de Csiszàr–Kullback–Pinsker : $H_M(f) \geq \|f - M\|_{L^1}^2 / 2\|M\|_{L^1}$.

Le raisonnement de Boltzmann est essentiellement correct et il n'est pas difficile de le transformer en un argument rigoureux, montrant que les solutions suffisamment régulières de l'équation de Boltzmann s'approchent de l'équilibre maxwellien. Dans le cadre des solutions spatialement homogènes, T. Carleman mettait en forme ce raisonnement dès 1932 [26].

Cependant, Boltzmann n'avait aucun moyen de rendre l'argument qualitatif ; il faudra attendre près d'un siècle avant que l'on ose se poser la question de la vitesse de convergence vers l'équilibre gaussien, particulièrement pertinente car la validité de l'équation de Boltzmann n'est pas éternelle, et limitée dans le temps par des phénomènes tels que le théorème de récurrence de Poincaré.

6.3. L'approche probabiliste de Mark Kac

Au début des années 1950, Kac [67] cherche à comprendre la convergence vers l'équilibre pour l'équation de Boltzmann, et commence par simplifier le modèle. Kac oublie les positions, simplifie à outrance la géométrie des collisions, et invente un modèle stochastique *où l'aléatoire est présent dans l'interaction* : quand deux particules interagissent, on tire au hasard les paramètres qui décrivent la collision. Les positions étant absentes, les particules interagissent toutes les unes avec les autres, et il s'agit donc d'un modèle "de champ moyen". Ce modèle probabiliste simplifié est pour Kac l'occasion de formaliser mathématiquement la notion de propagation du chaos dans les équations de champ moyen, qui s'avèrera si féconde, et sera reprise plus tard par Sznitman [99] et bien d'autres.

Méfiant envers l'équation de Boltzmann, Kac souhaite expliquer la convergence par un raisonnement probabiliste microscopique, au niveau du système à N particules ; il cherche à obtenir des estimations de trou spectral uniformes en N. Son approche paraît aujourd'hui naïve dans ce qu'elle sous-estime la difficulté du traitement de la dimension N ; cependant le problème de la détermination du trou spectral optimal, résolu un demi-siècle plus tard, s'est avéré très intéressant [31, 66, 80]. À ce sujet on pourra consulter également [105, section 6] et [32] où l'on s'intéresse à la version entropique de ce programme "microscopique".[3]

En 1966 McKean [82] reprend le travail de Kac et dresse un parallèle avec la problématique du théorème central limite. Il introduit dans le sujet des outils de théorie de l'information, en particulier l'information de Fisher [41] qui mesure la

3. Le programme culmine avec un manuscrit récent de Mischler et Mouhot.

difficulté à estimer un paramètre tel que la vitesse des particules. Le programme sera complété par Tanaka [101] qui découvrira de nouvelles distances contractantes, et culminera avec les travaux de Carlen, Carvalho, Gabetta, Lu, Toscani sur le “théorème central limite pour l'équation de Boltzmann” [29, 30, 33, 34]. Cette théorie comprend aussi bien des théorèmes de convergence basés sur la combinatoire des interactions entre particules et des outils émanant de l'étude du théorème central limite (distances faibles...), que des contre-exemples présentant des convergences extrêmement lentes vers l'équilibre.

Ce programme stochastique permet donc de se dispenser du Théorème H ; en fait il a aussi permis de mettre en jour plusieurs autres fonctionnelles de Lyapunov : distance contractante (de transport optimal) de Tanaka, information de Fisher. Cependant du point de vue technique toute la théorie reste essentiellement confiné aux interactions maxwelliennes (dans lesquelles $\tilde{B}(v - v_*, \omega)$ dépend uniquement de l'angle entre $v - v_*$ et ω), et aux gaz spatialement homogènes. Le chapitre 4 de [104] est consacré aux propriétés particulières, fort élégantes au demeurant, de ces interactions.

Pour gagner en généralité et étudier des situations inohomogènes ou des interactions non maxwelliennes, la seule approche robuste connue à ce jour est basée sur le Théorème H.

6.4. Conjecture de Cercignani

Le Théorème H de Boltzmann est général et pertinent, il est donc naturel d'en chercher des raffinements quantitatifs. Au début des années 1980, C. Cercignani se demandait si l'on pouvait *minorer* la production d'entropie locale en fonction de la “non-gaussianité” de la distribution cinétique ; idéalement par un multiple de l'information $H_M(f)$. Il fallut attendre une dizaine d'années pour que Carlen et Carvalho [28], Desvillettes [44], sans répondre à la question de Cercignani, puissent cependant apporter des bornes inférieures quantitatives à la production d'entropie.

Une réponse plus précise a ce problème est obtenue dans mes articles [102] et [104] (le premier en collaboration avec G. Toscani). Sans perte de généralité on suppose que $\int f\, dv = 1$, $\int f\, v\, dv = 0$, $\int f\, |v|^2\, dv = d$, le cas général pouvant se déduire par changement d'échelle ou de référentiel.

Commençons par mentionner un résultat étonnant de simplicité, tiré de [105], qui s'applique dans un cas non physique : si $B(v - v_*, \sigma) \geq K\,(1 + |v - v_*|^2)$, alors

$$\mathrm{PE}\,(f) \geq \left(K_B\, \frac{|S^{d-1}|}{8}\, \frac{d-1}{1+2d}\right) T_f^*\, H_M(f), \tag{37}$$

où

$$T_f^* = \inf_{e \in S^{d-1}} \int f(v)\, (v \cdot e)^2\, dv.$$

La quantité T_f^* quantifie la non-concentration de f près d'un hyperplan ; elle se minore en fonction d'une information d'entropie, ou de régularité, ou bien automatiquement pour des distributions radialement symétriques. La concision du résultat masque une technique de preuve surprenante où l'on régularise f par un semigroupe auxiliaire de diffusion ; sous l'effet de ce semigroupe la variation de la production d'entropie de Boltzmann est essentiellement minorée par la production d'entropie de Landau, que l'on estime en termes de l'information de Fisher avant d'intégrer le long du semigroupe. On renvoie à [105] ou [109] pour les détails.

L'hypothèse de croissance quadratique en la vitesse relative n'est pas physiquement réaliste ; elle est cependant optimale au sens où il existe des contre-exemples [16] pour des noyaux à croissance $|v - v_*|^\gamma$, pour tout $\gamma < 2$. On peut ensuite travailler sur l'inégalité (37) pour en déduire une minoration plus faible s'appliquant à des sections efficaces réalistes, telles que le modèle des sphères dures ; la principale difficulté consiste à contrôler la quantité de production d'entropie qui est induite par les petites vitesses relatives ($|v - v_*| \leq \delta$) ; les logarithmes rendent ce contrôle délicat. On renvoie à [105] pour les détails. À la fin, on obtient, pour tout $\varepsilon > 0$ arbitrairement petit, l'inégalité

$$\mathrm{PE}\,(f) \geq K_\varepsilon(f)\,\big[H(f) - H(M^f)\big]^{1+\varepsilon}, \tag{38}$$

où M^f est la maxwellienne associée à f, i.e., celle dont les paramètres ρ, u, T correspondent à la densité, vitesse moyenne et température de f. La constante $K_\varepsilon(f)$ ne dépend que de ε, de la régularité C^r de f pour r assez grand, d'un moment $\int f|v|^s\,dv$ pour s assez grand, et d'une borne inférieure $f \geq K\,e^{-A|v|^q}$. Cela reste un problème ouvert de savoir si ces hypothèses peuvent être relaxées…

6.5. Convergence conditionnelle

L'inégalité (38) concerne une fonction $f = f(v)$ mais n'inclut pas de dépendance en espace ; c'est inévitable, puisque la variable x n'intervient pas dans l'étude de la production globale d'entropie. Bien sûr, (38) implique immédiatement (modulo les bornes de régularité) une convergence en $O(t^{-\infty})$ pour l'équation spatialement homogène, c'est à dire que la distance entre la distribution et l'équilibre tend vers 0 plus vite que t^{-k} pour tout k ; cependant cette inégalité ne résout pas le problème inhomogène. L'obstacle à surmonter est la *dégénérescence de la production d'entropie sur les états hydrodynamiques.* Une clé de l'étude en temps grand de l'équation de Boltzmann consiste donc à montrer que l'on ne passe pas trop de temps dans un état hydrodynamique, ou approximativement hydrodynamique. Pour éviter ce piège, nous ne pouvons compter que sur le transport, représenté par l'opérateur $v \cdot \nabla_x$. Grad [56] l'avait compris dès 1965, dans un texte au demeurant passablement obscur : *"the question is whether the deviation from a local Maxwellian, which is fed by molecular streaming in the presence of spatial inhomogeneity, is sufficiently strong to ultimately wipe out the inhomogeneity"* (…) *"a valid proof of the approach to equilibrium in a spatially varying problem requires just the opposite of the procedure that is followed in a proof of the H-Theorem, viz., to show that the distribution function does not approach too closely to a local Maxwellian."*

Dans les années 2000, Desvillettes et moi-même [46] redécouvrions ce principe formulé par Grad, et établissions une version de l'**instabilité de l'approximation hydrodynamique** : si le système devient, à un moment donné, proche d'être hydrodynamique, sans être pour autant à l'équilibre, alors les phénomènes de transport le font sortir de cet état hydrodynamique. Ceci est quantifié, sous des hypothèses de forte régularité, par l'étude des variations secondes d'un carré de norme, $\|f-M^f\|^2$, entre $f = f(t,x,v)$ et la maxwellienne *locale* associée,

$$M^f(t,x,v) = \rho(t,x)\,\frac{1}{(2\pi T(t,x))^{d/2}}\exp\left(-\frac{|v-u(t,x)|^2}{2T(t,x)}\right),$$

$$\rho(t,x) = \int f(t,x,v)\,dv, \qquad u(t,x) = \frac{1}{\rho(t,x)}\int f(t,x,v)\,v\,dv,$$

$$T(t,x) = \frac{1}{d\,\rho(t,x)}\int f(t,x,v)|v-u(t,x)|^2\,dv.$$

En quelque sorte, M^f est la meilleure approximation possible de f par un état hydrodynamique, et l'étude des variations de $\|f - M^f\|$ permet de vérifier que f ne peut rester trop longtemps proche de l'état hydrodynamique.

En mettant bout à bout (de manière particulièrement technique, et à l'aide de nombreuses inégalités fonctionnelles) le Théorème H quantitatif et l'instabilité de l'approximation hydrodynamique, nous aboutissions à la **convergence conditionnelle** : une solution de l'équation de Boltzmann vérifiant des bornes de régularité uniforme converge vers l'équilibre en $O(t^{-\infty})$. Ce résultat est constructif au sens où les constantes de temps impliquées ne dépendent que des bornes de régularité, de la forme de l'interaction et des conditions aux limites. La convergence repose sur un système d'inégalités qui font intervenir simultanément l'entropie et la distance aux états hydrodynamiques : par exemple, l'une d'entre elles s'écrit

$$\frac{d^2}{dt^2}\|f-M^f\|^2_{L^2} \geq K\int |\nabla T|^2\,dx - \frac{C}{\delta^{1-\varepsilon}}\left(\|f-M^f\|^2_{L^2}\right)^{1-\varepsilon} - \delta[H(f)-H(M)]. \quad (39)$$

Pour comprendre l'apport d'une telle inégalité, supposons que f devienne hydrodynamique à un moment : alors $f = M^f$, et (38) ne sert plus à rien. Mais si la température est inhomogène, et si δ est très petit dans (39), alors on se retrouve avec $(d^2/dt^2)\|f - M^f\|^2_{L^2} \geq$ const., ce qui bien sûr empêche f de rester proche de M^f pendant trop longtemps. Une fois que f est sortie de l'approximation hydrodynamique, on peut ré-appliquer (38), et ainsi de suite. Ce raisonnement ne fonctionne que quand la température est inhomogène, mais on peut trouver d'autres inégalités faisant intervenir les gradients de vitesse macroscopique et de densité. On est ainsi passé d'un argument “passif” à un argument “actif”, où l'augmentation de l'entropie est forcée par des inégalités différentielles plutôt que par l'identification d'une limite.

Finissons cette section par quelques commentaires sur les hypothèses. La théorie de la régularité de l'équation de Boltzmann permet de ramener les bornes générales à des bornes bien particulières ; par exemple, on sait que la distribution cinétique est automatiquement minorée par un multiple de $e^{-|v|^q}$ si par exemple l'équation est posée dans le tore et que la solution est régulière. On sait aussi que des bornes sur les moments d'ordre bas permettent d'avoir des bornes sur les moments arbitrairement élevés, etc. Mais la régularité dans le cadre général reste un célèbre problème ouvert. Le résultat de convergence conditionnelle montre que c'est le dernier obstacle qui nous sépare des estimations quantitatives de convergence vers l'équilibre ; il unifie également les résultats déjà connus sur la convergence : aussi bien le cas de distributions spatialement homogènes, que celui de distributions proches de l'équilibre, sont des situations dans lesquelles on dispose d'une théorie de la régularité à peu près complète.

6.6. Hypocoercivité

Dans l'étude de la convergence vers l'équilibre pour l'équation de Boltzmann, nous assistons à une interaction subtile entre l'opérateur de collision (non linéaire, dissipatif dégénéré) et l'opérateur de transport (linéaire, conservatif). Aucun des deux, pris séparément, ne serait suffisant pour induire la convergence, mais la combinaison des deux y parvient. Cette situation est assez fréquente et rappelle la problématique de l'hypoellipticité en théorie de la régularité des équations aux dérivées partielles. Par analogie, le problème de l'*hypocoercivité* s'intéresse aux propriétés de convergence pour des équations éventuellement dégénérées.

Une étude quelque peu systématique de ces situations, aussi bien pour des équations linéaires que non linéaires, est effectuée dans mon mémoire [110] ; on pourra aussi consulter l'article introductif [108]. La stratégie générale consiste à construire des fonctionnelles de Lyapunov adaptées à la dynamique, en ajoutant à la fonctionnelle naturelle (comme l'entropie) un terme d'ordre inférieur bien choisi. Un cas d'école est le "théorème $A^*A + B$", inspiré du théorème des carrés de Hörmander, qui donne des conditions suffisantes sur les commutateurs entre les opérateurs A et B, avec B antisymétrique, pour que l'évolution $e^{-t(A^*A+B)}$ soit hypocoercive. Dans la variante la plus simple, l'une de ces conditions, réminiscente de la condition d'algèbre de Lie de Hörmander, est la coercivité de $A^*A+[A,B]^*[A,B]$.

La théorie de l'hypocoercivité a maintenant acquis sa vie propre avec déjà un certain nombre de résultats marquants, et continue son expansion, surtout dans un cadre linéaire. Ceci est vrai aussi bien dans en théorie cinétique, comme le travail [58] que l'on évoquera dans la section suivante, que en-dehors de la théorie cinétique, comme dans le travail de Liverani et Olla sur les limites hydrodynamiques de certains systèmes de particules [74].

Dans un cadre non linéaire, le principal résultat reste [110, Théorème 51] ; cet énoncé général permet de simplifier la preuve du théorème de convergence conditionnelle pour l'équation de Boltzmann, et d'y inclure de nouvelles interactions et conditions aux limites. On renvoie à [110, Partie III] pour plus de détails.

6.7. Régime linéarisé

Le taux de convergence vers l'équilibre peut être précisé au moyen de l'étude du linéarisé. Commençons par débusquer une faute logique classique : l'étude du linéarisé ne peut en aucun cas se substituer à l'étude non linéaire, puisque la linéarisation n'est valable précisément qu'à partir du moment où la distribution est très proche de l'équilibre.

L'étude de la convergence linéarisée suppose de surmonter trois principales difficultés :

- estimer quantitativement le trou spectral de l'opérateur de collision linéarisé ;
- effectuer une étude spectrale du linéarisé dans un espace qui convienne au problème non linéaire, afin d'effectuer un "raccord" entre l'étude non linéaire et l'étude linéarisée ;
- prendre en compte la dégénérescence hydrodynamique, dans une perspective hypocoercive : en effet, l'équation linéarisée est tout autant dégénérée que l'équation non linéaire.

Toutes ces difficultés ont été résolues dans la dernière décennie par C. Mouhot et ses collaborateurs Baranger, Gualdani et Mischler [11, 58, 83], au moins dans le cas emblématique des sphères dures. C'est ainsi que l'article récent [58] établit un résultat de convergence conditionnelle avec taux exponentiel $O(e^{-\lambda t})$ au lieu de $O(t^{-\infty})$, et le taux λ est estimé de manière constructive.

La convergence exponentielle n'est pas une caractéristique universelle de l'équation de Boltzmann : on ne l'attend que pour des potentiels durs ou modérément mous. Supposons pour fixer les idées que le noyau de collision se comporte comme $|v - v_*|^\gamma \, b(\cos\theta)$. Dans le cas où $b(\cos\theta)\, \sin^{d-2}\theta$ est intégrable (souvent obtenu par troncature angulaire aux collisions rasantes), l'opérateur de collision linéarisé n'admet un trou spectral que pour $\gamma \geq 0$. Une abondance de collisions rasantes permet d'étendre cette condition, comme l'ont montré Mouhot et Strain [84] : si $b(\cos\theta)\, \sin^{d-2}\theta \simeq \theta^{-(1+\nu)}$ pour $\theta \to 0$ (collisions rasantes importantes), alors l'opérateur de collision linéarisé n'admet un trou spectral que pour $\gamma + \nu \geq 0$. La théorie de la régularité est actuellement en cours de développement pour de telles équations (travaux de Gressman–Strain, Alexandre–Morimoto–Ukai–Xu–Yang), et on peut parier que d'ici quelques années la théorie linéarisée couvrira tous ces cas.

Pour des potentiels trop mous (ou pour le modèle de collisions coulombiennes de Landau), on n'a pas de trou spectral et le meilleur résultat espérable est une convergence en exponentielle fractionnaire $O(e^{-\lambda t^\beta})$, $0 < \beta < 1$. On trouvera de telles estimées dans les travaux de Guo et Strain [60].

6.8. Évolution qualitative de l'entropie

Un thème récurrent dans toute cette partie est la dégénérescence liée aux états hydrodynamiques, qui gêne la convergence vers l'équilibre. Au début des années 2000, Desvillettes et moi-même suggérions que cette dégénérescence se reflétait dans des oscillations de la production d'entropie. Jamais observées auparavant, ces oscillations ont été mises en évidence dans les simulations numériques très précises de F. Filbet. J'ai reproduit ci-dessous une courbe frappante, obtenue avec l'équation de Boltzmann dans une géométrie monodimensionnelle périodique.

Sur ce schéma, on a tracé le logarithme de la fonction H y en fonction du temps ; la décroissance globalement rectiligne corresond donc à une convergence exponentielle vers l'état d'équilibre. On a également séparé l'information cinétique

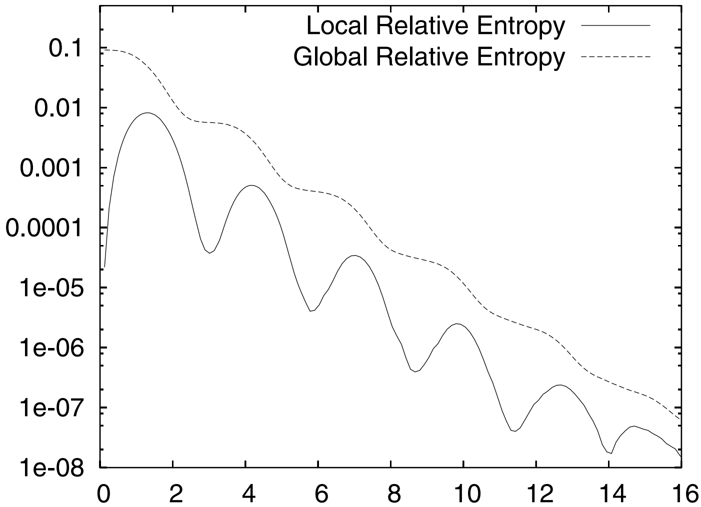


FIGURE 1. Évolution logarithmique de la fonction H cinétique et de la fonction H hydrodynamique, pour l'équation de Boltzmann dans une boîte périodique.

en information hydrodynamique et information “purement cinétique” :

$$\int f \log \frac{f}{M} = \left(\int \rho \log \frac{\rho}{T^{d/2}} \right) + \int f \log \frac{f}{M^f};$$

la deuxième quantité (information purement cinétique) est la courbe que l'on voit en-dessous de la courbe de la fonction H. Quand les deux courbes sont éloignées, la distribution est presque hydrodynamique, quand elles sont proches la distribution est presque homogène. Partant d'une distribution hydrodynamique, on s'en écarte tout de suite, conformément au principe d'instabilité de l'approximation hydrodynamique. Ensuite on distingue nettement des oscillations entre des états plutôt hydrodynamiques, associés à un ralentissement de la production d'entropie, et des états plutôt homogènes ; ces oscillations sont importantes compte tenu de la nature logarithmique du diagramme. Filbet, Mouhot et Pareschi [50] présentent d'autres courbes et tentent d'expliquer la fréquence d'oscillation dans un certain régime asymptotique.

Ici l'équation de Boltzmann révèle bien sa double nature, relevant aussi bien des transferts d'information via les collisions, que de la mécanique des fluides via l'opérateur de transport. C'est souvent le mariage entre les deux aspects qui s'avère délicat.

L'importance relative du transport et des collisions peut être modulée par les conditions aux limites ; dans un cadre périodique, il s'agit de faire varier la

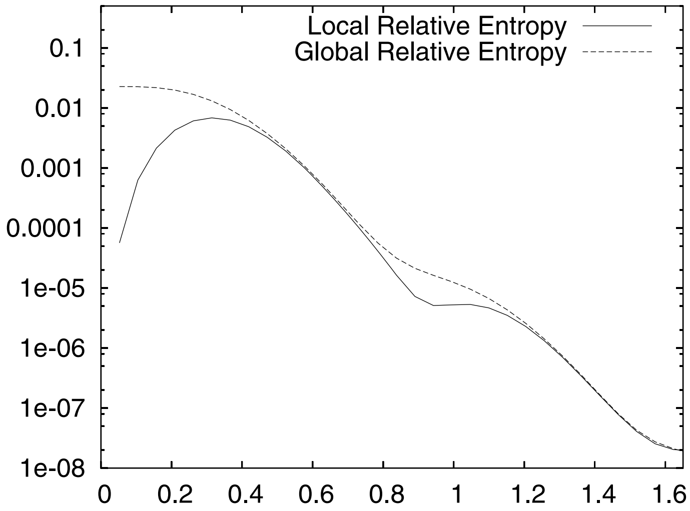


Figure 2. Même chose, dans une boîte plus petite.

taille de la boîte. Une boîte large permettra des variations spatiales importantes, laissant libre cours aux effets hydrodynamiques, comme dans la simulation ci-dessus. Cependant, on voit bien que même dans ce cas, et contrairement à une idée bien enracinée même chez les spécialistes, *le régime asymptotique n'est pas hydrodynamique*, au sens où le rapport entre l'entropie hydrodynamique et l'entropie cinétique totale n'augmente pas significativement au fur et à mesure que le temps passe, oscillant plutôt entre des valeurs minimum et maximum.

On peut se demander ce qui se passe dans une boîte plutôt petite. Une telle simulation est présentée ci-après.

La conclusion de cette figure est précisément à l'opposé de l'intuition selon laquelle les effets hydrodynamiques dominent en temps grand : bien au contraire, partant d'une situation hydrodynamique, on arrive rapidement à une situation presque homogène (à vue on a l'impression que au temps $t \simeq 0.7$ l'information hydrodynamique ne représente guère plus de 1% de l'information totale !). Les effets inhomogènes reprennent ensuite leurs droits (au temps $t = 1$ l'information est partagée en parts du même ordre), après quoi on redevient résolument homogène. Dans cet exemple, l'homogénéisation a agi plus vite que la convergence vers l'équilibre. On reviendra dans la section 8 sur ce graphique qui a suscité une certaine perplexité.

6.9. Deux problèmes non conventionnels

Je terminerai cette section en mentionnant deux curieux problèmes liés à la flèche du temps dans l'équation de Boltzmann, qui sont peut-être simplement des curiosités. Le premier est la classification de solutions éternelles de l'équation de

Boltzmann : j'avais tenté de montrer, au cours de ma thèse, que, au moins pour l'équation de Boltzmann spatialement homogène avec molécules maxwelliennes, il n'existe pas de solution éternelle d'énergie finie. La seconde consistait au contraire à chercher des solutions autosimilaires d'énergie infinie, ne convergeant pas vers l'équilibre maxwellien. Sur le premier problème, on pourra consulter [112] pour des résultats partiels ; la conjecture tient toujours, et Bobylev et Cercignani [18] ont pu montrer qu'il n'existe pas de solution éternelle ayant des moments finis à tous ordres. Quant au second problème, il a été résolu par les mêmes auteurs [17], en utilisant des techniques de transformée de Fourier.

7. Relaxation isentropique : s'accommoder de ses souvenirs

On considère maintenant l'équation de Vlasov avec potentiel d'interaction W :

$$\frac{\partial f}{\partial t} + v \cdot \nabla_x f - \left(\nabla W * \int f \, dv \right) \cdot \nabla_v f = 0. \tag{40}$$

Contrairement à l'équation de Boltzmann, l'équation (40) n'impose pas de flèche du temps, et reste inchangée sous l'action d'un renversement temporel. La constance de l'entropie correspond à une préservation d'information microscopique. La solution de l'équation de Vlasov au temps t permet théoriquement de reconstituer la condition initiale sans perte de précision, en résolvant simplement l'équation de Vlasov après avoir renversé les vitesses.

En outre, alors que l'équation de Boltzmann n'admet qu'un très petit nombre d'équilibres (les maxwelliennes déterminées par les lois de conservation), l'équation de Vlasov en admet une quantité considérable. Par exemple, *toutes* les distributions homogènes $f^0 = f^0(v)$ sont stationnaires. Il existe encore bien d'autres distributions stationnaires, par exemple la famille des ondes de Bernstein–Greene–Kruskal [14].

Pour toutes ces raisons, on ne voit rien a priori qui laisse supposer un comportement bien déterminé en temps grand ; aucune indication de flèche du temps. Pourtant, en 1946, L. Landau, sorti quelques années auparavant des geôles du régime communiste soviétique où son franc-parler l'avait mené, suggéra un comportement bien spécifique pour l'équation de Vlasov en temps grand. Il se basait sur une analyse de l'équation linéarisée autour d'un équilibre homogène. La prédiction de Landau a provoqué une secousse et un changement conceptuel qui suscitent encore aujourd'hui de vives discussions [93] ; à sa suite on se met à soupçonner que la convergence vers l'équilibre n'est pas forcément liée à une augmentation d'entropie.

Cette section est consacrée à un survol de la question de la convergence isentropique, en insistant sur le régime perturbatif qui est le seul sur lequel nous ayons des éléments solides. On trouvera plus de détails dans mon cours [111].

7.1. Analyse linéarisée

Étudions l'équation de Vlasov près d'un équilibre homogène $f^0(v)$. Si l'on pose $f(t,x,v) = f^0(v) + h(t,x,v)$, l'équation devient

$$\frac{\partial h}{\partial t} + v \cdot \nabla_x h + F[h] \cdot \nabla_v f^0 + F[h] \cdot \nabla_v h = 0, \tag{41}$$

où

$$F[h](t,x,) = -\iint \nabla W(x-y)\, h(t,y,v)\, dv$$

est la force induite par la distribution h.

En négligeant le terme quadratique $F[h] \cdot \nabla_v h$ dans (41), on obtient l'**équation de Vlasov linéarisée** près d'un équilibre homogène :

$$\frac{\partial h}{\partial t} + v \cdot \nabla_x h + F[h] \cdot \nabla_v f^0 = 0. \tag{42}$$

Avant d'examiner (42), considérons le cas sans interaction ($W = 0$), soit le **transport libre** $\partial_t f + v \cdot \nabla_x f = 0$. Cette équation se résout dans $\mathbb{T}^d_x \times \mathbb{R}^d_v$ en $f(t,x,v) = f_i(x - vt, v)$, où f_i est la distribution initiale. Passons en variables de Fourier, en posant

$$\tilde{g}(k,\eta) = \iint g(x,v)\, e^{-2i\pi k\cdot x}\, e^{-2i\pi \eta\cdot v}\, dx\, dv;$$

la solution du transport libre s'écrit alors

$$\tilde{f}(t,k,\eta) = \tilde{f}_i(k, \eta + kt). \tag{43}$$

Dès que $k \neq 0$, cette expression tend vers 0 quand $t \to \infty$, avec une vitesse déterminée par la régularité de f_i dans la variable de vitesse (principe de Riemann–Lebesgue). Tous les modes spatiaux non nuls relaxent donc vers 0 ; c'est l'action **homogénéisante** du transport libre.

L'équation (42) ne se résout pas si aisément ; cependant, si l'on pose $\rho(t,x) = \int h(t,x,v)\, dv$, alors on constate que les modes $\hat{\rho}(t,k)$ vérifient tous des équations indépendantes pour des valeurs distinctes de k. Cette remarquable propriété de *découplage* des modes est à la base de l'analyse de Landau. Pour chaque k on dispose d'une **équation de Volterra** sur le mode k :

$$\hat{\rho}(t,k) = \tilde{f}_i(k,kt) + \int_0^t K^0(k, t-\tau)\, \hat{\rho}(\tau,k)\, d\tau,$$

où

$$K^0(t,k) = -4\pi^2 \hat{W}(k) \tilde{f}^0(kt)\, |k|^2\, t.$$

La stabilité des équations de Volterra est un problème classique. Si u vérifie $u(t) = S(t) + \int_0^t K(t-\tau)\, u(\tau)\, d\tau$, alors le taux de décroissance de u est dicté par le pire de deux taux : le taux de décroissance de S bien sûr, et d'autre part la largeur de la plus grande bande $\{0 \leq \Re\xi \leq \Lambda\}$ qui ne rencontre pas de solution de l'équation $K^L = 1$, où K^L est la transformée de Laplace de K. Si $\Lambda > 0$, on a donc stabilité exponentielle pour le linéarisé.

Adapté à notre contexte, ce résultat mène au **critère de stabilité de Penrose**, dont on va énoncer la version multidimensionnelle. Pour tout $k \in \mathbb{Z}^d$, on définit $f_k^0 : \mathbb{R} \to \mathbb{R}_+$ par

$$f_k^0(r) = \int_{k^\perp} f^0 \left(\frac{k}{|k|} r + z \right) dz;$$

en clair, f_k^0 est la marginale de f^0 dans la direction k. Le critère de Penrose [89] demande que pour tout $k \in \mathbb{Z}^d$,

$$\forall \omega \in \mathbb{R}, \qquad (f_k^0)'(\omega) = 0 \implies \hat{W}(k) \int \frac{(f_k^0)'(v)}{v - \omega} \, dv < 1.$$

Si ce critère (essentiellement optimal) est satisfait, alors il y a stabilité exponentielle pour le linéarisé : la force décroît exponentiellement vite, de même que toutes les inhomogénéités de la densité spatiale $\int h \, dv$.

Le critère de stabilité de Penrose est vérifié dans de nombreuses situations : dans le cas d'une interaction coulombienne, dès que les marginales de f^0 sont croissantes à gauche de 0, décroissantes à droite (autrement dit, si $(f_k^0)'(z)/z < 0$ pour $z \neq 0$) ; en particulier si f^0 est une fonction décroissante de $|v|$, une gaussienne par exemple. Toujours dans le cas coulombien, en dimension 3 ou plus, le critère est vérifié si f^0 est isotrope. Dans le cas de l'attraction newtonienne, les choses sont plus complexes : par exemple, pour une distribution gaussienne, la stabilité dépend de la masse et de la température de la distribution. Ceci reflète la célèbre **instabilité de Jeans**, selon laquelle l'équation de Vlasov est linéairement instable aux longueurs plus grandes que

$$L_J = \sqrt{\frac{\pi T}{\mathcal{G} \rho^0}},$$

avec $\mathcal{G}$ la constante de gravitation universelle, ρ^0 la masse de la distribution f^0 et T sa température. C'est cette instabilité qui est responsable de la tendance des corps massifs à se regrouper en “clusters” (galaxies, amas de galaxies, etc.).

En résumé, l'équation de Vlasov linéarisée autour d'un équilibre homogène stable (au sens de Penrose) prédit un amortissement exponentiel de la force, de manière apparemment irréversible. Cette découverte faisait revenir la problématique de la flèche du temps dans la théorie de l'équation de Vlasov.

L'étude de l'équation de Vlasov linéarisée se trouve dans tous les traités avancés de physique des plasmas, comme [72] ; cependant, le traitement y est systématiquement obscurci par l'usage d'intégrales de contour dans le plan complexe, qui proviennent de l'inversion de la transformée de Laplace. Ceci est évité dans la présentation de [86, section 3], basée sur la simple transformée de Fourier ; ou dans la version courte [111].

7.2. Amortissement Landau non linéaire

La linéarisation effectuée par Landau n'est peut-être pas une opération innocente, et depuis un demi-siècle des doutes ont été émis sur sa validité. En 1960, Backus [6] faisait remarquer que remplacer dans le terme de force $\nabla_v (f^0 + h)$ par

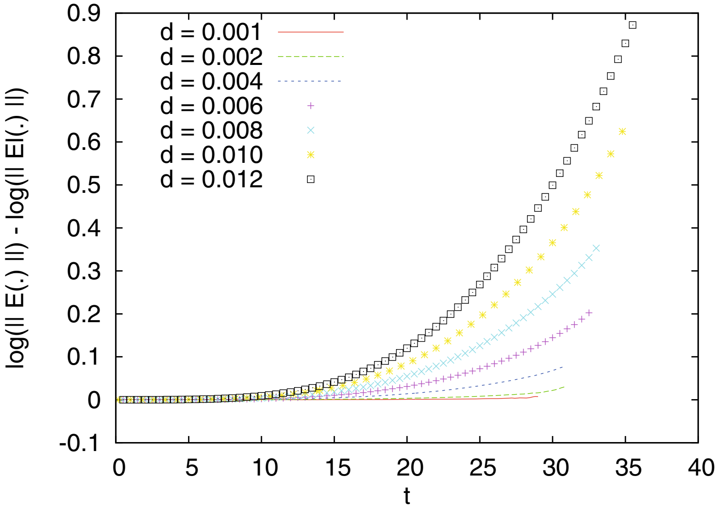


FIGURE 3. Pour une évolution de Vlasov, rapport logarithmique entre les normes de l'énergie suivant l'équation non linéaire, et suivant l'équation linéaire, pour différentes amplitudes de perturbation.

$\nabla_v f^0$ serait conceptuellement simple si $\nabla_v h$ restait petit pour tous temps ; mais si l'on remplace h par la solution de l'équation linéarisée, on constate que son gradient en vitesses croît linéairement en temps, devenant arbitrairement grand. Ceci, suggère Backus, "détruit la validité de la théorie linéaire". L'argument de Backus est contestable car $\nabla_v h$ est multiplié par $F[h]$ que l'on s'attend à voir décroître exponentiellement ; il n'empêche que des considérations heuristiques [87] suggèrent l'échec de l'approximation linéaire au bout d'un temps $O(1/\sqrt{\delta})$ où δ est la taille de la perturbation initiale. La courbe ci-après (tracée par F. Filbet) représente le logarithme du quotient entre l'énergie calculée selon l'équation non linéaire, et celle qui est obtenue selon l'équation linéaire, pour différentes valeurs de l'amplitude δ de perturbation ; on voit clairement que même si δ est petit, on finit par arriver dans un régime où les effets non linéaires ne peuvent être négligés.

Il y a d'autres raisons pour se méfier de la linéarisation. D'abord, le terme éliminé, $F[h] \cdot \nabla_v h$, est de plus haut degré en termes de dérivées de h en vitesse. Ensuite, la linéarisation élimine la conservation de l'entropie, et privilégie l'état particulier f^0, ce qui rend caduque la discussion sur la réversibilité.

En 1997, Isichenko [64] jette un pavé dans la mare en argumentant que la convergence vers l'équilibre ne peut être en général plus rapide que $O(1/t)$ pour l'équation non linéaire. Il est aussitôt contredit par Caglioti et Maffei [25] qui construisent des solutions exponentiellement amorties de l'équation non linéaire. Les simulations numériques (voir ci-dessous) ne sont pas très fiables en temps très grand, et le besoin d'un théorème se fait sentir.

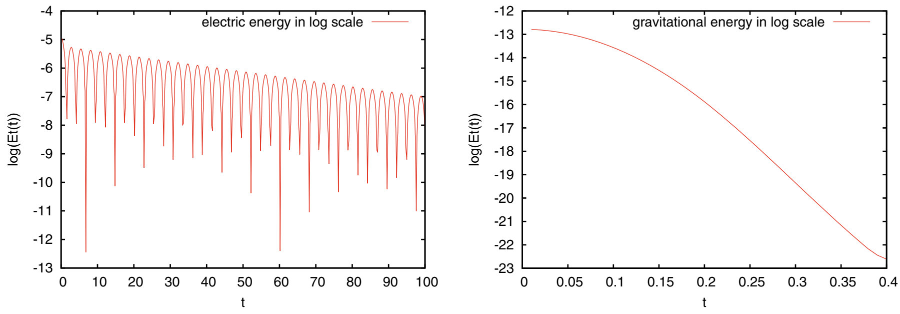


FIGURE 4. Évolution de la norme du champ de force, pour des interactions électrostatique (à gauche) et gravitationnelle (à droite). Dans le cas électrostatique, les oscillations rapides sont appelées ondes de Langmuir.

En 2009, Mouhot et moi-même établissons un tel résultat [86]. Si le potentiel d'interaction W n'est pas trop singulier, au sens où $\hat{W}(k) = O(1/|k|^2)$ (cette hypothèse autorise tout juste les interactions de Coulomb et Newton !), et si f^0 est un équilibre homogène analytique vérifiant la condition de stabilité de Penrose, alors il y a stabilité non linéaire dynamique : partant d'une donnée initiale f_i analytique telle que $\|f_i - f^0\| = O(\delta)$, avec δ très petit, on a décroissance de la force en $O(e^{-2\pi\lambda|t|})$, pour tout $\lambda < \min(\lambda_0, \lambda_i, \lambda_L)$, où λ_0 est la largeur de la bande d'analyticité complexe de f^0 autour de $\mathbb{R}^d_v$, λ_i est la largeur de la bande d'analyticité complexe de f_i dans la variable v, et λ_L est le taux de convergence de Landau. En somme, l'amortissement linéaire implique l'amortissement non linéaire, avec une perte arbitrairement petite sur le taux de convergence.

Le théorème établit aussi la convergence faible de $f(t, \cdot)$ vers un état asymptotique homogène $f_\infty(v)$. Plus précisément, l'équation étant invariante par renversement du temps, il y a un profil asymptotique $f_{+\infty}$ pour $t \to +\infty$, et un autre profil $f_{-\infty}$ pour $t \to -\infty$. Si l'on voit l'équation de Vlasov comme un système dynamique, on a donc un comportement remarquable : les trajectoires homoclines/hétéroclines sont si nombreuses qu'elles emplissent un plein voisinage de f^0 en topologie analytique.

L'amortissement non linéaire de l'équation de Vlasov repose sur le **confinement** et le **mélange**. Le confinement est indispensable : on sait que l'amortissement Landau n'a pas lieu dans tout l'espace, même pour l'équation linéarisée [52, 53] ; dans notre cas il est automatique car l'espace des phases est le tore. Le mélange a lieu à cause du phénomène de vitesse différentielle : des particules qui ont des vitesses différentes se déplacent à des vitesses différentes dans l'espaces des phases ; ici c'est presque une tautologie. Un exemple de système non mélangeant est l'oscillateur harmonique, où les trajectoires portées par des variables d'action différentes se déplacent à une vitesse angulaire constante. Certains des autres ingrédients sous-jacents à l'étude non linéaire sont

- une ré-interprétation du problème en termes de régularité : au lieu de montrer qu'il y a amortissement, on montre que $f(t, x, v)$ est "aussi régulière" que la solution du transport libre, uniformément en temps ;
- des estimations de "déflexion" : une particule placée dans un champ de forces exponentiellement décroissant suit une trajectoire asymptotique au transport libre, en un sens que l'on peut quantifier précisément ;
- le rôle stabilisant de la réponse avec retard, *en échos*, du plasma : quand un des modes du plasma est perturbé, la réaction des autres modes n'est pas instantanée, mais survient avec un léger retard, car l'effet des modes se compense en-dehors de certains instants de résonance ;
- un schéma de Newton, qui tire parti de ce que le système de Vlasov linéarisé est en quelque sorte complètement intégrable ; la vitesse de convergence de ce schéma permet de compenser la perte de décroissance qui vient avec la résolution du linéarisé.

Tous ces ingrédients sont décrits plus en détail dans [111]. La place particulière du schéma de Newton et de la complète intégrabilité forment un pont inattendu avec la théorie KAM (Kolmogorov–Arnold–Moser). En quelque sorte, l'équation de Vlasov non linéaire, dans le régime perturbatif, hérite certaines des bonnes propriétés de l'équation complètement intégrable de Vlasov linéarisé.

Du point de vue physique, l'information va vers les **petites échelles cinétiques** : les oscillations de la fonction de distribution s'amplifient quand le temps devient grand, et deviennent *invisibles.* Lynden-Bell [76, 77] l'avait bien compris, et utilise une formule frappante pour l'expliquer : *"A [galactic] system whose density has achieved a steady state will have information about its birth still stored in the peculiar velocities of its stars".*

Ces oscillations, bien visibles sur les figures ci-dessous, sont à la fois une nuisance du point de vue technique, et le mécanisme physique fondamental qui produit l'impression d'irréversibilité. On note la différence avec le mécanisme dit de *radiation*, dans lequel l'énergie est émise à échelle macroscopique et s'en va à l'infini : ici au contraire l'énergie disparaît littéralement dans le néant…

7.3. Régularité glissante

Le théorème d'amortissement non linéaire repose sur une ré-interprétation nouvelle en termes de régularité, qui mérite quelques commentaires. Commençons par parler de la **cascade** associée au transport libre, représentée sur le diagramme ci-dessous :

Cette image, qui se déduit de la formule (43), montre que les fréquences qui comptent varient au cours du temps : il y a un mouvement d'ensemble vers les hautes fréquences cinétiques, et ce mouvement est d'autant plus rapide que la fréquence spatiale est élevée. Plus précisément, le mode spatial de fréquence k oscille en vitesse avec une période en $O(1/|k|t)$. L'enjeu de l'amortissement Landau consiste à montrer que cette cascade, bien que déformée, est globalement préservée par l'effet des interactions, qui *couplent* les différents modes.

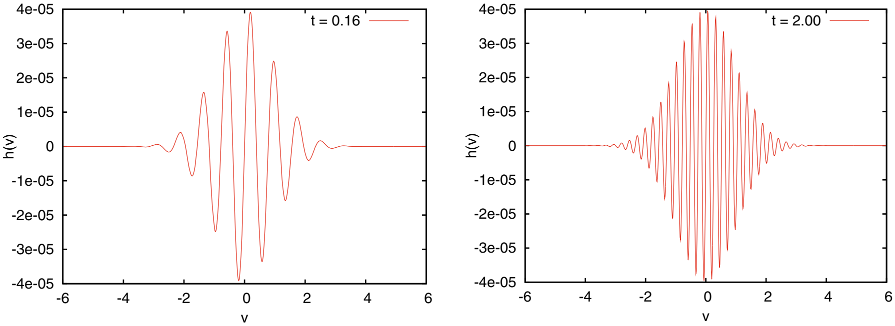


FIGURE 5. Une tranche de la fonction de distribution (par rapport à un équilibre homogène) pour l'amortissement Landau gravitationnel, à deux temps différents.

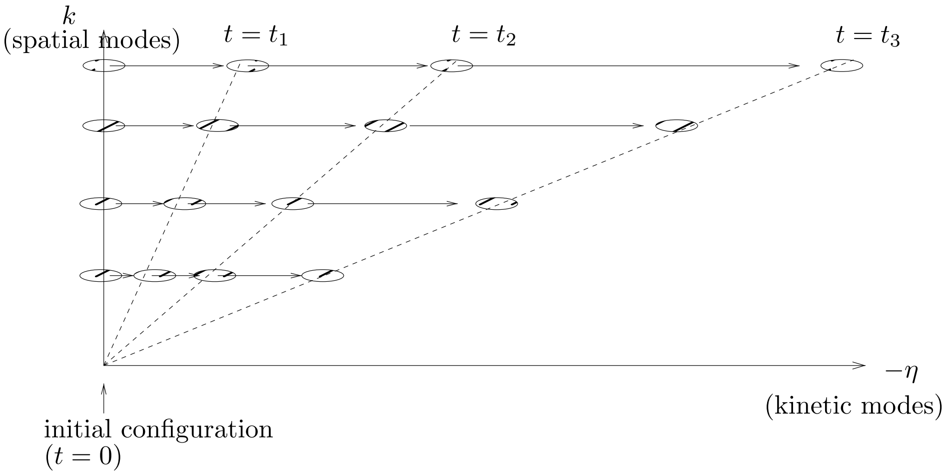


FIGURE 6. Évolution de l'énergie dans l'espace des fréquences le long du transport libre ou d'une perturbation de ce dernier ; les marques indiquent la localisation de l'énergie dans l'espace des phases.

Ces fortes oscillations empêchent tout espoir d'obtenir des bornes uniformes en temps, par exemple en régularité analytique au sens usuel. Une idée maîtresse dans [86] consiste à se concentrer sur les modes de Fourier qui comptent dans la solution du transport libre, et donc à suivre la cascade au fur et à mesure que le temps passe. On appelle ce concept **régularité glissante**. La régularité glissante vient avec une dégradation des bornes de régularité en vitesse, mais en même temps avec une *amélioration de la régularité en position*, dès que l'on fait des moyennes en vitesse. Notre interprétation de l'amortissement Landau est donc un **transfert**

de régularité de la variable v vers la variable x, la régularité de la force allant en s'améliorant, ce qui implique que son amplitude s'éteint.

La norme analytique utilisée dans [86] est assez complexe : elle a de bonnes propriétés algébriques qui permettent de suivre les erreurs obtenues par composition, elle s'adapte bien à la géométrie du problème, et suit le transport libre pour mesurer la régularité glissante :

$$\|f\|_{\mathcal{Z}_\tau^{\lambda,(\mu,\gamma);p}} = \sum_{k\in\mathbb{Z}^d}\sum_{n\in\mathbb{N}^d} \frac{\lambda^n}{n!}\, e^{2\pi\mu|k|}\,(1+|k|)^\gamma \left\|\big(\nabla_v + 2i\pi\tau k\big)^n \hat{f}(k,v)\right\|_{L^p(dv)} \qquad (44)$$

(Ici $\hat{f}$ désigne la transformée de Fourier dans la variable de position, pas de vitesse.) L'exposant λ quantifie la régularité analytique en vitesse, les exposants μ et γ (par défaut $\gamma = 0$) quantifient la régularité en position, et le paramètre τ est à prendre comme un décalage en temps. On renvoie à [86] pour une étude des propriétés remarquables de ce type de normes, et aussi pour des résultats de comparaison aux normes plus naïves dans lesquelles le théorème d'amortissement non linéaire est énoncé.

Le résultat principal de [86] consiste à prouver une borne uniforme du type

$$\|f(t,\cdot) - f^0\|_{\mathcal{Z}_t^{\lambda,\mu;1}} = O(\delta).$$

Cette borne implique l'amortissement Landau, mais contient aussi beaucoup plus d'informations : par exemple, elle montre que les fréquences spatiales élevées relaxent plus vite ; elle entraîne aussi la stabilité orbitale non linéaire sous condition de Penrose, un problème qui avait jusqu'ici résisté à toutes les méthodes classiques.

7.4. Échos non linéaires et régularité critique

La célèbre expérience de l'écho plasma [78, 79] décrit l'interaction de deux ondes engendrées par des perturbations spatiales distinctes. Si l'on envoie une première perturbation au temps initial avec une fréquence k, il s'ensuit des oscillations à fréquence cinétique $|k|t$, oscillations qui ne s'atténuent pas au cours du temps et au contraire deviennent de plus en plus frénétiques. Si maintenant au temps τ on fait intervenir une seconde perturbation avec fréquence ℓ, alors on engendrera des oscillations à la fréquence cinétique $|\ell|(t-\tau)$. Les deux trains d'oscillations seront invisibles l'un à l'autre, par effet de moyenne, sauf quand ils auront la même fréquence cinétique ; ceci se produit à un temps t tel que $kt + \ell(t-\tau) = 0$, soit

$$t = \frac{\ell\tau}{k+\ell}; \qquad (45)$$

il est sous-entendu ici que k et ℓ sont colinéaires et de direction opposée, avec $|\ell| > |k|$. En un certain sens, dans l'asymptotique en temps grand, la réaction à la seconde perturbation τ s'effectue à un temps t qui est strictement plus grand que t. Ce *retard* est capital pour expliquer la stabilité non linéaire de l'équation de Vlasov. Pour se donner une idée de ce gain, comparer l'inégalité $u(t) \le A + \int_0^t \tau\, u(\tau)\, d\tau$, qui implique pour u une croissance essentiellement en $O(e^{t^2})$, à l'inégalité $u(t) \le A + t\,u(t/2)$, qui implique une croissance très lente en $O(t^{\log t})$.

On peut proposer comme caricature des estimées pour l'équation de Vlasov–Poisson la famille d'inégalités

$$\varphi_k(t) \leq a(kt) + \frac{c\,t}{k^2}\varphi_{k+1}\left(\frac{kt}{k+1}\right).$$

Ici $\varphi_k(t)$ représente en gros la norme du mode k de la densité spatiale au temps t ; $a(kt)$ représente l'effet de la source (oublions le terme linéaire représenté par une équation de Volterra), le coefficient t traduit le fait que le couplage se fait par l'intermédiaire du gradient de f en vitesses, et que ce gradient croît linéairement avec le temps ; $1/k^2$ est la transformée de Fourier du potentiel d'interaction, on note à ce sujet que l'interaction entre modes est d'autant plus dangereuse que le potentiel est singulier ; on n'a gardé que l'interaction entre le mode k et le mode $k+1$; enfin l'argument du mode $k+1$ n'est pas t mais $kt/(k+1)$, ce qui représente un léger retard par rapport à t, comme dans la formule des échos. Une résolution explicite montre que

$$\varphi_k(t) \lesssim a(kt)\,\exp((ckt)^{1/3}).$$

Ces estimations peuvent être adaptées à l'équation originelle de Vlasov–Poisson ; on trouve ainsi, dans la résolution de l'équation linéarisée autour d'une solution non stationnaire, une *perte de régularité/décroissance* qui est en exponentielle fractionnaire. Sous de bonnes hypothèses (aussi fortes que la condition de Penrose dans le cas gravitationnel, plus fortes dans le cas coulombien) on trouve effectivement $\exp((kt)^{1/3})$; dans des cas plus généraux la croissance reste comme une exponentielle fractionnaire en kt. Comme elle reste sous-exponentielle, cette perte de régularité peut être compensée par la décroissance exponentielle issue du problème linéaire.

La perte de régularité dépend essentiellement de l'interaction, alors que le gain linéaire dépend surtout de la régularité des données : exponentiel pour des données analytiques, polynomial pour des données C^r, exponentiel fractionnaire pour des données Gevrey. La discussion précédente suggère donc qu'il est possible d'étendre le théorème d'amortissement non linéaire à des données Gevrey. Par exemple, dans le cas gravitationnel, l'exposant critique $1/3$ correspond à une régularité critique Gevrey-3. On rappelle qu'une fonction est dite Gevrey-ν si ses dérivées successives ne croissent pas plus vite que $O(n!^\nu)$. Quitte à perdre arbitrairement peu sur ν, il est équivalent de demander que sa transformée de Fourier décroisse en exponentielle fractionnaire $\exp(c\,|\xi|^{1/\nu})$.

7.5. Spéculations

Le théorème d'amortissement Landau non linéaire ouvre un grand nombre de questions. D'abord son extension à d'autres géométries que $\mathbb{T}^d$ est un vrai défi, car on perd alors la magique transformation de Fourier. L'extension à des équilibres inhomogènes est encore un rêve lointain ; en fait la stabilité linéaire des ondes de Bernstein–Greene–Kruskal n'est toujours pas connue !

Ensuite, on a vu que l'on sait traiter en l'amortissement en régularité Gevrey, en revanche l'extension à des régularités plus basses telle que la régularité C^r est un problème ouvert. On a déjà souligné le parallèle avec la théorie KAM, dans laquelle on sait traiter des problèmes en classe C^r ; mais dans KAM, la perte de régularité est seulement polynomiale, ici elle est bien plus sévère. Certaines variantes du problème KAM mènent à une perte de régularité en exponentielle fractionnaire, et alors c'est également un problème ouvert de travailler en régularité plus basse que Gevrey. Dans l'immédiat, la seule avancée en régularité C^r suggérée par [86] est la possibilité de prouver l'amortissement sur des échelles de temps bien plus grandes que l'échelle non linéaire ($O(1/\delta)$ au lieu de $O(1/\sqrt{\delta})$, voir [86, section 13] ; ce développement semble dépendre d'une conjecture originale concernant les constantes optimales survenant dans certaines inégalités d'interpolation). On reparlera dans la section 8 de stratégie permettant de contourner conceptuellement cette limitation de très haute régularité.

Quelle que soit la régularité optimale, il est exclu d'obtenir un amortissement Landau dans l'espace d'énergie naturel associé aux lois de conservations physiques. En effet, Lin et Zeng [73] montrent que l'amortissement Landau non linéaire est faux si l'on a strictement moins de deux dérivées, dans un sens approprié.

Enfin, même si l'amortissement Landau n'est qu'un phénomène perturbatif, il faut noter que son importance conceptuelle reste considérable parce que c'est à l'heure actuelle le seul îlot que nous parvenons à explorer dans l'océan des questions ouvertes ayant trait à la relaxation isentropique. Par sa découverte, Landau a fait prendre conscience que des systèmes physiques pouvaient relaxer sans qu'il y ait pour autant d'irréversibilité et d'augmentation d'entropie. Dans les années 1960, Lynden-Bell [76, 77] invoquait cette avancée conceptuelle pour résoudre le paradoxe de la relaxation des galaxies, qui apparaissent dans un état approximativement quasi-stationnaire alors que les temps de relaxation associés à l'équation de Vlasov galactique sont largement supérieurs à l'âge de l'univers. Depuis, le principe de la **relaxation violente**, relaxation du champ de forces sur quelques temps caractéristiques de la dynamique, est bien accepté par les astrophysiciens, sans que l'on ait aucune explication théorique à lui avancer. Il y a là un défi scientifique majeur.

8. Faible dissipation

Entre le modèle de Boltzmann qui fait la part belle aux collisions et celui de Vlasov qui les néglige complètement, nous trouvons un compromis particulièrement intéressant dans le modèle de Landau (ou Fokker–Planck–Landau), à faible dissipation :

$$\frac{\partial f}{\partial t} + v \cdot \nabla_x f + F[f] \cdot \nabla_v f = \varepsilon\, Q_L(f, f), \tag{46}$$

où Q_L est l'opérateur de Landau (21).

En physique des plasmas classique, le coefficient ε vaut $(\log \Lambda)/(2\pi\Lambda)$, où Λ est le paramètre plasma, d'ordinaire très grand (entre 10^2 et 10^{30}). Dans une approche particulaire, le coefficient ε est une fluctuation par rapport à la limite de champ moyen, proportionnelle à $\log N/N$. Les effets entropiques irréversibles modélisés par l'opérateur de collision ne se font donc sentir notablement que sur de grands temps $O(1/\varepsilon)$. Par ailleurs, des effets régularisants se font sentir instantanément, même s'ils sont faibles. L'intérêt de l'étude est donc multiple :

- c'est un modèle physique plus réaliste que l'équation de Vlasov "pure", sans collisions ;
- il permet de quantifier, en fonction du petit paramètre ε, les vitesses relatives des phénomènes d'homogénéisation (amortissement Landau) et de convergence entropique ;
- il permet de contourner l'obstacle de la régularité Gevrey auquel se heurte l'étude du modèle sans collision.

Tout reste à faire en la matière et je vais me contenter d'esquisser un programme de longue haleine.

8.1. Scénario plausible

Partant d'une perturbation d'équilibre homogène à décroissance en vitesses très rapide, on doit, au cours de l'évolution temporelle par (46), rester proche du régime homogène ; ceci est dans l'esprit de résultats d'Arkeryd, Esposito et Pulvirenti [5] sur l'équation de Boltzmann faiblement inhomogène. Dans le cadre homogène, l'opérateur du membre de droite a sans doute les mêmes propriétés régularisantes qu'un laplacien en vitesses, du moins localement (les propriétés de régularisation deviennent très faibles aux grandes vitesses, mais on a imposé une décroissance très forte en vitesses). Admettant que cela reste vrai dans un cadre faiblement inhomogène, on se retrouve avec une équation hypoelliptique, qui va régulariser dans toutes les variables, plus vite bien sûr dans la variable de vitesse que dans la variable de position.

La **régularisation hypoelliptique en classe Gevrey** n'a été que très peu étudiée, mais par des arguments dimensionnels on peut penser que dans ce contexte il y a régularisation en classe Gevrey-$1/\alpha$, avec vitesse $O(\exp((\varepsilon t)^{-\alpha/(2-\alpha)}))$ en v, et $O(\exp((\varepsilon t)^{-3\alpha/(2-3\alpha)}))$ en x.

D'un autre côté, dans la classe Gevrey-$1/\alpha$, pour $\alpha > 1/3$ on doit avoir décroissance vers le régime homogène en $O(\exp -t^{\alpha})$.

En combinant les deux effets, on obtiendra l'homogénéisation sur une échelle de temps $O(\varepsilon^{-\zeta})$, avec $\zeta < 1$, ce qui est un taux plus rapide que le taux d'augmentation de l'entropie en $O(\varepsilon^{-1})$.

Quand on fait les comptes, le coefficient ζ que l'on peut espérer est décevant, de l'ordre de 8/9. Parmi les étapes utilisées, le maillon le plus faible semble être la régularisation Gevrey en x, qui est extrêmement coûteuse et peut-être pas optimale puisque cette régularité n'est pas nécessaire dans l'analyse linéaire. Ceci motive le développement d'une version du théorème d'amortissement non linéaire en basse régularité en x. Si l'on se passe de cette régularité, le coefficient devient bien meilleur, de l'ordre de 1/6...

8.2. Réexamen des simulations

Avec cette interprétation, nous pouvons maintenant revenir sur la figure 2 : l'usage d'une petite boîte spatiale renforce l'effet de l'opérateur $v \cdot \nabla_x$ au détriment de l'opérateur de collision, de sorte que l'on est dans un régime faiblement dissipatif. (Le champ de force est nul.) Alors en temps grand l'homogénéisation intervient plus rapidement que la relaxation entropique. Ceci n'explique pas tout, pour deux raisons : d'abord, dans cette figure la condition initiale est fortement (et non faiblement) inhomogène ; ensuite l'opérateur de Boltzmann ne régularise pas. Cependant on veut bien croire que c'est l'homogénéisation par amortissement Landau qui se manifeste en premier dans cette figure, avant que les collisions ne fassent leur travail d'augmentation de l'entropie. (Comment décrire la sortie temporaire du régime homogène, semble un mystère.)

9. Métastatistiques

J'utilise ici le mot “métastatistiques” pour parler de statistiques sur la fonction de distribution, qui elle-même a un contenu statistique. Cette section sera courte car nous n'avons guère que des spéculations en la matière.

Le théorème de Hewitt–Savage, un avatar du théorème de Krein–Milman, décrit les mesures de probabilité symétriques en un grand nombre de variables comme des combinaisons convexes de mesures chaotiques :

$$\mu^\infty = \int_{P(\mathcal{Y})} \mu^{\otimes\infty}\, \Pi(d\mu),$$

où Π est une mesure de probabilité sur $P(\mathcal{Y})$, l'espace des mesures de probabilité sur l'espace macroscopique. En somme, une incertitude microscopique peut se décomposer en deux niveaux : outre le chaos à profil macroscopique fixé, il y a l'incertitude sur le profil macroscopique, c'est-à-dire le choix du profil μ, qui s'effectue avec la mesure de probabilité Π.

Maintenant y a-t-il une mesure de probabilité Π naturelle sur l'espace des profils admissibles ? Idéalement, une telle mesure serait **invariante par la dynamique**. Dans le cadre de l'équation de Boltzmann, la question ne se pose pas vraiment : seules restent en lice des mesures triviales portées par des équilibres maxwelliens. En revanche, dans le cadre de l'équation de Vlasov, la construction de mesures invariantes non triviales est un problème passionnant. De telles mesures refléteraient la nature hamiltonienne de l'équation de Vlasov, étudiée pour des interactions simplifiées par Ambrosio et Gangbo [4].

Un candidat assez sérieux au statut de mesure invariante est la **mesure entropique** de Sturm [98], issue de la théorie du transport optimal, formellement de la forme $\mathbb{P} = e^{-\beta H_\nu}$. Sa définition complexe a empêché jusqu'ici qu'on réussisse à prouver son invariance. Il ne doit pas être très difficile de modifier la construction pour ajouter un terme d'énergie. La mesure de Sturm est définie sur un espace compact, et il y a peut-être des subtilités à l'étendre dans un contexte cinétique où l'espace des vitesses est non borné. Mais la pire difficulté vient sans doute de la singularité des mesures typiques : on s'attend à ce que $\mathbb{P}$-presque toute mesure soit totalement étrangère à la mesure de Lebesgue, et portée par un ensemble de codimension 1. Ceci semble fermer la porte à toute étude statistique d'amortissement basée sur la régularité, et accroît le mystère.[4]

Robert [92] et d'autres ont tenté de faire une théorie statistique de l'équation de Vlasov, en partant de la notion d'entropie, essayant de prédire l'état asymptotique *probable* de l'évolution dynamique. La théorie a remporté quelques succès, cependant elle reste controversée. En outre, comme l'état asymptotique est obtenu par une limite faible, la question se pose de savoir s'il faut imposer une égalité ou une inégalité sur les contraintes faisant intervenir des fonctionnelles non linéaires de la densité. À ce sujet on pourra consulter [103].

Ensuite, cette théorie ne tient presque pas compte de l'équation d'évolution sous-jacente, postulant une certaine universalité par rapport à l'interaction. Isichenko [64] a fait remarquer que l'état asymptotique en temps grand, s'il existe, doit dépendre de détails fins de la distribution initiale et de l'interaction, alors que les mesures construites par la théorie statistique ne dépendent que d'invariants : énergie, entropie, voire d'autres fonctionnelles de la forme $\iint A(f)\,dx\,dv$. Cette objection a trouvé de la substance avec les contre-exemples construits dans [86, section 14], qui montrent que la transformation $f(x,v) \to f(x,-v)$ peut modifier l'état asymptotique final, alors qu'elle préserve tous les invariants connus de la dynamique. L'objection est peut-être surmontable, car ces contre-exemples sont construits en régularité analytique, c'est-à-dire dans une classe qui doit être invisible à un traitement statistique ; mais ces contre-exemples montrent la subtilité du problème, et renforcent le sentiment de difficulté de construction de mesures invariantes.

10. Paradoxes perdus

Dans cette dernière section je vais passer en revue une série de paradoxes plus ou moins célèbres associés à la flèche du temps et aux équations cinétiques, et présenter la résolution communément admise de ces paradoxes. Un certain nombre d'entre eux font intervenir l'infini, source classique de paradoxes comme l'"hôtel Hilbert" disposant d'un nombre infini de chambres, où l'on peut toujours loger un nouvel arrivant même s'il est déjà complet. À notre échelle, ce paradoxe reflète notre incapacité de nous rendre compte de l'apparition ou disparition d'une particule par rapport à la quantité gigantesque qui composent notre univers. La limite $N \to \infty$ (ou l'asymptotique $N \gg 1$, si comme Boltzmann on préfère éviter de manipuler les infinis) étant le fondement de la mécanique statistique, il n'est pas étonnant que ce paradoxe se manifeste.

4. D'après une communication personnelle de Mouhot, la mesure de Sturm pourrait être trop singulière pour convenir.

Dans toute la suite, quand on parle de temps positif ou négatif, ou de configuration pré-collisionnelle ou post-collisionnelle, on fait référence au temps microscopique absolu des équations de Newton.

10.1. Paradoxe de Poincaré–Zermelo

Poincaré [90] en 1895 met en doute la théorie de Boltzmann au motif qu'elle semble contredire les propriétés fondamentales des systèmes dynamiques. Zermelo [114] peu après développe ce point et note que l'augmentation inexorable de l'entropie interdit le retour du système à l'état initial, qui est pourtant prédit par le théorème de récurrence (à une erreur arbitrairement petite près).

La même objection peut s'appliquer au problème de l'amortissement Landau : si la distribution tend vers un équilibre homogène, elle ne reviendra jamais près de son état initial.

Du point de vue mathématique, ce raisonnement ne s'applique évidemment pas, puisque l'équation de Boltzmann fait intervenir un nombre infini de degrés de liberté ; ce n'est que pour un nombre de particules fixé que le théorème de récurrence s'applique. Du point de vue physique, la réponse est un peu plus subtile. D'une part, le temps de récurrence diverge quand le nombre N de particules tend vers l'infini, et cette divergence est probablement monstrueusement rapide ! Pour un système de taille macroscopique, même petit, le théorème de récurrence ne s'applique simplement jamais, il met en oeuvre des temps bien plus grands que l'âge de l'univers. D'autre part, la validité de l'équation de Boltzmann n'est pas éternelle : à N fixé, la qualité de l'approximation va se dégrader avec le temps, car le chaos (simple ou pré-collisionnel) n'est préservé qu'approximativement. Quand le temps de récurrence de Poincaré aura lieu, l'équation de Boltzmann aura cessé d'être valable depuis bien longtemps ! ! [5]

10.2. Conservation microscopique du volume

Le théorème de récurrence de Poincaré est basé sur la conservation du volume dans l'espace des phases microscopiques (préservation de la mesure de Liouville). L'entropie est directement fonction du volume des états microscopiques admissibles, comment peut-elle augmenter si le volume des états microscopiques est constant ?

La réponse à cette question peut sembler surprenante : on peut argumenter que l'augmentation de l'entropie n'a pas lieu *malgré* la conservation du volume microscopique, mais *à cause* de cette conservation ; plus précisément, c'est elle qui empêche l'entropie de diminuer. En effet, partons au temps initial de toutes les configurations typiques associées à une distribution f_i. Après un temps t, ces configurations typiques ont évolué et sont maintenant associées à une distribution f_t, la transition entre f_i et f_t étant régie par l'équation de Boltzmann. Les configurations typiques associées à f_t sont donc *au moins aussi nombreuses* que les

5. Dans la vraie vie, je pense que la validité de l'équation de Boltzmann est plus longue, car des facteurs aléatoires microscopiques, comme des fluctuations quantiques, doivent intervenir pour "renouveler" l'équation ; mais ceci ne remet pas le raisonnement en cause.

configurations typiques associées à f_i, ce qui veut bien dire que l'entropie ne peut diminuer.

Dans un modèle microscopique irréversible, on aura typiquement une contraction de l'espace des phases microscopique, lié à un phénomène dissipatif. L'argument précédent ne s'applique alors plus, et on peut imaginer que l'entropie diminue, au moins pour certaines données initiales. C'est de fait ce qui se produit par exemple dans les modèles de gaz granulaires subissant des collisions inélastiques.

10.3. Apparition spontanée de la flèche du temps

Comment, partant d'une équation microscopique qui ne privilégie aucune direction du temps, l'équation de Boltzmann peut-elle prédire une évolution inexorable vers les temps positifs ?

La réponse est simple : il n'y a pas d'évolution inexorable vers les temps positifs, et la double direction du temps est préservée. Simplement, il y a eu un choix particulier de la donnée initiale (instant de préparation de l'expérience), qui a fixé un temps particulier, disons $t = 0$. À partir de là, on a une double flèche du temps, l'entropie augmente pour les temps positifs, et diminue pour les temps négatifs.

10.4. Paradoxe de Loschmidt

Le paradoxe de Loschmidt [75] formalise l'apparente contradiction qu'il y a dans la coexistence d'une dynamique microscopique réversible et d'une évolution irréversible de l'entropie. Supposons que l'on parte d'une configuration initiale donnée, et qu'au temps t on arrête le gaz, et on renverse les vitesses de toutes les particules. Cette opération ne change pas l'entropie, et partant de cette nouvelle donnée initiale on peut laisser la dynamique agir à nouveau. Par réversibilité microscopique, au bout d'un temps $2t$ on sera revenu au point de départ ; mais l'entropie n'aura cessé d'augmenter, d'où contradiction.

On peut résoudre ce paradoxe de plusieurs manières, qui reviennent toutes à la même constatation : la *dégradation de la notion de chaos* entre le temps initial et le temps $t > 0$. Au niveau mathématique, on ne sait prouver que la convergence faible de $\mu_t^{1;N}$ vers $f(t,\cdot)$ quand $N \to \infty$, alors que la convergence est supposée uniforme au temps initial. En fait, on conjecture que la donnée (μ_t^N) vérifie la propriété (encore à définir...) de *chaos pré-collisionnel*, alors que la donnée initiale était supposée vérifier une propriété de chaos complète. Quand on inverse les vitesses, on transforme l'hypothèse de chaos pré-collisionnel en chaos post-collisionnel, et l'équation pertinente n'est plus l'équation de Boltzmann, mais l'équation de Boltzmann “inverse”, dans laquelle on a mis un signe négatif devant l'opérateur de collision. L'entropie croît alors vers les temps négatifs et non plus vers les temps positifs, et toute contradiction disparaît.

Pour dire les choses de manière plus informelle : au temps initial, les particules sont toutes étrangères les unes aux autres. Après un temps t, les particules qui viennent de se rencontrer se connaissent encore, celles qui vont se rencontrer ne se connaissent pas : les particules ont une mémoire du passé et pas du futur. Quand on inverse les vitesses, les particules ont une mémoire du futur et pas du passé, et le temps se met à s'écouler à l'envers !

La légende dit que Boltzmann, confronté au paradoxe du renversement des vitesses, a répondu “Vas-y, renverse-les !” Derrière la boutade, se cache une observation profonde : le renversement des vitesses est une opération qui nous est inaccessible car elle nécessite une connaissance microscopique du système ; et précisément la notion d'entropie émerge de ce que nous ne pouvons agir que macroscopiquement dessus. À partir des années 1950, les expériences d'écho de spin permettaient de voir le paradoxe sous un autre angle [10].

10.5. Non-validité universelle de l'équation de Boltzmann

Ce paradoxe est une variante du précédent. Ayant compris que l'équation de Boltzmann ne s'applique pas après renversement des vitesses, on va exploiter ce fait pour mettre Monsieur Boltzmann en défaut. On refait l'expérience précédente et on choisit comme donnée initiale la distribution obtenue après renversement des vitesses au temps t. On laisse alors le temps agir, et l'équation pertinente n'est bien sûr pas l'équation de Boltzmann.

Ce paradoxe montre effectivement qu'il y a des configurations microscopiques qui *ne mènent pas* à l'équation de Boltzmann. Cependant, et c'est ainsi que Boltzmann argumenta, ces configurations sont rares : précisément, elles font apparaître des corrélations entre vitesses pré-collisionnelles. Ceci n'est pas plus rare que des corrélations entre vitesses post-collisionnelles, mais c'est plus rare que de n'avoir pas de corrélations du tout ! L'équation de Boltzmann est approximativement vraie si l'on part d'une configuration *typique*, c'est-à-dire tirée selon une loi microscopique “fortement chaotique”, mais elle n'est pas vérifiée pour *toutes* les configurations initiales. Une fois ces grands principes énoncés, le travail quantitatif reste à faire.

10.6. Arbitraire de la procédure de Boltzmann

Pour établir l'équation de Boltzmann, on exprime les probabilités de rencontre de particules en fonction des probabilités pré-collisionnelles, ce qui est arbitraire. Si l'on avait utilisé à la place les probabilités post-conditionnelles, on aurait obtenu une équation différente, avec un signe négatif devant l'opérateur de collision ! Pourquoi alors faire confiance à Boltzmann ?

La réponse est encore la même, bien sûr, et dépend du côté de l'origine où l'on se place : pour les temps positifs, ce sont les probabilités pré-collisionnelles qui sont presque factorisées, alors que pour les temps négatifs, ce sont les probabilités post-collisionnelles.

10.7. Démon de Maxwell

Maxwell imagina une expérience de pensée dans laquelle un démon malicieux s'installe dans une boîte à deux compartiments, et manie adroitement un petit clapet de manière à ce qu'il y ait un flux de boules allant du compartiment de droite vers le compartiment de gauche, et pas l'inverse. Ainsi le système évolue vers plus d'ordre, et l'entropie diminue.

Bien sûr, ceci ne peut être considéré comme une objection à la loi de croissance de l'entropie, et l'expérience est destinée à nous faire réfléchir : d'abord le démon devrait faire partie du modèle, et être lui-même soumis à des lois mécaniques réversibles, tenant compte de l'énergie qu'il faut mettre en jeu pour reconnaître qu'une particule s'approche et évaluer sa vitesse, du travail cérébral qu'il effectue, etc. Si l'on fait le bilan complet, on se heurtera, soyons-en sûrs, à la seconde loi de la thermodynamique. [6]

Notons à ce sujet que récemment, des expériences de démon de Maxwell ont pu être réalisés avec des *gaz granulaires* : comme je l'ai moi-même vu avec stupéfaction sur un film expérimental, on part d'un récipient avec deux compartiments séparés verticalement, et une ouverture en haut permettant la communication, on remplit les deux compartiments de particules inélastiques en nombre à peu près égal, on agite le tout automatiquement, et peu à peu l'un des compartiments se vide au profit de l'autre. Un principe sous-jacent est que dans le compartiment le plus plein, l'abondance des collisions entraîne par dissipation d'énergie un refroidissement, et les particules sautent moins haut, ce qui rend de plus en plus difficile leur évasion du compartiment plein. On retrouve à cette occasion le principe déjà mentionné selon lequel une dynamique microscopique dissipative (irréversible) ne mène pas forcément à une augmentation de l'entropie, bien au contraire.

10.8. Convergence et réversibilité

Ce paradoxe est une variante du paradoxe de Loschmidt ; il s'applique aussi bien à la thématique de la relaxation entropique de Boltzmann, qu'à l'amortissement de Landau non linéaire : comment peut-on avoir la convergence quand $t \to +\infty$ si on a réversibilité de la dynamique ? ? La réponse est d'une simplicité enfantine : il y a aussi convergence quand $t \to -\infty$. Pour Vlasov, ceci se fait avec la même équation, et on a donc un phénomène d'homoclinie/hétéroclinie généralisé. Pour Boltzmann, l'équation change selon que l'on considère des temps qui sont antérieurs ou postérieurs à la donnée chaotique.

10.9. Stabilité et réversibilité

Ce paradoxe est plus subtil et s'applique à l'amortissement Landau non linéaire : stabilité asymptotique et réversibilité de la dynamique entraînent automatiquement une instabilité, ce qui semble contradictoire.

Détaillons l'argument. Si l'on a stabilité en temps $t \to +\infty$, soit $f_\infty(v)$ un profil asymptotique stable, que l'on supposera pair. Prenons une solution $\overline{f}(t,x,v)$, inhomogène, qui converge vers $f_\infty(v)$. Choisissons alors pour donnée initiale $f(T,x,-v)$ avec T très grand, on aura ainsi une donnée arbitrairement proche

6. Le Démon de Maxwell a fait l'objet de nombreuses discussions, en particulier par Smoluchowski, Szilard, Gabor, Brillouin, Landauer et Bradbury ; il a aussi inspiré des romanciers comme Pynchon. Un récent article de Binder et Bradbury suggère de traquet de tels concepts dans les mécanismes du vivant.

de $f_\infty(-v) = f_\infty(v)$, et qui nous ramène au bout d'un temps T à la donnée $\overline{f}(0,x,v)$, assez éloignée de $f_\infty(v)$. Autrement dit, la distribution f_∞ est *instable.* Comment cela est-il compatible avec la stabilité ? ?

La réponse, comme expliquée par exemple dans [25], tient dans la topologie : dans le théorème de stabilité asymptotique (amortissement Landau non linéaire), la convergence en temps grand a lieu *au sens de la topologie faible*, avec des oscillations frénétiques dans la distribution de vitesses, qui se compensent localement. Quand on dit qu'une distribution f^0 est stable, cela veut dire que si l'on part près de f^0 au sens de la topologie forte (par exemple analytique ou Gevrey), alors on restera près de f^0 au sens de la topologie faible. La stabilité asymptotique combinée avec la réversibilité impliquent donc l'*instabilité au sens de la topologie faible*, ce qui est parfaitement compatible avec la stabilité au sens de la topologie forte.

10.10. Relaxation conservative

Ce problème est de nature assez générale. L'équation de Vlasov vient avec une préservation de la quantité d'incertitude microscopique (conservation de l'entropie). En outre la distribution au temps $t > 0$ permet de reconstruire exactement la distribution au temps $t = 0$: il suffit de résoudre l'équation de Vlasov après renversement des vitesses. On peut dire que l'équation de Vlasov n'oublie rien ; or la convergence consiste précisément à oublier les péripéties de l'évolution dynamique !

La réponse tient encore dans la convergence faible et les oscillations. Une information va se loger dans ces oscillations, information qui est *invisible* car en pratique nous ne mesurons jamais la fonction de distribution complète, mais des moyennes de cette fonction de distribution (rappelons-nous la citation de Lynden-Bell reproduite dans la fin de la section 7.2). Chaque observable va converger vers sa valeur limite, et il y aura un "oubli". Le champ de force, obtenu comme moyenne de la distribution cinétique, converge vers 0 sans que cela soit contradictoire avec la préservation de l'information : l'information quitte les variables spatiales pour aller dans les variables cinétiques. En particulier, l'entropie spatiale $\int \rho \log \rho$ (où $\rho = \int f\,dv$) tend vers 0, alors que l'entropie cinétique totale $\int f \log f$ est conservée (mais ne converge pas ! L'information est conservée pour tout temps, mais à cause de la convergence faible il y a une perte d'information dans le passage à la limite $t \to \infty$).

De même, dans l'amortissement Landau non linéaire, l'énergie d'interaction $\int W(x-y)\,\rho(x)\,\rho(y)\,dx\,dy$ tend vers 0, et elle est convertie en énergie cinétique (qui peut croître ou décroître en fonction de l'interaction).

10.11. Expérience des échos

Dans cette célèbre expérience [78, 79], on prépare un plasma en état d'équilibre, et on l'excite au temps initial par une impulsion de fréquence spatiale k. Au bout d'un temps τ, après relaxation du plasma, on l'excite à nouveau avec une fréquence spatiale ℓ, colinéaire et de direction opposée à k, d'amplitude plus grande. On attend ensuite et on observe une réponse spontanée du champ électrique du plasma, appelée **écho**, qui se produit à la fréquence spatiale $k + \ell$ et autour du temps $t_e = (|\ell|/|k + \ell|)\tau$.

Cette expérience montre que la distribution cinétique du plasma a gardé trace des impulsions passées : même si le champ de force s'est amorti jusqu'à devenir négligeable, les oscillations cinétiques de la distribution restent présentes, et évoluent au cours du temps. La première impulsion subsiste sous la forme d'oscillations très rapides de période $(|k|t)^{-1}$, la seconde sous forme d'oscillations de période $(|\ell|(t - \tau))^{-1}$. Un calcul que l'on trouvera par exemple dans [111, section 7.3] montre que la distribution continue à osciller rapidement en vitesse, et la force associée reste négligeable, jusqu'à ce que les deux trains d'oscillations se compensent presque exactement, ce qui se manifeste par l'écho.

Références

[1] Alexandre, R., Desvillettes, L., Villani, C., et Wennberg, B. : Entropy dissipation and long-range interactions. Arch. Rat. Mech. Anal. **152**, 327–355 (2000).

[2] Alexandre, R., et Villani, C. : On the Boltzmann equation for long-range interaction and the Landau approximation in plasma physics. Comm. Pure Appl. Math. **55**, 1, 30–70 (2002).

[3] Alexandre, R., et Villani, C. : On the Landau approximation in plasma physics. Ann. Inst. H. Poincaré Anal. Non Linéaire **21**, 1, 61–95 (2004).

[4] Ambrosio, L., et Gangbo, W. : Hamiltonian ODE's in the Wasserstein space of probability measures. Comm. Pure Appl. Math. **51**, 18–53 (2007).

[5] Arkeryd, L., Esposito, R., et Pulvirenti, M. : The Boltzmann equation for weakly inhomogeneous data. Comm. Math. Phys. **111**, 3, 393–407 (1987).

[6] Backus, G. : Linearized plasma oscillations in arbitrary electron distributions. J. Math. Phys. **1**, 178–191, 559 (1960).

[7] Balescu, R. : Irreversible processes in ionized gases. Phys. Fluids **3**, 52–63 (1960).

[8] Balescu, R. *Statistical Mechanics of Charged Particles.* Wiley-Interscience, New York, 1963.

[9] Balian, R. : Entropy, a protean concept. Poincaré Seminar **2**, 119–145 (2003). Birkhäuser Verlag, Basel.

[10] Balian, R. : Information in statistical physics. Studies in Hist. and Phil. of Mod. Phys. **36**, 323–353 (2005).

[11] Baranger, C., et Mouhot, C. : Explicit spectral gap estimates for the linearized Boltzmann and Landau operators with hard potentials. Rev. Matem. Iberoam. **21**, 819–841 (2005).

[12] Batt, J., et Rein, G. : Global classical solutions of the periodic Vlasov-Poisson system in three dimensions. C. R. Acad. Sci. Paris Sér. I Math. **313**, 6, 411–416 (1991).

[13] Ben Arous, G., et Zeitouni, O. : Increasing propagation of chaos for mean field models. Ann. Inst. H. Poincaré Probab. Statist. **35**, 1, 85–102 (1999).

[14] Bernstein, I.B., Greene, J.M., et Kruskal, M.D. : Exact nonlinear plasma oscillations. Phys. Rev. **108**, 3, 546–550 (1957).

[15] Binney, J., et Tremaine, S. : *Galactic Dynamics*, second ed. Princeton Series in Astrophysics. Princeton University Press, 2008.

[16] Bobylev, A.V., et Cercignani, C. : On the rate of entropy production for the Boltzmann equation. J. Statist. Phys. **94**, 3–4, 603–618 (1999).

[17] Bobylev, A.V., et Cercignani, C. : Exact eternal solutions of the Boltzmann equation. J. Statist. Phys. **106**, 5–6, 1019–1038 (2002).

[18] Bobylev, A.V., et Cercignani, C. : Self-similar solutions of the Boltzmann equation and their applications. J. Statist. Phys. **106**, 5–6, 1039–1071 (2002).

[19] Bogoliubov, N.N. : Problems of a dynamical theory in statistical physics. Studies in Statistical Mechanics **1**, 1–118 (1962). Traduction de l'original russe de 1946.

[20] Bolley, F., Guillin, A., et Villani, C. : Quantitative concentration inequalities for empirical measures on non-compact spaces. Probab. Theory Related Fields **137**, 3–4, 541–593 (2007).

[21] Boltzmann, L. : Weitere Studien über das Wärmegleichgewicht unter Gasmolekülen. *Sitzungsberichte Akad. Wiss.* 66 (1872), 275–370. Traduction : Further studies on the thermal equilibrium of gas molecules, in Kinetic Theory **2**, 88–174, Éd. S.G. Brush, Pergamon, Oxford (1966).

[22] Boltzmann, L. : *Lectures on gas theory.* University of California Press, Berkeley, 1964. Traduit en anglais par Stephen G. Brush. Réimpression de l'édition de 1896–1898. Dover Publications, 1995.

[23] Bouchut, F. : Introduction à la théorie mathématique des équations cinétiques. In *Session "L'Etat de la Recherche" de la SMF* (1998). "Kinetic equations and asymptotic theory", F. Bouchut, F. Golse, M. Pulvirenti, coll. "Series in Appl. Math.", Gauthier-Villars, 2000.

[24] Braun, W., et Hepp, K. : The Vlasov dynamics and its fluctuations in the $1/N$ limit of interacting classical particles. Commun. Math. Phys. **56**, 125–146 (1977).

[25] Caglioti, E., et Maffei, C. : Time asymptotics for solutions of Vlasov–Poisson equation in a circle. J. Statist. Phys. **92**, 1–2, 301–323 (1998).

[26] Carleman, T. : Sur la théorie de l'équation intégrodifférentielle de Boltzmann. Acta Math. **60**, 369–424 (1932).

[27] Carleman, T. : *Problèmes Mathématiques dans la Théorie Cinétique des Gaz.* Almqvist & Wiksell, 1957.

[28] Carlen, E.A., et Carvalho, M.C. : Strict entropy production bounds and stability of the rate of convergence to equilibrium for the Boltzmann equation. J. Statist. Phys. **67**, 3–4, 575–608 (1992).

[29] Carlen, E.A., Carvalho, M.C. et Gabetta, E. : Central limit theorem for Maxwellian molecules and truncation of the Wild expansion. Comm. Pure Appl. Math. **53**, 370–397 (2000).

[30] Carlen, E.A., Carvalho, M.C. et Gabetta, E. : On the relation between rates of relaxation and convergence of Wild sums for solutions of the Kac equation. J. Funct. Anal. **220**, 362–387 (2005).

[31] Carlen, E.A., Carvalho, M.C. et Loss, M. : Determination of the spectral gap for Kac's master equation and related stochastic evolution. Acta Math. **191**, 1–54 (2003).

[32] Carlen, E.A., Carvalho, M.C., Le Roux, J., Loss, M., et Villani, C. Entropy and chaos in the Kac model. *Kinet. Relat. Models* **3**, 1, 85–122 (2010).

[33] Carlen, E.A., Gabetta, E., et Toscani, G. : Propagation of smoothness and the rate of exponential convergence to equilibrium for a spatially homogeneous Maxwellian gas. Comm. Math. Phys. **199**, 3, 521–546 (1999).

[34] Carlen, E.A., et Lu, M. : Fast and slow convergence to equilibrium Maxwellian molecules via Wild sums J. Stat. Phys. **112**, 1–2, 59–134 (2003).

[35] Carleson, L. : Some analytic problems related to statistical mechanics. In *Euclidean Harmonic Analysis*, J.J. Benedetto, éd. (Univ. of Maryland, 1979), Lect. Notes Math. **779**, 5–45.

[36] Cercignani, C. : On the Boltzmann equation with cutoff potentials. Phys. Fluids **10**, 2097 (1967).

[37] Cercignani, C. : On the Boltzmann equation for rigid spheres. Transport Theory Statist. Phys. **2**, 3, 211–225 (1972).

[38] Cercignani, C. : *The Boltzmann equation and its applications*. Springer-Verlag, New York, 1988.

[39] Cercignani, C. : *Rarefied gas dynamics*. Cambridge University Press, Cambridge, 2000. From basic concepts to actual calculations.

[40] Cercignani, C., Illner, R., et Pulvirenti, M. : *The Mathematical Theory of Dilute Gases*. Springer, New York, 1994.

[41] Cover, T.M., et Thomas, J.A. : *Elements of information theory*. John Wiley & Sons Inc., New York, 1991. A Wiley-Interscience Publication.

[42] Delcroix, J.-L., et Bers, A. : *Physique des plasmas* (2 Vol.). InterEditions/ CNRS Éditions, 1994.

[43] Dembo, A., et Zeitouni, O. : *Large deviations techniques and applications*, 2nde éd. Springer-Verlag, New York, 1998.

[44] Desvillettes, L. : Entropy dissipation rate and convergence in kinetic equations. Comm. Math. Phys. **123**, 4, 687–702 (1989).

[45] Desvillettes, L., et Villani, C. : On the spatially homogeneous Landau equation for hard potentials. Part I : existence, uniqueness and smoothness. Comm. Partial Differential Equations **25**, 1–2, 179–259 (2000).

[46] Desvillettes, L., et Villani, C. : On the trend to global equilibrium for spatially inhomogeneous kinetic systems : the Boltzmann equation. Invent. Math. **159**, 2, 245–316 (2005).

[47] DiPerna, R., et Lions, P.-L. : On the Cauchy problem for the Boltzmann equation : Global existence and weak stability. Ann. of Math. (2) **130**, 312–366 (1989).

[48] Dobrušin, R.L. : Vlasov equations. Funktsional. Anal. i Prilozhen. **13**, 2, 48–58, 96 (1979).

[49] Dudley, R.M. : *Real analysis and probability*, vol. 74 de *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2002. Réédition révisée de l'original de 1989.

[50] Filbet, F., Mouhot, C., et Pareschi, L. : Solving the Boltzmann equation in $N \log_2 N$. SIAM J. Sci. Comput. **28**, 3, 1029–1053 (2006).

[51] Glassey, R.T. : *The Cauchy problem in kinetic theory*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1996.

[52] Glassey, R., et Schaeffer, J. : Time decay for solutions to the linearized Vlasov equation. Transport Theory Statist. Phys. **23**, 4, 411–453 (1994).

[53] Glassey, R., et Schaeffer, J. : On time decay rates in Landau damping. Comm. Partial Differential Equations **20**, 3–4, 647–676 (1995).

[54] Golse, F. : From kinetic to macroscopic models. In *Session "L'Etat de la Recherche" de la SMF* (1998), "Kinetic equations and asymptotic theory", F. Bouchut, F. Golse, M. Pulvirenti, coll. "Series in Appl. Math.", Gauthier-Villars, 2000.

[55] Golse, F., et Saint-Raymond, L. : The Navier–Stokes limit of the Boltzmann equation for bounded collision kernels. Invent. Math. **155**, 1, 81–161 (2004).

[56] Grad, H. : On Boltzmann's H-theorem. J. Soc. Indust. Appl. Math. **13**, 1, 259–277 (1965).

[57] Grad, H. : Principles of the kinetic theory of gases. In *Flügge's Handbuch des Physik*, vol. XII. Springer-Verlag, 1958, pp. 205–294.

[58] Gualdani, M.P., Mischler, S. et Mouhot, C. : Factorization for non-symmetric operators and exponential H-theorem. Preprint, available online at http ://hal.archives-ouvertes.fr/ccsd-00495786

[59] Guo, Y. : The Landau equation in a periodic box. Commun. Math. Phys. **231**, 391–434 (2002).

[60] Guo, Y., et Strain, R.M. : Exponential decay for soft potentials near Maxwellian. Arch. Ration. Mech. Anal. **187**, 2, 287–339 (2008).

[61] Horst, E. : On the asymptotic growth of the solutions of the Vlasov–Poisson system. Math. Methods Appl. Sci. **16**, 75–86 (1993).

[62] Illner, R., et Pulvirenti, M. : Global validity of the Boltzmann equation for a two-dimensional rare gas in vacuum. Comm. Math. Phys. **105**, 2, 189–203 (1986). "Erratum and improved result", Comm. Math. Phys. **121**, 1, 143–146 (1989).

[63] Illner, R., et Pulvirenti, M. : A derivation of the BBGKY-hierarchy for hard sphere particle systems. Transport Theory Statist. Phys. **16**, 7, 997–1012 (1987).

[64] Isichenko, M. : Nonlinear Landau damping in collisionless plasma and inviscid fluid. Phys. Rev. Lett. **78**, 12, 2369–2372 (1997).

[65] Jabin, P.-E., et Hauray, M. : N-particles approximation of the Vlasov equations with singular potential. Arch. Ration. Mech. Anal. **183**, 3, 489–524 (2007).

[66] Janvresse, E. : Spectral gap for Kac's model of Boltzmann equation. Ann. Probab. **29**, 1, 288–304 (2001).

[67] Kac, M. : Foundations of kinetic theory. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability,* 1954–1955*, vol. III* (Berkeley and Los Angeles, 1956), University of California Press, pp. 171–197.

[68] Landau, L.D. : Die kinetische Gleichung für den Fall Coulombscher Wechselwirkung. *Phys. Z. Sowjet* 10 (1936), 154. English translation : The transport equation in the case of Coulomb interactions, in *Collected papers of L.D. Landau*, edited and with an introduction by D. ter Haar, Pergamon Press, 1965, pp.163–170.

[69] Landau, L.D. : On the vibration of the electronic plasma. J. Phys. USSR **10**, 25 (1946). Traduction en anglais dans JETP **16**, 574. Reproduit dans *Collected papers of L.D. Landau*, édité par et avec une introduction de D. ter Haar, Pergamon Press, 1965, pp. 445–460 ; et dans *Men of Physics : L.D. Landau*, Vol. 2, Pergamon Press, D. ter Haar, ed. (1965).

[70] Lanford, O.E. : Time evolution of large classical systems. In *Dynamical systems, theory and applications* (*Rencontres, Battelle Res. Inst., Seattle, Wash.,* 1974). Springer, Berlin, 1975, pp. 1–111. Lecture Notes in Phys., Vol. 38.

[71] Lenard, A. : On Bogoliubov's kinetic equation for a spatially homogeneous plasma. Ann. Physics **10**, 390–400 (1960).

[72] Lifshitz, E.M., et Pitaevskiĭ, L.P. : *Course of theoretical physics* [*"Landau-Lifshits"*]. *Vol. 10.* Pergamon Press, Oxford, 1981. Translated from the Russian by J.B. Sykes and R.N. Franklin.

[73] Lin, Z., et Zeng, C. : BGK waves and nonlinear Landau damping. Preprint, 2010.

[74] Liverani, C., et Olla, S. : Toward the Fourier law for a weakly interacting anharmonic crystal. Preprint, 2010.

[75] Loschmidt, J. : Über den Zustand des Wärmegleichgewichtes einse Systems von Körpern mit Rücksicht auf die Schwerkraft. Wien. Ber. **73**, 128 (1876).

[76] Lynden-Bell, D. : The stability and vibrations of a gas of stars. Mon. Not. R. astr. Soc. **124**, 4, 279–296 (1962).

[77] Lynden-Bell, D. : Statistical mechanics of violent relaxation in stellar systems. Mon. Not. R. astr. Soc. **136**, 101–121 (1967).

[78] Malmberg, J., et Wharton, C. : Collisionless damping of electrostatic plasma waves. Phys. Rev. Lett. **13**, 6, 184–186 (1964).

[79] Malmberg, J., Wharton, C., Gould, R., et O'Neil, T. : Plasma wave echo experiment. Phys. Rev. Letters **20**, 3, 95–97 (1968).

[80] Maslen, D. : The eigenvalues of Kac's master equation. Math. Zeit. **243**, 291–331 (2003).

[81] Maxwell, J.C. : On the dynamical theory of gases. Philos. Trans. Roy. Soc. London Ser. A **157**, 49–88 (1867).

[82] McKean, H.J. : Speed of approach to equilibrium for Kac's caricature of a Maxwellian gas. Arch. Rational Mech. Anal. **21**, 343–367 (1966).

[83] Mouhot, C. : Rate of convergence to equilibrium for the spatially homogeneous Boltzmann equation with hard potentials. Comm. Math. Phys. **261**, 629–672 (2006).

[84] Mouhot, C., et Strain, R.M. : Spectral gap and coercivity estimates for linearized Boltzmann collision operators without angular cutoff. J. Math. Pures Appl. **87**, 5, 515–535 (2007).

[85] Mouhot, C., et Villani, C. : Regularity theory for the spatially homogeneous Boltzmann equation with cut-off. Arch. Rational Mech. Anal. **173**, 2, 169–212 (2004).

[86] Mouhot, C., et Villani, C. : On Landau damping. À paraître dans *Acta Math.* Disponible en ligne à `arXiv :0904.2760`

[87] O'Neil, T. : Collisionless damping of nonlinear plasma oscillations. Phys. Fluids **8**, 12, 2255–2262 (1965).

[88] Neunzert, H. : An introduction to the nonlinear Boltzmann–Vlasov equation. In *Kinetic theories and the Boltzmann equation*, C. Cercignani, ed., vol. 1048 of Lecture Notes in Math., Springer, Berlin, Heidelberg, 1984, pp. 60–110.

[89] Penrose, O. : Electrostatic instability of a non-Maxwellian plasma. Phys. Fluids **3**, 258–265 (1960).

[90] Poincaré, H. : Le mécanisme et l'expérience. Revue de Métaphysique et de Morale **I**, 534–537 (1893).

[91] Pulvirenti, M. : From particle to transport equations. In *Session "L'Etat de la Recherche" de la SMF* (1998). "Kinetic equations and asymptotic theory", F. Bouchut, F. Golse, M. Pulvirenti, coll. "Series in Appl. Math.", Gauthier-Villars, 2000.

[92] Robert, R. : Statistical mechanics and hydrodynamical turbulence. In *Proceedings of the International Congress of Mathematicians*, Vol. 1, 2 (Zürich, 1994) (Basel, 1995), Birkhäuser, pp. 1523–1531.

[93] Ryutov, D.D. : Landau damping : half a century with the great discovery. Plasma Phys. Control. Fusion **41**, A1–A12 (1999).

[94] Saint-Raymond, L. : *Hydrodynamic limits of the Boltzmann equation.* Lectures at SISSA, Trieste (2006), Springer Monograph, à paraître.

[95] Shannon, C.E., et Weaver, W. : *The mathematical theory of communication.* University of Illinois Press, Urbana, Illinois, 1949.

[96] Spohn, H. : *Large scale dynamics of interacting particles.* Texts and Monographs in Physics. Springer-Verlag, Berlin, 1991.

[97] Strain, R.M. : On the linearized Balescu–Lenard equation. Comm. Partial Differential Equations **32**, 1551–1586 (2007).

[98] Sturm, K.-Th. : Entropic measure on multidimensional spaces. In *Stochastic analysis, random fields and applications VI*, Dalang, R., Dozzi, M., and Russo, F., Eds. Progress in Probability, Birkhäuser.

[99] Sznitman, A.-S. : Equations de type de Boltzmann, spatialement homogènes. Z. Wahrsch. Verw. Gebiete **66**, 559–562 (1984).

[100] Sznitman, A.-S. : *Topics in propagation of chaos.* In *École d'Été de Probabilités de Saint-Flour XIX* (1989), Lect. Notes in Math. **1464** (1991), Springer, Berlin, pp. 165–251.

[101] Tanaka, H. : An inequality for a functional of probability distributions and its application to Kac's one-dimensional model of a Maxwellian gas. Z. Wahrscheinlichkeitstheorie und Verw. Gebiete **27**, 47–52 (1973).

[102] Toscani, G., et Villani, C. : Sharp entropy dissipation bounds and explicit rate of trend to equilibrium for the spatially homogeneous Boltzmann equation. Comm. Math. Phys. **203**, 3, 667–706 (1999).

[103] Turkington, B. : Statistical equilibrium measures and coherent states in two-dimensional turbulence. Comm. Pure Appl. Math. **52**, 7, 781–809 (1999).

[104] Villani, C. : *A review of mathematical topics in collisional kinetic theory.* In *Handbook of Mathematical Fluid Dynamics I*, 71–305, S. Friedlander et D. Serre, éds, North-Holland, Amsterdam, 2002.

[105] Villani, C. Cercignani's conjecture is sometimes true and always almost true. Commun. Math. Phys. **234**, 455–490 (2003).

[106] Villani, C. : Mathematics of granular materials. J. Stat. Phys. **124**, no. 2–4, 781–822 (2006).

[107] Villani, C. *H*-Theorem and beyond : Boltzmann's entropy in today's mathematics. Actes du colloque "Boltzmann's legacy", Erwin-Schrödinger Institute (Vienne, juillet 2007).

[108] Villani, C. : Hypocoercive diffusion operators. Texte de ma conférence au Congrès International des Mathématiciens (Madrid, août 2006).

[109] Villani, C. : Entropy production and convergence to equilibrium. Notes pour une série de cours à l'Institut Henri Poincaré, Paris (automne 2001). In *Entropy methods for the Boltzmann equation*, 1–70, Lecture Notes in Math., Vol. 1916, Springer, Berlin, 2008.

[110] Villani, C. : Hypocoercivity. Mem. Amer. Math. Soc. **202**, no. 950 (2009).

[111] Villani, C. : *Landau damping.* Notes de cours, CEMRACS 2010. Disponible sur www.math.univ-lyon1.fr/~villani

[112] Villani, C. : Is there any backward solution of the Boltzmann equation ? Travail non publié, disponible sur www.math.univ-lyon1.fr/~villani

[113] Vlasov, A.A. : On the oscillation properties of an electron gas. Zh. Èksper. Teoret. Fiz. **8**, 291–318 (1938).

[114] Zermelo, E. : Über einen Satz der Dynamik und die mechanische Wärmetheorie. Annalen der Physik **54**, 485–494 (1896).

Cédric Villani
Université de Lyon
& Institut Henri Poincaré
11 rue Pierre et Marie Curie
F-75231 Paris Cedex 05, FRANCE

Time, 145–172
© 2013 Springer Basel

Poincaré Seminar 2010

Equalities and Inequalities: Irreversibility and the Second Law of Thermodynamics at the Nanoscale

Christopher Jarzynski

Abstract. The reason we never observe violations of the second law of thermodynamics is a matter of statistics: When very many degrees of freedom are involved, the odds are overwhelmingly stacked against the possibility of seeing significant deviations away from the mean behavior. As we turn our attention to smaller systems, however, statistical fluctuations become more prominent. In recent years it has become apparent that the fluctuations of systems far from thermal equilibrium are not mere background noise, but satisfy strong, useful, and unexpected properties. In particular, a proper accounting of fluctuations allows us to rewrite familiar inequalities of macroscopic thermodynamics as equalities. This review describes some of this progress, and argues that it has refined our understanding of irreversibility and the second law.

1. Introduction

On anyone's list of the supreme achievements of the nineteenth-century science, both Maxwell's equations and the second law of thermodynamics surely rank high. Yet while Maxwell's equations are widely viewed as done, dusted, and uncontroversial, the second law still provokes lively arguments, long after Carnot published his *Reflections on the Motive Power of Fire* (1824) and Clausius articulated the increase of entropy (1865). The puzzle at the core of the second law is this: how can microscopic equations of motion that are symmetric with respect to time-reversal give rise to macroscopic behavior that clearly does not share this symmetry? Of course, quite apart from questions related to the origin of "time's arrow", there is a nuts-and-bolts aspect to the second law. Together with the first law, it provides a set of tools that are indispensable in practical applications ranging from the design of power plants and refrigeration systems to the analysis of chemical reactions.

The past few decades have seen growing interest in applying these laws and tools to individual microscopic systems, down to nanometer length scales. Much of this interest arises at the intersection of biology, chemistry and physics, where there has been tremendous progress in uncovering the mechanochemical details of biomolecular processes. [1] For example, it is natural to think of the molecular complex $\phi 29$ – a motor protein that crams DNA into the empty shell of a virus – as a nanoscale machine that generates torque by consuming free energy. [2] The development of ever more sophisticated experimental tools to grab, pull, and otherwise bother individual molecules, and the widespread use of all-atom simulations to study the dynamics and the thermodynamics of molecular systems, have also contributed to the growing interest in the "thermodynamics of small systems", as the field is sometimes called. [3]

Since the rigid, prohibitive character of the second law emerges from the statistics of huge numbers, we might expect it to be enforced somewhat more leniently in systems with relatively few degrees of freedom. To illustrate this point, consider the familiar gas-and-piston setup, in which the gas of $N \sim 10^{23}$ molecules begins in a state of thermal equilibrium, inside a container enclosed by adiabatic walls. If the piston is rapidly pushed into the gas and then pulled back to its initial location, there will be a net increase in the internal energy of the gas. That is,

$$W > 0, \tag{1}$$

where W denotes the work performed by the agent that manipulates the piston. This inequality is not mandated by the underlying dynamics: there certainly exist microscopically viable N-particle trajectories for which $W < 0$. However, the probability to observe such trajectories becomes fantastically small for large N. By contrast, for a "gas" of only a few particles, we would not be surprised to observe – once in a rare while, perhaps – a negative value of work, though we still expect Eq. (1) to hold on average:

$$\langle W \rangle > 0. \tag{2}$$

The angular brackets here and below denote an average over many repetitions of this hypothetical process, with the tiny sample of gas re-equilibrated prior to each repetition.

This example suggests the following perspective: as we apply the tools of thermodynamics to ever smaller systems, the second law becomes increasingly blurred. Inequalities such as Eq. (1) remain true on average, but statistical fluctuations around the average become ever more important as fewer degrees of freedom come into play.

This picture, while not wrong, is incomplete. It encourages us to dismiss the fluctuations in W as uninteresting noise that merely reflects poor statistics (small N). As it turns out, these fluctuations themselves satisfy rather strong, interesting and useful laws. For example, Eq. (2) can be replaced by the *equality*,

$$\langle e^{-W/k_B T} \rangle = 1, \tag{3}$$

where T is the temperature at which the gas is initially equilibrated, and k_B is Boltzmann's constant. If we additionally assume that the piston is manipulated in a time-symmetric manner, e.g., pushed in at a constant speed and then pulled out at the same speed, then the statistical distribution of work values $\rho(W)$ satisfies the symmetry relation

$$\frac{\rho(+W)}{\rho(-W)} = e^{W/k_B T}. \tag{4}$$

The validity of these results depends neither on the number of molecules in the gas, nor (surprisingly!) on the rate at which the process is performed.

I have used the gas and piston out of convenience and familiarity, but the predictions illustrated here by Eqs. (3) and (4) – and expressed more generally by Eqs. (15) and (30) below – are not specific to this particular example. They apply to any system that is driven away from equilibrium by the variation of mechanical parameters, under relatively standard assumptions regarding the initial equilibrium state and the microscopic dynamics. Moreover, they belong to a larger collection of recently derived theoretical predictions, which pertain to fluctuations of work, [4–9] entropy production, [10–18] and other quantities [19, 20] in systems far from thermal equilibrium. While these predictions go by various names, both descriptive and eponymous, the term *fluctuation theorems* has come to serve as a useful label encompassing the entire collection of results. There is by now a large body of literature on fluctuation theorems, including reviews and pedagogical treatments. [3, 21–38]

In my view these are not results that one might naturally have obtained, by starting with a solid understanding of macroscopic thermodynamics and extrapolating down to small system size. Rather, they reveal genuinely new, nanoscale features of the second law. My aim in this review is to elaborate on this assertion. Focusing on those fluctuation theorems that describe the relationship between work and free energy – these are sometimes called *nonequilibrium work relations* – I will argue that they have refined our understanding of dissipation, hysteresis, and other hallmarks of thermodynamic irreversibility. Most notably, when fluctuations are taken into account, inequalities that are related to the second law (e.g., Eqs. (5), (24), (28), (35)) can be rewritten as equalities (Eqs. (15), (25), (30), (31)). Among the "take-home messages" that emerge from these developments are the following:

- Equilibrium information is subtly encoded in the microscopic response of a system driven far from equilibrium.
- Surprising symmetries lurk beneath the strong hysteresis that characterizes irreversible processes.
- Physical measures of dissipation are related to information-theoretic measures of irreversibility.
- The ability of thermodynamics to set the direction of time's arrow can be quantified.

Moreover, these results have practical applications in computational thermodynamics and in the analysis of single-molecule manipulation experiments, as discussed briefly in Section 8.

Section 2 of this review introduces definitions and notation, and specifies the framework that will serve as a paradigm of a thermodynamic process. Sections 3–6 address the four points listed above, respectively. Section 7 discusses how these results relate to fluctuation theorems for entropy production. Finally, I conclude in Section 8.

2. Background and setup

This section establishes the basic framework that will be considered, and introduces the definitions and assumptions to be used in later sections.

2.1. Macroscopic thermodynamics and the Clausius inequality

Throughout this review, the following will serve as a paradigm of a nonequilibrium thermodynamic process.

Consider a finite, classical system of interest in contact with a thermal reservoir at temperature T (e.g., a rubber band surrounded by air), and let λ denote some externally controlled parameter of the system (the length of the rubber band). I will refer to λ as a *work parameter*, since by varying it we perform work on the system. The notation $[\lambda, T]$ will specify an equilibrium state of the system. Now imagine that the system of interest is prepared in equilibrium with the reservoir, at fixed $\lambda = A$, that is in state $[A, T]$. Then from time $t = 0$ to $t = \tau$ the system is perturbed, perhaps violently, by varying the parameter with time, ending at a value $\lambda = B$. (The rubber band is rapidly stretched.) Finally, from $t = \tau$ to $t = \tau^*$ the work parameter is held fixed at $\lambda = B$, allowing the system to re-equilibrate with the thermal reservoir and thus relax to the state $[B, T]$.

In this manner the system is made to evolve from one equilibrium state to another, but in the interim it is generally driven away from equilibrium. The Clausius inequality of classical thermodynamics [39] then predicts that the external work performed on the system will be no less than the free energy difference between the terminal states:

$$W \geq \Delta F \equiv F_{B,T} - F_{A,T} \tag{5}$$

Here $F_{\lambda,T}$ denotes the Helmholtz free energy of the state $[\lambda, T]$. When the parameter is varied slowly enough that the system remains in equilibrium with the reservoir at all times, then the process is reversible and isothermal, and $W = \Delta F$.

Eq. (5) is the essential statement of the second law of thermodynamics that will apply in Sections 3–6 of this review. Of course, not all thermodynamic processes fall within this paradigm, nor is Eq. (5) the broadest formulation of the Clausius inequality. However, since complete generality can impede clarity, I will focus on the class of processes described above. Most of the results presented in the following sections apply also to more general thermodynamic processes – such as those involving multiple thermal reservoirs or nonequilibrium initial states – as I will briefly mention in Section 7.

Three comments are now in order, before moving down to the nanoscale.

(1) As the system is driven away from equilibrium, its temperature may change or become ill-defined. The variable T, however, will always denote the *initial* temperature of the system and thermal reservoir.
(2) No external work is performed on the system during the re-equilibration stage, $\tau < t < \tau^*$, as λ is held fixed. In this sense the re-equilibration stage is somewhat superfluous: Eq. (5) remains valid if the process is considered to end at $t = \tau$ – even if the system has not yet re-equilibrated with the reservoir! – provided we always define ΔF to be a free energy difference between the equilibrium states $[A, T]$ and $[B, T]$.
(3) While in general it is presumed that the system remains in thermal contact with the reservoir for $0 < t < \tau$, the results discussed in this review are also valid if the system is isolated from the reservoir during this interval.

2.2. Microscopic definitions of work and free energy

Now let us "scale down" this paradigm to small systems, with an eye toward incorporating statistical fluctuations. Consider a framework in which the system of interest and the thermal reservoir are represented as a large collection of microscopic, classical degrees of freedom. The work parameter λ is an additional coordinate describing the position or orientation of a body – or some other mechanical variable such as the location of a laser trap in a single-molecule manipulation experiment [27] – that interacts with the system of interest, but is controlled by an external agent. This framework is illustrated with a toy model in Fig. 1. Here the system of interest consists of the three particles represented as open circles, whose coordinates z_i give distances from the fixed wall. The work parameter is the fourth particle, depicted as a shaded circle at a distance λ from the wall.

Let the vector $\mathbf{x}$ denote a microscopic state of the system of interest, that is the configurations and momenta of its microscopic degrees of freedom; and let $\mathbf{y}$ similarly denote a microstate of the thermal reservoir. The Hamiltonian for this collection of classical variables is assumed to take the form

$$\mathcal{H}(\mathbf{x}, \mathbf{y}; \lambda) = H(\mathbf{x}; \lambda) + H_{\text{env}}(\mathbf{y}) + H_{\text{int}}(\mathbf{x}, \mathbf{y}) \tag{6}$$

where $H(\mathbf{x}; \lambda)$ represents the energy of the system of interest – including its interaction with the work parameter – $H_{\text{env}}(\mathbf{y})$ is the energy of the thermal environment, and $H_{\text{int}}(\mathbf{x}, \mathbf{y})$ is the energy of interaction between system and environment. For the toy model in Fig. 1, $\mathbf{x} = (z_1, z_2, z_3, p_1, p_2, p_3)$ and we assume

$$H(\mathbf{x}; \lambda) = \sum_{i=1}^{3} \frac{p_i^2}{2m} + \sum_{k=0}^{3} u(z_{k+1} - z_k) \tag{7}$$

where $u(\cdot)$ is a pairwise interaction potential, $z_0 \equiv 0$ is the position of the wall, and $z_4 \equiv \lambda$ is the work parameter.

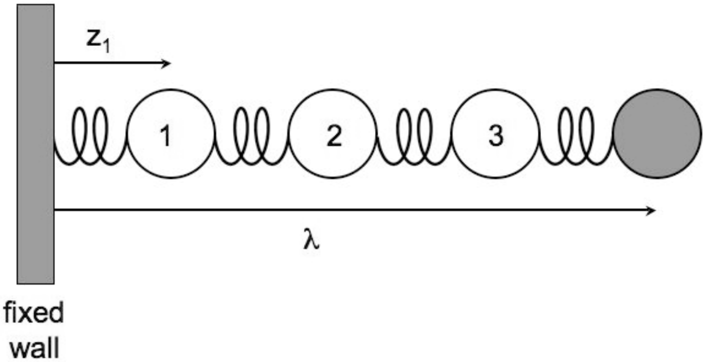


FIGURE 1. Illustrative model. The numbered circles constitute a three-particle system of interest, with coordinates (z_1, z_2, z_3) giving the distance of each particle from the fixed wall, as shown for z_1. The shaded particle is the work parameter, whose position λ is manipulated externally. The springs represent particle-particle (or particle-wall) interactions. The system of interest interacts with a thermal reservoir whose degrees of freedom are not shown.

Now imagine a process during which the external agent manipulates the work parameter according to a protocol $\lambda(t)$. As the parameter is displaced by an amount $\mathrm{d}\lambda$, the change in the value of H due to this displacement is

$$đW \equiv \mathrm{d}\lambda \, \frac{\partial H}{\partial \lambda}(\mathbf{x};\lambda) \tag{8}$$

Since $\mathrm{d}\lambda \cdot \partial H/\partial\lambda$ is the work required to displace the coordinate λ against a force $-\partial H/\partial\lambda$, we interpret Eq. (8) to be the work performed by the external agent in effecting this small displacement. [40] Over the entire process, the work performed by the external agent is:

$$W = \int đW = \int_0^\tau \mathrm{d}t \, \dot{\lambda} \, \frac{\partial H}{\partial \lambda}(\mathbf{x}(t);\lambda(t)) \tag{9}$$

where the trajectory $\mathbf{x}(t)$ describes the evolution of the system of interest. This will be the microscopic definition of work that will be used throughout this review. (For discussions and debates related to this definition, see [37, 40–49].)

Let us now focus on the free energy difference ΔF appearing in Eq. (5). In statistical physics an equilibrium state is represented by a probability distribution rather than by a single microscopic state. If the interaction energy $H_{\rm int}$ in Eq. (6) is sufficiently weak – as usually assumed in textbook discussions of macroscopic systems – then this distribution is given by the Boltzmann-Gibbs formula,

$$p^{\rm eq}_{\lambda,T}(\mathbf{x}) = \frac{1}{Z_{\lambda,T}} \exp[-H(\mathbf{x};\lambda)/k_B T] \tag{10}$$

where

$$Z_{\lambda,T} = \int \mathrm{d}\mathbf{x}\, \exp[-H(\mathbf{x};\lambda)/k_B T] \tag{11}$$

is the classical partition function. If H_{int} is too large to be neglected, then the equilibrium distribution takes the modified form

$$p^{\mathrm{eq}}_{\lambda,T} \propto \exp(-H^*/k_B T) \quad , \quad H^*(\mathbf{x};\lambda) = H(\mathbf{x};\lambda) + \phi(\mathbf{x};T) \tag{12}$$

where $\phi(\mathbf{x};T)$ is the free-energetic cost of inserting the system of interest into its thermal surroundings, e.g., associated with the rearrangement of water required to accommodate the presence of a biomolecule. For purpose of this review, the distinction between Eqs. (10) and (12) is not terribly relevant. I will use the more familiar Eq. (10), which applies to the weak-coupling limit (small H_{int}), with the understanding that all the results discussed below are equally valid in the case of strong coupling, provided H is replaced by H^*. See [50] for a more detailed discussion. The free energy associated with this equilibrium state is

$$F_{\lambda,T} = -k_B T \ln Z_{\lambda,T} \tag{13}$$

With these elements in place, imagine a microscopic analogue of the process described in Section 2.1. The system of interest is prepared in equilibrium with the reservoir, at $\lambda = A$. From $t = 0$ to $t = \tau$ the system evolves with time as the work parameter is varied from $\lambda(0) = A$ to $\lambda(\tau) = B$. By considering infinitely many repetitions of this process, we arrive at a statistical ensemble of realizations of the process, which can be pictured as a swarm of independently evolving trajectories, $\mathbf{x}_1(t)$, $\mathbf{x}_2(t), \ldots$. For each of these we can compute the work, W_1, $W_2, \ldots$ (Eq. (9)). Letting $\rho(W)$ denote the distribution of these work values, it is reasonable to expect that Eq. (5) in this case becomes a statement about the mean of this distribution, namely

$$\langle W \rangle \equiv \int \mathrm{d}W\, \rho(W)\, W \geq \Delta F. \tag{14}$$

As suggested earlier, this inequality is correct but it is not the entire story.

2.3. The need to model

Although the laws of macroscopic thermodynamics can be stated without reference to underlying equations of motion, when we study how these laws might apply to a microscopic system away from equilibrium we must typically specify the equations we use to model its evolution. These equations represent approximations of physical reality, and the choice inevitably reflects certain assumptions. Eq. (6) suggests one approach: treat the system and reservoir as an isolated, classical system evolving in the full phase space $(\mathbf{x}, \mathbf{y})$ under a time-dependent Hamiltonian $\mathcal{H}(\mathbf{x}, \mathbf{y}; \lambda(t))$. The results discussed in Sections 3–6 can all be obtained within this framework. Alternatively, we can treat the reservoir implicitly, by writing down effective equations of motion for just the system variables, $\mathbf{x}$. Examples include

Langevin dynamics, the Metropolis algorithm, Nosé-Hoover dynamics and its variants, the Andersen thermostat, and deterministic equations based on Gauss's principle of least constraint. [23, 34] As with the Hamiltonian approach, the results discussed below can be derived for each of these model dynamics. This suggests that the results themselves are rather robust: they do not depend sensitively on how the microscopic dynamics are modeled.

Since the aim of this review is to describe what the second law of thermodynamics "looks like" in the presence of fluctuations, full-blown derivations of fluctuation theorems and work relations will not be provided. However, in Sections 3 and 4, in addition to describing various work relations and their connections to the second law, I will sketch how several of them can be derived for the toy system shown in Fig. 1, in the physical context mentioned by the final comment in Section 2.1: the system is thermally isolated during the interval $0 < t < \tau$. The aim here is to convey some idea of the theoretical foundations of these results, without exploring the technical details that accompany an explicit treatment of the reservoir. [50]

3. Equilibrium information from nonequilibrium fluctuations

Thermodynamics accustoms us to the idea that irreversible processes are described by inequalities, such as $W \geq \Delta F$. One of the surprises of recent years is that if we pay attention to fluctuations, then such relationships can be recast as equalities. In particular, the *nonequilibrium work relation* [6, 7] states that

$$\langle e^{-W/k_BT} \rangle = e^{-\Delta F/k_BT}, \tag{15}$$

where (as above) T is the initial temperature of the system and thermal reservoir, and angular brackets denote an ensemble average over realizations of the process. This result has been derived in various ways, using an assortment of equations of motion to model the microscopic dynamics [6–9, 17, 18, 50–63], and has been confirmed experimentally. [64–67] In the following paragraph I will sketch how it can be obtained for the toy model of Fig. 1.

Imagine that after preparing the system in equilibrium at $\lambda = A$ we disconnect it from the thermal reservoir. Then from $t = 0$ to $t = \tau$ the three-particle system of interest evolves under the Hamiltonian $H(\mathbf{x}; \lambda(t))$ (Eq. (7)) as we displace the fourth particle from $\lambda = A$ to B using a protocol $\lambda(t)$. A realization of this process is described by a trajectory $\mathbf{x}_t \equiv \mathbf{x}(t)$ obeying Hamilton's equations. Combining Eq. (9) with the identity $\mathrm{d}H/\mathrm{d}t = \partial H/\partial t$ (see [68], Section 8-2), we get $W = H(\mathbf{x}_\tau; B) - H(\mathbf{x}_0; A)$. We then evaluate the left side of Eq. (15) by averaging over initial conditions, using Eq. (10):

$$\begin{aligned} \langle e^{-W/k_BT} \rangle &= \int \mathrm{d}\mathbf{x}_0 \, p^{\mathrm{eq}}_{A,T}(\mathbf{x}_0) \, e^{-W/k_BT} \\ &= \frac{1}{Z_{A,T}} \int \mathrm{d}\mathbf{x}_\tau \left| \frac{\partial \mathbf{x}_\tau}{\partial \mathbf{x}_0} \right|^{-1} e^{-H(\mathbf{x}_\tau;B)/k_BT} = \frac{Z_{B,T}}{Z_{A,T}}. \end{aligned} \tag{16}$$

On the second line, the variables of integration have been changed from initial conditions to final conditions. By Liouville's theorem, the associated Jacobian factor is unity, $|\partial \mathbf{x}_\tau / \partial \mathbf{x}_0| = 1$, which brings us to the desired result, $Z_{B,T}/Z_{A,T} = e^{-\Delta F/k_B T}$ (Eq. (13)). (Note that the system is generally out of equilibrium at $t = \tau$; see comment (2) at the end of Section 2.1.)

This gist of the calculation can be extended to the more general case in which the system and reservoir remain in contact during the interval $0 < t < \tau$ [6, 50]. The steps are essentially the ones in Eq. (16), only carried out in the full phase space $(\mathbf{x}, \mathbf{y})$, and care must be taken if the interaction energy $H_{\text{int}}(\mathbf{x}, \mathbf{y})$ is strong. [50] For derivations of Eq. (15) in which the presence of the reservoir is modeled implicitly, using non-Hamiltonian equations of motion, see [6–8, 17, 18, 32, 52–63].

Recall that the work performed during a reversible, isothermal process depends only on the initial and final states, $W = \Delta F \equiv F_{B,T} - F_{A,T}$, and not on the sequence of equilibrium states that mark the journey from $[A, T]$ to $[B, T]$. The nonequilibrium work relation extends this statement to irreversible processes:

$$-k_B T \ln \langle e^{-W/k_B T} \rangle = \Delta F. \tag{17}$$

That is, the value of the nonlinear average on the left depends only on equilibrium states $[A, T]$ and $[B, T]$ (since these determine ΔF), and not on the intermediate, out-of-equilibrium states visited by the system. This implies that we can determine an equilibrium free energy difference by observing a system driven away from equilibrium, provided we repeat the process many times: the value of ΔF is to be found not in a single measurement of work, but in its statistical fluctuations. The idea that far-from-equilibrium fluctuations encode useful equilibrium information is further extended by Eqs. (25), (30) and (31) below, but before getting to those results I will briefly draw attention to a few points related to Eq. (15).

First, Eq. (15) is closely related, but not equivalent, to an earlier work relation derived by Bochkov and Kuzovlev [4, 5, 69, 70], which can be written as

$$\langle e^{-W_0/k_B T} \rangle = 1. \tag{18}$$

This result does not involve ΔF, and uses a definition of work that differs from Eq. (9). [32, 42, 71] contain a more detailed discussion of the precise relationship between Eqs. (15) and (18), as well as between Eqs. (25), (30), and their counterparts in [4, 5, 69, 70].

With minimal effort we can use Eq. (15) to obtain two *inequalities* that are closely related to the second law of thermodynamics. Combining Eq. (15) with Jensen's inequality, [72] $\langle \exp x \rangle \geq \exp \langle x \rangle$, we get

$$\langle W \rangle \geq \Delta F, \tag{19}$$

as already anticipated (Eq. (14)).

A stronger and less expected result follows almost as immediately from Eq. (15): [31]

$$P[W < \Delta F - \zeta] \equiv \int_{-\infty}^{\Delta F-\zeta} \mathrm{d}W\, \rho(W) \leq \int_{-\infty}^{\Delta F-\zeta} \mathrm{d}W\, \rho(W)\, e^{(\Delta F-\zeta-W)/k_BT}$$
$$\leq e^{(\Delta F-\zeta)/k_BT} \int_{-\infty}^{+\infty} \mathrm{d}W\, \rho(W)\, e^{-W/k_BT} = e^{-\zeta/k_BT} \qquad (20)$$

Here, P is the probability to observe a value of work that falls below $\Delta F - \zeta$, where ζ is an arbitrary positive value with units of energy. Eq. (20) tells us that the left tail of the distribution $\rho(W)$ becomes exponentially suppressed in the thermodynamically forbidden region $W < \Delta F$, a bit like the evanescent piece of a quantum-mechanical wave function in a classically forbidden region. Thus we have no hope to observe a value of work that falls much more than a few k_BT below ΔF. This is gratifyingly consistent with everyday experience, which teaches us not only that the second law is satisfied on average, in the sense of Eq. (19), but that it is *never* violated on a macroscopic scale.

For sufficiently slow variation of the work parameter, the central limit theorem suggests that $\rho(W)$ is approximately Gaussian. In this case Eq. (15) implies [6]

$$\Delta F = \langle W \rangle - \frac{\sigma_W^2}{2k_BT} \qquad (21)$$

where σ_W^2 is the variance of the work distribution. This is the result that one expects from linear response theory. [73–76]

Because Eq (15) unequivocally implies that $\langle W \rangle \geq \Delta F$, it might at first glance appear that this represents a microscopic, first-principles derivation of the second law, and thus clarifies the microscopic origins of irreversibility. This is not the case, however. In all derivations of Eq. (15) and related work relations (e.g., Eqs. (25), (30), (31)), the arrow of time is effectively inserted by hand. Specifically, a quite special statistical state (the Boltzmann-Gibbs distribution, p^{eq}) is assumed to describe the system at a particular instant in time ($t = 0$), and attention is then focused on the system's evolution at later times only ($t > 0$). If instead the evolution of the system leading up to the equilibrium state at $t = 0$ had been considered, then all the inequalities associated with the second law would have been obtained, but with their signs reversed. This emphasizes the importance of boundary conditions (in time), and touches on the deep connection between irreversibility and causality [77–79].

Gibbs already recognized that if one accepts an initial equilibrium state given by $p^{\mathrm{eq}} \propto e^{-H/k_BT}$, then various statements of the second law follow from properties of Hamiltonian dynamics (see Chapter XIII of [80]). Similar results can be obtained if the initial equilibrium state is represented by any distribution that is a decreasing function of energy [81]. Interestingly, however, for a microcanonical initial distribution, inequalities related to the second law of thermodynamics can be violated, at least for systems with one degree of freedom [82, 83].

Let us now return to the picture of our ensemble as a swarm of trajectories, $\mathbf{x}_1(t)$, $\mathbf{x}_2(t), \ldots$ described by the time-dependent phase space density,

$$f(\mathbf{x}, t) \equiv \Big\langle \delta\Big[\mathbf{x} - \mathbf{x}_k(t)\Big]\Big\rangle, \tag{22}$$

and let us define a *weighted* density

$$g(\mathbf{x}, t) \equiv \Big\langle \delta\Big[\mathbf{x} - \mathbf{x}_k(t)\Big]\, e^{-w_k(t)/k_B T}\Big\rangle \tag{23}$$

where $w_k(t)$ is the work performed up to time t during the kth realization. If we visualize each trajectory $\mathbf{x}_k(t)$ as a particle moving through phase space, and $\mu_k(t) = \exp[-w_k(t)/k_B T]$ as a time-dependent "mass" that the particle carries on its journey, then $f(\mathbf{x}, t)$ and $g(\mathbf{x}, t)$ can be interpreted as a normalized particle density and mass density, respectively. Both are initially described by the canonical distribution, $f = g = p^{\mathrm{eq}}_{A,T}$, but for $t > 0$ the system is no longer in equilibrium:

$$f_t \equiv \Big\langle \delta\Big[\mathbf{x} - \mathbf{x}_k(t)\Big]\Big\rangle \neq p^{\mathrm{eq}}_{\lambda(t),T}(\mathbf{x}, t)\ , \quad t > 0. \tag{24}$$

By the simple trick of reweighting each trajectory by $\mu_k(t)$, this inequality is transformed into an equality, namely [9]

$$g_t \equiv \Big\langle \delta\Big[\mathbf{x} - \mathbf{x}_k(t)\Big]\, e^{-w_k(t)/k_B T}\Big\rangle = \frac{1}{Z_{A,T}} e^{-H(\mathbf{x};\lambda(t))/k_B T}. \tag{25}$$

Note that the right side is proportional to $p^{\mathrm{eq}}_{\lambda(t),T}$, and that we recover Eq. (15) by setting $t = \tau$ and integrating over phase space.

To sketch a derivation of Eq. (25) for our toy model (Fig. 1), note that the ordinary density $f(\mathbf{x}, t)$ satisfies the Liouville equation, $\partial f/\partial t + \{f, H\} = 0$, using Poisson bracket notation [68] and assuming that the system is isolated from the reservoir for $0 < t < \tau$. The left side of the Liouville equation is just the total time derivative of $f(\mathbf{x}(t), t)$ along a Hamiltonian trajectory. For the weighted density $g(\mathbf{x}, t)$, an additional term accounts for the time-dependent weight: [7, 9]

$$\frac{\partial g}{\partial t} + \{g, H\} = -\frac{\dot{w}}{k_B T}\, g, \tag{26}$$

where $\dot{w} = \dot{\lambda}\, \partial H/\partial \lambda$. It is now a matter of substitution to show that for the initial conditions $g_0 = p^{\mathrm{eq}}_{A,T}$, the right side of Eq. (25) solves Eq. (26). For derivations of Eq. (25) (or equivalent results) in which the reservoir is modeled using stochastic and other non-Hamiltonian dynamics, see [7, 9, 18, 26, 32, 60].

Eq. (25) reveals the following: even as it is driven away from equilibrium, the swarm of trajectories retains information about the equilibrium state $p^{\mathrm{eq}}_{\lambda(t),T}$, and the key to unlocking this information is to attach a statistical, time-dependent weight $\exp[-w_k(t)/k_B T]$ to each realization. This reweighting procedure was described and illustrated by Jarzynski [7, 84], and obtained in terms of path averages by Crooks [18], but the elegant formulation given by Eq. (25) is due to Hummer and Szabo [9, 26], who recognized it as a consequence of the Feynman-Kac theorem of stochastic processes. This naturally brings to mind an analogy with the path-integral formulation of quantum mechanics, in which a wave function is constructed as a sum over paths, each contributing a phase $\exp(iS/\hbar)$. The reweighting procedure outlined above has a similar flavor to it, but with real weights $\exp[-w_k(t)/k_BT]$ rather than complex phases. In the quantum-mechanical case, the sum over paths produces a solution to the Schrödinger equation, while here we get the construction of an equilibrium distribution from nonequilibrium trajectories. Hummer and Szabo [9] have used Eq. (25) to derive a method of constructing an equilibrium potential of mean force (a free energy profile along a reaction coordinate that differs from the work parameter λ) from nonequilibrium data. This method has been confirmed experimentally by Berkovich *et al.* [85]

4. Macroscopic hysteresis and microscopic symmetry

The second law of thermodynamics is manifested not only by inequalities such as $W \geq \Delta F$, but also by the time-asymmetry inherent to irreversible processes. *Hysteresis loops* neatly depict this asymmetry. As an example, imagine that we rapidly stretch an ordinary rubber band, then after a sufficient pause we contract it, returning to the initial state. For this process we get a classic hysteresis loop by plotting the tension $\mathcal{T}$ versus the length L of the rubber band (Fig. 2). Hysteresis conveys the idea that the state of the rubber band follows one path during the stretching stage, but returns along a different path during contraction. Quantitatively, the second law implies that the enclosed area is non-negative, $\oint \mathcal{T}\,\mathrm{d}L \geq 0$.

Similar considerations apply to the analogous stretching and contraction of single molecules [86], only now statistical fluctuations become important: the random jigglings of the molecule differ from one repetition of the process to the next. In the previous section we saw that when fluctuations are taken into account, the relationship between work and free energy can be expressed as an equality rather than the usual inequality. The central message of the present section has a similar ring: with an appropriate accounting of fluctuations, the two branches of an irreversible thermodynamic cycle (e.g., the stretching and contraction of the single molecule) are described by unexpected symmetry relations (Eqs. (30), (31)) rather than exclusively by inherent asymmetry (Eqs. (28), (35)).

To develop these results, it is useful to imagine two distinct processes, designated the *forward* and the *reverse* process. [8] The forward process is the one defined in Section 2, in which the work parameter is varied from A to B using a protocol $\lambda_F(t)$ (the subscript F has been attached as a label). During the reverse process, λ is varied from B to A using the time-reversed protocol,

$$\lambda_R(t) = \lambda_F(\tau - t). \tag{27}$$

At the start of each process, the system is prepared in the appropriate equilibrium state, corresponding to $\lambda = A$ or B, at temperature T. If we perform the two processes in sequence, the forward followed by the reverse, allowing the system to equilibrate with the reservoir at the end of each process, then we have a thermodynamic cycle that exhibits hysteresis. The Clausius inequality applies separately

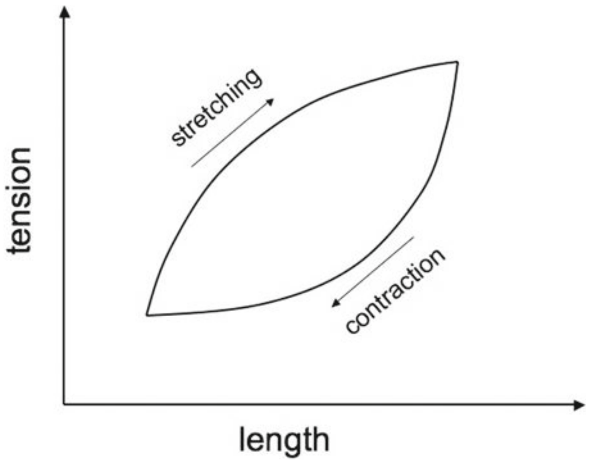


FIGURE 2. Hysteresis loop for the irreversible stretching and contraction of rubber band. During the stretching stage, the temperature and tension of the rubber band are higher than would have been the case if the process were performed reversibly, while during the contraction stage they are lower. As a result, $W > 0$ over the entire cycle. The hysteresis loop illustrates the idea that the system evolves through one sequence of states during the forward process, but follows a different path back during the reverse process. The statistical expression of this statement is given by Eq. (35).

to each stage:

$$-\langle W \rangle_R \leq \Delta F \leq \langle W \rangle_F \tag{28}$$

where ΔF is defined as before (Eq. (5)) and the notation now specifies separate averages over the two processes. Of course, Eq. (28) implies that the average work over the entire cycle is non-negative,

$$\langle W \rangle_F + \langle W \rangle_R \geq 0. \tag{29}$$

This illustrates the Kelvin-Planck statement of the second law: no process is possible whose sole result is the absorption of heat from a reservoir and the conversion of all of this heat into work. [87]

Statistically, the forward and reverse processes are described by work distributions $\rho_F(W)$ and $\rho_R(W)$. While Eq. (28) applies to the *means* of these distributions, Crooks [17] has shown that their *fluctuations* satisfy

$$\frac{\rho_F(+W)}{\rho_R(-W)} = e^{(W-\Delta F)/k_B T} \tag{30}$$

As with Eq. (15) (which is an immediate consequence of Eq. (30)), this result remains valid even when the system is driven far from equilibrium, and has been verified in a number of experiments. [65–67, 86, 88]

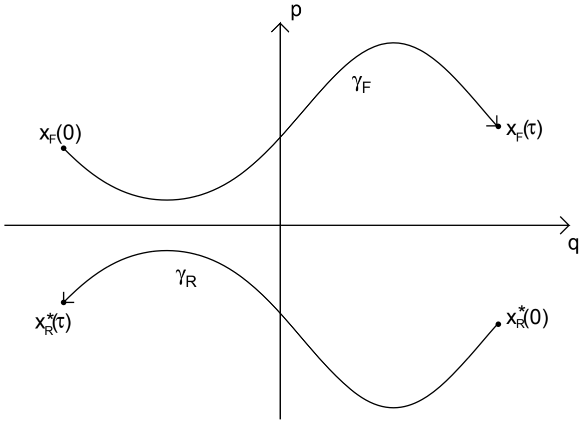


FIGURE 3. A conjugate pair of trajectories, γ_F and γ_R.

While Crooks's fluctuation theorem, Eq. (30), is a statement about distributions of work values, at its heart is a stronger result about distributions of *trajectories*: [8]

$$\frac{\mathcal{P}_F[\gamma_F]}{\mathcal{P}_R[\gamma_R]} = e^{(W_F - \Delta F)/k_B T} \tag{31}$$

Here, the notation $\gamma_F \equiv \{\mathbf{x}_F(t)\,;\, 0 \leq t \leq \tau\}$ denotes a trajectory that might be observed during a realization of the forward process, and γ_R is its *conjugate twin*,

$$\mathbf{x}_R(t) = \mathbf{x}_F^*(\tau - t) \tag{32}$$

where $\mathbf{x}^*$ is the microscopic state obtained by reversing all the momenta of $\mathbf{x}$, as illustrated schematically in Fig. 3. Simply put, the trajectory γ_R represents what we would see if we were to film the trajectory γ_F, and then run the movie backward. Eq. (31) then states that the probability of observing a particular trajectory when performing the forward process, $\mathcal{P}_F[\gamma_F]$, relative to that of observing its conjugate twin during the reverse process, $\mathcal{P}_R[\gamma_R]$, is given by the expression on the right side of the equation, where $W_F \equiv W[\gamma_F]$ is the work performed in the forward case.

To derive Eq. (31) for our toy model, assuming as before that the reservoir is removed for $0 < t < \tau$, note that the ratio of probabilities to observe the Hamiltonian trajectories γ_F and γ_R is simply the ratio of probabilities to sample their respective initial conditions from equilibrium. [79] Thus

$$\begin{aligned}\frac{\mathcal{P}_F[\gamma_F]}{\mathcal{P}_R[\gamma_R]} &= \frac{Z_{B,T}}{Z_{A,T}}\, e^{[H(\mathbf{x}_R(0);B) - H(\mathbf{x}_F(0);A)]/k_B T} \\ &= \frac{Z_{B,T}}{Z_{A,T}}\, e^{[H(\mathbf{x}_F(\tau);B) - H(\mathbf{x}_F(0);A)]/k_B T} = e^{(W_F - \Delta F)/k_B T},\end{aligned} \tag{33}$$

using Eqs. (32) and (7) to replace $H(\mathbf{x}_R(0); B)$ by $H(\mathbf{x}_F(\tau); B)$. We get to the final result by observing that the quantity inside square brackets on the second line is the net change in H during the forward process, which (for a thermally isolated system, see Section 3) is the work performed on the system. As with the results of Section 3, numerous derivations of Eqs. (30) and (31) exist in the literature, corresponding to various models of the system and reservoir. [8, 17, 18, 32, 52, 54, 58, 59, 61–63, 89].

To gain some appreciation for this result, recall that a system in equilibrium satisfies *microscopic reversibility* [90] (closely related to *detailed balance* [17]): any sequence of events is as likely to occur as the time-reversed sequence. Using notation similar to Eq. (31) this condition can be written,

$$\mathcal{P}^{\mathrm{eq}}[\gamma] = \mathcal{P}^{\mathrm{eq}}[\gamma^*], \tag{34}$$

where γ and γ^* are a conjugate pair of trajectories (of some finite duration) for a system in equilibrium. By contrast, as depicted by the two branches of a hysteresis loop, an essential feature of thermodynamic irreversibility is that the system does not simply retrace its steps when forced to return to its initial state. This idea is expressed statistically by the inequality

$$\mathcal{P}_F[\gamma_F] \neq \mathcal{P}_R[\gamma_R], \tag{35}$$

that is the trajectories we are likely to observe during one process are not the conjugate twins of those we are likely to observe during the other process. Eq. (31), which replaces this inequality with a stronger equality, can be viewed as an extension of the principle of microscopic reversibility, to systems that are driven away from equilibrium by the variation of external parameters.

5. Relative entropy and dissipated work

Information theory and thermodynamics enjoy a special relationship, evidenced most conspicuously by the formula,

$$I[p^{\mathrm{eq}}] = S/k_B, \tag{36}$$

where $I[p] \equiv -\int p \ln p$ is the *information entropy* associated with a statistical distribution p. When p describes thermal equilibrium (Eq. (10)), its information entropy I coincides with the thermodynamic entropy, S/k_B (Eq. (36)). This familiar but remarkable result relates a measure of our ignorance about a system's microstate (I), to a physical quantity defined via calorimetry (S).

In recent years, another set of results have emerged that, similarly, draw a connection between information theory and thermodynamics, but these results apply to irreversible processes rather than equilibrium states. Here the relevant information-theoretic measure is the *relative entropy* [91, 92] between two distributions (Eq. (37)), and the physical quantity is *dissipated work*, $W - \Delta F$. This section describes these results in some detail, but the central idea can be stated succinctly as follows. The irreversibility of a process can be expressed as an inequality between a pair of probability distributions, either in trajectory space or in phase space (Eqs. (35), (40), (24)). Using the relative entropy to quantify the difference between the two distributions, we find in each case that this measure relates directly to dissipated work (Eqs. (38), (41), (43)).

For two normalized probability distributions p and q on the same space of variables, the relative entropy, or *Kullback-Leibler divergence*, [91]

$$D[p|q] \equiv \int p \ln\left(\frac{p}{q}\right) \geq 0 \tag{37}$$

quantifies the extent to which one distribution differs from the other. $D = 0$ if and only if the distributions are identical, and $D \gg 1$ if there is little overlap between the two distributions. Note that in general $D[p|q] \neq D[q|p]$.

Because relative entropy provides a measure of distinguishability, it is a handy tool for quantifying irreversibility. Recall that hysteresis can be expressed statistically by the inequality $\mathcal{P}_F[\gamma_F] \neq \mathcal{P}_R[\gamma_R]$ (Eq. (35)), where the trajectory-space distributions $\mathcal{P}_F$ and $\mathcal{P}_R$ represent the system's response during the forward and reverse processes. We can then use the relative entropy $D[\mathcal{P}_F|\mathcal{P}_R]$ to assign a value to the extent to which the system's evolution during one process differs from that during the other. From Eq. (31) it follows that [79]

$$D[\mathcal{P}_F|\mathcal{P}_R] = \frac{W_F^{\mathrm{diss}}}{k_B T} \tag{38}$$

where

$$W_F^{\mathrm{diss}} \equiv \langle W \rangle_F - \Delta F \tag{39}$$

is the average amount of work that is dissipated during the forward process. (Similarly, $D[\mathcal{P}_R|\mathcal{P}_F] = W_R^{\mathrm{diss}}/k_B T$.)

While distributions in trajectory space are abstract and difficult to visualize, a result similar to Eq. (38) can be placed within the more familiar setting of *phase space*. Let $f_F(\mathbf{x}, t)$ denote the time-dependent phase space density describing the evolution of the system during the forward process (Eq. (22)), and define $f_R(\mathbf{x}, t)$ analogously for the reverse process. Then the densities $f_F(\mathbf{x}, t_1)$ and $f_R(\mathbf{x}, \tau - t_1)$ are snapshots of the statistical state of the system during the two processes, both taken at the moment the work parameter achieves the value $\lambda_1 \equiv \lambda_F(t_1) = \lambda_R(\tau - t_1)$. The inequality

$$f_F(\mathbf{x}, t_1) \neq f_R(\mathbf{x}^*, \tau - t_1) \tag{40}$$

then expresses the idea that the statistical state of the system is different when the work parameter passes through the value λ_1 during the forward process, than when it returns through the same value during the reverse process. (The reversal of momenta in $\mathbf{x}^*$ is related to the conjugate pairing of trajectories, Eq. (32).) Evaluating the relative entropy between these distributions, Kawai, Parrondo and Van den Broeck [93] found that

$$D[f_F|f_R^*] \leq \frac{W_F^{\mathrm{diss}}}{k_B T}, \tag{41}$$

where the arguments of D are the distributions appearing in Eq. (40), for any choice of λ_1. This becomes an *equality* if the system is isolated from the thermal environment as the work parameter is varied during each process. As with Eq. (38), we see that an information-theoretic measure of the difference between two distributions, f_F and f_R^*, is related to a physical measure of dissipation, $W_F^{\rm diss}/k_BT$.

Eqs. (38) and (41) are closely related. The phase-space distribution $f_F = f_F(\mathbf{x}, t_1)$ is the projection of the trajectory-space distribution $\mathcal{P}_F[\gamma_F]$ onto a single "time slice", $t = t_1$, and similarly for f_R^*. Since the relative entropy between two distributions decreases when they are projected onto a smaller set of variables [91, 93] – in this case, from trajectory space to phase space – we have

$$D[f_F|f_R^*] \leq D[\mathcal{P}_F|\mathcal{P}_R] = \frac{W_F^{\rm diss}}{k_BT}. \tag{42}$$

In the above discussion, relative entropy has been used to quantify the difference between the forward and reverse processes (hysteresis). It can equally well be used to measure how far a system is removed from equilibrium at a given instant in time, leading again to a link between relative entropy and dissipated work (Eq. (43) below).

For the process introduced in Section 2, let $f_t \equiv f(\mathbf{x}, t)$ denote the statistical state of the system at time t, and let $p_t^{\rm eq} \equiv p_{\lambda(t),T}^{\rm eq}(\mathbf{x})$ be the equilibrium state corresponding to the current value of the work parameter. It is useful to imagine that f_t continually "chases" $p_t^{\rm eq}$: as the work parameter is varied with time, the state of the system (f_t) tries to keep pace with the changing equilibrium distribution ($p_t^{\rm eq}$), but is unable to do so (except in the reversible, isothermal limit). Vaikuntanathan and Jarzynski [94] have shown that

$$D[f_t|p_t^{\rm eq}] \leq \frac{\langle w(t)\rangle - \Delta F(t)}{k_BT} \tag{43}$$

where $\Delta F(t) \equiv F_{\lambda(t),T} - F_{A,T}$. In other words, the average work dissipated up to time t, in units of k_BT, provides an upper bound on the degree to which the system lags behind equilibrium at that instant. This result can be obtained from either Eq. (25) or Eq. (41). [94] If we take $t = \tau^*$, allowing the system to relax to a final state of equilibrium (see Section 2.1), then the left side of Eq. (43) vanishes and once again we recover the Clausius inequality.

As mentioned, relative entropy is an asymmetric measure: in general $D[p|q] \neq D[q|p]$. Feng and Crooks [95] have discussed the use of two symmetric measures of distinguishability to quantify thermodynamic irreversibility. The first is the Jeffreys divergence, $D[p|q]+D[q|p]$. When applied to forward and reverse distributions in trajectory space, this gives the average work over the entire cycle (see Eq. (38)):

$$\text{Jeffreys}(\mathcal{P}_F;\mathcal{P}_R) = \frac{W_F^{\rm diss} + W_R^{\rm diss}}{k_BT} = \frac{\langle W\rangle_F + \langle W\rangle_R}{k_BT}. \tag{44}$$

The second measure is the Jensen-Shannon divergence,

$$\text{JS}(p;q) = \frac{1}{2}\left(D[p|m] + D[q|m]\right), \tag{45}$$

where $m = (p + q)/2$ is the mean of the two distributions. When evaluated with $p = \mathcal{P}_F$ and $q = \mathcal{P}_R$, this leads to a more complicated, nonlinear average of $W_F^{\rm diss}$ and $W_R^{\rm diss}$ (see Eq. 7 of [95]). Feng and Crooks nevertheless argue that the Jensen-Shannon divergence is the preferred measure of time-asymmetry, as it has a particularly nice information-theoretic interpretation. I will return to this point at the end of the following section.

6. Guessing the direction of time's arrow

Sir Arthur Eddington introduced the term "arrow of time" to describe the evident directionality associated with the flow of events. [96] While time's arrow is familiar from daily experience – everyone recognizes that a movie run backward looks peculiar! – Eddington (among others) argued that it is rooted in the second law of thermodynamics. For a macroscopic system undergoing an irreversible process of the sort described in Section 2.1, the relationship between the second law and the arrow of time is almost tautological: $W > \Delta F$ when events proceed in the correct order, and $W < \Delta F$ when the movie is run backward, so to speak. For a microscopic system, fluctuations blur this picture, since we can occasionally observe violations of the Clausius inequality (Eq. (5)). Thus the sign of $W - \Delta F$, while correlated with the direction of time's arrow, does not fully determine it. These general observations can be made precise, that is the ability to determine the direction of time's arrow can be quantified.

To discuss this point, it is convenient to consider a hypothetical guessing game [79]. Imagine that I show you a movie in which you observe a system undergoing a thermodynamic process as λ is varied from A to B. Your task is to guess whether this movie depicts the events in the order in which they actually occurred, or whether I have filmed the reverse process (varying λ from B to A) and am now (deviously) showing you the movie of that process, run backward. In the spirit of a *Gedankenexperiment*, assume that the movie gives you full microscopic information about the system – you can track the motion of every atom – and that you know the Hamiltonian function $H(\mathbf{x};\lambda)$ and the value $\Delta F = F_{B,T} - F_{A,T}$. Assume moreover that in choosing which process to perform, I flipped a fair coin: heads $= F$, tails $= R$.

We can formalize this task as an exercise in statistical inference. [95] Let $L(F|\gamma)$ denote the likelihood that the movie is being shown in the correct direction (i.e., the coin landed on heads and the forward process was performed), given the microscopic trajectory γ that you observe in the movie. Similarly, let $L(R|\gamma)$ denote the likelihood that the reverse process was in fact performed and the movie is now being run backward. Since these are the only possibilities, the likelihoods associated with the two hypotheses (F, R) sum to unity:

$$L(F|\gamma) + L(R|\gamma) = 1 \tag{46}$$

Now let W denote the work performed on the system, for the trajectory depicted in the movie. If $W > \Delta F$, then the first hypothesis (F) is in agreement with the Clausius inequality, while the second hypothesis (R) is not; if $W < \Delta F$, it is the other way around. Therefore for a macroscopic system the task is easy, as the sign of $W - \Delta F$ determines the direction of time's arrow. Formally,

$$L(F|\gamma) = \theta(W - \Delta F) \tag{47}$$

where $\theta(\cdot)$ is the unit step function.

For a microscopic system we must allow for the possibility that Eq. (5) might be violated now and again. Bayes' Theorem then provides the right tool for analyzing the likelihood:

$$L(F|\gamma) = \frac{P(\gamma|F)\,P(F)}{P(\gamma)}. \tag{48}$$

Here $P(F)$ is the prior probability that I carried out the forward process, which is simply $1/2$ since I flipped a fair coin to make my choice, and $P(\gamma|F)$ is the probability to generate the trajectory γ when performing the forward process; in the notation of Section 4, this is $\mathcal{P}_F[\gamma]$. Finally, $P(\gamma)$ is (effectively) a normalization constant. Writing the analogous formula for $L(R|\gamma)$, then combining these with the normalization condition Eq. (46) and invoking Eq. (31), we get [31, 97, 98]

$$L(F|\gamma) = \frac{1}{1 + e^{-(W-\Delta F)/k_BT}}. \tag{49}$$

This result quantifies your ability to determine the arrow of time from the trajectory depicted in the movie. The expression on the right is a smoothed step function. If the W surpasses ΔF by many units of k_BT, then $L(F|\gamma) \approx 1$ and you can say with high confidence that the movie is being shown in the correct direction, while in the opposite case you can be equally confident that the movie is being run backward. The transition from one regime to the other – where time's arrow gets blurred, in essence – occurs over an interval of work values whose width is a few k_BT. (Indeed, when viewed from a distance – that is, when plotted as a function of W on an axis with macroscopic units of energy – Eq. (49) is indistinguishable from Eq. (47).) What is remarkable is that this transition does not depend on the details of either the system or the protocol $\lambda(t)$. Eq. (49) was derived by Shirts *et al.* [97] and later by Maragakis *et al.* [98] in the context of free energy estimation, where the interpretation is somewhat different from the one I have discussed here.

Finally, returning to the point mentioned at the end of the previous section, the Jensen-Shannon divergence has the following interpretation in the context of our hypothetical guessing game: $\mathrm{JS}(\mathcal{P}_F;\mathcal{P}_R)$ is the average gain in information (regarding which process was performed) obtained from observing the movie [95]. When the processes are highly irreversible, this approaches its maximum value, $\mathrm{JS} \approx \ln 2$, corresponding to one bit of information. This makes sense: by watching the movie you are able to infer with confidence whether the coin I flipped turned up heads (F) or tails (R). Feng and Crooks [95] have argued that this interpretation has surprisingly universal implications for biomolecular and other nanoscale machines. Namely, about $4 - 8\,k_BT$ of free energy must be dissipated per operating cycle to guarantee that the machine runs reliably in a designated direction (as opposed to taking backward and forward steps with equal probability, as would necessarily occur under equilibrium conditions).

Finally, time's arrow has unexpected relevance for the convergence of the exponential average in Eq. (15). Namely, the realizations that dominate that average are precisely those "during which the system appears as though it is evolving backward in time". [79] A detailed analysis of this assertion involves both hysteresis and relative entropy, thus nicely tying together the four strands of discussion represented by Sections 3–6. [79]

7. Entropy production and related quantities

This review has focused on far-from-equilibrium predictions for work and free energy (Eqs. (15), (25), (30), (31)) and how these inform our understanding of irreversibility and the second law of thermodynamics. Because the second law is often taken to be synonymous with the increase of entropy, we might well wonder how these predictions relate to statements about entropy.

As a point of departure, for macroscopic systems we can use the first law ($\Delta U = W + Q$) and the definition of free energy ($F = U - ST$) to write,

$$\frac{W - \Delta F}{T} = \Delta S - \frac{Q}{T} = \Delta S_{\mathrm{tot}}, \tag{50}$$

where ΔS_{tot} is the combined entropy change of the system and reservoir. If we extend this to microscopic systems, accepting it as the *definition* of ΔS_{tot} for a single realization of a thermodynamic process, then the results discussed in Sections 3–6 can formally be rewritten as statements about the fluctuations of entropy production.

When multiple thermal reservoirs are involved, one can generalize Eq. (6) in an obvious way by including terms for all the reservoirs, e.g., $\mathcal{H} = H + \sum_k (H_{\mathrm{env}}^k + H_{\mathrm{int}}^k)$. Working entirely within a Hamiltonian framework, the results of Section 3, notably Eqs. (15), (19) and (20), can then be written in terms of entropy production, and generalized further by dropping the assumption that the system of interest begins in equilibrium. [99] Esposito, Lindenberg and Van den Broeck have recently shown that in this situation the value of ΔS_{tot} is equal to the statistical correlation that develops between the system and the reservoirs, as measured in terms of relative entropy [100].

The Hamiltonian framework is often inconvenient for studying nonequilibrium steady states. Among the many tools that have been introduced for the theoretical analysis and numerical simulation of such states, *Gaussian thermostats* – the term refers to a method of modeling nonequilibrium systems based on Gauss's principle of least constraint [101] – have played a prominent role in recent developments in nonequilibrium thermodynamics. The term *fluctuation theorem* was originally applied to a property of entropy production, observed in numerical investigations of a sheared fluid simulated using a Gaussian thermostat [10–13]. Since fluctuation theorems for entropy production have been reviewed elsewhere [21, 22, 24, 29, 30, 32, 33, 35, 36], I will limit myself to a brief summary of how these results connect to those of Sections 3–6.

The *transient* fluctuation theorem of Evans and Searles [11] applies to a system that evolves from an initial state of equilibrium to a nonequilibrium steady state. Letting $p_\tau(\Delta s)$ denote the probability distribution of the entropy produced up to a time $\tau > 0$, it states that

$$\frac{p_\tau(+\Delta s)}{p_\tau(-\Delta s)} = e^{\Delta s/k_B}. \tag{51}$$

This is clearly similar to Eq. (30), except that it pertains to a single thermodynamic process, rather than a pair of processes (F and R). Eq. (51) implies an *integrated* fluctuation theorem,

$$\langle e^{-\Delta s/k_B} \rangle = 1, \tag{52}$$

that is entirely analogous to Eq. (15), and from this we in turn get analogues of Eqs. (19) and (20):

$$\langle \Delta s \rangle \geq 0 \ , \qquad P[\Delta s < -\xi] \leq e^{-\xi/k_B} \tag{53}$$

Now consider a system that is in a nonequilibrium steady state from the distant past to the distant future, such as a fluid under constant shear [10], and let $\sigma \equiv \Delta s/\tau$ denote the entropy production rate, time-averaged over a single, randomly sampled interval of duration τ. The *steady-state* fluctuation theorem of Gallavotti and Cohen [12, 13] asserts that the probability distribution $p_\tau(\sigma)$ satisfies

$$\lim_{\tau\to\infty} \frac{1}{\tau} \ln \frac{p_\tau(+\sigma)}{p_\tau(-\sigma)} = \frac{\sigma}{k_B}. \tag{54}$$

The integrated form of this result is [21]

$$\lim_{\tau\to\infty} \frac{1}{\tau} \ln \left\langle e^{-\tau\sigma/k_B} \right\rangle_\tau = 0, \tag{55}$$

where the brackets denote an average over intervals of duration τ, in the steady state. Formal manipulations then give us

$$\langle \sigma \rangle_\tau \geq 0 \ , \qquad \lim_{\tau\to\infty} \frac{1}{\tau} \ln P_\tau[\sigma < -\epsilon] \leq -\epsilon, \tag{56}$$

where $P_\tau[\sigma < -\epsilon]$ is the probability to observe a time-averaged entropy production rate less than $-\epsilon$, during an interval of duration τ. The resemblance between Eqs. (54)–(56), and Eqs. (30), (15), (19), (20), respectively, should be obvious, although viewed as mathematical statements they are different.

The microscopic definition of entropy production in Eqs. (51)–(56) depends on the equations of motion used to model the evolution of the system. In the early papers on fluctuation theorems, entropy production was identified with phase space contraction along a deterministic but non-Hamiltonian trajectory [10–13]. These results were then extended to encompass stochastic dynamics, first by Kurchan [14] for diffusion, and then by Lebowitz and Spohn [15] for Markov processes in general. Maes subsequently developed a unified framework based on probability distributions of "space-time histories" [16], that is trajectories. In all these cases, the validity of the fluctuation theorem ultimately traces back to the idea that trajectories come in pairs related by time-reversal, and that the production of entropy is intimately linked with the probability of observing one trajectory relative to the other, in a manner analogous with Eq. (31).

As an aside, it is intriguing to note that multiple fluctuation theorems can be valid simultaneously, in a given physical context. This idea was mentioned in passing by Hatano and Sasa [19] in the context of transitions between nonequilibrium states, and has been explored in greater detail by a number of authors since then [32, 54, 103, 104].

Finally, for nonequilibrium steady states there exist connections between relative entropy and entropy production, analogous to those discussed in Section 5. If relative entropy is used to quantify the difference between distributions of steady-state trajectories and their time-reversed counterparts, then the value of this difference can be equated with the thermodynamic production of entropy. This issue has been explored by Maes [16], Maes and Netočný [105], and Gaspard [106].

8. Conclusions and outlook

The central message of this review is that far-from-equilibrium fluctuations are more interesting than one might have guessed: they tell us something new about how the second law of thermodynamics operates at the nanoscale. In particular, they allow us to rewrite thermodynamic inequalities as equalities, and they reveal that nonequilibrium fluctuations encode equilibrium information.

The last observation has led to practical applications in two broad settings. The first is the development of numerical methods for estimating free energy differences, an active enterprise in computational chemistry and physics [23]. While traditional strategies involve equilibrium sampling, Eqs. (15), (25) and (30) suggest the use of nonequilibrium simulations to construct estimates of ΔF. This is an ongoing area of research [107, 108], but nonequilibrium methods have gradually gained acceptance into the free energy estimation toolkit, and are being applied to a variety of molecular systems; see [109] for a recent example.

Nonequilibrium work relations have also been applied to the analysis of single-molecule experiments, as originally proposed by Hummer and Szabo [9] and pioneered in the laboratory by Liphardt *et al.* [64] Individual molecules are driven away from equilibrium using (for instance) optical tweezers or atomic-force microscopy, and from measurements of the work performed on these molecules, one can reconstruct equilibrium free energies of interest [27, 110]. For recent applications of this approach, see [111–114].

It remains to be seen whether the understanding of far-from-equilibrium fluctuations that has been gained in recent years will lead to the formulation of a unified "thermodynamics of small systems", that is, a theoretical framework based on a few propositions, comparable to classical thermodynamics. Some progress, however, has been made in this direction.

For stochastic dynamics, Seifert and colleagues [32, 54, 115–117] – building on earlier work by Sekimoto [37, 118] – have developed a formalism in which microscopic analogues of all relevant macroscopic quantities are precisely defined. Many of the results discussed in this review follow naturally within this framework, and this has helped to clarify the relations among these results. [32] Evans and Searles [22] have championed the view that fluctuation theorems are most naturally understood in terms of a dissipation function, Ω, whose properties are (by construction) independent of the dynamics used to model the system of interest. More recently, Ge and Qian [119] have proposed a unifying framework for stochastic processes, in which both the information entropy $-\int p \ln p$ and the relative entropy $\int p \ln(p/q)$ play key roles.

[32] and [119] make a connection to earlier efforts by Oono and Paniconi [120] to develop a "steady-state thermodynamics" organized around non equilibrium steady states. While the original goal was a phenomenological theory, the derivation by Hatano and Sasa of fluctuation theorems for transitions between steady states [19, 121] has encouraged a microscopic approach to this problem [122, 123]. In the absence of a universal statistical description of steady states analogous to the Boltzmann-Gibbs formula (Eq. (10)) this has proven to be highly challenging.

This review has focused exclusively on classical fluctuation theorems and work relations, but the quantum case is also of considerable interest. While quantum versions of these results have been studied for some time [124–127], the past two to three years have seen a surge of interest in this topic [128–142]. Quantum mechanics of course involves profound issues of interpretation. It can be hoped that in the process of trying to specify the quantum-mechanical definition of work [132], or dealing with open quantum systems [131, 137–142], or analyzing exactly solvable models [130, 133, 135, 136], or proposing and ultimately performing experiments to test far-from-equilibrium predictions [134], important insights will be gained. Applications of nonequilibrium work relations to the detection of quantum entanglement [143] and to combinatorial optimization using quantum annealing [144] have very recently been proposed.

Finally, there has been a rekindled interest in recent years in the thermodynamics of information-processing systems and closely related topics such as the apparent paradox of Maxwell's demon [145]. Making use of the relations described in this review, a number of authors have investigated how nonequilibrium fluctuations and the second law are affected in situations involving information processing, such as occur in the context of memory erasure and feedback control. [146–150]

Acknowledgment

I gratefully acknowledge financial support from the National Science Foundation (USA), under DMR-0906601. This article originally appeared in Annu. Rev. Cond. Mat. Phys. **2**, 329–351 (2011).

References

[1] Kolomeisky, A.B., Fisher, M.E.: Annu. Rev. Phys. Chem. **58**, 675–95 (2006).

[2] Smith, D.E., Tans, SJ., Smith, S.B., Grimes, S., Anderson, DL., Bustamante, C.: Nature **413**, 748–52 (2001).

[3] Bustamante, C., Liphardt, J., Ritort, F.: Phys. Today **58** (July), 43–48 (2005).

[4] Bochkov, G.N., Kuzovlev, Yu.E.: Zh. Eksp. Teor. Fiz. **72**, 238–47 (1977).

[5] Bochkov, G.N., Kuzovlev, Yu.E.: Zh. Eksp. Teor. Fiz. **76**, 1071 (1979).

[6] Jarzynski, C.: Phys. Rev. Lett. **78**, 2690–2693 (1997).

[7] Jarzynski, C.: Phys. Rev. E **56**, 5018–5035 (1997).

[8] Crooks, G.E.: J. Stat. Phys. **90**, 1481–87 (1998).

[9] Hummer, G., Szabo, A.: Proc. Natl. Acad. Sci. (USA) **98**, 3658–61 (2001).

[10] Evans, D.J., Cohen, E.G.D., Morriss, G.P.: Phys. Rev. Lett. **71**, 2401–04 (1993).

[11] Evans, D.J., Searles, D.J.: Phys. Rev. E **50**, 1645–48 (1994).

[12] Gallavotti, G., Cohen, E.G.D.: Phys. Rev. Lett. **74**, 2694–97 (1995).

[13] Gallavotti, G., Cohen, E.G.D.: J. Stat. Phys. **80**, 931–970 (1995).

[14] Kurchan, J.: J. Phys. A: Math. Gen. **31**, 3719–29 (1998).

[15] Lebowitz, J.L., Spohn, H.: J. Stat. Phys. **95**, 333–65 (1999).

[16] Maes, C.: J. Stat. Phys. **95**, 367–92 (1999).

[17] Crooks, G.E.: Phys. Rev. E **60**, 2721–26 (1999).

[18] Crooks, G.E.: Phys. Rev. E **61**, 2361–66 (2000).

[19] Hatano, T., Sasa, S.-I.: Phys. Rev. Lett. **86**, 3463–66 (2001).

[20] Adib, A.B.: Phys. Rev. E **71**, 056128/1–5 (2005).

[21] Gallavotti, G.: Statistical Mechanics: A Short Treatise. Springer-Verlag, Berlin (1999).

[22] Evans, D.J., Searles, D.J.: Adv. Phys. **51**, 1529–85 (2002).

[23] Frenkel, D., Smith, B.: Understanding Molecular Simulation: From Algorithms to Applications. Academic Press, San Diego (2002).

[24] Maes, C.: Séminaire Poincaré **2**, 29–62 (2003).

[25] Ritort, F.: Séminaire Poincaré **2**, 193–226 (2003).

[26] Hummer, G., Szabo, A.: Acc. Chem. Res. **38**, 504–13 (2005).

[27] Ritort, F.: J. Phys.: Condens. Matter **18**, R531–83 (2006).

[28] Kurchan, J.: J. Stat. Mech.: Theory and Experiment P07005/1–14 (2007).

[29] Harris, R.J., Schütz, G.M.: J. Stat. Mech.: Theory and Experiment P07020/1–45 (2007).

[30] Gallavotti, G.: Eur. Phys. J. B **64**, 315–20 (2008).

[31] Jarzynski, C.: Eur. Phys. J. B **64**, 331–40 (2008).

[32] Seifert, U.: Eur. Phys. J. B **64**, 423–431 (2008).

[33] Sevick, E.M., Prabhakar, R., Williams, S.R., Searles, D.J.: Annu. Rev. Phys. Chem. **59**, 603–33 (2008).

[34] Evans, D.J., Morriss, G.: Statistical Mechanics of Nonequilibrium Liquids. 2nd ed. Cambridge University Press, Cambridge (2008).

[35] Esposito, M., Harbola, U., Mukamel, S.: Rev. Mod. Phys. **81**, 1665–1702 (2009).

[36] Kurchan, J.: Lecture Notes of the Les Houches Summer School: Volume 90, August 2008 (arXiv:0901.1271v2) (2010).

[37] Sekimoto, K.: Stochastic Energetics. Lecture Notes in Physics 799. Springer, Berlin (2010).

[38] Boksenbojm, E., Wynants, B., Jarzynski, C.: Physica A (2010), (in press)

[39] Finn, C.B.P.: Thermal Physics. 2nd ed. Chapman and Hall, London (1993).

[40] Uhlenbeck, G.E., Ford, G.W.: Lectures in Statistical Mechanics. American Mathematical Society, Providence (1963). Chapter 1, Section 7.

[41] Jarzynski, C.: Acta Phys. Pol. B **29**, 1609–22 (1998).

[42] Jarzynski, C.: C. R. Physique **8**, 495–506 (2007).

[43] Vilar, J.M.G., Rubi, J.M.: Phys. Rev. Lett. **100**, 020601/1–4 (2008).

[44] Peliti, L.: J. Stat. Mech.: Theory and Experiment P05002/1–9 (2008).

[45] Peliti, L.: Phys. Rev. Lett. **101**, 098901 (2008).

[46] Vilar, J.M.G., Rubi, J.M.: Phys. Rev. Lett. **101**, 098902 (2008).

[47] Horowitz, J., Jarzynski, C.: Phys. Rev. Lett. **101**, 098903 (2008).

[48] Vilar, J.M.G., Rubi, J.M.: Phys. Rev. Lett. **101**, 098904 (2008).

[49] Zimanyi, E.N., Silbey, R.J.: J. Chem. Phys. **130**, 171102/1–3 (2009).

[50] Jarzynski, C.: J. Stat. Mech.: Theory and Experiment P09005/1–13 (2004).

[51] Sun, S.X.: J. Chem. Phys. **118**, 5769–75 (2003).

[52] Evans, D.J.: Mol. Phys. **101**, 1551–45 (2003).

[53] Imparato, A., Peliti, L.: Europhys. Lett. **70**, 740–46 (2005).

[54] Seifert, U.: Phys. Rev. Lett. **95**, 040602/1–4 (2005).

[55] Oberhofer, H., Dellago, C., Geissler, P.L.: J. Chem. Phys. **109**, 6902–15 (2005).

[56] Cuendet, M.A.: Phys. Rev. Lett. **96**, 120602/1–4 (2006).

[57] Cuendet, M.A.: J. Chem. Phys. **125**, 144109/1–12 (2006).

[58] Schöll-Paschinger, E., Dellago, C.: J. Chem. Phys. **125**, 054105/1–5 (2006).

[59] Chelli, R., Marsili, S., Barducci, A., Procacci, P.: Phys. Rev. E **75**, 050101(R)/1–4 (2007).

[60] Ge, H., Qian, M.: J. Math. Phys. **48**, 053302/1–11 (2007).

[61] Ge, H., Jiang, D.-Q.: J. Stat. Phys. **131**, 675–89 (2008).

[62] Williams, S.R., Searles, D.J., Evans, D.J.: Phys. Rev. Lett. **100**, 250601/1–4 (2008).

[63] Chetrite, R., Gawędzki, K.: Commun. Math. Phys. **282**, 469–518 (2008).

[64] Liphardt, J., Dumont, S., Smith, S.B., Tinoco I. Jr., Bustamante, C.: Science **296**, 1832–35 (2002).

[65] Douarche, F., Ciliberto, S., Petrosyan, A.: J. Stat. Mech.: Theory and Experiment P09011/1–18 (2005).

[66] Blickle, V., Speck, T., Helden, L., Seifert, U., Bechinger, C.: Phys. Rev. Lett. **96**, 070603/1–4 (2006).

[67] Douarche, F., Ciliberto, S., Petrosyan, A., Rabbiosi, I.: Europhys. Lett. **70**, 593–99 (2005).

[68] Goldstein, H.: Classical Mechanics. 2nd ed. Addison-Wesley, Reading, Massachusetts (1980).

[69] Bochkov, G.N., Kuzovlev, Yu.E.: Physica A **106**, 443–79 (1981).

[70] Bochkov, G.N., Kuzovlev, Yu.E.: Physica A **106**, 480–520 (1981).

[71] Horowitz, J., Jarzynski, C.: J. Stat. Mech.: Theory and Experiment P11002/1–14 (2007).

[72] Chandler, D.: Introduction to Modern Statistical Mechanics. Section 5.5. Oxford University Press, New York (1987).

[73] Hermans, J.: J. Phys. Chem. **114**, 9029–32 (1991).

[74] Wood, R.H., Mühlbauer, C.F., Thompson, P.T.: J. Phys. Chem. **95**, 6670–75 (1991).

[75] Hendrix, D.A., Jarzynski, C.: J. Chem. Phys. **114**, 5974–81 (2001).

[76] Speck, T., Seifert, U.: Phys. Rev. E **70**, 066112/1–4 (2004).

[77] Cohen, E.G.D., Berlin, T.H.: Physica **26**, 717–29 (1960).

[78] Evans, D.J., Searles, D.J.: Phys Rev E **53**, 5808–15 (1996).

[79] Jarzynski, C.: Phys. Rev. E **73**, 046105/1–10 (2006).

[80] Gibbs, J.W.: Elementary Principles in Statistical Mechanics. Scribner's, New York (1902).

[81] Campisi, M.: Studies in History and Philosophy of Modern Physics **39**, 181–94 (2008).

[82] Sato, K.: J. Phys. Soc. Japan **71**, 1065–66 (2002).

[83] Marathe, R., Parrondo, J.M.R.: arXiv:0909.0617v2, (2010).

[84] Jarzynski, C.: Proc. Natl. Acad. Sci. (USA) **98**, 3636–38 (2001).

[85] Berkovich, R., Klafter, J., Urbakh, M.: J. Phys.: Condens. Matter **20**, 345008/1–7 (2008).

[86] Collin, D., Ritort, F., Jarzynski, C., Smith, SB., Tinoco, I. Jr., Busamante, C.: Nature **437**, 231–34 (2005).

[87] Planck, M.: 1927. Treatise on Thermodynamics, 3rd ed. (translated from 7th German edition). Longmans, Green and Co., Inc., London. Page 89, (1927), (Dover edition, 1945).

[88] Douarche, F., Joubaud, S., Garnier, N.B., Petrosyan, A., Ciliberto, S.: Phys. Rev. Lett. **97**, 140603/1–4 (2006).

[89] Cleuren, B., Van den Broeck, C., Kawai, R.: Phys. Rev. Lett. **96**, 050601/1–4 (2006).

[90] Tolman, R.C.: Proc. Natl. Acad. Sci. (USA) **11**, 436–39 (1925).

[91] Cover, T.M., Thomas, J.A.: Elements of Information Theory. John Wiley and Sons, Inc. New York (1991).

[92] Qian, H.: Phys. Rev. E **63**, 042103/1–4 (2001).

[93] Kawai, R., Parrondo, J.M.R., Van den Broeck, C.: Phys. Rev. Lett. **98**, 080602/1–4 (2007).

[94] Vaikuntanathan, S., Jarzynski, C.: EPL **87**, 60005/1–6 (2009).

[95] Feng, E.H., Crooks, G.E.: Phys. Rev. Lett. **101**, 090602/1–4 (2008).

[96] Eddington, A.S.: The Nature of the Physical World. Macmillan (1928).

[97] Shirts, M.E., Bair, E., Hooker, G., Pande, V.S.: Phys. Rev. Lett. **91**, 140601/1–4 (2003).

[98] Maragakis, P., Ritort, F., Bustamante, C., Karplus, M., Crooks, G.E.: J. Chem. Phys. **129**, 024102/1–8 (2008).

[99] Jarzynski, C.: J. Stat. Phys. **96**, 415–27 (1999).

[100] Esposito, M., Lindenberg, K., Van den Broeck, C.: New J. Phys. **12**, 013013/1–10 (2010).

[101] Evans, D.J., Hoover, W.G., Failor, B.H., Moran, B., Ladd, A.J.C.: Phys. Rev. A **28**, 1016–21 (1983).

[102] Touchette, H.: Phys. Reports **478**, 1–69 (2009).

[103] Chernyak, V.Y., Chertkov, M., Jarzynski, C.: J. Stat. Mech.: Theory and Experiment P08001/1–19 (2006).

[104] Esposito, M., Van den Broeck, C.: Phys. Rev. Lett. **104**, 090601/1–4 (2010).

[105] Maes, C., Netočný, K.: J. Stat. Phys. **110**, 269–310 (2003).

[106] Gaspard, P.: J. Stat. Phys. **117**, 599–615 (2004).

[107] Park, S., Schulten, K.: J. Chem. Phys. **120**, 5946–61 (2004).

[108] Chipot, C., Pohorille, A.: Free Energy Calculations. Springer, Berlin (2007).

[109] Chami, F., Wilson, M.R.: J. Am. Chem. Soc. **132**, 7794–7802 (2010).

[110] Bustamante, C.: Quar. Rev. Biophys. **38**, 291–301 (2005).

[111] Harris, N.C., Song, Y., Kiang, C.-H.: Phys. Rev. Lett. **99**, 068101/1–4 (2007).

[112] Greenleaf, W.J., Frieda, K.L., Foster, D.A.N., Woodside, M.T., Block, S.M.: Science **319**, 630–33 (2008).

[113] Junier, I., Mossa, A., Manosas, M., Ritort, F.: Phys. Rev. Lett. **102**, 070602/1–4 (2009).

[114] Shank, E.A., Cecconi, C., Dill, J.W., Marqusee, S., Bustamante, C.: Nature **465**, 637–41 (2010).

[115] Speck, T., Seifert, U.: J. Phys. A: Math. Gen. **38**, L581–88 (2005).

[116] Schmiedl, T., Seifert, U.: J. Chem. Phys. **126**, 044101/1–12 (2007).

[117] Schmiedl, T., Speck, T., Seifert, U.: J. Stat. Phys. **128**, 77–93 (2007).

[118] Sekimoto, K.: Prog. Theor. Phys. Suppl **130**, 17–27 (1998).

[119] Ge, H., Qian, H.: Phys. Rev. E **81**, 051133/1–5 (2010).

[120] Oono, Y., Paniconi, M.: Prog. Theor. Phys. Suppl **130**, 29–44 (1998).

[121] Trepagnier, E.H., Jarzynski, C., Ritort, F., Crooks, G.E., Bustamante, C.J., Liphardt, J.: Proc. Natl. Acad. Sci. (USA) **101**, 15038–41 (2004).

[122] Komatsu, T.S., Nakagawa, N., Sasa, S.-I., Tasaki, H.: Phys. Rev. Lett. **100**, 230602/1–4 (2008).

[123] Komatsu, T.S., Nakagawa, N., Sasa, S.-I., Tasaki, H.: J. Stat. Phys. **134**, 401–23 (2009).

[124] Yukawa, S.: J. Phys. Soc. Japan **69**, 2367–70 (2000).

[125] Kurchan, J.: arXiv:cond-mat/0007360v2, (2000).

[126] Tasaki, H.: arXiv:cond-mat/0009244v2, (2000).

[127] Mukamel, S.: Phys. Rev. Lett. **90**, 170604/1–4 (2003).

[128] Talkner, P., Lutz, E., Hänggi, P.: Phys. Rev. E **75**, 050102(R)/1–2 (2007).
[129] Talkner, P., Hänggi, P.: J. Phys. A: Math. Theor. **40**, F569–71 (2007).
[130] Teifel, J., Mahler, G.: Phys. Rev. E **76**, 051126/1–6 (2007).
[131] Quan, H.T., Yang, S., Sun, C.P.: arXiv:0804.1312v1, (2008).
[132] Talkner, P., Hänggi, P., Morillo, M.: Phys. Rev. E **77**, 051131/1–6 (2008).
[133] Talkner, P., Burada, P.S., Hänggi, P.: Phys. Rev. E **78**, 011115/1–11 (2008).
[134] Huber, G., Schmidt-Kaler, S., Deffner, S., Lutz, E.: Phys. Rev. Lett. **101**, 070403/1–4 (2008).
[135] Van Zon, R., Hernández de la Peña, L., Peslherbe, G.H., Schofield, J.: Phys. Rev. E **78**, 041103/1–11 (2008).
[136] Van Zon, R., Hernández de la Peña, L., Peslherbe, G.H., Schofield, J.: Phys. Rev. E **78**, 041104/1–14 (2008).
[137] Crooks, G.E.: J. Stat. Mech.: Theor. Experiment P10023/1–9 (2008).
[138] Quan, H.T., Dong, H.: arXiv:0812.4955v1, (2008).
[139] Talkner, P., Campisi, M., Hänggi, P.: J. Stat. Mech.: Theor. Experiment P02025/1–12 (2009).
[140] Deffner, S., Lutz, E.: arXiv:0902.1858v1, (2009).
[141] Andrieux, D., Gaspard, P., Monnai, T., Tasaki, S.: New J. Phys. **11**, 043014/1–25 (2009).
[142] Campisi, M., Talkner, P., Hänggi, P.: Phys. Rev. Lett. **102**, 210401/1–4 (2009).
[143] Hide, J., Vedral, V.: Phys. Rev. A **81**, 062303/1–5 (2010).
[144] Ohzeki, M.: Phys. Rev. Lett., in press (2010).
[145] Maruyama, K., Nori, F., Vedral, V.: Rev. Mod. Phys. **81**, 1–23 (2009).
[146] Piechocinska, B.: Phys. Rev. A **61**, 062314/1–9 (2000).
[147] Sagawa, T., Ueda, M.: Phys. Rev. Lett. **100**, 080403/1–4 (2008).
[148] Dillenschneider, R., Lutz, E.: Phys. Rev. Lett. **102**, 210601/1–4 (2009).
[149] Sagawa, T., Ueda, M.: Phys. Rev. Lett. **102**, 250602/1–4 (2009).
[150] Sagawa, T., Ueda, M.: Phys. Rev. Lett. **104**, 090602/1–4 (2010).

Christopher Jarzynski
Department of Chemistry and Biochemistry
and Institute for Physical Science and Technology
University of Maryland
College Park MD 20742, USA

Time, 173–186

Poincaré Seminar 2010

Time Measurement in the XXI$^{\text{st}}$ Century

Christophe Salomon

Abstract. After a brief historical review on the evolution of man-made clocks, we describe the principle and performances of modern atomic clocks. These clocks are based on transitions between long-lived states of laser cooled atoms and reach an error of about one picosecond per day, i.e., 1 second over 3 billion years or 5 seconds over the age of the universe. In a little more than 4 centuries after the invention of the pendulum,, clocks have gained 13 orders of magnitude and over the last 50 years the gain is about a factor 10 every 10 years. We then discuss fundamental physics tests that are possible with these ultrastable clocks operating on the Earth or onboard satellites, such as relativity tests or searches for a temporal drift of the fundamental constants of physics. As an example, the clock gravitational shift test planned for the European ACES/PHARAO space mission is presented. Finally we argue that the fluctuations of the Earth gravitational potential through Einstein's red shift will soon become a serious limit to the precision of time keeping on Earth and we propose to circumvent this limitation by installing such clocks onboard high-orbiting satellites.

The quest for precision: a brief history

Since the most ancient times, humans have always searched to measure time or more precisely time intervals with greater and greater precision. Initially natural phenomena that were observed to be very regular such as the Moon rotation around the Earth or the Earth rotation around the Sun or on itself have been used to create calendars and time scales. People have soon tried to realize themselves instruments that were more appropriate to their daily life or more precise than the natural phenomena. One of the most ancient known instrument to measure time intervals is the Egyptian sand-glass. However it is only at the beginning of the 17$^{\text{th}}$ century with the discovery of the pendulum by Galileo and its refinement by Huygens that these instruments began to reach a precision of a few tens of seconds per day (Figure 1).

The small oscillations of the pendulum are independent of the motion's amplitude and are remarkably regular, i.e., periodic (Figure 2). The unique property

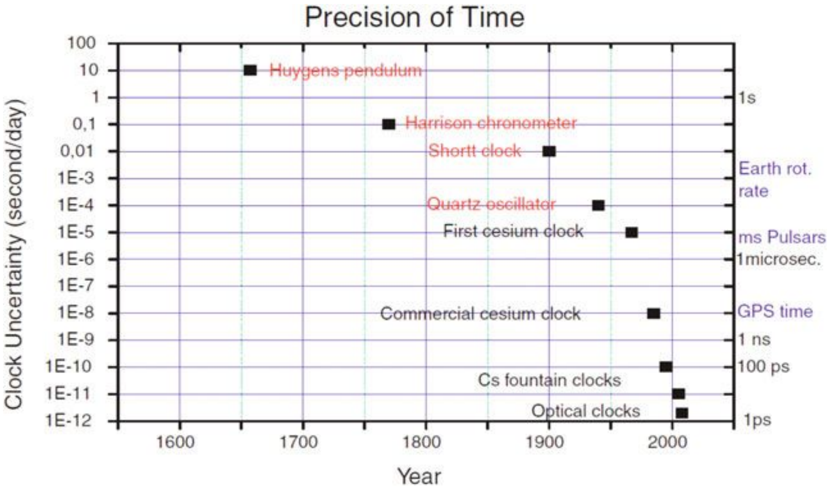


FIGURE 1. Improvement over the last 4 centuries of the precision of time. Before 1950, clocks used mechanical systems as pendulum or spring watches. Since the middle of the 20$^{\text{th}}$ century, the most precise clocks are atomic clocks. Recent atomic clocks have an error that does not exceed 1 second every 3 billion years or about 5 seconds over the age of the universe.

of the pendulum is that the period of the motion depends only on a small number of parameters, the wire length l and the local gravity acceleration $g : T = 2\pi(l/g)^{1/2}$. This simple formula also indicates the limits of this instrument. If the wire length changes for instance through temperature variations or if the local gravity changes, for instance by moving to higher elevation, the pendulum clock will lose its precision. Driven by transoceanic navigation and the determination of the longitude, and by the high money prizes promised by several sovereigns, the art of developing mechanical clocks, watches, and chronometers with ever increasing precision has flourished in the two centuries following the discovery of the pendulum. Fighting against temperature and humidity changes with clever tricks over more than 40 years, Harrison produced in 1759 a chronometer that had an error of only 0.1 second per day. His chronometer was tested at sea by a British vessel that made a round-trip Portsmouth-les Antilles in slightly less than 3 months. At return the onboard chronometer had accumulated less than 5 seconds error in comparison with fixed instruments. Harrison received for this feat a 40000 pound prize, a considerable amount at the time!

The next breakthrough was the invention of the quartz oscillator in 1918. The periodic system is no longer a mechanical system but an oscillating electromagnetic signal generated by the piezoelectric effect in quartz. The mechanical vibration of the quartz crystal creates an oscillatory electric field at a well-defined

FIGURE 2. The Galileo/Huygens pendulum. The small oscillations are remarkably regular and periodic. When counting the number of oscillations of the pendulum, we measure a time interval, thus realizing a clock. The shorter the pendulum period, the greater number of oscillations in a time interval and the more precise will be the determination of this time interval. Modern clocks use electromagnetic signals where the period is 10^{10} to 10^{15} times shorter than the typical 1 second period of a pendulum.

frequency that is a million times higher than the pendulum frequency (a few million oscillations per second). The electric field is electronically amplified until it reaches the threshold of self-sustained oscillations when the amplifier gain exceeds the losses in the system. Quartz oscillators exist in every modern electronic device and have invaded modern societies. Their precision is sufficient for most common life applications.

However, as the pendulum, but to a much smaller degree, a quartz oscillator is subject to temperature drifts. In order to control and suppress this effect physicists have invented the atomic clock, a device that combines the properties of the quartz oscillator with the properties of an atom. Quantum mechanics indicates that the energy of an atom can only take discrete values. The energy is quantized and the difference in energy between two atomic states is well defined and no longer depends on temperature (to a very good approximation). Furthermore atoms are universal: a cesium atom in different locations (Paris, New York, Tokyo) will have the same energy spectrum and properties unlike pendulum or quartz oscillators that are impossible to realize in an identical fashion. The first cesium clock was realized in the United Kingdom by Essen and Parry in 1955. Its operation principle

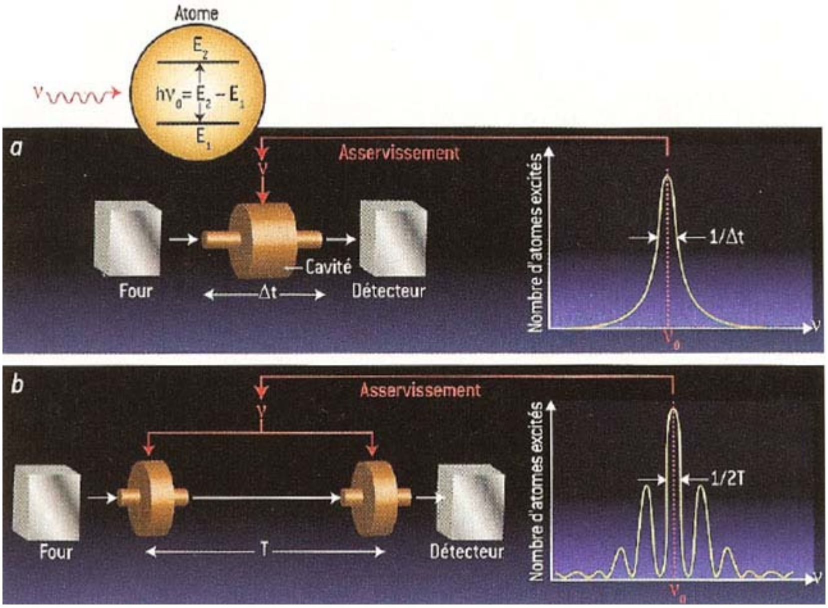


FIGURE 3. Principle of an atomic clock. An electromagnetic radiation with frequency ν is sent to an ensemble of atoms with two energy levels E_1 (ground state) and E_2 (excited state). The energy difference is $h\nu_o$ where h is Planck's constant. The cesium atoms exiting an oven form an atomic beam that traverses a cavity where the microwave radiation is applied. A detector records the number of atoms transferred from the ground state to the excited state when the radiation frequency ν is close to the atomic frequency ν_0. When ν is scanned around ν_0 the number of excited atoms has the shape of a resonance curve centered in ν_0. The resonance width is inversely proportional to the interaction time Δt between the radiation and the atoms. The method invented by N. Ramsey uses two interaction zones separated in space so that atoms experience two interactions separated in time by an interval T. A quantum interference phenomenon then produces a resonance with a sinusoidal shape (called Ramsey fringes) having a width $1/2T$. If the atoms are slow, the width of the fringes is narrow and the better will be the clock.

is described in Figure 3 and its daily error was about $10\,\mu s$, much lower than the error of quartz crystals and mechanical devices.

The 13$^{\text{ème}}$ Conférence Générale des Poids et Mesures in 1967 has chosen the cesium atom for providing the current definition of the second in the international system of Units. "The second is the duration of 9 192 631 770 periods of the radiation that corresponds to the transition between two hyperfine energy levels of the ground electronic state of cesium 133". Since 1967, several thousands of commercial cesium clocks have been produced for a number of applications, the satellite Global Positioning System (GPS) being one prominent example. From 1990 on and during the next following years, laser cooled atoms have enabled to increase the precision of cesium clocks by two additional orders of magnitude. At a temperature of one microkelvin, cesium atoms have a thermal speed of only 7 mm/s. They can be used in an atomic fountain so that the interaction time with the exciting radiation can approach 1 second (Figure 4). This duration is typically 100 to 1000 times longer than in the atomic beam machine introduced by Essen and Parry. Today the LNE-SYRTE atomic fountains and several others worldwide operate in a routine manner with an error of about 10 picoseconds per day. The relative frequency stability reaches $7\,10^{-17}$ after 10 days of averaging [1, 2]. About 15–20 fountain devices spread around the world are regularly compared at a distance by the GPS system and contribute to the realization of the Temps Atomique International (TAI), the worldwide time reference. One interesting application of TAI is the long term monitoring of binary pulsars. As discovered by Hulse and Taylor, pulsars emit very regular pulses of radiation at millisecond rate and constitute remarkably stable natural clocks of gravitational nature. The slow decay of their orbital period has been a clear signature of emission of gravitational waves. Since their discovery, several hundred binary pulsars have been detected and the most stable among them (more than 20) are regularly compared to TAI [3]. Even if the stability of pulsar time does not match that of modern atomic clocks, it is extremely interesting to compare over the long term a gravitation-based time scale realized from pulsars with a quantum physics based time scale realized by atomic clocks.

In the last two years, new developments on atomic clocks have enabled researchers to reach an error level near 1 picosecond per day, i.e., 1 second over 3 billion years or 5 seconds over the age of the universe. These new clocks are optical clocks. Instead of using an electromagnetic field that oscillates at a frequency of about 10^{10} Hertz as in a cesium clock, optical clocks use visible or ultraviolet light near 10^{15} Hertz. The pendulum now beats 10^{15} times per second, five orders of magnitude faster than in the microwave domain with the hope to gain in the future this factor on the clock performance. This visible light is produced by a laser and this laser is servo-locked to an atomic transition in the visible or UV range. This transition uses a long-lived excited state in the atom so that its lifetime is not a limit to the operation of the clock. Optical clocks now display a relative frequency stability of about $1\,10^{-17}$ for an averaging time of 3000 seconds and an accuracy of $9\,10^{-18}$ [4]. With such devices, a group at NIST (USA) has been able to show that general relativity effects can be detected at small distance and velocity scales (human scales are meter distances and meters/second speeds). Lifting up by 33 cm one optical clock with respect to another identical one nearby, the NIST group has shown that the lower clock runs slower by a tiny fraction, 4 parts in 10^{17}, in accordance with Einstein prediction for the gravitational shift of clocks near the surface of the Earth [5]. This experiment shows the extreme sensitivity of optical

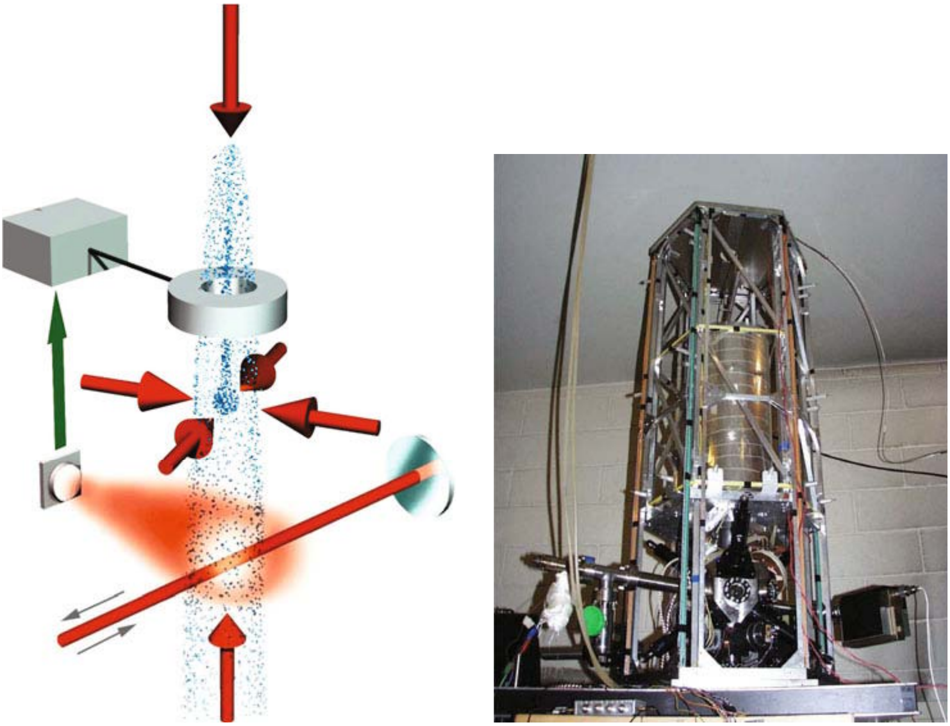


FIGURE 4. (a) An atomic fountain. Cesium atoms are cooled using laser light to a temperature of 1 microKelvin and launched upwards at a speed of 4 m/s. They cross on their way up and down a cavity that contains the microwave field with frequency ν near 9.2 GHz that excites them. The time separation between the two interactions reaches 0.5 second and is about 100 times longer than in a thermal atomic beam clock as illustrated in Figure 3. (b) A cold atom fountain clock from LNE-SYRTE at Paris Observatory.

clocks. However it is not a precision test of Einstein's redshift because of the small height difference. We show below that by comparing ground-based and satellite clocks gravitational shift tests can be performed with high sensitivity.

Here an immediate question arises: how to count such ultrafast light oscillations? The answer came from two of the three 2005 Nobel laureates, T.W. Haensch and J.L. Hall [6]. They have constructed a frequency divider that enables to divide optical frequencies towards the microwave domain where frequencies can be counted with fast detectors. The system uses a laser that emits ultrashort pulses of near infrared light of duration $\sim$ 30 femtoseconds at a very regular rate of typically 100 MHz. When properly controlled, this laser emits in a broad frequency range a comb of several million lines that are integer multiple of this 100 MHz frequency and spaced in the frequency domain by exactly 100 MHz. If this 100 MHz frequency comes from a very stable clock, then all lines of the comb are equally stable, and constitute individual clocks that cover most of the visible part and near infrared part of the electromagnetic spectrum. This simple laser device has brought a revolution in quantum optics, accelerating considerably the development of optical frequency metrology and spectroscopy.

As shown in Figure 1, the improvement of clocks has been spectacular. In a little more than 4 centuries, modern atomic clocks have gained 13 orders of magnitude with respect to the Huygens pendulum. Over the last 50 years the gain is about a factor 10 every 10 years. No one knows if such a pace will continue in the future!

Cold atom clocks and the PHARAO Space clock

The principle of fountain clocks is illustrated in Figure 4 [7]. The motion of atoms in the gravity field is used to increase the interaction time between the oscillatory field and the atoms to about 0.5 second, providing a Fourier transform limited resonance width of 1 Hz. This is accomplished using the elegant method proposed by N. Ramsey in 1952. It is not necessary that the exciting field is continuously applied to the atoms. Two successive interactions are sufficient. Because of quantum interference, the resonance curve displays a width no longer given by the inverse of the time duration spent by the atoms in the interrogating field but by the time between the two successive interactions. On the Earth, a natural geometry is thus to make an atomic fountain where the laser cooled atoms are launched upwards against gravity. Atoms then cross the cavity fed by a microwave field where they experience a first interaction with the field. They further travel up until gravity reverses their velocity, and fall through the cavity producing a second interaction with the field, typically half a second later for a 60 cm high trajectory above the cavity. With a resonance width $\delta\nu = 1\,\text{Hz}$, the quality factor of the clock is $\nu/\delta\nu \simeq 10^{10}$. Quantum limited detection with a few 10^6 atoms provides a relative frequency stability of $2\,10^{-14}\,\tau^{-1/2}$ where τ is the measurement time in seconds [1, 2]. By comparison between two fountains, a relative frequency stability of $7\,10^{-17}$ after 10 days of averaging has been measured.

In order to further enhance the interaction time, it is tempting to launch the atoms to much higher elevations and a 10 meter tall fountain is being assembled in Stanford for this purpose. However such dimensions brings new constraints (compensation of the residual magnetic field with large magnetic shields, temperature uniformity, ...) and the gain in frequency resolution is only proportional to the square root of the fountain height (factor of 3 for the Stanford fountain). It appears more radical to get rid of gravity acceleration by operating the clock onboard a satellite. This is the aim of the PHARAO project (Projet d'Horloge Atomique par Refroidissement d'Atomes en Orbite) conducted by the French space agency, CNES, the LNE-SYRTE at Paris Observatory, and LKB at École Normale Supérieure [8].

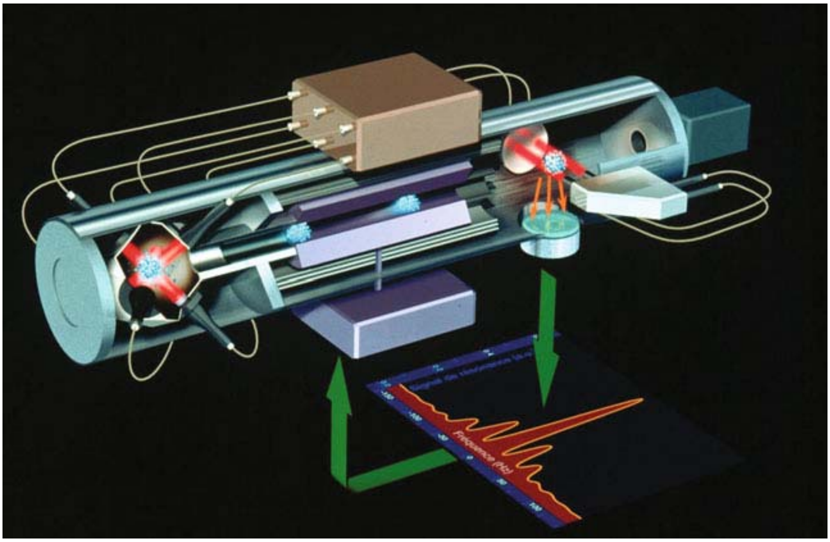


FIGURE 5. Principle of the cold atom space clock PHARAO. Cesium atoms are collected in optical molasses and laser cooled in a first vacuum chamber (left). In a microgravity environment atoms are then launched slowly through a cavity where they undergo the two successive interactions with a microwave field tuned near the hyperfine cesium frequency ν_0=9 192 631 770 Hz. The excited atoms are subsequently detected by laser induced fluorescence. For a launch velocity of 10 cm/ s, the expected resonance width is 0.2 Hz, or 5 times narrower than in an atomic fountain in Earth gravity.

The principle of the microgravity PHARAO clock is illustrated in Figure 5 and is very similar to the Essen and Parry design but with ultra slow atoms produced by laser cooling. Thanks to the absence of gravity onboard a satellite the atoms keep a constant velocity through the device and the two microwave interactions are spatially separated. In a compact set-up with a 1 meter total length, we expect to produce an atomic resonance 5 to 10 times narrower than in a fountain and about 10 000 times narrower than in a commercial cesium clock.

The PHARAO clock will be a core element of the European space mission ACES (Atomic Clock Ensemble in Space) under the responsibility of ESA [10]. The ACES payload will be installed (Figure 8) in 2015 onboard the International Space Station (ISS) that is orbiting around the Earth at an elevation of 400 kms, with an orbital period of 5400 seconds. By clear sky, it is easy to see the ISS with the naked eye about one hour after sunset. Illuminated by the sun the ISS appears as a bright satellite scanning the observer's sky in about 5 minutes.

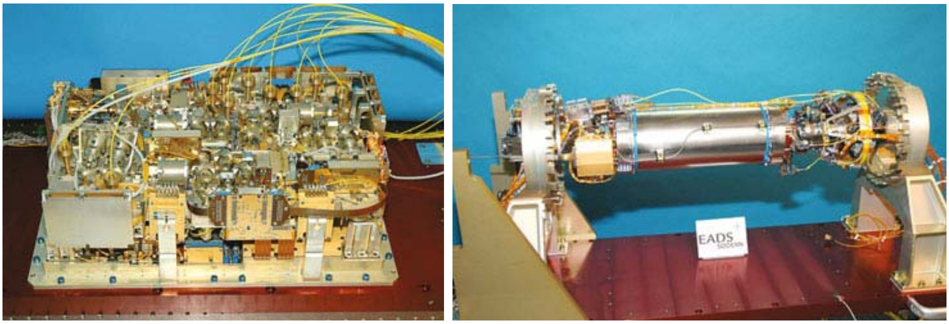


FIGURE 6. (a) The PHARAO clock optical bench. It is based on 8 laser diodes that are stabilized in temperature, frequency, and intensity. The laser beams are guided to the vacuum chamber using optical fibers (yellow). The bench volume is 30 liters and mass 20 kg. (b) Vacuum chamber of the PHARAO clock where cesium atoms interact with the microwave field. The tube length is 900 mm and weight 45 kg. 3 layers of magnetic shields reduce the influence of the varying Earth magnetic field along the satellite orbit. Both the laser bench and cesium tube have been realized by the EADS-SODERN company.

In addition to the PHARAO clock, ACES will carry a second atomic clock, a space hydrogen maser developed in Switzerland, a high accuracy time transfer system developed in Germany, a GPS/GALILEO receiver, a laser time transfer system and support equipments. The overall mass is 220 kg and power consumption 450 Watts.

Fundamental physics tests

The scientific objectives of the ACES mission will cover several domains. First, the combined PHARAO-maser clocks will realize a high stability time scale in space. This time scale will be compared to time scales realized on the ground with a network of high stability clocks. Knowing the gravitational potentials of the space and ground clocks, it will be possible to perform a precision measurement of the clock gravitational shift. Measured on the ground, the frequency of the PHARAO clock will appear upshifted by an amount gH/c^2 where g is the gravity acceleration, H the mean distance between the ISS and the considered ground clock, and c the speed of light. With a 400 kms elevation difference we expect a relative frequency shift of $+4.5\,10^{-11}$. With the PHARAO clock accuracy of 10^{-16}, the Einstein effect will be tested with a precision of $2\ 10^{-6}$, an improvement by a factor 70 over the current best determination of this effect by the NASA Gravity Probe A 1976 rocket mission [9].

Second, with a 51 degree inclination, the ISS will fly over time and frequency laboratories spread around the world several times per day. Repeated frequency

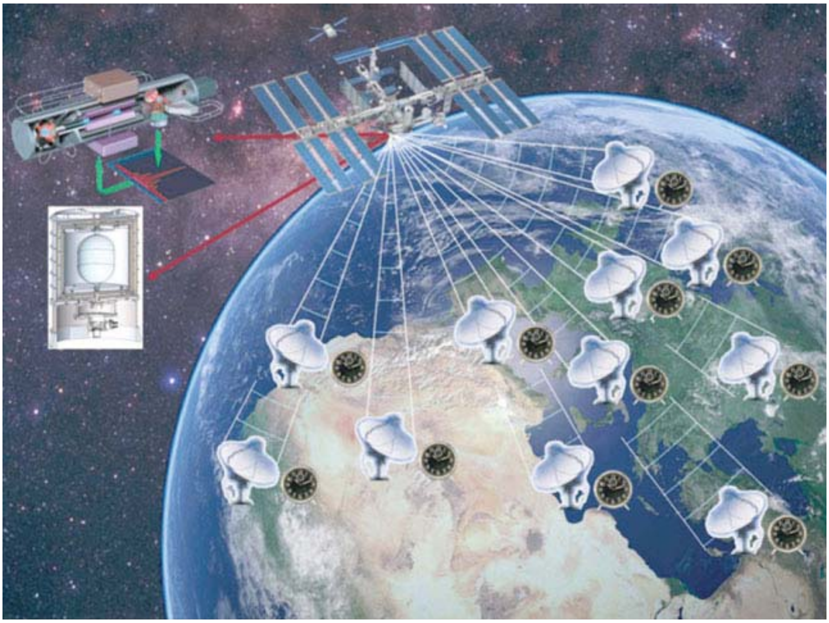


FIGURE 7. The space mission ACES. The PHARAO clock and a space hydrogen maser will be installed on the International Space Station in 2015. Ultra precise time comparisons of the space clocks with a network of high stability ground clocks will provide tests in fundamental physics and general relativity. A GPS/GALILEO receiver will provide the ISS orbital parameters required for a precision measurement of the gravitational shift (Einstein effect) and also allow applications in Earth observation and navigation.

comparisons between the space clocks and the ground clocks that operate with different atoms and or different atomic transitions will enable to test the stability of the fundamental constants of physics. As first proposed by P. Dirac in 1937, it is interesting to test whether the dimensionless constants of physics are truly constant over time. For instance the fine structure constant $\alpha = 1/137035999074(44)$ that characterizes the electromagnetic interaction and is responsible for the stability of atoms and molecules is a dimensionless number. Similarly the ratio between the proton mass and the electron mass is also a pure number. We know that the Universe is expanding and that its expansion is accelerating under the effect of elusive dark energy. We also know that 95 percent of the mass of the universe is of unknown origin, an uncomfortable situation! Is this due to new particles that couple to matter or to a change of the laws of physics at some distance scale? On the theory side, attempts to unify gravitation that is a classical theory with the Standard Model that is a quantum theory do predict that the fundamental con-

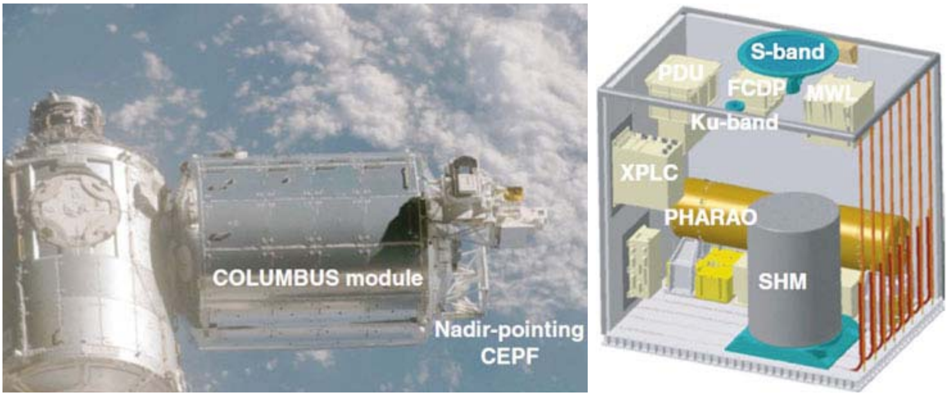


FIGURE 8. Left: the European Columbus module of the ISS and the nadir oriented platform where the ACES equipments will be installed. Right: global view of the ACES payload with the PHARAO clock, the space hydrogen maser, the time transfer system operating in the microwave domain, and the support equipments.

stants may vary in time [11]. The high precision of modern atomic clocks can bring this fundamental test to a new level of precision and perhaps discover that indeed fundamental constants may slowly vary over cosmological scales ! Such a finding would constitute a violation of Einstein's equivalence Principle that underpins our present understanding of the laws of physics. The objective of the ACES mission is to test the time stability of fundamental constants at a level of 10^{-17}/year or $3\,10^{-18}$/year over a three year mission duration.

A few applications

In addition to the fundamental tests described above and to a significant improvement of the International Atomic Time made possible by these clock comparisons on a global scale at a level of 10^{-17}, the ACES mission has also more applied goals. The first one deals with geodesy. After having tested to a high level the validity of the gravitational clock shift, ultra stable clocks can be used to perform a new type of "relativistic geodesy". Take for instance two identical clocks at rest at a distance in two different gravitational potentials. If their frequencies are compared by the ACES mission or, for continental distances by time transfer using optical fibers, the frequency difference between the two clocks will directly provide the difference in gravitational potential between the two clocks. Near the Earth the sensitivity coefficient is 10^{-16} per meter of elevation difference. Current optical clocks with 10^{-17} accuracy will provide gravitational potential differences at 10 cm level, a value already competitive with space-based measurements done on a global scale by space mission such as GRACE, CHAMP or GOCE. At 10^{-18}, clocks will probe gravitational potential variations at the 1 cm level. At this level,

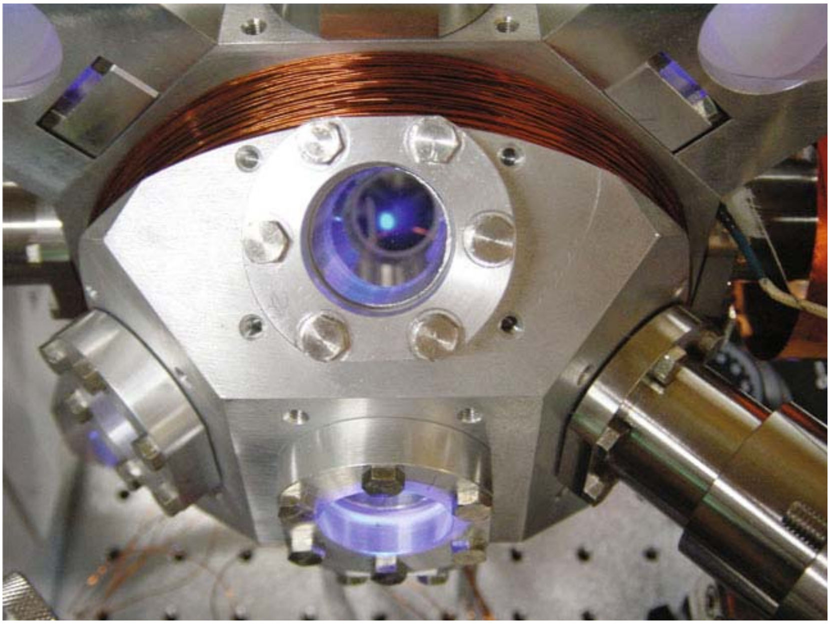


FIGURE 9. An optical clock using strontium atoms developed at Paris Observatory. The blue fluorescence of strontium atoms is visible at the center of the vacuum chamber. This type of clock is one of the most promising to reach a frequency stability in the $10^{-18} - 10^{-19}$ range. Courtesy Pierre Lemonde.

the local potential on the Earth is no longer stable; it varies with the ocean tides, with the atmospheric pressure and winds and also with the amount of water stored in the ground in the neighborhood of the clock. Clocks thus will become a new type of geodetic sensor but, conversely, one can predict that these gravitational fluctuations will limit the progress of ultra precise time realizations on the Earth. A solution to this problem is to install these ultrastable clocks in space where the Earth potential fluctuations decrease very fast with the distance to Earth.

A second class of applications of the ACES mission relies on the GPS/GALILEO receiver installed onboard the ACES platform and connected to the ultrastable ACES timescale for studying the Earth atmosphere. The radio-occultation method that is already used in meteorological satellites to monitor the temperature and humidity of the troposphere can be also employed on the ISS that has an advantageous orbit [12]. The satellites of the GPS constellation permanently emit microwave signals that are dephased and attenuated by the Earth atmosphere. In the radio-occultation method, when a GPS satellite rises over the horizon of the ACES GPS receiver antenna, its GPS signal is traversing a long sample of the atmosphere and is delayed and attenuated by a quantity that depends on the water vapor content and temperature of the atmosphere. Geodesy scientists have developed models that enable to reconstruct vertical temperature and humidity profiles with 500 meters vertical resolution. These maps are then used in global meteo models for wheather forecasts. Similarly GPS signals are reflected on the ocean surface and the reflected signals contain informations on the height and direction of the waves.

Future

The space mission ACES is in its last realization phase before the launch towards the ISS. Engineering models of the instruments have been realized and tested in the last 2 years. The PHARAO clock flight model is being fabricated in industry and will be delivered for tests at CNES Toulouse during 2012. After a year of performance tests, the clock will be delivered to the ACES prime contractor EADS Astrium in Friedrichshafen for assembly on the ACES palette with the other instruments. In parallel the mission exploitation scenario and data analysis are being prepared. The launch to the ISS will be made in 2015 by a Japanese H2B rocket and the ISS robotic arm will install the ACES palette on its allocated site. The mission duration is from 18 months to 3 years.

If we turn back to the exponential increase of clock performances in Figure 1, the comparison between microwave and optical clocks shows that it is now time to change the definition of the second from the current cesium atom to another species that will provide a better time standard. Indeed optical clocks such as the aluminium ion clock developed at NIST and the strontium lattice clock at LNE-SYRTE (Figure 9) have now surpassed the cesium cold atom standards by up to a factor 20 [4, 13]. After all, the current definition based on cesium was a good choice as it served for more than 45 years ! Candidate atoms, both trapped ions and neutral atoms in optical lattices, are numerous and progress in this domain is very fast with a stability range of 10^{-18}–10^{-19} in view! Finally, as pointed out above, ground based clocks will become geodetic sensors, ultimately limited by fluctuations in the Earth potential. Installing one or several ultrastable optical clocks in space would circumvent this problem and enable further improvements in precision time together with more refined fundamental physics tests [14].

References

[1] Santarelli, G., et al.: Phys. Rev. Lett. **82**, 4619 (1999).

[2] Bize, S., et al.: J. Phys. B: At. Mol. Opt.Phys. **38** S449 (2005).

[3] Verbiest, J.P.W., et al.: arXiv:0908.0244v1.

[4] Chou, C.W., et al.: Phys. Rev. Lett. **78**, 070802 (2010).

[5] Chou, C.W., Hume, D.B., Rosenband, T., and Wineland, D.J.: Science **329**, 1630 (2010).

[6] Haensch, T.W.: Review of Modern Physics **78**, 1297 (2006); Hall, J.L.: Review of Modern Physics **78**, 1279 (2006).

[7] Clairon, A., Salomon, C., Guellati, S., and Phillips, W.D.: Europhys. Lett. **16**, 165 (1991).

[8] Laurent, P., et al.: *Applied Physics B: Lasers and Optics* **84**, 683 (September 2006). See also: http://smsc.cnes.fr/PHARAO/Fr/.

[9] Vessot, R.F.C., et al.: Phys. Rev. Lett. **45**, 2081(1980).

[10] Cacciapuoti, L., and Salomon, C.: European Physical Journal (Special topics) **172**, 57–68 (2009).

[11] Damour, T., Piazza, F., Veneziano, G.: Phys. Rev. Lett. **89**, 081601 (2002).

[12] Beyerle, G., and Hocke, K. : Geophys. Res. Lett. **28**, 1895 (2001).

[13] Westergaard, P.G., et al.: Phys. Rev. Lett. **106**, 210801(2011).

[14] Schiller, S., et al.: Nuclear Physics B (Proc. Suppl.) **166**, 300–302 (2007).

Christophe Salomon
Laboratoire Kastler Brossel
Ecole Normale Supérieure
CNRS and UPMC
24, rue Lhomond
F-75231 Paris, France

Time, 187–215

Poincaré Seminar 2010

Time's Arrow and Eddington's Challenge

Huw Price

Abstract. When Sir Arthur Eddington died in 1944, TIME magazine noted that "one of mankind's most reassuring cosmic thinkers" had passed away: "Sir Arthur," TIME said, had "discoursed on his cosmic subject with a wit and clarity rare among scientists." One of Eddington's favorite cosmic subjects was "time's arrow", a term he himself introduced to the literature in his 1928 book, The Nature of the Physical World – though without his celebrated clarity about what it actually means, as his philosophical critics were later to note. What is clear is that Eddington thought that there is something essential about time that physics is liable to neglect: the fact that it "goes on", as he often puts it. Despite the best efforts of many writers to pour cold water on this idea, similar claims are still made today, in physics as well as in philosophy. All sides in these debates can profit, in my view, by going back to Eddington. Eddington appreciates some of the pitfalls of these claims with greater clarity than most of their contemporary proponents; and also issues a challenge to rival views that deserves to be better known.

1. A head of his time

The phrase 'time's arrow' seems to have been first introduced to physics by Sir Arthur Eddington, in *The Nature of the Physical World* (1928), [14] based on his Gifford Lectures in Edinburgh the previous year. Eddington's work is little-known to contemporary readers, but he was one of the leading scientific writers of his day. He even reached the cover of TIME magazine, in 1934. (See Figure 1 – the inscription beneath Eddington's name reads "His universe expanded into popularity.")

When Eddington died, ten years later, TIME reported that "one of mankind's most reassuring cosmic thinkers" had passed away:

> Death came at 61 to cool, unruffled Sir Arthur Stanley Eddington, Cambridge University astronomer.... To scientists, Sir Arthur was affectionately known as the senior partner in the firm of "Eddington & Jeans, Interpreters of the Universe." Shy, neat, reed-nosed Sir Arthur looked

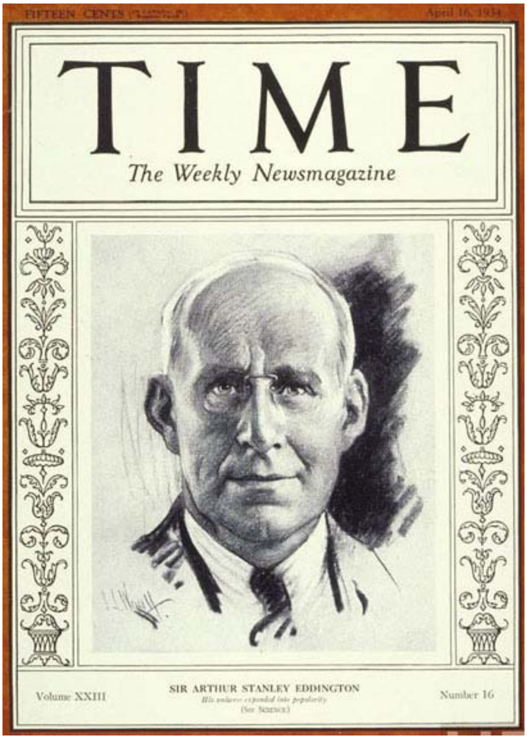


FIGURE 1. Sir Arthur Eddington (1882–1944).

precisely like the British university don he was, and he discoursed on his cosmic subject with a wit and clarity rare among scientists.

Eddington was astute, as well popular, and for those of us interested in the physics and philosophy of time, wit and clarity are not the only reasons to go back to his work. About time's arrow itself, in fact, his famous clarity is sometimes missing. What he himself means by 'time's arrow' is not always entirely clear. I think it is fair to say that he was actually discussing several different notions, and does not completely succeed in distinguishing them, or understanding the connections between them. But there are gems, too, and in some respects Eddington was well ahead not only of all his contemporaries, but also of most writers since.

In this paper, I want to try to provide some clarity, and review progress, concerning some of the issues Eddington discusses, under the heading 'time's arrow'. In some respects, as I'll explain, we have made a lot of progress since Eddington's day. If we haven't found all the answers, at least we have a better understanding where the true puzzles lie. In other respects, I think, progress has not been fast, or extended very far – but it might be encouraged, I suggest, by reminding ourselves of some of the elements of Eddington's discussion of these problems.

2. Introducing 'time's arrow'

Let's begin with the passage in Eddington's book in which the term 'time's arrow' makes its first appearance:

> *Time's Arrow.* The great thing about time is that it goes on. But this is an aspect of it which the physicist sometimes seems inclined to neglect. In the four-dimensional world ... the events past and future lie spread out before us as in a map. The events are there in their proper spatial and temporal relation; but there is no indication that they undergo what has been described as "the formality of taking place" and the question of their doing or undoing does not arise. We see in the map the path from past to future or from future to past; but there is no signboard to indicate that it is a one-way street. Something must be added to the geometrical conceptions comprised in Minkowski's world before it becomes a complete picture of the world as we know it. ([14], p. 34)

Here already we can usefully distinguish two kinds of elements, which Eddington takes to be missing from Minkowski's four-dimensional picture of the world (in which time and space are treated in much the same way). One missing element is what Eddington elsewhere calls "happening", or "becoming", or the "dynamic" quality of time – the fact that time "goes on", as he puts it in the passage above. Time seems in *flux,* to use a much older term, in a way in which space is not, and Eddington is objecting that this aspect of time is missing from the four-dimensional picture.

The other missing ingredient – which Eddington himself doesn't distinguish from the dynamic aspect of time, but which is usefully treated as a distinct idea – is something to give a *direction* to the time axis in Minkowski's picture; something to distinguish past from future, as we might say.

Eddington thinks that both these elements are missing from Minkowski's world, and that physics should be trying to put them back in. Since he doesn't distinguish them, he takes for granted that we should be looking for some one element, which would do both jobs. The reason to keep them distinct, in my view, is that in principle we might agree with Eddington about one but not about the other. For example, we might be persuaded that flux, or "becoming" is entirely subjective, and best left out of physics. Famously, this was the view of Hermann Weyl:

> The objective world simply *is,* it does not *happen.* Only to the gaze of my consciousness, crawling upward along the life line of my body, does a section of this world come to life as a fleeting image in space which continuously changes in time. ([33], p. 116)

Yet we might also think that something objective distinguishes one direction of time from the other, even if it is not some objective process of flux or "happening". (Similarly, perhaps, the other way round: we might think there is something objective about "happening", but treat it as just a conventional matter which way things happen – change is fundamental, but not the *direction* of change!)

So we already have two ideas – not distinguished by Eddington himself – under the heading 'time's arrow'. And we haven't yet got to most important one, which is the time-asymmetry summed-up by the *second law of thermodynamics.* As we shall see later, Eddington sometimes seems deeply puzzled about the relation between this thermodynamic arrow of time and the arrow he calls the "going on" of time. He sometimes seems to think that there might be something more to the latter, accessible only to consciousness. But he seems clear, at least, that the "going on" of time depends on the thermodynamic asymmetry. Considering a region in which matter has already reached thermodynamic equilibrium, he says:

> In such a region we lose time's arrow. You remember that the arrow points in the direction of increase of the random element. When the random element has reached its limit and becomes steady the arrow does not know which way to point. It would not be true to say that such a region is timeless; the atoms vibrate as usual like little clocks; by them we can measure speeds and durations. Time is still there and retains its ordinary properties, but it has lost its arrow; like space it extends, but it does not "go on." ([14], p. 39)

Similarly, returning to this theme in an important lecture published in *Nature* in 1931, he says that in the distant future

> the whole universe will reach a state of complete disorganisation – a uniform featureless mass in thermodynamic equilibrium. This is the end of the world. Time will *extend* on and on, presumably to infinity. But there will be no definable sense in which it can be said to *go* on. ([15], p. 449)

For most of what follows, I shall focus exclusively on the thermodynamic arrow, and especially on the puzzling conflict between the time-asymmetry of thermodynamic phenomena and the time-symmetry of the underlying microphysics on which these phenomena depend. My aim is to provide a guide to the current status of this puzzle, distinguishing the central issue from various issues with which it tends to be confused. In particular, I'll show that there are two competing conceptions of what is needed to resolve the puzzle of the thermodynamic asymmetry, which differ with respect to the number of distinct time-asymmetries they take to be manifest in the thermodynamic arrow. According to one conception, we need two time-asymmetries to explain the thermodynamic arrow; according to the other, we need only one.

I shall offer some reasons for preferring the one-asymmetry view. On this conception of the origin of the thermodynamic arrow, the remaining puzzle concerns the ordered distribution of matter in the early universe. The puzzle of the thermodynamic arrow thus becomes a puzzle for cosmology. As I'll explain, Eddington's 1931 paper looks surprisingly modern, in this context – Eddington seems to have been the first to put his finger on some of the key points that need to be made at this juncture.

At the end of the paper I shall return to some of Eddington's other ideas – the "going on" of time, and the role of consciousness. I think these ideas haven't stood the test of time quite so well, but that there are still some gems here that deserve to be better known. Even where Eddington takes the wrong turn, we can learn a lot by following such an astute and engaging thinker down a few dead ends.

3. The puzzle of the thermodynamic arrow

By the end of the nineteenth century, on the shoulders of Maxwell, Boltzmann and many lesser giants, physics had realised that there is a deep puzzle behind the familiar phenomena described by the new science of thermodynamics. On the one hand, many such phenomena show a striking temporal bias. They are common in one temporal orientation, but rare or non-existent in reverse. On the other hand, the underlying laws of mechanics show no such temporal preference. If they allow a process in one direction, they also allow its temporal mirror image. Hence the puzzle: if the laws are so even-handed, why are the phenomema themselves so one-sided?

What has happened to this puzzle since the 1890s? I suspect that many contemporary physicists regard it as a dead issue, long since laid to rest. Didn't it turn out to be just a matter of statistics, after all? However, while there are certainly would-be solutions on offer – if anything, as we'll see, too many of them – it is far from clear that the puzzle has actually been solved. Late in the twentieth century, in fact, one of the most authoritative writers on the conceptual foundations of statistical mechanics could still refer to an understanding of the time-asymmetry of thermodynamics as 'that obscure object of desire'. [31]

One of the obstacles to declaring the problem solved is that there are several distinct approaches, not obviously compatible with one another. Which of these, if any, is supposed to be *the* solution, now in our grasp? Even more interestingly, it turns out that not all these would-be solutions are answers to the same question. There are different and incompatible conceptions in the literature of what the puzzle of the thermodynamic asymmetry actually is – about what *exactly* we should be trying explain, when we try to explain the thermodynamic arrow of time.

What the problem needs is therefore what philosophers do for a living: drawing fine distinctions, sorting out ambiguities, and clarifying the logical structure of difficult and subtle issues. My aim here is to bring these methods to bear on the puzzle of the time-asymmetry of thermodynamics. I want to distinguish the true puzzle from some of the appealing false trails, and hence to make it clear where physics stands in its attempt to solve it.

Little here is new, but it is surprisingly difficult to find a clear guide to these matters in the literature, either in philosophy or in physics. Accordingly, I think the paper will serve a useful purpose, in helping non-specialists to understand the true character of the puzzle discovered by those nineteenth century giants, the extent to which it has been solved, and the nature of the remaining issues.[1]

4. The true puzzle – a first approximation and a popular challenge

Everyone agrees, I think, that the puzzle of the thermodynamic arrow stems from the conjunction of two facts (or apparent facts – one way to dissolve the puzzle would be to show that one or other of the following claims is not actually true):

1. There are many common and familiar physical processes, collectively describable as cases in which entropy is increasing, whose corresponding time-reversed processes are unknown or at least very rare.
2. The dynamical laws governing such processes show no such T-asymmetry – if they permit a process to occur with one temporal orientation, they permit it to occur with the reverse orientation.

As noted, some people will be inclined to object at this point that the conjunction is *merely* apparent. In particular, it may be objected that we now know that the dynamical laws are not time-symmetric. Famously, T-symmetry is violated in weak interactions, by the neutral K meson. Doesn't this eliminate the puzzle?

No. If the time-asymmetry of thermodynamics were associated with the T-symmetry violation displayed by the neutral K meson, then anti-matter would show the reverse of the normal thermodynamic asymmetry. Why? Because PCT-symmetry guarantees that if we replace matter by anti-matter (i.e., reverse P and C) and then view the result in reverse time (i.e., reverse T), physics remains the same. So if we replaced matter by anti-matter but didn't reverse time, any intrinsic temporal arrow or T-symmetry violation would reverse its apparent direction. In other words, physicists in anti-matter galaxies find the opposite violations of T-symmetry in weak interactions to those found in our galaxy. So if the thermodynamic arrow were tied to the T-symmetry violation, it too would have to reverse under such a transformation.

But now we have both an apparent falsehood, and a paradox. There's an apparent falsehood because (of course) we don't think that anti-matter behaves anti-thermodynamically. We expect stars in anti-matter galaxies to radiate just like our own sun (as the very idea of an anti-matter galaxy requires, in fact). And there's a paradox, because if this were the right story, what would happen to particles which are their own anti-particles, such as photons? They would have to behave both thermodynamically and anti-thermodynamically!

Here's another way to put the point. The thermodynamic arrow is not just a T-asymmetry, it is a PCT-asymmetry as well. There are many familiar process whose PCT-reversed processes are equally compatible with the underlying laws, but which never happen, in our experience. We might be tempted to explain this asymmetry as due to the imbalance between matter and anti-matter, but the above reflections show that this is not so. So instead of the puzzle of the T-asymmetry of

[1]For those interested in more details, I discuss these topics at greater length elsewhere.[23–26]

thermodynamics, we could speak of the puzzle of the PCT-asymmetry of thermodynamics. Then it would be clear to all that the strange behaviour of the neutral K meson is not relevant. Knowing that we could if necessary rephrase the problem in this way, we can safely rely on the simpler formulation, and return to our original version of the puzzle.

5. Four things the puzzle is not

Some of the confusions common in debates about the origins of the thermodynamic asymmetry can be avoided distinguishing the genuine puzzle from various pseudo-puzzles with which it is liable to be confused. In this section I'll draw four distinctions of this kind.

5.1. The meaning of irreversibility

The thermodynamic arrow is often described in terms of the 'irreversibility' of many common processes – e.g., of what happens when a gas disperses from a pressurised bottle. This makes it sound as if the problem is that we can't make the gas behave in the opposite way – we can't make it put itself back into the bottle. Famously, Loschmidt's reversibility objection rested on pointing out that the reverse motion is equally compatible with the laws of mechanics. Some responses to this problem concentrate on the issue as to why we can't actually reverse the motions (at least in most cases). [29]

This response misses the interesting point, however. The interesting issue turns on a numerical imbalance in nature between 'forward' and 'reverse' processes, not case-by-case irreversibility of individual processes. Consider a parity analogy. Imagine a world containing many left hands but few right hands. Such a world shows an interesting parity asymmetry, even if any individual left hand can easily be transformed into a right hand. Conversely, a world with equal numbers of left and right hands is not interestingly P-asymmetric, even if any individual left or right hand cannot be reversed. Thus the interesting issue concerns the numerical asymmetry between the two kinds of structures – here, left hands and right hands – not the question whether one can be transformed into the other.

Similarly in the thermodynamic case, in my view. The important thing to explain is the numerical imbalance in nature between entropy-increasing processes and their T-reversed counterparts, not the practical irreversibility of individual processes.

5.2. Asymmetry in time versus asymmetry of time

Writers on the thermodynamic asymmetry often write as if the problem of explaining this asymmetry is the problem of explaining 'the direction of time'. This may be a harmless way of speaking, but we should keep in mind that the real puzzle concerns the asymmetry of physical processes *in* time, not an asymmetry *of time itself*. By analogy, imagine a long narrow room, architecturally symmetrical end-to-end. Now suppose all the chairs in the room are facing the same end. Then there's a puzzle about the asymmetry in the arrangement of the chairs, but not a puzzle about the asymmetry of the room. Similarly, the thermodynamic asymmetry is an asymmetry of the 'contents' of time, not an asymmetry of the container itself.

It may be helpful to make a few remarks about the phrase 'direction of time'. Although this expression is in common use, it is not at all clear what it could actually mean, if we try to take it literally. Often the thought seems to be that there is an objective sense in which one time direction is future (or 'positive'), and the other past (or 'negative'). But what could this distinction amount to? It is easy enough to make sense of idea that time is *anisotropic* – i.e., different in one direction than in the other. For example, time might be finite in one direction but infinite in the other. But this is not enough to give a *direction* to time, in above sense. After all, if one direction were objectively the future or positive direction, then in the case of a universe finite at one end, there would be two possibilities. Time might be finite in the past, and or finite in the future. So anisotropy alone doesn't give us *direction.*

Similarly, it seems, for any other physical time-asymmetry to which we might appeal. If time did have a direction – an objective basis for a privileged notion of positive or future time – then for any physical arrow or asymmetry in time, there would always be a question as to whether that arrow pointed forwards or backwards. And so no physical fact could answer this question, because for any candidate, the same issue arises all over again. Thus the idea that time has a real direction seems without any physical meaning. (Of course, we can use any asymmetry we like as a basis for a conventional labelling – saying, for example, that we'll regard the direction in which entropy is increasing as the positive direction of time. But this is different from discovering some intrinsic directionality to time itself.)

As we shall see later, Eddington wrestled with these issues. They go a long way to explain why he felt tempted by the view that the true source of the direction of time – the fact that it "goes on", as he put it – is something accessible in consciousness but not in physical instruments. But for present purposes, since our immediate focus is the thermodynamic asymmetry, I shall assume that it is a conventional matter which direction we treat as positive or future time. Moreover, although it makes sense to ask whether time is anisotropic, it seems clear that this is a different issue from that of the thermodynamic asymmetry. As noted, the thermodynamic asymmetry is an asymmetry of physical processes *in* time, not an asymmetry of time itself.

5.3. Entropy gradient, not entropy increase

If it is conventional which direction counts as positive time, then it is also conventional whether entropy increases or decreases. It increases by the lights of the usual convention, but decreases if we reverse the labelling. But this may seem ridiculous. Doesn't it imply, absurdly, that the thermodynamic asymmetry is merely conventional?

No. The crucial point is that while it is a conventional matter whether the entropy gradient slopes up or down, the gradient itself is objective. The puzzling asymmetry is that the gradient is monotonic – it slopes in the same direction everywhere (so far as we know).

It is worth noting that in principle there are two possible ways of contrasting this monotonic gradient with a symmetric world. One contrast would be with a world in which there are entropy gradients, but sometimes in one direction and sometimes in the other – i.e., worlds in which entropy sometimes goes up and sometimes goes down. The other contrast would be with worlds in which there are no significant gradients, because entropy is always high. If we manage to explain the asymmetric gradient we find in our world, we'll be explaining why the world is not symmetric in one of these ways – but which one? The answer is not obvious in advance, but hopefully will fall out of a deeper understanding of the nature of the problem.

5.4. The term 'entropy' is inessential

A lot of time and ink has been devoted to the question how entropy should be defined, or whether it can be defined at all in certain cases (e.g., for the universe as a whole). It would be easy to get the impression that the puzzle of the thermodynamic asymmetry depends on all this discussion – that whether there's really a puzzle depends on how, and whether, entropy can be defined, perhaps.

But in one important sense, these issues are beside the point. We can see that there's a puzzle, and go a long way towards saying what it is, without ever mentioning entropy. We simply need to describe in other terms some of the many processes which show the asymmetry – which occur with one temporal orientation but not the other. For example, we can point out that there are lots of cases of big difference in temperatures spontaneously equalising, but none of big differences in temperature spontaneously arising. Or we can point out that there are lots of cases of pressurised gases spontaneously leaving a bottle, but none of gas spontaneously pressurising by entering a bottle. And so on.

In the end, we may need the notion of entropy to generalise properly over these cases. However, we don't need it to see that there's a puzzle – to see that there's a striking imbalance in nature between systems with one orientation and systems with the reverse orientation. For present purposes, then, we can ignore objections based on problems in defining entropy. (Having said that, of course, we can go on using the term entropy with a clear conscience, without worrying about how it is defined. In what follows, talk of entropy increase is just a placeholder for a list of the actual phenomena which display the asymmetry we're interested in.)

5.5. Summary

For the remainder of our discussion of the *thermodynamic* 'arrow of time', I take it (i) that the asymmetry in nature is a matter of numerical imbalance between temporal mirror images, not of literal reversibility; (ii) that we are concerned with an asymmetry of physical processes in time, not with an asymmetry in time itself; (iii) that the objective asymmetry concerned is a monotonic *gradient,* rather than an increase or a decrease; and (iv) that if need be the term ‘entropy’ is to be thought of as a placeholder for the relevant properties of a list of actual physical asymmetries.

6. What would a solution look like? Two models

With our target more clearly in view, I now want to call attention to what may be the most useful distinction of all, in making sense of the many things that physicists and philosophers say about the thermodynamic asymmetry. This is a distinction between two very different conceptions of *what it would take* to explain the asymmetry – so different, in fact, that they disagree on *how many* distinct violations of T-symmetry it takes to explain the observed asymmetry. On one conception, an explanation needs two T-asymmetries. On the other conception, it needs only one.

Despite this deep difference of opinion about what a solution would look like, the distinction between these two approaches is hardly ever noted in the literature – even by philosophers, who are supposed to have a nose for these things. So it is easy for advocates of the different approaches to fail to see that they are talking at cross-purposes – that in one important sense, they disagree about what the problem is.

6.1. The two-asymmetry approach

Many approaches to the thermodynamic asymmetry look for a dynamical explanation of the second law – a dynamical cause or factor, responsible for entropy increase. Here are some examples, old and new:

1. *The H-theorem.* Oldest and most famous of all, this is Boltzmann’s development of Maxwell’s idea that intermolecular collisions drive gases towards equilibrium.
2. *Interventionism.* This alternative to the H-theorem, apparently first proposed by S.H. Burbury in the 1890s, [6, 7] attributes entropy increase to the effects of random and uncontrollable influences from a system’s external environment.
3. *Indeterministic dynamics.* There are various attempts to show how an indeterministic dynamics might account for the second law. A recent example is a proposal that the stochastic collapse mechanism of the GRW approach to quantum theory might also explain entropy increase. [1, 2]

I stress two points about these approaches. First, if there is something dynamical which makes entropy increase, then it needs to be time-asymmetric. Why? Because otherwise it would force entropy to increase (or at least not to decrease) in both directions – in other words, entropy would be constant. In the H-theorem, for example, this asymmetry resides in the assumption of molecular chaos. In interventionism, it is provided by the assumption that incoming influences from the environment are ‘random’, or uncorrelated with the system’s internal dynamical variables.

The second point to be stressed is that this asymmetry alone is not sufficient to produce the observed thermodynamic phenomena. Something which forces entropy to be non-decreasing won't produce an entropy gradient unless entropy starts low. To give us the observed gradient, in other words, this approach also needs a low entropy boundary condition – entropy has to be low in the past. This condition, too, is time-asymmetric, and it is a separate condition from the dynamical asymmetry. (It is not guaranteed by the assumption of molecular chaos, for example.)

So this approach is committed to the claim that it takes *two* T-asymmetries – one in the dynamics, and one in the boundary conditions – to explain the observed asymmetry of thermodynamic phenomena. If this model is correct, explanation of the observed asymmetry needs an explanation of both contributing asymmetries, and the puzzle of the thermodynamic arrow has become a double puzzle.

6.2. The one-asymmetry model

The two-asymmetry model is not the only model on offer, however. The main alternative was first proposed by Boltzmann in the 1870s,[3] in response to Loschmidt's famous criticism of the H-theorem. To illustrate the new approach, think of a large collection of gas molecules, isolated in a box with elastic walls. If the motion of the molecules is governed by deterministic laws, such as Newtonian mechanics, a specification of the microstate of the system at any one time uniquely determines its entire trajectory. The key idea of Boltzmann's new approach is that in the overwhelming majority of possible trajectories, the system spends the overwhelming majority of the time in a high entropy macrostate – among other things, a state in which the gas is dispersed throughout the container. (Part of Boltzmann's achievement was to find the appropriate way of counting possibilities, which we can call the *Boltzmann measure.*)

Importantly, there is no temporal bias in this set of possible trajectories. Each possible trajectory is matched by its time-reversed twin, just as Loschmidt had pointed out, and the Boltzmann measure respects this symmetry. Asymmetry arises only when we apply a low entropy condition at one end. For example, suppose we stipulate that the gas is confined to some small region at the initial time t_0. Restricted to the remaining trajectories, the Boltzmann measure now provides a measure of the likelihood of the various possibilities consistent with this boundary condition. Almost all trajectories in this remaining set will be such that the gas disperses after t_0. The observed behaviour is thus predicted by the time-symmetric measure, once we conditionalise on the low entropy condition at t_0.

On this view, then, there's no time-asymmetric factor which causes entropy to increase. This is simply the most likely thing to happen, given the combination of the time-symmetric Boltzmann probabilities and the single low entropy restriction in the past. More below on the nature and origins of this low entropy boundary condition. For the moment, the important thing is that although it is time-asymmetric, so far as we know, this is the only time-asymmetry in play, according to Boltzmann's statistical approach. There's no need for a second asymmetry in the dynamics.

7. Which is the right model?

It is important to distinguish these two models, but it would be even more useful to know which of them is right. How many time-asymmetries should we be looking for, in trying to account for the thermodynamic asymmetry? This is a big topic, but I'll mention two factors, both of which seem to me to count in favour of the one-asymmetry model.

The first factor is simplicity, or theoretical economy. If the one-asymmetry approach works, it simply does more with less. In particular, it leaves us with only one time-asymmetry to explain. True, this would not be persuasive if the two-asymmetry approach actually achieved more than the one-asymmetry approach – if the former had some big theoretical advantage that the latter lacked. But the second argument I want to mention suggests that this can't be the case. On the contrary, the second asymmetry seems redundant.

Redundancy is a strong charge, but consider the facts. The two-asymmetry approach tries to identify some dynamical factor (collisions, or external influences, or whatever) that causes entropy to increase – that makes a pressurised gas leave a bottle, for example. However, to claim that one of these factors *causes* the gas to disperse is to make the following 'counterfactual' claim: *If the factor were absent, the gas would not disperse* (or would do so at a different rate, perhaps). But how could the absence of collisions or external influences *prevent* the gas molecules from leaving the bottle?

Here's a way to make this more precise. In the terminology of Boltzmann's statistical approach, we can distinguish between *normal* initial microstates (for a system, or for the universe as a whole), which lead to entropy increases much as we observe, and *abnormal* microstates, which are such that something else happens. The statistical approach rests on the fact that normal microstates are vastly more likely than abnormal microstates, according to the Boltzmann measure.

In these terms, the above point goes as follows. The two-asymmetry approach is committed to the claim that the universe begins in an abnormal microstate. Why? Because in the case of normal initial microstates, entropy increases anyway, without the mechanism in question – so the required counterfactual claim is not true.

It is hard to see what could justify this claim about the initial microstate. At a more local level, why should we think that the initial microstate of a gas sample in an open bottle is normally such that if it weren't for collisions (or external influences, or whatever), the molecules simply wouldn't encounter the open top of the bottle, and hence disperse?

Thus it is doubtful whether there is really any need for a dynamical asymmetry, and the one-asymmetry model seems to offer the better conception of what it would take to solve the puzzle of the thermodynamic asymmetry. But if so, then the various two-asymmetry approaches – including Boltzmann's own H-theorem, which he himself defended in the 1890s, long after he first proposed the statistical approach – are looking for a solution to the puzzle in the wrong place, at least in part.

For present purposes, the main conclusion I want to emphasise is that we need to make a choice. The one-asymmetry model and the two-asymmetry model represent are two very different views of *what it would take* to explain the thermodynamic arrow – of what the problem is, in effect. Unless we notice that they are different approaches, and proceed to agree on which of them we ought to adopt, we can't possibly agree on whether the old puzzle has been laid to rest.

8. The Boltzmann-Schuetz hypothesis – a no-asymmetry solution?

If the one-asymmetry view is correct, the puzzle of the thermodynamic arrow is really the puzzle of the low entropy boundary condition. Why is entropy so low in the past? After all, in making it unmysterious why entropy doesn't decrease in one direction, the Boltzmann measure equally makes it mysterious why it does decrease in the other – for the statistics themselves are time-symmetric.

Boltzmann himself was one of the first to see the importance of this issue. In a letter to *Nature* in 1895, he suggests an explanation, based on an idea he attributes to 'my old assistant, Dr Schuetz'. [4] He notes that although low entropy states are very unlikely, they are very likely to occur eventually, given enough time. If the universe is very old, it will have had time to produce the kind of low entropy region we find ourselves inhabiting simply by accident. 'Assuming the universe great enough, the probability that such a small part of it as our world should be in its present state, is no longer small,' as Boltzmann puts it.

It is one thing to explain why the universe contains regions like ours, another to explain why we find ourselves in such a region. If they are so rare, is not it more likely that we'd find ourselves somewhere else? But Boltzmann suggests an answer to this, too. Suppose, as seems plausible, that creatures like us couldn't exist in the vast regions of near-equilibrium between such regions of low entropy. Then it is no surprise that we find ourselves in such an unlikely place. As Boltzmann himself puts it, 'the ... H curve would form a representation of what takes place in the universe. The summits of the curve would represent the worlds where visible motion and life exist.'

Figure 2 shows what Boltzmann calls the H curve, except that this diagram plots entropy rather than Boltzmann's quantity H. Entropy is low when H is high, so the summits of Boltzmann's H curve are the troughs of the entropy curve. The universe spends most of its time very close to equilibrium. But occasionally – much more rarely than this diagram actually suggests – a random re-arrangement of matter produces a state of low entropy. As the resulting state returns to equilibrium, there's an entropy slope, such as the one on which we (apparently) find ourselves, at a point such as A.

Why do we find ourselves on an uphill rather than a downhill slope, as at B? In another paper, Boltzmann offers a remarkable proposal to explain this, too. [5]

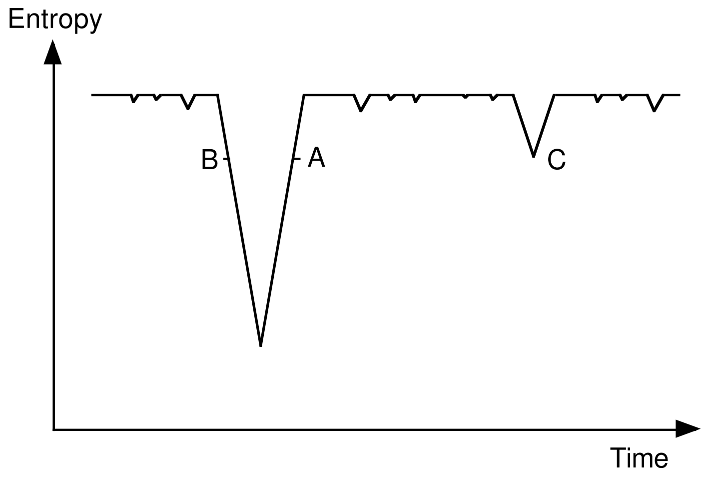


FIGURE 2. Boltzmann's entropy curve.

Perhaps our perception of past and future depends on the entropy gradient, in such a way that we are bound to regard the future as lying 'uphill'. Thus the perceived direction of time would not be objective, but a product of our own orientation in time. Creatures at point B would see the future as lying in the other direction, and there's no objective sense in which they are wrong and we are right, or vice versa. Boltzmann compares this to the discovery that spatial up and down are not absolute directions, the same for all observers everywhere.

We shall return to this aspect of the Boltzmann-Schuetz hypothesis – its elimination of an objective *direction* of time – in a moment, and compare it to Eddington's views on the matter. For moment, however, what matters about the hypothesis is that it offers an explanation of the local asymmetry of thermodynamics in terms which are symmetric on a larger scale. So it is a no-asymmetry solution – the puzzle of the thermodynamic asymmetry simply vanishes on the large scale.

9. The big problem

Unfortunately, however, this clever proposal has a sting in its tail, a sting so serious that it now seems almost impossible to take the hypothesis seriously. The problem flows directly from Boltzmann's own link between entropy and probability. In Figure 1, the vertical axis is a logarithmic probability scale. For every downward increment, dips in the curve of the corresponding depth are exponentially more improbable. So a dip of the depth of point A or point B is much more likely to occur in the form shown at point C – where the given depth is very close to the minimum of the fluctuation – than in association with a much bigger dip, as at A and B. Hence if our own region has a past of even lower entropy, it is much more improbable than it needs to be, given its present entropy. So far, this point seems to have been appreciated already in the 1890s, in exchanges between Boltzmann and Zermelo. What doesn't seem to have appreciated is its devastating consequence, namely, that according to the Boltzmann measure it is much easier to produce fake records and memories, than to produce the real events of which they purport to be records.

Why does this consequence follow? Well, imagine that the universe is vast enough to contain many separate fluctuations, each containing everything that we see around us, including the complete works of Shakespeare, in all their twenty-first century editions. Now imagine choosing one of these fluctuations at random. It is vastly more likely that we'll select a case in which the Shakespearean texts are a product of a spontaneous recent fluctuation, than one in which they were really written four hundred years earlier by a poet called William Shakespeare. Why? Simply because entropy is much higher now than it was in the sixteenth century (as we normally assume that century to have been). Recall that according to Boltzmann, probability increases exponentially with entropy. Fluctuations like our twenty-first century – 'Shakespearian' texts and all – thus occur much more often in typical world-histories than fluctuations like the lower-entropy sixteenth century. So almost all fluctuations including the former don't include the latter. The same goes for the rest of history – all our 'records' and 'memories' are almost certainly misleading.

To make this conclusion vivid we can take advantage of the fact that in the Boltzmann picture, there is not an objective direction of time. So we can equally well think about the question of 'what it takes' to produce what we see around us from the reverse of the normal temporal perspective. Think of starting in what we call the future, and moving in the direction we call towards the past. Think of all the apparently miraculous accidents it takes to produce the kind of world we see around us. Among other things, our bodies themselves, and our editions of Shakespeare, have to 'undecompose', at random, from (what we normally think of as) their future decay products. That's obviously extremely unlikely, but the fact that we're here shows that it happens. But now think of what it takes to get even further back, to a sixteenth century containing Shakespeare himself. The same kind of near-miracle needs to happen many more times. Among other things, there are several billion intervening humans to 'undecompose' spontaneously from dust.

So the Boltzmann-Schuetz hypothesis implies that our apparent historical evidence is almost certainly unreliable. One of the first authors to make this point in print was Carl Friedrich von Weizsäcker, in 1939. [32] Weizsäcker notes that 'improbable states can count as documents [i.e., records of the past] only if we presuppose that still less probable states preceded them.' He concludes that 'the most probable situation by far would be that the present moment represents the entropy minimum, while the past, which we infer from the available documents, is an illusion.'

Weizsäcker also notes that there's another problem of a similar kind. The Boltzmann-Schuetz hypothesis implies that as we look further out into space, we should expect to find no more order than we already have reason to believe in. But we can now observe vastly more of the universe than was possible in Boltzmann's day, and there seems to be low entropy all the way out.

So the Boltzmann-Schuetz hypothesis faces some profound objections. Fortunately, modern cosmology goes at least some way to providing us with an alternative – an option noted with great prescience by Eddington himself, in the 1931 paper to which we have already referred. Anticipating Weizsäcker by several years, Eddington sets out very clearly why we need this alternative, and cannot rely simply on chance fluctuations, as Boltzmann and Schuetz had suggested:

> [I]t is practically certain that a universe containing mathematical physicists will at any assigned date be in the state of maximum disorganization which is not inconsistent with the existence of such creatures. ([15], 452)

In other words, a region of space and time containing some intelligent creatures – mathematical physicists, as Eddington puts it, setting the bar for intelligence a little higher than usual – will almost certainly contain very little of the ordered past those creatures *believe* themselves to have; and very little additional order, when they investigate new realms.

But what is the alternative to relying on chance to produce the low entropy we observe? In Eddington's view, it amounts to this:

> We are thus driven to admit anti-chance; and apparently the best thing we can do with it is to sweep it up into a heap at the beginning of time. ([15], 452)

Eddington does not regard this as a novel suggestion. On the contrary, he regards it as implicit in the physics of "the last three-quarters of a century":

> There is no doubt that the scheme of physics as it has stood for the last three-quarters of a century postulates a date at which either the entities of the universe were created in a state of high organisation, or pre-existing entities were endowed with that organisation which they have been squandering ever since. Moreover, this organisation is admittedly the antithesis of chance. It is something which could not occur fortuitously. ([14], 84)

One of the remarkable developments in recent decades has been that cosmology now tells us quite a lot about this 'heap at the beginning of time'.

10. Initial smoothness

The observed thermodynamic asymmetry requires that entropy was low in the past. Low entropy requires concentrations of energy in usable forms, and presumably there are many ways such concentrations might have existed in the universe.

On the face of it, we seem to have no reason to expect any particularly neat or simple story about how it works in the real world – about where the particular concentrations of energy we depend on happen to originate. Remarkably, however, modern cosmology suggests that all the observed low entropy is associated with a single characteristic of the early universe, soon after the big bang – in other words, a single 'heap' of low entropy at the beginning of time, pretty much as Eddington had proposed.

The crucial thing seems to be that matter is distributed extremely smoothly in the early universe. This provides a vast reservoir of low entropy, on which everything else depends. In particular, smoothness is necessary for galaxy and star formation, and most familiar irreversible phenomena depend on the sun.

Why does a smooth arrangement of matter amount to a low entropy state? Because in a system dominated by an attractive force such as gravity, a uniform distribution of matter is highly unstable (and provides a highly usable supply of potential energy). However, about 10^5 years after the big bang, matter seems to have been distributed smoothly to very high accuracy.

One way to get a sense how surprising this is, is to recall that we've found no reason to disagree with Boltzmann's suggestion that there's no objective distinction between past and future – no sense in which things really happen in the direction we think of as past-to-future. Without such a distinction, there's no objective sense in which the big bang is not equally the end point of a gravitational collapse. Somehow that collapse is coordinated with astounding accuracy, so that the matter involved manages to avoid forming large agglomerations (in fact, black holes), and instead spreads itself out very evenly across the universe. (By calculating the entropy of black holes with comparable mass, Penrose [22] has estimated the odds of such a smooth arrangement of matter at 1 in $10^{10^{123}}$.)

In my view, this discovery about the cosmological origins of low entropy is one of the great achievements of twentieth century physics. It is a remarkable discovery in two quite distinct ways, in fact. First, it is the only anomaly necessary to account for the low entropy we find in the universe, at least so far as we know. So it is a remarkable theoretical achievement – it wraps up the entire puzzle of the thermodynamic asymmetry into a single package, in effect. Second, it is astounding that it happens at all, according to existing theories of how gravitating matter should behave (which suggests, surely, that there is something very important missing from those theories).[2]

[2] True, it is easy to fail to see how astounding the smooth early universe is, by failing to see that the big bang can quite properly be regarded as the end point of a gravitational collapse. But anyone inclined to deny the validity of this way of viewing the big bang faces a perhaps even more daunting challenge: to explain what is meant by, and what is the evidence for, the claim that time has an objective direction!

11. Open questions

Why is the universe smooth soon after the big bang? This is a major puzzle, but – if we accept that the one-asymmetry model – it is the only question we need to answer, to solve the puzzle of the thermodynamic arrow. So we have an answer to the question with which we began. What has happened to the puzzle noticed by those nineteenth century giants? It has been transformed by some of their twentieth century successors into a puzzle for cosmology, a puzzle about the early universe.

It is far from clear how this remaining cosmological puzzle is to be explained. Indeed, there are some who doubt whether it needs explaining. [9, 10, 30] But these issues are beyond the scope of this paper. I want to close by calling attention to some open questions associated with this understanding of the origins of the thermodynamic asymmetry, and by making a case for an unusually sceptical attitude to the second law.

One fascinating question is whether whatever explains why the universe is smooth after the big bang would also imply that the universe would be smooth before the big crunch, if the universe eventually recollapses. In other words, would entropy would eventually decrease, in a recollapsing universe?[3] It is often dismissed on the grounds that a smooth recollapse would require an incredibly unlikely 'conspiracy' among the components parts of the universe, to ensure that the recollapsing matter did not clump into black holes. However, as we have already noted, this incredible conspiracy is precisely what happens towards (what we usually term) the big bang, if we regard that end of the universe as a product of a gravitational collapse. The statistics themselves are time-symmetric. If something overrides them at one end of the universe, what right do we have to assume that the same does not happen at the other? Until we understand more about the origins of the smooth early universe, it seems best to keep an open mind about a smooth late universe.

Another interesting and open question is whether a future low entropy boundary condition would have effects *now.* Events at the present era provide us with evidence of a low entropy past. Could there also be evidence of a low entropy future? The answer depends on our temporal distance from such a future boundary condition, in relation to the relaxation time of cosmological processes. It has been argued that a symmetric time-reversing universe would require more radiation in the present era than we actually observe – radiation which in the reversed time sense originates in the stars and galaxies of the opposite end of the universe. [17] But because of its anti-thermodynamic character, from our point of view, it is doubtful whether this radiation would be detectable, at least by standard means [23].

Some people dismiss the question whether entropy would reverse in a recollapsing universe on the grounds that the current evidence suggests that the

[3]This possibility is often called the "Gold universe", though as Larry Schulman pointed out to me, the attribution of this proposal to Gold [18], e.g., by Davies [13] as well as by me [23], seems to depend on an extrapolation from Gold's own views.

universe will not recollapse. However, it seems reasonable to expect that when we find out why the universe is smooth near the big bang, we'll be able to ask a theoretical question about what that reason would imply in the case of universe which did recollapse. Moreover, as a number of writers [20, 21] have pointed out, much the same question arises if just a bit of the universe recollapses – e.g., a galaxy, collapsing into a black hole. This process seems to be a miniature version of the gravitational collapse of a whole universe, and so it makes sense to ask whether whatever constrains the big bang also constrains such partial collapses.

12. Scepticism about the second law

In my view, the moral of these considerations is that until we know more about why entropy is low in the past, it is sensible to keep an open mind about whether it might be low in the future. The appropriate attitude is a kind of healthy scepticism about the universality of the second law of thermodynamics.

The case for scepticism goes like this. What we've learnt about why entropy increases in our region is that it does so because it is very low in the past (for some reason we don't yet know), and the increase we observe is the most likely outcome consistent with that restriction. As noted, however, the statistics underpinning this reasoning are time-symmetric, and hence the predictions we make about the future depend implicitly on the assumption that there is no corresponding low entropy boundary condition in that direction. Thus the Boltzmann probabilities don't enable us to predict without qualification that entropy is unlikely to decrease, but only that it is unlikely to decrease, *unless there is the kind of boundary condition in the future that makes entropy low in the past.* In other words, the second law is likely to continue to hold so long as there is not a low entropy boundary condition in the future. But it can't be used to exclude this possibility – even probabilistically!

Sceptics about the second law are unusual in the history of thermodynamics, and I would like to finish by giving some long-overdue credit to one of the rare exceptions. Samuel Hawksley Burbury (1831–1911) was not one of the true giants of thermodynamics. However, he made an important contribution to the identification of the puzzle of the time-asymmetry of thermodynamic phenomena. And he was more insightful than any of his contemporaries – and most writers since, for that matter – in being commendably cautious about declaring the puzzle solved.

Burbury was an English barrister. He read mathematics at Cambridge as an undergraduate, but his major work in mathematical physics came late in life, when deafness curtailed his career at the Bar. In his sixties and seventies, he thus played an important in discussions about the nature and origins of the second law. In a review of Burbury's monograph *The Kinetic Theory of Gases* for *Science* in 1899, the reviewer describes his contribution as follows:

> [I]n that very interesting discussion of the Kinetic Theory which was begun at the Oxford meeting of the British Association in 1894 and

> continued for months afterwards in *Nature,* Mr. Burbury took a conspicuous part, appearing as the expounder and defender of Boltzmann's H-theorem in answer to the question which so many [had] asked in secret, and which Mr. Culverwell asked in print, '*What is the H-theorem and what does it prove?*' Thanks to this discussion, and to the more recent publication of Boltzmann's *Vorlesungen über Gas-theorie,* and finally to this treatise by Burbury, the question is not so difficult to answer as it was a few years ago. [19]

It is a little misleading to call Burbury a defender of the H-theorem. The crucial issue in the debate referred to here was the source of the time-asymmetry of the H-theorem, and while Burbury was the first to put his finger on the role of assumption of molecular chaos, he himself regarded this assumption with considerable suspicion. Here's how he puts it in 1904:

> Does not the theory of a general tendency of entropy to diminish[4] take too much for granted? To a certain extent it is supported by experimental evidence. We must accept such evidence as far as it goes and no further. We have no right to supplement it by a large draft of the scientific imagination.[8]

Burbury's reasons for scepticism are not precisely those which seem appropriate today. Burbury's concern might be put like this. To see that the dynamical processes routinely fail to produce entropy increases towards the past is to see that it takes an extra ingredient to ensure that they do so towards the future. We're then surely right to wonder whether that extra ingredient is sufficiently universal, even towards the future, to guarantee that the second law will always hold. As the first clearly to identify the source of the time-asymmetry in the H-theorem, Burbury was perhaps more sensitive to this concern than any of his contemporaries.

At the same time, however, Burbury seems never to have distanced himself sufficiently from the H-theorem to see that the real puzzle of the thermodynamic asymmetry lies elsewhere. The interesting question is not whether there is a good dynamical argument to show that entropy will always increase towards the future. It is why entropy steadily *decreases* towards the past – in the face, note, of such things as the effects of collisions and external influences, which are 'happening' in that direction as much as in the other! As we've seen, this re-orientation provides a new reason for being cautious about proclaiming the universal validity of the second law. Once we regard the fact that entropy decreases towards the past as itself a puzzle, as something in need of explanation, then it ought to occur to us that whatever explains it might be non-unique – and thus that in principle, there might be a low entropy boundary condition in the future, as well as in the past.

It is interesting to compare Burbury's scepticism about the second law to Eddington's view that it holds "the supreme position among the laws of Nature":

[4]Burbury is apparently referring to Boltzmann's quantity H, which does decrease as entropy increases.

> If someone points out to you that your pet theory of the universe is in disagreement with Maxwell's equations – then so much the worse for Maxwell's equations. If it is found to be contradicted by observation – well, these experimentalists do bungle things sometimes. But if your theory is found to be against the second law of thermodynamics I can give you no hope; there is nothing for it but to collapse in deepest humiliation. This exaltation of the second law is not unreasonable. There are other laws which we have strong reason to believe in, and we feel that a hypothesis which violates them is highly improbable; but the improbability is vague and does not confront us as a paralysing array of figures, whereas the chance against a breach of the second law (i.e., against a decrease of the random element) can be stated in figures which are overwhelming. ([14], 74–75)

The appropriate reply to Eddington, in my view, is that he himself accepts that there is "anti-chance" in play in the universe, needed to explain the low entropy past. If someone's "pet theory of the universe" proposes violations of the second law because it proposes that there might be anti-chance in the future, as well as the past, then Eddington's appeal to overwhelming probabilities simply misses the point. We can't appeal to probability to refute the hypothesis that the relevant probabilities are not reliable everywhere – that simply begs the question against the hypothesis.

So Burbury has the better of this argument, in my view. Eddington falls into a trap that has snared many later thinkers. What makes the trap so tempting is the lure of a temporal *double standard* – a temptation to reason differently about the past and about the future. But that's the very temptation we need to resist, if we want to understand where the differences in question come from.

13. Where now for the flow of time?

To finish, I want to return to the question as to whether time has some other arrow, not captured by the thermodynamic asymmetry. This issue has close connections with one of the oldest debates in philosophy. On one side of the debate are philosophers who think of time just as we experience it, a live process of flow and change. ("All is flux", as Heraclitus of Ephesus put it, around 500BC.) On this view, time really *passes* (or "goes on", as Eddington put it). And the present moment seems to have a special status, as the moving boundary between fixed past and open future. (This is deeply connected to our sense of freedom. Our actions, freely chosen, seem to play a small part in bringing about the future.)

On the other side of the debate are philosophers who think of time the way we describe it in history, as a fixed and unchanging series of events, lined up in a particular order – laid out as in a map, as Eddington says. Here, the original credit often goes to Parmenides of Elea, another of the early Greek philosophers, who argued that existence is uniform and timeless, and that change is impossible. In recent work, this view is often called the Block Universe picture.

Which is the right view of time, the 'dynamic' Heraclitan view, or the 'static' Block Universe? This is still a live issue in philosophy, and Heraclitus and Parmenides both have their contemporary champions. However, the two sides sometimes seem to be talking past one another – and in part, I think, this is because they don't pay enough attention to precisely what the issues are.

13.1. Three ingredients in 'temporal passage'

In particular, I think it is helpful to distinguish three distinct elements in dynamic conception. These elements tend to be bundled together, but in principle they can be separated, and defended in almost any combination. (We already encountered two of them in Section 2, when we considered what Eddington has in mind when he discusses 'time's arrow'.)

1. The view that the *present moment* is objectively distinguished, and that reality is objectively divided into past, present and future.
2. The view that time has an objective *direction* – that it is an objective matter which of two non-simultaneous events is the *earlier* and which the *later.*
3. The view that there is something objectively "flow-like" about time – that time really "goes on", as Eddington puts it.

Philosophers who defend the Block Universe picture tend to be steadfast in their rejection of (1) and (3), and a little bit more open-minded about (2). In physics, however, it easy to find famous critics of all three elements.

About (1), for example, there is a well-known remark that Einstein makes in a letter to the bereaved family of his old friend, Michele Besso (a few months before his own death). Einstein offers the consoling thought that past, present and future are all equally real – only from our human perspective does the present seem special, and the past seem lost: "We physicists know that the distinction between past, present and future is only a stubbornly persistent illusion," as he puts it. (For this, Karl Popper called him the Parmenides of modern physics.)

Concerning (2), we have already encountered Boltzmann's proposal: "For the universe, the two directions of time are indistinguishable, just as in space there is no up and down." While for (3), we noted a famous remark by Hermann Weyl, which rejects the idea of an objective flow of time:

> The objective world simply is, it does not happen. Only to the gaze of my consciousness, crawling upward along the world-line of my body, does a section of the world come to life as a fleeting image in space which continuously changes in time.

The Block Universe seems to be the majority view among contemporary physicists, though it is easy to find dissenters: for example, physicists who argue that the fact that modern physics *seems* to favour the Block conception is an indication that something is missing from physics.[5]

13.2. Eddington on 'becoming'

As we have seen, Eddington wanted to be a dissenter, too, though he struggles with question as to what the missing ingredient – "becoming" – could actually be, and what reason we really have for thinking that it is something objective. His discussion of these matters deserves a much more thorough examination than I can give it here, but I want to highlight two features of it: first, the role that he takes consciousness to play, in providing a reason to think of becoming as something real, and second a challenge he issues to his opponents.

Eddington begins his chapter on 'Becoming' with a wonderfully characteristic[6] presentation of the view he wants to oppose:

> When you say to yourself, "Every day I grow better and better", science churlishly replies –
>
> "I see no signs of it. I see you extended as a four-dimensional worm in space-time; and, although goodness is not strictly within my province, I will grant that one end of you is better than the other. But whether you grow better or worse depends on which way up I hold you. There is in your consciousness an idea of growth or 'becoming' which, if it is not illusory, implies that you have a label 'This side up'. I have searched for such a label all through the physical world and can find no trace of it, so I strongly suspect that the label is non-existent in the world of reality."
>
> ... Taking account of [entropy], the reply is modified a little, though it is still none too gracious –
>
> "I have looked again and, in the course of studying a property called entropy, I find that the physical world is marked with an arrow which may possibly be intended to indicate which way up it should be regarded. With that orientation I find that you really do grow better. Or, to speak precisely, your good end is in the part of the world with most entropy and your bad end in the part with least. Why this arrangement should be considered more creditable than that of your neighbour who has his good and bad ends the other way round, I cannot imagine." ([14], 87)

In response to this challenge, Eddington goes on to propose that in the case of becoming, consciousness gives us an insight in the nature of reality which physics otherwise misses. These two passages give some of the flavour of the viewpoint:

> Unless we have been altogether misreading the significance of the world outside us – by interpreting it in terms of evolution and progress, instead

[5]I have heard this view expressed by Lee Smolin, George Ellis, Chris Fuchs and David Mermin, for example.

[6]That is, witty and clear!

> of a static extension – we must regard the feeling of 'becoming' as ... a true mental insight into the physical condition which determines it. ([14], 89)
>
> We have direct insight into "becoming" which sweeps aside all symbolic knowledge as on an inferior plane. If I grasp the notion of existence because I myself exist, I grasp the notion of becoming because I myself become. It is the innermost Ego of all which *is* and *becomes.* ([14], 97)

As he puts it in another passage:

> The view here advocated is tantamount to an admission that consciousness, *looking out through a private door,* can learn by direct insight an underlying character of the world which physical measurements do not betray. ([14], 91, my emphasis)

Eddington appreciates, of course, that this is not an easy position for a physicist to accept: "The physicist ... does not look kindly on private doors, through which all forms of superstitious fancy might enter unchecked." But he stresses the alternative:

> But is he [i.e., the physicist who renounces private doors] ready to forgo that knowledge of the going on of time which has reached us through the door, and content himself with the time inferred from sense-impressions which is emaciated of all dynamic quality? ([14], 91)

And at this point, backing up this rhetorical question, he issues what I want to call *Eddington's Challenge:*

> No doubt some will reply that they are content; to these I would say – Then *show your good faith by reversing the dynamic quality of time* (which you may freely do if it has no importance in Nature), and, just for a change, give us a picture of the universe passing from the more random to the less random state ... If you are an astronomer, tell how waves of light hurry in from the depths of space and condense on to stars; how the complex solar system unwinds itself into the evenness of a nebula. ... If you genuinely believe that a contra-evolutionary theory is just as true and as significant as an evolutionary theory, *surely it is time that a protest should be made against the entirely one-sided version currently taught.*" ([14], 91–92, my emphasis)

In my view, Eddington is wrong about becoming. I side with Einstein, Boltzmann and Weyl, in rejecting all three elements of the view that time really "goes on". Nevertheless, I think that Eddington's Challenge deserves to be better known. In thinking about how to meet it, we may learn a lot about the consequences of the revolution that has taken place in our understanding of time, since the late nineteenth century. I suspect that friends of the Block Universe have not done enough to free themselves from the shackles of the old viewpoint; and in the long run, the best arguments for the Block view might flow from a recognition of the advantages of thinking about time in the revolutionary way.

13.3. Meeting Eddington's Challenge

The first response to Eddington's Challenge should be to appeal to Boltzmann, I think. The Boltzmann of the Boltzmann-Schuetz hypothesis is well ahead of Eddington, in offering us a picture in which the entropy gradient is a local matter in the universe as a whole, entirely absent in most eras and regions (and with no single preferred direction in those rare locations in which it is to be found). Combined with Eddington's own view that the asymmetries he challenges his opponent to consider reversing – asymmetries of inference and explanation, for example – have their origin in the entropy gradient, this means that Boltzmann has an immediate answer to the Challenge. Of course we can't "[reverse] the dynamic quality of time" *around here,* for we live within the constraints of the entropy gradient in the region in which we are born. But we can tell you, in principle, how to find a region in the picture is properly reversed; and that shows that the fixity of our own perspective does not reflect a fundamental asymmetry in nature. Analogously (Boltzmann might add), the fact that people in Northern Europe cannot live with their feet pointed to the Pole Star does not prove a spatial anisotropy. If you want to live with your feet pointing that way, you simply need to move elsewhere.

Eddington himself associates the entropy gradient quite closely with the "time of consciousness":

> It seems to me, therefore, that consciousness with its insistence on time's arrow and its rather erratic ideas of time measurement may be guided by entropy-clocks in some portion of the brain. ... Entropy-gradient is then the direct equivalent of the time of consciousness in both its aspects. ([14], 101)

So the Boltzmann-Schuetz hypothesis certainly threatens the veracity of Eddington's "private door".

In broader terms, however, Eddington's Challenge has not been taken up. Most advocates of the Block view – even those explicit about the possibility that time might have no intrinsic direction – have not explored the question as to what insights might follow from Boltzmann's 'Copernican' shift in our perspective. I want to conclude with some brief remarks on this issue. It seems to me that there are at least two domains in physics in which we might hope to vindicate Boltzmann's viewpoint, by exhibiting the advantages of the atemporal perspective it embodies.

13.4. Eddington's Challenge in cosmology

The first domain is cosmology. There are two aspects to the relevance of Boltzmann's viewpoint in this context. First, and closest to Boltzmann's own concerns, there is the project of understanding the origin of the entropy gradient, in our region. As I have already noted, one of the great advances in physics over recent decades has been the realisation that this problem seems to turn on the question as to why gravitational entropy was low, early in the history of the known universe – in particular, why matter was smoothly distributed, to a very high degree, approximately 100,000 years after the Big Bang. As we try to explain this feature of the early stages of the known universe, Boltzmann's hypothesis ought to alert us to the possibility that it is non-unique – ought to open our eyes to a new range of cosmological models, in which there is no single unique entropy gradient.

There is some recent work which takes this possibility seriously – see, e.g., [12] (from which Figure 3 is borrowed) and [11]. However, there is much more contemporary work in which it is either overlooked, or dismissed for what, with Boltzmann's symmetric viewpoint clearly in mind, can be seen to be fallacious reasons. For example, the possibility that entropy might decrease 'towards the future' is dismissed on statistical grounds, with no attempt to explain why this is a good argument towards the future, despite the fact that (i) it is manifestly a bad argument towards the past, and (ii) that the relevant statistical considerations are time-symmetric. (This is the 'double standard' of which I accused Eddington at the end of Section 12. See [23] for a discussion of other cases of this fallacy, and the role of the timeless viewpoint in avoiding them.)

These considerations point in the direction of the second and broader aspect of the relevance of Boltzmann's atemporal viewpoint in cosmology. It alerts us to the possibility that the usual model of 'explanation-in-terms-of-initial-conditions' might simply be the wrong one to use in the cosmological context, where the features in need of explanation are larger and more inclusive than anything we encounter in the familiar region of our 'home' entropy gradient. Here, the point connects directly with Eddington's Challenge, in the way noted above. We can concede our local practices of inference and explanation are properly time-asymmetric, as Eddington observes; while insisting that symmetry might prevail on a larger scale.

13.5. Eddington's Challenge in microphysics

Even more interestingly, in my view, there is the possibility that the 'pre-Copernican' viewpoint might be standing in the way of progress needed in fundamental physics – that is, that there might be explanations to which this viewpoint is at least a major obstacle, if not an impenetrable barrier. Here the most interesting candidate, in my view, is the project of realist interpretations and extensions of quantum mechanics. Discussions of hidden variable models normally take for granted that in any reasonable model, hidden states will be independent of future interactions to which the system in question might be subject. The spin of an electron will not depend on what spin measurements it might be subject to in the future, for example. Obviously, no one expects the same to be true in reverse. On the contrary, we take for granted that the state of the electron may depend on what has happened to it in the past. But how is this asymmetry to be justified, if the gross familiar asymmetries of inference, influence and explanation are to be associated with the entropy gradient, and this is a local matter? Are electrons subject to different laws in one region of the universe than in another, or "aware" of the prevailing entropy gradient in their region? On the contrary, in Boltzmann's picture: we want microphysics to provide the universal background, on top of which the statistical asymmetries are superimposed.

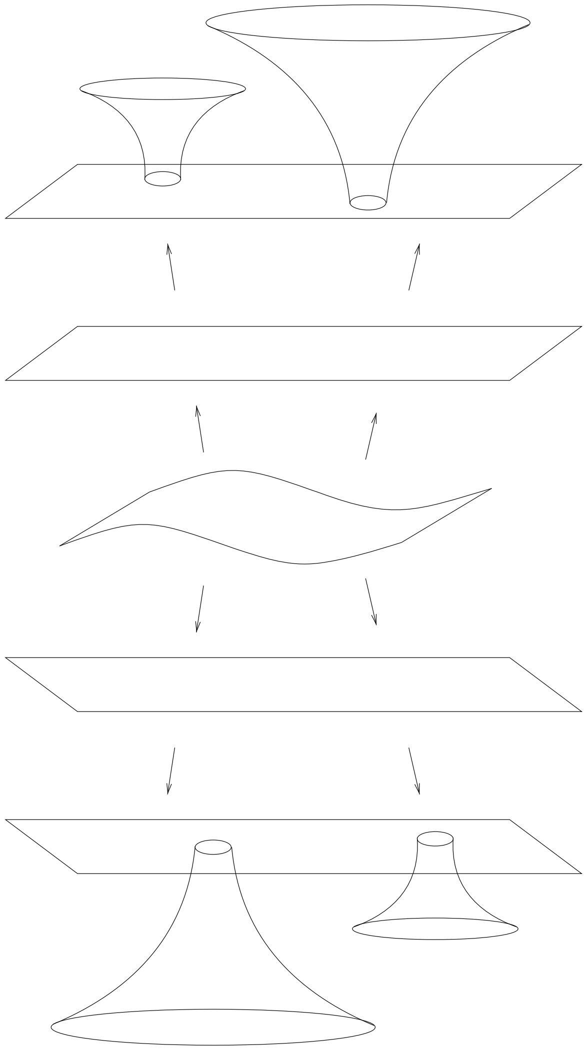


Figure 3. "The ultra-large-scale structure of the universe. Starting from a generic state, it can be evolved both forward and backward in time, as it approaches an empty de Sitter configuration. Eventually, fluctuations lead to the onset of inflation in the far past and far future of the starting slice. The arrow of time is reversed in these two regimes." [12]

When we explore these issues, it might turn out that the apparently puzzling assumption that hidden variables cannot depend on future interactions is just a manifestation of the time-asymmetry of our ordinary causal notions, grounded entirely in the asymmetry of our own viewpoint – in other words, that the assumption is just a kind of perspectival gloss on underlying dynamical principles which are symmetric in themselves. If so, there would be no new *physical* mileage to be gained by adopting the atemporal viewpoint. Certainly, we would understand better what belonged to the physics and what to our viewpoint, but no new physics would be on offer as a result.

However, the more intriguing possibility is that there is a new class of physical models on offer here – models which are being ignored not for any genuinely good reason, but only because they seem to conflict with our ordinary asymmetric perspective. If that's the case – see [16] and [28] for some recent discussion – and if the models presently excluded have the potential they seem to have in accounting for some of the puzzles of quantum mechanics, then Boltzmann's viewpoint will prove to be truly revolutionary; and Eddington's Challenge will be well and truly met.

Acknowledgment

I am grateful to Ikaros Bigi and Martin Faessler for permission to reprint material from [27]. My research is supported by the Australian Research Council and the University of Sydney.

References

[1] Albert, D.Z.: British Journal for the Philosophy of Science **45**, 669 (1994).

[2] Albert, D.Z.: *Time and Chance* (Harvard University Press, Cambridge, 2000).

[3] Boltzmann, L.: Sitzungsberichte, K. Akademie der Wissenschaften in Wien, Math.-Naturwiss **75**, 67 (1877).

[4] Boltzmann, L.: Nature **51**, 413 (1895).

[5] Boltzmann, L.: Wied. Ann. **60**, 392 (1897).

[6] Burbury, S.H.: Nature **51**, 78 (1894).

[7] Burbury, S.H.: Nature **51**, 320 (1895).

[8] Burbury, S.H.: Philosophical Magazine, Series 6, **8**, 43 (1904).

[9] Callender, C.: Metascience **11**, 68 (1997).

[10] Callender, C.: British Journal for the Philosophy of Science **49**, 135 (1998).

[11] Carroll, S.: *From Eternity to Here: The Quest for the Ultimate Theory of Time.* (Dutton, New York, 2009).

[12] Carroll, S., and Chen, J.: arXiv:hep-th/0410270v1.

[13] Davies, P.C.W.: *The Physics of Time Asymmetry.* (Surrey University Press, London, 1974).

[14] Eddington, A.: *The Nature of the Physical World.* (Cambridge University Press, Cambridge, 1928).

[15] Eddington, A.: Nature **127**, 447 (1931).

[16] Evans, P., Price, H., and Wharton, K.B.: British Journal for the Philosophy of Science. Online at doi:10.1093/bjps/axr052 (2012).

[17] Gell-Mann, M., and Hartle, J.: in *Physical Origins of Time Asymmetry,* eds. J. Halliwell, J. Perez-Mercader and W. Zurek (Cambridge University Press, Cambridge, 1994).

[18] Gold, T.: American Journal of Physics **30**, 403 (1962).

[19] Hall, E.H.: Science, New Series **10**, 685 (1899).

[20] Hawking, S.W.: Physical Review D **33**, 2489 (1985).

[21] Penrose, R.: in *General Relativity: an Einstein Centenary,* eds. S.W. Hawking and W. Israel (Cambridge University Press, Cambridge, 1979).

[22] Penrose, R.: *The Emperor's New Mind.* (Oxford University Press, Oxford, 1989).

[23] Price, H.: *Time's Arrow and Archimedes' Point.* (Oxford University Press, New York, 1996).

[24] Price, H.: *British Journal for the Philosophy of Science* **53**, 83 (2002).

[25] Price, H.: in *Time, Reality and Experience.* Ed. C. Callender (Cambridge University Press, Cambridge, 2002).

[26] Price, H.: in *Contemporary Debates in the Philosophy of Science.* Ed. C. Hitchcock (Basil Blackwell, Oxford, 2004).

[27] Price, H.: in *Time and Matter.* Eds. I. Bigi and M. Faessler (World Scientific, Singapore, 2006).

[28] Price, H.: Studies in History and Philosophy of Modern Physics 43, 75 (2012).

[29] Ridderbos, T.M., and Redhead, M.: Foundations of Physics **28**, 1237 (1998).

[30] Sklar, L.: *Physics and Chance.* (Cambridge University Press, Cambridge, 1993).

[31] Sklar, L.: in *Time's Arrows Today.* Ed. S. Savitt (Cambridge University Press, Cambridge, 1995).

[32] von Weizsäcker, C.: Annalen der Physik (5 Folge) **36**, 275 (1939), reprinted in translation in *The Unity of Nature* (Farrar Straus Giroux, New York, 1980).

[33] Weyl, H.: *Philosophy of Mathematics and Natural Science.* (Princeton University Press, Princeton, NJ, 1949).

Huw Price
Centre for Time
Main Quad A14
University of Sydney
NSW 2006, Australia

and

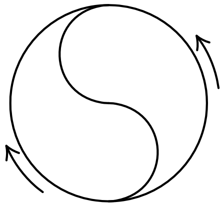
Faculty of Philosophy
University of Cambridge
Sidgwick Avenue
Cambridge CB3 9DA, UK

Time, 217–217

Image of Time's Irreversibility

Catherine de Mitry

An 'S' inscribed in a circle



Imagine that once upon a time, Time was walking in a circle, always turning in the same direction, from right to left through the top, endlessly, happily. It had noticed a fork in the circle, offering another path. A choice: to keep running on its outer cycle, habitual, therefore seemingly natural to it; or to slide into the 'S', unknown, therefore seemingly forbidden to it. So, one day, what happened? While walking along just before the fork, was Time tempted? Was Time distracted? It veered inwards and followed the 'S'. It traversed the circle and went through the center. When it emerged to retake its cycle, it quickly realized that its course had been reversed. It was now turning from left to right through the top, and felt embarrassed to have gone through the looking glass. It tried to get back by returning through the center. That was now impossible for it, as the fork had vanished. There was no longer any place to untake the 'S'; the choice no longer existed. Time was thus condemned to keep turning in its circle, but in that other direction nowadays called: clockwise.

Is that the 'S' of Serpent, Science, or Sapience?

Catherine de Mitry
www.catherinedemitry.com

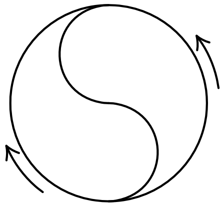
Time, 219–219

Poincaré Seminar 2010

Image de l'irréversibilité du Temps

Catherine de Mitry

Un “S” inscrit dans un cercle



Imaginez qu'il y a longtemps, le Temps marchait le long d'un cercle, tournant toujours dans le même sens, de la droite vers la gauche en passant par le haut, inlassablement, heureusement, en ayant pourtant remarqué, à l'emplacement d'une fourche, une autre voie possible. Il y avait là un choix : ou continuer le cycle en se tenant à l'extérieur, ce qui lui était habituel et lui semblait donc naturel, ou emprunter le “S” en se glissant à l'intérieur, ce qui lui était inconnu et lui semblait donc défendu. Or un jour, que se passa-t-il ? Tandis qu'il cheminait un peu avant la fourche, le Temps eut-il une tentation ? Le Temps eut-il une distraction ? Il se pencha vers l'intérieur et emprunta le “S”. Dès lors, il traversa le cercle et dépassa le centre. Mais, lorsqu'il ressortit et retrouva son cycle, il constata très vite que son cours s'était inversé, puisqu'il tournait maintenant de la gauche vers la droite en passant par le haut, et il fut gêné de voir l'autre côté des choses. Il chercha donc à revenir, en retournant par l'intérieur. Mais cela lui fut impossible, car la fourche avait disparu, il n'y avait plus d'endroit pour reprendre le “S”, le choix n'existait plus. Ainsi le Temps fut condamné à tourner toujours sur son cercle, mais dans cet autre sens que l'on nomme aujourd'hui : le sens des aiguilles d'une montre.

Est-ce le “S” de Serpent, de Science, de Sagesse ?

Catherine de Mitry
www.catherinedemitry.com