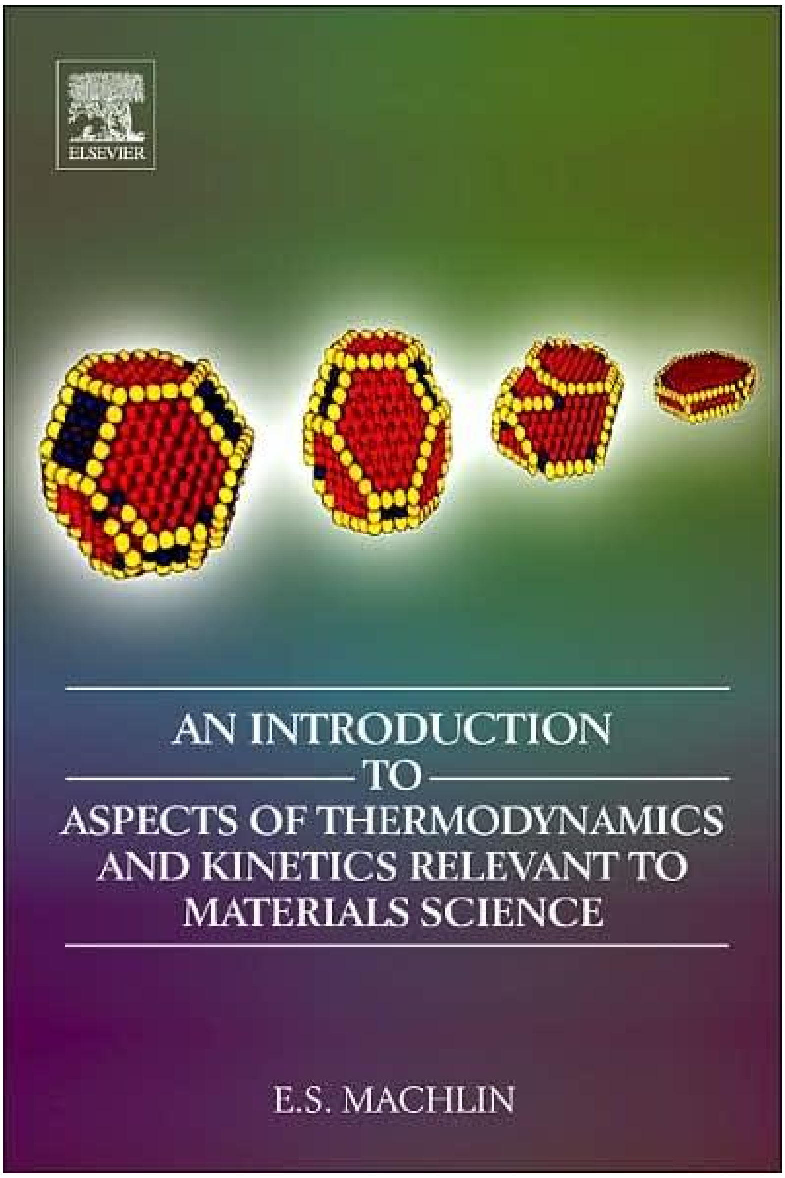
An Introduction to Aspects of Thermodynamics and Kinetics Relevant to Materials Science

by Eugene Machlin



- ISBN: 008046615X
- Publisher: Elsevier Science & Technology Books
- Pub. Date: September 2007

Preface to Third Edition

This book is based on a set of notes developed over a period of many years for an introductory course that the author taught to seniors and entering graduate students in materials science. This course has also been taken by students at IBM Fishkill Facility, who have a varied background, in a program leading to a master's degree in materials science. Our pedagogical philosophy in the teaching of materials science courses has been to take a materials generic approach to each of the core courses in this program. Thus, this book is based on such an approach, which is incidentally consistent with a recommendation in the recent report entitled "Materials Science and Engineering for the 1990's: Maintaining Competitiveness in the Age of Materials" of the Committee on Materials Science and Engineering of the National Research Council. Reference to specific materials, however, will be found in both the text and the problems, inasmuch as each material is unique in its properties.

This book is not a textbook on thermodynamics or kinetics. It assumes that the student has had prior courses in both subjects and is familiar with the essentials of these disciplines that are covered in books, such as, H.B. Callen, Thermodynamics and an Introduction to Thermostatics or B.S. Bokstein, M.I. Mendelev and D.J. Srolovitz, Thermodynamics and Kinetics in Materials Science. This book is about the application of thermodynamics and kinetics to solve problems in Materials Science.

Because this book is on the introductory level, it has not been possible to provide a detailed discussion of each topic covered in depth. Indeed, it is possible to write a book on the subject of each chapter. Thus, in many places detailed derivations of equations used in the text are absent or provided in appendices. Emphasis has been placed upon providing the student with a physical understanding of the phenomenon under discussion, with the mathematics presented as a guide to such an understanding. Thus, the researcher seeking comprehensive knowledge of a field will need to search further. It is hoped that the references and bibliography listed at the end of each chapter will aid in this search.

The problems have been used not only to provide practise in quantitative application of principles, but also to give examples of applications of the general subject matter to problems having current interest and to emphasize the important physical concepts.

The last decade has seen a remarkable advance in computer power. A consequence has been the ability to simulate complex phenomena and to obtain numerical solutions to non-linear problems. Thus, various questions which heretofore were unanswered now have been resolved. For example, we now know that solvability theory yields answers that are consistent with simulation results, even if we have as yet been unable to perform the needed experiments. Further, we are able to solve non-linear differential equations and to calculate the ground state energies of many materials from first principles. Still another change has occurred in the scope of Materials Science that requires a pedagogical response. A knowledge of soft matter, including the matter of molecular biology, has become necessary for the professional functioning of many materials scientists, especially those involved in nanomaterials. Thus, more examples involving soft matter have been included and chapters devoted to the thermodynamics of micelles and pattern formation have been added.

Eugene S. Machlin

Table of Contents

I. Thermodynamics of Phases Having Constant Composition

II. Thermodynamics of Solid Solutions

III. Free Energy and Phase Diagrams

A. Free Energy and Phase Diagrams-Binary Systems

B. Heterogeneous Chemical Equilibria and Phase Diagrams

IV. Thermodynamics of Interfaces

V. Heterophase and Homophase Fluctuations

VI. Thermodynamics of Defects

VII. Concepts in Kinetics

VIII.Diffusion

IX. Nucleation and Growth Kinetics

X. Solid/Solid Interface Migration Kinetics

XI. Growth of Phases - Diffusion or Interface Reaction Control

XII. Morphological Instability and Growth of Phases

XIII.Thermodynamics, Kinetics and Patterns

XIV. Thermodynamics of Micelles

CHAPTER I

Thermodynamics of Phases having Constant Composition

Introduction

Application of thermodynamics and kinetics to phenomena of interest in materials science is the theme of this book. Most materials in the phrase "materials science" consist of phases or mixtures of phases. By "phase" is meant the homogeneous configuration of atoms or molecules corresponding to a liquid; or to some crystalline solid, such as a solid having the body-centered-cubic (bcc) structure; or to an amorphous solid, such as glass; or to a vapor. A phase has a set of properties including thermodynamic potentials, such as entropy, enthalpy and Gibbs free energy. The phase may or may not be in an equilibrium state. In the equilibrium state the thermodynamic properties of a given phase are not functions of its past history, but are unique. A state of equilibrium is defined for a closed system by the condition that the entropy is at a maximum. Related to this condition it is also defined by the existence of a minimum in one of the thermodynamic potentials listed in Table 1.1 along with the corresponding variables that are constrained to be constant in the minimization procedure.

At equilibrium, there are a variety of variational properties of the thermodynamic potentials. For example, the bulk modulus is given by $-V[\partial^2 G/\partial V^2]_T$. A listing of the relations between derivatives of thermodynamic parameters and physical properties of materials is given in Table 1.2. The symbols in the table are defined in the list of symbols at the end of this chapter. From examination of this table it is apparent that a knowledge of the dependence of a thermodynamic potential on pressure or temperature of a material will provide a basis for determining a number of material properties. However, there is a more important reason for

Table 1.1. Thermodynamic potentials corresponding to stable equilibrium.

Thermodynamic potential	**Constrained variables**
Gibbs free energy (G)	P,T,N
Helmholtz free energy (F)	V,T,N
Enthalpy (H)	P,S,N
Internal energy (E)	V,S,N
Grand (Ω)	V,T,μ

Table 1.2. Summary of thermodynamic relations.[a]

X	**Y**	**Z**	$\left(\frac{\partial Y}{\partial X}\right)_Z$	**X**	**Y**	**Z**	$\left(\frac{\partial Y}{\partial X}\right)_Z$
T	V	p	αV	T	p	V	α/β
T	S	p	C_p/T	T	S	V	$C_p/T - \alpha^2 V/\beta$
T	V	p	$C_p - \alpha p V$	T	U	V	$C_p - \alpha^2 VT/\beta$
T	H	p	C_p	T	H	V	$C_p - \alpha^2 VT/\beta + \alpha V/\beta$
T	F	p	$-\alpha p V - S$	T	F	V	$-S$
T	G	p	$-S$	T	G	V	$\alpha V/\beta - S$
p	V	T	$-\beta V$	T	p	S	$C_p/\alpha VT$
p	S	T	$-\alpha V$	T	V	S	$-\beta C_p/\alpha T + \alpha V$
p	U	T	$\beta p V - \alpha VT$	T	U	S	$\beta p C_p/\alpha T - \alpha p V$
p	H	T	$V - \alpha VT$	T	H	S	$C_p/\alpha T$
p	F	T	$\beta p V$	T	F	S	$\beta p C_p/\alpha T - \alpha p V - S$
p	G	T	V	T	G	S	$C_p/\alpha T - S$

U in the table corresponds to our internal energy E.

[a] From J. Lumsden, **Thermodynamics of Alloys**, Institute of Metals, London, 1952.

wanting this knowledge which relates to the relative stability of competing allotropes or polymorphs of a given material.

In this chapter concerned with the thermodynamics of invariant composition phases, such as elements, molecules, "line" intermetallic compounds, monodisperse colloidal assemblies of hard spheres and Langmuir monolayers, the primary subject we shall address is the relative stability of allotropes and polymorphs of these materials. The set of constraints that apply to the Gibbs free energy are those applicable to most of the phenomena we shall study. Consequently, this thermodynamic potential will occupy an important role in this chapter and in this book. As noted above, materials may exist in either the solid, liquid or vapor states. When in the vapor state, and approximating the ideal gas, a monatomic gas consists of atoms that do not interact and that move freely throughout the container that defines the volume of the gas. Thus, in this state the gas has no potential energy.* Its energy is in the form of kinetic energy of the atoms. In the solid or glassy state the atoms are confined to potential wells of time invariant coordinates and oscillate in these wells. In the extremum condition at the absolute zero of temperature the energy of the solid is, except for the zero point energy, all in the form of potential energy. The liquid state is more akin to the glassy state in that the atoms oscillate in potential wells, but the potential wells may change coordinates with an increasing tendency to do so as temperature increases. In both solids and liquids at equilibrium at a finite temperature the potential and kinetic energies have time independent finite values. However, the specific heat can approach an infinite value at a particular type of phase transformation. (For example, see Figure A4.1.) The potential energy is the sum over the interaction energies between atoms or molecules situated at the minima in their potential wells. The kinetic energy is the sum over the motional energy in vibrations, rotations and translation of the atoms or molecules. Electrons also may contribute to both forms of energy since as temperature increases, at least in conducting materials, there are empty quantum states available for the outer electrons to occupy with a net increase in energy of each atom. These aspects of the states of materials affect their thermodynamic properties in that the configurational and thermal contributions to the latter differ in type in the different states and materials.

---

* This statement is not strictly true in that gravity provides a very small, negligible contribution to the potential energy of the molecules in the ideal gas.

1. Thermodynamic potentials

It is useful to obtain some physical insight into the materials aspect of the thermodynamic potentials. Since the Gibbs free energy of a phase is given by

$$\mathrm{G = E + PV - TS = H - TS} \tag{1.1}$$

then a study of the material origins of the energy and the entropy will be helpful in solving problems involving the equilibrium between phases at constant temperature and pressure.

A review of some of the methods of calculating the free energy of materials is given in Reference 1. An example of a first-principle-based calculation of energies is provided in Reference 2. From a reading of this reference it becomes apparent that first-principle evaluation of the terms yielding the temperature dependence of the free energy involves a significant computer burden, and skills that are not normally practised by scientists involved in the synthesis of materials – a group likely to be interested in the subject matter of this book. There are simpler methods for estimating the temperature dependence of the free energy of a material. Appendix 1 provides brief descriptions of these methods and references to instructions for their use.

1.1. Energy

A key goal of chemistry, and now materials science, is the ability to predict the energy of materials. In 6 years prior to this writing (2000–2006) enormous progress has been made in the ability to calculate these values at absolute zero temperature knowing only the atomic numbers of the elements involved and the likely low-energy configurations of the atoms. The accuracy of the absolute energy values so calculated is stated to be 0.1 eV/atom (9.65 KJ/mole) although the uncertainty in a comparison of relative values can be much smaller. For example, the experimental differences in cohesive energy between allotropes of elements is on the order of 0.04 eV/atom (3.86 KJ/mole) and usually less than 0.1 eV/atom and ab initio methods have successfully calculated the relative stability of allotropes in several cases. Given this rate of progress it appears likely and reasonable to expect that over the lifetime of this edition it will be possible to improve this accuracy.

The state of the art at this writing with respect to the prediction of ground state (stable at 0°K) energies of constant composition phases is as follows. The ground state energies for most of the elements are known and may be found in data banks (see bibliography). Most of these data were first deduced from experiments. The ab initio values followed and agreed with the experimental values within the 0.1 eV/atom uncertainty. The energies of allotropes of the elements or the difference in energy between these allotropes and the stable structure (known as the lattice stability energy) are also known and exist in data banks. Most of these energies have been obtained by extrapolation from experimental values for alloys in the composition range where the alloys are stable.

It is an interesting fact that first-principle calculations reveal that when the fcc (bcc) structure is stable then the bcc (fcc) structure is unstable, at least for transition metals.[3] This instability is revealed as a negative value of a shear modulus. However, it appears that for a few simple metal elements, Li, K, Rb, Ca and Sr, minima in the energy versus Bain strain (tetragonal distortion, see Figure 1.1) exist simultaneously for both the bcc and fcc structures, with one of them being the stable ground state structure.[4] For other than the transition metals good agreement between the lattice stabilities evaluated from experimental data and from ab initio calculations is obtained. For the transition metals the agreement is not good, although if the same method used to obtain experimentally based values (by extrapolation from values for alloys) is used for the ab initio based data, then acceptable agreement is obtained.[5]

The other materials of constant composition, such as intermetallic compounds and molecules, also reveal agreement between observed and ab initio based predictions of energies of formation,* but with a larger absolute value of the uncertainty. The rms uncertainty in the energy of formation is about 0.17 eV/atom (16 KJ/mole) for molecules formed from the following list H, C, N, O, F, Al, Si, P, S, Cl.[6] For intermetallic and inorganic compounds an evaluation of the accuracy of ab initio predictions of energy of formation has not been carried out to the best of my knowledge. However, it has been claimed that the uncertainty is on the order of 10%. As noted, Appendix 1 provides a description of some ab initio and also

---

* The difference in energy between that for the compound and the atomic composition weighted average of those for the pure components.

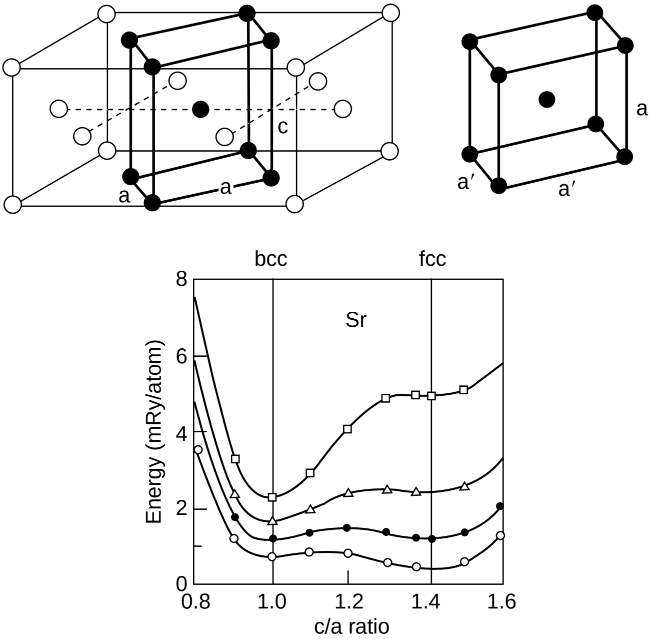


Figure 1.1. Showing above the bc cell which has a c/a value of $\sqrt{2}$ in the fcc structure (left) and of 1 in the bcc structure (right). The strain transforming one to the other structure (varying c/a) is known as the Bain strain. The figure below shows the ab initio deduced energy as a function of the Bain strain for the element Sr for several different atomic volumes. Each curve corresponds to a constant atomic volume (largest atomic volume curve is the lowest one). From J. Phys. Condens. Matter. 8, 799(1996) with permission. © 1996 IOP Publishing Ltd.

empirical, less computer costly, and quicker methods of obtaining approximate values of ground state energies of formation.

Figure 1.1 shows the results of ab initio calculations[4] of the energy of Sr as a function of the Bain strain. It is interesting to note that as the atomic volume of Sr is decreased, as for example by application of hydrostatic pressure, the stable structure changes from a fcc one to a bcc one. Indeed, experimentally, the ambient stable structure of Sr, fcc, is changed to the bcc one on application of the pressure of 3.5 GPa. Another interesting relation is exhibited in this figure. The curvatures at the minima are proportional to the shear constant C'. One may note that for smaller atomic volumes (i.e. as the fcc structure becomes metastable with respect to the bcc one with increasing pressure) the curvature at the fcc c/a minimum decreases. Or stated differently, C' (fcc) decreases and actually becomes zero at a sufficient pressure. When a shear constant of a metastable phase, such as the fcc structure of Sr at a pressure above about 3 GPa, approaches a low value, the lattice becomes prone to undergo a diffusionless (martensitic) cooperative transformation to the stable structure. Another feature revealed in Figure 1.1 is that once pressure induced a transformation to the bcc structure, release of the pressure would allow the bcc structure to remain metastable at atmospheric pressure conditions at low temperature, i.e. an energy barrier exists to the bcc → fcc transformation under these conditions.

The above description of stability and metastability of the bcc and fcc structures is incomplete because it relates to only one path in traversing between these two structures. True stability and metastability requires a local minimum in the free energy (energy at 0°K) with respect to all possible perturbations from the state at the minimum. We have considered in the above only one type of perturbation, the Bain strain.

1.2. Entropy

Entropy is much more complex, and interesting to the enquiring mind, than energy. To obtain some knowledge of its physical significance we need to exit the thermodynamic discipline and enter that of statistics,[7] as did Boltzmann who provided the first clue as to the meaning of entropy on the molecular level and Gibbs on the macroscopic level. Following Gibbs (see Appendix 2) the entropy of a "canonical"* ensemble of systems is

$$S = -k\Sigma_i \, [p_i \ln p_i] \tag{1.2}$$

where k is Boltzmann's constant and p_i is the fraction of the total very large number of specimens of the same system that are in the ith quantum state, distinguishable from the jth quantum state. If there are g distinguishable quantum states, all of which are equally probable, in the ensemble of specimens, then $p_i = 1/g$, which, when substituted in equation 1.2, yields the following statistical definition of entropy**

$$S = k \ln g \tag{1.2a}$$

If we are interested in the difference in entropy between two states of a classical systems of non-interacting particles and can evaluate the number of different ways that the system having entropy S can be arranged at constant energy and other extensive parameters of the system then the difference in entropy between these two states of the system is given by

$$S_1 - S_2 = k \ln(W_1/W_2) \tag{1.2b}$$

---

* An ensemble of systems which are all in contact with a large heat bath having a constant average energy.

** $p_i = e^{-Ei/kT}/\Sigma_i e^{-Ei/kT}$, where the denominator is called the partition function Q.

where k is Boltzmann's constant and W is the number of different ways . This is the fundamental relation that we shall use to describe the entropy associated with various systems we will be considering in the remainder of this book.

Conceptually, we may compartmentalize the various contributions to the entropy of a substance as follows.

1. Configurational entropy: In a crystal having an invariant composition and more than one component this contribution to the entropy is referred to as the entropy of disorder. This contribution refers to the distinguishable ways the atoms can be arranged on the lattice sites of the crystal. For a molecule, the distinguishable ways of arranging the orientation of the molecule also contributes to the configurational entropy. Polymers have configurational entropy even for the case of a polymer corresponding to a linear chain of a single type mer. This configurational entropy depends upon the stationary states of the polymer, which in reality may be complex. At constant energy, it is apparent that there are many distinguishable configurations of a linear chain polymer of a single type mer. A derivation, following Kittel,[8] *of the configurational entropy of an unreal model polymer in which the mers are confined to a plane is given in Appendix 3.*

2. Thermal entropy: At any finite temperature, the atoms of a solid are in continuous oscillating motion about their equilibrium positions. Thus, at any time there exists uncertainty concerning the exact positions of the atoms. From the theory of such thermal motion – the theory of phonons – certain conclusions can be drawn concerning the influence of material properties, such as cohesive energy, on the entropy of thermal origin.

For Einsteinian monatomic solids (i.e. consisting of 3N independent harmonic oscillators), the thermal entropy is given by

$$S_T = 3Nk\left\{\frac{\theta_E/T}{\left[\exp(\theta_E/T)\right]-1} + \ln\left(\frac{\exp[\theta_E/T]}{\left[\exp(\theta_E/T)\right]-1}\right)\right\} \tag{1.3a}$$

where the Einstein characteristic temperature, $\theta_E = h\omega/2\pi k$. Also, ω is the frequency of the harmonic oscillator, h is Planck's constant, k is Boltzmann's constant, and T is the absolute temperature.

For $T \gg \theta_E$, this expression simplifies to

$$S_T = 3\,Nk[\ln(T/\theta_E)] \tag{1.3b}$$

Thus, the thermal entropy of an Einsteinian monatomic solid should be a monotonic increasing function of T. Further, by dividing the temperature by a number characteristic to each solid it should be possible to make the thermal entropy – temperature functions of different monatomic solids superpose. This characteristic

Table 1.3. Approximate Debye θ values for the elements (°K).

Ar	85	Co	119	Hf	200	bcc Mn	$\sim$370	Re	275	Th	140
Ag	220	hcp Co	390	Hg	100	Mo	380	Rh	370	Tm	167
Al	395	fcc Co	385	Ho	162	Na	150	Ru	400	hcp Ti	365
As	285	Cr	490	In	129	Nb	250	Sb	210	bcc Ti	300
Au	177	Cs	45	Ir	285	Nd	145	Sc	$\sim$400	Tl	94
B	1250	Cu	315	K	100	Ne	63	Si	625	U	160
Ba	110	Dy	157	Li	360	Ni	390	Sm	150	Va	335
Be	920	bcc Fe	432	La	132	Os	250	(grey) Sn	260	W	310
Bi	120	fcc Fe	420	Lu	166	Pb	88	(white) Sn	170	Y	214
C(diamond)	1860	Er	167	Mg	318	Pd	290	Sr	140	Zn	235
C(graphite)	$\sim$400	Ga	240	α Mn	$\sim$380	Pr	141	Ta	230	hcp Zr	260
Ca	230	Ge	360	β Mn	$\sim$380	Pt	233	Te	180	bcc Zr	212
Cd	175	Gd	170	fcc Mn	$\sim$355	Rb	61	Tb	170		

From R.J. Weiss, **Solid State Physics for Metallurgists**, Pergmon Press, 1963.

number is, ofcourse, the Einstein characteristic temperature, θ_E, roughly equal to the Debye temperature (see Table 1.3).

Some insight, from the viewpoint of a materials scientist, can be achieved by noting that the frequency of a harmonic oscillator is given by $(K/M)^{1/2}$, where K is the force constant and M is the mass of the oscillator. Dimensional analysis suggests that the force constant of the "spring" should be proportional to the quotient of the cohesive energy by the square of a characteristic distance, such as the interatomic distance d, i.e. $K \propto E/d^2$ or $K \propto E/\varpi^{2/3}$, where ϖ is the atomic volume. A similar relation can be derived on the assumption that the energy of the solid is given by a sum over interatomic potentials.

The Einstein model is a poor approximation of the vibrational modes in a solid. A more realistic, but still inadequate, model was suggested by Debye. In a Debye solid the density of normal modes depends upon the square of the frequency up to some maximum frequency determined by the constraint that the total number of normal modes equals 3N. Although Debye's model was an improvement on Einstein's it is still not realistic in that the density of normal modes departs drastically from the frequency squared dependence in real crystals. The atomic nature of each crystal exerts a significant effect on the normal mode-frequency spectrum, which is investigated in books discussing phonons (see bibliography). Although we shall require little use of the results of this discipline in the remainder of this book, a very brief discussion of Einstein and Debye theories is given in Appendix 4.

The main results of the theories of lattice vibrations applicable to the present subject are as follows. The thermal entropy is a continuous monotonic function of the temperature, increasing with the temperature. It is also a function of the atomic mass and the cohesive energy, increasing as the former increases and decreasing as the latter increases for a given type of bonding and crystal structure. Variation of

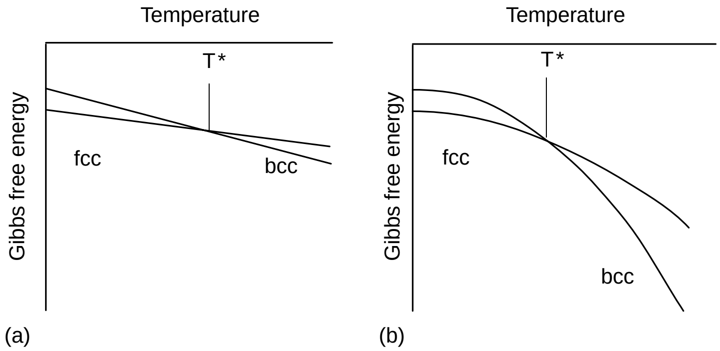


Figure 1.2. Temperature dependence of Gibbs free energy.

the bonding type exerts a strong effect as will be discussed later. Further, as may be expected from the above analysis, the thermal entropy at a given temperature will be crystal structure dependent.

The entropy of thermal origin plays a controlling role in many polymorphic transitions in the solid state. As we will note later, with increasing temperature, a first-order polymorphic transition* is between a phase that has a lower entropy to one that has a higher entropy (see Figure 1.2). In view of the relation between thermal entropy and cohesive energy discussed above, then for transitions controlled by thermal entropy the transition is between a phase of higher cohesive energy and one of lower cohesive energy.

For gases and liquids the entropy of thermal origin is given by the Sackur–Tetrode relation. The model yielding the Sackur–Tetrode equation for the translational entropy of an ideal gas needs no repetition here. It is described in numerous textbooks on statistical mechanicsas, e.g., that of Hill.[9] We will derive this equation in another way that sheds light on the configurational portion of the translational entropy which is not apparent in the usual derivation. First, we note that the "thermal wavelength", $\Lambda = (h^2/2\pi mkT)^{1/2}$, represents the uncertainty in the position of the particle corresponding to the thermal wave. We consider a cell of volume Λ^3. In effect this is the volume associated with the uncertainty of the particle's position. The number of such cells in a volume V is then $N_\Lambda = V/\Lambda^3$. The number of different ways that N alike molecules can be arranged in these cells is then $N_\Lambda!/[(N_\Lambda - N)!N!]$. Hence, the configurational entropy associated with N particles in the volume V is $k\ln\{N_\Lambda!/[(N_\Lambda - N)!N!]\}$. With $N_\Lambda \gg N$ we may use Stirling's approximation, $\ln M! = M\ln M - M$, with the result that this configurational entropy equals $Nk\ln(V/N\Lambda^3) + Nk$. To get the translational entropy of the ideal gas we add to this term the entropy due to the motional degrees of freedom for N molecules, 3 Nk/2, to obtain the Sackur–Tetrode equation for the translational entropy of an ideal gas

---

* A first-order transformation is defined as one in which there is a discontinuity in the first derivative of the free energy with respect to an intensive parameter, such as temperature.

$$\text{Ideal monatomic gas} \quad S = Nk \ln\left[\left(\frac{2\pi mkT}{h^2}\right)^{3/2} \frac{Ve^{5/2}}{N}\right]$$

The Sackur–Tetrode relation has been applied to liquids with schemes of correcting the relation by a host of investigators. In this application V/N is taken to be the volume per molecule in the liquid, which is an appreciably smaller quantity than it is for the ideal gas. However, the volume Λ^3 is still much smaller than V/N so that the derivation given above is still applicable to the liquid. The point in question is whether all of V/N is available to be occupied by a molecule because the size of the molecule is itself on the order of the volume available per molecule. One group subtracts the effective excluded volume per molecule from V/N as determined from measured pair correlation functions and theory. Another group, which is concerned with obtaining "back of the envelope" estimates of the translational entropy defines the volume as that available for occupation by the center of gravity of the molecule in its translational motion within the bounds of the encircling molecules. This more approximate method suffers from an inability to define a hard sphere volume. For the latter case the volume available might approach that of Λ^3. A cursory examination of the latter possibility for liquid metals indicates that for these dense liquids that Λ^3 is still much smaller than the volume of liquid available to the center of gravity of a metal atom in the liquid to move within the bounds of its neighbors.

An examination of the literature reveals that the accuracy of any of the methods to evaluate the translational entropy of liquids, including molecular dynamics ab initio ones, cannot yield uncertainties better than about 5 J/mol/°C, except for the simple metals and other simple cases. Often, the error involved in predicting the translational entropy of liquids is much larger than this value. (More than 50 years of attempts to improve the theory have not produced better predictions.) Thus, this topic will not be explored here in greater detail.

3. Electronic entropy: Because electrons at the Fermi level in metals have unoccupied states into which they can be excited by thermal energy and because of the corresponding uncertainty in the distribution of such electron states there is a contribution to the total entropy of electronic origin in metals, which is absent in insulators. This contribution is generally small and usually can be neglected in problems that will be encountered in this text.

4. Magnetic entropy: The atoms of many elements have net magnetic moments. These moments may have their vectors oriented randomly, as in a paramagnetic substance, or may have them aligned along a specific direction, as in a ferromagnetic substance. In the latter case the magnetic entropy is zero because there is no uncertainty governing the orientation of the magnetic moment vectors.

In the former case, it is

$$S_{mag} = Nk\ln(2J + 1) \tag{1.4}$$

where J is the quantum number of net spin on an atom.[10]

The problem of describing the magnetic entropy for the case of partial ordering of the spins is known as the Ising problem. An analogous problem exists in the determination of site occupation probabilities by different components of alloys. We shall postpone an analysis of the Ising problem to a later section. Magnetic entropy exerts a controlling effect in the low temperature polymorphic transition found in iron and in many magnetic alloys and compounds.

As noted, the free energies, enthalpies and entropies of many elements and stoichiometric solids have been measured and are collected in various data banks. The bibliography at the end of this chapter provides a list of these sources. Also, within the past few years relative to the time of this writing progress has been made in obtaining reliable ab initio deduced values of thermal entropy of simple solids (see Bibliography).

There is an interesting experiment that the reader can perform to be convinced of the reality of separation of the entropy into configurational and thermal contributions. The human lip is a sensitive thermometer. We suggest that you use it to determine changes in the temperature of a wide rubber band as it is stretched and unstretched rapidly. The rapid stretching or unstretching of a rubber band approaches an isentropic process in which the total entropy is conserved because there is no transfer of heat to or from the rubber band in the period during which the rapid stretching or unstretching is carried out. From Appendix 3 we can determine that the stretching of an elastomer, such as a rubber band, should decrease the configurational component of the entropy. Since the stretching is isentropic, then the thermal entropy must increase to compensate for the decrease in the configurational entropy. But, the thermal entropy is a single valued function of the temperature, increasing with an increase in the temperature. Thus, the temperature of a stretched rubber band just after stretching should be higher than just before the stretching process. Similarly, the temperature of the rubber band just after rapid unstretching should be lower than it was just before the unstretching process. Carry out the experiment and determine whether the above predictions are consistent with your observations.

2. Polymorphism

2.1. Phase transitions on varying the temperature

Let us now consider several aspects of the concepts discussed above. From the cohesive energy values listed in data banks for non-transition elements it is possible to deduce that usually the fcc structure (the ideally close-packed structure) is more stable than the bcc structure (a less close-packed structure).* (In this context, the term "packing" is meant to correspond to nearest-neighbor coordination number rather than atomic volume. The lower this coordination number the more open is the "packing" in the present usage.) Then, equation 1.3 yields that the entropy of thermal origin for these elements is higher for the bcc crystal structure relative to that for the fcc structure. This result is illustrated in Figure 1.2a. As shown there, the energy and entropy of each phase has been assumed to be independent of temperature (i.e. the entropy equals $-[\partial G/\partial T]_P$ and hence a constant entropy corresponds to a constant slope of the line representing the dependence of the free energy G on temperature T). In general, this assumption is not representative of the real behavior. The entropy of a phase increases with increasing temperature and the dependence of free energy on temperature corresponds more to that shown in Figure 1.2(b). The specific heat is proportional to the curvature of the G(T) line, i.e.

$$C_p = -T[\partial^2 G/\partial T^2]_P.$$

Since, the specific heat of materials has positive definite values, then the curvature of the G(T) line must be negative.

We may deduce, from the knowledge that the free energy is a minimum for stable equilibrium, that the fcc structure is the relatively stable one at temperatures below T* and that the bcc structure is stable relative to the fcc structure at temperatures exceeding T*. At the latter temperature, the bcc and fcc structures are in equilibrium with each other. This conclusion can be generalized to imply that the close-packed structure is stable at low temperature, whereas the less close-packed structures become stable at higher temperatures, at least for the case of simple metals (non-transition elements). This "rule" is also applicable to ionically bonded solids,[11] where it is stated in terms of coordination number. The latter increases with increasing close packing. However, covalently bonded solids and transition metals do not obey this "rule".

Covalently bonded solids differ from metallic and ionic solids primarily in that the effective spring constants associated with the normal modes are much stiffer for the former type because of the additional bond-bending constraints present only in the covalently bonded solids. Thus, it is to be expected that the transition from mainly covalent (the lower entropy phase) to mainly metallic bonding (the higher entropy phase) in a given solid occurs as temperature increases, e.g. gray-Sn → white-Sn transition at 13°C.

Because $S_{fcc} \neq S_{bcc}$ at T* and $S = -[\partial G/\partial T]_{P\ldots}$, the transformation that occurs at T* in Figure 1.2 is a first-order transformation, i.e. the discontinuity occurs in the first derivative of the free energy. Higher order transformations occur in Nature. The ferromagnetic–paramagnetic transformation in bcc iron is a second-order transformation, for example, as is the order–disorder transformation in beta brass and NH_4Cl. In this case, the discontinuity at T* occurs in the 2nd derivative of the free energy with respect to an intensive parameter, such as temperature, i.e. C_p (see Figure A4.1).

---

* Physically, this result stems from the fact that the contribution of the outer electron density to the cohesive energy increases with increase in this density for these elements.

It is possible to supercool a liquid below the equilibrium freezing point temperature with no discontinuous change in its properties, as well as to superheat a solid above its melting point, again with no discontinuous change in its properties. Thus, the freezing or melting transition at equilibrium is not the consequence of the onset of some physical catastrophe in the phase with change in temperature through the transition temperature. Rather, the first-order transition occurs because the entropies of the two phases differ and the Gibbs free energy of the solid equals that of the liquid at the melting point (see statistical model of transition in Reference 12).

Liquids have a higher entropy than their corresponding solid states, in part because there is an additional contribution to the thermal entropy in liquids not present in solids. This added factor corresponds to the uncertainty in specification of the positions of the potential wells about which the atoms oscillate, due to rapid diffusion of these wells. For a monatomic material, this difference in entropy, the “entropy of fusion”, can be derived and a derivation is given later in this chapter. The entropy of fusion of metallic elements, and compounds that upon melting retain their molecules in undissociated form,* is about 2 cal/mol/°C (Richard’s rule), which is just the value derived later in the chapter. The fact that the vibrational frequency spectrum of these atoms is not too different in the liquid and solid states is consistent with this result.

Covalently bonded elements (C, Si, Ge) and compounds which dissociate into atoms or ions rather than molecules upon melting exhibit higher values (i.e. 6–6.8 cal/mol/°C) than the Richard’s rule value for the entropy of fusion. This result stems partially from the fact, mentioned above, that a covalently bonded solid will have a lower entropy than its metallic form due to the stiffer spring constants of the former type and to the additional fact that these elements are metallically bonded in the liquid state. Indeed, the elements Ni and Si, which have nearly the same value of the melting point, also have nearly equal entropy values at the melting point in their liquid states. Covalently bonded solid Si has the smaller entropy value of 14.8 cal/mol/°C as compared to that of 20.6 cal/mol/°C for metallically bonded solid Ni at their nearly equal melting points. A different and incorrect explanation for the entropy of transition of covalently bonded solids to metallically bonded ones was described in the first edition of this book. This explanation is not applicable because it provides a temperature independent difference between the entropies of these two bonding types for a given solid, whereas, in fact, this entropy difference is temperature dependent.

---

* The dissociation of molecules into atoms increases the configurational entropy due to an increase in the number of different ways the components of the molecules can exist in the liquid and hence, when dissociation takes place on melting, the entropy of fusion will be larger than that given by Richard’s rule.

Amorphous solids or glasses are in a "frozen-in", non-equilibrium, state rather than an equilibrium state. Nevertheless, this class of polymorphs have an energy and entropy that characterize them thermodynamically. Such solids usually have a larger specific volume, a more positive energy and a larger entropy than their crystalline counterparts. However, there may be exceptions to the latter statement in certain cases, which we will consider in a later chapter. The entropy of the amorphous solid is not as high as that for the liquid because the diffusional contribution to the thermal entropy, that contribution due to diffusion of the potential wells, is larger in the liquid (it is absent in glasses below the "glass" temperature).* The latter effect is often neglected.

Polymers and biopolymers, such as proteins, also may undergo phase transitions on varying the temperature. The coil to globule transition and the coil to lamellar crystal transition in polymers are examples of a second-order and first-order transition, respectively, while the denaturation transition in proteins is another example of a second-order transition. The remarks concerning the relative values of the entropy of each phase in the transition above apply also to these transitions with one exception. If the entropy of solution of the polymer in water is negative, then the entropy of water must also be taken into account. In this case the higher entropy configuration of the system, polymer in water, consists of the globule in water. Hence, on decrease in temperature the transition is from the globule to the coil configuration in the water solution. Normally, the entropy of the coil configuration is higher than that of the globule phase, but if the entropy of solution of the polymer in water is negative then in water the reverse order describes the relative entropies of the system of polymer and water.**

2.2. Phase transitions on varying the pressure

It is also possible to vary the pressure P. Hence, by $[\partial G/\partial P]_T = V$, and the fact that the molar volumes of competing phases are not equal, in general, we obtain the possibility of having polymorphic transitions with increasing pressure. This concept is described graphically in Figure 1.3. As shown, with increasing pressure the phase having the smaller molar volume will be stabilized. This transformation is

---

* The configurational entropy of a frozen liquid approaches that of a crystal at a temperature denoted the Kauzmann temperature, after Kauzmann who first noted this fact.

** A negative entropy of solution in water implies an ordering of the water molecules adjacent to the polymer due to an interaction between them. The lowest entropy of the water–polymer solution corresponds then to the highest interface area between polymer and the water, i.e. a coil configuration.

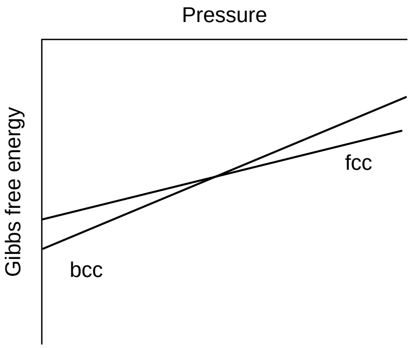


Figure 1.3. The dependence of Gibbs free energy G, on pressure P.

first-order because the first derivative of the free energy with respect to an intensive variable, in this case pressure, is discontinuous, i.e. V is discontinuous at the transformation pressure. Below the transition pressure, the phase with the larger molar volume is stable (i.e. it has the more negative values of Gibbs free energy), while above the transition pressure, the phase with the smaller molar volume is the stable phase. These curves also have a curvature to them given by

$$[\partial^2 G/\partial P^2]_T = [\partial/\partial P]_T[\partial G/\partial P]_T = [\partial V/\partial P]_T = -\beta V,$$

where β is the compressibility. Since β and V are positive quantities, the curvature must be negative in sign.

2.3. P–T phase diagram

A P–T phase diagram, such as that shown in Figure 1.4a, reveals the regions of stability of the phases present in a single component system for these intensive variables. The existence of a number of stable solid phases in different regions of the P–T diagram is termed polymorphism. Polymorphic transitions may occur on changing the temperature or the pressure. It is often desirable, in the search to synthesize some particular phase, to be able to predict the corresponding P–T phase diagram. The state of the art is such that this prediction can now be accomplished via ab initio methods.

There is a relationship called the Clapeyron equation that relates the slope of a phase boundary in a P–T phase diagram to the changes in molar enthalpy and volume and the transformation temperature. It is derived as follows. At equilibrium between two phases, denoted by ′and ″, the molar free energies of the two phases must be the same,

$$G' = G'' \tag{1.5}$$

Substituting from equation 1 we are then able to write

$$H' - T_E S' = H'' - T_E S''$$

or

$$\Delta H_t = T_E\, \Delta S_t \tag{1.6}$$

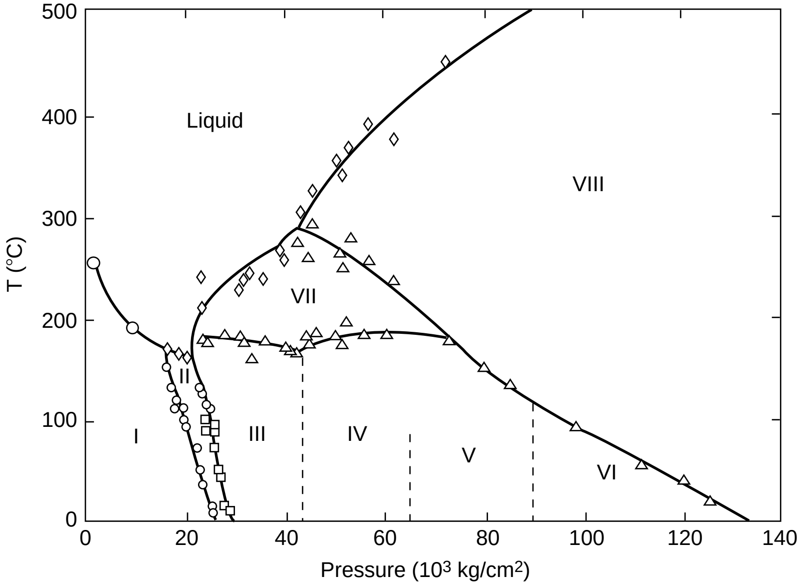


Figure 1.4a. Phase diagram of bismuth. From H.M. Strong. Am. Scientist. 48, 58(1960) with permission.

Along the phase boundary both T and P must be varied in such a way that equation 1.5 is satisfied, or that $dG' = dG''$. But for the single component system

$$dG = -SdT + VdP.$$

Thus, we may write

$$-S'dT + V'dP = -S''dT + V''dP$$

or

$$(\partial P/\partial T)_E = (S' - S'')/(V' - V'') = \Delta S_t/\Delta V_t$$

We can substitute for ΔS_t from equation 1.6 to obtain

$$(\partial P/\partial T)_E = \Delta H_t/(T_E \Delta V_t) \qquad (1.7)$$

Equation 1.7 is the Clapeyron equation.*

---

* It should be noted that equation 1.7 holds only for fluid phases that do not withstand a shear stress. For nonhydrostatically stressed solids the appropriate Clapeyron equation depends upon the imposed boundary conditions and system geometry.[13]

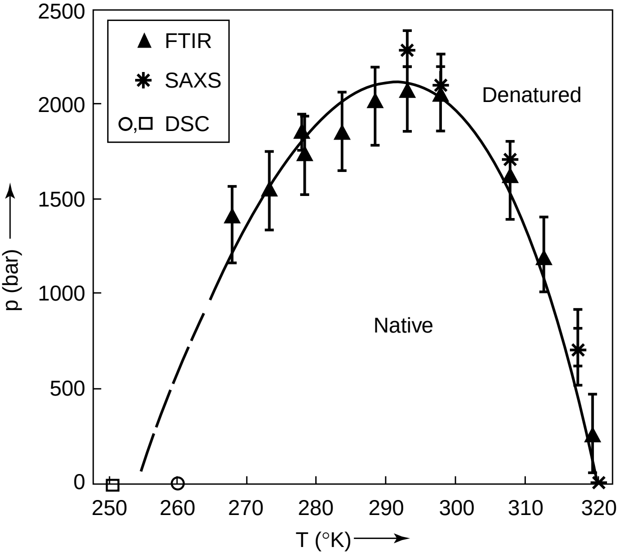


Figure 1.4b. Pressure–temperature stability diagram for Staphylococcal nuclease at p_H 5.5, as measured using SAXS, FTIR and DSC. From Chemphyschem. 4, 359(2003) with permission. © 2003 Wiley-VCH Verlag GmbH & Co KGaA, Weinheim.

We have noted that on increase in temperature the entropy increases either within a given phase or upon transformation to a new phase. Thus, the heat of transformation as defined in equation 1.7 must be positive. The sign of the slope of a phase boundary in the P–T diagram, must consequently depend upon the sign of ΔV, the change in molar volume on transformation from the low to the high temperature phase. Normally, this volume will increase and hence ΔV will be positive in value. However, there are many cases where the low temperature phase is a covalently bonded open structure and the high temperature phase is a closer-packed structure, such as a metallic phase. In this case, the sign of ΔV will be negative. For example, covalent Si is stable in the open diamond cubic structure as a solid and upon melting becomes a metallic liquid of smaller molar volume.

Not only do inorganic substances reveal polymorphic transitions but organic ones also exhibit polymorphic transitions. Figure 1.4(b) shows the typical polymorphic transition (denatured – native) found in the protein Staphylococcal nuclease. The denatured state is the state of the uncoiled polymer-like strand while the native state is the globular state in which the strand is coiled into the globular shape. There may still be short-range order in the uncoiled state, i.e. the long-range order of the coiled state is lost but there may still be a few α-helices and β-sheets in the uncoiled strand. The elliptical shape of the boundary between the two states is a common feature in the P–T diagram for proteins exhibiting these two polymorphs.

2.4. Effect of magnetic entropy on polymorphic transitions

Besides coordination number (packing) and bonding type other physical factors can affect the energy and entropy of phases. One such quantity is the net atomic magnetic moment due to electron spin. When there is an ordered state of the atomic magnetic moments their contribution to the entropy of the phases containing them is zero and the spin ordered state is a more stable array energetically than the spin disordered state. When these atomic magnetic moments are arranged randomly in a solid there is a contribution to its total entropy given by equation 1.4. The changes in magnetic entropy with temperature can give rise to polymorphic transitions, as occurs in iron. In iron, the stable polymorph at low temperature has the bcc structure. Above 910°C and below 1390°C the fcc structure is the stable structure for iron, and above 1390°C the bcc structure again becomes stable up to the melting point of iron.

The clue to understanding this behavior was first provided by Zener,[14] who noted that the curvature in the plot of ($G_{bcc} - G_{fcc}$) versus temperature exhibited a maximum in the vicinity of the Curie temperature for bcc iron. Thus, one may interpret the low temperature transition as follows. At low temperature bcc iron is ferromagnetic and hence the atomic magnetic moments are fully ordered in this structure. This enthalpic contribution of magnetic origin to the stability makes the bcc structure stable relative to the more close-packed fcc structure at T = 0°K. However, in the fcc structure, this ordered state of spins does not exist and hence there is a magnetic contribution to the entropy of the fcc structure. Thus, the magnetic entropy in the bcc structure and its total entropy is less than that for the fcc structure. Consequently, the free energy versus temperature curves near 0°K are as shown in Figure 1.5. The higher entropy of the fcc structure brings about the first transition at 910°C. The complete loss of short-range magnetic order of electron spins in the bcc structure does not occur until the temperature exceeds 910°C. Because, in the absence of any magnetic spin order, such as is the case at very high temperature above the Curie temperature, the bcc structure has a higher entropy than the fcc structure, as already noted, it will eventually become more stable than the fcc structure. This fact is responsible for the transition that occurs at 1390°C. The latter transition obeys the general rule, which has its origin in the dependence of

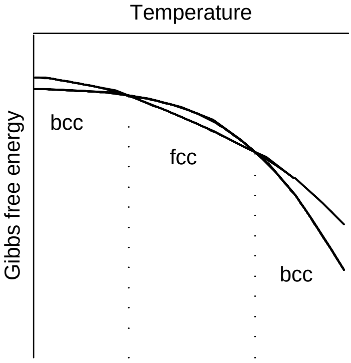


Figure 1.5. Schematic dependence of Gibbs free energy versus temperature for iron.

entropy on coordination number, while the low temperature transition has its origin in the different magnetic properties of bcc and fcc iron.

The rapid change in entropy that occurs below the first transition temperature (910°C) is indicated by the rapid change in the specific heat of iron below this temperature that is shown in Figure A4.1. Note the approach to infinity of the specific heat at the Curie temperature.

2.5. Melting transition in hard sphere system

Hoover and Ree[15] by careful Monte Carlo simulation of a hard sphere fluid in which each sphere is constrained to have its center within a fcc cell of volume equal to total volume divided by the number of atoms were able to obtain the equation of state for the fcc hard sphere system. They also were able to obtain the equation of state for the hard sphere fluid from a Monte Carlo simulation. From these data they were able to calculate the entropy of both the hard sphere fluid and the fcc hard sphere system as a function of the volume fraction. The results are presented in Table 1.4 as differences from the value for the ideal gas. (φ in the table is the volume fraction.) At the melting transition the Gibbs free energy at constant pressure and temperature must be equal, i.e. $(PV - TS)_{\text{fluid}} = (PV - TS)_{\text{fcc}}$.[*] Using this relation and their results Hoover and Ree found a first-order transition between the fluid phase branch at $\varphi = 0.494$ and the solid fcc phase branch at $\varphi = 0.545$ at which values the chemical potentials and pressures in the two phases were equal. Thus, from Table 1.4 we are able to set down the entropy of fusion as $-4.911 + 5.978 = 1.067$ in units of k.

Table 1.4. Entropies of ideal hard sphere system

$\varphi/0.74$	S_{fluid}	S_{fcc}	S_{com}
0.00	0.000	−1.000	1.000
0.05	−0.155	−1.061	0.906
0.10	−0.326	−1.155	0.829
0.15	−0.516	−1.278	0.762
0.20	−0.726	−1.428	0.702
0.25	−0.961	−1.608	0.647
0.30	−1.225	−1.816	0.591
0.35	−1.524	−2.061	0.537
0.40	−1.864	−2.352	0.488
0.45	−2.253	−2.695	0.442
0.50	−2.702	−3.101	0.399
0.55	−3.226	−3.583	0.357
0.60	−3.843	−4.155	0.312
0.65	−4.577	−4.806	0.229
0.66	−4.741	−4.933	0.192
0.67	−4.911	−5.058	0.148
0.68	−5.088	−5.184	0.107
0.69	−5.272	−5.311	0.072
0.70	−5.463	−5.440	0.044
0.71	−5.662	−5.570	0.023
0.72	−5.870	−5.703	0.009
0.73	−6.087	−5.838	0.001
0.74	−6.313	−5.978	0.000

Among the results that also may stand the test of time are those of the Monte Carlo results of Hoover and Ree[15] concerning the communal entropy of a hard sphere fluid. Their results for the communal entropy for hard spheres, disks and rods are shown in

---

[*] The energy of an ideal hard sphere system is zero.

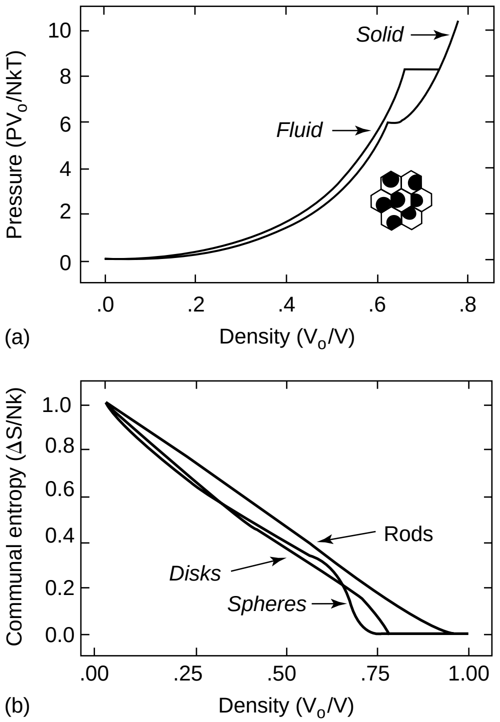


Figure 1.6a. Pressure–density (V_o is volume at close packing of spheres) of hard spheres. Communal entropy versus density for hard spheres, disks and rod. From J. Chem. Phys. 49, 3609(1968) with permission. © 1968 American Institute of Physics.

Figure 1.6. The communal entropy being considered here is that of Kirkwood and is given by $k \ln[Q_N/Q]$ where these Q values are determined as follows. There are N particles of hard sphere diameter σ in a volume V. This volume is divided up into N cells arranged in a fcc lattice. The number of different configurations which have the center of a particle in each of these cells is Q. The total number of different configurations in an ideal gas including those which have no particle centers in some of the cells is Q_N. Thus, the communal entropy, which involves the difference in these numbers, is the number of configurations which do not have the center of a particle in each cell, i.e includes those that have the centers of two particles in one cell, etc. This communal entropy turns out to be a linear function of the volume fraction of particles decreasing as the latter increases. The communal entropy is a measure of the density fluctuations in the fluid.

The entropy of fusion, 1.067 k, is much larger than the communal entropy of the coexisting fluid phase, 0.148 k. Hence, some other entropy is responsible for the entropy of fusion. We shall find the answer at the end of the next chapter where we are able to perform an analytic evaluation of the equilibrium between fluid and fcc hard sphere phases. One very interesting result of the calculations of Hoover and Ree is that the entropy of the fcc phase becomes larger than that of the fluid phase beyond a density relative to the close-packed state equal to 0.70 (volume fraction of 0.518), i.e. there is more room for a sphere to move in the solid phase than in the fluid phase at volume fractions exceeding about 0.518!*

* This insight was first reported by L. Onsager, Ann. N.Y. Acad. Sci. 51, 627–659(1949).

Ree and Hoover[16] also expressed the equation of state of the fluid phase in terms of virial coefficients* they evaluated numerically from the Monte Carlo simulation. Carnahan and Starling[17] found that they could simplify this relation to one involving integral values of the virial coefficients to yield

$$(pV/NkT) = (1 + \varphi + \varphi^2 - \varphi^3)(1 - \varphi)^{-3}$$

The equation of state found for the fcc solid phase is more complicated.[18]

The excess entropy relative to that for an ideal gas, as derived from the Carnahan–Starling equation of state, is then $Nk\phi\,(3\varphi - 4)(1 - \varphi)^{-2}$.

The example of an ordered structure stabilized by translational entropy treated in this section is an idealized example of similar translational entropy stabilized ordered structures in Nature in the form of liquid crystals, and various biological structures, such as assemblies of nucleosome core particles.[19] The ordered structures stabilized by translational entropy are so numerous that a book can be written on the subject.

2.6. Polymorphs of a 2D system

In this section we consider a monolayer of molecules having an amphiphilic head and a chain-like tail adsorbed on water. Contrary to the hard sphere assembly these molecules involve both energy and entropy contributions to the free energy of their assembly. At very low densities (the area per molecule is then much larger than the cross-sectional area a*, of an isolated molecule) an adsorbed monolayer behaves as a 2D gas. In this state, an amphiphilic molecule can exhibit its conformational entropy without interference from neighboring molecules. Compression of this layer then produces a first-order phase transition to a "liquid expanded" (LE) condensed phase. The energy of interaction between the molecules acts to stabilize the LE phase although in this transformation there is little change in configurational entropy. The area per molecule in the gas phase is on the order of 300–1500 A^2 while in the LE phase it is on the order of, but larger than, a*. The insets in Figure 1.7 provide a schematic illustration of the degree of order associated with the tail (chain) section of the amphiphilic molecule. A second phase transition upon further compression of the layer produces a "liquid condensed" (LC) state for which the change in area is from a* to an area barely larger than the cross-sectional area of a fully extended (all-trans) chain. Whether or not a transition occurs from the LE to the LC phase depends upon the existence of or lack of an attractive interaction between the amphiphilic molecules. A thermodynamic analysis due to

---

* Virial coefficients are the coefficients in the series expansion for $pV/NkT = (1 + B/V + C/V^2 + \cdots)$.

Andelman et al.[20] may shed some light on this matter.

The four major terms in the free energy of a monolayer are

$$F = F_{tr} + F_{conf} + F_{att} + F_{head}$$

The $F_{tr} = -TS_{tr}$ corresponds to the translational entropy of the molecule. To a first approximation, in the gas phase, F_{tr} per molecule is

$$F_{tr}/N \approx -kT \ln[(a - a^*)/\Lambda^2]$$

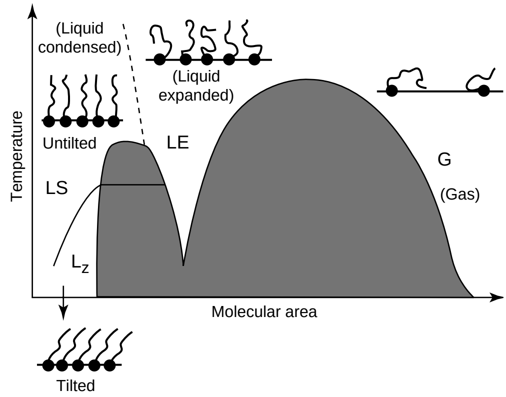


Figure 1.7. Schematic phase diagram at low and intermediate surface coverage.

where N is the number of molecules in the monolayer and a − a* represents the "free area" per chain, giving rise to a lateral pressure

$$\Pi_{tr} = -(\partial F_{tr}/\partial a)/N = kT/(a - a^*)$$

The second term F_{conf} represents the conformational free energy of the chain, which involves both an energetic contribution (on the order of kT) and an entropic one corresponding to the number of different chain conformations Ω. The latter is

$$S_{conf} = Nk \ln \Omega$$

(For an isolated chain ($a > a^*$) of (n + 2) C—C bonds, Ω can be estimated from a 3-state rotational-isomer model: $\Omega = 3^n$. Should a become less than a*, Ω decreases eventually equalling unity when all the chains are in the all-trans state.) F_{att} corresponds to the free energy due to an attractive interaction between the chains and is a function of a. In the gas phase because of the large distance between molecules the attractive interaction takes the form of a van der Waal's interaction

$$F_{att}/N \approx -\text{const}/a$$

where the const term is temperature dependent. Finally, the last term includes the interaction between the head groups. For the moment we will assume F_{head} is a constant, independent of a and T.

The nature of the LE phase is such that each chain can express its conformational entropy. Hence, the main factors involved in the transition from the

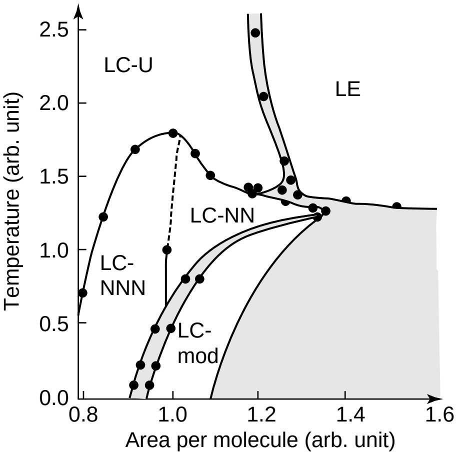


Figure 1.8. Phase diagram for model based on grafted stiff Lennard–Jones chains. LC-U, condensed phase, untilted chains; LC-NN, condensed phase, chains tilted towards nearest neighbors; LC-NNN condensed phase, chains tilted towards next nearest neighbors; LC-Mod, condensed phase with superstructure and intermediate tilt direction.

gas to the LE phase are F_{tr} and F_{att}. In the LE phase, F_{tr} is negligible. Hence, the free energies controlling in the LE to the LC phase transition are F_{conf} and F_{att}. Further, it is possible to show that F_{att} varies as $-d/a^k$, where $k \geqslant 1$ and d is a constant (except for a slight dependence on temperature). The LE to LC transition involves a loss in conformational entropy and a gain in the number of monomer contacts between adjacent chains. A schematic phase diagram for low surface coverage is shown in Figure 1.7. A more realistic phase diagram based on a model of grafted stiff Lennard–Jones chains is shown in Figure 1.8. The caption to Figure 1.8 provides some microscopic descriptions of the various phases. As indicated there the phases are differentiated by the configuration of the amphiphile chains. However, there are a number of additional phases found in monolayers. A summary of such phases can be found in Andelman et al.[20]

Shin et al.[21] have used explicit models for the flexible chains to evaluate the free energy terms more precisely than above. Cantor and McIlroy[22] have used a different model for the same problem. Both arrive at roughly the same conclusions already stated above. One other conclusion is that for rigid rod molecules only a single phase transition occurs, that from the gas phase to the LC phase. The introduction of a flexible group of four methylene units between the rigid chain and the head group restores the LE → LC transition.

From the phase diagrams it is apparent that self-assembly of 2D phases in Langmuir monolayers may be achieved by altering the temperature or varying the molecular area or density of amphiphiles.

Comparison of the 3D hard sphere system and the 2D system of this section shows that the phase transition in the former was controlled by the translational entropy while in the latter it is controlled by the energetic interactions between the molecules. However, for a hard disk system confined to a 2D surface a transition from a disordered array to an ordered array can be stabilized by translational entropy as occurs for the 3D system of hard spheres.[23]

2.7. Disordered, short-ranged-ordered and long-range-ordered solid solutions of invariant composition

Consider the atom positions in the common crystal structures, such as the fcc, the bcc and the hcp structures, of binary solid solutions or the equivalent of a distribution of up and down spins on these lattice sites. At stoichiometric compositions, the unlike atoms (spins) can be rearranged on these atom positions to form what are called superlattices. Figure 1.9 illustrates some of the possible superlattices in the fcc and bcc structures for various stoichiometric compositions.

If the unit cell of a superlattice is translated along each of the unit cell dimensions to build up a macroscopic crystal, the latter will be said to be a long-range-ordered structure. Such long-range-ordered structures can transform to the disordered parent solid solution on increasing temperature. This transformation can be of nth order, as measured by a discontinuity in the nth derivative of the free energy with respect to temperature or to any other independent variable.

Let us first consider the significance of long-range order. Long-range order may be differentiated from short-range order as follows. Consider the unit cell of the long-range-ordered superlattice. Translation of this unit cell along each of the axes of the unit cell by the corresponding lattice parameter results in the construction of the 3D-ordered crystal. The lattice sites corresponding to a particular atom of the unit cell make up a lattice of sites that are occupied by only that type atom in the long-range-ordered crystal. The degree of long-range order may

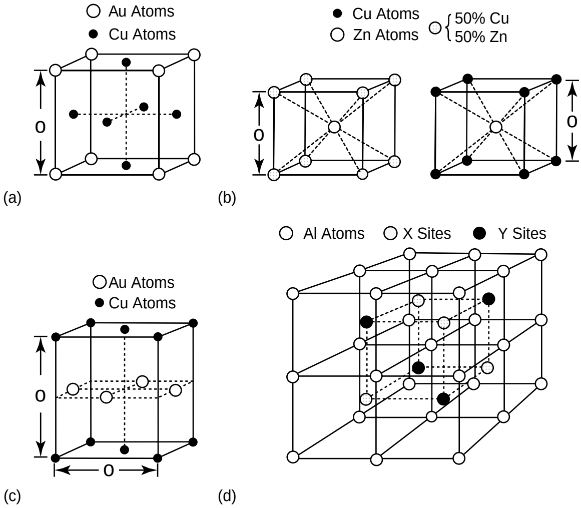


Figure 1.9. Various types of ordered superlattices: (a) Cu_3Au, (b) disordered and ordered structure of β brass, (c) Cu–Au, (d) the structure of Fe_3Al and FeAl.

thus be related to the fractional occupancy of a lattice of sites by the type of atom that occupies this type site in the long-range-ordered unit cell. A typical definition of the long-range order parameter is then

$$\eta = (\rho^* - c)/(1 - f^*)$$

where ρ^* is the fraction of the total number of * lattice type sites occupied by the atom in this type site in the superlattice unit cell, c is the concentration of this type atom in the alloy and f^* is the fraction of the total number of atoms per superlattice unit cell corresponding to the site in question. In the event that there are more than two types of atoms (i.e. a ternary or higher system) then more than one long-range order parameter is required to describe the long-range order in the long-range-ordered phase.

Short-range order corresponds to the number of unlike atoms that are nearest neighbors to a given atom relative to this number in the completely disordered alloy. Thus, the short-range order parameter can be described by

$$\sigma = (N_{AB}(\mathrm{sro}) - N_{AB}(\mathrm{random}))/(N_{AB}(\mathrm{lro}) - N_{AB}(\mathrm{random}))$$

where N_{AB} refers to the number of unlike atom bonds in the short-range-ordered condition (sro), the number in the perfectly long-range-ordered condition (lro) or the number in the random disordered condition (random). Another definition of short-range order parameter involves the correlation parameters for different coordination spheres about an atom. For example,

$$^{LL^*}\varepsilon_{AB}(\rho_l) = {}^{LL^*}p_{AB}(\rho_l) - {}^{L}p_A{}^{L^*}p_B$$

defines the correlation parameter $^{LL^*}\varepsilon_{AB}(\rho_l)$ in terms of the probability, $^{LL^*}p_{AB}(\rho_l)$, that sites L are occupied by the atom A and that sites L* located a distance ρ_1 from atom A (in the lth coordination sphere) are occupied by atoms B, while $^{L}p_A$ and $^{L^*}p_B$ are the probabilities for the occupation of L and L* sites, as defined for the case of the long-range order parameter, by atoms A and B, respectively.

If the probabilities of occupation of sites of a given type were independent of the arrangement of atoms on the other sites, then the correlation factor would equal zero because the first term on the right-hand side of the above equation would equal the second term. When the correlation factor is not equal to zero then there can be both short- and long-range order in a long-range-ordered phase. For disordered alloys in which the long-range order parameter is equal to zero the existence of short-range order is related to a non-zero value of the correlation factor. The short-range order parameter in disordered alloys can be related to the correlation factor, as follows

$$\begin{aligned}\alpha(1) &= 1 - p_{AB}(\rho_l)/(C_A C_B)\\ &= -\varepsilon_{AB}(\rho_l)/(C_A C_B)\end{aligned}$$

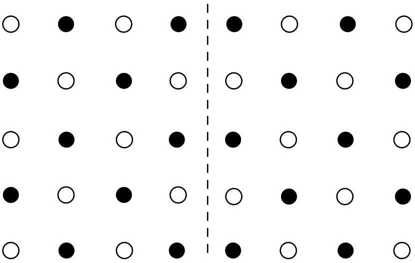


Figure 1.10. Anti-phase domain boundary.

There are two manifestations of partial long-range order. Both correspond to a value of η less than unity but greater than zero. However, one achieves this result in the absence of anti-phase boundaries and the other brings this about by the presence of anti-phase boundaries between perfectly short-range-ordered domains. Figure 1.10 illustrates a case where there is complete short-range order in each domain, but where the long-range order parameter equals zero. It is believed that most ordering systems exhibit both types of partial long-range order at equilibrium.

Superlattices will generally transform from the long-range ordered state to a disordered state as the temperature is increased. This transformation can be either of first order or of higher order in the discontinuity in the derivative of the free energy with respect to temperature. In the copper–gold system, the order–disorder transformations that occur are all first-order ones. However, that in the copper–zinc system, at the equiatomic composition, is a second-order transformation. In the former, the long-range order parameter changes discontinuously at the transformation temperature, while in the latter there is a continuous decrease in the long-range order parameter as the temperature is increased. The reason for these differing behaviors is related to the crystal structure of these superlattices. This subject has been considered in detail by Landau and Lifshitz (see Bibliography). Let us now examine examples of the various types of superlattices.

It may not be apparent, but for the unit cells shown in Figure 1.9 the arrangement of atoms on the basis lattice is such as to maximize the number of bonds between unlike atoms. However, although for many cases the driving force for long-range order may be a tendency to maximize the number of bonds between unlike atoms, other factors contribute to the formation of a superlattice arrangement of the atoms. For example, consider the Cu–Pt superlattice which consists of (111) planes containing only Cu atoms alternating with planes containing only Pt atoms. In this case, the number of bonds between unlike atoms is the same in the superlattice arrangement as in a random arrangement of the atoms on the fcc basis lattice. Thus, models of the thermodynamics of the long-range order phenomenon which limit their description of the energy to a sum over bond energies are likely to be insufficient descriptions in some cases. Modern models of the thermodynamics of partially disordered superlattices thus allow for the possibility of including many body terms in the energy.

We have given no indication how to express the energy or entropy of a partially long-range-ordered solid solution on the long-range order parameter. This is not a simple matter and has not been rigorously solved for 3Ds. However, good approximations have been developed. They are of two kinds. One makes use of computer simulation to evaluate the energy and entropy, and the other of analytic methods of approximating the energy and entropy. Reviews of the latter may be found in the articles by De Fontaine.[24]

In the simplest treatment of the order–disorder transition, the Bragg–Williams model, it is assumed that the distribution of atoms on one sub-lattice is independent of the distribution of atoms on the other sub-lattices. This is the type of approximation that belongs to the class called the mean-field approximation. In this mean-field approximation the probability of occupancy of one lattice site is assumed to be independent of the probability of occupation of any other lattice site. Thus, short-range order is neglected in the Bragg–Williams model. The result is a poor quantitative description of the order parameters. However, a review of the procedure used to evaluate the free energy of the partial long-range-ordered system is instructive and is given below.

Consider, the beta-brass system. In this case, there are two simple cubic sub-lattices, each one occupied by one of the two components, Cu or Zn, in the completely long-range-ordered superlattice. The nearest-neighbor sites of one lattice belong to the other lattice. An assumption in this model is that the energy is described as a sum over nearest-neighbor bond energies. There are four classes of such bonds: A/I–A/II, A/I–B/II, B/I–A/II, B/I–B/II, where A, B denote the two types of atoms and I,II denote the corresponding two sub-lattices occupied by these atoms, respectively, in the superlattice. Denoting the corresponding number of such bonds by N(1), N(2), N(3) and N(4), respectively, then

$$N(1) = Nz\rho(1 - \rho)$$

$$N(2) = Nz\rho^2$$

$$N(3) = Nz(1 - \rho)^2$$

$$N(4) = Nz\rho(1 - \rho)$$

where, N equals half the total number of atoms in this equiatomic system, z is the coordination number (number of nearest neighbors, equal to 8 in this case), and ρ is the fraction of A(B) atoms on the I(II) sub-lattice. Thus, the energy of the partially long-range-ordered system is

$$E = Nz\rho\,(1 - \rho)(E_{AA} + E_{BB}) + Nz(\rho^2 + (1 - \rho)^2)E_{AB}$$

The entropy of configuration corresponding to the Bragg–Williams approximation is based upon the assumption that the atoms on a given type of lattice sites are distributed randomly and is then

$$S = k \ln W = k \ln\{N!/[(N\rho)!(N(1-\rho))!]\}^2$$

Using Stirling's approximation for the factorial of large numbers, then

$$S = -2\,Nk(\rho \ln \rho + (1-\rho)\ln(1-\rho))$$

Since the energy of the superlattice is given by NzE_{AB} and the entropy of configuration of the superlattice is zero then the disordering energy and entropy are

$$\Delta E = -2Nz\rho(1-\rho)[E_{AB} - (E_{AA} + E_{BB})/2]$$

$$\Delta S = -2Nk(\rho \ln \rho + (1-\rho)\ln(1-\rho))$$

Minimizing the free energy $\Delta F = \Delta E - T\Delta S$ yields the equilibrium value of the occupation probability, ρ, and by its relation to the long-range order parameter η, the equilibrium value of the latter. Long-range order disappears above a critical temperature, equivalent to the Curie temperature in ferromagnetism, given by

$$kT_C = -4[E_{AB} - (E_{AA} + E_{BB})/2]$$

Figure 1.11 shows the equilibrium value of the long-range order parameter as a function of the reduced temperature in the Bragg–Williams model. The transition at T_C is second order. That this model is grossly inadequate needs to be restated, because it appears to describe the partial long-range order in one system, beta brass, at least qualitatively, fairly well. Better models take short-range order into account. In mean-field models, such as the Bragg–Williams model, the probability of an ij bond being of type A_i B_j, $p(A_iB_j)$, equals the product of the independent probabilities $p(A_i)p(B_j)$ that the ith site is occupied by an A atom and the jth site by a B atom. In the Bethe model, which takes short-range order into account this relation is not obeyed, but requires the independent parameter $p(A_iB_j)$. The Kikuchi cluster model extends the Bethe model to account for the probabilities of occupation of clusters of more than two sites. Both models also account for the overcounting, i.e. only one type of atom occupies a given site in a given configuration. A detailed description of these models is provided in the book of Ducastelle.[25]

Monte Carlo methods may be used to obtain good estimates of entropies of partially ordered phases. However, the accuracy of the energy value depends upon the model used to calculate it for a given configuration of the ordering units. The

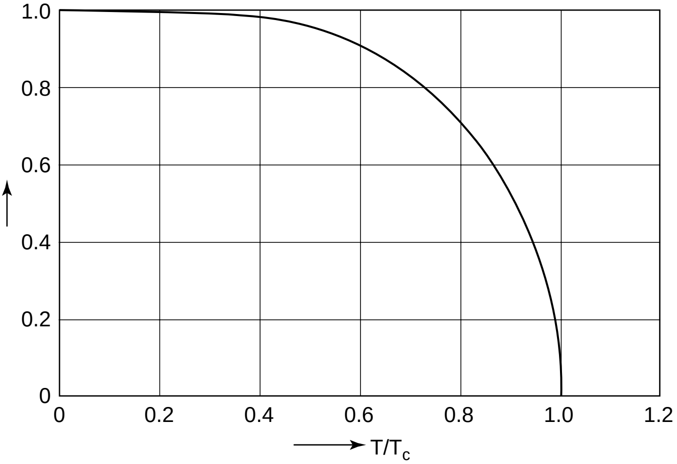


Figure 1.11. Long-range order parameter as a function of reduced temperature in the Bragg–Williams model. From J.D. Fast, **Entropy**, Philips Technical Library, 1962 with permission.

Monte Carlo method merely provides a configuration for this model to evaluate. In effect, this statement is valid for all the methods used in treating the order–disorder transition. Hence, let us consider methods for evaluating the energies of a disordered phase.

Many first-principle-based approximations for the energy of disordered solid solutions of alloy phases have been used in the past. In order to achieve the accuracy required for the comparison of ordered and disordered solutions at an invariant composition an impractical computer usage time is required for most of the first-principle methods. With the recent development of efficient computer codes for calculating the ground state energies of arbitrary configurations the most promising procedure for use in order–disorder problems is that involving cluster expansion. In one mode of using this procedure the coefficients in the cluster expansion for the energy of a configuration of some composition in a binary system are evaluated by fitting them to the energies of a sufficient number of configurations of different compositions. The latter values are first obtained using the ab initio codes. With values for these coefficients the energy of some disordered phase in the system may be evaluated. Thus, using a Monte Carlo procedure for generating configurations the stable configuration at some temperature can be evaluated. A detailed description of this procedure that has been automated for the general user is provided in Reference 26. At this writing, the applications of this methodology have been few in number, but the promise in its utility is such that there is little doubt that this number will start to increase exponentially.

The systems that involve large deviations in atomic size between constituents require corrections to the simple cluster expansion ab initio method described above. These corrections involve modifications of the relations to include the elastic strain energy in these systems, the differences in interatomic distances between the different pairs of components which corresponds to a relaxation of the energy from a common average interatomic distance, and the vibrational entropy. A detailed account of what is necessary to carry through an ab initio cluster expansion treatment to describe correctly the thermodynamics of such systems is given in Reference 27. With these corrections it appears that the calculated order–disorder transition temperatures are in good agreement with experiment for the Cu–Au system.

For both ab initio and semi-empirical modes of evaluating free energies of disordered and ordered solid solutions the effect of vibrational entropy has been neglected until recently. However, from the original study of Zener[28] and the recent work of Fulst[29] and others it has become clear that this parameter can have a significant effect on the relative stability of competing phases.

Until at least 2005, most evaluations of the energies of disordered or partially ordered solutions used interatomic potentials consistent with first principle for the material type of the solution and fitted to data that were consistent with experimentally measured parameters. For example, metallic solutions used pseudopotentials or embedded atom potentials, covalently bonded solutions used a potential suitable for this bonding type, such as a Stillinger or Tersoff potential, ionic solutions used Coulomb's potential between charged species and some repulsive potential, molecular solid solutions used potentials corresponding to the main bonding mode (van der Waal, hydrogen, dipole).

Mean-field models are currently the predominant ones used in the investigation of order–disorder transitions in mineral systems. Such disorder–order transitions are ubiquitous in mineral systems. For example, Al/Si ordering on tetrahedral sites occurs in the aluminosilicates, Mg/Ca ordering is found in garnets, and simultaneous Al/Si and Mg/Al ordering in pyroxenes, micas and amphiboles. Computer codes using models for interatomic potentials, covalent bond angle potentials, long-range Coulomb interactions between charged entities, and electron polarizability via a shell model for the oxygen anion have been developed, such as the GULP lattice energy code (http://gulp.curtin.edu.au/), to minimize the lattice energy varying atomic coordinates and unit cell parameters. These codes already incorporate parameters in the potentials that have previously been obtained by fitting to a variety of data. These models, such as GULP, are good when the ordering does not involve changing the coordination number of a component (i.e. by transferring an atom from a tetrahedral to an octahedral site). When the ordering process does involve such a transfer then it is necessary to use a first-principle-based model to evaluate the energies of the configurations. The systemization of application of ab initio codes to evaluation of mineral ground state energies has not reached the stage it has for the alloy systems.

A few examples illustrating current treatments of the order–disorder competition for stability in several different classes of materials are briefly described in the following.

An ab initio cluster expansion study of the Ti–Al system, but one that neglects elastic strain and vibrational entropy, yielded the result that the predicted order–disorder transition temperature for the DO_{19} Ti_3Al phase was 1850°K as compared to the experimental value of 1450°K. The transition is a first-order one. It was stated that possible reasons for this discrepancy are neglect of vibrational entropy, neglect of elastic strain in both energy and entropy, and the limited precision of calculations based on the local density approximation method (LDA). Since it was found that the change in vibrational entropy on ordering in the Cu–Au system is on the same order as the configuration entropy change, one may expect a large effect on the order–disorder transition temperature due to the change in vibrational entropy. Hence, even for ab initio procedures we need to take into account all contributions to the free energy in considering order–disorder transitions given the small difference expected in the free energies of the competing configurations.

For the mineral muscovite the order–disorder transition involves the redistribution of Al and Si atoms on tetrahedral sites. In modelling this transition use was made of the Bragg–Williams model for the interactions between first, second, third and fourth neighbor sites. The parameters for this model were obtained by fitting procedures which are described in detail in Reference 30. These parameters were then used in a Monte Carlo procedure to evaluate various quantities, including the ordering parameter. The result for the order parameter within one sheet of tetrahedrons versus temperature is shown in Figure 1.12 along with a representation of the long-range-ordered structure in the tetrahedron sheet. In the disordered state the Al and Si atoms are distributed at random on the tetrahedral sites.

An order–disorder transition may occur in strands of DNA such as one having the sequence: 5′GAGATGACTCATCTC3′. It is apparent that this strand could form a hairpin configuration since the bases reading in from both ends are complementary. Calorimetric and circular dichroism spectroscopy evidence confirms the suggestion that a transition from an ordered arrangement to a random coil configuration occurs upon heating such a strand of DNA.[31] This writer is not aware of any model of this transition in the literature. However, similar transitions from globule to random coil in polymers have been modelled.

2.8. Polymorphism in oligomers and polymers

Phase transitions occur in oligomers and polymers. Figure 1.13 represents one phase diagram provided by a simulation showing the effect of van der Waal's bonding between mers along the polymer via the bonding parameter ε and of the stiffness of the bond angle bending via the parameter κ. These phase transitions are between

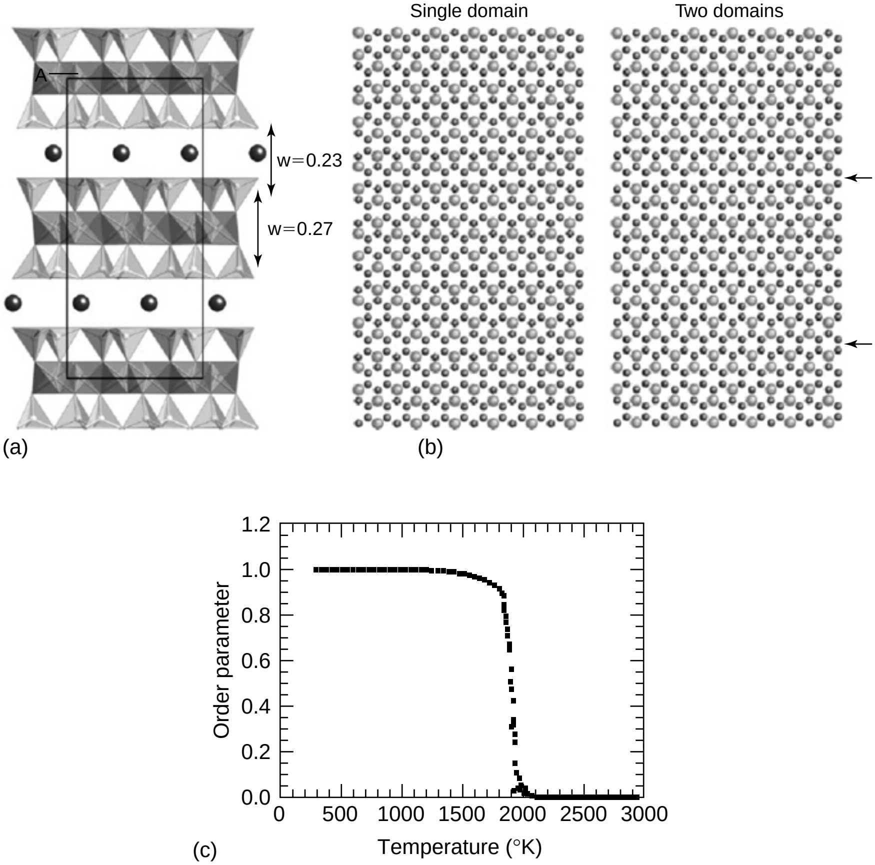


Figure 1.12. Showing the unit cell of muscovite in (a), the arrangement of Al (large) and Si (small) atoms in the topmost sheet of tetrahedral (b) with the Al atoms situated in plane AA in the stable long-range-ordered structure and in (c) the order parameter versus temperature as deduced in a Monte Carlo procedure revealing a second-order transition at about 1800°K. From Phys. Chem. Mineral. 28, 534(2001) with permission. © 2001 Springer-Verlag.

various thermodynamic states of isolated oligomer or polymer macromolecules. Both experimental techniques and computer simulations have contributed to our knowledge of these phase transitions and thermodynamic states. One experimental system in which isolated macromolecules are found is that of a dilute solution of such macromolecules. The effect of the solvent can be neglected if there is no interaction between macromolecule and the solvent species. However, this is not always the case, especially in water as a solvent will be illustrated. We shall assume

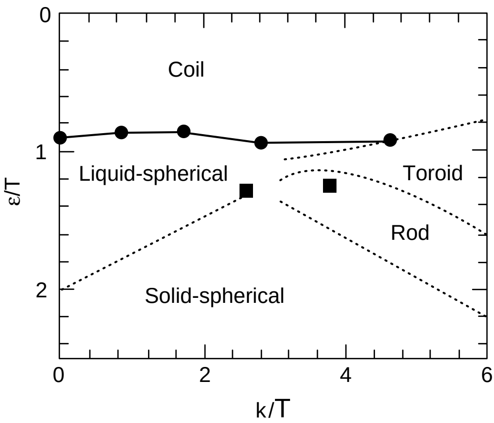


Figure 1.13. Phase diagram showing effect of square potential well depth ε and bond bending stiffness κ on stability of polymer. From J. Chem. Phys. 109, 5070(1998) with permission. © 1998 American Institute of Physics.

implicitly in the following that the solvent species can be neglected, but when it cannot we will consider the effects of the solvent explicitly.

In both oligo- and polymeric materials the first stage in the self-organization process to a more condensed state involves the folding of a chain of mers into various configurations. Folding of a polymer chain can produce a variety of structures and be involved in phase transitions of the polymer chain. The first phase transition of a random coil on decreasing the temperature or equivalently altering the properties of a solution containing random polymer coils is either to a glassy globule, or to a lamellar crystalline state, or in certain oligomers to a condensed helix, disk, closed ring or other geometric unit. We shall investigate the factor(s) responsible for the difference in structures of these condensed states. The coil to glassy globule transition occurs at the temperature T_θ. The globule may or may not undergo additional phase transitions at lower temperatures. We shall investigate these possible transitions and their causes. These polymer coil to globule transitions have been of interest to the biological community because of the analogous transitions that occur upon protein folding.

2.8.1. Phase transitions of isolated polymers and oligomers

It was determined some time ago that the free coil, collapsed globule and lamellar crystal represent different thermodynamic states.[32] Sometime later this proposal was examined quantitatively and found to be valid in that the coil to collapsed globule is a phase transition of the second order with the collapsed globule in a disordered state[33]. This conclusion is supported by numerous simulations and experiment.[34] However, coils with sufficient stiffness can undergo a first-order phase transition from the coil to a lamellar crystalline state.[35] Although this transition was demonstrated in a computer simulation the existence of lamellar crystals of polymers cooled from the melt or precipitated from a solution is beyond doubt.[36] Other possible phase transitions on decreasing the temperature below T_θ have been found in some of the simulations. One such, for the case of chains with stiffness, is from the collapsed globule to a crystalline state.[35,37] The collapse transition to a non-crystalline disordered state was suggested to be equivalent to a gas–liquid transition while that from the collapsed state to a crystalline or ordered state is equivalent to a liquid to solid transition.[38] Further possible polymorphic transitions of the displacive type are also suggested in this study.[38] Not all of these transitions have been found experimentally although evidence for some of them have been reported.[34] In the following subsections we will consider in detail each of these phase transitions separately.

2.8.1.1. Coil to globule transition

There are at least two mechanisms proposed for the details of the collapse process. One due to de Gennes[39] proposes the prior formation of sausages along the coil before the final production of a collapsed globule. This process is shown in Figure 1.14 taken from a computer simulation[40] of the dynamics of the process of collapse. The other mechanism is termed "crumpled globule" model by its proposers.[41] The latter is one in which the collapse takes place in a fractal mode and is illustrated in Figure 1.15.

The driving force for coil collapse is an effective attractive force between monomers of the polymer chain. In the de Gennes model in the initial stage the monomers rapidly form adjacent blobs along the chain with the density of monomers near to the equilibrium value of the globule. These blobs then grow by incorporating

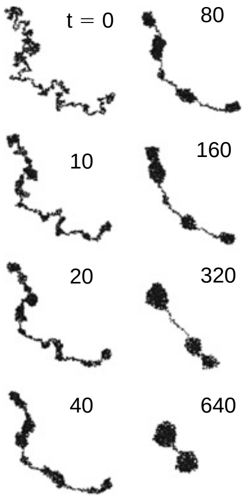


Figure 1.14. Showing successive stages in an MD simulation of collapse of a homopolymer coil. From Europhys. Letters. 59, 391(2002) with permission. © 2002 EDP Sciences.

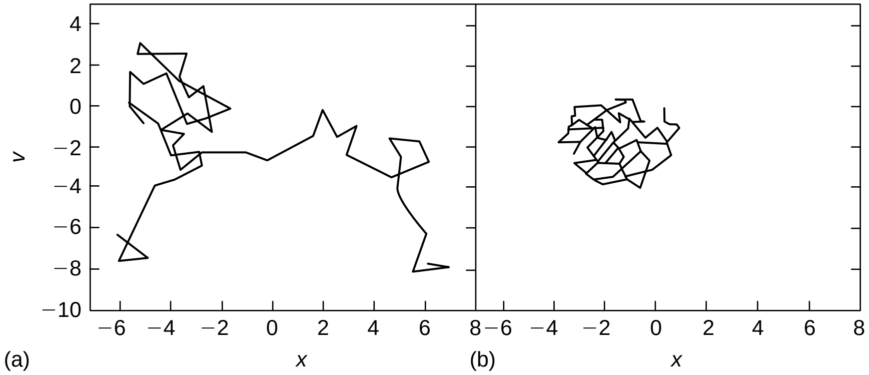


Figure 1.15. Simulation of coil and globule configurations. From J. Chem. Phys. 103, 2615(1995) with permission. © 1995 American Institute of Physics.

neighboring blobs as illustrated in Figure 1.14. A characteristic of this mode of collapse is that monomers that are distantly separated from each other along the chain are likely to be nearest neighbors in the globule. Although this phase transition has the appearance of a nucleation and growth process it isn't one. The formation of nodules and their growth does not occur isothermally. Rather, it occurs with decreasing temperature with each stage representing an equilibrium configuration, i.e. the transition is second order and not first order.

In the crumpled globule model nearest-neighbor monomers along the chain tend to be nearest neighbors in the blobs that form in the crumpling process. In the second stage of this process, chain quasi-knotting occurs while the crumpled globule continues to shrink through topological relaxation to the final equilibrium state. Hence, in this model, nearest-neighbor monomers along the chain tend to be nearest neighbors in the equilibrium state. Figure 1.15 shows two simulated configurations of the polymer chain, one at an early stage of the crumpling process and the other at a stage close to the equilibrium state of the crumpled globule. The crumpling process in this model corresponds to an increase in the number of folds along the chain as the ends of the polymer chain come towards each other. Experimental results that distinguish between the two models are not available at this writing. For both models, however, the phase transition produced by simulations is second order *as long as the attractive driving force is between monomers alone.*

In the above phase transitions of an isolated homopolymer, as the temperature decreases the entropy associated with each succeeding phase is less than that of the phase stable at the higher temperature. This situation also holds for the polymer *N*-isopropyl-acrylamide (PNIPAM) in water. However, because the entropy of dissolution of this polymer in water is negative, the phase with the higher system entropy is the globule while the coil state has the lower system entropy! Thus, the coil to globule transformation requires the temperature to be increased from that in which the coil state is stable. The coil to globule and globule to coil transitions for isolated PNIPAM in water have been studied.[42] The results indicated that two additional transitions take place in the process of condensing the coil state. The additional thermodynamic states found are illustrated schematically in Figure 1.16. The final globule has a density

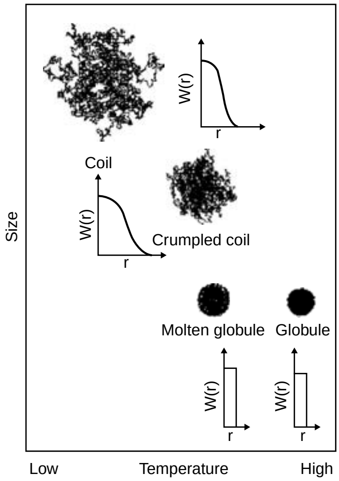


Figure 1.16. Schematic illustration of the states found in the collapse from coil to globule of PNIPAM. From Macromolecules. 31, 2972(1998) with permission. © 1998 American Chemical Society.

of about 0.34 gm/ml, which suggested to these authors that this configuration still contains within it about 66% water. As shown, there also exists a crumpled coil and a molten globule state. It is interesting to note that these investigators' description of their view of the crumpled coil corresponds to the de Gennes description of sausages along the chain. However, in their case the crumpled coil corresponds to a thermodynamic state and not to an intermediate stage in the collapse process. It is not known whether this behavior is general or is limited to polymers which produce a decrease in entropy on dissolution and thereby include solvent molecules within the condensed states. However, there have been simulations which reveal a similar sequence of thermodynamic states during the collapse from the coil state.

The results of various simulations for "normal" high temperature coil to low temperature globule states indicate agreement on the following features of the transitions. First, there is a marked reduction in the radius of gyration during the collapse. For polymer lengths exceeding a few hundred mers the decrease in entropy that accompanies this collapse is about 40–50 J/mol-mer/°K. As polymer lengths decrease the decrease in radius of gyration occurs over a wider temperature range below T_θ. In the absence of stiffness in the polymer chain the state at the completion of the collapse is disordered and liquid like in that it supports fluctuations in density.[43] Further decrease in temperature in this case can lead to the development of order[44] or to a denser solid glassy state[43]. For some simulations the ordered phase appears upon a first-order transition.[38] For other simulations the transition to a more ordered state involves a second-order transition.[44] The transition to the denser solid glassy phase is first order.[43] In the presence of sufficient chain stiffness the collapse of the coil directly into a lamellar crystal may occur without the intermediate disordered condensed state.[35]

In the absence of a negative entropy of dissolution the translational entropy of the isolated coil in solution equals that of a self-avoiding random walk coil and is higher than that of the corresponding condensed states. The thermodynamic state of an isolated coil has been the subject of many studies[32,33] over the years and is by now well known. The main concepts are that chain connectivity is responsible for long-range correlations between the bonds, the coil obeys the statistics of a self-avoiding random walk and collapses below the temperature T_θ when the effective persistence length* approaches unity in a good solvent. At and below this temperature Gaussian conformations describe the coil state. The entropy of translation of the chain is given by $k \ln \Gamma_N$ where Γ_N is the number of configurations of a self-avoiding walk of N steps. From the theory of self-avoiding walks in the ideal gas approximation (see de Gennes[32]), $\Gamma_N = cN^{\gamma-1}z^N$, where $\gamma = 7/6$ in 3D and z should be somewhat less than 108 due to excluded volume constraints. A comparison of the entropy given by the above with an exact numerical calculation for $N \leq 6$ yielded $c = 1.26$ and $z = 85.2$.[45] The entropy of translation is then about 37 J/°K/mol-mer. As noted above the entropy change in going from the coil state to the solid glassy state is between about 40 and 50 J/mol-mer/°K.[44] Further the configurational entropy change in the first-order transition between the liquid-like and solid glassy states is about 0.4 J/mol-mer/°K.[43] Finally, the entropy of a globule produced from a single coil is not very high. A mean-field theory of a small globule[46] yields a maximum value of $0.35\,k_B$ or about 3 J/°K/mol-mer. The relation that yields this result has the form in 3D of

---

* The mean fractional length of the coil that is linear.

$$S/N = -k\ln k - (1-k)\ln(1-k) - sk + (3/2\ln 2)k\ln k$$

where k is the number of contacts between mers per mer ($=K/N$), $-sk$ is the entropy lost in forming k contacts and the last term is the entropy lost in forming k loops. A more rigorous derivation of the entropy for both small and large collapsed globules can be found in Grosberg and Kokhlov.[33] However, the consistency between these values from different sources is heartening.

Treatments of the coil to globule transitions are complex and outside the scope of this book. We have discussed the liquid-like and solid glassy states, as well as the possibility of more ordered states assembled in the aftermath of a coil collapse. Among the more ordered forms are lamellar crystals, as well as ordered arrays having a variety of geometric characteristics.

The coil collapse is initiated by the transition between self-avoiding random walk conformations and Gaussian conformations at T_θ. The additional transitions have as their driving force an effective attraction between mers or bonds.

2.8.1.2. Coil to lamellar crystal transition

There are several questions we will attempt to answer in this subsection. These are: Will an isolated polymer chain crystallize? What factor(s) are responsible for the formation of polymer crystals? Why is the shape of the polymer single crystal grain lamellar?

The answer supplied by computer simulations to the first question is yes. Simulations of the coil → globule → lamellar crystal transition have shown that the intermediate globule state can be bypassed when the polymer coil has sufficient stiffness. Without such stiffness the transition to a crystal from either the coil or globule state cannot take place. This stiffness need only be manifested as a resistance to folding of the polymer chain. A phase diagram revealing this effect of chain stiffness is shown in Figure 1.17. Here e_g and e represent the energies of kinks of “gauche bonds” and of an effective mer–mer contact energy including polymer–polymer, polymer-solvent and solvent–solvent interactions. The total energy in this model is given by

$$E = -n_{pp}e + n_g e_g$$

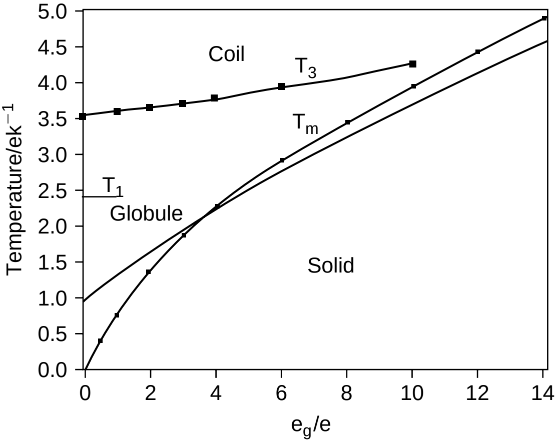


Figure 1.17. Phase diagram of a 343-unit polymer. T_q is the collapse temperature of the coil, T_M is the disorder–order transition temperature between the globule and the ordered solid phase. Abscissc is ratio of e_g/e as below. Ordinate is normalized temperature. From J. Chem. Phys. 108, 2134(1998) with permission. © 1998 American Institute of Physics.

where n_{pp} is the number of polymer–polymer contacts and n_g is the number of kinks in the polymer chain. This model is thus a simplified representation of a semi-flexible polymer in solution. In this phase diagram it should be noted that both phase transitions are second-order except for the coil to crystal transition for $e_g/e > 13$, which is a first order one.

What drives the crystallization of a polymer chain? Many polymer chains remain as disordered solid glasses at low temperature and do not transform to a crystal. Chain stiffness obviously provides for an increment in energy that opposes a decrease in free energy due to the enhanced entropy of the disordered glass relative to a crystal. The crystal however can have a lower total energy than the disordered glass due to a greater number and strength of bonding interactions providing that the crystal has sufficiently long straight lengths of chain to minimize the positive energy of strained bonds.

At this point it will be instructive to consider the coil to crystal transition of a long isolated polymer chain. In practise, crystallization usually occurs as a heterogeneous nucleation event. (We will explore the theory of heterogeneous nucleation in Chapters V and IX.) The first stage involves contact between the coil and a substrate surface. This contact could occur first either at the end of a coil or at some mer along the coil. There are many reaction paths between the isolated polymer coil in dilute solution state and that adsorbed state we will call the activation state. We shall consider one such state as follows. Any mer, not secondarily bonded to another mer along the polymer chain, that makes contact with the surface loses most of its entropy of translation, but gains bonding energy with the surface (energy of adsorption). At high temperature the entropic induced positive increment to the free energy (due to the loss of translation entropy of the mers contacting the surface) is larger than the decrease due to the energy of adsorption. Hence, the step of adsorption of any mer, not secondarily bonded to another mer, onto the surface must result in an increase in the free energy. A contact between two separated adsorbed mers along the chain decreases the free energy by the free energy of the secondary bond formed in the contact. As successive adjacent mers along the loop contact, the loop starts to close. Spontaneous closure of the loop will occur when the free energy continually decreases during loop closure. Such spontaneous closure will occur when the free energy decrement due to the secondary bond is larger in magnitude than the sum of the free energy increment due to translational entropy loss and strain energy increment due to increment in the curvature of the loop with closure of the loop. It is important to note that this increment per contact made depends upon the length of the loop, i.e. when the loop is completely closed the total increment in strain energy (corresponding to a surface energy) obtained is the same independent of the original loop length.

If only the mers on the first stem and the mer defining the loop and the first point of contact with the stem are adsorbed prior to the start of loop closing then the equation defining the minimum loop length for spontaneous loop closure is

$$\Delta G = (n + 1)(g_b + g_{ads} - T\Delta S_{tr}) + \Delta g_{st} \leqslant 0$$

where g_b is the free energy of a mer–mer bond, g_{ads} is the free energy of adsorption of a mer, ΔS_{tr} is the loss of translational entropy per mer due to mer adsorption onto the surface and then formation of a bond and Δg_{st} is the strain free energy per mer during loop closing. Thus, there is a minimum number of mers per stem n^*, or stem length, needed for spontaneous decrease in the free energy after adsorption of a length larger than this minimum length which decreases as temperature decreases, defined by setting $\Delta G = 0$ in the above relation. Further, in the above model, the minimum work of nucleating a crystal equals $(n^* + 1)(g_{ads} - T\Delta S_{tr})$. (In the above the g values and ΔS_{tr} are negative quantities while Δg_{st} is a positive quantity.)

As mentioned above there are many possible reaction paths. The most likely is the one yielding the smallest work of nucleation. Thus, in the reaction path described above, the rise in free energy from the origin represents the difference between the decrement due to loss in translational entropy and the increment due to the energy of adsorption of the mers. The drop in the free energy represents the process that occurs during the bonding of the second stem to the first stem. It will be a jagged plot because an adsorption step produces a net increase in the free energy, whereas a mer–mer bonding step decreases the net free energy. However, the latter step is a function of the number of mers in a stem because the strain energy and loss in translational entropy in the closed loop between the stems must be overcome and shared by the bonded mers between two stems. Hence, a plot of the free energy versus number of adsorbed mers would appear as shown in Figure 1.18.

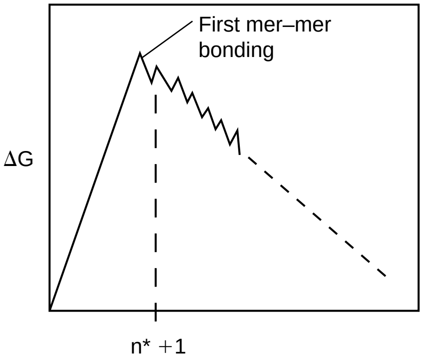


Figure 1.18. Schematic free energy versus number of mers adsorbed on surface.

According to the above concept, crystallization of the isolated polymer coil is favored by bond stiffness because the latter acts to decrease the effective translational entropy per mer of the coil. This effect results in a decrease in the work of nucleation of a crystal since this translational entropy is the effective barrier defining the work of nucleation. It also acts to increase the strain energy associated with the loops formed in the globule between mer–mer contacts and hence, to make the free energy of the equilibrium globule more positive. Since these loops are replaced by bonded stems in the crystal, the free energy of the crystal becomes more stable relative to the globule as the stiffness increases. In effect, the foregoing is the explanation of the phase diagram shown in Figure 1.17. Incidentally, the minimum lamellar thickness, which is proportional to n*, also decreases with decreasing magnitude of the translational entropy per mer of the coil.

At this point it is useful to consider the magnitudes of the quantities in the above equation. We already know that $T\Delta S_t$ will have a value at 300°K of up to 12 KJ/mol-mer. Let us consider the other quantities for polyethylene shown in Figure 1.19. From the figure there appear to be two hydrogen bonds to the carbon atom of the mer and one each per hydrogen atom of the mer. Since these bonds are each shared by two atoms there are a total of two hydrogen bonds per mer. This corresponds to a value of g_b of about -2.72 KJ/mol-mer and g_{ads} of -2.72 KJ/mol-mer. This value of the total binding energy is insufficient to overcome the self-avoiding random walk translational entropy loss term at 300ºK of a coil lacking stiffness of about 12 KJ/mol-mer. However, for a stiff coil this entropy loss term would be reduced. In fact, the entropy of fusion of polyethylene is 9.8 J/mol-mer/°K. We may take this value to be a good approximation to ΔS_{tr}. Hence, we need to compare 300×9.8 or 2.9 KJ/mol-mer to the total of -5.4 KJ/mol-mer. Thus, it appears that hydrogen bonding is sufficient to provide the bonding needed to crystallize polyethylene. From the above analysis it appears that hydrogen bonding alone will often be insufficient to bring about condensation of unstiff polymers. Other sources of bonding are then required. Table 1.5 provides a list of such bonding sources and estimates of associated bonding energy.

With the above we have provided answers to two of the questions posed at the beginning of this sub section. Yes, it is thermodynamically feasible for an

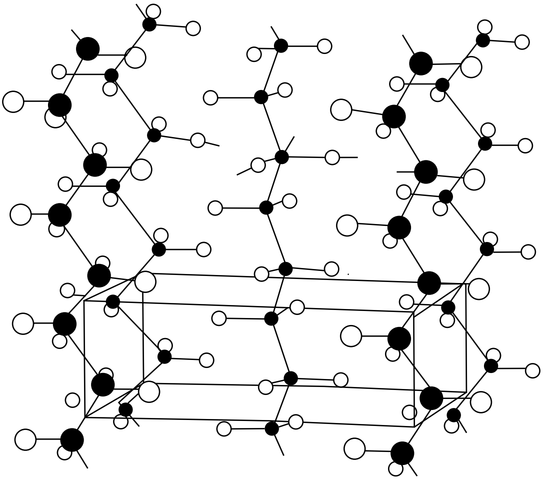


Figure 1.19. Illustration of polyethylene stems in crystalline arrangement showing the hydrogen bonds about one mer.

Table 1.5. Bonding sources and energies.

Type	**Bonding energy (KJ/mol)**
Ion–ion	∼58
Ion–dipole	−8 to +8
Dipole–dipole	−2 to +2
Ion–induced dipole	0.25
Dispersion and stacking of aromatic rings	0 to 40
H-bonding, per H bond	−5.4 ± 2.5
Hydrophobic, per CH_2	−5.4 ± 2.1
Hydration force, per residue	−0.4 to 0.4

isolated polymer of sufficient length and stiffness to crystallize. The factors driving the formation of polymer crystals are mer–mer bonding energy and mer surface adsorption energy. The factors opposing crystallization of polymer coils are the translational entropy per mer of the coil and the strain energy associated with the "bent" bonds at the ends of the stems. We will now attempt to provide an answer to the third question: Why is the crystal shape lamellar?

It is apparent that the minimum thickness or length of stem will not propagate a lamella because there is no driving force for spontaneous growth of parallel rows of stems of minimum thickness length. The driving force and rate of such growth increases with stem length beyond the minimum thickness length. However, the probability of obtaining an initial adsorbed stem length decreases with stem length. Hence, there will be some stem length larger than the minimum length at which the rate of growth of a folded 2D bonded sequence of stems will be a maximum. Nucleation and growth of another layer of folded stems can take place on top of the first layer to allow growth in a direction perpendicular to the 2D layer of folded stems. Nucleation of such a layer on the surface adjacent to the folds at the end of the chain stems is not likely because mer–mer bonding of fold to fold is hindered and weaker than the bonding of straight chain stems to each other. Hence, growth in a direction normal to either the (110) or $(\bar{1}10)$ planes will occur faster that along the *c* direction in Figure 1.20. These factors then account for the lamellar nature of the crystal shape finally produced.

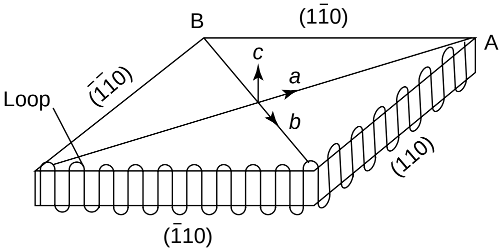


Figure 1.20. A schematic illustration of a lamellar crystal, such as polyethylene. The surface showing the stems is a {110} surface for the latter crystal.

There still remains one other factor that affects the probability of crystallization of a polymer chain. Does the topology of the polymer allow for easy or difficult formation of a loop between adjacent chain stems at the surface. The loop just referred to is shown schematically in Figure 1.20. A measure of the difficulty of such loop formation is the energy of the surface formed by such loops. Fortunately, computer simulations are able to provide estimates of this surface energy. One such estimate for polyethylene[47] is $108 \pm 30\,\mathrm{erg/cm^2}$. The average number of bond lengths in such loops was evaluated to be less than 10 and most of such loops were in the {110} plane, i.e. in the plane parallel to the surface being deposited for polyethylene. Contributing to this low surface energy is the relative ease of rotation about a primary bond. Thus, rotation of the bond with a minimum of steric hindrance allows a short loop to result with relatively low excess energy. Of course, there are many polymers which cannot easily form such loops and, hence, the ground states for such polymers do not include the lamellar crystal. However, other ground states are possible for these polymers, some of which are considered in the following sub section.

Lamellar crystals can also be formed from short chains of monodisperse oligomers.[45] Folds are not necessary to produce lamellar crystals when the chain lengths are less than l_{min}. Figure 1.21 shows the results of a computer simulation of the deposition of monodisperse oligomers each containing 8 mers onto a seed crystal.

In conclusion, the reader is warned that although the foregoing describes the essential physics governing the formation of lamellar polymer crystals, others have different views concerning the reaction path or the modes of describing the thermodynamic factors involved in such crystallization.[48] The standard mode in textbooks follows the lead of Lauritzen and Hoffman in using standard heterogeneous nucleation theory as applied to the lamellar crystalline embryo. However, this approach masks the physics. For example, the role of chain stiffness in promoting crystallization is not apparent although with thought its role can be elucidated in the parameters of nucleation theory. In any case, the approach used in this sub section clarifies the roles of the thermodynamic and kinetic factors affecting polymer crystallization.

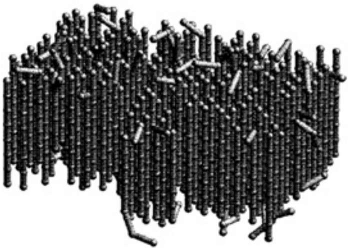


Figure 1.21. Extended chain crystal deposited from dilute solution in Monte Carlo simulation. From J. Chem. Phys. 108, 4305(1998) with permission. © 1998 American Institute of Physics.

2.8.1.3. Coil to helix transition

Whereas the coil to crystal transition occurred in nature before it was discovered by man the coil to helix transition, for the case where the secondary structure is not torsionally predetermined in homopolymers, was designed and accomplished by man although its occurrence in proteins was well known previously. This fact suggests that a design to produce the coil to helix transition in homopolymers can be described. Let us attempt to provide such a description at this point keeping in mind our interest in the thermodynamics and kinetics of this process:

1. *The attainment of backbone helical topology is facilitated by:*
   - *(a) backbone meta-connectivity,*
   - *(b) easy rotation about primary bonds.*
2. *The coil to helix transition is facilitated by:*
   - *(a) maximizing area of solvophobic inner circumference regions per degree of freedom,*
   - *(b) pi–pi attraction between overlapping aromatic rings,*
   - *(c) maximizing van der Waal type interaction.*

Following such a design strategy Moore and his students[49] made use of prior knowledge concerning phenylacetylene macrocycles to choose *m*-phenylacetylene as the backbone for their further attempts to induce a coil to helix transition in a homopolymer. Their success yielded the first reversible homopolymeric coil to helix transition ever observed. It is to be emphasized that this transition is from a random coil, such as shown in Figure 1.22 as "unfolded", to the helical "folded" state, as accomplished by changing the solution from the "good" state to the "poor" state, i.e. by altering the chloroform concentration in acetonitrile from 100 to less than 60 vol%. This effect of the solution on the transition cannot occur in torsionally determined homopolymer structures and represents the effect of a solvophobic interaction on the topology of the oligomer in analogy to the effect of solution p_H

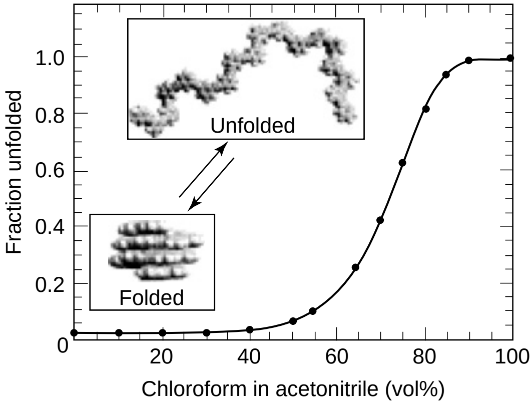


Figure 1.22. Fraction of phenylacetylene oligomer unfolded as revealed by titration curve. From MRS Bulletin, April 2000, Fig. 3, p. 39 with permission.

on secondary structure in proteins. The sigmoidal nature of titration curves such as that in Figure 1.22 is well-known to indicate a two-state transition. In this case, the transition is between the coil and helix states.

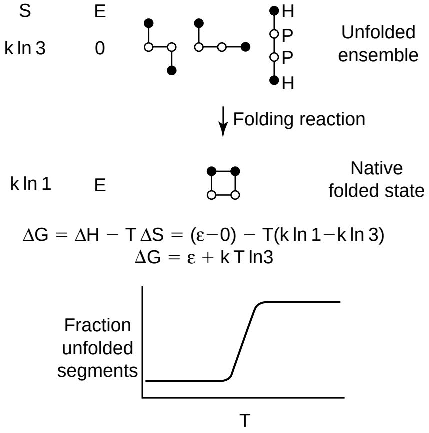


Figure 1.23. 2-D lattice tetramer toy model. From Chemical Reviews 101 (12) 3893(2001) with permission. © 2001 American Chemical Society.

A simple thermodynamic analysis of the chain–helix transition is illustrated in Figure 1.23. Consider the polymer consisting of segments of solvophobic mers (H) and polar residues (P) arranged as shown in the figure. For every four mers involved in this 2D chain there are only three distinguishable configurations having the same energy. The fourth configuration, that of the helix, has a lower energy ε. Hence, the thermodynamics of the transition of coil to helix is simple and given in the figure. The transition is defined of course by $\Delta G = 0$. This model can be generalized to yield the equation

$$\Delta G = \varepsilon + kTN \ln \omega$$

where N is the number of mers along the chain and ω is the number of distinguishable conformations available to each segment.

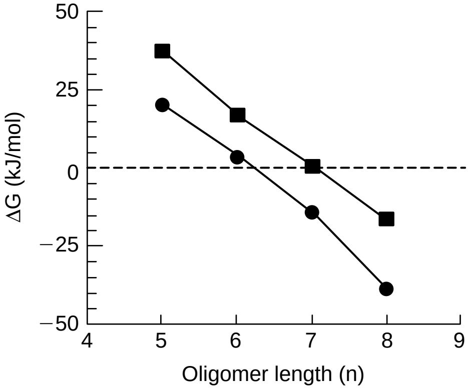


Figure 1.24. Calculated free energy of helix minus that of coil vs oligomer length for side chain R = CH_3 in chloroform and R = H in water. From Science. 277, 1793(1997) with permission. © 1997 Science.

Nelson in his thesis[49] evaluated the difference in free energy between coil and helix states for phenylacetylene oligomers using standard analytical chemistry procedures and models. His result is shown in Figure 1.24. The constant negative slope of this curve suggests that the addition of a mer to the oligomer length enhances the stability of the helix relative to the coil by a constant amount. It is apparent from this figure that mer lengths greater than ten are needed to complete the coil to helix transition in a solvophobic solvent. Since overlap occurs at mer lengths longer than five this result suggests that more than five overlaps are required for the solvophobic effect.

There are packaged programs which can yield approximate values of thermodynamic parameters for the various oligomers. However, educated guesses based upon the broad principles cited previously will often work as well to predict the relative stability of coil and helix structures. The small differences in free energy needed to bring about phase transitions in a short time suggest that the nucleation barriers, if they exist at all, are small so that the transition is reversible without the onset of hysteresis. Kinetic studies[31] reveal that the coil to helix transition can take place in oligomers without lengthy side chains within several tens of nanoseconds. Hence, there is no need for catalysts to speed up the coil to helix transition in simple oligomers or peptides. Nucleation is not the problem in this transition as it is in the coil to crystal transition.

3. Summary

To summarize, in this chapter we have introduced the concept that the most stable phase, at constant temperature and pressure, has the most negative value of the Gibbs free energy. Also, we have recalled that at equilibrium any system, at constant temperature and pressure, has its lowest possible value for the Gibbs free energy. We have found it possible to partition the various thermodynamic potentials into various categories according to their origin. For example, the entropy *can be partitioned into thermal and configurational contributions. We have found that the vibrational entropy of an ordered array of hard spheres can be larger than that for a random array. Also, at equilibrium, a system constrained to a constant composition proceeds, from low to high temperature (from low to high pressure), through a sequence of phases in such a way that the entropy (volume) continually increases (decreases). If the increase in entropy (decrease in volume) is discontinuous at distinct temperatures (pressures) corresponding to transitions from the regime of stable equilibrium of one phase to that of another, the transitions are of first order. For first-order transitions, it is possible to represent the regimes of phase stability in a P–T phase diagram. The Clapeyron equation yields the dependence of the transition temperature on pressure for polymorphic transitions. We have also defined the state of thermodynamic equilibrium and have been introduced to the concepts of long- and short-range order, the equilibrium values of which are determined by minimizing the free energy with respect to the order parameter. Further, although all materials at equilibrium obey the thermodynamic relations the values of the thermodynamic parameters differ between them. The origins of polymorphism in various materials, including soft materials, were investigated from the viewpoint of their effects on the energy and entropy of the systems concerned.*

Appendix 1

Strategies for calculating free energies

The preferred method to obtain needed free energy values is that involving the least computational burden and that provides the desired accuracy. In the order of increasing computational burden are: the method of Calphad which uses experimental data to extrapolate into the phase space of interest; the class of non-ab initio methods which are empirically or semi-empirically based schemes of predicting the desired thermodynamic potentials; the class of ab initio and first-principle procedures. In the second scheme each class of materials has its own group of predictive schemes. Also, the first-principle procedures used for different classes of materials often differ. Hence, we consider the procedures according to the class of materials. Also, since the Calphad procedure is so well established and codified we refer readers to the book by N. Saunders and A.P. Miodownik, **Calphad, (Calculation of Phase Diagrams): A Comprehensive Guide** Pergamon Press, Oxford, UK, 1998, for instructions as to its use.

Alloys, intermetallic compounds and inorganic compounds

The first (post Pauling) widely used method of predicting heats of formation is that of Miedema. (A.R. Miedema, R. Boom and F.R. de Boer, J. Less Common Met. 41, 283(1975), 45, 237(1976), 46, 67, 271(1976); F.R. de Boer, R. Boom, W.C.M. Mattens, A.R. Miedema and A.K. Niessen, **Cohesion in Metals: Transition Metal Alloys**, eds. F.R. de Boer and D.G. Pettifor, North-Holland, Amsterdam, 1988.) The accuracy of the predicted values varies from good to poor depending upon the combination of elements involved. Recently, Chen, Podloucky, Rogl and Wolf (to be published) found that correcting an electronegativity value for Zr improved many of the poor predictions.

Energies of formation may be predicted using interatomic potentials fitted to various physical parameters. For simple metal systems (those excluding transition elements and other elements likely to yield covalent bonding) central force type interatomic potentials are adequate. Of these, three may be mentioned: Finnis–Sinclair potentials (M. Finnis, and J.E. Sinclair, Phil. Mag. A, 50, 45(1984)), embedded atom potentials (M.S. Daw, and M.I. Baskes, Phys. Rev. B 29, 6443(1984)) and electronegativity modified Mie potentials (E.S. Machlin, Mat. Res. Soc. Symp. Proc. 19, 67, (1983)). For application to alloy systems containing transition metals non-central potentials are necessary. Among these the following may be cited: J.A. Moriarty, Phys. Rev. B, 42, 1609(1990); J.A. Moriarty, and M. Widom, Phys. Rev. B, 56, 7905(1997); S. Znam, D. Nguyen-Manh, D.G. Pettifor and V. Vitek, Phil. Mag. 83(4), 415–438(2003). A collection of interatomic potentials for various materials may be found in A.M. Stoneham and R. Taylor, **Handbook of Interatomic Potentials**, AERE-Harwell, Osfordshire, 1981. Interatomic potentials useful for predicting lattice stabilities and energies of solution in mineral systems may be found in A. Bosenick, M.T. Dove, E.R. Myers, E.J. Palin, C.I. Sainz-Diaz, B.S. Guiton, M.C. Warren, M.S. Craig and S.A.T. Redfern, Mineralogical Magazine, 65(2), pp. 193–219(April 2001). Another method based on defined electronegativities can be used for clay type minerals (P. Vieillard, Clay Clay Mineral. 50(3), 352(2002)). A basic reference to the treatment of ionic solids is G.R.A. Catlow and W.C. Mackrodt, in **LECTURE NOTES IN PHYSICS**, Vol. 166, Springer, Berlin, 1982. For covalently bonded semiconductor materials two applicable interatomic potentials are found in F.H. Stillinger and T.A. Weber, Phys. Rev. B 31, 5262(1985) and J. Tersoff, Phys.Rev. B 38, 9902(1988); B 39, 5566(1989). Computer code providing a more recent improvement on these and other potentials for covalent materials may be found at www-math.mit.edu/∼bazant/EDIP/.

To obtain estimates of free energy at temperature, of either a random solution or a perfectly ordered compound, a quick procedure is to use values of the heats of formation as obtained by one of the above described methods, to use the appropriate configurational entropy and to estimate the vibrational entropy using an interatomic potential that provides good estimates of the elastic constants. The vibrational entropy will be significant only if there is a significant hardening of a soft mode or a significant softening of a hard mode on alloying. A very rough estimate of the vibrational entropy can be obtained using either the Einsteinian entropy or Debye entropy from a knowledge of the change in cohesive energy and the change in atomic volume relative to the respective weighted values for the precursors. A more computer intensive method makes use of quasi-harmonic lattice theory (J.M. Rickman and R. LeSar, Ann. Rev. Mater. Sci. 32, 195(2002)). Links to computer codes making use of quasi-harmonic lattice theory to evaluate free energies of ionic and polar solids (SHELL) and metals (EAMLD) are available at http://www.chm.bris.ac.uk/pt/allan/Research/shell/Welcome.html. Still more computer intensive procedures make use of thermodynamic integration (see Rickman and LeSar reference above).

The ATAT (Alloy Theoretic Automated Toolkit) ab initio code may be obtained from http://cms.northwestern.edu/atat/. This code automates the steps necessary to compute the stable ground state structure of alloy phases. At this writing it does not include code for vibrational entropy or elastic strain induced by different size atoms in the alloy system. The

VASP code is a particularly efficient ab initio code for calculating the ground state energies of alloy phases and inorganic compounds. It can be obtained at http://cms.univie.ac.at. There are many other codes and procedures for evaluating ground state energies well known to the practitioners. Ab initio evaluated databases may be found at: http://alloy.phys.cmu.edu/ and http://www-lab.imr.tohoku.ac.jp/~marcel/enthalpy/enthlp.html.

Organic materials

The energetics of organic molecules can be predicted using a variety of computer programs. A listing of such programs can be found on the web sites: www.ccl.net/cca/documents/chamotlabs/Software.shtml and www.netsci.org/Resources/Software/list.html. It is worthwhile to read (Summary Report, Symposium on Computational Thermochemistry, Computers in Chemistry Division, American Chemical Society, 212th National Meeting, Orlando, Florida, August 25–29, 1996) to obtain the conclusions of experts concerning the accuracy achievable with these programs as compared with the accuracy needed in various applications. (There are publications which proclaim the accuracy and applicability of computational chemistry and physics in this area. However, after reading the above cited reference it is wise to question such professions of faith and examine the facts. Consequently, we will need to examine the applications of these programs that are cited in the remainder of this book in order to evaluate critically conclusions based on predicted results.) The book T. Clark, **A Handbook of Computational Chemistry**, Wiley-Interscience, New York, 1985, also provides instructions in the use of some of these programs. A web site to links to sites concerned with chemical thermodynamics is http://www.psigate.ac.uk/newsite/chemistry-gateway.html; click on *Computational Chemistry* link. Some thermodynamic databases for organic molecules are Beilstein on-line, Chemical Abstracts and NIST. The web site for the latter is http://www.nist.gov/srd/thermo.html. A bibliography relating to quantum chemistry computational procedures is given near the close of this chapter.

Reference to a typical force field used in some molecular dynamics programs, such as AMBER and CHARMM, help provide some insight into the evaluation of energies of molecules in these programs. A typical force field is as follows

$$\begin{aligned} V_{\text{total}} &= \frac{1}{2}\sum_{\text{bonds}} K_r(r - r_{\text{eq}})^2 + \frac{1}{2}\sum_{\text{angles}} K_\theta(\theta - \theta_{\text{eq}})^2 \\ &\quad + \frac{1}{2}\sum_{\text{torsions}} V_n[1 + \cos(n\phi - y)] \\ &\quad + \sum_{i=2}^{N}\sum_{j=1}^{i=1}\left[\underset{\substack{\text{Van der Waals}\\ \text{(Lennard–Jones)}}}{\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}}} \quad + \quad \underset{\text{Coulombic}}{\frac{q_i q}{\varepsilon R_{ij}}}\right] + \underset{\text{H-bonding}}{\sum_{i=2}^{N}\sum_{j=1}^{i=1}\left[\frac{C_{ij}}{R_{ij}^{12}} - \frac{D_{ij}}{R_{ij}^{10}}\right]} \end{aligned}$$

The coefficients in this equation are adjusted empirically. Typically, the accuracy obtainable is on the order of a few kcal/mol which may or may not be sufficient to satisfy the needs of the application.

Appendix 2

Entropy of gases

To relate entropy to statistically based quantities we follow the lead of Fast (ENTROPY, Philips Technical Library, 1962) who noted that from thermodynamics the change in entropy upon increase in volume of a gas of N molecules from V_1 to V_2 is

$$\Delta S = Nk \ln(V_2/V_1) \tag{A2.1}$$

and upon mixing of n_1 molecules of gas 1 of volume V_1 and n_2 molecules of gas 2 of volume V_2 by removing a partition separating the two volumes is

$$\Delta S = k\{n_1 \ln[(V_1 + V_2)/V_1] + n_2 \ln[(V_1 + V_2)/V_2]\} \tag{A2.2}$$

Now consider that the volume V_1, containing gas is partitioned to produce many equal volumes, the number of which is much larger than the number of molecules in the volume, and the size of which is such that no cell contains more than one molecule. As a result of the constant thermal motion of the molecules, different cells become occupied with time. Suppose there are N_1 cells and n molecules. Thus, n cells are occupied and $N_1 - n$ are unoccupied. The total number of different configurations is then

$$W_1 = N_1!/[n!(N_1 - n)!]$$

because the first molecule can fill N_1 boxes, the second $N_1 - 1$, and so on, but not all these configurations are different because there are n indistinguishable molecules and $N_1 - n$ indistinguishable cells.

Using Stirling's approximation for large numbers, $\ln N! = N \ln N - N$ then

$$\ln W_1 = -n \ln(n/N_1) - (N_1 - n)\ln[(N_1 - n)/N_1]$$

Suppose the volume containing the n molecules were increased to V_2. This is equivalent to increasing the number of cells to N_2, each cell maintaining the same cell volume, (i.e. $V_2/V_1 = N_2/N_1$). Then, obviously

$$\ln W_2 = -n \ln(n/N_2) - (N_2 - n)\ln[(N_2 - n)/N_2]$$

and carrying out the algebra we find that

$$\ln(W_2/W_1) = n \ln(V_2/V_1) \tag{A2.3}$$

Now let us consider n_1 type 1 molecules in volume V_1 and n_2 type 2 molecules in volume V_2, which are allowed to mix by removing a partition between the two volumes. Before the mixing we had

$$\ln W_1 = -n_1 \ln(n_1/N_1) - (N_1 - n_1)\ln[(N_1 - n_1)/N_1]$$

$$\ln W_2 = -n_2 \ln(n_2/N_2) - (N_2 - n_2)\ln[(N_2 - n_2)/N_2]$$

Thus, before the mixing the total number of different configurations is the product of W_1 and W_2. After the partition is removed we have n_1+n_2 molecules distributed over N_1+N_2 cells and the number of different configurations is given by

$$W' = (N_1 + N_2)!/\{n_1!\, n_2!\, [N_1 + N_2 - (n_1 + n_2)]!\}$$

We are interested in the ratio $W'/(W_1W_2)$ which can be shown to be given by

$$\ln[W'/(W_1W_2)] = n_1 \ln[(V_1 + V_2)/V_1] + n_2 \ln[(V_1 + V_2)/V_2] \qquad \text{(A2.4)}$$

for the same approximation $n_1, n_2 \ll N_1, N_2$ and $V_2/V_1 = N_2/N_1$. Comparing (A2.4) with (A2.2) and (A2.3) with (A2.1) allows us to write

$$\ln[W_{\text{final}}/W_{\text{initial}}] = \Delta S/k$$

If the initial state is taken as the reference state at 0°K for equilibrium at complete order the third law of thermodynamics yields $S_o = 0$. Hence,

$$S = k \ln W$$

Although this relation was developed on the basis of change in the entropy due to change of volume of a gas or to mixing of two gases, it is more general in application. In the next section we derive the relation between entropy and statistical parameters for a canonical ensemble of systems, which need not be limited to gases.

Canonical ensemble and entropy

Let us consider an ensemble of N total number of systems immersed in a constant temperature bath with which each system can exchange only energy subject to the constraints that the total energy of the ensemble is constant and N is constant. This type of ensemble is called “canonical”. Let n_i be the number of systems in the ensemble in the ith quantum state. A complexion is a way of arranging the members of the ensemble of systems, each member in a different box. The number of complexions corresponding to some particular distribution of quantum states $\{n_i\}$ is then

$$W\{n_i\} = N!/\Pi_i n_i! \qquad \text{(A2.5)}$$

The most probable set of quantum states $\{n_i\}$ maximizes $W\{n_i\}$ subject to the constraints already described. Thus,

$$\delta \ln W\{n_i\} = 0 \qquad \text{(A2.6)}$$

$$\delta(NE) = \delta\Sigma_i n_i E_i = 0 \qquad (A2.7)$$

$$\delta N = \delta\Sigma_i n_i = 0 \qquad (A2.8)$$

where E_i is the energy of a system of the ensemble in the ith state, NE is the total energy of the ensemble and E is the average energy over the energies of the systems in the ensemble. Using Lagrange's method of undetermined multipliers we multiply the latter two equations by $-\beta$ and $-\alpha$, respectively and add them to the first equation to obtain, after suitable use of Stirling's approximation for the factorial of a large number (i.e. $N! = N \ln N - N$),

$$\Sigma_i\, \delta\, n_i(\alpha + \beta E_i + \ln[n_i/N]) = 0 \qquad (A2.9)$$

But, use of the undetermined multipliers can make only two of the terms in the parentheses ( ), say the jth and the kth ones equal to zero. The δn_i are completely independent of each other. Hence, the latter equation can only be satisfied if each term in the parentheses ( ) is equal to zero. Thus,

$$p_i = n_i/N = \exp[-\alpha - \beta E_i] \qquad (A2.10)$$

is the canonical distribution function and the fraction of the systems of the ensemble in quantum state i.

We now set about to determine the values of the undetermined multipliers by comparing functions we shall derive to thermodynamic functions. We sum p_i over all states using the normalization condition

$$\Sigma_i p_i = 1 = [\exp(-\alpha)]\Sigma_i \exp(-\beta E_i) \qquad (A2.11)$$

Solving for $[\exp(-\alpha)]$ and substituting we obtain

$$p_i = \exp(-\beta E_i)/\Sigma_i \exp(-\beta E_i) \qquad (A2.12)$$

The sum in the denominator is called the partition function and is denoted by Z.

The average energy of the ensemble is given by

$$E = \Sigma_i p_i E_i \qquad (A2.13)$$

We rewrite an equation obtained above into the following form

$$\Sigma_i p_i E_i = -\beta^{-1} \ln[\Sigma_i \exp(-\beta E_i)] - \beta^{-1}\Sigma_i p_i \ln p_i \qquad (A2.14)$$

The left-hand side equals the energy E. We now draw attention to the thermodynamic relation

$$E = F - T(\partial F/\partial T)_V \qquad (A2.15)$$

Let $L = \beta^{-1} = kT$ and then we can write the above relation as

$$E = F - L(\partial F/\partial L)_V \tag{A2.16}$$

If we now equate F to $-L \ln \Sigma_i \exp(-E_i/L)$ we will note that on taking the derivative of F with respect to L at constant volume V, which is equivalent to taking it with respect to E_i since holding the latter constant corresponds to holding the volume constant, we obtain

$$\begin{aligned}(\partial F/\partial L)_{Ei} &= -[\ln \Sigma_i \exp(-E_i/L) + (1/L)\Sigma_i E_i \exp(-E_i/L)/\Sigma_i \exp(-E_i/L)] \\ &= [F - E]/L \end{aligned} \tag{A2.17}$$

In arriving at the latter result we have made use of the fact that the average energy E is given by

$$E = \Sigma_i E_i \exp(-E_i/L)/\Sigma_i \exp(-E_i/L) \tag{A2.18}$$

We thus see that our statistical relation for the Helmholtz free energy F satisfies the thermodynamic relation between E and F given above providing that $\beta^{-1} = kT$. Further, since the entropy S satisfies

$$E = F + TS \tag{A2.19}$$

and we have already identified the first two terms in the statistical equation (A2.10) as E and F. Hence, the last term must equal TS. Thus,

$$S = -k\Sigma_i p_i \ln p_i \tag{A2.20}$$

Appendix 3

Consider a polymer having N total mers constrained to lie in a plane and having the mers making 90° angles with each other, with the mers lying at 45° to the horizontal axis, as illustrated in Figure A3.1. Let the total x-projected length of the polymer equal **L** and the x-projected length of each mer equal **a**. If the number of mers pointed to the right exceeds that directed to the left by 2n then

$$2n = L/a$$

Thus, the number of mers pointed to the right is given by

$$\# \rightarrow = N/2 + n$$

while the number directed to the left is given by

$$\# \leftarrow = N/2 - n.$$

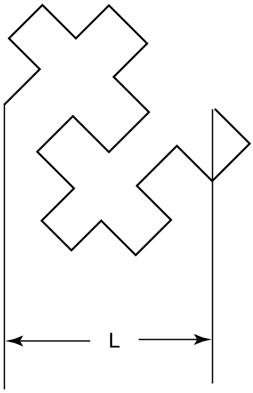


Figure A3.1. Schematic two dimensional polymer illustrating horizontal component of end-to-end length L, as compared total length of polymer.

It is now possible to set down an expression for the entropy of the polymer of x-projected length L as follows

$$S = k \ln\{2N!/[(N/2 + n)!(N/2 - n)!]$$

Appendix 4

Specific heat

For a crystal containing N atoms there will be 3N independent normal modes of oscillation each of frequency ω_i. Einstein considered the case of 3N independent harmonic oscillators of the same frequency, the average energy of which is

$$3Nh\omega/[\exp(h\omega/kT) - 1].$$

Using the definition of the specific heat $C_V = (\partial E/\partial T)_V$, then

$$C_V = 3Nkx^2e^x(e^x - 1)^{-2}$$

is obtained, where $x = h\omega/kT$. For large values of T, $C_V = 3Nk = 3R$, in agreement with experiment. This latter result holds approximately for solid compounds also, with the proviso that the dimensions are energy/atom or energy/g atom. However, the variation of C_V with T, especially at low temperatures, does not agree with experiment.

Debye suggested that the solid be treated as a continuous isotropic medium, which leads to the existence of a range of frequencies from zero to some maximum ω_m, to a density of normal modes that varies as the square of the frequency up to some maximum frequency such that the total number of normal modes equals 3N. He then found that

$$E = [9NkT/(x_m)^3]\int_0^{X_m}\{x^3/(e^x - 1)\}dx$$

from which the specific heat is found to be

$$C_v = [9Nk/(2x_m^3)]\int_0^{X_m}[x^4/(\cosh x - 1)]dx$$

This relation reduces to 3Nk at high temperature and at low temperature to

$$12\pi^4Nk/[5x_m^3] = CT^3$$

both of which are in agreement with experiment. Since the Debye temperature $\theta = h\omega_m/k$, it is possible to use the experimental data for the specific heat at low temperature to evaluate a value for θ. The latter temperature may also be obtained from the dependence of C_V at higher temperature (i.e. 0.5 to 0.75θ), from elastic constant data and from thermal expansion data. The relation of the Debye temperature to the elastic constants is given by

$$\theta = (hc/2\pi k)(6\pi^2 N/V)^{1/3}$$

where V is the volume of the crystal and the average elastic wave velocity for an isotropic crystal is given by

$$c^3 = 3/[2/(c_t)^3 + 1/(c_l)^3]$$

and c_t and c_l are the transverse and longitudinal wave velocities, respectively, i.e. $c_t = (C_t/r)^{1/2}$ and $c_l = (C_l/r)^{1/2}$, where r is the density C_t is the transverse modulus and C_l is the longitudinal modulus.

Since the entropy and energy are related to the specific heat via the relations

$$S = \int_0^T (C/T)\,dT \quad \text{and} \quad E = \int_0^T C\,dT$$

it is possible from the Debye theory for C to derive the related values of S and E.

The Debye model is itself an approximation for the specific heat of thermal origin. Real crystals do not have the density of normal modes depend on the square of the frequency.

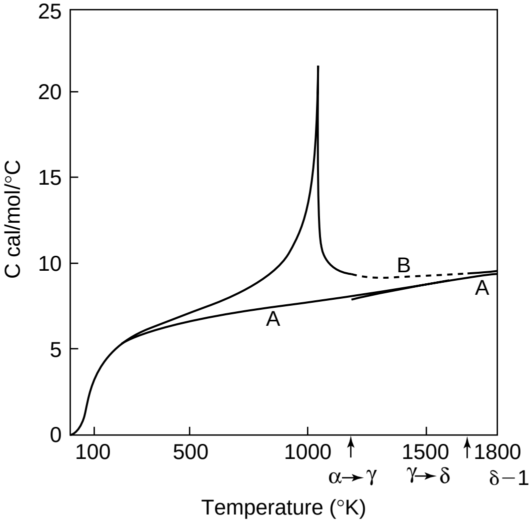


Figure A4.1. The measured specific heat of iron from T = 0–1808°K (solid curves). Curve A is the estimated non-magnetic specific heat of bcc iron and curve B is the estimated total specific heat of bcc iron in the range over which fcc iron is stable, 1183–1673°K. The discontinuities occur at the α → γ transition temperature.

The assumption that solids are continuous isotropic media is too simple. The atomic nature of the solid must be taken into account. This subject forms the modern field of investigation of phonons in solids.

In the above we have not considered any contribution to the specific heat other than that due to the vibrations of the atoms. There are of course other contributions, such as electronic and magnetic ones.

At this point it is of interest to note that in second-order transformations, such as the ferromagnetic–paramagnetic one in iron, the specific heat exhibits a discontinuity at the transformation temperature, and an anomalous dependence on temperature, such as is shown in Figure A4.1. This dependence is due to the disordering of the atomic magnetic moments. The theory of such disordering is known as the theory of the Ising problem. An exact solution to this problem for 3D has not yet been obtained, although there are very good approximate solutions. An analogous problem is that of disordering of a long-range-ordered superlattice.

References

1. J.M. Rickman and R. LeSar, Ann. Rev. Mater. Sci. 32, 195(2002).
2. M. Asta and V. Ozolins, Phys. Rev. B 64, 094104(2001).
3. L.G. Wang, M. Sob and Z. Zhang, J. Phys. Chem. Sol. 64, 863(2003).
4. V.L. Sliwko, P. Mohn, K. Schwarz and P. Blaha, J. Phys. Conden. Mat. 8, 799(1996).
5. Y. Wang et al., Calphad 28, 79(2004).
6. P. Winget and T. Clark, J. Comput. Chem. 25, 725(2004).
7. B. Widom, **Statistical Mechanics**, Cambridge University Press, Cambridge, 2002.
8. C. Kittel, **Introduction to Solid State Physics**, John Wiley & Sons, Inc. New York, 1953.
9. T.L. Hill, **An Introduction to Statistical Thermodynamics**, Dover Publications, New York, 1986.
10. R.J. Weiss, **Solid State Physics for Metallurgists**, Pergamon Press, New York, 1963, p. 276.
11. R. Roy, J. Am. Cer. Soc. 60, 350(1977).
12. V.I. Yukalov, Phys. Rev. B 32, 436(1985).
13. W.C. Johnson and H. Schmalzried, Acta metall mater. 40, 2337(1992).
14. C. Zener, Trans. A.I.M.M.E. 203, 619(1955).
15. W.G. Hoover and F.H. Ree, J. Chem. Phys. 49, 3069(1968).
16. F.H. Ree and W.G. Hoover, J. Chem. Phys. 40, 939(1964).
17. N.F. Carnahan and K.E. Starling, J. Chem. Phys. 51, 635(1969).
18. K.R. Hall, J. Chem. Phys. 57, 2252(1972).
19. S. Fraden and R.D. Kamien, Biophys. J. 78, 2189(2000).
20. D.A. Andelman et al. in **Micelles, Membranes, Microemulsions & Monolayers**, eds. W. Gelbart, A. Ben-Shaul and D. Roux, Springer-Verlag, Berlin, 1994.
21. S. Shin, Z.-G. Wang and S.A. Rice, J. Chem. Phys. 92, 1427(1990).
22. R.S. Cantor and P.M. McIlroy, J. Chem. Phys. 90,4423 and 4431(1989).
23. C.F. Tejero and J.A. Cuesta, Phys. Rev. E 47, 490(1993).
24. D. de Fontaine, Solid State Phys. 34, 73(1979); ibid 47, 33(1994).

25. F. Ducastelle, **Order and Phase Stability in Alloys**, North-Holland, New York, 1991.
26. A. van de Walle, M. Asta and G. Ceder, Calphad 26 (4), 539(2002).
27. V. Ozolin, C. Wolverton and A. Zunger, Phys. Rev. B 57, 6427(1998).
28. C. Zener, in **Thermodynamics in Physical Metallurgy**, ed. C. Zener, ASM, Cleveland, OH, 1950, pp. 16–27.
29. L. Anthony, J.K. Okamoto and B. Fultz, Phys. Rev. Lett. 70, 1128(1993); B. Fultz, L. Anthony, L.J. Nagel, R.M. Nicklow and S. Spooner, Phys. Rev. B 52, 3315(1995); L.J. Nagel, L. Anthony and B. Fultz, Philos. Mag. Lett. 72, 421(1995); L. Anthony, L.J. Nagel, J.K. Okamoto and B. Fultz, Phys. Rev. Lett. 73, 3034(1994).
30. E.J. Palin et al., Phys. Chem. Miner. 28, 534(2001).
31. W. Cao and L. Lai, Biophys. Chem. 80, 217(1999).
32. P.J. Flory, **Principles of Polymer Chemistry**, Cornell Univ. Press, Ithaca, London, 1953; P.G. de Gennes, **Scaling Concepts in Polymer Physics**, Cornell Univ. Press, Ithaca, 1979; I.M. Lifshitz, Sov. Phys. JETP 28, 1280(1969); Zh. Eksp. Teor. Fiz. 68, 2408(1968); I.M. Lifshitz, A.Yu. Grosberg and A.R. Khokhlov, Rev. of Mod. Phys. 50, 683(1978); H.S. Chan and K.A. Dill, Ann. Rev. Biophys. Biophys. Chem. 20, 447(1991).
33. A.Yu. Grosberg and D.V. Kuznetsov, Macromolecules 25, 1970(1992); A.Yu. Grosberg and A.R. Khokhlov, **Statistical Physics of Macromolecules**, American Institute of Physics, New York, 1994.
34. B. Chu, J. Polymer Sci. Part B Polym. Phys. 31, 2019(1993); B. Chu, Q. Ying and A.Y. Grosberg, Macromolecules 28, 180(1995).
35. J.P.K. Doye, R.P. Sear and D. Frenkel, J. Chem.Phys. 108, 2134(1998).
36. K.H. Storcks, J. Am. Chem. Soc. 60, 1753(1938); R. Jaccodine, Nature 176, 305(1955); P.H. Till, J. Polym. Sci. 24, 301(1957); A. Keller, Phil. Mag. 2, 1171(1957); any later textbook on polymers.
37. A. Kolinski, J. Skolnick and R. Yaris, Proc. Natl. Acad. Sci. USA 83, 7267(1986); S. Doniach, T. Garel and H. Orland, J. Chem. Phys. 105, 1601(1996); U. Bastolla and P. Grassberger, J. Stat. Phys. 89, 1061(1997); P.J. Flory, Proc. Roy. Soc. A 234, 60(1956).
38. Y. Zhou, C.K. Hall and M. Karplus, Phys. Rev. Lett. 77, 2822(1996); Y. Zhou, M. Karplus, J.M. Wichert and C.K. Hall, J. Chem. Phys. 107, 10691(1997).
39. P.G. de Gennes, J. de Phys. Lett. 46, L 635(1985).
40. C.F. Abrams, N. Lee and S. Obukhov, Europhys. Lett. 59, 391(2002).
41. A.Y. Grosberg, S.K. Nechaev and E.I. Shakhnovich, J. Phys. (France) 49, 2095(1988).
42. X.Wang, X. Qui and C. Wu, Macromolecules 31, 2972(1998).
43. V.G. Rostiashvili, G. Migliorini and T.A. Vilgis, arXiv/cond-mat/0105607.
44. A. Irback and E. Sandelin, J. Chem. Phys. 110, 12256(1999).
45. C.-M. Chen and P.G. Higgs, J. Chem. Phys. 108, 4305(1998).
46. V.S. Pande, A. Yu. Grosberg and T. Tanaka, J. Chem. Phys. 107, 5118(1997).
47. S. Balijipalli and G.C. Rutledge, J. Chem. Phys. 109, 6523(1998).
48. J.I. Lauritzen and J.D. Hoffman, J. Res. Bur. Stand. 64A, 73(1960); J.D. Hoffman and J.I. Lauritzen, ibid 65A, 297(1961); D.H. Sadler and G.H. Gilmer, Polymer 25,1446(1984), Phys. Rev. Lett. 56, 2708(1986), Phys. Rev. B 38, 5654(1988); Jonathan P.K. Doye, arXiv:cond-mat/9910111 v1 7Oct1999.
49. J.C. Nelson, J.G. Saven, J.S. Moore and P.J.G. Wolynes, Science 277, 1793(1997); J.C. Nelson, Ph.D. Thesis at http://sulfur.scs.uiuc.edu/Theses/Nelson_J/jnelsonCh1.html (there are also Chapters 2–5 separately indexed.)

Bibliography

Statistical thermodynamics

1. J. Willard Gibbs, **Elementary Principles in Statistical Mechanics**, Ox Bow Press, Woodbridge, CT, 1981.
2. E. Schrodinger, **Statistical Thermodynamics**, Cambridge University Press, Cambridge, 1946.
3. L.D. Landau and E.M. Lifshitz, **Statistical Physics**, 3rd ed., Pergamon Press, New York, 1980.
4. R.H. Fowler and E.A. Guggenheim, **Statistical Thermodynamics**, MacMillan, Cambridge, 1939.
5. C. Kittel, **Elementary Statistical Physics**, John Wiley & Sons, New York, 1958.

Bonding theories

1. D.Pettifor, **Bonding and Structure of Molecules and Solids**, Clarendon Press, Oxford, 1995.

Interatomic potentials
First principle

1. J. Hafner, **From Hamiltonians to Phase Diagrams**, Springer, Berlin, 1987.
2. J.A. Moriarty, Phys. Rev. B 38, 3199(1988).

Empirical

Metals

1. S.M. Foiles, Phys. Rev. B 32, 7685(1985). (Embedded atom method).
2. A.M. Stoneham and R. Taylor, **Handbook of Interatomic Potentials**, AERE-Harwell, Osfordshire, 1981.
3. M. Finnis and J. E. Sinclair, Phil. Mag. A, 50, 45(1984).
4. E.S. Machlin, Mat. Res. Soc. Symp. Proc. 19, 67(1983).

Semiconductors

1. F.H. Stillinger and T.A. Weber, Phys. Rev. B 31, 5262(1985).
2. J. Tersoff, Phys. Rev. B 38, 9902(1988); B 39, 5566(1989).
3. www-math.mit.edu/∼bazant/EDIP/

Ionic Crystals

1. G.R.A. Catlow and W.C. Mackrodt in **Lecture Notes in Physics**, Vol. 166, Springer, Berlin, 1982.

Topical Review

1. MRS Bulletin, 21 (2) (1996).

Phonons

1. M. Born and K. Huang, **Dynamical Theory of Crystal Lattices**, Clarendon Press, Oxford, 1954.
2. W. Cochran, **The Dynamics of Atoms in Crystals**, Crane, Russak, New York, 1973.

Data Banks for thermodynamic quantities of solids

1. F*A*C*T*, McGill University/Ecole Polytechnique, Montreal, Quebec H3C 3A7, Canada. (Att. Prof. A.D. Pelton.)
2. Manlabs Inc., 21 Erie Street, Cambridge, Mass. 02139 (Dr. L. Kaufman.)
3. Laboratoire de Thermodynamique et Physico-Chemie Metallurgique associe au CRNS, Institut National Polytechnic, B.P. 44-38401 St. Martin d'Heres, Grenoble, France. (Prof. I. Ansara and C. Bernard.)
4. MTDS, National Physical Laboratory, Teddington, Middlesex, TW11 OLW, England.
5. Institut fur Theoretische Huttenkunde, Technische Hochschule, Aachen, Germany.
6. THERMOCALC, Royal Institute of Technology, Department of Metallurgy, 10044 Stockholm 70, Sweden. (Prof. Mats Hillert.)

Quantum chemistry

1. W.J. Hehre, L. Radom, P.V.R. Schleyer and J.A. Pople, **AB Initio Molecular Orbital Theory**, Wiley-Interscience, New York, 1986. This is a best available reference on accuracy of ab initio methods, limited to small molecules.
2. A. Szabo and N.S. Ostlund, **Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory**, McGraw-Hill, New York, 1989. This is an excellent introduction to Hartree Fock theory, good overviews of CI, CC, and MPPT methods.
3. T. Clark, **A Handbook of Computational Chemistry**, Wiley-Interscience, New York, 1985. This describes the usage of common semiempirical programs and Gaussian 82, good bibliography of MNDO and MINDO/3 calculations listed by molecule.
4. J.B. Foresman and Æ Frisch, **Exploring Chemistry with Electronic Structure Methods**, Gaussian Inc., Pittsburgh, 1993. This is a tutorial on doing quantum chemistry with the Gaussian program, which includes a summary of the theoretical background of the methods.

Molecular mechanics and dynamics

1. J.M. Haile, **Molecular Dynamics Simulation: Elementary Methods**. John Wiley & Sons, New York, 1992. A good introduction to the theory and methods of molecular dynamics simulations, contains source for 2D and 3D hard sphere programs.
2. M.P. Allen and D.J. Tildesley, **Computer Simulation of Liquids**. Clarendon, Oxford, 1991. A detailed treatment of the theoretical underpinnings and statistical mechanics of molecular dynamics and Monte Carlo simulations.
3. C.L. Brooks, M. Karplus and B.M. Pettitt, **Proteins: A Theoretical Perspective of Dynamics, Structure, and Thermodynamics**. Wiley-Interscience, New York, 1991. A good introduction to the methods and potential energy functions used in protein simulations.
4. J.A. McCammon and S.C. Harvey, **Dynamics of Proteins and Nucleic Acids**. Cambridge University Press, New York, 1987. A general overview of biochemical MD simulations.
5. For a recent review on Ginzburg–Landau theories see G. Gompper. Ber. Bunsenges. Phys. Chem. 100:264–271, 1996.

Monte Carlo techniques

1. K. Binder and D.W. Heermann, **Monte Carlo Simulation in Statistical Physics**, Springer, Berlin, 1990.
2. K. Binder, ed. **Monte Carlo and Molecular Dynamics Simulations in Polymer Science**, Oxford University Press, Oxford, 1995.
3. **Monte Carlo and Molecular Dynamics of Condensed Matter Systems**, K. Binder and G. Ciccotti eds. Societa italiana di Fisica, Bologna, 1996.

Related Web sites

General starting point for chemistry materials:
http://www.yahoo.com/Science/Chemistry/Organizations/Professional/American_Chemical_Society/ACS
http://www-chem.harvard.edu/webchem.html

Chemistry resources

http://www.ChemCenter.org/resources.html
Ohio State Computational Chemistry Mailing List (CCL):
to subscribe send name, address, and affiliation to chemistry-request@osc.edu
http://www.osc.edu/chemistry.html
gopher infomeister.osc.edu 73

Directory to chemistry related mailing lists

http://bionmr1.rug.ac.be/chemistry/overview.html
Good physical chemistry site:
http://www-wilson.ucsd.edu/education/qm/qm.html
Sandia National Labs, Quantum Chemistry Tutorial:
http://midway.ca.sandia.gov/~mecolv.qc.html
Pacific Northwest Laboratory basis set server:
http://www.emsl.pnl.gov:2080/forms/basisform.html
NIH Molecular Modeling home page:
http://www.nih.gov/molecular_modeling/mmhome.html

Quantum chemistry program exchange, on-line catalog, ordering info for software

gopher://gopher.gdb.org/1ftp%3aqcpe6.chem.indiana.edu%40/ftp qcpe6.chem.indiana.edu
Brookhaven protein databank, thousands of crystal structures of proteins and nucleic acids: http://pdb.pdb.bnl.gov/
Computational Chemistry Project, Australian internet project on computational chemistry: http://chem2.chem.swin.oz.au/CAUT.html

Books on computer simulation

1. D.W. Heermann, **Computer Simulation Methods**, Springer, Berlin, 1986.
2. **Molecular Dynamics Simulation of Statistical-Mechanical Systems**, G. Ciccotti and W.G. Hoover Eds., North-Holland, Amsterdam, 1986. This book Contains chapters written by experts of the field on many aspects of simulation. A place to dig in into specific topics, with plenty of technical details.

3. M.P. Allen and D.J. Tildesley, **Computer Simulation of Liquids**, Oxford University Press, Oxford, 1987. This is a classic how to book, containing a large number of technical details on a vast array of techniques.
4. R.W. Hockney and J.W. Eastwood, **Computer Simulation Using Particles**, CRC Press UK, London, 1988. This book mostly addresses simulation methods where particles are immersed in an external field as well as with each other. Such methods are commonly used for instance in plasma physics. A section on molecular dynamic simulations on liquids and solids is however present, and the discussion of the cell method in section 8–4 is particularly useful.
5. W.G. Hoover, **Computational Statistical Mechanics**, Elsevier, Amsterdam, 1991. This book focuses on the foundations of simulation from the statistical mechanics point of view, and contains interesting chapters on ensembles, hydrodynamics, non-equilibrium molecular dynamics, with several examples.
6. J.M. Haile, **Molecular Dynamics Simulation**, Wiley, New York, 1992. An excellent and recent general introduction to molecular dynamics. Very deep in explaining the fundamentals, including "philosophical" aspects. Probably the best starting point. A good amount of Fortran 77 code is included.
7. D.C. Rapaport, **The Art of Molecular Dynamics Simulation**, Cambridge University Press, Cambridge, 1995. Excellent recent "recipe book", describing in detail a large number of techniques, and containing a large amount of C code. The book has a home page. G. Ciccotti, D. Frenkel and I.R. McDonald (eds.), **Simulation of Liquids and Solids**, North-Holland, 1987.
8. J.P. Hansen and I.R. McDonald, **Theory of Simple Liquids**, 2nd ed. Academic, Boston, 1986. A classic book on liquids, with a chapter devoted to computer simulation.
9. D. Frenkel and B. Smit, **Understanding Molecular Simulation: From Algorithms to Applications**, Academic, Boston, 1996.

Thermodynamics texts and references

For a recent review on Ginzburg–Landau theories see G. Gompper. Ber. Bunsenges. Phys. Chem. 100:264–271, 1996.

A collection of key papers on simulation from the origins to 1986

1. W.W. Wood, in Ciccotti and Hoover, p. 3.
2. B.J. Alder and T.E. Wainwright, J. Chem. Phys. 27, 1208(1957).
3. J.B. Gibson, A.N. Goland, M. Milgram and G.H. Vineyard, Phys. Rev. 120, 1229(1960).
4. A. Rahman, Phys. Rev. 136, A 405(1964).
5. L. Verlet, Phys. Rev. 159, 98(1967); Phys. Rev. 165, 201(1967).
6. J.-P. Hansen and L. Verlet, Phys. Rev. 184, 151(1969).
7. S. Yip, in Ciccotti and Hoover, p. 523.
8. H.J.C. Berendsen, in Ciccotti and Hoover, p. 496; Comp. Phys. Commun. 44, 233(1987).

Computer techniques specific to proteins and other biological molecules

1. S. Dixon, J. Blaney and D. Weininger, Presented at the Third York Meeting – Characterizing and Satisfying the Steric and Chemical Restraints of Binding Sites, (March 1993) pp. 29–30.

2. A.W.R. Payne and R.C. Glen, Molecular Recognition Using a Binary Genetic Search Algorithm, J. Mol. Graph. 11, 74(1993).
3. Comparative Molecular Field Analysis (COMFA) 1. Effect of Shape on Binding of Steroids to Carrier Proteins, R.D. Cramer III, D.E. Patterson and J.D. Bunce, J. Am. Chem. Soc. 110, pp. 5959–5967(1988).
4. P.J. Goodford, A Computational Procedure for Determining Energetically Favorable Binding Sites on Biologically Important Macromolecules, J. Med. Chem. 28, pp. 849–857(1985).
5. D.S. Goodsell and A.J. Olson, Automated Docking of Substrates to Proteins by Simulated Annealing, Proteins 8, pp. 195–202(1990).
6. N. Pattabiraman, M. Levitt, T.E. Ferrin and R. Langridge, Computer Graphics in Real-Time Docking with Energy Calculation and Minimization, J. Comput. Chem. 6, pp. 432–436(1985).

Problems

1. Use the method of Villars* to evaluate the structure likely to be stable for five hypothetical binary compounds of your choice.
2. Ammonium Chloride has the CsCl structure(B2) in which the NH_4 ions occupy the body-centered sites and the Cl ions the corner sites of the unit cell. There are two possible orientations for the NH_4 ions. Above a critical temperature these orientations are randomly distributed among the NH_4 ions. Below the critical temperature there is a partial long-range order to the distribution of these orientations. Use a Bragg–Williams type model to calculate the configurational entropy of a partially long-range-ordered crystal having an excess of one orientation over the other, i.e. in the perfectly ordered crystal there is only one orientation while in the random case there are equal numbers of the two orientations.
3. Use Figure 1.3 to determine whether the liquid phase of Bi has a lower or higher specific volume than each of solid phases I, VII and VIII.
4. Suppose by radiation that defects can be introduced into a solid and thereby raise its Gibbs free energy. Use free energy temperature curves to determine the effect of these defects on the melting point of the irradiated solid.
5. Given a thin film coherent with its substrate. The latter is elastically much stiffer than the film. The film will not deform plastically, but only elastically up to its melting point. Also, it is known that the elastic energy per unit volume in a thin film with principal stresses, t_1 and t_2, in the plane of the film, and $t_3(=0)$ normal to the plane of the film is given by $(1/2E)[t_1^2 + t_2^2 + t_3^2] - (n/E)[t_1t_2 + t_2t_3 + t_3t_1]$. Assume $E = 200\,\text{GPa}$, $n = 0.3$, molar volume $= 10\,\text{cm}^3/\text{mol}$ and that the difference in thermal expansion coefficient between substrate and film equals 0.00001/°C. If the film is deposited coherently at 500°C, the melting point in the absence of stress is 1500°C and the difference in free energy between liquid and solid is given by $\Delta g = \Delta s(T_M - T)$ with $\Delta s = 6\,\text{cal/mol/°C}$, calculate the change in the melting point of the stressed solid thin film when it is heated from its deposition temperature up to its melting point as compared to the melting point of the unstressed film.

---

* P. Villars, J. Less Common Met. 92, 215(1983); 99, 33(1984); 102, 199(1984); 109, 93(1985).

6. Suppose that in the structure corresponding to some compound ABO_4 there are two sites (types I and II) that only one of the 4 atoms of type O per molecule can occupy, per unit cell corresponding to one molecular unit. The energy of the crystal varies linearly with site occupation varying from $-E$ when a random distribution of this O type atom exists on the I and II type sites to $-E - \Delta E$ (per mole of molecules) when only the II type sites are completely occupied by this O type atom and the I type sites are vacant in the crystal. Assume the occupation of any site is independent of the occupation of other sites. Evaluate the site occupancy as a function of temperature.
7. Does the entropy of an amorphous solid have a value closer to that for the crystalline solid or to that for the amorphous liquid of the same material? Justify your answer.
8. It has been found that exposure to x-rays or to a magnetic field induces an insulator to metal transition in a manganite material. Assuming that the material was near its thermodynamic transition temperature at the time of exposure to x-rays, provide a physical basis for the assumption that x-ray or magnetic field exposure can increase the free energy of the insulator crystal structure while not affecting that of the metallic structure. *Hint*: the insulator structure is anti-ferromagnetic.
9. If the insulator–metal transition in the previous problem was from an anti-ferromagnetic structure of the spins to a ferromagnetic one would you expect a discontinuity in the electrical resistivity to occur during the transition? Provide a basis for your answer.
10. Derive the condition such that an increase in entropy of the gaseous system of N non-interacting spheres of radius r and n non-interacting spheres of radius R in a box of volume V, where $R > r$ and $n < N$ will bring about the self-assembly of the large spheres in a close-packed array along an interior planar surface of the box. *Hint*: There is a volume about each large sphere equal to $4\pi R^2 r$ from which the little spheres are excluded when the large sphere is not in contact with the box or other large spheres. Each such contact to a large sphere increases the volume available to be occupied by the small spheres.

List of Symbols for Table 1.2

α: Coefficient of thermal expansion
β: Compressibility
C_p: Specific heat at constant pressure
C_V: Specific heat at constant volume.

CHAPTER II

Thermodynamics of Solid Solutions

Introduction

We consider the following subjects in this chapter: random solid solutions of binary systems; derivation of the free energy–composition relation for such solutions and the free energy of mixtures of solid solutions. This knowledge is then used to determine the equilibrium distribution of phases in a binary system at a given temperature. We then investigate the graphical significance of the partial molar free energy; non-random solutions; and various models for describing the energies and entropies of these solid solutions. Models that evaluate the excess free energy due to difference in atom or ion radius or a difference in bond length between components, the strain free energy, is then investigated. Factors other than atomic size that affect the enthalpy and entropy of solid solutions are briefly considered. Finally, the activity and activity coefficient of solutions are defined and some behavior of these parameters is described.

1. Random solutions in binary systems

1.1. Free energy of a random binary solution

The free energy of a solution is given by

$$G = E + PV - TS = H - TS \tag{2.1}$$

We assume that the enthalpy of the random solution, and its first and second derivatives with respect to composition are finite and continuous functions of composition. These assumptions are in agreement with experimental measurements of the composition dependence of the enthalpy of solutions and with first principle calculations.

In addition to the entropy contributions that exist in the pure components there will be a positional contribution to the entropy of a random binary solution. The latter contribution stems from the fact that the number of different ways that the two types of atoms can be arranged on the N lattice sites differs from unity and is given by

$$W = N!/[(NX_1)!(NX_2)!] \tag{2.2}$$

for a random distribution, where X_i is the atom fraction of component i. We make use of the Boltzmann relation between the entropy of configuration, S_{conf}, and W, which is

$$S_{conf} = k \ln W \tag{2.3}$$

and Stirling's approximation for the factorial of a large number. (Reference to equation 1.2 of Chapter I yields the fact that W corresponds to the number of microstates g and that we have implicitly assumed in the above that each microstate is as probable as any other. This expression for W can be understood as follows. Assume that there are N distinguishable atoms that are to be arranged on N lattice sites. There are then N different ways of choosing atoms to fill the first lattice site, $N - 1$ different ways of choosing atoms to fill the second lattice site, and so on, so that the number of different ways of placing N distinguishable atoms on N lattice sites is then N!. But, there are only two types of distinguishable atoms, 1 and 2. Thus, to arrive at the number of different ways of distributing these two types of atoms on the N sites we must divide N! by $(NX_1)!$ and $(NX_2)!$, which are the number of different ways of arranging NX_1 distinguishable atoms on NX_1 sites and NX_2 distinguishable atoms on NX_2 sites, respectively.)

The positional entropy or entropy of mixing for a random solution is then

$$S_{conf} = -Nk[X_1 \ln X_1 + X_2 \ln X_2]^* \tag{2.4}$$

We can now obtain the following relation for the free energy of a solution.

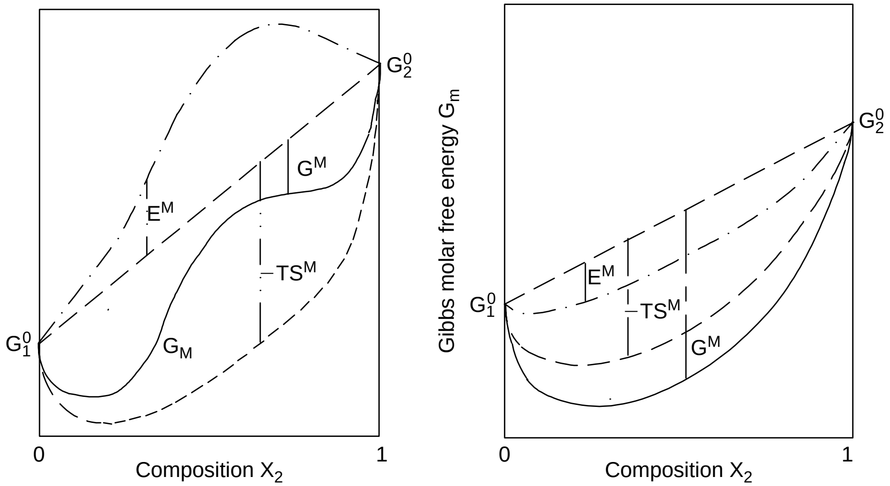
$$G = Nh(X_2) + NkT[X_1 \ln X_1 + X_2 \ln X_2] - NTs_T(X_2) \tag{2.5}$$

(In the above the units of h and s_T, the enthalpy and non-configurational entropy, respectively, are per atom, and N is the total number of atoms in the solution.)

Figure 2.1 shows graphically how G depends on composition. It is important to grasp the fact that the value of the first derivative of the enthalpy, dH/dX, is finite for all values of composition, while that for the entropy term, $-T(dS/dX)$, approaches negative or positive infinity at $X_2 = 0$ and $X_2 = 1$, respectively, due to the contribution of the positional entropy! The aforesaid fact governs the shapes

---

* This expression is valid when the components have nearly the same atomic or molecular volume. If this volume differs between components then the appropriate relation is $\Delta S_{mix} = -kn_1 \ln[V_1/(V_1 + V_2)] - kn_2 \ln[V_2/(V_1 + V_2)]$, where n_i and V_i are the number of i molecules and total volume of i molecules in the mixture, respectively.



(a) (b)

Figure 2.1. Molar free energy versus composition diagrams illustrating various molar and relative integral molar quantities.

assumed by the free energy–composition curves! As shown, adjacent to the terminal compositions, the free energy–composition curves must always have positive curvature values. The sign of the curvature at intermediate values of the composition may be either positive or negative depending upon whether the enthalpy of mixing

$$N(h - X_1 h_{1o} - X_2 h_{2o})$$

has a negative or positive value, respectively. The significance of these facts will be demonstrated after the next section in which the free energy of a mixture of two solutions is derived analytically.

To summarize, the free energy of a binary solution must be represented as a function of composition by a single continuous curve having either of the shapes described by the full lines in Figures 2.1(a) and (b).

1.2. Free energy of a mixture of solutions

We wish to derive the free energy of a mixture of two binary solutions. One of these solutions has the composition defined by X_2' and the other has the composition defined by X_2''. These solutions have the corresponding free energies per atom g' and g'', respectively. The points representing such solutions on a free energy–composition map are shown in Figure 2.2. Let us further suppose that the relative

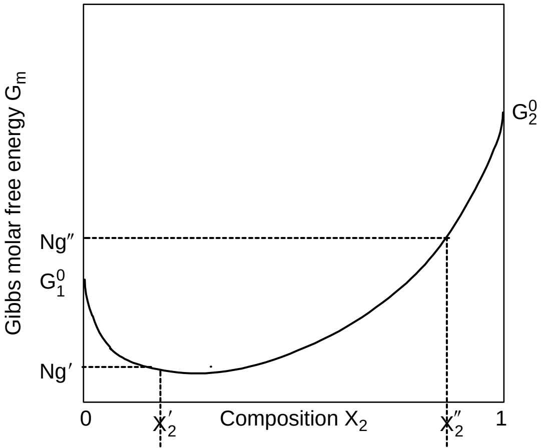

Gibbs molar free energy G_m
Ng''
G_1^0
Ng'
G_2^0
0
X_2'
Composition X_2
X_2''
1

Figure 2.2. Illustrating the molar free energies of two solutions corresponding to the composition, X_2' and X_2''.

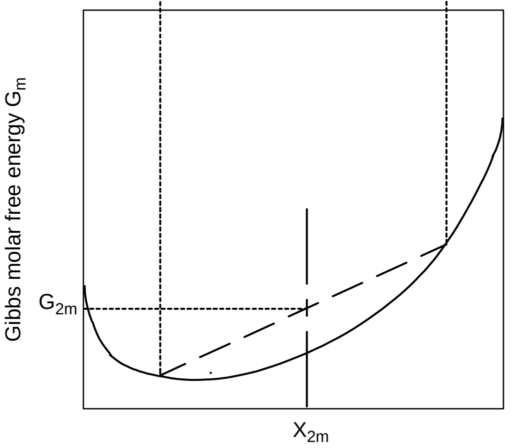

Gibbs molar free energy G_m
G_{2m}
X_{2m}

Figure 2.3. Illustrating that the mixture of solutions “and” having the average composition X_{2m} is not stable with respect to the solution having this same composition.

amounts of these two solutions are such as to yield the average composition for the mixture given by X_{2m}.

According to the tie-line principle derived from conservation of atoms of a given kind, the total number of atoms of the solution denoted by the symbol ″ is

$$N'' = N(X_{2m} - X_2')/(X_2'' - X_2')$$

Similarly, the total number of the ′ denoted solution is given by

$$N' = N(X_2'' - X_{2m})/(X_2'' - X_2')$$

Hence, the free energy of the mixture is given by

$$G_{mixt} = g'N' + g''N''$$

Substituting for N' and N'' yields

$$G_{mixt} = N(g'X''_2 - g''X'_2)/(X''_2 - X'_2) + N(g'' - g')X_{2m}/(X''_2 - X'_2)$$

This relation is identical to the equation for the straight line passing through the points (Ng', X'_2) and (Ng'', X''_2)! Hence, in a free energy–composition map, points representing the free energy of mixtures of solutions lie along the line joining the points corresponding to the component solutions. From this result, it follows that for the case where the free energy–composition curve for a phase has a positive curvature that the free energy of any mixture of two compositions of this phase is not stable with respect to the solution having the average composition of the mixture. This situation is illustrated in Figure 2.3. In this figure the intersection of the vertical line with the dashed line yields the free energy of a mixture of solid solutions. The compositions of these component solutions are defined by the intersections of the dashed line with the full curve. The average composition of these solutions is that of the vertical line, X_{2m}.

It follows from the above analysis that when the free energy–composition curve of a solution has a negative curvature then a mixture of solutions can be stable relative to the solution corresponding to the average composition of the mixture. This fact is illustrated in Figure 2.4, where the dashed line corresponds to the mixture that has the lowest free energy in the composition range extending from X'_2 to X''_2. Further, our analysis has shown that the negative curvature of the free energy–composition curve is a consequence of a positive value of the enthalpy of mixing, i.e.

$$H^M = H_m - X_1H_{1mo} + X_2H_{2mo} > 0$$

Here, H_m is the molar enthalpy (enthalpy of one mole) of the solution, and subscripts io denote pure i. The enthalpy of mixing, H^M, is also known as the relative integral molar enthalpy.

1.3. Phase boundary compositions corresponding to miscibility gap compositions of a random binary solution

When the enthalpy of mixing is positive, as noted in the previous section, there will be a miscibility gap consisting of a mixture of solid solutions that represents the equilibrium situation at low temperatures. The phase boundary compositions

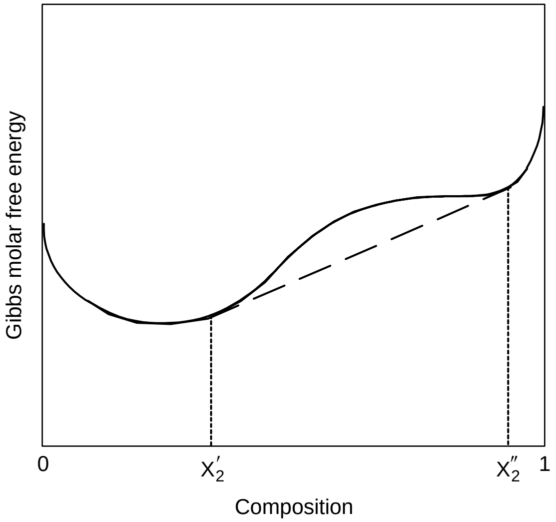


Figure 2.4. The dashed line represents the free energy of mixtures of solutions that are stable over the range of average compositions from X_2' to X_2'', relative to the solid solution having the average composition. The composition of the solutions in the mixture are X_2' and X_2''.

can be derived from two considerations: graphical or analytic. According to Figure 2.4 the following relation must be valid at equilibrium

$$dG_m/dX_2\Big|_{X_2=X_2'} = dG_m/dX_2\Big|_{X_2=X_2''} = (G_m'' - G_m')/(X_2'' - X_2') \qquad (2.6)$$

(Cautionary note. See Problem 1 at end of chapter.)
Or by substitution of

$$G^M = G_m - (X_1G_{1mo} + X_2G_{2mo}) \qquad (2.7)$$

the relations defining the phase boundaries become

$$dG^M/dX_2\Big|_{X_2=X_2''} + G_{2mo} - G_{1mo} = (G_m'' - G_m')/(X_2'' - X_2') \qquad (2.8)$$

and $$dG^M/dX_2\Big|_{X_2=X_2''} + G_{2mo} - G_{1mo} = (G_m'' - G_m')/(X_2'' - X_2') \qquad (2.9)$$

Another way of defining the compositions of the coexisting phases in the miscibility gap at equilibrium is to use the fact that then the partial molar free energies of a given component are equal in the coexisting phases, i.e.

$$\bar{G}_i' = \bar{G}_i'' \quad (i = 1, 2) \tag{2.10}$$

Consider the molar free energy G_m. It is related to the partial molar free energies by

$$G_m = X_1\bar{G}_1 + X_2\bar{G}_2 \tag{2.11}$$

We now take the differential of equation 2.11 to obtain

$$dG_m = X_1 d\bar{G}_1 + X_2 d\bar{G}_2 + \bar{G}_1 dX_1 + \bar{G}_2 dX_2 \tag{2.12}$$

By the Gibbs–Duhem relation*, $X_1 d\bar{G}_1 + X_2 d\bar{G}_2 = 0$. Using this result and then multiplying the resulting equation by X_1/dX_2 under the constraint that $X_1 + X_2 = 1$. We obtain

$$\begin{aligned} X_1 dG_m/dX_2 &= -X_1\bar{G}_1 + X_1\bar{G}_2 \\ &= \bar{G}_2 - (X_1\bar{G}_1 + X_2\bar{G}_2) \\ &= \bar{G}_2 - G_m \end{aligned} \tag{2.13}$$

We rearrange equation 2.13 to the form

$$\bar{G}_2 = G_m + X_1 dG_m/dX_2 \tag{2.14}$$

Similarly it can be shown that

$$\bar{G}_1 = G_m + X_2 dG_m/dX_1 \tag{2.15}$$

Equations 2.14 and 2.15 are interpreted graphically in Figure 2.5.

The relations developed in the foregoing are independent of any model of the solution. If either equations 2.8 and 2.9 or 2.14 and 2.15 are to be used for the prediction of phase boundaries then it is apparent that it is necessary to know the molar free energy of the solution as a function of composition at the temperatures

---

* This form of the Gibbs–Duhem relation used here is applicable to fluid phases and not to non-hydrostatistically stressed phases. For the latter case read Reference 14 to determine the applicable relation.

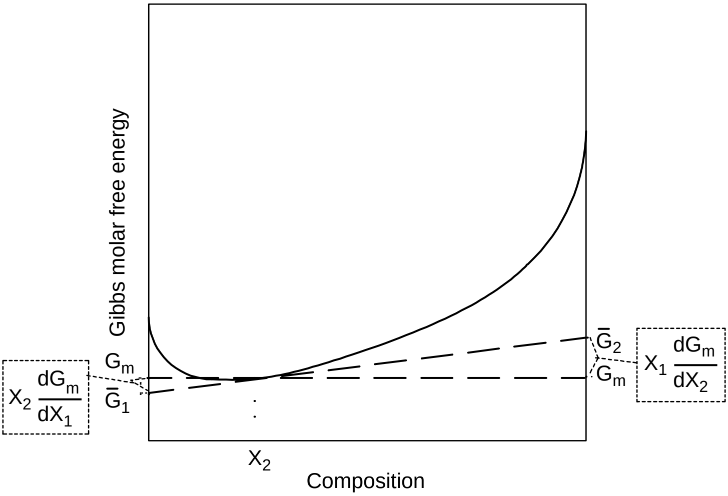


Figure 2.5. Illustrating the graphical significance of the partial molar free energy and its relation to the molar free energy. Refer to equations 2.14 and 2.15 of the text.

of interest. It is useful at this point to make some simplifying assumptions governing the solution for pedagogical reason.

1.4. Regular solutions

Consider a regular solution in which the entropy of configuration is the same as for the random solution and the enthalpy of mixing can be expressed as a symmetric function of the composition for a binary system, i.e. $H^M = bX_1X_2$. In this case where both H^M and S^M are symmetric functions of the composition then the relative integral free energy must also be a symmetric function of the composition, as shown in Figure 2.6. The symmetric function shown in Figure 2.6, corresponding to $b > 0$, will yield a stable miscibility gap between the compositions X' and X''. As shown in this figure at the phase boundary compositions the value of dG^M/dX_2 is zero. Consequently, by equations 2.8 and 2.9 then $G_{2mo} - G_{1mo}$ must equal the quantity $(G''_m - G'_m)/(X''_2 - X'_2)$ as well. But, for the regular solution the value of dG^M/dX_2 is given by $b(1 - 2X_2) + RT\ln(X_2/X_1)$, where $b > 0$, for the case described in Figure 2.6. Hence, the phase boundary compositions of the miscibility gap are given by

$$X_2 = X_1 \exp[-b(1-2X_2)/RT] \tag{2.16}$$

for both the coexisting compositions.

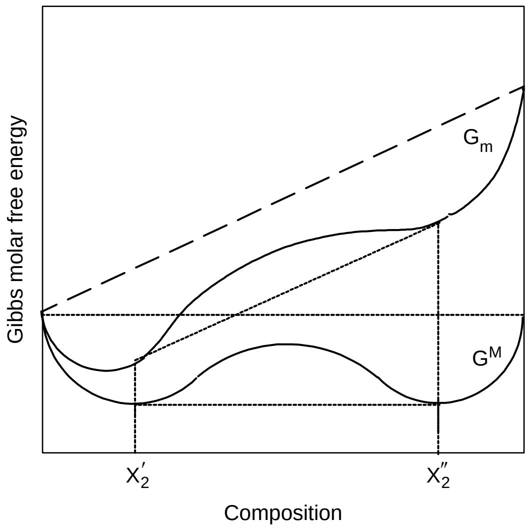


Figure 2.6. Illustrating how the relative integral molar free energy becomes a symmetric function of composition for a regular solution.

Obviously, there must be a temperature above which the curvature of the free energy–composition relation is everywhere positive. The temperature at which the curvature first changes from positive to negative at some composition on decreasing temperature is called the critical temperature. This temperature can be derived by differentiating equation 2.16 with respect to composition, X_2, then setting $dT/dX_2 = 0$ and $X_2 = 0.5$. The result is $T_c = b/2R$. Of course, this result is applicable only to a regular solution. A miscibility gap and the critical temperature is illustrated in Figure 2.7.

The regular solution approximation is applicable to many semiconductor solid solutions and less applicable to metallic alloys. There are, of course, other analytic descriptions giving the composition and temperature dependence of the enthalpies and entropies of solid and liquid solutions. The values of the parameters in such relations are obtained primarily by fitting to experimental data. For some systems ab initio calculations can yield acceptable values for such parameters. This topic will be discussed in greater detail in the next chapter.

1.5. Equilibrium between terminal solutions of different crystal structure

Finally, let us consider the case of equilibrium between two solid solutions having different crystal structures. In this case there will be a free energy–composition

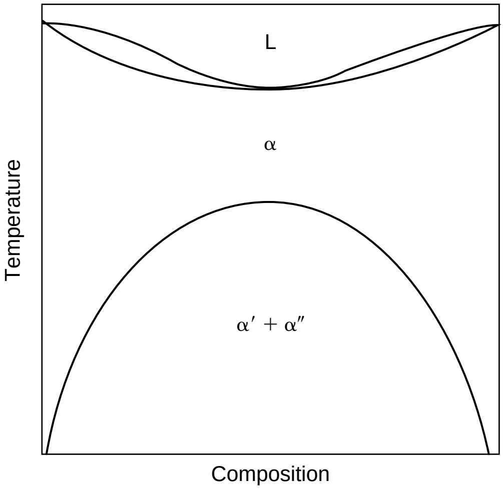

L
α
α′ + α″
Temperature
Composition

Figure 2.7. Illustrating a miscibility gap type phase diagram.

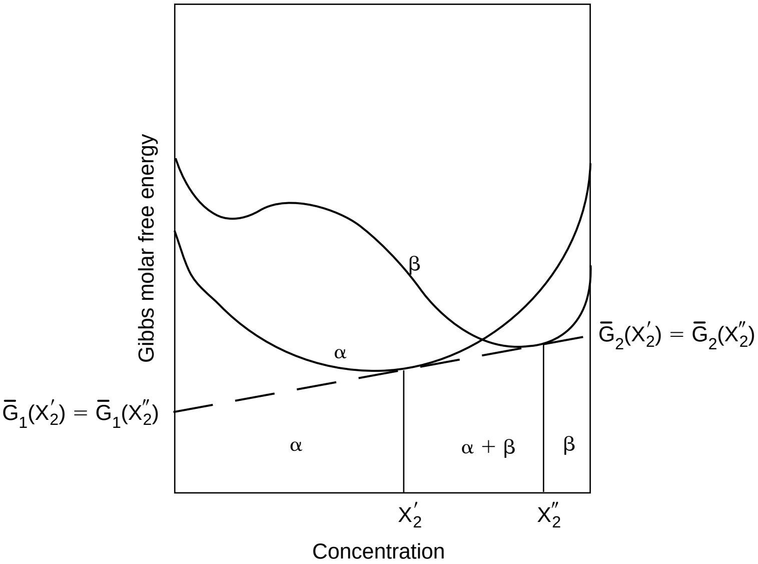

Gibbs molar free energy
β
α
$\bar{G}_2(X_2') = \bar{G}_2(X_2'')$
$\bar{G}_1(X_2') = \bar{G}_1(X_2'')$
α
α + β
β
X_2'
X_2''
Concentration

Figure 2.8. Illustrating the equilibrium developed between two different structures α and β.

curve for each solution, as shown in Figure 2.8. The common tangent to the two curves (the dashed line) satisfies the requirement described by equation 2.10 and thus yields the compositions of the coexisting phases in the stable two phase region (i.e. the compositions between the vertical dotted lines).

2. Non-random binary solid solutions

2.1. Positional entropy or entropy of mixing

The positional entropy of non-random solutions cannot be described by equation 2.4. Also, a rigorous analytic description of the positional entropy of a non-random solution does not yet exist. Consequently, to obtain a thermodynamic description of a non-random solution it is necessary to resort to approximations of the positional entropy.

The first attempts to describe the positional entropy (the entropy of mixing distinguishable units, be they atoms or molecules) used the mean-field approximation. Although this method can successfully predict the second-order transition in beta-brass between ordered and disordered phases it is inadequate in describing the temperature dependence of the partial order. The mean-field approximation in the order-disorder problem for a binary alloy of components A and B consists in the assumption that the probability of having an A_iB_j pair between the ith and jth sites in the lattice network is the product of the probability, X_A, that the ith site is occupied by A, and X_B, the probability that the jth site is occupied by a B atom, where X is the atom fraction of the component denoted by the subscript. The Bragg–Williams model for the order–disorder transition belongs to the mean-field category.

The Bethe approximation in which short-range order is taken into account is an improvement on the mean-field approximation, as is that of the quasi-chemical approximation, in that it uses a pair interchange energy to evaluate the short-range order and, hence, the probability of finding an A_iB_j pair. The Kikuchi cluster variation method[2] is a way of obtaining a still more accurate approximation to the configurational entropy. The Kikuchi idea is to group the nodes of the Markov network* involved in the problem into basic (possibly overlapping) clusters, and then to compute an approximation to the Gibbs free energy by summing the free energies of the basic clusters, minus the free energy of over-counted intersections of clusters, minus the free energy of over-counted intersections of intersections, and so on. Kikuchi's concept has been exploited by De Fontaine and his student Sanchez[3] to describe a host of disorder–order transitions in metallic systems.

Computer simulation procedures also are capable of providing results for the entropy of position. Primary among these is the Monte Carlo method which by counting the number of configurations yielding the equilibrium value of the energy provides a direct evaluation of the positional entropy.

---

* A definition of the Markov network may be found at Wikipedia.org.

2.2. Other contributions to the entropy of solution

Depending upon the nature of the solution (alloy, semiconductor, molecular, ionic) there will be contributions to the entropy of solution in addition to the entropy of mixing. For example, in a molecular solid solution rotation of individual components of the molecules may be possible. The change in such rotations upon solution contributes to the relative integral entropy of solution. Similarly, changes in the vibrational spectra of the atoms or molecules or of the spin states of atoms upon solution can also contribute to the relative integral entropy of solution. Further, the breaking apart of molecules into their component units (atoms) can also contribute to the entropy of solution in liquid solutions. Also, in water solutions, as already mentioned in Chapter I, the interaction between solute and water can affect the entropy of solution. The state of the art at this writing is at the point of attempting to evaluate the entropy of solution either via ab initio procedures or via semiempirical procedures.

3. Concept of strain energy in solid solutions

3.1. Metallic alloys

The lattice parameter of a pure crystalline element defines the equilibrium spacing between the atoms corresponding to the absence of strain energy in the lattice. If the distance between these atoms is changed from that of the equilibrium spacing then the lattice is said to be strained and the increment in energy associated with this change in lattice parameter is termed strain energy. When solute atoms are introduced to form a dilute solid solution, x-ray evidence demonstrates that the distance between solvent atoms has altered from the value corresponding to the equilibrium spacing in the pure crystalline element. One view of the change in energy associated with the change in solvent interatomic distances is that it corresponds to the increment in energy associated with strain of the solvent–solvent interatomic bonds. This concept has certain predictive value of use in the interpretation and prediction of phase diagrams and thus a more detailed investigation of it is justified. However, it must not be inferred that the only contribution to the change in lattice parameter upon the introduction of a solute atom is due solely to strain. There are additional effects on the lattice parameter associated with a change in configuration of the solute atom's outer electrons and those of its neighbors that are not included in the strain energy concept. Such effects are minimized when the difference in electronegativity and dsp character of the outer electrons between solute and host are negligible in the case of alloy solid solutions and when the solute ions have the same valence as that of the host for the case of ionic solid solutions.

Thus, the analysis of strain energy given in this section is limited to solid solutions where these other effects are minimal, e.g. Ag–Cu, (K, Na)Cl solid solutions.

The original evaluation of the strain energy due to the introduction of a solute atom into a pure solid solvent host is due to Friedel[4]. Eshelby[5] then provided a result that did not make approximations for certain elastic parameters. We shall make use of the results of the latter's analysis. The assumption is made that the introduction of a solute atom into a vacant site in the interior of the host lattice exerts a pressure on the surrounding atoms. This pressure acts to distort the host lattice. The distortion produced is assumed to be the same as that calculated from the application of elasticity theory to the equivalent continuum problem. Thus, if the atomic distance between host atoms in the pure host is $2{}^oR_1$ then the radius of the hole, into which the solute atom is introduced and at which the latter exerts its pressure on the host lattice, is ${}^oR_1$. It is assumed that this pressure acts to make the hole dilate until the outward pressure due to the solute atom is equal to the inward pressure due to the distorted host on the solute atom. Hence, the stress-free solute atom is considered to be a compressible solid with a radius equal to ${}^oR_2$. The interesting results of this model are:

1. The state of stress in the host about the solute atom is pure shear.

2. The total strain-free energy due to the introduction of one solute atom into the host is

$$G_{st} = 2\mu_1 C_6({}^oV_2 - {}^oV_1)^2/3V_2$$

where $C_6 = 3K_2/(3K_2 + 4\mu_1)$, *K is the bulk modulus,* μ *is the shear modulus, V is the atomic volume, the superscript* ^{o} *denotes the pure state. The strain-free energy due to the introduction of solute atoms sufficient to produce the composition* X_2 *is then*

$$G^{M,st} = NG_{st}X_2\{1 - K_2X_2/K_1\}$$

for $X_2 << 1$ *and where* $(1 - K_2X_2/K_1)$ *is a consequence of the change in the strain energy due to the interaction of the introduced solute atom with all the other solute atoms in the host. When there is no other contribution to the relative integral molar-free energy, other than the entropy of mixing, then*

$$G^M = G^{M,st} + NkT[X_1 \ln X_1 + X_2 \ln X_2]$$

3. The strain entropy, S_{st}, *is given by* $-dG_{st}/dT$, *which upon the substitution of representative values is closely approximated by the value of the first term in the derivative,* dG_{st}/dT, *namely* $[2C_6({}^oV_2 - {}^oV_1)^2/3V_2](d\mu_1/dT)$. *Since, the sign of the first* [ ] *bracket is positive, then the sign of* S_{st} *depends upon the sign of the derivative in the second bracket, which is negative. Hence, the sign of* S_{st} *is* *positive, i.e. the strain entropy is a positive quantity. Also, it is an excess quantity in that*

$$S_m = X_1{}^oS_1 + X_2{}^oS_2 - Nk[X_1 \ln X_1 + X_2 \ln X_2] + S^{M,st}$$

The strain entropy term prevents the solution from belonging to the regular solution category, because for the latter the relative integral molar entropy is equal solely to the entropy of mixing, i.e. the excess molar entropy is zero for a regular solution. (An excess quantity is given by the difference between the relative integral molar term and that for an ideal solution.)

In the absence of chemical effects between unlike atoms, as may be induced by a difference in electronegativity or d band centers of gravity if the components are transition elements or hybridization between overlapping unlike bands (i.e. d band of one component with p band of other component), the above model provides a close description of the solid solution thermodynamic quantities of metallic alloys. For example, Figure 2.9 shows a comparison between experiment and the results of using the above model to calculate the phase boundary compositions of the solid state miscibility gap in the copper–silver binary system. (The calculation is based on setting the parameter b in equation 2.16 equal to

$$NG_{st}(1 - K_2X_2/K_1)/(1 - X_2)$$

and solving for X_2 in 2.16.)

A much more sophisticated evaluation of the strain energy in Cu–Ag has been performed using ab initio procedures by Sanchez et al.[6] which yield a better agreement for the solubility limit of Ag in Cu (5 versus 4 at% (xptl)) but worse agreement for the solubility limit of Cu in Ag (7 at% versus 13 (xptl)) than the simple model results (9 and 16 at%, respectively) shown in Figure 2.9. Hence, use of the simple model is justified where applicable.

One conclusion of the above discussion of strain energy is that a difference in atomic volumes between components contributes a positive term to the free energy and to the enthalpy. Since a positive value of the enthalpy of mixing leads to the formation of a miscibility gap, and, hence, limited solid solubility, it follows that a sufficient difference in atomic volumes between components will restrict the mutual solid solubility of the components. The larger is the positive enthalpy of mixing the less is the solubility (composition of the solute component at the phase boundary). This result is the explanation for the empirical rule of Hume-Rothery and Raynor[7] according to which the solid solubility of a component in a host is very limited when the difference in atomic radii between host and solute exceeds 15%.

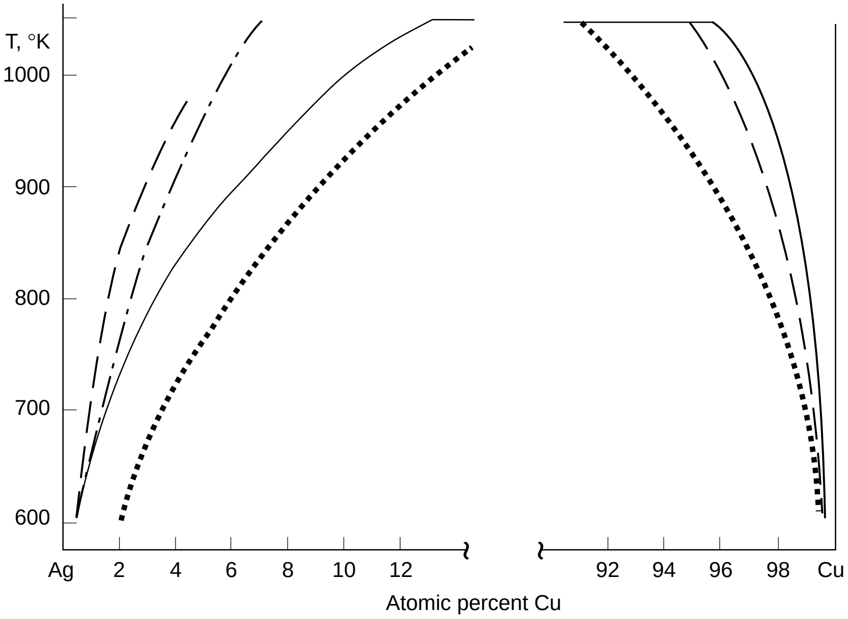


Figure 2.9. Dotted lines represent our calculated phase boundaries of the miscibility gap. The heavy dashed (relaxed) and dot-dash (unrelaxed) lines are the calculations of Sanchez et al.[6] The full lines correspond to the experimental phase diagram, which is taken from M. Hansen. **Constitution of Binary Alloys**, McGraw-Hill Book Co., New York, 1958 with permission.

3.2. Semiconductor alloys

Strain energy exerts an important effect in semiconductor alloys in that it tends to destabilize random solid solutions relative to long-range-ordered ones. (The optical properties of these solutions differ sufficiently to make this difference a practical concern.) One approach to an evaluation of the equilibrium atomic arrangement in multicomponent semiconductor alloys has been to assume that the strain energy can be calculated in terms of the directed bond strain. Again, it is assumed that the bonds in the elemental crystal of the same crystal structure represent the strain-free bonds. The force constant associated with each bond is calculated from the elastic constants using a model for the latter.[8a] That for a bond between unlike atoms is taken to be the mean of that for the bonds between the like atoms of the couple. Using this procedure, and minimizing the energy as a function of the average lattice parameter, expressions for the energy of disordered and long-range-ordered semiconductor alloys may be calculated. First principle calculations have also been used to determine the energies of various ordered structures in III–V semiconductor alloys.[8b]

4. Factors other than atomic size deviation that affect the enthalpy and entropy of solid solutions

It must be remembered that factors other than size affect the solid solubility as suggested above, i.e. chemical interactions between the unlike atoms. An illustration of the latter fact is given by the Darken–Gurry plot, Figure 2.10, which describes the limit of solid solubility in a plot of atom radius difference against the Pauling electronegativity difference, for the case of metallic alloy solid solutions. As shown in this figure the solid solubility is extremely limited when the electronegativity difference exceeds about 0.4, even if the difference in atomic size is negligible. It may be recalled

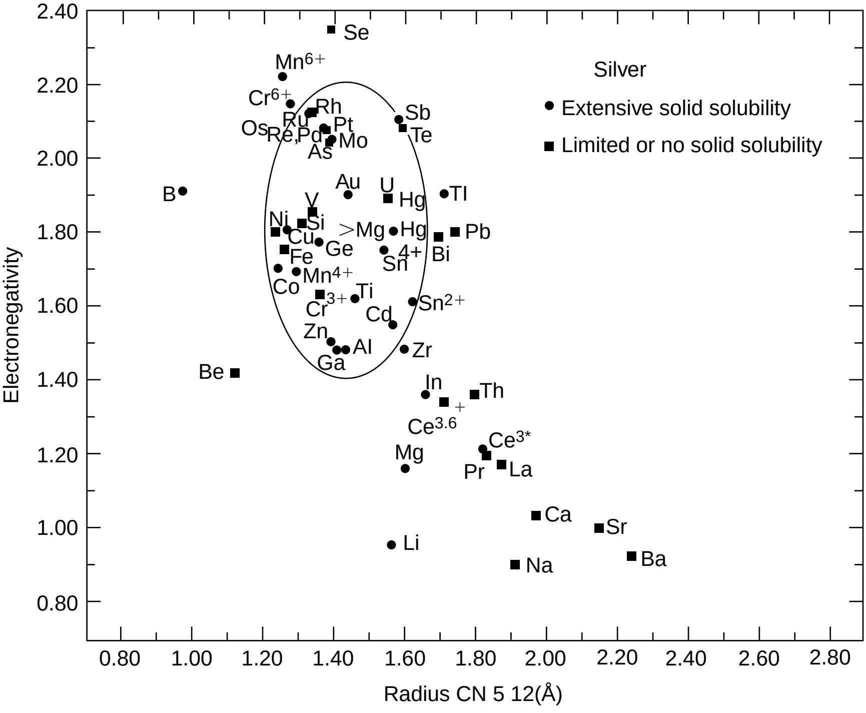


Figure 2.10. Darken–Gurry plot for various solutes dissolved in silver. Circles indicate alloys in which extensive solid solutions are found; squares indicate alloys in which limited or no solid solubility is found. From Pearson, **The Crystal Chemistry and Physics of Meals and Alloys**, Wiley-Interscience, New York, 1972 with permission.

that the electronegativity, according to Pauling, is a measure of the ability of an atom to attract electrons. There are various measures of the electronegativity, most of which are derived from the heats of formation of molecules. However, there are more modern measures of the electronegativity for solids. One such measure for covalently bonded solids is the dielectric-based ionicity.[8c]

The effect of an electronegativity difference on the solid solubility can be understood using tetrahedrally coordinated $A^N B^{8-N}$ covalently bonded semiconductors to illustrate the relation between the heat of formation of the semiconductor from its elemental components and the AB bond ionicity.[8] This ionicity is roughly proportional to the difference in Pauling electronegativity between components. The greater the bond ionicity the more negative is the energy of formation of the AB phase. The same considerations apply to other covalently bonded structures. Hence, if in an AB binary system it is possible for covalently bonded intermediate phases to appear, the greater the ionicity of the AB bond (the greater the electronegativity difference) the more stable will be the intermediate phase. The relative stability of the intermediate and terminal phases will then affect the solid solubility limits (phase boundaries) of the terminal phases, as illustrated in Figure 2.11.

The same results apply when the intermediate phases are not covalently bonded semiconductors in that the electronegativity difference is roughly proportional

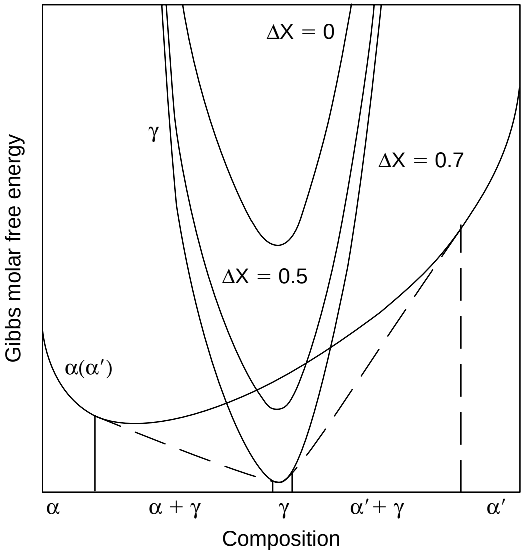


Figure 2.11. Illustrating the effect of an intermediate phase on the phase equilibria in a binary system. Changes in the stability of the intermediate phase will result in changes in the compositions defining the phase boundaries. The ΔX values represent the difference in electronegativity between the pure components.

to the parameter that, for the solid in question, determines the energy of formation of the intermediate phase. For example, the energy of formation of transition metal alloys is governed mainly by the difference in centers of gravity of the d bands of the component transition metals.[9] The latter difference is roughly proportional to the electronegativity difference. Thus, in general, the greater the electronegativity difference is the more stable is the intermediate phase relative to the terminal phases and the less is the limit of solubility of the terminal phases for the solute component.

This effect of electronegativity difference on the solid solubility is not a consequence of an effect on the thermodynamic properties of the terminal phase, but rather of an effect on the thermodynamic properties of the intermediate phase in equilibrium with the terminal phase. This role differs from that of the atomic size difference on the solid solubility in that the latter is the result of an effect on the thermodynamic properties of the terminal phase itself.

The thermodynamic properties of the terminal phase may be affected by other than atomic size difference. Empirically, Hume-Rothery showed that for alloys based on copper, silver and gold that the phase diagrams could almost be superimposed using outer electron concentration in place of atom composition as the abscissa. The effect of valence of the solute was originally explained qualitatively by Jones and Mott[10] and then given a quantitative explanation in the work of Blandin.[11] Other factors are known to affect the thermodynamic properties of metallic solid solutions, such as the overlapping of the d and p levels of the transition and simple metal components, respectively, in their alloys. However, a more thorough discussion of this matter would take us far outside the scope of this book.

Strain energy, ionicity (difference in electronegativity) also affect the equilibrium solubility of donor and acceptor species in semiconductor solid solutions. However, the problem of describing the thermodynamic values of solutes in semiconductor compounds and ionic crystals is much more complicated in that a given solute species can exist simultaneously as a charged or neutral species in substitutional or interstitial sites and sometimes in two oxidation states. Thus, in these solutions each distinguishable species of the same element must be considered a different solute species. This subject has significance for II–VI and some III–V semiconductor alloys where it is desired to increase the solubility of certain dopants on the substitutional sites as charged species in order to enhance the conductivity.

5. Activity and activity coefficient of solutions

The partial molar free energy of a solute component, $\bar{G}_i$, is related to the activity of that component, a_i, by the relation

$$\bar{G}_i - G_i^o = RT \ln a_i$$

where G_i^o is the free energy of one mole of component i in its standard state. Since the standard state is arbitrary, the activity is arbitrary. Usually, for solid solutions, the standard state is chosen to be the pure component. Also, the activity coefficient is defined by $\gamma_i = a_i/X_i$. According to Figure 2.5 and the corresponding section in this chapter, the partial molar free energy of a solution at some composition X_i in the binary system is given graphically by the intercept of the line that is tangent to the molar free energy of the solution at this composition with the coordinate axis representing pure i. Thus, for compositions where the molar free energy versus composition curve has a positive curvature the partial molar free energy of the solute component will increase with increase in the composition of the solute component while that for the solvent component will decrease. A transition from positive curvature to negative curvature of the molar free energy versus composition curve implies that the term $d\ln\gamma_i/dX_i$ must be negative and exceed in magnitude beyond some composition X_i the value of the positive quantity $d\ln X_i/dX_i = 1/X_i$. Any value of the activity coefficient differing from unity implies that the solution is no longer ideal. When the activity coefficient of the solute equals a constant the solution obeys Henry's law. Strain energy and chemical interactions contribute nonzero terms to the solute dependence of the activity coefficient, but of opposite sign.

In general, the results we have derived concerning the thermodynamics of solutions apply, in principle, not only to metallic solid solutions, but to ionic solid solutions, to covalently bonded solid solutions and to polymer solutions. We shall discuss the case of polymer solutions in the next section.

6. Polymer alloys

At one time it was thought that polymers did not form solutions. This deduction was based on two facts and one assumption. One fact is that the relative integral free energy for the solution must be a negative quantity if it is to form. The other fact is that the entropy of mixing for a high molecular weight polymer solution is very small. The assumption was that the enthalpy of mixing was necessarily a positive quantity. That the entropy of mixing of a high molecular weight polymer solution is small can be obtained from equation 2.4, on the assumption that the two polymer components of the solution have the same molecular weight and shape. N in equation 2.4 is the number of molecules. Thus, the ratio of the entropy of mixing for the polymer solution to that for a metallic solution is the reciprocal of the number of mers in the polymer molecule. For a normal high molecular weight polymer molecule the configurational entropy is a negligible quantity. For components that have different numbers of mers per molecule the relation for the entropy of mixing per mer in the solution becomes

$$s_{conf} = -k[(\Phi_1/r_1)\ln \Phi_1 + (\Phi_2/r_2)\ln \Phi_2]$$

where r_i is the number of mers in strand type i and Φ_i is the site fraction occupied by component i.

That polymer solutions do form is now well recognized. Robeson[12] has reported the existence of more than 160 miscible pairs of polymers. It is apparent that the assumption concerning the sign of the enthalpy of mixing for polymer solutions is wrong and that the latter can be a negative quantity. It is interesting to consider the possible origins of a negative enthalpy of mixing in polymer solutions. Polymers have satisfied primary bonds. Thus, the only bonding contribution to the enthalpy of mixing must involve secondary bonds, such as dipole–dipole, hydrogen bonding and van der Waal type bonding. As an example, when the radicals on the component polymers are conducive to hydrogen bonding, as in DNA or PVC/polyester mixtures, it is possible that the enthalpy of mixing of such component polymers will be negative. Because many useful products can be made from polymer alloys this field has experienced a concentration of activity (see Bibliography).

7. Equilibrium in a stressed solid

In stressed solids the criteria defining equilibrium that have been considered in this chapter may no longer be applicable. In particular, equilibrium may not be described by equations 2.8–2.10. Indeed, the concept that the chemical potential of a species in the solid at equilibrium is everywhere the same value is no longer valid in a non-hydrostatically stressed solid. Gibbs showed that for a homogeneous, single-component solid under non-hydrostatic stress in equilibrium with three fluids, each at a pressure equal to one of the non-equal principal stresses of the solid, there will be three different values of the chemical potential of the solid component in the fluids, one for each fluid.*

Larche and Cahn[13] have developed this concept further based on use of a lattice network, in which the the total number of lattice sites (total number of substitutionally sited atoms and vacancies) in an element of volume is conserved, and which does not contain surfaces or dislocations in this reference volume. In this network the thermodynamic properties are functions of the elastic strain and local composition as defined by the network and individual chemical potentials of substitutionally sited species may be meaningless. (A Bravais solid where lattice sites are occupied by substitutional atoms is one example of such a network solid.) Equilibrium may exist in the presence of a non-homogeneous distribution of stress and composition in a network solid. In the network solid the useful thermodynamic quantities are the diffusion potential, to be defined shortly, the Helmholtz-free energy density (per unit volume) and the stress tensor. At equilibrium in the stressed network solid, *the diffusion potential is constant throughout the solid and the divergence of the stress tensor is zero*, the latter to achieve mechanical equilibrium.

---

* It should be mentioned that this situation can occur only if vacancies cannot be created or destroyed at the solid/fluid interfaces, otherwise diffusion of vacancies between interfaces can occur to relieve the principal stresses, i.e. the solid is no longer at equilibrium.

The diffusion potential, M_{IK} is defined by

$$M_{IK} = (1/\rho_o')(\partial f'/\partial C_{IK})_{\theta,E_{ij}}$$

where ρ_o' is the molar density (number per unit volume) of lattice sites in the reference state relative to which the strain is defined, f' is the specific Helmholtz-free energy (energy per unit volume) in the reference state, θ is the absolute temperature, E_{ij} is the strain defined by $E_{ij} = 0.5(u_{ij} + u_{ji})$, where $u_{ij} = \partial u_i/\partial x_j$, u_i is the displacement in the i direction and $(\partial/\partial C_{IK})$ is the variation per unit composition increase in species I, and equal decrease in species K, holding the composition of all other substitutional species on that type site fixed. For binary solutions, $C = C_1$ and $(\partial/\partial C_{12}) = (\partial/\partial C)$. In the case of equilibrium of a network solid (i.e. a closed system) with a fluid

$$M_{IK} = \mu_I^L - \mu_K^L$$

where μ_i^L is the chemical potential of i in the fluid.

For interstitially sited components, the diffusion potential is the chemical potential of the interstitial species.

Inhomogeneous solid solutions are often under conditions of internal stress due to the difference in atomic volume between species and the siting of substitutional solute on lattice sites. Thus, for these systems, equilibrium is reached when each diffusion potential is a constant throughout the system. *At incoherent interfaces, where vacancies can be created and destroyed, all species have defined chemical potentials, which, at equilibrium, are constant everywheres along the interface. At coherent interfaces, where vacancies cannot be created or destroyed, individual chemical potentials for substitutional species are not defined as is the case also in the network solid itself.*

A fluid cannot resist shear stress and has a constant hydrostatic pressure throughout at equilibrium. Chemical potentials of species in a fluid are meaningful and the concepts in the other sections of this chapter apply to fluids. Dislocations can generate and move readily in some solids at temperatures above about one half the absolute melting point. These solids are effectively fluids because they relax stresses rapidly and again chemical potentials of species are meaningful in them.

Larche and Cahn[13] have considered many applications of these concepts. A complete review here is outside the scope of this book. However, we shall undertake to describe some results of their analysis in later chapters in this book. It should be noted also that both the Gibbs–Duhem and the Clausius–Clapeyron equations are altered in non-hydrostatically stressed solids from the corresponding equations applicable to fluids.[14]

8. Remarks about liquid solutions

Liquid solutions may exhibit phenomena not described in this chapter about solid solutions. For example, in water as a solvent, amphiphilic molecules may form micelles above a critical concentration, a transition which is not a phase transition. The interaction between one part of the solute molecule, the hydrophobic end, and water causes a rearrangement of the water molecules when the hydrophobic end inside the micelle no longer is adjacent to these molecules. This transfer increases the entropy of the water molecules while at the same time the configurational entropy and positional entropy of the amphiphilic molecules is reduced. Water, as any polar liquid, can be induced by a hydrophobic molecule to have an orientational short-range order. Although the thermodynamics of solutions does not distinguish between liquid and solid solutions the interpretation of the origin of the relevant thermodynamic parameters may vary between them. In other words, the phenomenological thermodynamic description of solutions does not distinguish between liquid and solid solutions, except for modifications due to elastic strain. However, the statistical thermodynamic description may vary widely between liquid and solid solutions.

9. Hard sphere fluid in a cell model

It is possible using the relationships developed in this chapter to investigate the effectiveness of a model for a liquid or fluid phase that is suggested by the concept that a fluid or liquid of ideal monodisperse hard spheres is a solution of hard spheres and empty sites in a volume given by the product of the number of sites and the volume of a sphere.

For the volume fraction, φ, of particles in the fluid near $\varphi = 0.5$ we may approximate the entropy of disorder of the particles by the usual entropy of mixing term involving the mixing of $n = N\varphi$ particles and $N - n$ empty sites on N lattice sites, i.e. this contribution to the entropy of the fluid is then

$$S_{\text{disorder}}(\text{fluid}) = -Nk[\varphi \ln \varphi + (1 - \varphi)\ln(1 - \varphi)]$$

There is another contribution to the entropy of the fluid, that of the vibrational entropy of the spheres within the free volume provided by the stationary neighboring spheres. This contribution to the entropy of the fluid is then

$$S_{tr}(\text{fluid}) = N\varphi k \ln[v_f(\text{fluid})/\Lambda^3]$$

The parameter Λ is the thermal wavelength and v_f(fluid) is the free volume per sphere in the fluid phase. In the solid, fcc, phase the entropy of vibration is the same as that of the fluid except that v_f(fcc) replaces that for the fluid phase. Hence, the Helmholtz free energy of the fluid phase is

$$F(\text{fluid}) = -3N\varphi kT/2 - TS_{\text{disorder}}(\text{fluid}) - TS_{tr}(\text{fluid})$$

The free energy of the fcc phase is

$$F(\text{fcc}) = -3N\varphi kT/2 - TS_{tr}(\text{fcc})$$

Fortunately, the free volumes in units of σ^3 may be obtained from studies of the statistics of voids in the two phases. For the fluid phase

$$v_f(\text{fluid}) = \exp[-21.96\varphi + 5.575]$$

(P.G. Debenedetti and T.M. Truskett, Fluid Phase Equilibria 549(1999) 158–160.) and for the fcc phase

$$v_f(\text{fcc}) = \{2^{5/2}[(2^{-5/6}\{4\pi/3\}^{1/3}\varphi^{-1/3}) - 1]^3\}$$

(P. Ziherl and R.D. Kamien, arXiv:cond-mat/0007256v1, 14 July 2000.)

To obtain the volume fractions that are in equilibrium we should plot the molar Helmholtz free energy versus volume (mole) fraction according to the discussion in this chapter relating to Figure 2.8. The reduced free energies (F/NkT) except for the terms in common ($3\varphi/2$ and $\varphi \ln[\Lambda^3]$) are plotted in Figure 2.12.* As shown, the common tangent to these curves occurs at $\varphi = 0.494$ and 0.546. These values are also those obtained on analytically evaluating the partial molar free energies for the two phases and setting them equal. These values are to be compared to the values obtained by Hoover and Ree of 0.494 and 0.545 using a Monte Carlo procedure to evaluate the entropies.

* The constant term in common subtracts out and those proportional to φ merely skew the curves without changing the lowest common tangent and its values of φ.

What is the source of the entropy of fusion? To find the entropy of fusion we must evaluate the entropy of the fluid at $\varphi = 0.494$ and that for the fcc solid at $\varphi = 0.546$ at the same number of spheres. The difference between these two entropy values is the entropy of fusion. Carrying out this calculation we have

$$
\begin{aligned}
&-(k/0.496)[0.496\ln 0.496 + (1 - 0.496)\ln(1 - 0.496)] \\
&\quad + k\ln v_f(\text{fluid})_{\varphi=0.496} - k\ln v_f(\text{fcc})_{\varphi=0.545} \\
&= 1.057k
\end{aligned}
$$

This value can be compared to the value obtained by Hoover and Ree of 1.067k. On the basis of this agreement then the main contribution to the entropy of fusion appears to be a configurational entropy term for the fluid evaluated by mixing the actual number of spheres in a fluid that coexists with the solid phase in a lattice of cells the number of which is given by the actual volume divided by the volume of a sphere. A minor contribution comes from the lower free volume in this fluid over that in the coexisting fcc solid, a contribution stemming from the vibrational configurations in these free volumes and which acts to counteract the effect of the configurational entropy term. From the agreement for both the coexisting volume

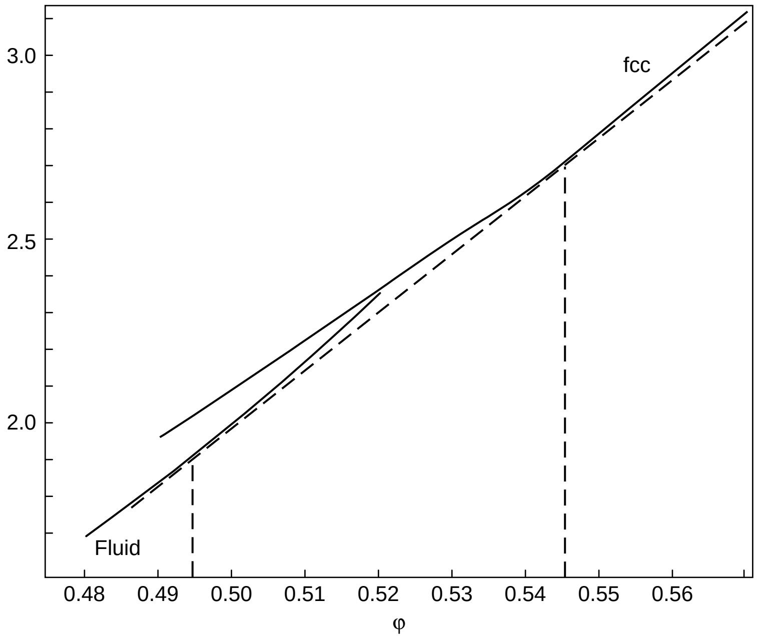


Figure 2.12. Plots of F(fluid)/NkT and F(fcc)/NkT minus the terms in common ($-3\varphi/2$ and $\varphi\ln\Lambda^3$) showing the densities of common tangency.

fractions and the entropy of fusion we may conclude that the present analytical method of evaluating these parameters are as accurate as the Monte Carlo-based procedure of Hoover and Ree.

10. Summary

To summarize, we have investigated the forms that the free energy–composition curves can take in binary systems. We have derived the relation for the free energies of mixtures of two solutions and thereby have been able to describe when such a mixture will be stable and when the solution will be stable relative to the mixture of solutions. Also, we have graphically interpreted the significance of the partial molar free energy. Further, we have discussed the concepts of short-range order, strain energy and chemical interactions between the components of a homogeneous solid solution. We have introduced the concepts of polymer alloys and of equilibrium in stressed network solids. In the latter, chemical potentials may not have meaning, but diffusion potentials replace them as the quantity that is constant throughout coexistant phases in equilibrium.

Appendix 1

Definitions of thermodynamic quantities for constant T and P condition

Partial molar quantities

$$\bar{Q} = (\partial Q/\partial n_i)_{P,T,n.} \quad (k \neq i)$$

where n_i is the number of moles of i. (Replacing n by the number of atoms N and with Q, the total Gibbs-free energy, yields the chemical potential.)

Relative integral molar quantities

$$Q^M = Q_m - \sum(X_i {}^oQ_i)$$

(Superscript o refers to the pure state, subscript i to the component i and subscript m to the value for one mole of substance.)

Relative partial molar quantities

$$Q_i^M = \bar{Q}_i - {}^oQ_i$$

Excess quantities

$$\text{Integral: } Q^{xs} = Q^{M} - Q^{id}$$

$$\text{Partial: } \bar{Q}^{xs} = Q_i^{M} - Q_i^{Mid}$$

(Superscript id refers to the ideal solution value defined below.)

Types of solutions

$$\text{Ideal: } H^{M} = 0;\ S^{M} = -R\sum[X_i \ln X_i]$$
$$\text{Regular: } H^{M} = bX_1X_2\ \text{(binary solution) with } b \neq 0;\ S^{M} = S^{Mid}$$

Binary solutions – Correlating functions

$$\alpha_1 = G_1^{Mxs}/(X_2)^2 = [RT\ln\gamma_1]/(X_2)^2$$
$$\alpha_2 = [RT\ln\gamma_2]/(X_1)^2 = -X_2\alpha_1/X_1 + (X_1^{-2})\int_0^{X}\alpha_1\,dX_1$$
$$\beta_1 = S_1^{M,xs}/X_2^2 = -d\alpha_1/dT$$

Relations between Quantities

$$Q_m = X_1\bar{Q}_1 + X_2\bar{Q}_2$$
$$Q^{M} = X_1Q_1^{M} + X_2Q_2^{M}$$

$\bar{Q}_1 = Q_m + (1 - X_1)(\partial Q_m/\partial X_1)_{X_2 \ldots}$ (Multicomponent solution; for binary solution substitute total derivative in place of partial derivative.)

$Q_1^{M} = Q^{M} + (1 - X_1)(\partial Q/\partial X_1)_{X_2 \ldots}$ (Multicomponent solution; for binary solution substitute total derivative in place of partial derivative.)

$$X_1\,d\bar{Q}_1 + X_2\,d\bar{Q}_2 = 0 \quad \text{(Gibbs–Duhem)}$$

$$(\partial Q_m/\partial X_2)_{X_{1\ldots}} = \bar{Q}_2 - \bar{Q}_1$$

$$Q_1^{M,xs} = Q_1^{M} - Q_1^{M,id}$$

References

1. H. Bethe, Proc. Roy. Soc. A 150, 552(1935).
2. R. Kikuchi, Phys. Rev. 81, 988(1951); R. Kikuchi and S.G. Brush, J. Chem. Phys. 47, 195(1967).

3. D. de Fontaine, Solid State Physi. 50, 73(1979); T. Mohri, J.M. Sanchez and D. de Fontaine, Acta Met. 33, 1171(1985).
4. J. Friedel, Adv. in Phys. 3, 446(1954).
5. J.D. Eshelby, Solid State Phys. 3, 79(1956).
6. J.M. Sanchez, J-P. Stark and V.L. Moruzzi, Phys. Rev. B 44, 5411(1991).
7. W. Hume-Rothery and G.V. Raynor, **The Structure of Metals and Alloys**, The Institute of Metals, London, 1962.
8. a) P.N. Keating, Phys. Rev. 145, 637(1966); b) A. Zunger in **Handbook on Semi Conductors**, ed. T.S. Moss, Elsevier Science B.V., Amsterdam, 1994, p. 1399; c) J.C. Phillips, **Bonds and Bands in Semiconductors**, Academic Press, New York, 1973.
9. D.G. Pettifor, Calphad, 1, 305(1977).
10. N.F. Mott and H. Jones, **The Theory of the Properties of Metals and Alloys**, Oxford University Press, Oxford, 1936.
11. A. Blandin in **Alloying Behavior and Effects of Concentrated Solid Solutions**, ed. T.B. Massalski, Gordon and Breach, NY, 1963; in **Phase Stability in Metals and Alloys**, ed. P.S. Rudman, J. Stringer and R.I. Jaffee, McGraw-Hill, New York, 1967.
12. L.M. Robeson, in **Polymer Compatibility and Incompatibility: Principles and Practice**, MMI SYMP. SER, 3, ed. K. Solc, Cooper Station, New York (1981).
13. F. Larche and J.W. Cahn, Acta Met. 21, 1051(1973), 26(53), 1579(1978); 30, 1835(1982).
14. W.C. Johnson and H. Schmalzried, Acta Mettall. Mater. 40, 2337(1992).
15. A.G. Khachaturyan, **Theory of Structural Transformations in Solids**, John Wiley, New York, 1983.

Bibliography

General:

1. R.A. Swalin, **Thermodynamics of Solids**, 2nd ed., John Wiley, New York, 1972.
2. D. Gaskell, **Introduction to Metallurgical Thermodynamics**, Hemisphere Publ. Corp., New York, 1981.

Metal Solutions:

L.S. Darken and R.W. Gurry, **Physical Chemistry of Metals**, McGraw-Hill, New York, 1953.

Semiconductor Solutions:

G.B. Stringfellow, **Organometallic Vapor–Phase Epitaxy, Theory and Practice**, Academic Press, New York, 1989, Chapter 3. This chapter is an excellent summary of the thermodynamic developments in semiconductor solid solutions.

Polymer Solutions:

1. J.W. Barlow and D.R. Paul, Ann. Rev. Mat. Sci. 11, 299(1981).
2. L.A. Utracki, **Polymer Alloys and Blends**, Hanser, Oxford University Press, New York, 1989.

Problems

1. Show why the relation $dG/dX_2|_\alpha = dG/dX_2|_\beta$ is insufficient to define the compositions of the coexisting phases α and β at thermodynamic equilibrium.
2. Why does the functional dependence of free energy on composition in a binary system exhibit either a positive curvature at all compositions, or a sequence in which the curvature alternates from positive to negative and then back to positive as the composition varies from pure solvent to pure solute?
3. Why is there a thermodynamic driving force for impurities to form a solution with an absolutely pure element?
4. What thermodynamic factor is responsible for the existence of a positive curvature to the dependence of free energy of a binary solid solution on composition?
5. Evaluate the activity coefficient corresponding to a regular solution where the factor b equals 1000 cal/g atom at $X_2 = 0.3$.
6. Show where, in a plot of molar free energy versus composition for a single solution, the points corresponding to the partial molar free energy of the solute component (2) and that of the solvent component (1) lie, for a composition of the solution equal to $X_2 = 0.75$.
7. Demonstrate in a plot of molar free energy versus composition that the partial molar free energies in two different phases are not equal at compositions differing from the solvus compositions for both phases, whereas show that they are equal at the latter compositions.
8. Use the appropriate relation in the Appendix to show that a positive partial excess molar enthalpy of any component will yield a miscibility gap in a binary phase diagram.
9. The Ge–Si binary phase diagram reveals complete mutual solid solubility of the components. However, the tetrahedral covalent radii are 1.225 and 1.173Å for Ge and Si, respectively. Calculate the critical temperature corresponding to the expected miscibility gap based on the assumption that strain energy provides the only contribution to the excess enthalpy of solution.
10. All pseudo-binary solid solutions of semiconductor binary compounds reveal positive values for the energy of mixing, whereas the energy of formation of the binary compounds from their elements is negative. Also, despite the positive energies of mixing in the pseudo-binary solutions, in many cases ordered arrangements of the components are found in these solutions. Provide reasonable explanations for these facts.
11. Si normally acts as an n-type donor in GaAs occupying Ga-type sites. Given that the Si bond is smaller than the Ga–As bond and assuming that the solubility of Si in GaAs is limited by the strain energy developed on solution of Si speculate how the coaddition of each of the following ternary solutes would affect the solubility of Si in GaAs: Be, Sb, Te providing a reasonable explanation for each effect.

CHAPTER III

Free Energy and Phase Diagrams

Introduction

In the previous chapter the bases for the development of an understanding of the relationship between free energy–composition diagrams and phase diagrams were explored. We shall in Section A below attempt to provide an understanding of the factors responsible for the effect of composition on the solid/liquid equilibria; the influence of intermediate phases on phase diagrams; the concept of metastable phases; the development of phase diagrams from the temperature dependence of the free energy–composition curves; and, finally, methods for the prediction of ternary phase diagrams. In Section B we shall present first the basic principles regarding heterogeneous chemical equilibria and then we shall relate these principles to ternary phase diagrams and equilibrium between contacting compounds.

A. FREE ENERGY AND PHASE DIAGRAMS – BINARY SYSTEMS

1. Solid–liquid equilibria

1.1. Phase boundary compositions and their distribution coefficient

From the condition governing the equilibrium between phases, the relations for the solid–liquid equilibrium in a binary alloy can be set down as follows

$$\bar{G}_1^S = \bar{G}_1^L; \quad \bar{G}_2^S = \bar{G}_2^L \tag{3.1}$$

But, $\bar{G}_i = {}^{xs}G_i + {}^{o}G_i + RT \ln X_i$ (actually, the latter relation acts to define the excess molar free energy of component i, ${}^{xs}G_i$, in terms of the partial molar free energy of component i and the molar free energy of component i in its pure state, ${}^{o}G_i$). Substitution into equation 3.1 and solution of the resulting equations yield

$${}^{S}X_2/{}^{L}X_2 = \exp\{[({}^{oL}G_2 - {}^{oS}G_2) + ({}^{xsL}G_2 - {}^{xsS}G_2)]/RT\} \tag{3.2a}$$

$$^{S}X_1/^{L}X_1 = \exp\{[(^{oL}G_1 - ^{oS}G_1) + (^{xsL}G_1 - ^{xsS}G_1)]/RT\} \qquad (3.2b)$$

Let us postpone consideration of the significance of these equations for the distribution coefficient (i.e. ratio of solidus to liquidus compositions) until we have considered another concept that will enable us to understand which factors are responsible for the change in "average melting point" on addition of solute to a host solvent.

1.2. Composition dependence of "average" of solidus and liquidus temperatures

Figure 3.1 shows the free energy–composition curves for a solid phase and a liquid phase. Attention is directed to the composition X* at which the two phases have both the same composition and same values of free energy, i.e. the composition at which the two curves intersect. From our previous discussion, we are able to conclude that X* lies in the stable two phase solid plus liquid region between the solidus composition ^{S}X and the liquidus composition ^{L}X for the temperature corresponding to the free energy curves. Thus, X*(T) may be considered to be an average of the solidus and liquidus compositions at the temperature T and hence the inverse function T(X*) may be considered to be an average of the solidus and liquidus temperatures corresponding to the composition X*. For the sake of brevity we

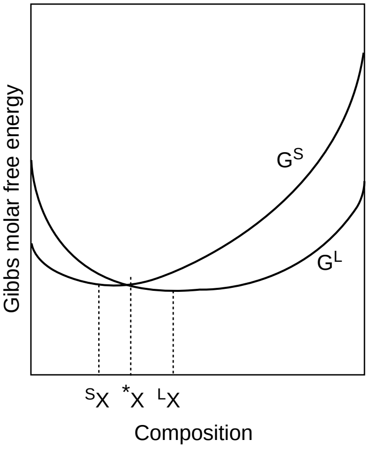


Figure 3.1. Illustrating definition of composition where $G^S = G^L$. Since free energy curves correspond to a particular temperature T then T(*X) is also defined.

shall call T(X*) the "average melting point". An attempt will now be made to derive the temperature dependence of the composition X*.

The function X*(T) is defined by

$$G_m^L(X^*) = G_m^S(X^*) \tag{3.3}$$

Equation 3.3 can also be written as follows

$$^*X_1\bar{G}_1^L + {}^*X_2\bar{G}_2^L = {}^*X_1\bar{G}_1^S + {}^*X_2\bar{G}_2^S \tag{3.4}$$

Let $\Delta G_i = {}^LG_i - {}^SG_i$. Also, by the definition for the excess free energy, i.e.

$$\bar{G}_i = {}^oG_i + {}^{xs}G_i + RT\ln X_i$$

Substitution into equation 3.4 yields

$$^*X_1(\Delta^oG_1 + \Delta^{xs}G_1) + {}^*X_2(\Delta^oG_2 + \Delta^{xs}G_2) = 0 \tag{3.5}$$

Taking the derivative of equation 3.5 with respect to T yields

$$\begin{aligned} 0 = & -{}^*X_1(\Delta^oS_1 + \Delta^{xs}S_1) - {}^*X_2(\Delta^oS_2 + \Delta^{xs}S_2) \\ & + (\Delta^oG_2 - \Delta^oG_1 + \Delta^{xs}G_2 - \Delta^{xs}G_1)[d^*X_2/dT] \end{aligned}$$

Solving yields

$$[dT/d^*X_2] = \frac{\{\Delta^oG_2 - \Delta^oG_1 + \Delta^{xs}G_2 - \Delta^{xs}G_1\}}{({}^*X_1[\Delta^oS_1 + \Delta^{xs}S_1] + {}^*X_2[\Delta^oS_2 + \Delta^{xs}S_2])} \tag{3.6}$$

To simplify, we consider the case where

$$\Delta S = \Delta^oS_1 = \Delta^oS_2$$

(Thus, the following applies only to those binary systems for which the approximation that the entropy of fusion is the same for both components is valid.) It is also reasonable to assume that $\Delta^{xs}S \ll \Delta^oS$. Further, we will make use of the approximation that near the melting point, ^MT

$$\Delta^oG_i = ({}^MT_i - T)\Delta^oS_i$$

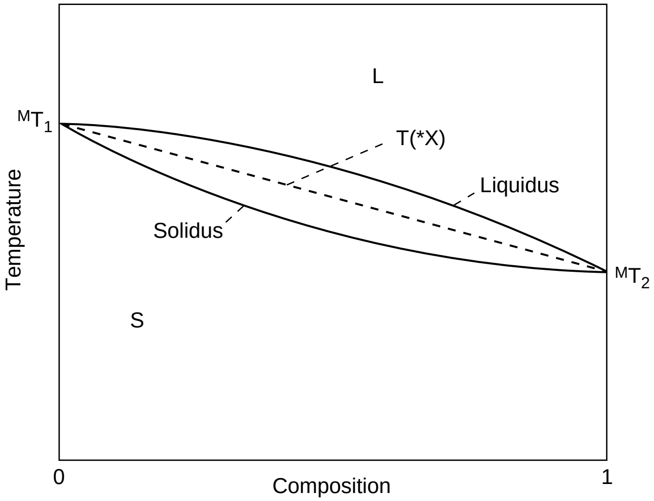


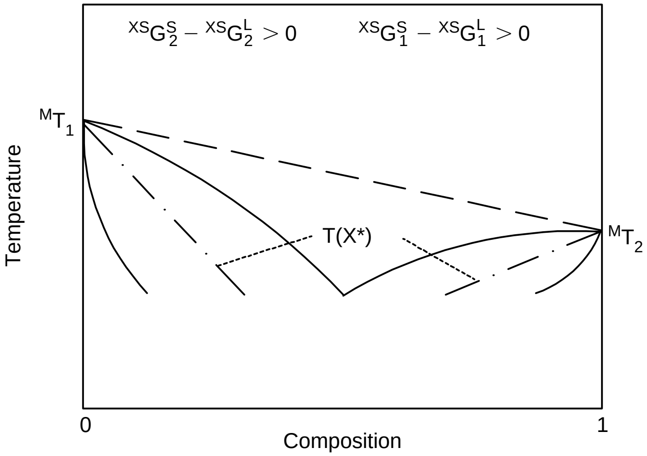
Figure 3.2. Schematic phase diagram for ideal liquid and solid solutions.

With these substitutions in equation 3.6 we obtain

$$[dT/d{*}X_2] = (^MT_2 - {}^MT_1) + \{(\Delta^{xs}G_2 - \Delta^{xs}G_1)/\Delta^oS\} \tag{3.7}$$

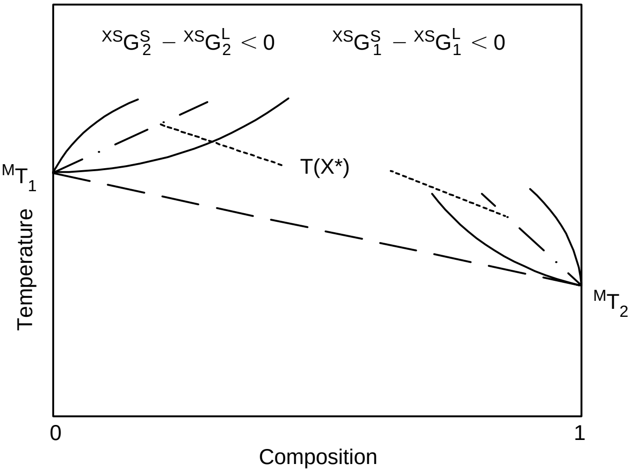
Thus, the first term on the right-hand side of equation 3.7 is the slope of the line joining the melting points of the two pure components in the phase diagram (see Figure 3.2). The second term on the right-hand side of equation 3.7 equals zero for an ideal solution. Hence, for an ideal solution, the liquidus and solidus curves must lie on either side of the line joining the melting points of the two components, as shown in Figure 3.2!

If the binary system is not ideal then we may use equation 3.7 to determine the behavior of the "average melting point". Let component 1 denote the solvent. For $X_2 \ll X_1$ the assumption that the absolute value of the excess free energy of component 1 is less than that of component 2 is a good approximation. Thus, the sign of the deviation of the slope of the "average melting point" from the line joining the melting points of the two components is given by the sign of $\Delta^{xs}G_2$. Generalizing, at a given composition, if the excess free energy of any component in its solid phase is more positive than that in its liquid phase, then for increasing composition of that component there will be a negative deviation of the "average melting point" relative to that for the ideal solution, as shown in Figure 3.3. With this relationship between these excess free energies reversed, the "average melting point" will exceed that for the ideal solution, as shown in Figure 3.4.



$^{XS}G_2^S - {}^{XS}G_2^L > 0$
$^{XS}G_1^S - {}^{XS}G_1^L > 0$
$^MT_1$
$^MT_2$
T(X*)
Temperature
0
1
Composition

Figure 3.3. Schematic section of a phase diagram illustrating the negative deviation of T(X*) when the partial molar excess free energies for the solid exceed those for the liquid.



$^{XS}G_2^S - {}^{XS}G_2^L < 0$
$^{XS}G_1^S - {}^{XS}G_1^L < 0$
$^MT_1$
$^MT_2$
T(X*)
Temperature
0
1
Composition

Figure 3.4. Schematic section of a phase diagram illustrating the positive deviation of T(X*) when the partial molar excess free energies for the solid are less than those for the liquid.

In Chapter IV we will make use of equations 3.2 and 3.7 to provide a basis for the prediction of the Gibbs adsorption of solute to interfaces on the assumption that the local atomic or molecular environments at interfaces approximate to those in liquids. Also, we shall show in the next section why sometimes the sign of the deviation between the "average melting point" and the weighted mean pure component melting point will correlate to the existence of phase separation or order in isomorphic phase diagrams.

2. Origin of eutectic phase diagram in isomorphic systems

When the enthalpy of mixing H^M (the relative integral molar enthalpy) has a positive value there will be phase separation or a miscibility gap in an isomorphic system. It is also true that if the partial excess molar enthalpy is positive for $0 < X_2 < 1$ then there must be a miscibility gap in the system. (The partial molar excess enthalpy for the solute component 2 corresponding to some composition X_2^* is given by the intercept with the ordinate at $X_2 = 1$ of the tangent to the molar enthalpy curve at that composition for $0 < X_2^* < 1$.)

Thus, the conditions required to produce an eutectic type phase diagram are that ${}^{xs}H_2^s > 0$ and that ${}^{xs}H_2^s > {}^{xs}H_2^1$. These conditions are satisfied by the case where the *only* contribution to the partial molar excess enthalpies is that due to atomic (ionic) size induced strain energy. This situation is found to occur unambiguously in phase diagrams involving ionically bonded solids in which the cation is isovalent. For example, Figure 3.5 shows the phase diagrams between first column fluorides and CsF, where the deviation in ionic radius between Cs and the other first column elements increases in the sequence: Rb, K, Na, Li. These phase diagrams qualitatively obey equations 3.6 and 2.27 in that as the deviation in ionic radius between Cs and the other cations increases the negative deviation of the “average melting point” from the weighted mean melting points of the pure components increases. Also, the solid solubility of one component in the other decreases in this sequence. The lack of first neighbor interactions between cations in ionically bonded solids eliminates many interactions that occur in alloy systems. A consequence of this fact is that it is possible to illustrate the effect of the strain energy in phase diagrams of ionic solids by limiting the cation to an isovalent sequence.

In ionically bonded solids, solutes differing in valence from that of the solvent will have corresponding excess partial molar enthalpies that are positive, because the introduction of such solutes into the lattice requires the concomitant production of defects to maintain electroneutrality. Further, this enthalpy will be less positive in the liquid state where the excess charge of the solute can be easily screened by motion of the anions relative to it. Thus, except for the cases where long-range interactions, excepting the electrostatic interaction, can exist between cations (e.g. magnetic interactions), it is expected that the phase diagrams of ionically bonded systems will not exhibit peritectic behavior or positive deviations of the “average melting point” from the weighted mean pure component melting point. Examination of ceramic phase diagrams involving ionically bonded solids

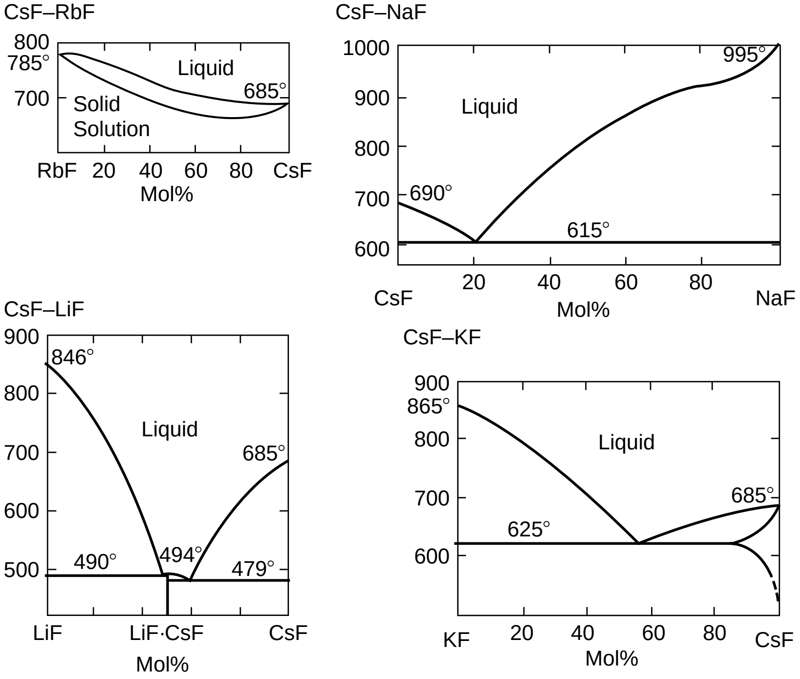


Figure 3.5. Phase diagrams of CsF with the other isovalent fluorides. The deviation in ionic size increases in the sequence Rb, K, Na, Li. Phase diagrams are from E.M. Levin, C.R. Robbins, and H.F. McMurdie, **Phase Diagrams for Ceramists**, The American Ceramic Society, Columbus, 1964 with permission.

reveals that this expectation is obeyed (i.e. most phase diagrams exhibit eutectic type behavior of the solidus and liquidus phase boundaries).

Nearest-neighbor and/or many-body interactions are prevalent in alloy and covalently bonded systems. Hence, it is much more common to observe positive deviations of the "average melting point" from that of the ideal solution in their phase diagrams than in ionic systems. One source of this behavior stems from bonding interactions that are characteristic of the solid state and are absent in the liquid state. In multistructural systems the first term in equation 3.6 provides a negative deviation of the "average melting point" from the weighted mean melting point of the pure components. For these systems $\{(G_2^{o,s} - G_2^{o,l}) - (G_1^{o,s} - G_1^{o,l})\}$ is always negative when 2 is the solute. The solute in multistructural (non-isomorphous) phase diagrams always has a crystal structure that is metastable with respect to the stable structure of the solvent (component 1).

3. Solid–solid equilibria equivalent to the solid–liquid case

The relationships derived above can be applied to the equilibria between polymorphic phases that extend into the composition coordinate merely by substituting for the subscript S that corresponding to the phase that is stable at low temperature in the pure host, for the subscript L that corresponding to the phase stable at high temperature and interpret ${}^{o}T_M$ as the equilibrium transformation temperature between these two structures. It is usually difficult to predict how solute affects the relative stability of the two competing polymorphs. However, if one of these competing polymorphs has the same crystal structure as the stable structure of the solute, then the effect of solute is to stabilize this structure. Examples illustrating this effect are given in Figures 3.6 and 3.7. The stable structure of Ni is fcc and, as shown, Ni as a solute stabilizes the fcc structure of Fe. The stable structure of Mo is bcc and Mo as a solute stabilizes the bcc structure of Fe. This particular role of solute is exerted through the term $\Delta^{o}G_2$ in equation 3.6.

The term $\Delta^{xs}G_2$ in equation 3.6 also can exert significant effects on the relative stability of competing polymorphs. For example, the effect of solute on the magnetic interaction between Fe atoms is manifested by means of this term.[1] Also, the difference in strain energies induced in the competing polymorphs by interstitial solutes, as described quantitatively in equation 2.27, enters into this term. For example, interstitial solutes, such as carbon and nitrogen, are larger than the cavities into which they fit in either bcc or fcc iron. However, the misfit is greater for the bcc structure. As a consequence, carbon or nitrogen as a solute in iron

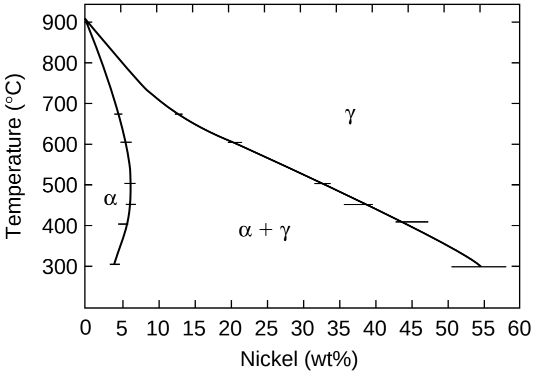


Figure 3.6. Fe–Ni phase diagram. From Metall. Trans. 11, 1151(1980) with permission.

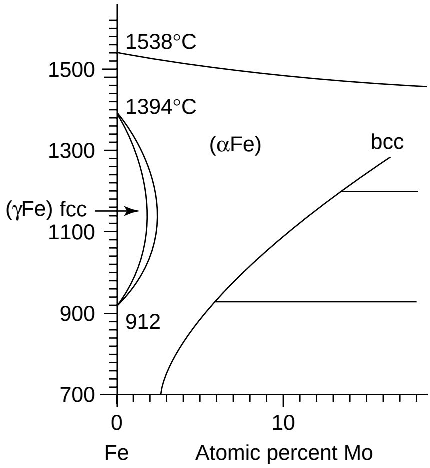


Figure 3.7. Fe–Mo phase diagram. Extracted from **Binary Alloy Phase Diagrams**, Vol. 1, ASM, 1986 with permission.

stabilizes the fcc structure, as shown in Figure 3.8. Similar effects can be shown to exist in ceramic phase diagrams and in phase diagrams for covalently bonded systems.

The solid state transformations differ from the melting transition in monostructural systems in one important aspect. The entropy of the transformation between two given crystal structures can differ markedly between elements. Thus, equations 3.2 and 3.6 are appropriate descriptions, whereas equation 3.7 is not, of the dependence on solute of the transformation temperature. For example, although for most normal metals $S_{bcc} - S_{fcc} > 0$, as suggested in Chapter 1, the reverse is valid for the Periodic Groups V and VI transition elements and for the low temperature transition in iron. This effect of the entropy of transformation is also likely to be valid for solid state transformations in ionic and covalently bonded solids.

4. Intermediate phases

Intermediate phases centered about stoichiometric compositions will also have associated free energy–composition curves that tend to have minima at the corresponding stoichiometric composition, as shown in Figure 2.12. Normally, the curvature in the vicinity of the minimum free energy will tend to be large, although

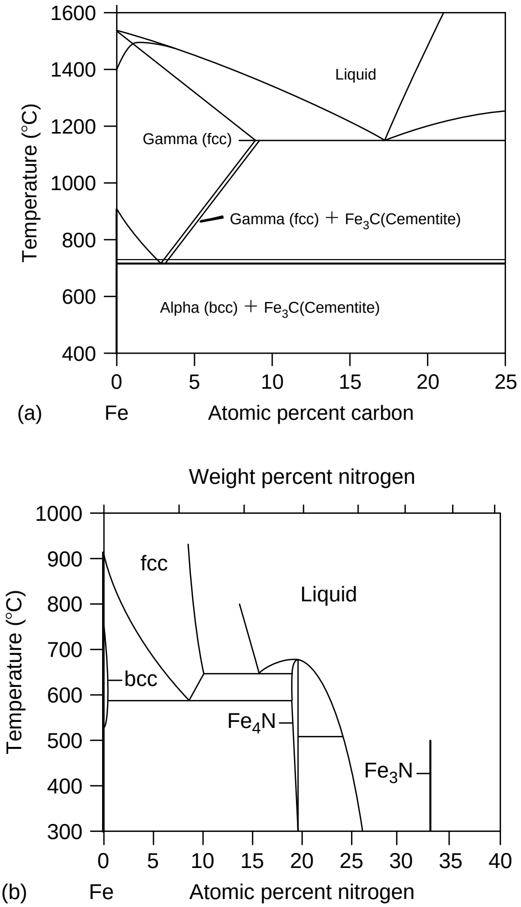


Figure 3.8. (a) Fe–C phase diagram. Note both stable and metastable diagrams and (b) Fe–N phase diagram.

this is not a necessary condition. If any part of the free energy–composition curve for the intermediate phase falls at a lower free energy value than that corresponding to the most stable arrangement for the terminal phases, at any composition, then the intermediate phase becomes stable relative to the partially stable situation for the terminal phases, as shown in Figure 2.12.

Examination of Figure 2.12 leads to the conclusion that any intermediate phase that is stable must have a range of composition over which this phase is stable. The larger is the curvature in the vicinity of the minimum in the free energy–composition curve for the intermediate phase the smaller is the composition range of stability of this phase. There have been attempts in the past to categorize intermediate phases into two groups: one exhibiting a large range of compositional stability and the other exhibiting a negligible one. It is questionable that the population distribution frequency as a function of the compositional range of stability is bi-modal. At least, this author has not seen any evidence concerning the modality of this distribution. In any case, it should be apparent that the limits of composition over which the intermediate phase is stable are functions also of equilibrium with the adjoining phases (i.e. of where the common tangents lie) and not only of the curvature of the free energy–composition curve for the intermediate phase at its minimum.

In general, in alloy systems, intermediate phases tend to have more unlike nearest-neighbor bonds than do solid solutions at the same composition. Hence, if there is a tendency for the unlike nearest-neighbor bonds to be more stable than the average for the like atom bonds then there is a tendency for the intermediate phase to be stabilized. This tendency to stabilize unlike atom bonds can be related to several factors, depending upon the nature of the components. Alloy theory is concerned with the identification and description of these factors. There are intermediate phases, however, which depend upon factors other than those related to the maximization of the number of unlike bonds for their stability. Indeed, in the phase diagrams of ceramic systems we encounter intermediate phases the stability of which have different origins than in alloy systems. For example, the spinel phase $MgAl_2O_4$ is an intermediate phase in the MgO–Al_2O_3 binary system and does not depend for its stability on Mg–Al bonds, which do not exist in the spinel structure. An exploration of the origins of the stability of intermediate phases would be of interest, but is not relevant to the present objective. Here, we are concerned only with the effect of stable intermediate phases upon the phase diagram.

If the intermediate phase is only just slightly more stable than the most stable mixture of its bounding phases, as occurs for example in the Ni–Al system where the intermediate phase Ni_3Al has the terminal Ni base solid solution and another intermediate phase NiAl as its bounding phases (see Figure 3.9), then it is possible that there will be relatively large compositional ranges of stability for all three phases. On the other hand, the more general situation is that the intermediate phase is sufficiently more stable than the most stable mixture of its bounding phases so that these compositional ranges of stability are limited. If one of the bounding phases is a terminal solid solution then the solvus line corresponding to the limit of solubility of this phase will fall close to the terminal composition, i.e. it is said in this case that the solubility of the terminal solid solution is limited.

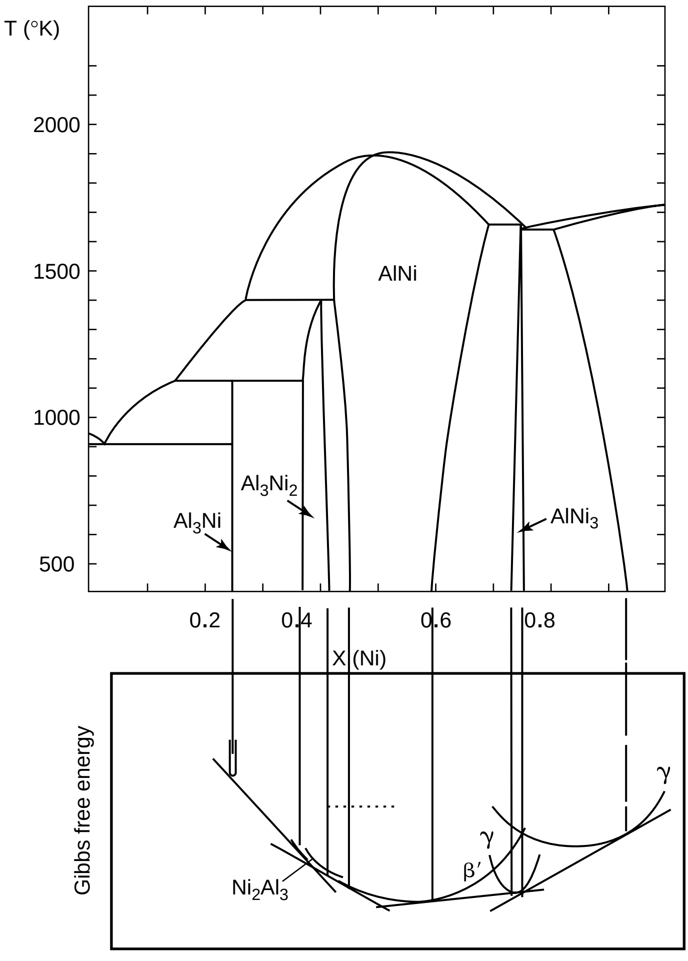


Figure 3.9. Ni–Al phase diagram showing relation to free energy–composition curves (400°K) for all but one of the various phases. Wide composition range of phase stability must coincide with small deviations in free energy between coexisting phases.

5. Metastability

The concept of the metastability of a phase in a binary system and some consequences of this metastability can now be discussed. Consider the free energy–composition diagram shown in Figure 3.10. As shown, the intermediate phase is not stable relative to the mixture of terminal phases alpha and beta. However, suppose that the kinetics of formation of the beta phase are so slow relative to that for the

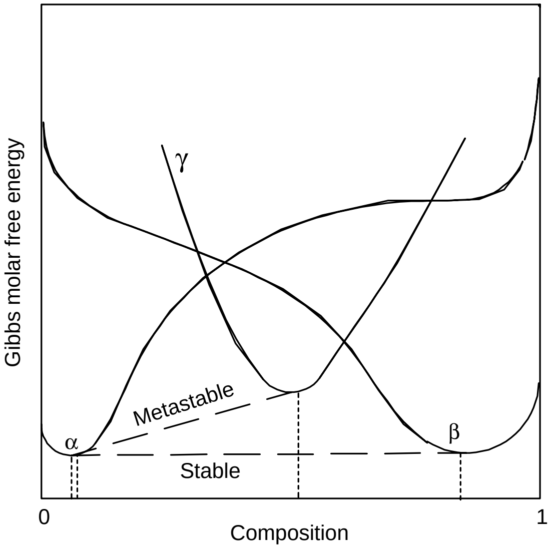


Figure 3.10. Illustrating the partial equilibrium possible between a metastable phase and a stable phase.

intermediate phase, gamma, as to allow the gamma phase to form. In this case, it is possible to attain a metastable equilibrium between the alpha and gamma phases described by the common tangent to the free energy–composition curves for these two phases. One consequence of the fact that the gamma phase is unstable relative to the mixture of the alpha and beta phases is that *the solvus line for the alpha phase in metastable equilibrium with the gamma phase must be contained within the solvus lines corresponding to the stable equilibrium between the alpha and beta phases.* A more general rule of phase diagram construction follows from this property of metastability – **the solvus lines separating single phase regions from two phase regions in a binary system must extrapolate into two phase regions.** The student is urged to construct the free energy–composition curves corresponding to the two solid phases and liquid phase for an eutectic phase diagram and for two temperatures – one above and one below the eutectic temperature as a means of appreciating the basis for this rule.

6. Temperature dependence of free energy–composition curves and phase diagrams

It should be apparent by now that the phase boundary compositions at a given temperature in some system of binary components are determined by the points of tangency of the lowest common tangents to the free energy–composition curves of the

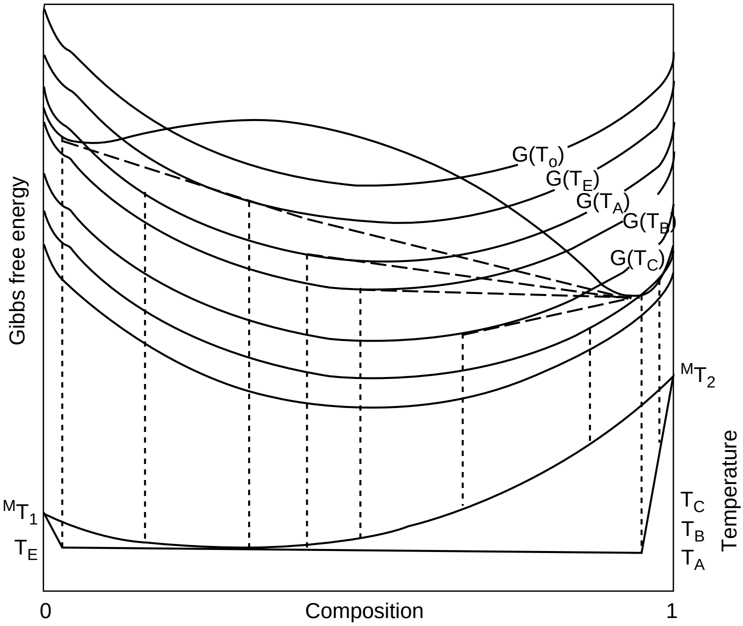


Figure 3.11. Free energy–composition curves at various temperatures for a liquid phase relative to a solid phase (held stationary for convenience in drawing only). The relation between the common tangent points and the solidus and liquidus phase boundaries is also shown. The eutectic temperature T_E and its related significance in the free energy curves are also apparent. $T_o < T_E < T_A < T_B < T_C$.

possible phases in the system. Also, whether a single phase or mixture of phases will be stable at some composition and temperature in the phase diagram is determined by the line in the free energy–composition diagram that represents the lowest free energy for these conditions. Obviously, a variation in temperature involves changes in the free energy–composition curves of the possible phases and hence in the identity of the envelope that represents the lowest possible free energy in the system, as is illustrated in Figure 3.11. Consequently, derivation of a phase diagram from free energy–composition curves for possible phases requires that the latter curves be known as a function of temperature. It is easier to grasp how phase diagrams develop from free energy–composition curves using a dynamic mode of illustrating this relationship. Such a dynamic mode is available in the form of computer programs (see Bibliography).

The development of phase diagrams based on the assumption that the solutions obey regular solution criteria is edifying and demonstrated in Figure 3.12. The parameter p_o shown both on the ordinate and abscissa (top) corresponds to the parameter b in equation 2.16.

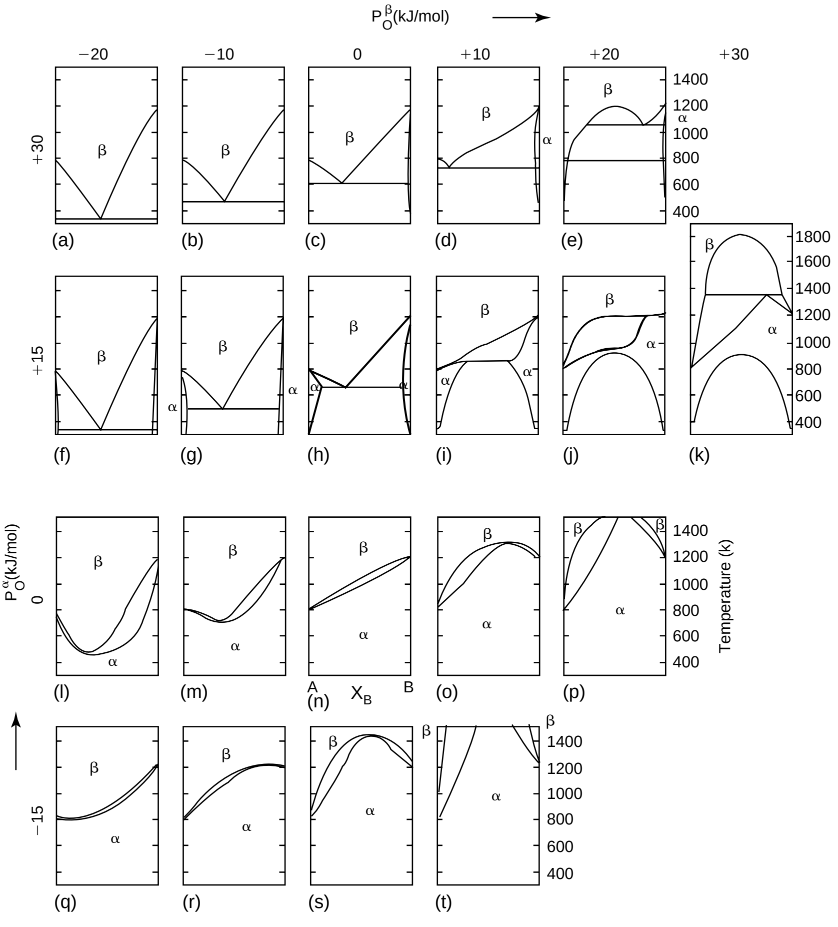


Figure 3.12. Sequence of phase diagrams corresponding to changes in regular solution parameters ($p_o \equiv b$) for liquid and solid solutions. From Pelton and Thompson, Prog. Solid State Chem. 10(part 3), 119(1976) with permission.

7. Prediction of phase diagrams

7.1. CALPHAD procedures

It is apparent from the above that given a knowledge of the composition and temperature dependence of the free energy of phases it should then be possible to derive the corresponding phase diagram. Although there exist some thermodynamic data for binary systems, these data are far from sufficient to allow such calculations to be made for most of the binary systems. Hence, a different point of view has arisen concerning a strategy of prediction of phase diagrams. This strategy is based on the existence of a large body of binary phase diagram data and is as follows.

From the knowledge that the free energy is both composition and temperature dependent, functions have been assumed to give this dependence analytic form. The composition dependence of several of these functions for the excess free energy is described in Table 3.1. The Gibbs free energy for one phase in a binary system is usually written in the form

$$G = \sum X_i {}^{o}G_i + RTX_i \ln X_i + {}^{xs}G \quad (i = 1, 2)$$

Since equilibrium is determined by the equality of partial molar free energies, it is more instructive to consider the equivalent expression for the latter quantity for a phase as follows

$$\bar{G}_i = {}^{o}G_i + RT \ln X_i + {}^{xs}G_i \tag{3.8}$$

where $${}^{xs}\bar{G}_i = {}^{xs}G + (1 - X_i)d^{xs}G/dX_i$$

as can be ascertained by substituting equation 3.8 into equation 2.16. It is now apparent that by equating the partial molar free energies for a given component, in the two phases being equilibrated, the parameters characteristic of the elements in the two phases appear as differences in such equations. This difference (${}^{o}G_i$(alpha) − ${}^{o}G_i$(beta)) is called the lattice stability of component i for the two phases alpha and beta and values for this quantity are listed in data banks.

The unknown terms in the resulting equations for the compositions of the solvus lines for the two phases in equilibrium are the lattice stabilities and the partial molar excess free energies. A power series is assumed for the temperature dependence of each of the lattice stabilities and for the parameter, such as C_p in the Redlich equation in Table 3.1. Thus, for example, if we consider two phases

Table 3.1. Proposed relations for excess free energy.

Margules	${}^{xs}G = X_1X_2 \sum_{p=0}^{n} A_p X_2^p$
Borelius	${}^{xs}G = X_1X_2 \sum_{p=0}^{n} B_p X_2^{n-p} X_2^p$
Redlich and Kister	${}^{xs}G = X_1X_2 \sum_{p=0}^{n} C_p (X_1 - X_2)^p$
Bale and Pelton	${}^{xs}G = X_1X_2 \sum_{p=0}^{n} D_p (P_p(X_1 - X_2))$

$P_p(X_1 - X_2)$ = Legendre polynomial of degree p of argument $(X_1 - X_2)$.

involved in a two phase equilibrium, and if we consider terms only up to the quadratic, then there will be 14 adjustable parameters to be fitted to the two solvus lines in the phase diagram representing this two phase equilibrium. (From the following equation for the dependence of the molar free energy of a pure component as a function of temperature

$$ {}^{o}G = m_1 - m_2T + m_3\,(T - \ln T\,) - m_4T^2/2 + \cdots $$

we deduce that there are 4 parameters for each component in the lattice stability contribution to the free energy and three for each phase in the excess free energy contribution. Thus, there will be a total of 4*2+3*2 = 14 adjustable parameters.) The equations necessary to evaluate these parameters are obtained both from known thermodynamic data and from the phase diagram using the equality of partial molar free energies at the solvus compositions. If there are not sufficient independent data then the temperature dependence is limited to fewer terms in the expansion.

In order to minimize the computational problem, it has been customary to assume that there is no solubility range for the intermediate phases and to treat them as compounds. It would be useful to have experimental values for the free energy of formation of such compounds. Various procedures have developed for estimating values of the energy of formation of such compounds and these are outlined in Appendix 1 to Chapter I. Using such estimated values it is then possible by a trial-and-error procedure to fit the parameters required to describe quantitatively any binary phase diagram, consistent with the assumption that the intermediate phases are compounds.

Using procedures similar to those described above, various centers around the world have developed data banks containing values for the adjustable parameters that are required to yield values for the various lattice stabilities of elements and the excess free energy of binary systems. These data now provide a basis for the prediction of multicomponent phase diagrams using the following procedure.

The free energy of a phase in the multicomponent system is assumed to be given by

$$ G = \sum_{i=1}^{n} X_i\,{}^{o}G_i + RT\sum_{i=1}^{n} X_i \ln X_i + \sum_{k=2}^{n}\ \sum_{j=1}^{\binom{n}{k}} {}^{xs}G_{j,k} $$

where $\binom{n}{k}$ is the binomial coefficient $= n!/[(n - k)!k!]$.

Knowing the terms for the binary interactions for each phase in a two phase equilibrium it is then possible to evaluate the equilibrium either by calculating the minimum free energy for the mixture of phases or by equating the appropriate partial molar free energies. At first, it is assumed that there are no ternary excess free energy terms in the first attempt to predict a ternary phase diagram and only if there are sufficient data to allow for a correction to the predicted phase diagram are the ternary terms fitted to these data.

The CALPHAD procedure in evaluating the free energy of a phase assumes that any deviation from a random mixture of the components of a phase may be described by the excess free energy of mixing. Although this is a viable procedure for the large majority of phases, it tends to lead to complicated expressions for the latter when a phase involves either partial long-range order or significant short-range order for the distribution of the components on the lattice sites. In the latter case, it is better to attempt to obtain a more accurate description of the entropy of configuration and a simpler expression for the excess free energy relative to this modified entropy. The latter is the procedure followed by the group using the CV method.[2]

Various software packages have been developed for the calculation of phase diagrams consistent with the CALPHAD philosophy. Among these are ChemSage,[5] MTDATA,[6] Thermo-Calc[7] and Pandat.[12] A review of thermochemical database systems that were available in 1990 is provided by Bale and Eriksson.[8] Since then, their review has been complemented by a site on the World Wide Web.[9]

One of the currently active research fields is that of the prediction of microstructure on solidification involving computer simulation. The ability to calculate multicomponent phase diagrams has been incorporated into the attack on this problem based on the fact that local thermodynamic equilibrium is closely approximated at the solid–liquid interface.[10]

7.2. Thermodynamic integration and quasiharmonic lattice dynamics methods of evaluating free energy

The thermodynamic integration method to evaluate the free energy of some state requires a value for the free energy in some other state, a known dependence of the energy on a variable intensive parameter, and the relations

$$F = -kT \ln z$$

$$z = \int \exp(-E(X)/kT)dX$$

where F is the Helmholtz free energy, E, the energy and X is an intensive parameter. The integration is carried out from the value of X, where F or E is known, to the value of X corresponding to the state for which the value of F is desired. The paths along which the integration is carried involve constant values of the other intensive parameters. There are various methods for evaluating the E(X): molecular dynamics or Monte Carlo methods, which require the forces acting on the atoms as input, first-principle methods to evaluate ground state energies and use of quasiharmonic lattice dynamics to evaluate E(T). Reference is made to review articles[13] for more details concerning the multitude of methods of using both the thermodynamic integration method and quasiharmonic lattice dynamics to determine free energies of various states.

8. Coherent equilibrium

In the foregoing it has been assumed implicitly that the coexisting phases are unstressed because the condition governing the equilibrium between coexisting phases was taken to be the equality of the chemical potential of a species in the coexisting phases, for each species, i.e. equation 3.1. However, when the phases in equilibrium have a coherent interface (i.e. a continuity of lattice rows and planes across the coherent interface) then both phases are stressed. Under this circumstance, as already noted in the previous chapter, chemical equilibrium is no longer governed by this equality of the chemical potential of a given species in each coexisting phase. (It is not possible, in general, to accomplish a transfer of atoms i between the lattices of coherent coexisting phases keeping the number of each other type j atom, as well as the intensive parameters stress or strain, constant. The latter constraints, and also the condition of coherency, imply a conservation of the total number of lattice sites in each coexisting phase. Further, the lattice parameter of a phase usually changes with a change in composition.) Rather, it is governed, in part, by the equality of the diffusion potential. As implied, this statement is not a sufficient statement of the chemical equilibrium. A more complete discussion of the stability of homogeneous, coherent systems has been given by Huh and Johnson[3] for the case of invariant morphology. For the constraints of either constant strain or constant edge traction and an invariant morphology the sets of intensive parameters of the coexisting phases at equilibrium for a binary system are given by the common tangent to the free energy, where the latter is either a function of strain or edge traction. However, only for the constant strain case, where the constraint imposes the same strain on both the coexisting phases, does the common tangent to the free energy–composition curve for that strain determine the compositions of the coexisting phases at stable equilibrium. For the case of an externally imposed constant edge traction, where the two coexisting phases at equilibrium can have different values of homogeneous stress, it is the common tangent to the free energy surface as a function of composition and stress that determines the sets of composition and stress values that hold for the coexisting phases at stable equilibrium. In this case, the tie lines lie in the stress–composition plane for a given temperature, just as they do for a ternary system when the phases in equilibrium are essentially fluid type phases.

For a given set of externally controlled thermodynamic variables it has been suggested that a coherent two phase system can have more than one equilibrium solution (i.e. a Williams point[4]). However, according to Huh and Johnson[3] only one stable equilibrium state is to be expected for a coherent two phase system possessing a consulate critical point, i.e. a homogeneous solid solution for which the relations

$$[\partial^2 M_{12}/\partial C^2]_{\varepsilon,t} = 0$$
$$[\partial^3 M_{12}/\partial C^3]_{\varepsilon,t} > 0$$

both hold simultaneously, where C is the composition, ε is the normalized strain and t is the normalized temperature.

It is not given that morphological stability exists for coherent systems, as was assumed above. Indeed, the competition between strain energy and interface energy usually induces a change in the morphology, whenever the latter is determined initially by other factors.[11]

9. Summary for Section A

Summarizing, we have determined the conditions governing the effect of solute addition upon the solidus and liquidus temperatures and found that a more positive excess free energy for the solid phase relative to the liquid phase results in a decrease in the "average melting temperature". Further, the relations governing the distribution coefficient for solidus to liquidus compositions have been derived. One origin of the eutectic phase diagram has been shown to be the existence of a positive enthalpy of mixing in the solid state. The origin of intermediate phases in phase diagrams has been explored, as well as the effect of metastability on solvus composition. The integral relation between free energy–composition curves for possible phases as a function of temperature and the phase diagram has been demonstrated. Also, a strategy for the prediction of phase diagrams has been outlined. Finally, some aspects of coherent phase equilibrium have been considered.

B. HETEROGENEOUS CHEMICAL EQUILIBRIA AND PHASE DIAGRAMS

1. Thermodynamics of heterogeneous chemical reactions

It is worthwhile summarizing here the thermodynamic relations governing heterogeneous chemical reactions because a knowledge of this discipline is often of use in materials science, ceramics and metallurgy. We note first that the vapor pressures of substances should be exponentially related to the reciprocal temperature, as can be derived using the Clapeyron relation as follows. Also we should note that in the transformation from a solid or a liquid (a condensed phase) to a vapor phase, the change in molar volume is to a good approximation equal to that for the vapor phase itself, i.e. $V_v - V_s = V_v$. Also, for an ideal gas, the molar volume of the vapor is given by

$$V_v = RT/p$$

where p represents the partial pressure of the vapor. Substitution into the Clapeyron equation and rearranging terms yields

$$d \ln p/d(1/T) = -\Delta H/R$$

Integration then yields that

$$p = p_o e^{-\Delta H/RT}$$

Thus, as the temperature increases the vapor pressure of a substance in equilibrium with it increases greatly.

Let us now consider reactions between two different substances, e.g. the reaction between aluminum and oxygen which forms the oxide Al_2O_3. The reaction can be written as follows

$$(4/3)Al + O_2 = (2/3)Al_2O_3 \tag{3.9}$$

The equilibrium constant for this reaction is $K = 1/p(O_2)$ and this corresponds to a certain value of the standard free energy for this reaction and temperature, where the standard free energy of reaction refers to all components being in their standard reference states at the temperature and at *1 atm pressure*. Values for the standard free energy of reaction ΔG^o are given in various books (see Bibliography). A convenient method of providing such information is via Ellingham diagrams, which presents the standard free energy of reactions with a given component, such as O_2, as a function of temperature (see Figure 3.13).

The Ellingham diagram has certain other features which are useful and which we shall demonstrate. Consider the reaction given above. Suppose that we are interested in the values corresponding to 300°C. According to the Ellingham diagram, the standard free energy for the above reaction at this temperature is -996 kJoules/mol ($-237{,}000$ cal/mol). If now we draw a straight line through the coordinate (0, 0°K) in the upper left-hand region of the diagram and the point defined by $\Delta G^o = -996$ kJoules/mol and $T = 300°C$, it will be found to intercept the outermost line to the right and down at the value of $p(O_2) = 10^{-90}$ atm. The significance of this result is that the latter is the value of the partial pressure of O_2 in equilibrium with Al and Al_2O_3. A partial pressure of O_2 less than this value will reduce the oxide and a partial pressure of O_2 above this value will oxidize the Al. For example, since even ultra high vacuum chambers contain a partial pressure of O_2

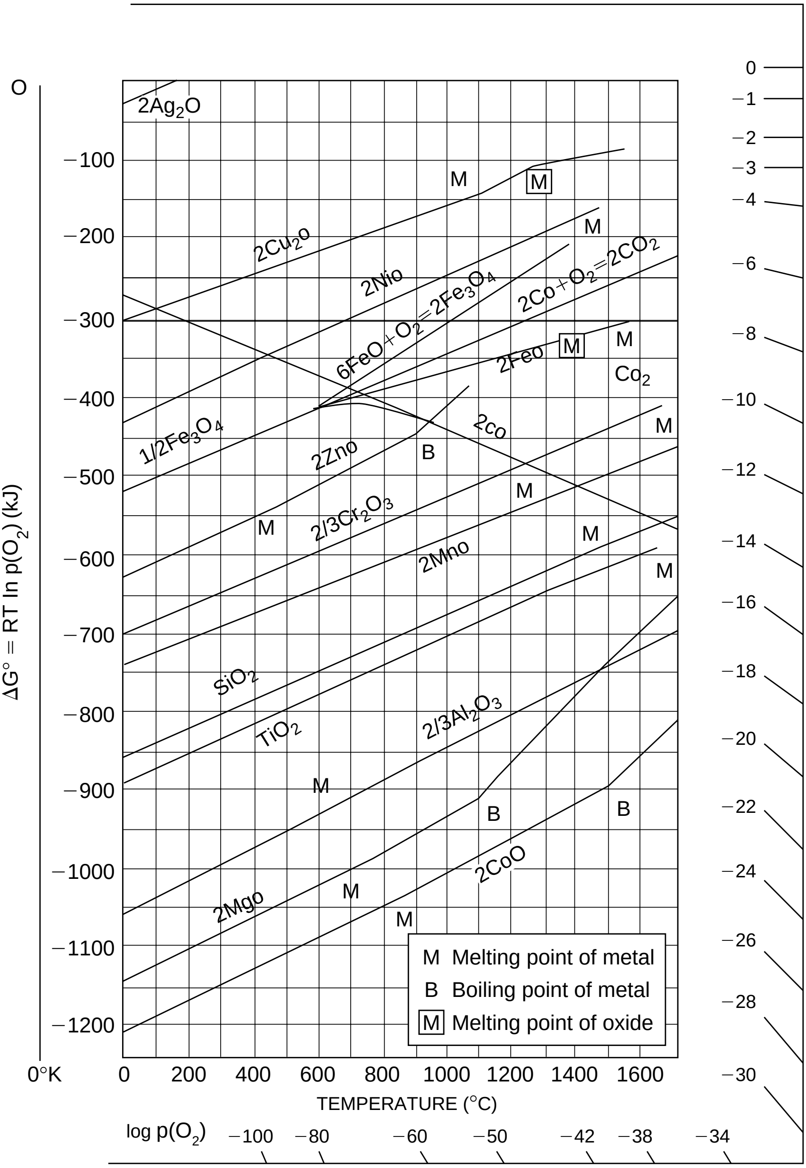


Figure 3.13. Ellingham diagram for several oxides. From Gaskell in Physical Metallurgy, eds. R.W. Cahn and P. Haasen, North Holland Physics Publishing, New York, 1983 with permission.

more than this value it is obvious that the surface of aluminum conducting stripes formed in vacuo for integrated circuits is covered with an oxide.

There may be some question in the reader's mind as to the origin of the equilibrium partial pressures of O_2 and partial pressure ratios in the Ellingham diagram. The basis for this relation is simple. It must be recalled that the standard free energy of a reaction corresponds to vapor phases having a partial pressure of 1 atm. The standard free energy change for the reaction given by equation 3.9 is $-237{,}000$ cal/mol as noted above. At equilibrium, the change in free energy for this reaction would of course equal zero. But, the standard state of the O_2 gas in this reaction, 1 atm pressure does not correspond to the partial pressure in equilibrium with this reaction, i.e. the partial pressure that would make the change in free energy equal to zero. Thus, to equation (3.9) must be added the reaction

$$O_2(p_{eq}) = O_2(p^o = 1) \tag{3.10}$$

for which the change in free energy, considering that O_2 acts as an ideal gas, is

$$\Delta G' = RT\ln(p^o/p_{eq}) \tag{3.11}$$

Thus, for equilibrium, $\Delta G^o + \Delta G' = 0$, or in this case

$$RT\ln(p^o/p_{eq}) - 237{,}000 = 0$$

Rearranging the terms and using the fact that $p^o = 1$ then

$$p_{eq} = e^{(-237{,}000/RT)}$$

Substituting 573°K for T and solving yields $p_{eq} = 1.5*10^{-90}$ atm, the result we found previously using the Ellingham diagram.

In general, if a reaction involves m_i moles for the ith reactant for a total of n gases of which j are reactants (to be denoted by the superscript r) and $n-j$ are products (to be denoted by the superscript pr) then we would have

$$\prod_{i=j+1}^{n-j} (p^{pt}_{eq,i})^{m_i} \Big/ \prod_{i=1}^{j} (p^{r}_{eq,i})^{m_i} = e^{(-\Delta G/RT)} = K \tag{3.12}$$

where K is the equilibrium constant.

For the case where one or more of the elemental constituents taking place in a reaction is a solute in a solid solution then the reaction must be corrected to account for the free energy of solution of this solute in its solvent. The correction is analogous to that already made for the change in state of a gas from the standard state of 1 atm to the equilibrium pressure. In the case that the solute is a reactant rather than a product, we must add to the reaction equation the reaction

$$[A] = A^o$$

for which the free energy of the reaction is $-RT \ln a_A$, where a_A is the activity of component A in its solvent. (The activity of A in its pure standard state is equal to unity.) Since, at equilibrium, the total free energy of the reaction is zero, we would then obtain that

$$RT \ln a_{A,eq} = \Delta G^o$$

If more than one solute is involved or there are also constituents in gaseous states then the corresponding equations would have to be added to obtain the relation between their partial pressures, the activities of the dissolved constituents and the standard free energy of reaction for the equilibrium situation.

The Ellingham diagram is also useful in indicating which metal will reduce which oxide. Any metal corresponding to a reaction having more negative free energy of formation values than those for the formation of some other oxide will reduce the latter oxide. For example, Ti will reduce Cr_2O_3.

References

1. M. Hillert, T. Wada and H. Wada, J. Iron Steel Inst. 205, 539(1967); H. Harvig, G. Kirchner and M. Hillert, Met. Trans. 3, 329(1972).
2. See Bibliography, Chapter II.
3. J.-Y. Huh and W.C. Johnson. Acta Metall. Mater. 43, 1631(1995).
4. R.O. Williams Calphad 8, 1(1984); J.W. Cahn and F.C. Larche, Acta Metall. 32, 1915 (1984).
5. G. Eriksson and K. Hack, Met. Trans. 21B, 1013(1990).
6. R.H. Davies et al., Applications of Thermodynamics in the Synthesis and Processing of Materials. eds. P. Nash and B. Sundman, TMS, Warrendale, PA, 1995, p. 371.
7. B. Sundman, B. Jansson and J.-O. Andersson Calphad 9, 153(1985).
8. C.W. Bale and G. Eriksson, Canad. Metall. Quart. 29, 105(1990).
9. C.W. Bale, “Web sites in Inorganic Chemical Thermodynamics”, http://www.crct.polymtl.ca/fact/websites.html.
10. T. Kraft and Y.A. Chang J. Met. 49 (12), 20(1997).
11. J.K. Lee and J.H. Choy, MRS Symp. Proc. 398, 439(1996).
12. S.-L. Chen, F. Zhang, S. Daniel, F.-Y. Xie, X.-Y. Yan, Y.A. Chang, R. Schmid-Fetzer and W.A. Oates, J. Met. 55 (12), pp. 48–51(2003).
13. J.M. Rickman and R. LeSar, Ann. Rev. Mater. Sci. 32, 195(2002); D. Alfe et al., Int. J. Quantum Chem. 77, 871(2000).

Bibliography

1. A.D. Pelton in Physical Metallurgy, eds. R.W. Cahn and P. Haasen, 3rd edition, North Holland Physics Publishing, New York, 1983.
2. A.M. Alper, ed., Phase Diagrams – Materials Science and Technology, Vol. 1–5, Academic Press, New York 1970–1978.
3. P. Gordon, Principles of Phase Diagrams in Materials Systems, McGraw-Hill, New York, (1968).
4. Free Energy Curves, Code 435, Institute of Metals, Old Post Road, Brookfield, Vt 05036.
5. O. Kubaschewski and C.B. Alcock, Metallurgical Thermochemistry, 5th edition, Pergamon Press, New York, 1979.

Coherent Equilibrium

1. See references cited in Reference 3 above.
2. F.C. Larche, Ann. Rev. Mater. Sci. 20, 83(1990).

Problems

1. Describe the behavior of the liquidus and solidus phase boundaries for the following three conditions: ideal liquid and solid solutions; excess free energy of solute component for the solid exceeds that for the liquid and vice versa.
2. If the excess free energy required to insert one solute atom in the solvent is 0.216 ev/atom what is the approximate limit to the solid solubility at 500ºK?
3. Draw a series of phase diagrams corresponding to more positive values of the difference ${}^{xs,l}G_{solute}{}^{xs,s}G_{solute}$. In this case, do you expect the onset of a miscibility gap in the solid state or the formation of an intermetallic compound or ordered phase?
4. Demonstrate that the strain energy due to the insertion of carbon atoms into the interstitial sites at 0, 0, 1/2 in both the bcc and fcc crystal structures is greater for the bcc structure.
5. Using a free energy–composition diagram demonstrate how it is possible in some cases for the stable composition range of an intermediate phase **not** to include the stoichiometric composition.
6. How does the free energy–composition curve for a superlattice differ from that for an intermetallic compound?
7. If there is a large electronegativity difference between the components of a binary system is it likely or unlikely that there will be intermediate phases in this system?
8. Why is the solvus composition of a terminal phase in equilibrium with a metastable phase always richer in solute than that corresponding to the stable equilibrium? (Hint. Use free energy–composition diagrams to help provide the answer to the question.)
9. Draw the free energy–composition curves for the phases that appear in Figure 3.6 corresponding to 600°C and 1000°C.

10. Figure 3.5 reveals that an intermediate phase is stable in the LiF–CsF system. Speculate as to whether the Li and Cs ions occupy points on the same sub-lattice in a random fashion or are arranged on different sub-lattices in an ordered array? Provide a justification for your answer.
11. If atomic jumps are effectively frozen-in at absolute temperatures below half the absolute melting point, and if you assume nature chooses the polymorph that is most stable how would you predict the range of stability of amorphous solids in eutectic type phase diagram systems?
12. In an atmosphere consisting of a partial pressure of O_2 equal to 10 atm, above what temperature will SiO_2 decompose to Si and O_2?

CHAPTER IV

Thermodynamics of Interfaces

Introduction

For the materials scientist one of the more important set of properties is that associated with surfaces and interfaces. This chapter treats their thermodynamic aspects. The excess free energy associated with interfaces provides a driving force for a variety of kinetic processes. These relations are developed quantitatively. Two contrasting approaches to the concept of the surface or interface, those of Gibbs' and Guggenheim, are described, and applied to the phenomenon of adsorption in this chapter. Also, the various types of interfaces are investigated briefly and expressions governing their local equilibria are developed. Finally, applications are made of these concepts to some important problems in materials science.

1. Concept of surface quantities

Let us perform a thought experiment as follows. Separate a solid crystal into two halves along some plane (say plane A), parallel to a crystal plane. The separation plane A is chosen so as not to contain lattice points nor equilibrium positions of atoms. In the approximation where the cohesive energy is given by a sum over bond energies, the energy difference, energy of original solid minus energy of resulting halves, is then equal to the sum of the energies of the bonds that intersect the separation plane A in the original solid. The specific surface energy, e_s, per unit area of surface, in this approximation is then one-half of the negative of the bond energy sum over the bonds intercepting unit area of the separation plane A. This follows from Gibbs' definition of a specific surface quantity as being the excess per unit area of surface of the object having the surface over that given by the product of the specific quantity per unit volume by the volume of the object.

Thus, the specific surface energy is given by

$$e_s = (2E_{1/2} - E^o)/(2A) \tag{4.1}$$

where A is the area of plane A, $E_{1/2}$ is the energy of one of the product halves and E^o is the energy of the original solid.

2. An approximate model for evaluating the surface energy

The specific surface energy has been calculated by Friedel et al.,[1] and many others subsequently, using a nearest-neighbor bond approximation for the cohesive energy and the assumption that the specific surface energy corresponds to one-half the sum of the energies of nearest-neighbor bonds that have been broken in the formation of the unit area of surface in the process of separating a crystal into two halves along some crystal plane with indices h, k, l, with the result that

$$e_s = -2(2h + l)(h^2 + k^2 + l^2)^{-1/2} e_b / a^2 \text{ (fcc)} \tag{4.2}$$

where $h \geqslant k \geqslant l$, e_b = energy/bond which is equal to cohesive energy/(Nz/2), z is the coordination number and a is the lattice parameter.

Values of specific surface energy deduced using the above relation are shown below and are compared to measured values of the average specific surface free energy.

			e_s (erg/cm^2)			
			Calculated		**Experimental***	
Metal	**e_b (10^{-12} erg/bond)**	**a^2 (10^{-16} cm^2)**	**(111)**	**(110)**	**(100)**	
Ag	−0.792	16.6	1650	1350	1905	1690
Au	−0.98	16.58	2100	1715	2425	2175
Cu	−0.94	13.02	2500	2080	2890	2260

* Data from reference 24 corrected to the temperature of absolute zero.

It is apparent that the bond model yields a good approximation for the value of the specific surface energy and thus the concept that the excess energy associated with the surface is related to the defect of bonds across the surface has rough predictive value, at least for metals and covalently bonded materials. A calculation of the work required to separate an ionic solid into two halves has been performed[2] using a sinusoidal function to approximate the force resisting the separation with the result that $e_s \approx Ya/4\pi^2$, where Y is Young's modulus and a is the interatomic spacing, with a similar match to experimental values.

2.1. Wulff plot of surface energy

Herring[3] has shown for any monatomic crystal with a center of symmetry that: (a) if the total energy of the crystal can be expressed as a sum over bond energies and (b) if the surface atoms are at perfect periodic lattice points then the plot of surface energy versus orientation must consist of portions of spheres that pass through the origin when extended. Figure 4.1 illustrates such a polar plot of surface energy versus orientation for the case of surface planes having a common [001] rotation axis.

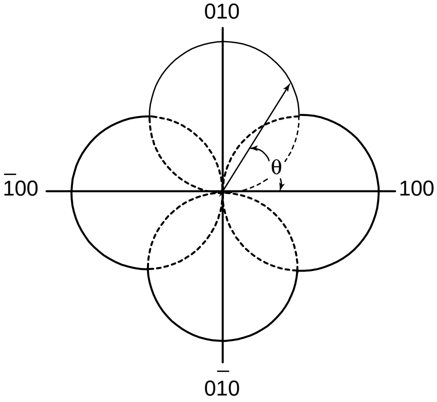


Figure 4.1. Wulff plot showing surface energy σ, as a function of polar orientation θ.

Mackenzie[4] has calculated specific surface energies using this concept. The result, as schematically illustrated in Figure 4.1, is that at certain low index planes, the specific surface energy has a minimum value relative to orientations just displaced from the low index orientations. Thus, the specific surface energy–orientation plot consists of such singular cusps and regions where the specific surface energy, and its first and second derivatives with respect to orientation are continuous functions. These polar plots of specific surface energy versus orientation are called Wulff plots.

The difference in surface energy between various crystal planes can exert important effects. For example, consider a thin polycrystalline film of silicon deposited on an inert substrate. If the average grain diameter is larger than the film thickness, then the difference in free energy between the various silicon grains in the thin film is primarily the difference in the excess free surface energy. For small diameter grains and a thin film this excess free energy can change the melting points of the various grains significantly. This effect is illustrated qualitatively in Figure 4.2. As shown, the specific free energy of the small grains in the thin film will vary according to the orientation of the surface plane. (The molten silicon–thin film interface energy shows a much smaller relative change in value with variation in orientation of the crystal plane parallel to the interface.) A consequence of this variation is that the melting point of differently oriented grains will differ. The orientation exhibiting the lowest surface energy will have the highest melting point, as shown. In the case of silicon, this plane is the (100) plane. Hence, it should be possible to retain selectively the (100) oriented grain in a molten thin film on slowly raising the temperature. Fortunately, the shapes of such grains are anisotropic, so that they can be aligned along ridges. By slowly moving a temperature gradient transversely along the thin molten film and parallel to the ridges, it should then be possible to grow out a single crystal thin film having a (100) orientation. Indeed, this objective has sometimes been achieved in SOI (silicon on insulator) technology via the use of a traveling molten zone and is a phenomenon belonging to the class called “Artificial Epitaxy” or “Graphoepitaxy”.

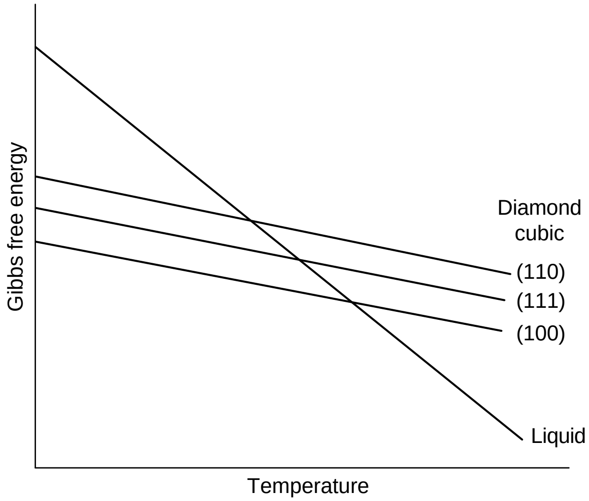


Figure 4.2. Showing the melting points of crystals having different surface facets in the diamond cubic structure.

Also, as suggested by the above analysis, the difference in surface energies can produce a driving force for the solid-state growth of grains in thin films having the lowest surface energy orientation at the expense of neighboring grains having surface orientations of higher specific surface energy. An example where this effect operates is in the heat treatment used to produce (100) oriented Fe–Si alloy transformer sheet. It may also be useful in attempts to produce desired epitaxial single crystal films.

3. Surface reconstruction

An implication of the description of crystal surfaces given in the previous sections is that the atoms at the surface are at lattice sites of the bulk crystal. In general, this concept is not valid. Although, in many cases the atoms at the surface are at lattice sites of the bulk crystal structure, the lattice parameters of this crystal structure are distorted. Normally, the lattice planes parallel to the surface are closer to each other than in the bulk crystal, although oscillations in this spacing have also been observed on passing from the surface layer into the bulk. However, in many other cases, the surface atoms are in a different crystallographic arrangement as compared to that described by the bulk crystal structure. In the latter case, the surface is said to have undergone surface reconstruction. In both cases, the driving force for these deviations from the sites defined by the bulk crystal structure is the tendency to decrease the free energy of the material. With a reliable interatomic potential for a pure material it should be possible to calculate the energies of various surface arrangements and thereby determine the most stable one of these arrangements. (Since the first edition of this book this application of interatomic potentials has become an active area in computational materials science. At this writing semiconductors and simple metals have sufficiently accurate interatomic potentials to reproduce experimental surface reconstructions.) The energy change associated with these relaxations is small compared to the increment in energy that comprises the surface energy itself. Thus, it is not surprising that the rough bond model used in the previous sections yields a reasonable estimate for the magnitude of the surface energy even if it neglects surface relaxations of interatomic distances or reconstruction of unit cells into others. The interested reader is referred to the monograph by Somorjai (see Bibliography) for additional material on this subject.

4. Some particle size effects

4.1. Effect of particle size on difference in pressure between a small spherical isotropic solid and an external fluid in local equilibrium

Let the system consist of one component and let dn molecules be transferred from a fluid phase to a spherical, isotropic solid phase in such a way that the change in Helmholtz free energy for the total system is zero, that the temperature is constant and the specific Helmholtz free energy in each phase is essentially constant (i.e. the latter is not strictly true as will be discussed later.) Thus, there is an increase in volume of the solid phase, which also involves an increase in the solid–fluid interface area. There are two main contributions to the change in free energy of the system: a volume contribution $-(p_b - p_f)\delta v$ and a surface contribution $\sigma\delta A$. (For a single component system, the surface tension σ equals the specific surface excess Helmholtz free energy f^s.)* Summing these contributions yields at equilibrium that

$$0 = -(p_b - p_f)\delta v + \sigma\delta A$$

or

$$(p_b - p_f) = \sigma\delta A/\delta v = 2\sigma/r \quad \text{(for the spherical particle)} \tag{4.3}$$

---

* The surface tension is used as a name for the surface free energy σ in the literature, although Gibbs used it for the work to stretch a surface. We shall use the former sense of σ and discuss the work to stretch a surface in the section on surface stress.

(For r small, $2\sigma/r >> p_f$, i.e. for $\sigma = 1000\,\text{erg/cm}^2$ and $r = 10^{-6}\,\text{cm}$, then $p_b = 1000\,\text{atm}$.) In the above we have neglected second order terms such as

$$V_b \delta f_b \approx (4\pi r^3/3)\kappa p_b \delta p_b \approx (8\pi r/3)\sigma p_b \kappa \delta r$$

Substituting reasonable values into the ratio of $V_b\delta f_b/\sigma\delta A$ we obtain a value of about 0.001, which substantiates the neglect of the change in free energy of the solid due to the change in pressure.

4.2. Effect of particle size on difference in pressure between a crystal particle that develops surface facets and external fluid in local equilibrium

We shall not derive the result considered in this section, but make use of it. For a derivation the reader is referred to Defay et al.[5] (see Bibliography). If the corners of the facets on a single crystal particle are joined by lines to a point O in the crystal and the normal distances to these facets from this point are denoted by $h^{(i)}$, then it can be shown that at equilibrium the point O exists such that

$$\frac{2\sigma^{(1)}}{h^{(1)}} = \frac{2\sigma^{(2)}}{h^{(2)}} = \cdots \frac{2\sigma^{(N)}}{h^{(N)}} = p_b - p_f \tag{4.4}$$

These relations are known as Wulff's relations.

Equation 4.4 states that the pressure inside a particle is a function of its size and not the radius of curvature as is implied by equation 4.3. Relations 4.3 and 4.4, which were derived for single component systems, apply also to multicomponent systems.

4.3. Dependence of equilibrium vapor pressure on particle size

Consider particles all having the same size and crystal facets or isotropic spherical particles having the same radius in equilibrium with a vapor phase. The particles and vapor may have many components. For each component at equilibrium, the chemical potentials in the coexisting phases are equal, i.e.

$$\mu_{ib}(T, p_b) = \mu_{iv}(T, p_v) \tag{4.5}$$

But

$$p_b = p_v + 2\sigma/r \quad (= p_v + 2\sigma^{(\gamma)}/h^{(\gamma)})$$

Hence, expanding the left hand term in equation 4.5 yields

$$\mu_{ib}(T, p_b) = \mu_{ib}(T, p_v) + 2v_{ib}\sigma/r \tag{4.6}$$

where v_{ib} is the derivative of the chemical potential with respect to pressure. Thus, the chemical potential in a small particle is also a function of the size of the particle.

Now, assuming the vapor acts like an ideal gas

$$\mu_{iv}(T, p_v) = \mu_{iv}^{*} + RT \ln p_{iv} \tag{4.7}$$

where the superscript * denotes the standard state dependent on temperature only. Consider that there is an equilibrium displacement such that the temperature and composition of the solid phase are maintained constant. For this displacement

$$\delta\mu_{ib}(T, p_b) = v_{ib}\delta p_b \tag{4.8}$$

But, by equation 4.5

$$\delta\mu_{ib}(T, p_b) = \delta\mu_{iv}(T, p_v)$$

Hence

$$\delta p_b = \delta\mu_{iv}(T, p_v)/v_{ib} = (RT/v_{ib})\delta(\ln p_{iv}) \tag{4.9}$$

But

$$\delta p_b = \delta p_v + \delta(2\sigma/r)$$

Hence

$$\delta(2\sigma/r) = (RT/v_{ib})\delta(\ln p_{iv}) - \delta p_v \tag{4.10}$$

Integrating from $1/r = 0$ to $1/r$ and assuming that v_{ib} is constant during this integration yields

$$2\sigma/r = (RT/v_{ib})\ln(p_{iv}/p_{iv}^{o}) - (p_v - p_v^{o}) \tag{4.11}$$

For small particles, $(p_v - p_v^{o}) << 2\sigma/r$ so that a good approximation is

$$\ln(p_{iv}/p_{iv}^{o}) = (2\sigma/r)(v_{ib}/RT) \tag{4.12}$$

This relation also holds for a single component system.*

---

* The prior derivation assumes the absence of shear stresses in the particles.

Thus, the vapor pressure in equilibrium with particles of radius of curvature r increases as the radius r decreases. This provides the driving force for vapor transport of atoms from small particles to large particles, or from above surfaces having small radius of curvature to surfaces having large radius of curvature. Further, the dependence of chemical potential on curvature provides the driving force for bulk or surface diffusion from regions of small radius of curvature (or negative radius of curvature) to regions of large radius of curvature, as occurs during the process of sintering.*

There will always be a driving force for the establishment of equilibrium between a particle and its vapor. If there are particles of different size, each attempting to establish local equilibrium with its vapor, it is obvious that the vapor pressure cannot be uniform everywhere. Rather, the local equilibrium vapor pressures will set up gradients of vapor pressure that bring about transport of atoms from the high vapor pressure regions to the low vapor pressure regions. This transport will distort the vapor pressures corresponding to local equilibrium and the ever present tendency to maintain local equilibrium results in the evaporation and consequent disappearance of small particles and the growth of large particles. Such growth of large particles at the expense of small particles occurs also when the particles (precipitates) are embedded in a host matrix and attempt to approach local equilibrium with their environment. This process is called Ostwald Ripening and is considered below and in Chapter XI.

4.4. Dependence of solvus composition on precipitate particle size

Consider spherical particles of a precipitate of radius r containing only pure solute B, which are in equilibrium with a flat solvent solid solution by means of being heated in a box, so that the same vapor pressure of B is in equilibrium with both the particles and the solid solution. Thus

$$\mu_B|_{\text{solution}} = \mu_B|_{\text{vapor}} = \mu_B|_{\text{particles}} \tag{4.13}$$

Now, the change in chemical potential of B in the solid solution with change in concentration of B is given by

$$(\mu_B - \mu_{Bo})|_{\text{solution}} = RT\ \ln(C/C_o) \approx RT(C - C_o)/C_o$$

---

* Should the particles have facets, then the transport obviously occurs from facets having high values of σ/h to facets having smaller values of this quantity.

(For small deviations of C from C_o, where the latter composition corresponds to that in equilibrium with flat particles of B.) But

$$\mu_B\big|_{\text{particles}} = \mu_{Bo}\big|_{\text{particles}} + 2\,\sigma v_B/r$$

Thus

$$RT(C - C_o)/C_o = 2\sigma v_B/r \tag{4.14}$$

where v_B is the volume per B atom in solid B, i.e. $(\partial\mu_B/\partial p) = v_B$

This result states that the solubility limit of a solid solution in equilibrium with particles of pure solute depends on the radius of the latter particles. A graphical illustration of this result is shown in Figure 4.3. In the derivation of phase equilibria we normally neglect the effect of surface energy. As shown above, this neglect is not justified if any one of the phases is in the form of small particles.

The result just derived can be shown to apply to the case of any precipitate phase that is in equilibrium with a host phase, even if the precipitate is itself a solution or compound phase.[6] That is, the phase boundary compositions of a phase in local equilibrium with another solid phase are dependent on the size of the latter particles. The latter result can be demonstrated by yet another approach, as follows. It can be shown[6] that the difference in pressure across an interface is given by

$$P' - P'' = K\sigma$$

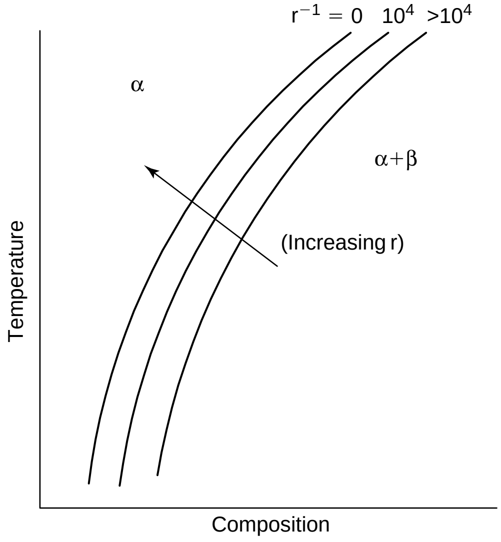


Figure 4.3. Showing the schematic effect of curvature on the solvus composition is equilibrium with particles that vary in size.

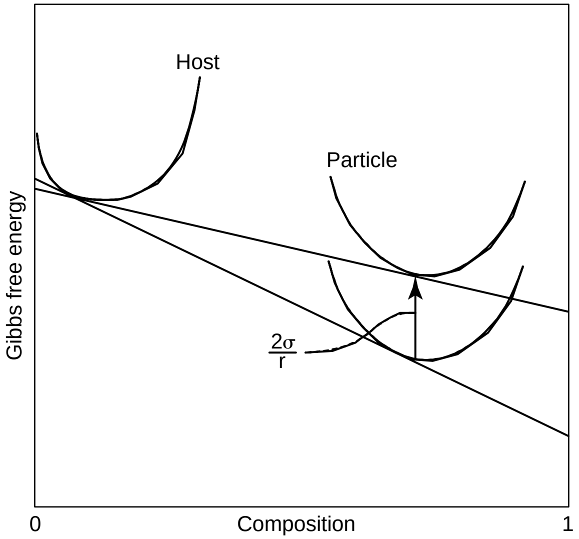


Figure 4.4. Schematic illustration of the effect of particle size on the free energy–composition curve for particles of an intermediate phase as precipitates in the host phase and the effect on the solvus composition (r is the radius of the particle).

where K is the mean curvature (i.e. 2/r, for a spherical particle). If the interface is between a host phase and a small included particle of an intermediate compound phase, it is usually assumed that the pressure in the host phase is that in equilibrium with its environment. Thus, the increase in pressure relative to the host phase is sensed by the particle itself. The effect of this increase in pressure can be demonstrated graphically. Figure 4.4 shows the free energy–composition curves corresponding to the pressure of the environment for both host and particle phases and also that for a spherical particle phase subjected to an increase in pressure relative to that of the environment corresponding to $2\sigma/r$. If common tangents are drawn, as shown, then it is apparent that the solvus composition of the host phase is dependent on the radius of the particle phase and that the partial molar energy (chemical potential) of the solute at the solvus composition increases with decrease in the particle radius.

All of the above applies to precipitates that involve incoherent equilibrium. The exchanges of atoms are not accomplished at a constant number of lattice sites and the phases involved are essentially fluid phases. Thus, it is correct to equate the chemical potentials of a constituent in the various phases in the equilibria considered in this section.

5. Adsorption

5.1. Gibbs' adsorption

Let us now consider the surface–volume equilibrium for a multicomponent system. We follow the method of Gibbs in defining surface excess quantities as being equal to the excess of a quantity over that given by the sum of the volume densities of the quantity multiplied by the volume of the corresponding phase, i.e.

$$Q^s = Q - \Sigma_i q_i V_i \qquad (4.15)$$

Let there be a large block of the single phase in contact with a thermal reservoir at temperature T and n chemical reservoirs each capable of maintaining the chemical potentials within the single phase block at their specified values μ_1, μ_2, … μ_n. Let us now create new surface area in this block (by separating the block in such a way as not to affect the specific quantities q_i applicable to the block. The work dW is done to create the new surface area dA (reversibly), with the reservoirs maintaining the temperature T and the chemical potentials μ_i. We define $\sigma \equiv dW/dA$. The production of area dA will be accompanied by a transfer of dN_i moles of component i from the ith reservoir, maintaining the chemical potential at μ_i. Since μ_i in the bulk was unaffected by the creation of dA, the dN_i corresponds to a surface excess defined by $\Gamma_i = dN_i/dA$. The change in Helmholtz free energy of the block is

$$dF = dW + \Sigma \mu_i \, dN_i = \sigma dA + \Sigma \mu_i \Gamma_i \, dA \qquad (4.16)$$

This change must be a surface excess by the constraints, i.e. $dF = dF^s$. Per unit area then f^s $(= dF^s/dA)$ is then

$$f^s = \sigma + \Sigma \mu_i \Gamma_i \qquad (4.17)$$

We now allow a change dT in the temperature of the reservoir and $d\mu_i$ in the chemical potential of the reservoirs at constant surface area. The change in Helmholtz free energy of the system, block and surface, is

$$dF = -SdT + \Sigma \mu_i \, dN_i \qquad (4.18)$$

But

$$S = S_b + S^s \quad \text{and} \quad dN_i = dN_{ib} + d^sN_i$$

Also

$$dF - dF_b = dF^s = -S^s dT + \Sigma\mu_i\, d^sN_i$$

Dividing by the area of the surface of the block then

$$df^s = dF^s/A = -s^s\, dT + \Sigma\mu_i\, d\Gamma_i \tag{4.19}$$

But

$$df^s = d\sigma + \Sigma\mu_i d\Gamma_i + \Sigma\Gamma_i\, d\mu_i$$

by previous equation. Hence

$$d\sigma = -s^s\, dT - \Sigma\Gamma_i\, d\mu_i \tag{4.20}$$

This is the Gibbs' adsorption relation between the variables σ, T and μ. For T constant, we obtain

$$\left.\frac{\partial\sigma}{\partial\mu_i}\right|_{T,\mu_j} = -\Gamma_i \tag{4.21}$$

which can be interpreted as meaning that upon increase of the quantity i in the surface, then σ is decreased on raising μ_i or if by raising μ_i, the work to produce new surface, σ, is decreased, then an excess of i will be found in the surface.

Perhaps this relation can be more easily appreciated by applying the Gibbs' equation to a binary system at constant T. Then

$$(d\sigma/dX_2) = -\Gamma_1(d\mu_1/dX_2) - \Gamma_2(d\mu_2/dX_2) \tag{4.22}$$

But, by the Gibbs'–Duhem relation $X_1 d\mu_1/dX_2 + X_2 d\mu_2/dX_2 = 0$, then

$$(d\sigma/dX_2) = (\Gamma_1 X_2/X_1 - \Gamma_2)d\mu_2/dX_2 \equiv -\Gamma d\mu_2/dX_2$$

$$= -(\Gamma_2 - \Gamma_1 X_2/X_1)d\mu_2/dX_2 \equiv -\Gamma d\mu_2/dX_2 \tag{4.23}$$

$$(\Gamma \equiv \Gamma_2 - \Gamma_1 X_2/X_1)$$

(Gibbs' chose the dividing plane of the surface such that $\Gamma_1 = 0$.)

We have already learned in Chapter II that in the composition range where a solid solution is stable the sign of $d\mu_2/dX_2$ must be positive. Hence, in stable solid solutions there will be an excess of the solute 2 at the surface when the addition of solute acts to decrease the surface tension.

The Gibbs' adsorption isotherm that has been derived for a surface applies as well to any interface, such as a grain boundary in a single-phase system. Application to multiphase systems is somewhat complicated (see Cahn in Reference 10).

5.2. Guggenheim's pseudo-thermodynamic model of an interface phase

Contrary to the conclusions we are led to make on the basis of Gibbs' approach to the thermodynamics of interfaces, we now start with the hypothesis that we can describe an interface as a separate phase with homogeneous properties. This approach was originally developed by Guggenheim[7] and more extensively by others[5].

Thus, the surface phase will have a Gibbs' free energy, ${}^{s}G_{m}$, per mole that depends upon composition in the normal way. However, the value of this molar energy at each terminal composition is related to the Gibbs' molar free energy in the corresponding bulk phase by

$$
\begin{aligned}
{}^{s}G_{1o} - {}^{b}G_{1o} &= \sigma_{1o}{}^{s}A_{1o} \qquad (4.24)\\
{}^{s}G_{2o} - {}^{b}G_{2o} &= \sigma_{2o}{}^{s}A_{2o}
\end{aligned}
$$

(${}^{s}A_{1o}$ is the area per mole of pure component 1 of surface phase s.)

We are now concerned with evaluation of the distribution coefficient per solute for equilibrium between the bulk b and surface phases. Let the virtual transfer of atoms correspond to the exchange of a solvent atom in the boundary phase with a solute atom in the bulk phase at constant number of atoms in each of the surface and bulk phases. For this equilibrium, the change in free energy of the surface phase in the exchange must equal that for the bulk phase. Thus

$$
{}^{s}\mu_2 - {}^{s}\mu_1 = {}^{b}\mu_2 - {}^{b}\mu_1 \qquad (4.25)
$$

or

$$
{}^{s}\bar{G}_2 - {}^{s}\bar{G}_1 = {}^{b}\bar{G}_2 - {}^{b}\bar{G}_1
$$

describes this equilibrium. But

$$
\bar{G}_i = G_{io} + RT \ln X_i + {}^{xs}G_i \qquad (4.26)
$$

Hence

$$\begin{aligned}
&{}^s\bar{G}_2 - {}^s\bar{G}_1 = {}^sG_{2o} + RT\ln {}^sX_2 + {}^{s,xs}G_2 - {}^sG_{1o} - RT\ln {}^sX_1 - {}^{s,xs}G_1 = \\
&{}^b\bar{G}_2 - {}^b\bar{G}_1 = {}^bG_{2o} + RT\ln {}^bX_2 + {}^{b,xs}G_2 - {}^bG_{1o} - RT\ln {}^bX_1 - {}^{b,xs}G_1
\end{aligned} \tag{4.27}$$

Rearranging terms yields

$$\begin{aligned}
{}^sX_2/{}^sX_1 = ({}^bX_2/{}^bX_1)\exp\{&-[({}^sG_{2o} - {}^bG_{2o}) - ({}^sG_{1o} - {}^bG_{1o}) \\
&+ ({}^{s,xs}G_2 - {}^{b,xs}G_2) - ({}^{s,xs}G_1 - {}^{b,xs}G_1)]/(RT)\}
\end{aligned} \tag{4.28}$$

$${}^sX_2/{}^bX_2 = ({}^sX_1/{}^bX_1)\exp\{-[(\sigma_{2o}{}^sA_{2o} - \sigma_{1o}{}^sA_{1o}) + \Delta^{xs}G_2 - \Delta^{xs}G_1]/(RT)\}$$

Thus, we obtain again a distribution coefficient relation equivalent to that between liquid and solid phases (see equation 3.2) except that the terms have somewhat different significance. Segregation of solute 2 to the interface phase is promoted by a value of ${}^{b,xs}G_2 >> {}^{s,xs}G_2$ and ${}^{s,xs}G_1 >> {}^{b,xs}G_1$. Note that these quantities are evaluated at the respective concentrations of the interface (${}^sX_i$) and bulk (${}^bX_i$) phases. If ${}^{b,xs}G_2$ is due to strain energy, then certainly the similarity of the liquid–solid distribution coefficient to that of the surface/bulk one is apparent. This analogy has proved to be useful in predicting the likely segregation of solute to interfaces on the basis of experimental phase diagrams for the system in question.

The approach of this section can be illustrated graphically, as follows. We draw the free energy–composition curves for the host phase and for the host surface phase, as shown in Figure 4.5. We note that the equilibrium being considered involves an exchange of a solvent atom in the surface phase with a solute atom in the host phase or vice versa. In either case, the condition describing the equilibrium is given by equation 4.25 and the relation described by equation 4.25 is shown in Figure 4.5, where the straight lines shown have the same slope.

It must be emphasized that the relations obtained in this section are not rigorous. Further, they have been developed only for the special case of exchange of a bulk sited atom with one in the so-called surface phase. The possibility exists that surface adsorption can occur without the necessity of such an exchange. For example, interstitial sited atoms in the bulk can move to the surface region without requiring surface atoms to move into the bulk. There are more than enough vacant interstitial sites in both regions to accommodate the transfer of interstitial sited atoms from one region to the other. Nevertheless, the relations 4.28 are useful where applicable.

A correlation between the segregation ratio, ${}^sX_2/{}^bX_2$, and the solid solubility has been found.[8] We can develop the bases for this correlation using relations

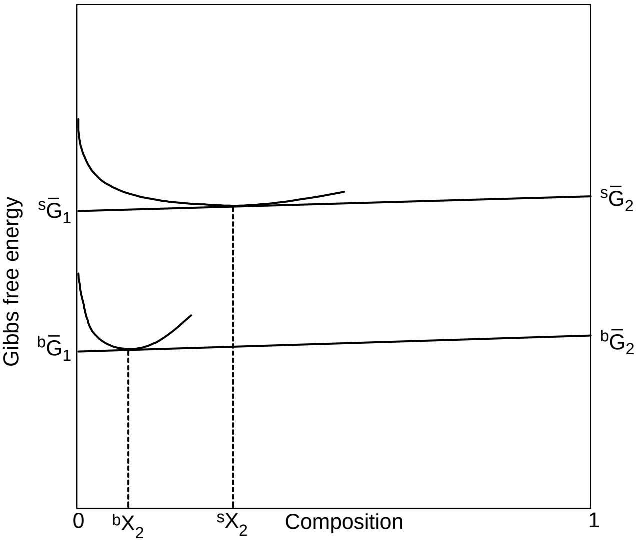


Figure 4.5. Free energy–composition curves for a bulk and its surface phase illustrating the relation for the equilibrium between these phases.

already considered. In particular, we note that for a regular solution that the difference in partial excess molar free energies is given by

$$^{xs}G_2 - {}^{xs}G_1 = -RT\ \ln(^*X_2/^*X_1)$$

where $^*X_i$ are the compositions at the solvus or solid solubility limit.

Substitution of this relation in equation 4.28 yields for the segregation ratio

$$\beta = {}^sX_2/{}^bX_2 = ({}^sX_1/{}^bX_1)(^*X_1/^*X_2)\exp[-\Delta G_s/RT] \qquad (4.29)$$

where $\Delta G_s = [(\sigma_{2o}{}^sA_{2o}/{}^sN_{2o} - \sigma_{1o}{}^sA_{1o}/{}^sN_{1o}) + ({}^{s,xs}G_2 - {}^{s,xs}G_1)]$

Now, if we consider only the case of dilute solutions, then equation 4.29 becomes

$$\beta = (1/^*X_2)\exp[-\Delta G_s/RT]$$

Figure 4.6(a) shows a plot of experimental values of the enrichment ratio for adsorption at grain boundaries versus the solid solubility in metallic hosts. The scatter of a factor of 5 in the enrichment ratio is in part a consequence of the fact

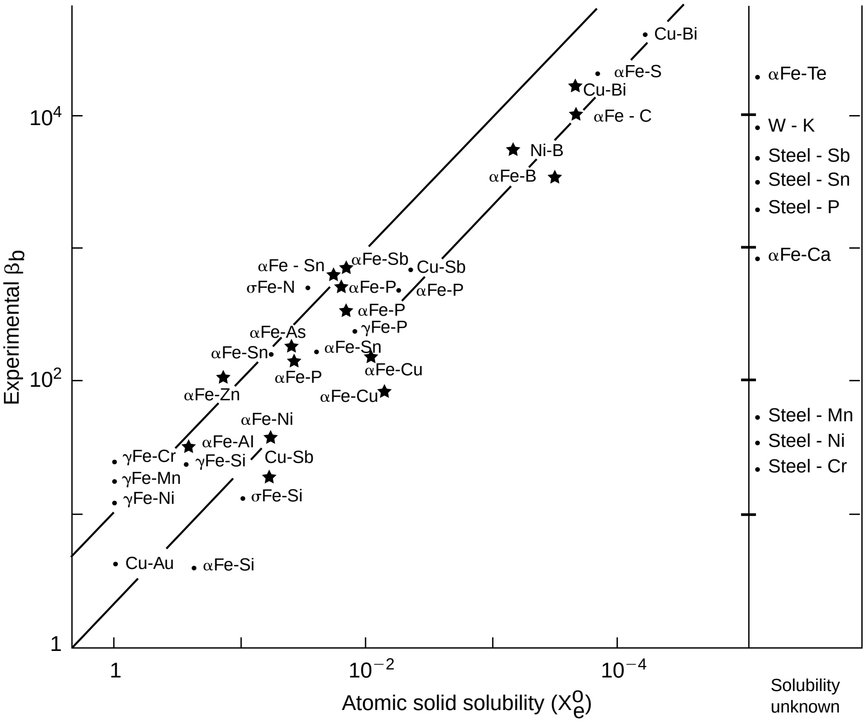


Figure 4.6a. Correlation of measured grain boundary enrichment ratios with the atomic solid solubility. From E.D. Hondros and M.P. Seah in Physical Metallurgy, eds. R.W. Cahn and P. Haasen, 3rd edition, North-Holland Physics Publishing, New York, 1983 with permission.

that ΔG_s varies with grain boundary misorientation.[9] Incidentally, the data at a value of the solubility equal to unity (complete solid solubility) appear to obey the correlation although the theory should apply only to dilute solution. Little is known about the segregation of solutes to surfaces in ceramics, except that such segregation can be marked and have significant effects on properties. Since the above relations are general and not specific to any material, there is no reason not to expect them to operate in ceramics, as well as they do in metals.

The enrichment of a surface in solute can lead to surface phase transitions. Little is known about this phenomenon. However, it has been found that the segregation of sulfur on Pt and H_2 on Mo leads to several transitions from one surface unit cell to another. Presumably, further studies will develop many more examples of this phenomenon.

A significant study of Gibbs adsorption to grain boundaries has been carried out by Seidman[10] using both simulations (Monte Carlo) and experiment. He has demonstrated elegantly that Gibbs' adsorption is an anisotropic function of all 5°

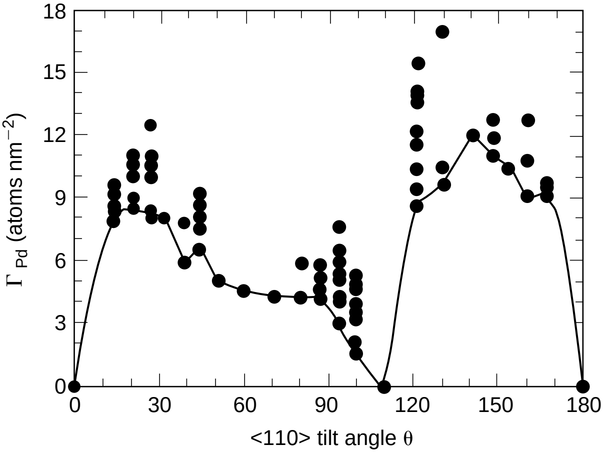


Figure 4.6b. The Gibbsian interfacial excess of Pd versus the tilt angle (μ) for the 21(110) symmetric tilt boundaries in Ni 4 at.% Pd at 800ºK. The curve is drawn through the lowest Γ_{Pd} for each GB. The columns of points are the different Γ_{Pd} values for the multiple structures of some GBs. From Ann. Rev. Mater. Res. 32, 235(2002). © 2002 Annual Reviews.

of freedom of a grain boundary. Thus, the enrichment ratio reported in Figure 4.6(a) represent an average grain boundary segregation. A plot of the Gibbs' adsorption excess Γ_{Pd} as a function of tilt angle for Pd as a solute in Ni is shown in Figure 4.6(b). The multiple points at a constant tilt angle correspond to values of Γ_{Pd} for different grain boundary structures produced by different values of the rigid-body translation vector and local atomic relaxations in addition to the 5° of freedom that define each grain boundary.

5.3. Other adsorption isotherms

An implicit assumption of the derivations in the previous section is that all sites at the surface are available to be occupied by the solute species. If this assumption is not met then other adsorption isotherms are possible. In particular, if the fraction of the total number of sites at the interface that can be occupied by solute equals ${}^{s}X_{2o}$ then replacement of ${}^{s}X_1$ by ${}^{s}X_{2o} - {}^{s}X_2$ in equation 4.28 yields the relation known as the Langmuir–McLean isotherm. The same substitution, but into equation 4.29, and with application to dilute solution only yields the "Truncated BET" isotherm. Distinguishing between these isotherms requires measurements of the composition dependence of the segregation ratio for dilute solution, which is difficult to accomplish.

6. Surface stress

A solid phase differs from a fluid phase in that a change in surface area can be accomplished not only by an exchange of atoms between the interior and the surface, but also by elastic strain. Let us consider the latter situation. A slab is shown in Figure 4.7 of thickness $d \ll L$, the width of the slab. There is a surface stress tensor $g_{\mu\nu}$, representing the force acting on unit length of a line in the surface of the slab, where the index μ can stand for the directions x, y, z and ν for x and y. There is a body stress acting in the bulk of the slab corresponding to $g_{\mu\nu}$ and induced by the latter equal to

$$p_{\mu\nu} = -2\, g_{\mu\nu}/d$$

At equilibrium, the free energy must be stationary with respect to any virtual distortion, such as a homogeneous strain $e_{\mu\nu}$. If A is the area of the face of the slab, then the change in energy in the bulk due to the strain is

$$Ad\Sigma\, p_{\mu\nu}e_{\mu\nu} = -2A\Sigma g_{\mu\nu}e_{\mu\nu}$$

and that in the surface is

$$2A\Sigma(\partial\sigma/\partial e_{\mu\nu})e_{\mu\nu} + 2\sigma\, \Sigma(\partial A/\partial e_{\mu\nu})e_{\mu\nu}$$

Now, for $\mu = x,y$ and $\nu = x,y$, $(\partial A/\partial e_{\mu\nu}) = A\delta_{\mu\nu}$, where $\delta_{\mu\nu}$ is Kronecker's delta (i.e. $\delta_{\mu\nu} = 0$ for $\mu \neq \nu$ and $\delta_{\mu\nu} = 1$ for $\mu = \nu$). Adding and noting that the strains $e_{\mu\nu}$ are arbitrary, we must then have that the coefficients of each $e_{\mu\nu}$ vanish or

$$-2Ag_{\mu\nu} + 2A(\partial\sigma/\partial e_{\mu\nu}) + 2\sigma A\delta_{\mu\nu} = 0$$

or

$$g_{\mu\nu} = \sigma\delta_{\mu\nu} + (\partial\sigma/\partial e_{\mu\nu})$$

In words, for uniaxial elastic strains the surface stress equals the surface tension (i.e. Gibbs' terminology in which the surface tension is a specific surface free energy) plus the variation

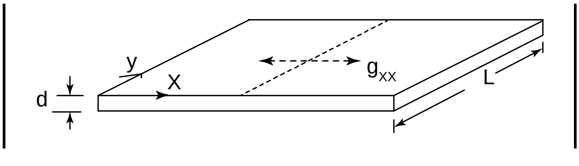


Figure 4.7. Definition of quantities for a surface slab (see text).

of the surface tension with respect to surface elastic strain. For metallic solids the latter quantity is not equal to zero for temperatures less than about one-half the absolute melting temperature. These solids act as fluids above this temperature. The other types of solids undergo plastic strain rates sufficient to relax surface stresses at other fractions of the absolute melting temperature. Above these fluid/rigid divides, the surface stress approximately equals the surface tension. Below this dividing temperature the contribution of the variation of surface tension with respect to elastic strain cannot be ignored.*

7. Surface energies or solid–gas and liquid–gas interface energies

Table 4.1 incorporates representative values of surface energies for various liquids. A generalization from these data is that the surface free energies of liquid salts and oxides (ionic materials) are lower than those for metals. Table 4.2 presents the same comparison for solid surface energies. Again the same conclusion can be drawn. If the interphase energy between an oxide of a metal and the metal is low, then it is apparent that the oxide phase will tend to cover the metal phase to provide the surface phase. Indeed, the observation that oxides do cover their metals after prolonged heating in a gas in equilibrium with the oxide suggests that the sum of the metal–oxide and oxide–vapor interface energies is less than that of the respective metal surface energies. This fact plays a significant role in the important technologies of brazing and soldering. The joining together of two similar or dissimilar metals by either of these methods requires a liquid metal or alloy “cement” to flow between the two solids to be joined. Given the low surface free energies of the oxides, for the “cement” to function it is necessary that either it have a lower interface energy to the oxide (than the oxide surface energy) or that the oxide be removed while the liquid “cement” is in contact both with the surfaces to be joined and the flux that removes the surface oxides.

There is a potential for application of the surface energy data described in this section that continually amazes this author. For example, the so-called C4 process of joining integrated circuit chips to their packages by a self-aligning procedure, whereby hundreds of electrical connectors separated by microscopic dimensions are made in one operation, makes clever use of these data. In this process, each molten ball of solder is prevented from spreading beyond its connector area by the

---

* Stresses exist along interfaces as well as along surfaces. They have been measured in thin films between layers of different metals and were found to be in the compressive state in Ag–Cu and Ag–Ni interfaces,[11] contrary to a prediction based on an embedded atom method.[12]

Table 4.1. Surface free energy of liquid metals and typical molten salts and inorganic materials.

Metal	**Surface energy (erg/cm^2)**	**Temperature (°K)**	**Metal**	**Surface energy (erg/cm^2)**	**Temperature (°K)**
Al	866	600(mp)	Mo	2250	2620(mp)
Ag	895	1000	Na	191	98(mp)
Au	1140	1063(mp)	Ni	1780	1455(mp)
Ba	225	720	Os	2500	3000(mp)
Be	1100	1500	Pb	450	327(mp)
Ca	335	850	Pd	1500	1547(mp)
Cs	60	29(mp)	Pt	1800	1773(mp)
Cd	590	321(mp)	Rb	76	39(mp)
Cr	1590	1950	Re	2700	3200(mp)
Co	1880	1495(mp)	Rh	2000	1966(mp)
Cb	1900	2473(mp)	Ru	2250	2250(mp)
Cu	1300	1083(mp)	Sb	380	640
Fe	1880	1535(mp)	Si	730	1410(mp)
Ga	720	30.3(mp)	Sn	550	232(mp)
Ge	610	960	Ta	2150	3020(mp)
Hf	1630	2230(mp)	Tc	2100	2220(mp)
Hg	475	20	Ti	1650	1730(mp)
In	560	156(mp)	Tl	440	318
Ir	2250	2454(mp)	U	1550	1132(mp)
K	101	64(mp)	V	1950	1710(mp)
Li	398	180(mp)	W	2400	3410(mp)
Mg	540	651(mp)	Zn	770	420(mp)even
Mn	1060	1245(mp)			

Salts	**Surface energy (erg/cm^2)**	**Temperature (°K)**	**Salts**	**Surface energy (erg/cm^2)**	**Temperature (°K)**
LiF	236	1121(mp)	$NaNO_3$	120	583(mp)
NaF	186	1269(mp)	B_2O_3	80	1173
UF_4	196	1309(mp)	FeO	585	1693
LiCl	126	883(mp)	Al_2O_3	700	2353
NaCl	114	1074(mp)	$ZnBr_2$	51	675(mp)
CuCl	92	703(mp)	$CdCl_2$	100	841(mp)
AgCl	179	728(mp)	$SnCl_2$	104	518(mp)
$CaBr_2$	120	1015(mp)	$PbCl_2$	138	771(mp)
Li_2CO_3	244	993(mp)	$GaCl_3$	27	351(mp)
$LiNO_3$	116	527(mp)	$HgCl_2$	56	550(mp)
Na_2SO_4	195	1157(mp)	$BaBr_2$	153	1127(mp)

Table 4.2. Surface free energies of solids.

Metals	**Surface energy (erg/cm^2)**	**Temperature (°C)**	**Metals**	**Surface energy (erg/cm^2)**	**Temperature (°C)**
Al	980	450	Pb	560	300
	1100	200	Pt	1950	1700
Ag	1100	950		2200	1300
	1200	850	Sn	670	200
Au	1400	1000		685	223
Bi	500	240	Ta	2480	2700
Cd	670	300	Ti	1940	1300
Co	2420	1400	Tl	560	280
	1970	1354	W	2800	2000
Cr	2090	1700	Zn	870	400
Cu	1520	1000	**Non-metals**		
	1780	925	NiO	1100	25
Fe(bcc)	1930	1475	MgO	1000	25
Fe(fcc)	2100	1350	KCl	110	25
Ga	770	20	NaCl(100)	300	25
In	630	140	Al_2O_3	905	1850
Mo	2630	2400	TiC	1190	1100
Nb	2150	2225	LiF(100)	340	25
Ni	1940	1400	CaF_2(111)	450	25
	2280	1060	$CaCO_3$(1010)	230	25

presence of oxides separating the connectors, which themselves are gold plated to prevent oxidation of the connector surfaces and to wet the solder.

8. Solid–liquid interfaces

Much has been written about the solid–liquid interface,[13] perhaps because it is involved in the technologically important problem of solidification. However, little is known about it. The one conclusion that can be deduced from a comparison of the data in Table 4.3 with those for the same metals in Table 4.1 is that the solid–liquid interface energy is about an order of magnitude smaller than the surface energy and smaller than the difference between solid and liquid surface energies. Thus, we have the general rule that *liquids wet their own solids*.

A rough rule for the solid–liquid interface energy for metals is that the molar interface energy (= $N^{1/3}\Omega^{2/3}\sigma_{sl}$, where N is Avogadro's number and Ω is the

Table 4.3. Solid–liquid interface energies.

Metal	Interface energy (erg/cm^2)	Temperature (°C)
Al	93	600
Au	132	1063
Bi	61	271
Cu	177	1083
Fe	204	1535
Ni	255	1455
Pt	240	1773
Sn	55	232

molar volume) equals about one-half the heat of fusion. For semi-metals and organic compounds the proportionality constant is about 0.3. The small value for the solid–liquid interface energy suggests that the excess volume associated with the interface must be small. This brings up an interesting point. In Gibbsian thermodynamics of surfaces, the surface excess of volume is zero by definition. In fact, it isn't. Cahn[14] has described a surface thermodynamic method in which the excess volume has meaning and for relating such excess quantities to measurable quantities. A fuller description of Cahn's method is outside the scope of this chapter because the main advantages of using Cahn's method are derived from more detailed applications to special cases, such as the case of interfaces when there are three-or-more coexisting phases.

9. Solid–solid interfaces

9.1. Grain boundaries

There are a variety of solid–solid interfaces as listed in the Table 4.4 for copper. It should be noted that the interface energy can vary from a very small number for coherent type interfaces up to about 1/3 the surface energy for high-angle grain boundaries. This is a general result for metals. For ionic and covalently bonded solids, the few measurements that have been made suggest that the ratio of the high angle grain boundary energy to the surface energy can reach a much higher value. In NiO, for example, this ratio approaches unity for [110] symmetric tilt boundaries.

The energy of very low angle tilt or twist boundaries is described quite well by dislocation theory (see any current textbook on dislocation theory). The tilt and twist small angle boundaries are illustrated in Figures 4.8 and 4.9, which show the dislocation content of such boundaries clearly. Also, implicit in these figures is the relationship of these boundaries to an axis of rotation of one grain relative to its

Table 4.4. Interface energies for copper.

Interface	Energy (erg/cm^2)
(High-angle grain boundary)	625
(Coherent twin boundary)	24
(Non-coherent twin boundary)	498
(Intrinsic stacking-fault)	78

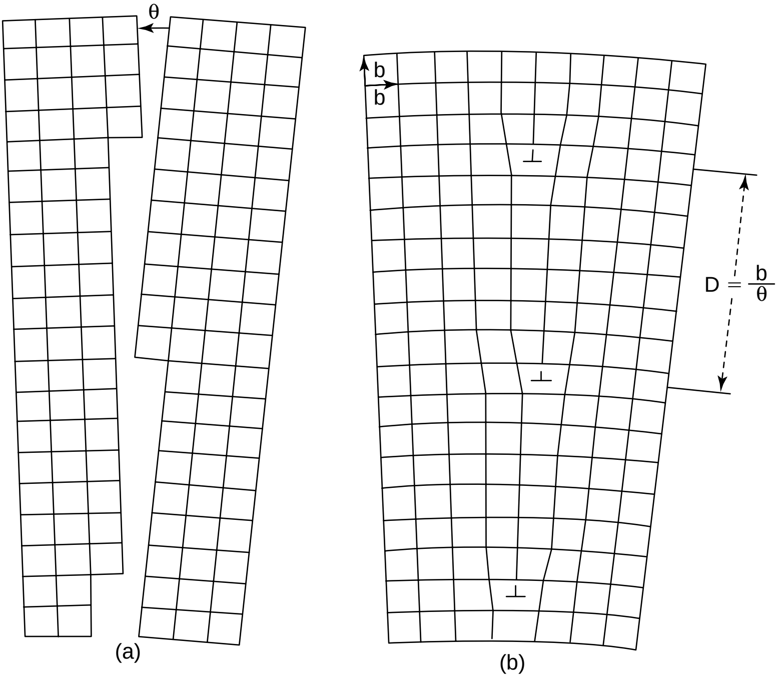


Figure 4.8. Schematic illustration of a tilt boundary formed by joining a bicrystal, each crystal of which has a common axis normal to the plane of the paper. One crystal is rotated relative to the other about this axis by the angle θ.

bounding grain. In the case of the tilt boundary this axis of rotation is in the plane of the grain boundary and in the case of the twist boundary it is perpendicular to the plane of the grain boundary. One explanation for the difference in grain boundary energies between ionic materials and metals for similar boundaries is that the Burger’s vector in ionic solids is about twice that in metals due to the larger unit cell.

Computer simulation of higher angle tilt and twist grain boundaries have provided some insight concerning the structure and energies of such boundaries.

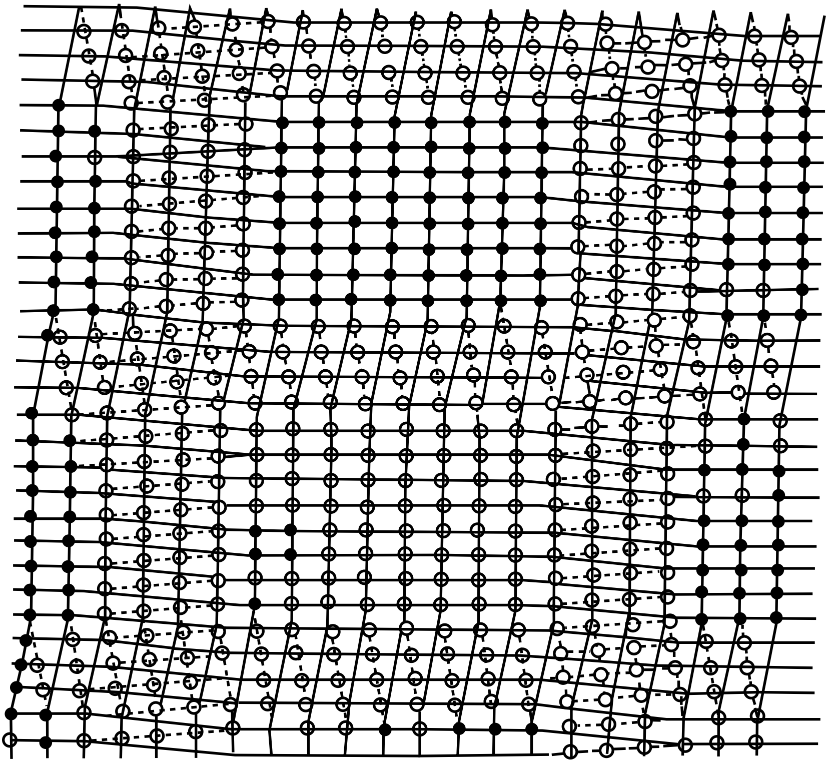


Figure 4.9. Schematic illustration of a twist boundary formed by rotating one-half of a bicrystal about an axis normal to the paper and boundary relative to the other half crystal.

This subject is still under development and the interested reader is urged to peruse the articles in the Bibliography relating to this subject at the end of this chapter. At this time, it appears that a reasonable model of such grain boundaries yields the result that the atoms along such tilt and twist boundaries tend to configure themselves into arrangements that describe polyhedra of the type that Bernal has suggested exist in liquids. The orientations at which the boundaries consist only of one type of polyhedra are those at which the grain boundary energy exhibits minima or cusps in the plot of energy versus tilt or twist orientation.

The energy of high-angle boundaries is less readily predicted by theory, although current computer simulation models of grain boundaries, using appropriate interatomic potentials and taking into account macroscopic relations that can occur due to relative translation and rotation of grains adjoining the boundary, yield fair approximations for such energies.[15] The high-angle boundary is not a well-defined species since it is a function of 5 variables (2 for the relative orientation of

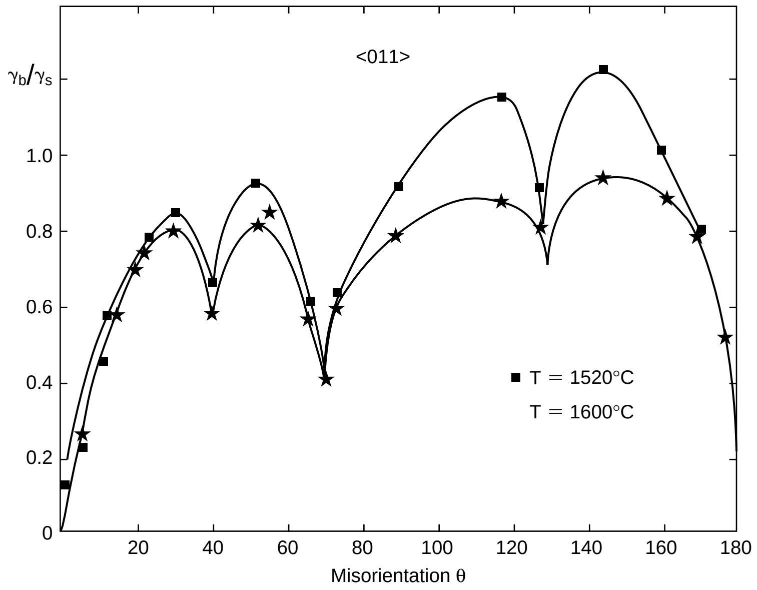


Figure 4.10. Relative <011> symmetrical tilt boundary energy in NiO as a function of the tilt angle. From G. Dhalenne et al. in Advances in Ceramics, Vol. 6, 1983, p. 139 with permission.

the grains adjoining the boundary and 3 to define the orientation of the grain boundary plane normal relative to the crystal axes of one of the bounding grains.) Hence, the structure of a high-angle boundary can vary enormously, as can its energy. Indeed, at special high-angle orientations a high degree of lattice sites at the boundary can be common to both adjoining crystals. Such special boundaries will have lower energy. Consequently, one expects the existence of low energy cusps in the polar plots of the orientation dependence of grain boundary energies – i.e., the grain boundary energy is expected to be orientation dependent. Indeed, experiment is consistent with this expectation, as shown in Figure 4.10. This feature, as will be shown in the following section, is responsible for the existence of torques acting on boundaries tending to reorient them into lower-energy orientations.*

The brief discussion of grain boundary energies in the previous paragraph is applicable to all materials. However, materials do differ in the details concerning their grain boundary structure and energy. In particular, ceramics (ionic) must obey the constraint of electroneutrality, with a consequent effect on the structure of their grain boundaries: the development of a cloud of defects adjacent to the grain boundary to screen any net charge on the boundary; or the presence of vacancies on the appropriate sublattices at the grain boundaries sufficient to provide the required electroneutrality, and the sensitivity of these grain boundaries to impurities. Also, the relatively small-free energy difference between crystalline and amorphous phases of covalently bonded solids and the high-grain boundary energies in some of these crystalline materials may allow thin films of the amorphous phases to become stable at such grain boundaries. Pertinent to this argument is the fact that liquid–solid interface energies in a given material are smaller in magnitude than general high-angle grain boundaries and, thus, one may expect amorphous–crystalline interface energies to be smaller than some of the high-angle grain boundary energies in covalently bonded solids. Indeed, molecular dynamic-based simulations of grain boundaries in silicon reveal a stable confined amorphous structure of $\sim$0.25 nm width at high-angle grain boundaries, whereas low-angle grain boundaries are crystalline and sharper.[16] This case is a variant of a general phenomenon in which a thin film of a bulk metastable phase becomes the stable phase at an interface as a consequence of its low-interface energies. Another variant of this phenomenon is discussed in Section 12 of this chapter.

---

* Calculations in Phys. Rev. B. 60, 2740(1999) imply an inconsistency with respect to the data in Figure 4.10 that is not as yet understood. Much more data of this kind for ceramic crystals are needed to resolve this inconsistency which relates to the torque term.

9.2. Interphase interfaces

In the case of grain boundaries, the interface is between two differently oriented crystals of the same homogeneous phase. Interfaces exist also between two different phases. These can include solid–vapor, solid–liquid and solid–solid interfaces. Solid–solid interphase interfaces occur between precipitates in solid hosts and the host phase. Such interfaces can be coherent, if there is an appropriate matching of the relative orientations of the two phases, of the crystal structures and of the lattice parameters. Table 4.5 lists some interphase interface energies.

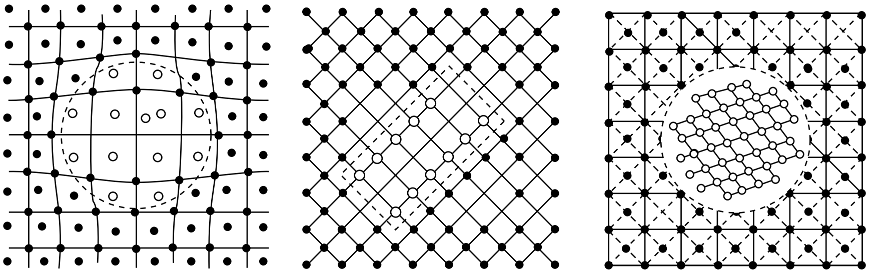
The interphase interface energies in the table are for non-coherent interfaces, and, as shown, these values can be large. Figure 4.11(a) illustrates a coherent

Table 4.5. Interphase interface energies.

Interphase system	**Interface energy (erg/cm^2)**	**Temperature (°C)**
Au/Al_2O_3	1725	1000
Ag/Al_2O_3	1630	700
Cu/Al_2O_3	1925	850
Ni/Al_2O_3	2140	1000
Pt/Al_2O_3	1050	1400
Ni/ThO_2	2000	1200

interface between a precipitate and its host phase. As shown, there is a continuity of rows and planes of atoms across the interface. Figure 4.11(c) illustrates a non-coherent interface. It is apparent that the non-coherent interface exhibits much greater misfit than does the coherent interface. It is no surprise, therefore, that the non-coherent interphase interface energies in Table 4.5 are large.

Figure 4.12 shows a lattice image electron micrograph for an interface between a silver host and a CdO precipitate, where the continuity of atomic rows across the interface is apparent. By computer simulation and comparison it is demonstrated that the outermost plane of the oxide consists of oxygen ions and not



(a) (b) (c)

Figure 4.11. Schematic illustration of a precipitate: (a) quasi-coherent or coherent with positive misfit; (b) semi-coherent and (c) non-coherent. From J.W. Martin, Micro-Mechanisms in Particle-Hardened Alloys, Cambridge University, 1980 with permission.

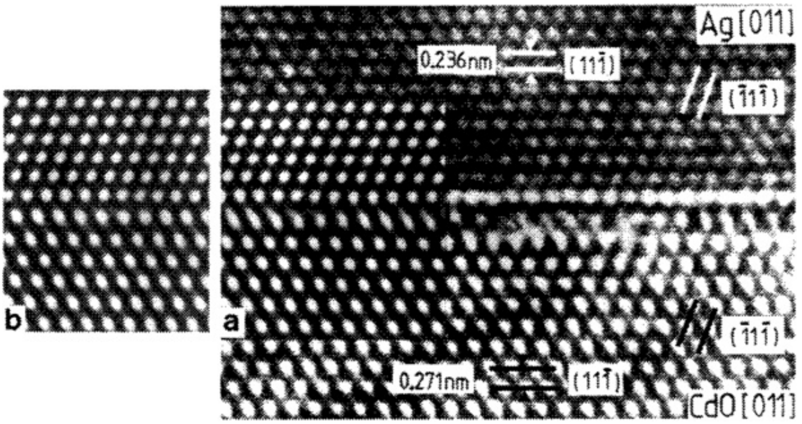


Figure 4.12. Lattice image TEM of Ag–CdO interface. Inserts (a) inside box and (b) show calculated contrast for (a) oxygen and (b) cadmium as the outermost plane of the oxide. From H.F. Fischmeister et al. Mat. Res. Soc. Symp. Proc. 122, Mat. Res. Soc., 1988 with permission.

cadmium. The interface energy of an interphase interface can usually be partitioned into two parts. One is of chemical origin involving the change in composition that occurs across the interface and the other is geometric in the sense that it includes the energy associated with the misfit between the two lattices at the interface. It is reasonable then that the geometric portion of the interphase interface energy for a coherent interface is negligible. Very little is known about the energy of interphase interfaces at this writing although it is an active area of research, in part, because such information is vital to the solution of joining problems involving dissimilar materials. The misfit energy associated with the deviation from coherence, such as occurs in a semi-coherent boundary containing misfit dislocations, can be estimated from dislocation theory.[17] Much remains to be done to elucidate the energy and structure of interphase interfaces.

9.3. Local equilibrium at grain boundary intersections

Consider the situation described in Figure 4.13. We shall allow, in turn, virtual motion of the grain boundaries from the position of the solid lines to those of the dotted lines and we shall calculate the change in free energy accompanying such motion. The change in interface area induces the change in free energy given by

$$(\sigma_1 - \sigma_2 \cos\varphi - \sigma_3 \cos\theta)\overline{\mathrm{OP}}$$

The change in free energy due to the change in grain boundary orientation is

$$\overline{\mathrm{AP}}(\partial\sigma_2/\partial\varphi)\delta\varphi + \overline{\mathrm{BP}}(\partial\sigma_3/\partial\theta)\delta\theta$$

But

$$\delta\varphi = (\overline{\mathrm{OP}}\sin\varphi)/\overline{\mathrm{AP}} \quad \text{and} \quad \delta\theta = (\overline{\mathrm{OP}}\sin\theta)/\overline{\mathrm{BP}}$$

Setting the total-free energy change equal to zero yields

$$\sigma_1 - \sigma_2 \cos\varphi - \sigma_3 \cos\theta + \sin\varphi(\partial\sigma_2/\partial\varphi) + \sin\theta\,(\partial\sigma_3/\partial\theta) = 0$$

Similar equations may be obtained for the virtual displacement of the boundaries in the direction of the vector contained in the stationary boundary and normal to the line of intersection of the three boundaries. The resulting relations can be described by the following equation

$$\Sigma_{i=1}^{3}\,(\sigma_i\mathbf{t}_i + (\partial\sigma_i/\partial\mathbf{t}_i)) = 0$$

where $\mathbf{t}_i$ is a unit vector in the plane of the ith grain boundary, perpendicular to the line of intersection of the grain boundaries and pointing away from this line. The derivative term on the left-hand side of the equation is called a "torque" term because it is a driving force tending to reorient the grain boundary. If these torque terms are zero in value then the above equation can be shown to reduce to

$$\sigma_1/\sin\theta_1 = \sigma_2/\sin\theta_2 = \sigma_3/\sin\theta_3$$

where θ_i is the angle opposite the i boundary. This relation is the same as found in mechanics for the vector resolution of forces acting at a vertex. This analogy leads to the simple relations describing the local equilibria shown in Figure 4.14.

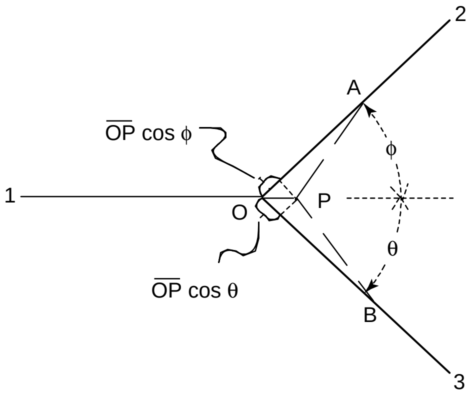
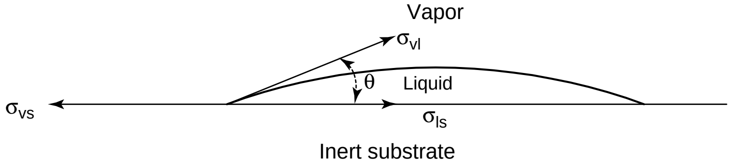
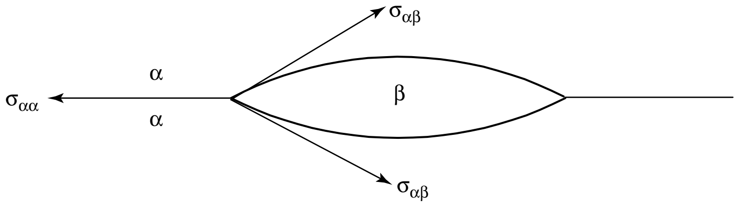

2
A
$\overline{OP}\cos\phi$
ϕ
1
P
O
θ
$\overline{OP}\cos\theta$
B
3

Figure 4.13. Schematic illustration of virtual displacement of grain boundaries and the changes in area and orientation that occur.


Vapor
σ_{vl}
θ
Liquid
σ_{vs}
σ_{ls}
Inert substrate

(a) $\sigma_{vs} = \sigma_{vl}\cos\theta + \sigma_{ls}$


$\sigma_{\alpha\beta}$
α
$\sigma_{\alpha\alpha}$
β
α
$\sigma_{\alpha\beta}$

(b) $\sigma_{\alpha\alpha} = 2\sigma_{\alpha\beta}\cos\frac{\theta}{2}$

Figure 4.14. Schematic illustration of the “law of forces” representation of the surface tensions in local equilibrium at an intersection.

10. Diffuse interfaces

In the above we have implicitly assumed that the interface is a region of transition between its bounding phases, with the thickness of the interface on the order of atomic dimensions. However, the interface can be much thicker. The interface thickness depends upon the circumstances and is not a parameter to be assumed. Cahn and Hilliard[18] were the first to derive an analytic relation describing the interface parameters as a function of distance normal to the interface between two infinite blocks that are in equilibrium with each other. They expanded the local free energy (free energy/unit volume), which because it is a scalar quantity includes only the even terms in the expansion. They applied this concept to an incompressible binary fluid. The interfacial energy they derived is

$$\sigma = N_v \int_{-\infty}^{\infty} [\Delta f(C) + K(dC/dx)^2]dx$$

where N_v is the number of molecules per unit volume, $\Delta f(C) = f(C) - C_1\mu_1 - C_2\mu_2$ and is shown in Figure 4.15.

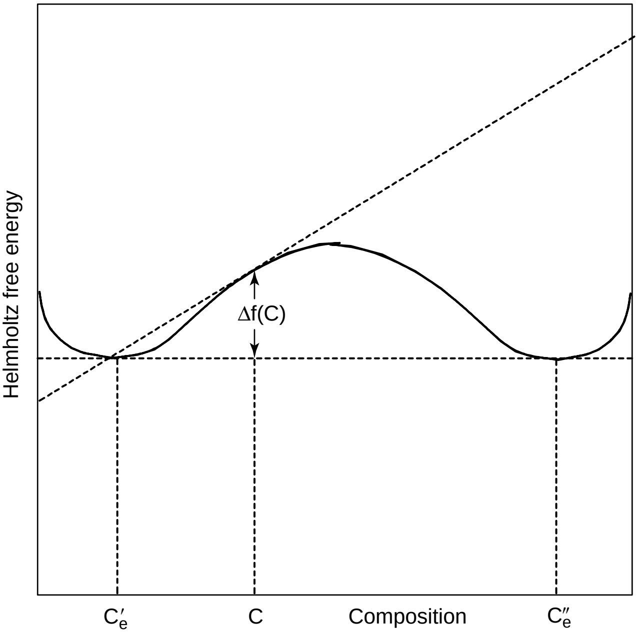


Figure 4.15. Defining compositions and free energies used in expression for thickness of diffuse interface.

At equilibrium the integrand in the equation for the interface energy must be at a minimum which is equivalent to the statement that the chemical potential of each species is constant throughout. This requirement then leads to the relation

$$(dC/dx) = [\Delta f(C)/K]^{1/2}$$

Substitution of this relation in that for the interface energy then yields

$$\sigma = 2N_v \int_{-\infty}^{\infty} \Delta f(C)dx = 2N_v \int_{C'}^{C''} [K\Delta f(C)]^{1/2}$$

Now suppose that the composition in the interface between the two phases in equilibrium with each other varied with distance normal to the interface, as shown in Figure 4.16. A thickness of the interface region can be defined to be

$$\delta_e = (C''_e - C'_e)/(dC/dx)_{max}$$

where the slope of the composition is the maximum in the transition region. Thus, the thickness can be written as

$$\delta = (C''_e - C'_e)(K/\Delta f(C))^{1/2}$$

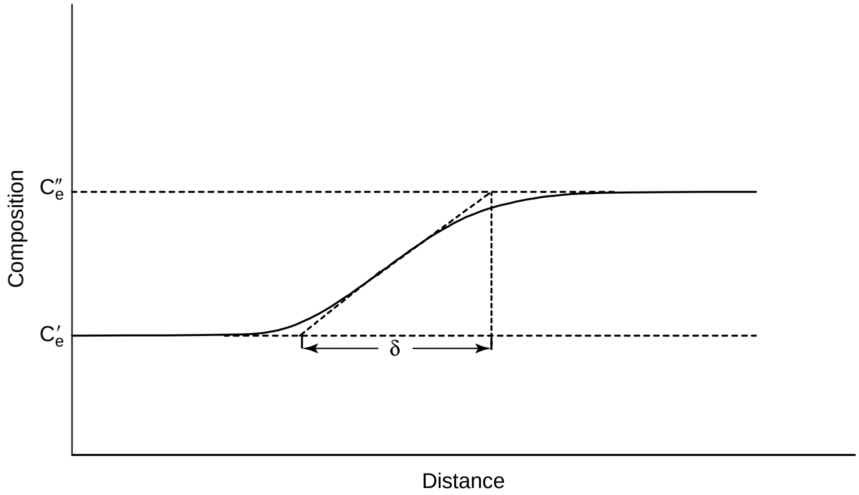


Figure 4.16. Illustrating the change in composition with distance at an interface between two phases having the same crystal structure.
]

For a miscibility gap system corresponding to the free energy curve shown in Figure 4.15, Cahn and Hilliard have evaluated the different quantities in numerator and denominator of this relation for interface thickness. In the vicinity of the critical temperature, if $\Delta f(C)$ can be expanded as a Taylor's series about T_C and C_C then

$$\Delta f(C) = -b(T_C - T)[(C - C_C)^2 - (C'' - C_C)^2] \\ + g[(C - C_C)^4 - (C'' - C_C)^4] + \cdots$$

and $[(C'' - C')/2]^2 = b(T_C - T)/2g$, while

$$\Delta f(C) = g\{[(C'' - C')/2]^2 - (C - C_C)^2\}^2$$

Substitution of these expressions into the relation for the interface thickness, δ, yields in the vicinity of the critical temperature

$$\delta = 2[2K/b(T_C - T)]^{1/2}$$

Thus, as the temperature approaches the critical temperature from below, the interface thickness increases and approaches infinity.

Lee and Aaronson[19] have provided equations for the interface energies of diffuse interfaces deduced from a discrete lattice plane model and determined the equivalence of their relations to the relation based on the continuum model given above. Based on these equations, Aaronson and LeGoues[20] calculated the energies of {111} and {100} diffuse interfaces in dilute Cu-based alloys supersaturated with a Co constituent. They found values of 180 mJ/m^2 for {111} and 190 mJ/m^2 for {100} at 873°K, in good agreement with values back calculated from the nucleation experiments of Servi and Turnbull[21] and LeGoues and Aaronson.[22] It should be remarked that the alloy systems which tend to exhibit diffuse interfaces (i.e. systems where all the pure components have the same crystal structure) are not likely to obey the regular solution approximation, which was assumed in the foregoing work. For these systems more accurate means of evaluating both the energy and entropy of the solid solution is needed. Asta[23] has used two such statistical–mechanical techniques to evaluate the energetics and entropy for the case of the Al–Li system which involve fitting to the phase diagram for the system and, in addition, for the Al–Ag system, the energetic parameters were derived from first principles total energy calculations. The results of these calculations are given in Figure 4.17(a) and (b) for Al–Li and in Figure 4.17(c) for Al–Ag. As shown, the calculated diffuse interface energy (denoted IPB energy in the figures) increases with decreasing temperature in the two-phase region and is consistent with the experimental values, where available. The difference in absolute values

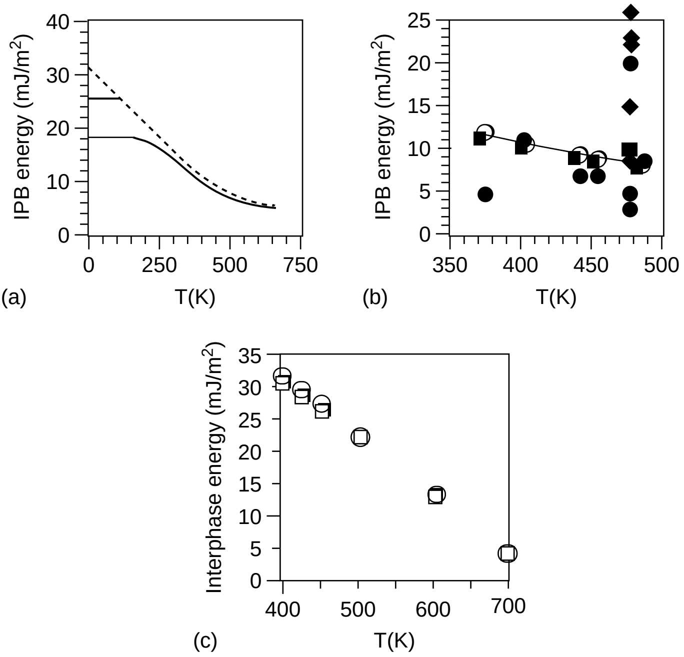


Figure 4.17. Diffuse interface energies: (a) Al–Li system, {100}(smallest values), {111} (largest values), {110} (intermediate values); (b) Al–Li system, open circles are {100} calculated values, filled points are values deduced from experimental kinetic data; (c) Al–Ag system, calculated values, {100} values as circles, {111} values as squares. From M. Asta, MRS Symp. Proc. 398, 281(1996) with permission.

between the interface energy values shown in this figure (5–30 mJ/m^2) with those deduced for the Cu–Co system (180–190 mJ/m^2) reflects the difference in critical temperatures between these specific Al- and Cu-based alloy systems (600−700°K versus 2768°K, respectively). The calculated widths of the diffuse interface are 24 Å in Al–Li[23] at 482°K, 16 Å in Al–Ag[23] at 450°K and about 5.4 Å in Cu–Co[19] at 0.75 T_C.

11. Methods of measuring interface energies

The methods for measuring the liquid–vapor interface energy have been described in detail[24] and need not be repeated here. Reviews of the measurements of liquid–solid, solid–vapor and solid–solid interfaces are rarer and, hence, some space will be devoted to brief descriptions of the methods used. The main technique for measuring solid–vapor interface energy for the case of metals and alloys is the zero creep method.[24] This technique involves the equilibration of a fine filament, both with respect to its environment and with respect to a load that tends to extend the filament at an elevated temperature. By systematically testing filaments with successive loads and measuring the extension rates of the filaments it is possible to deduce the load corresponding to zero creep rate of the filament. Since the surface energy of the filament itself acts to contract the filament, zero creep rate corresponds to a balance of the rate of change of $2\pi r\sigma$ with respect to length and the load, at constant volume, from which it is possible to calculate the surface energy in the event that the filament is a single crystal. Usually, the filament is a polycrystal in the form of a bamboo structure, in which the grain boundaries are normal to the filament's axis. Correction must be made for the contribution of the energies of the grain boundaries to the change in shape of the filament. In order to be able to deduce a value for the surface energy it is necessary to have a knowledge of the ratio of the energy of such grain boundaries to that of the surface. This ratio can be deduced from measurements of the contact angle of the thermal grooves developed where these grain boundaries intersect the surface. A variation of the zero creep method has been used in thin films to ascertain interface energies.[16]

Measurement of solid surface energies for non-metallic, brittle materials is accomplished using a controlled cleavage technique where the force required to form new surface by extension of a crack is balanced against the surface energy of the new surfaces produced by the crack extension.[25] The controlled cleavage technique has also been used to measure grain boundary energies in brittle materials.[26] Modern evaluation of this technique is in terms of fracture mechanics considerations in which the crack propagation resistance G_C is measured, where G_C is proportional to the specific energy of the surface created in an ideally brittle material. Another method of measuring solid surface energies is heat of solution calorimetry. Grain boundary energies are generally measured by equilibrating the grain boundaries and the surface and evaluating the groove angles. Finally, solid–liquid energies are measured by equilibrating a grain boundary in contact with the liquid phase in a temperature gradient. The theory of this technique may be obtained in reference[27]. Thus, it is apparent that most of the values depend in an absolute sense on the values of the solid surface energies.

12. Pseudomorphic stabilization of metastable phases in thin films

The fact that the geometric component of the interface energy in coherent type interfaces is negligible is put to use in the pseudomorphic stabilization of metastable phases. Since, the chemical component of the interface energy can be adjusted to be negative in coherent interphase interfaces we have another parameter that can be controlled in the attempt to achieve this objective. For certain combinations of substrate and thin film material it is possible to demonstrate that this system will have a lower-free energy if the thin film has a metastable crystal structure, suitably related to the crystal structure and parameters of the substrate, than if it had its stable crystal structure. For example, alpha-Sn is metastable with respect to the metallic beta-Sn above about 13.2°C. Alpha-Sn has the diamond cubic structure with lattice parameter equal to 6.489 Å at 25°C. A candidate substrate material with the same crystal structure and lattice parameter (6.4798 Å) is zinc-blend type InSb. The beta-Sn has a tetragonal crystal structure with lattice parameters of $a = 5.8311$ Å and $c = 3.1817$ Å. Thus, an appreciable elastic distortion is required for the beta-Sn lattice to match coherently with that of InSb. This elastic distortion has an associated energy which adds to the free energy of the substrate/beta-Sn film system an amount proportional to the thickness of the film. When this total distortional energy per unit area of the film exceeds the interface energy associated with the non-coherency between the beta-Sn film and the substrate then there will be a driving force for the film/substrate interface to generate misfit dislocations and non-coherency. On the other hand, the InSb substrate/alpha-Sn film system has no strain energy and the interface energy is nil or negative. The only positive contribution to the free energy *relative to the InSb/beta-Sn system* is the lattice stability energy (i.e. the difference in free energy between the alpha and beta polymorphs of Sn), which is also proportional to the thickness of the film.

These considerations lead to the possibility of stabilizing the metastable structure as described schematically in Figure 4.18. In this figure we have assumed that the free energy comparison is between a stable β phase having a non-coherent interface with the substrate and a metastable α phase having a coherent interface with the substrate. Thus, the sum of the interface and surface specific energies is larger for the stable phase than for the metastable phase as indicated in Figure 4.18. Hence, there is a range of film thickness over which the coherent α phase is stable with respect to the non-coherent β phase. Let us also consider the free energy comparison for the case that the interface with the substrate is coherent for both phases. In this case, if the strain energy per unit volume induced in the stable phase is less than the lattice stability energy, the strained, but coherent stable phase will be stable with respect to the coherent metastable phase. On the other hand, when the strain energy per unit volume exceeds the corresponding lattice stability energy then, again, the coherent metastable phase will have a range of stability up to the same thickness defined in Figure 4.18. These are simplified concepts. A more sophisticated approach has been undertaken by Bruinsma et al.[28] Pseudomorphic stabilization of metastable phases has been observed in multilayer thin films.[29]

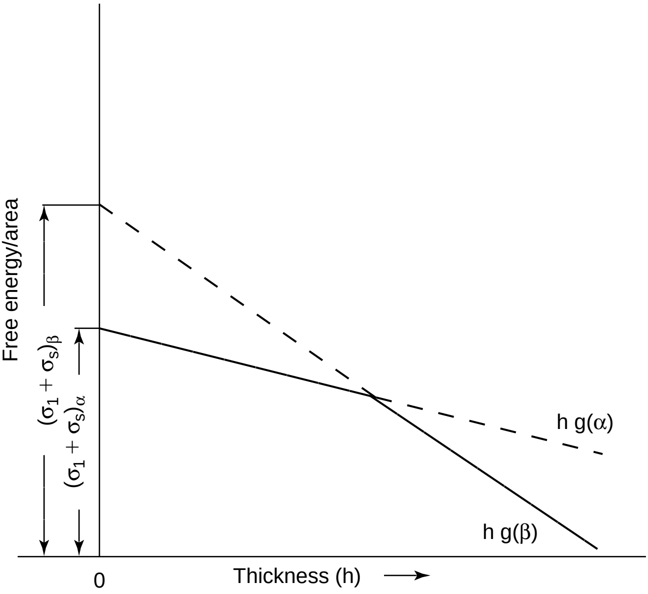


Figure 4.18. Illustrating the relative free energies of coherent metastable α and non-coherent stable β phases as a function of film thickness. Subscripts i and S represent interface and surface, respectively.

13. Wetting transition

Consider a grain boundary in a phase α that intersects another phase β. Local equilibrium, as noted in Figure 4.14(b) requires the dihedral angle θ to obey

$$\cos(\theta/2) = \sigma_{\alpha\alpha}/2\sigma_{\alpha\beta}$$

It may be possible for the denominator to become less than the numerator in this equation allowing for θ to equal zero (it can never become less than zero). If this function is temperature dependent in the two-phase region where a mixture of both phases are stable and it increases with increasing temperature then the temperature at which θ becomes zero corresponds to what is called a wetting transition temperature and a wetting transition is said to take place at this temperature.

Cahn[30] noted that as the bulk critical point of the liquid–gas transition is approached along the coexistence line that, for a liquid on a solid substrate, the difference between the solid–liquid and solid–gas interface energies and the liquid–gas interface energy will approach zero. Several models suggest that the difference approaches zero faster than the liquid–gas interface energy and, as implied by the equation corresponding to the equilibrium illustrated in Figure 4.14(a), the cosine of the contact angle will diverge at the critical point. Thus, a wetting transition is indicated at a temperature below the critical point. It has already been shown that in a binary system where the interface between phases is diffuse the interface energy approaches zero as the critical temperature is approached from below (see Section 10 of this chapter, last equation).

Another phenomenon, pre-wetting, can occur, which is not related to critical point behavior. Consider, the case of a single component system where both the solid–solid (grain boundary) and solid–liquid interface energies are temperature dependent so that $2\sigma_{SL}$ becomes smaller than σ_{GB} at a temperature T_W, which is less than the melting point temperature, T_M, as the temperature increases. In this case, the liquid phase should wet the grain boundary at temperatures between T_W and T_M. Examples of such pre-wetting in elemental solids are not known to this writer. However, from the published data for interface energies and their temperature dependence, one might expect pre-wetting of high-energy grain boundaries. For example, copper is reported to have a value of $2\sigma_{SL}$ equal to 354 erg/cm^2 and a value of σ_{GB} equal to 625 erg/cm^2 at a temperature of 925°C with a value of $d\sigma/dT$ equal to -0.10 erg/cm^2/°C. Hence, this high-angle grain boundary of copper should have liquid copper as a layer along the grain boundary from at least 925°C up to its melting point of 1083°C, if these values of interface energy are correct. The thickness of such a layer is given by the same considerations discussed in the previous section on pseudomorphic stabilization of metastable phases. Below its bulk melting point the liquid phase is not stable. However, it can be stabilized along a grain boundary by the difference in interface energies $\sigma_{GB} - 2\sigma_{SL}$. Equating this difference to the product of the thickness of the liquid layer and the difference between the free energies of the liquid and solid phases yields the layer thickness, i.e.

$$\sigma_{GB} - 2\sigma_{SL} = \Delta S \, \Delta T \, \delta$$

where ΔS is the entropy of fusion and ΔT is the difference between the melting point temperature and the temperature of the solid. Substituting in values appropriate to copper we find, if these values are correct, that at 925°C a liquid layer of about 25 Å thickness should exist. This result suggests the possibility of another T_W, determined not by the equality of σ_{GB} and $2\sigma_{SL}$, but by the temperature below which the thickness of the liquid phase becomes less than the spacing between atomic planes. (It is suggested that experiments to verify these predictions should be based on single grain boundaries of known interface energy and on interface energies of the applicable orientation of the solid surface in contact with its liquid, i.e. use of average values, as in the calculations above, is not justified.)*

---

* The phenomena of grain boundary melting[31] and of surface melting have been observed.[32]

Another interface stabilized bulk metastable phase (ISBMP) phenomenon is that also known as the effect of particle size on the melting point temperature. In this case, it is the difference between twice the solid and liquid surface energies that is set equal to the product of ΔS ΔT and r, the particle radius in the above relation. A measurement of ΔT versus r has been carried out for tin using a highly sensitive microcalorimeter[33] from which it is possible by plotting ΔT versus reciprocal r to determine a value of $\sigma_S - \sigma_L$ for tin as a function of temperature. It was found that the interface energy difference is not temperature dependent within 80°C of the bulk melting point and equals the value determined from independent measurements of the solid and liquid surface energies, as listed in Tables 4.1 and 4.2.

Pre-wetting along grain boundaries and wetting transition temperatures have been observed in binary systems[34] when the liquid and solid phases are essentially in local equilibrium, i.e. the phenomenon does not involve ISBMP, but it does involve wetting of grain boundaries by stable liquid phases. The ISBMP phenomenon may have been observed in binary systems[35] where it is believed the solute acts to change the interface energies so as to stabilize the metastable phase at bulk compositions and temperatures at which this liquid phase is not stable.

Two-dimensional phase transitions have been observed along grain boundaries[36] and are known to occur along surfaces, where they are known also as surface reconstructions. However, if the “wetting” phenomenon, as the name implies, involves the wetting of one bulk phase by another phase, which is thicker than twice the spacing between two adjacent atomic planes, then these 2D phase transitions are not wetting transitions.

Summarizing, solid surface energies reveal minima, in the form of cusps, in the Wulff plot of surface energy versus orientation. Orientations in the vicinity of these cusp orientations tend to have surfaces comprising cusp orientation surfaces separated by ledges of low-energy orientation. The arrangement of atoms at surfaces or interfaces need not mimic the arrangement in bulk. Indeed, reconstruction of the surface or interface region to produce surface or interface phases can and does occur for many materials. Capillarity effects due to the small size of particles lead to driving forces for atom transfer from small particles to large particles. Solute tend to partition in an equilibrium manner between the homogeneous bulk and interface regions according to the Gibbs’ adsorption equation. The relative values of interface energies can and do have important consequences. Local equilibrium between interfaces lead to the shapes of interfaces found in microstructure. The gradient implicitly associated with an interface can be treated by gradient thermodynamics to yield the thickness and energy of the interface. The difference in the excess energy of interfaces provides the basis for pseudomorphic stabilization of thin metastable phases and for the ISBMP. A wetting transition requires that the contact angle involved in the wetting of one bulk phase by another to be temperature dependent and to become equal to zero at the wetting transition temperature.

14. Soft matter interfaces

The foregoing discussion, except for the general principles, applies primarily to hard materials. Soft matter, such as polymers, also have interfaces. These interfaces differ in several aspects from those for hard materials. First, the interfaces in soft matter are generally diffuse. Then there are no dangling bonds at surfaces. Further, there is usually enough mobility for the surface moities to assume their lowest-energy configuration. Thus, it is not surprising that the surface energies of soft matter lie at the low end of the range of observed values. For example, values range from 16 erg/cm^2 for polytetrafluoroethylene to about 70 erg/cm^2 for albumin. Typical solid polymers have surface energies on the order of 35 erg/cm^2. Fluid polymer surface energies depend upon molecular weight of the polymer but range below about 30 erg/cm^2 with surface energy decreasing with increasing molecular weight. Interface energies between block polymers can approach zero. Liquid crystals also have low values of surface energy. A typical value is 14 erg/cm^2. Surface energy values for specific polymers can be found in the **Polymer Handbook**. Reference 37 reviews the theoretical contributions concerning the surface energy of polymers.

For proteins one of the more interesting interfaces is that involved in the "lock and key" matching of proteins. Here there is both a topological aspect as well as a complementary bonding aspect of the interface between the two entities that affects the free energy of their bonded configuration.[38]

References

1. J. Friedel, Acta Met. 1, 79(1953).
2. J.J. Gilman, J. Appl. Phys. 31, 2208(1960).
3. C. Herring, Phys. Rev. 82, 87(1949).
4. J.K. Mackenzie, J. Phys. Chem. Solids 23, 185(1962).
5. R. Defay, I. Prigogine, A. Bellemans and D.H. Everett in **Surface Tension and Adsorption**, Longman's, Green & Co., London, 1966.
6. R. Trivedi in **Lectures on the Theory of Phase Transformations**, ed. H.I. Aaronson, AIME, Warrendale, PA, 1975.
7. E.A. Guggenheim, **Thermodynamics**, North-Holland Publishing, Amsterdam, 1967.
8. E.D. Hondros and M.P. Seah in **Physical Metallurgy**, eds. R.W. Cahn and P. Haasen, 3rd edition, North-Holland Physics Publishing, NY, 1983.
9. P. Lejcek and S. Hoffman, CRC Cr. Rev. Sol. State & Mat. Sci. 20, 1(1995).
10. D. Seidman, Ann. Rev. Mater. Res. 32, 235(2002).
11. A.L. Shull and F. Spaepen, J. Appl. Phys. 80, 6243(1996).
12. P. Gumbsch and M.S. Daw, Phys. Rev. B 44, 3934(1991).
13. D.P. Woodruff, **The Solid–Liquid Interface**, Cambridge University Press, London, 1973.
14. J.W. Cahn in **Interfacial Segregation**, ASM, Metals Park, Ohio, 1979.

15. J. Wetzel and E.S. Machlin, Surf. Sci. 144, 124(1984).
16. D. Josell and F. Spaepen, Acta Metall. Mater. 41, 3017(1993).
17. J.H. van der Merwe, Proc. Phys. Soc. 63A, 613(1950); in **Treatise on Materials Science and Technology**, Vol. 2, Academic, New York, 1973.
18. J.W. Cahn and J.E. Hilliard, J. Chem. Phys. 28, 258(1958).
19. Y.W. Lee and H.I. Aaronson, Acta Met. 28, 539(1980).
20. H.I. Aaronson and F.K. LeGoues, Met. Trans. 23A, 1915(1992).
21. I. Servi and D. Turnbull, Acta Met. 14(161), 908(1966).
22. F.K. LeGoues and H.I. Aaronson, Acta Met. 32, 1855(1984).
23. M. Asta, MRS Symp. Proc. 398, 281(1996).
24. L.E. Murr, **Interfacial Phenomena in Metals and Alloys**, Addison-Wesley, Reading, MA, 1975.
25. J.J. Gilman, J. Appl. Phys. 31, 2208(1960).
26. W.H. Class and E.S. Machlin, Am. Cer. Soc. J. 49, 306(1966).
27. M. Gunduz and J.D. Hunt, Acta Met. 33, 1651(1985).
28. R. Bruinsma and A. Zangwill, J. de Physique. 47, 2055(1986).
29. S.A. Dregia, R. Banarjee and H.L. Fraser, MRS Symp. Proc. 481, 545(1998).
30. J. Cahn, J. Chem. Phys. 66, 3667(1977).
31. A.M. Alsayed, M.F. Islam, J. Zhang, P.J. Collings, A.G. Yodh, Science 309, 1207(2005).
32. Z.L. Wang et al., J. Phys. Chem. B. 102, 6145(1998).
33. S.L. Lai, P. Infante, G. Ramanath and L.H. Allen, MRS Symp. Proc. 398, 469(1996).
34. B. Straumal, T. Muschik, W. Gust and B. Predel, Acta Metall. Mater. 40, 939(1992). Also, see references cited in present reference.
35. L.-S. Chang, E. Rabkin, B.B. Straumal, S. Hofmann, B. Baretzky and W. Gust, Proceeding of International Workshop on Surface and Grain-Boundary Segregation, Balatonfoldvar, 1997, p. 97.
36. A. Bourret, J. Phys. France 51, 1(1990); W. Krakow, Acta Metall. Mater. 38, 1031 (1990); W. Krakow, Phil. Mag. A. 63, 2(1991).
37. G.T. Dee and B.B. Sauer, Adv. Phys. 47(2), 161(1998).
38. J-P Behr, **The Lock-and-Key Principle**, Vol. 1, The State of the Art – 100 Years On. J. Wiley & Sons, New York, 1995.

Bibliography

1. J.W. Gibbs', **The Scientific Papers of J. Willard Gibbs'**, Vol. 1, Dover, New York, 1961.
2. C. Herring in **Structure and Properties of Solid Surfaces**, R. Gomer and C.S. Smith, eds. University of Chicago Press, Chicago, 1953.
3. W.W. Mullins in **Metal Surfaces**, ASM, Metals Park, Ohio, 1963.
4. **Grain Boundary Structure and Properties**, eds. G.A. Chadwick and D.A. Smith, Academic Press, New York, 1976.
5. **Grain-Boundary Structure and Kinetics**, ed. R.W. Balluffi, ASM, Metals Park, Ohio, 1980.
6. C. Herring in **Physics of Powder Metallurgy**, ed. W.E. Kingston, McGraw-Hill, New York, 1951.

7. G.A. Somorjai, **Chemistry in Two Dimensions: Surfaces**, Cornell University Press, Ithaca, 1981.
8. A.P. Sutton and R.W. Balluffi, **Interfaces in Crystalline Materials**, Oxford University Press, Oxford, 1995.

Problems

1. On the assumption of unreconstructed surfaces, evaluate the relative surface energy for silicon (diamond cubic) of the 100, 110 and 111 surfaces. Take into account that the number of dangling bonds can vary depending between which of the possible parallel planes the surface is formed.
2. Why, at equilibrium, is it possible to form faceted surfaces on solids even though this corresponds to an increase in surface area per unit volume relative to a smoothly curved surface?
3. Why do surface torques exist near low-index surface orientations that tend to rotate the surface to another orientation?
4. Describe the dependence of equilibrium vapor pressure on the radius of the particles in equilibrium with the vapor?
5. What is the driving force for Ostwald Ripening?
6. In what way does the concentration at a two-phase boundary in the phase diagram depend upon the size of one of the phases?
7. If increase in solute concentration acts to decrease the surface tension will the solute segregate to the surface or desegregate from the surface over the range of concentration within the spinodal?
8. If increase in the solute concentration acts to decrease the average of the solidus and liquidus temperatures relative to the concentration weighted average of the melting points of the pure components will the solute segregate or desegregate from interfaces in this dilute alloy?
9. If the surface of an elemental solid is considered to be a separate phase will it melt at a lower or higher temperature than the bulk phase? Justify your answer. Is it valid to consider the surface to be a separate phase in this context?
10. Will surface stress and surface tension differ at a temperature more than about 0.5 the absolute melting temperature?
11. What is the relation between the surface tensions of the interfaces at local equilibrium for a drop of liquid on a flat solid surface, both in contact with a vapor phase.
12. In a 2D thin polycrystalline film, what shape will the grains assume to produce a state of metastable equilibrium? (That is any small deviation of the shape of the grains will increase the free energy.) *Hint*: Consider the grain boundary energy to be independent of orientation. In this case, what inclusive angles do three grain boundaries that intersect along a line define?
13. What group of elements in the periodic table will tend to segregate at the surfaces of metallic solids? *Hint*: Consider the physical basis for the excess energy associated with the surface to provide a clue to one group of such elements.
14. A solder must wet the parts it joins. In order to achieve such wetting (as small a contact angle as possible) the surface energy of the solder is usually much smaller than that of the parts it joins. Why?

15. Suppose it is desired to prevent phase separation in a thin film at a composition of the film where a two-phase mixture is stable at the deposition temperature. Not only is it desired to prevent phase stabilization, but it is also desired to stabilize the solid solution at a particular composition in the two-phase region. How would you use pseudomorphic stabilization to accomplish this objective?
16. Will particles of bismuth, smaller than 10^{-5}cm and at equilibrium with respect to change of phase, have the same crystal structure as bulk bismuth? *Hint*: See Figure 1.3 and Table 4.2 for the data required to solve the problem.
17. Estimate the grain boundary solute segregation ratio for the case of Ga as a solute in polysilicon. Describe the steps you took to obtain this estimate.
18. Derive an expression for the thickness of the liquid phase that may exist along a high-angle grain boundary at a temperature difference ΔT below the bulk melting point. Estimate this thickness for copper using data listed in tables in this chapter, for $\Delta T = 10°C$. Comment on the possibility that this effect may account for the frequent observation of an amorphous phase along grain boundaries in polycrystalline ceramics.
19. Evaluate $\Delta f(C)$ in the expression for the energy of a diffuse interface in terms of the composition and temperature under the assumption that the parent metastable solid solution is a regular solution.

CHAPTER V

Heterophase and Homophase Fluctuations

Introduction

The bases for consideration of fluctuations of various kinds in otherwise homogeneous phases have been developed in the previous chapters. In particular, heterophase fluctuations involve interfaces between different phases, as well as changes in the free energy of a system. Homophase or, more accurately, homostructural fluctuations are fluctuations in composition and/or density or strain. At equilibrium, there are distributions of both such fluctuations, called embryos. Nucleation of product phases from a metastable parent phase may occur by spontaneous growth of critically sized embryos belonging to an equilibrium distribution of embryos when the product phase is thermodynamically a more stable phase.

In the homophase system we can define a locus of spinodal compositions. A change of temperature from above to below the spinodal leads to instability of the host phase and the onset of composition fluctuations. In this regime equilibrium distributions of embryos cannot exist in the metastable host. Composition fluctuations develop to produce periodic waves of composition, one wavelength of which grows in amplitude to result in spinodal decomposition of the parent solution. The locus of spinodal compositions lies within the miscibility gap. Analogous to the spontaneous reaction along a composition path yielding “spinodal decomposition” there is one along a strain path that can decompose a host phase into periodic strain waves at constant composition. The onset of phonon softening with change in temperature makes the latter decomposition a realistic possibility. These homophase fluctuations will be considered after first considering heterophase fluctuations, which occur within a stable phase or in a metastable phase.

1. Heterophase fluctuations

In Chapter I we defined a phase, at equilibrium, to be a homogeneous material, whether a solid, liquid or vapor. However, this statement needs to be modified in that we need to state the dimension of the phase under discussion. If the dimension is such that the volume of the phase contains more than about several thousand atoms or molecules then the concept of homogeneous needs no modification. If the dimension is smaller than this scale then deviations from homogeneity due to the temperature induced motions of the atoms can occur in density, composition or local arrangement of the atoms such that these intensive parameters can vary from one such volume of the phase to another such volume, or in the course of time within a given sub-microscopic volume. Further, these temperature induced deviations from homogeneity, called fluctuations, occur while the phase is in its thermodynamic state of metastable or stable equilibrium. We will call fluctuations that yield atomic arrangements having a symmetry differing from the parent phase – heterophase fluctuations.

Consider a homogeneous host phase, alpha, containing N atoms. Suppose that the free energy of the system is increased by δg_n, when a particular small volume containing n atoms, where $n << N$, fluctuates to a configuration characteristic of a new phase, beta. However, since this beta embryo could be formed at any of the N sites, the free energy of the system of embryos and host phase must include the contribution of the entropy of mixing of the embryos among the N possible sites. It will be shown that the contribution of this entropy of mixing heterophase embryos among the "boxes" centered about the host sites initially reduces the total free energy of the system. In this development we follow the lead of Frenkel.*

Consider the free energy change in the host system due to these heterophase fluctuations

$$\Delta G_{\text{system}} = N_n \delta g_n - Tk \ln W \tag{5.1}$$

where N_n is the number of embryos of the phase, beta, which are distributed at random in the host alpha phase, each embryo containing n atoms; T is the absolute temperature; k is Boltzmann's constant; and W is the number of independent configurations of embryos and host atoms, each configuration having the same energy (see equation 1.2b).

Now, if the embryo has a spherical shape

$$W = N!/[N_n!(N - N_n)!] \tag{5.2}$$

To obtain the stable state of the system consisting of these embryos of beta phase distributed in the host alpha phase we minimize the free energy of the system with respect to the number of such embryos as follows

$$\frac{\partial G^{\text{system}}}{\partial N_n} = \delta g_n - kT \frac{\partial \ln W}{\partial N_n} = 0 \tag{5.3}$$

---

* See Bibliography. Frenkel was the first to use the concept of heterophase fluctuation.

But, by Stirling's approximation $N! \cong N \ln N - N$ for large N. Hence, substituting and taking the derivative we obtain

$$\frac{\partial G^{\text{system}}}{\partial N_n} = \delta g_n + kT \ln[N_n/(N - N_n)] = 0 \tag{5.4}$$

We now perform the second derivative to determine whether equation 5.4 represents a minimum or a maximum and obtain

$$\frac{\partial^2 G^{\text{system}}}{\partial N_n^2} = kT[(1/N_n) - 1/(N - N_n)] \tag{5.5}$$

Since $N_n < < N$, the right-hand side of equation equation 5.5 is positive definite and hence equation 5.4 represents a minimum in the free energy of the system of embryos and host. Solving equation 5.4 for N_n we obtain then that the number of embryos of phase beta containing n atoms at equilibrium is

$$N_n = N \exp[-\delta g_n/kT] \tag{5.6}$$

The dependence of the free energy of the system of embryos and host as a function of the number of embryos is shown in Figure 5.1. Thus, it is apparent that fluctuations initially act to decrease the free energy. To determine how many embryos exist at equilibrium it will be necessary to obtain an estimate for the free energy of formation of a particular embryo δg_n. Consider a heterophase embryo; it will have an interface between it and the surrounding host phase, alpha, which has an associated specific interface free energy $\sigma_{\alpha\beta}$. The beta phase will have a specific free energy (per unit volume) g_β, different from that of the host phase g_α. Suppose, for the sake of simplicity, the embryo has a spherical shape of radius r. In this case, the increase in free energy in the formation of one embryo in a specific volume of the host phase is

$$\delta g_n = (4/3)\pi r^3 (g_\beta - g_\alpha) + 4\pi r^2 \sigma_{\alpha\beta} \tag{5.7}$$

Here, we assume that there is no change in volume in the transformation.

If the host alpha is the stable phase at this temperature then δg_n has a positive value for all values of r (of n). If, on the other hand, alpha is a supersaturated or supercooled metastable phase relative to beta then δg_n will depend upon r as shown in Figure 5.2. It exhibits a maximum positive value at the radius r*. This maximum

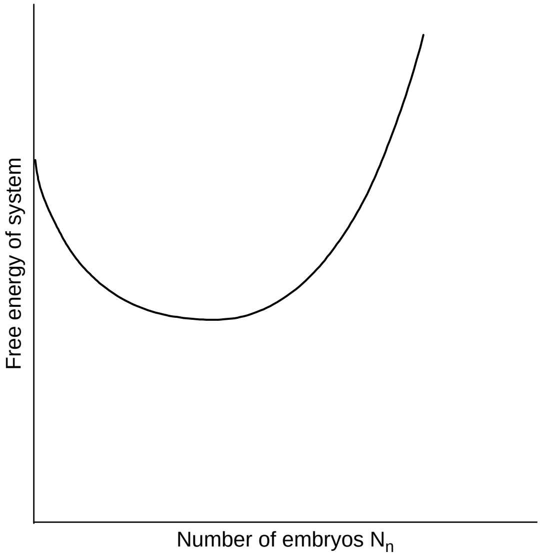


Figure 5.1. Dependence of free energy of system on number of embryos, each containing n atoms.

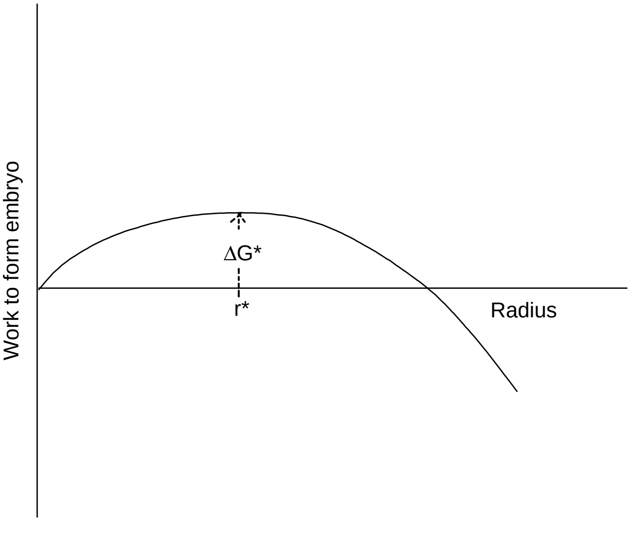


Figure 5.2. Dependence of the work to form an embryo on the radius of the embryo. At r* this work reaches a maximum and beyond this radius the embryo can grow spontaneously.

value is called the free energy of nucleation and is denoted by the symbol ΔG^*. For a spherically shaped embryo, differentiation of equation 5.7 with respect to r and appropriate rearrangement of terms yields that

$$\Delta G^* = 16\pi\sigma_{\alpha\beta}^3/[3(g_\beta - g_\alpha)^2] \tag{5.8}$$

and
$$r^* = -2\sigma_{\alpha\beta}/(g_\beta - g_\alpha) \tag{5.9}$$

For any other shape of the embryo defined by the shape factor Q, where the surface area of the embryo is equal to $Q\Omega^{2/3}$ and Ω is the volume of the embryo, the free energy of nucleation is given by

$$\Delta G^* = 4Q^3\sigma_{\alpha\beta}^3/[27(g_\beta - g_\alpha)^2] \tag{5.10}$$

with
$$\Omega^* = [-2Q\sigma_{\alpha\beta}/\{3(g_\beta - g_\alpha)\}]^3 \tag{5.11}$$

The reason why the critically sized embryo is called a nucleus follows from the fact that as the nucleus grows the free energy of the system decreases. Since the critically sized embryo can spontaneously grow into a particle of the stable product phase it is the nucleus of the growing particle. However, for any embryo smaller than the nucleus the free energy of the system must initially increase if it is to grow.

The classical concept of nucleation is based on the assumption that fluctuations develop in a metastable host and maintain an equilibrium distribution of sub-critically sized heterophase embryos. As nucleation drains off critically sized embryos, new ones are produced at a rate sufficiently fast as to maintain the equilibrium distribution. We shall consider the rate of nucleation and embryo formation in a later section.

It is instructive to insert values for the quantities that enter into the free energy of nucleation (or to form an embryo) to obtain a sense of the likelihood that such nuclei (or embryos) can exist in a host phase. If the case being considered is the formation of a solid from a liquid host phase, then we may use Richard's approximation to estimate $g_\beta - g_\alpha$ in the vicinity of the equilibrium melting temperature T_m, as follows

$$g_\beta - g_\alpha = (h_\beta - h_\alpha) - T(S_\beta - S_\alpha) \tag{5.12}$$

However, at T_m, the difference in free energy corresponding to the left-hand side of this equation must equal zero. Hence

$$h_\beta - h_\alpha = T_m(S_\beta - S_\alpha) \tag{5.13}$$

Substituting equation 5.13 into equation 5.12 yields

$$g_\beta - g_\alpha = (T_m - T)(S_\beta - S_\alpha) \tag{5.14}$$

By Richard's rule $(S_\beta - S_\alpha) = 2\,\text{cal/mol/°C}$. Suppose we take $|T_m - T|$ equal to 10°C. Also, from the table in the previous chapter, we obtain an estimate of a typical solid–liquid interface energy to be about $100\,\text{erg/cm}^2$. Substituting these values, using a typical molar volume of about $10\,\text{cm}^3/\text{mol}$, $N \approx 10^{23}$ and the appropriate conversion factors to achieve the same units we then obtain the following values

$$r^* = 2.39 * 10^{-6}\,\text{cm} = 23.9\,\text{nm}$$

$$\text{and } N^* = 10^{-7509}/\text{cm}^3$$

This number of embryos is much smaller than one in a host volume of about $1\,\text{cm}^3$ and hence quite unlikely to be observed. Obviously, for the circumstance being considered – that of a homogeneous distribution of embryos in the host phase – the only chance of observing embryos is when the barrier energy provided by the interface energy is very small or when the driving energy (the supercooling or supersaturation) is very large and $g_\beta - g_\alpha < 0$. Is the existence of heterophase embryos thus merely an academic question? We shall show in the next section that the existence of embryos at interfaces that act to catalyze their nucleation is highly probable.

1.1. Heterogeneous distributions of heterophase fluctuations

We have considered homogeneous distributions of heterophase embryos and nuclei in the previous section. In this one we shall show that the presence of interfaces in a host phase, which the embryo phase tends to "wet", can act to catalyze the presence of heterophase embryos and nuclei. We shall investigate two cases. In the first, the interface is between a material that does not interact chemically either with the host phase or with the embryo phase. In the second, the interface is a grain boundary in the host phase. Consider the situation shown in Figure 5.3. Local interface equilibrium in the absence of "torque" contributions and chemical interactions

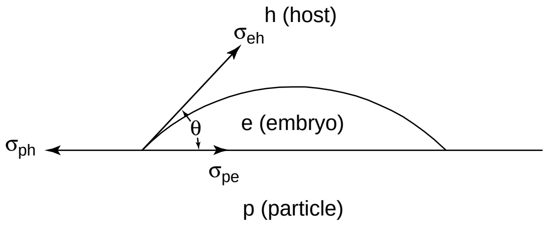


Figure 5.3. Illustrating an embryo formed at a particle (p)/host (h) interface and the local equilibrium at the host–particle–embryo intersection.

between the particle phase p, and either the host or embryo phases yields the following relation between the interface energies.

$$\sigma_{ph} = \sigma_{pe} + \sigma_{eh}\cos\theta \tag{5.15}$$

Using this relation between the specific interface energies the change in total interface energy due to the formation of one embryo at the original host–particle interface can be written as

$$\Delta G_s = 4\pi r^2 \sigma_{eh} f(\theta) \tag{5.16}$$

where $f(\theta) = (2 - 3\cos\theta + \cos^3\theta)/4$. It can be shown that the change in free energy in the volume occupied by the embryo is given by

$$\Delta G_b = (4\pi r^3/3)(g_e - g_h) f(\theta) \tag{5.17}$$

Now, the total change in free energy due to the formation of the embryo is the sum of ΔG_s and ΔG_b. To obtain the free energy of nucleation, we set the derivative of this sum with respect to the radius equal to zero with the result that

$$\Delta G^* = (16\pi\sigma_{eh}^3/[3(g_e - g_h)^2]) f(\theta) \tag{5.18}$$

and

$$r^* = -2\sigma_{eh}/(g_e - g_h) \tag{5.19}$$

Thus, the radius of the spherical cap of the nucleus is unaffected by nucleation on the substrate. However, the free energy of nucleation can be markedly affected depending upon the value of the function $f(\theta)$. As the embryo phase more effectively wets the

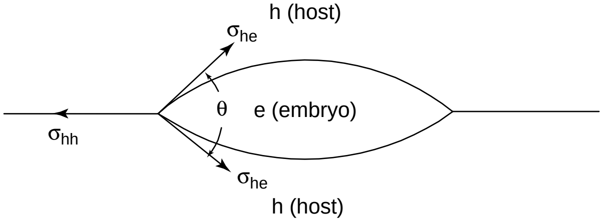


Figure 5.4. Embryo shown formed along a grain boundary with local equilibrium of the interface tensions.

substrate, the contact angle θ approaches zero, $f(\theta) \to 0$, and the free energy of nucleation $\Delta G^* \to 0$. Thus, depending upon the contact angle θ, heterophase fluctuations and nucleation at inert interfaces in the host phase may or may not occur readily.

Heterogeneous nucleation may also be produced at defects in metastable solids, such as dislocations, grain boundaries or point defects. Consider the case of grain boundaries as sites for the formation of heterophase fluctuations. Figure 5.4 defines the geometry of the situation. Local equilibrium at interface intersections requires that

$$\sigma_{hh} = \sigma_{gb} = 2\sigma_{eh}\cos(\theta/2) \tag{5.20}$$

Under this constraint it can be shown that the free energy change associated with the interfaces upon formation of the embryo is

$$\Delta G_s = 4\pi r^2 \sigma_{eh} f'(\theta) \tag{5.21}$$

where $f'(\theta) = f(\theta)/2$. Also, the free energy change in the volume occupied by the embryo is

$$\Delta G_b = [4\pi r^3 (g_e - g_h)/3] f'(\theta) \tag{5.22}$$

Again, the total free energy change due to the formation of one embryo is the sum of ΔG_s and ΔG_b. The free energy of nucleation is obtained as the maximum of this sum with respect to variation in the radius of the embryo's spherical cap r, and is given by equation 5.18 with $f(\theta)$ replaced by $f'(\theta)$. Thus, for embryo phases that wet grain boundaries in metastable host phases (i.e. for which $f'(\theta) \to 0$), the free energy to form such embryos can be sufficiently small as to make the appearance of heterophase fluctuations and nucleation of the stable phase at these grain boundaries highly probable.

Technologically, the result we have just obtained has great significance. For example, inert particles that are wetted by the product stable phase are used to seed supersaturated clouds in order to produce rain, to catalyze the production of fine-grained castings in supercooled metallic melts, etc.; wetted inert interfaces are the location at which nucleation of boiling occurs in superheated liquids; grain boundaries are one of the sites of nucleation of solid state transformations. Indeed, in practice, heterogeneous nucleation is the rule.

1.2. Effect of stress on fluctuation probability and embryo shape

If the parent and product phases are solids then an additional contribution to the increase in free energy due to the formation of one embryo arises either because the specific volumes of the two phases differ or because in the presence of a coherent interface between the host and embryo phase the lattice parameters of the two phases parallel to the interface differ. In the case of an incoherent interface between embryo and parent phase, Nabarro[1] has evaluated the increase in free energy due to the strain energy induced in both embryo and host. He has also evaluated the effect of shape of the embryo on this increase in free energy on the assumption that all the strain energy is borne by the host phase with the result that the strain energy per atom of embryo is

$$\Delta g_e = [2\mu_h(\Omega_e - \Omega_h)^2/3\Omega_e]E(y/r) \tag{5.23}$$

where E(y/r) is a function of the ratio of the half thickness y of a spheroid shaped embryo to its radius r and hence of the shape of the embryo, μ_h is the shear modulus of the host and Ω is the atomic volume. This function is illustrated in Figure 5.5. From the dependence of E(y/R) on y/R shown in this figure it is apparent that the minimum strain energy occurs when the embryo has the shape of a disc. However, the disc shape has a higher surface energy for the same volume as compared to a sphere, which, as shown, is associated with a higher strain energy. Thus, the shape adjusts to some compromise between a disc and a sphere to minimize the total overall increase in free energy on formation of the embryo.

If the parent and product phases are such as to accommodate a coherent interface between the two, then the possibility exists that for small embryos the total free energy of formation of an embryo having coherent interfaces with the parent phase will be less than that for the case where these interfaces are incoherent.* Although there is an enhanced strain energy for the embryo with coherent interface relative to the

* Guinier–Preston zones are typical embryos having coherent interfaces with the parent phase. See the Appendix for the factors controlling the equilibrium configuration of such coherent embryos.

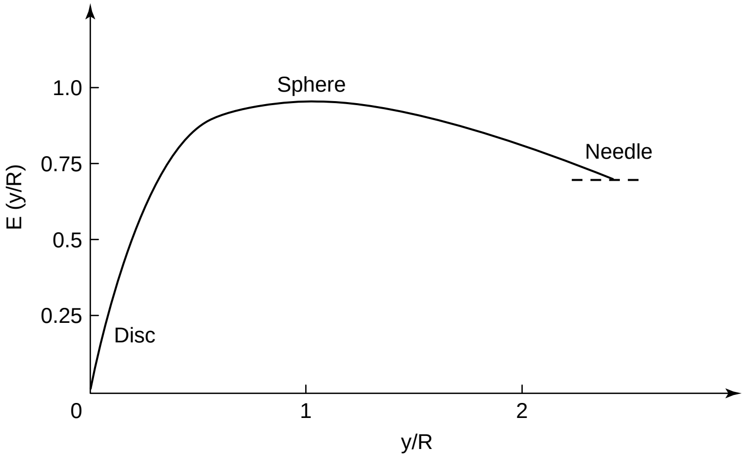


Figure 5.5. The function E(y/R) as calculated by Nabarro.[1]

one with non-coherent interfaces, the decrease in surface free energy due to the lower energy of the coherent interface (25 erg/cm^2) as compared to a non-coherent interface ($>$200 erg/cm^2) is more significant than the increase in strain energy, at least for sufficiently small embryos. This concept is illustrated in Figure 5.6. These same concepts hold also in describing the transition between a coherent interface and one containing an array of misfit dislocations for a thin film on an epitaxial substrate, i.e. the coherent interface is stable until the film becomes thicker than a critical value.

For the simple case of equal elastic moduli of host and embryo, similar atomic arrangements in both, and disregistry along only one atomic direction, the strain energy *per unit volume* of embryo is independent of the shape and is given approximately by

$$\Delta g_e = \mu\delta^2/(1 - \nu)^* \qquad (5.24)$$

where the relative disregistry in atomic spacing parallel to the coherent interface is

$$\delta = (a_h - a_e)/a_e \qquad (5.25)$$

The morphology of coherent embryos is not readily investigated using analytical methods and this accounts for the numerically based attempts to shed light on this subject. One such computational investigation was based on the Monte Carlo site exchange process to achieve an approach to equilibrium. It was

---

* ν is Poisson's ratio.

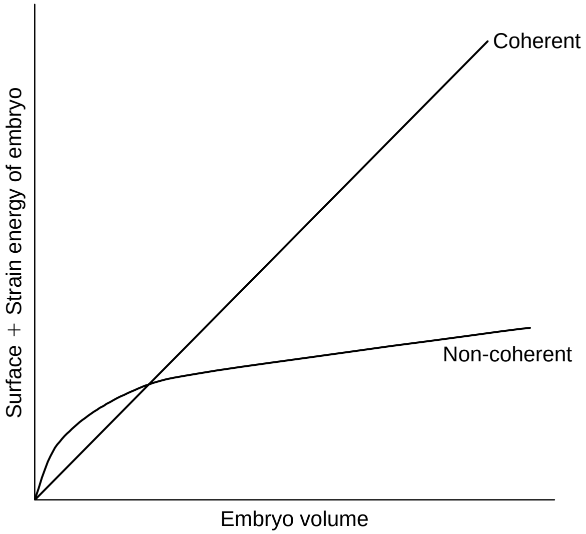


Figure 5.6. Showing that for small volumes coherent embryos have lower free energy per embryo of given volume than non-coherent ones.

found[2] that with a purely dilatational transformation (misfit) strain, a soft embryo tends to have a plate-like equilibrium shape, whereas a hard embryo takes up a shape of high symmetry such as a circle.* With either a tetragonal misfit strain of mixed signs or a pure shear misfit strain, an embryo takes up a plate-like shape whose major axis lies along the direction containing an invariant line in accordance with the continuum elasticity prediction. In the investigations cited above for coherent interfaces, the latter were assumed not to be diffuse, but discontinuous.

The concepts just described hold as well for heterogeneous type nucleation. For example, at grain boundaries, it is possible for the embryo to have a coherent interface with at least one of the contiguous grains. Also, considering the plenitude of low energy boundaries at special orientations, it is conceivable that the interface with the other grain, if not a coherent boundary is, at least, a low energy boundary.

The existence of coherent interfaces between host and product phases implies that there is a unique crystallographic relation between host and product phases. However, there is another possible reason for the observance of crystallographic relations between host and product phases which is based on the kinetics of growth of the product phases. Thus, observation of unique crystallographic relations between host and product phases may not be used as a proof of the mode of nucleation.

---

* A detailed analysis of the shapes of coherent embryos based on first-principle calculations and Monte Carlo simulation is provided in C. Wolverton, Model. Simul. Mater. Sci. Eng. 8, 323(2000). A brief summary is given in the Appendix.

2. Homophase fluctuations of composition in a metastable homogeneous phase

Consider a metastable phase of total volume V corresponding to Figure 5.7 at the composition C (atom fraction) and temperature T′. We wish to know the change in free energy associated with a fluctuation in composition in some small volume v from C to C + ΔC. Let the free energy per unit volume of the phase be f. Thus, the work to form this composition fluctuation is

$$\Delta F' = f(C + \Delta C)v + f[C - \Delta C\{v/(V - v)\}](V - v) - f(C)V \qquad (5.26)$$

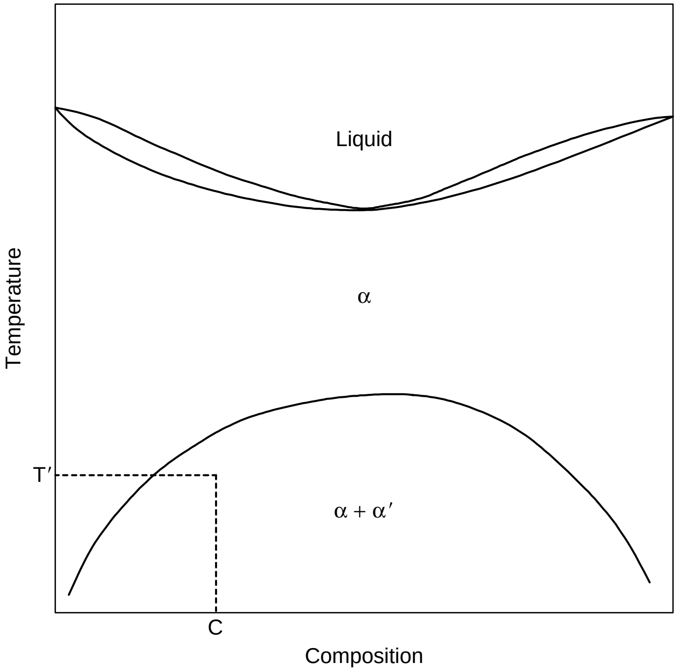


Figure 5.7. Phase diagram of miscibility gap system defining the metastable phase discussed in the text.

But, a Taylor's expansion yields

$$f(C - \Delta Cv/(V - v)) = f(C) + (\partial f/\partial C)[-\Delta Cv/(V - v)] + \cdots$$

and by substitution then

$$\Delta f'v = [f(C + \Delta C) - f(C) - (\partial f/\partial C)\Delta C]v \tag{5.27}$$

The graphical interpretation of equation 5.27 is shown in Figure 5.8. Another way to arrive at the same result is to consider the change in free energy in terms of the chemical potentials of the components. It can be shown then that

$$\Delta f'v = n[(1 - X)\mu_1(X) + X\mu_2(X)] - [(1 - X)\mu_1(X_{eq}) + X\mu_2(X_{eq})]$$

where n is the total number of atoms in the volume v of the fluctuation, X is the atomic composition corresponding to the fluctuation, μ_i is the chemical potential of component i and X_{eq} is a solvus composition. It is easy to show by rearranging terms

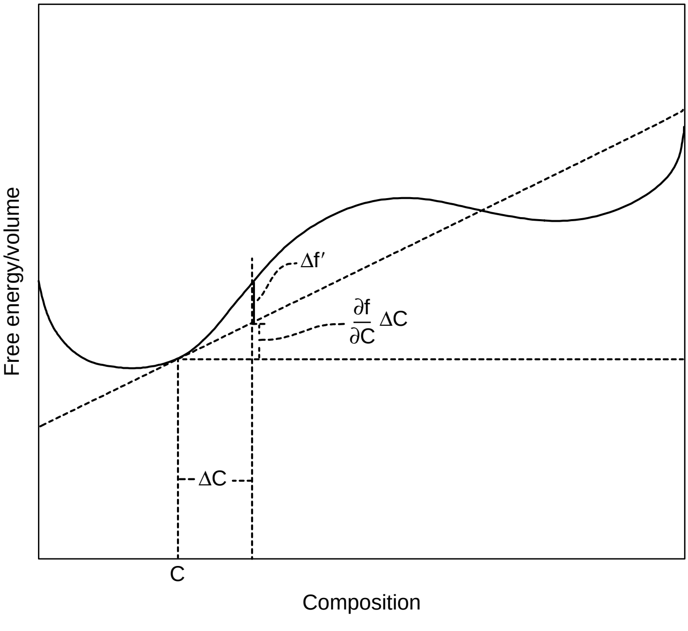


Figure 5.8. Graphical interpretation of the parameters in equation 5.27.

that this value of $\Delta f'$ equals that of equation 5.27. Thus, the work to produce a composition fluctuation in some volume, $v = n/C$, neglecting any contribution to an interface energy ranges from positive to negative when the composition of the host phase falls in the region where the free energy–composition curve has a positive curvature.

When the composition of the host phase falls in the range of compositions where the curvature of the free energy–composition curve has a negative value then ΔF has no maximum and is always negative in value. The turning point between the two cases is defined by the inflection point in the free energy–composition curve at which the second derivative

$$\partial^2 f/\partial C^2 = 0 \tag{5.28}$$

Equation 5.28 defines the conditions associated with what is called the spinodal. (The origin of the term "spinodal" is due to van der Waal. As shown in Figure 5.9, in a plot of the grand potential ($-PV$) against the chemical potential the spinodal occurs at the tip of a spine while in a plot of free energy against composition it satisfies equation 5.28!)

In the above we assumed that the concentration in the embryo of volume v was constant. A more likely condition is that the concentration varies continuously from some maximum value at the center of the embryo outward. Thus, if the matrix has a composition C_m, then equation 5.27 becomes

$$\Delta f' = f(C) - f(C_m) - (C - C_m)(\partial f/\partial C) \tag{5.29}$$

This $\Delta f'$ does not represent the total change in free energy due to the composition fluctuation because the contribution of surface energy to this change has not been included. In the previous chapter, we noted that for the situation under consideration there must be a contribution to the work to form the embryo associated with the gradient of composition. Hence, the work to form an embryo is

$$\Delta F_e = \int_V [\Delta f'(C) + K(\nabla C)^2] dV \tag{5.30}$$

where K is defined in the previous chapter (p. 146) and the composition in the embryo obeys the following equation[2] for the case of a spherically symmetric embryo

$$2K(\partial^2 C/\partial r^2) + (4K/r)(\partial C/\partial r) = \partial(\Delta f')/\partial C$$

where the boundary conditions are $\partial C/\partial r = 0$ at $r = 0$ (the center of the embryo) and $C = C_m$ at $r = \infty$. The above assumes no change in volume with changes in

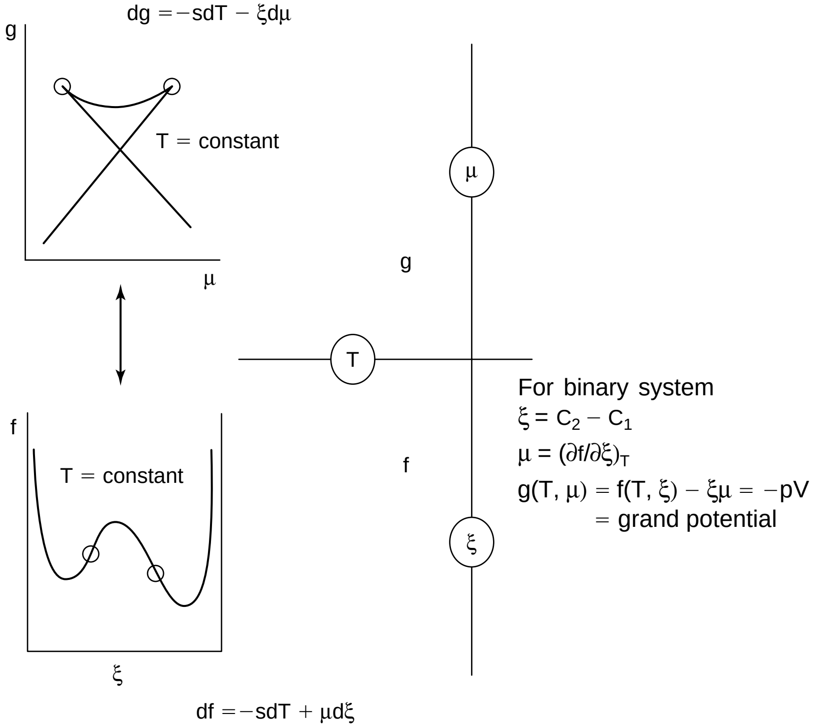


Figure 5.9. Schematic illustration of spine in plot of grand potential versus the chemical potential at position corresponding to zero value of second derivative in plot of free energy versus composition. Note that the g in this figure does not correspond to the g in the text.

composition. In the event there is a change in volume then a strain energy term must be included in the summation of terms within the bracket of the integral.

Cook et al.[3] have provided a discretized version of the integral involved in the work to form an embryo which allows a more accurate solution than does use of the continuum expression, whenever the strain energy is a significant term. Since the work of Cahn and Hilliard and Cook et al. there have been numerous experimental and theoretical studies of the subject of fluctuations in phase and composition in a metastable host. (A list of such studies up to 1992 may be found in References 4 and 5.) Most significant for the reader is the fact that experimental studies have verified and characterized fluctuations in composition in a metastable host prior to the decomposition of the host into stable precipitates distributed within a stable matrix. These fluctuations occur at compositions and temperatures in the phase diagram for which the value of $\partial^2 f/\partial C^2 > 0$, but at which the matrix is metastable.

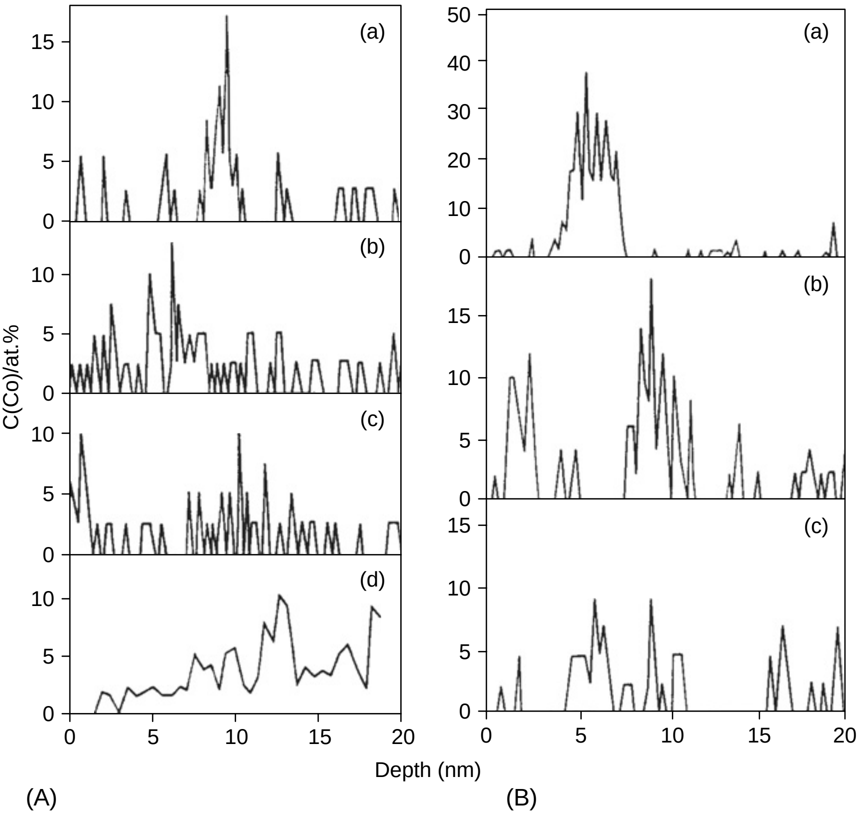


Figure 5.10. Cobalt concentration profiles as a function of depth of field evaporation in Cu–0.8 at.% Co: (A) after aging for 15 minutes at 510°C and (B) after aging 30 minutes at 510°C. From Jiang et al., Z. Metallkunde. 82, 192(1991) with permission.

Figures 5.10A and B reveal concentration versus length along a very thin field ion microscope (FIM) needle point produced from a Cu–0.8 at.% Co alloy that had been first equilibrated in the single phase field at elevated temperature and then quenched to 510°C and aged at this temperature for 15 and 30 minutes, respectively. Not shown is the result when aging is carried out for 60 minutes. In the latter case, narrow spikes representing precipitates containing more than 90 at.% Co are found and the fluctuations in composition shown in Figures 5.10A and B are absent. Figure 5.11 from a different investigation shows the composition profile obtained via FIM pulse spectrography through a precipitate in a Cu–1.0 at.% Co alloy aged for 10 minutes at 550°C after prior homogenization at 1273°K for 48 hours and step quenching into a molten salt bath at 550°C.

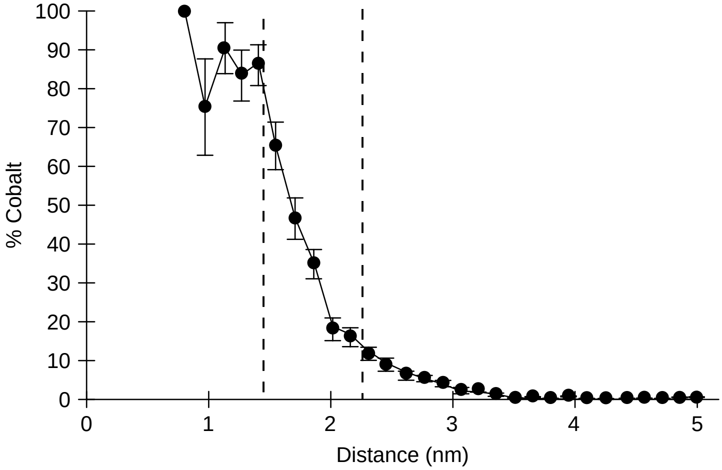


Figure 5.11. Concentration profile as a function of depth of field evaporation in Cu–1.0 at.% Co alloy after aging for 10 minutes at 550°C. From I. Rozdilsky, A. Cerezo, G.D.W. Smith and A. Watson, MRS Symp. Proc. 481, 521(1998) with permission.

A glance at the Cu–Co phase diagram shows that at these temperatures (510–550°C) the fcc solid solutions containing 0.8–1.0 at.% Co are metastable. Figures 5.10 show that prior to precipitation in these metastable hosts that composition fluctuations of the type that have been discussed and quantified by equation 5.30 do, in fact, exist. Further, we know that stable precipitates must form after sufficient time at these aging temperatures in what was originally a metastable fcc solid solution. (It is of interest to note that the precipitate corresponding to Figure 5.11 has a diffuse interface.) The question is how do these precipitates nucleate. Do they nucleate from embryos that already have the stable composition as described in Section 1 of this chapter or do they nucleate from embryos that have some composition intermediate to the stable composition and that of the metastable matrix?

Cahn and Hilliard[6] examined these questions in their classic series of papers. They predicted that close to the binodal phase boundary the critical nuclei have parameters close to the values predicted by classical nucleation theory. However, the interface becomes more and more diffuse the closer the conditions are to the spinodal conditions. *The work of nucleation instead of maintaining a finite value at the spinodal goes to zero as does the interface energy*. It would seem from Figures 5.10 and 5.11 that conditions for the 0.8 at.% Co–Cu alloy at 510°C are closer to the spinodal than they are for the 1 at.% Co–Cu alloy at 550°C for which classical nucleation occurs albeit with a diffuse interface.

3. Spinodals and their relationship to mode of metastable phase decomposition

For compositions and temperatures within the applicable coherent spinodal boundaries, the metastable parent phase is unstable. It decomposes spontaneously in a periodic "wavelike" mode to tend to produce the stable two phase mixture. Thus, there will be no equilibrium distribution of embryos, be they composition fluctuations or units of the second phase for these metastable solutions. For compositions and temperatures lying between the spinodal boundaries and the phase boundaries, the metastable parent phase can host a distribution of embryos because the work of nucleation is positive in this regime and the nature of these embryos has already been discussed in the previous section. Spontaneous spinodal decomposition has been observed not only in metals but in glasses[7] and polymers,[8] as well, and will be considered in Chapter IX.

Since the title of this section makes use of the term "homostructural", the reader may conclude that spinodal decomposition can only occur in "homostructural" phase diagrams. If so, the conclusion is wrong. The metastable parent phase will have chemical and coherent spinodal boundaries associated with it. They merely mimic the behavior of the free energy and the free energy of composition fluctuations as a function of composition and temperature in the metastable parent phase. Thus, a binary system in which stable precipitates and/or the pure solute phase have structures and symmetries that differ from those of the metastable parent phase can still exhibit spontaneous spinodal decomposition of a metastable parent phase prior the formation of nuclei of the stable product phases. (The fastest reaction will occur first in this competition of parallel processes.) For example, in the Cu–Ti system, the precipitate in equilibrium with the fcc Cu based solid solution at 350°C has an orthogonal structure at the composition containing 20 at.% Ti. When a metastable Cu–1.9 at.% Ti solid solution was annealed at a temperature where the two phase region is stable it decomposed by nucleation and growth, with the nucleus having the composition of the precipitate and a sharp interface with the matrix.[9] However, when a metastable 2.7 at.% Ti solid solution was annealed at the same temperature it first decomposed spinodally in a periodic wavelike mode,[10] with the maximum Ti composition increasing until it reached the 20 at.% level before nuclei of the stable precipitate structure appeared. The nature of the embryos in this system is unknown, although the sequence is consistent with that reported in the previous section of classical nucleation in the region where $\partial^2 f/\partial C^2 > 0$ and spinodal decomposition within the spinodal.

4. Easy embryo formers in stable host phases

In the foregoing we have discussed two types of embryos or fluctuations that appear in metastable host phases: (1) embryos having the crystal structure and composition of the stable precipitate with either a sharp or diffuse interface separating the embryo from the matrix and (2) fluctuations of composition in microscopic volumes where the composition varies from some maximum value in the center of the embryo to that of the matrix at the outer edges of the embryo. There are many other materials in which dynamic fluctuations of properties do exist in microscopic volumes which correspond to embryos of various types. Most of these fluctuations occur in just stable phases, i.e. stable phases at intensive parameters close in value to those at which a phase transition takes place. It is apparent that in this case there will be both a positive volume contribution as well as a positive surface contribution to the work of forming an embryo. For these embryos to be observed it is then necessary that both these contributions satisfy $\Delta F_\varepsilon/kT < \ln(V/v)$, where V is the volume examined by the embryo detecting apparatus, v is the volume of an embryo, ΔF_ε is the work to form one embryo and, k and T are Boltzmann's constant and temperature, respectively. Hence, only embryos with interface energy less than about 10 mJ/m^2 are likely to be observed. Among such embryo types are domain boundaries (e.g. ferromagnetic, ferroelectric, ordered). Also, if the embryo induces strain then the volume contribution to the work to form an embryo will be minimized in soft elastic hosts. Let us now consider examples of some of these various embryos.

For some materials when a phase transition temperature is approached from above, a shear elastic modulus tends to decrease markedly.[11] In this case, there are soft phonon modes associated with this elastic modulus yielding fluctuations of strain in microscopic volumes of the stable, but nearly unstable, host phase. For these fluctuations to be embryos they must somehow be related to the product phase that is more stable than the parent phase below the transition temperature. This correlation has been accomplished in a study of premartensitic phenomena in Ni_2MnGa.[12] Dynamic fluctuations of transverse TA_2 phonons in the parent cubic phase were correlated to the low temperature product structure. The nature of these phonons and their relationship to the product phase suggests the possibility that the transformation itself corresponds to a periodic array of $(220)[1\bar{1}0]$ shear displacements in the cubic parent structure producing a twinned array of tetragonal product phases. However, the parent phase does not become unstable at the transition temperature since the elastic moduli all have finite values at this temperature. It was suggested that the change in free energy as a function of the order parameter could be expressed in the expanded form of a conventional Landau model as

$$\Delta F = A\eta^2 - C\eta^4 + D\eta^6$$

where A is temperature dependent in the form $A = A'(T - T_1)$ and C and D are positive constants such that at the transition temperature T_M, $AD/C^2 = 1/4$ is

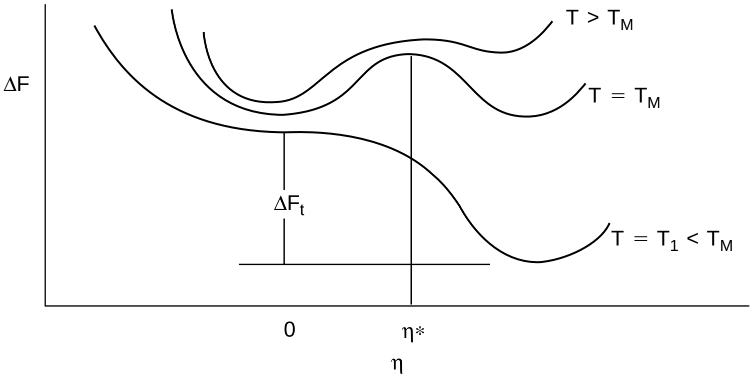


Figure 5.12. Showing the free energy change with change in the order parameter and temperature.

obeyed and also $T_1 < T_M$. This equation is illustrated graphically in Figure 5.12, $\Delta F(\eta)$ is shown as a function of η, with $\Delta F(\eta = 0) = 0$. As shown, at and below the transformation temperature, corresponding to zero and negative values of ΔF_t, the fluctuations in the order parameter in the parent phase would have to exceed η^* for the transformation to be able to proceed to the product configuration. In this regard it is interesting to note that the measured amplitude of the soft phonons increases as temperature approaches the transformation temperature from above. It is thus reasonable to expect that embryos of the product phase may exist above the transition temperature, at least dynamically. Indeed, they were found. It should be noted that the parent phase would be unstable at $T \leqslant T_1$ if somehow the transformation could be prevented on cooling to this temperature. Further, the reverse transformation, which regenerates the parent single crystal, transpires at about 8°C above T_M. The free energy at the product value of η includes the energy of interfaces and strain that form in the product and that are functions of η, and, thus, T_M is close to the equilibrium temperature between the single crystal parent phase and the product of the forward transformation.

Two types of fluctuations have been found in the colossal magnetoresistant material, $La_{1-x}Ca_xMnO_3$. One type consists of slowly fluctuating local Jahn–Teller distortions.[13] This fluctuation is associated with the tetragonal-orthorhombic phase transition found in these manganites with $x < 0.2$ and is observed in compositions with x slightly >0.2 above the Curie temperature of the ferromagnetic–paramagnetic transition. These fluctuations are embryos existing in a stable phase but of a phase that has a slightly higher free energy as evidenced by the phase transitions in this material. The other type of fluctuation consists of dynamic ferromagnetic clusters in the paramagnetic phase existing above the Curie temperature for x values between 0.2 and 0.5.[14] In the thermodynamic sense these clusters correspond to embryos of the slightly less stable ferromagnetic (metallic) state in the stable paramagnetic (insulating) host phase. It has been suggested that these two types of embryos are in fact the same physical entities, but at this writing this suggestion has not been proven. Ferroelectric embryos that have been found in the stable paraelectric phase of $K_{1-x}Li_xTa_{1-y}Nb_yO_3$:Cu,V[15] represent another example of this class of "easy embryo formers".

We have in this chapter given a simplified introduction to fluctuation theory. Actually the concepts and theory are much more sophisticated and complex. In particular, we have not provided a rigorous discussion of heterophase fluctuation theory. To obtain the benefit of rigor the reader is referred to the review article by V.I. Yukalov, Phys. Report. 208, 395(1991). This review also provides references to hundreds of articles describing observations of heterophase fluctuations. One such observation[16] not included in this review because it was made after the review was published, is that of reversible cooperative displacements of on the order of 100 atoms along grain boundaries. I believe that the phenomenon observed corresponds to a heterophase fluctuation, but that in this case the phases correspond to different 2D grain boundary structures for a grain boundary that separates two grains of having a constant orientation relation between them. As noted in Chapter IV there are many grain boundary structures possible for this situation because there are remaining degrees of freedom associated with the orientation of the grain boundary at a constant orientation relation between the crystals of the two grains. Shown in Figure 5.13 are a series of HRTEM (high-resolution transmission electron microscope) images having atomic resolution taken a fraction of a second apart in which it is apparent that cooperative motion of from 100 to 300 atoms at the grain boundary takes place and does so reversibly. In particular, a sequence starting at 15.37 seconds shows a cooperative jump of atoms from the crystal on the right to that on the left, resulting in a partial motion of the grain boundary to the right, that takes place between 15.37 and 15.73 seconds, as indicated by the images corresponding to these times from start of imaging. Further, the reverse jump takes place between 15.73 and 15.93 seconds after start of imaging. These patterns are frames from a video recording. The line in two of the images is drawn to facilitate viewing of the boundary and its partial motion. The white spots in the images represent rows of atoms. The

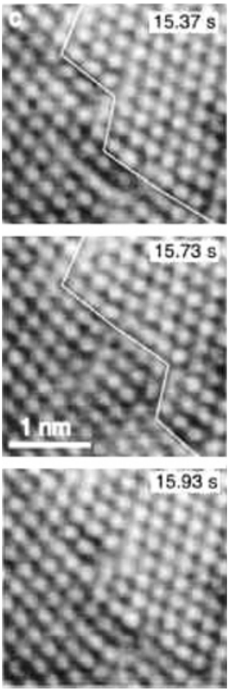


Figure 5.13. A sequence of HRTEM images at a grain boundary showing a heterophase fluctuation. From Phys. Rev. Lett. 88, 225501(2002) with permission. © 2002 Americal Physical Scocity.

sample is thin in the direction normal to the figure. What we see in these figures are the before and after results of a reversible cooperative heterophase fluctuation which occurs in less time than the intervals between these images. The blurry image in the 15.93-second image may be a consequence of motion of some of the atoms occurring during the image exposure. It is a heterophase fluctuation because the fluctuation is between two grain boundary structures (2D phases) which have nearly the same free energy. The authors of this study report many observations of such reversible fluctuations at a variety of grain boundaries.

Summarizing, in principle, fluctuations of phase in homogeneous stable or metastable hosts initially lead to a lowering of the free energy of the system before increase in their population increases the latter. However, in practice, the equilibrium concentration of heterophase fluctuations, called embryos, is so small as to be less than unity in a macroscopic specimen, except for the special cases of either a very small embryo–host interface energy ($<20\,\mathrm{erg/cm^2}$) or a very high supercooling or supersaturation of the host phase. Heterophase embryos are more likely to exist at interfaces, dislocations or other defects that lower the barrier to their formation.

It may or may not be possible to develop observable homophase concentration fluctuations in metastable solid solutions. The likelihood of producing such fluctuations increases the closer is the composition of the solid solution to the composition at which the solid solution becomes unstable. Contrary to the case of heterophase fluctuations, there exist conditions for the spontaneous generation of composition fluctuations that may grow spontaneously.

Easy forming embryos have been observed in stable phases in the vicinity of phase transitions. Such phases either have soft elastic moduli or the phase transition involves an interface energy with the host phase that is small or both.

Appendix 1

The capability of carrying out computations based on first principles has reached the point where Monte Carlo simulation based on first-principle evaluations of the energies of clusters of copper atoms in an aluminum solvent has provided the equilibrium configuration for Guinier–Preston zones reproduced in Figure A1.1. This study reveals the effect of edge energy and size on the shape of coherent precipitates. For the small precipitates in Figure A1.1a, b the controlling effect is minimization of the edge energy at the rim whereas for the larger size precipitate of the bilayer in Figure A1.1c with Cu/Al/Al/Al/Cu layers the controlling effect is the thermodynamic driving force.[A1] The Al layers are not shown in the figure. A controversy based on observations[A2] was resolved by this work.

A further first-principle examination of the relative effect of interface energy and strain energy on precipitates is provided in Reference A3. A result of this work is shown in Figure A1.2. This represents a generic simulation in a homophase system in which the interaction assumed is nearest neighbor only, assumed elastically soft direction is $<111>$ and the coherency strain is evaluated for assumed lattice parameter difference between components.

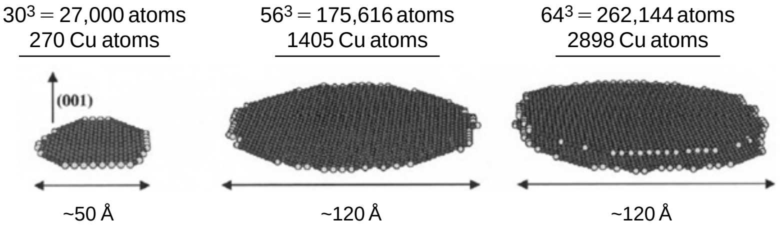

30³ = 27,000 atoms
270 Cu atoms
56³ = 175,616 atoms
1405 Cu atoms
64³ = 262,144 atoms
2898 Cu atoms
(001)
~50 Å
~120 Å
~120 Å

Figure A1.1. Successively larger Guinier–Preston zones in Al∼1% Cu alloy. From MRS Bull. 197, March 2001 with permission.

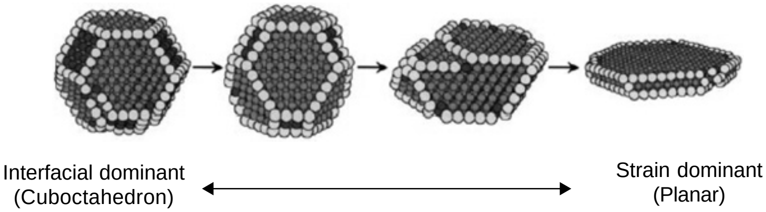

Interfacial dominant
(Cuboctahedron)
Strain dominant
(Planar)

Figure A1.2. Generic result of a Monte Carlo simulation showing the relative effects of interface energy and strain energy on precipitate shape for a soft elastic modulus along <111> in a fcc phase and a coherent interface. From C. Wolverton, Model. Simul. Mater. Sci. Eng. 8, 323(2000) with permission. Copyright Modelling and Simulation in Materials Science and Engineering, 2000.

The left-hand side of the figure represents the strain-free condition where only the interface energy controls the shape. As shown, the shape is consistent with a Wulff shape construction for minimum interface energy. The right-hand side corresponds to the case where the strain energy dominates and, as shown, the result is a plate with interface plane being normal to the soft elastic modulus.

A1. C. Wolverton, Philos. Mag. Lett. 79, 683(1999).
A2. V. Gerold, Scripta Metall. 22, 927(1988).
A3. C. Wolverton, Model. Simul. Mater. Sci. Eng. 8, 323(2000).

References

1. F.R.N. Nabarro, Proc. Roy. Soc. A 175, 519(1940); Proc. Phys. Soc. 52, 90(1940).
2. J.K. Lee and J.H. Choy, MRS Symp. Proc. 398, 439(1996).
3. H.E. Cook, D. deFontaine and J.E. Hilliard, Acta Met. 17, 765(1969).
4. H.I. Aaronson and F.K. LeGoues, Metall. Trans. 23A, 1915(1992).
5. V.I. Yukalov, Phys. Rep. (Review Section of Physics Letters) 208 (6), 395–489(1991), North-Holland.

6. J.W. Cahn and J.E. Hilliard, J. Chem. Phys. 31, 688(1959).
7. M. Tomozawa, R.K. McCrone and H. Herman, Phys. Chem. Glass. 11, 572(1970).
8. H.L. Snyder and P. Meakin, J. Chem. Phys. 79, 5588(1983).
9. L.V. Alvensleben and R. Wagner, in **Decomposition of Alloys: The Early Stages**, eds. P. Haasen, V. Gerold, R. Wagner and M.F. Ashby, Pergamon Press, New York, 1984, p. 143.
10. K.E. Biehl and R. Wagner, Proceedings of 27th International Field Emission Symposium, eds. Y. Yashiro and N. Igata, Tokyo, 1980, p. 267; Proceedings of International Conference on Solid–Solid Phase Transformations, TMS, Warrendale, PA, 1982, p. 185.
11. L.R. Testardi and T.B. Bateman, Phys. Rev. 154, 402(1967).
12. U. Stuhr, P. Vorderwisch, V.V. Kokorin and P.-A. Lindgard, Phys. Rev. B 56, 14360(1997).
13. P.G. Radaelli, D.E. Cox, M. Marezio, S.-W. Cheong, P.E. Schiffer and A.P. Ramirez, Phys. Rev. Lett. 75, 4488(1995); D. Louca, G.H. Kwei and J.F. Mitchell, Abstract V4.4, Fall, 1997 Meeting MRS.
14. J.Z. Sun, L. Krusin-Elbaum, A. Gupta, G. Xiao, P.R. Duncombe and S.S.P. Parkin, IBM J. Res. Dev. 42, 89(1998).
15. G. Bitton, M. Razvag and A.J. Agranat, Phys. Rev. B 58, 5282(1998).
16. K.L. Merkle, L.J. Thompson and F. Phillipp, Phys. Rev. Lett. 88, 225501(2002).

Bibliography

1. M. Volmer and A. Weber, Z. Phy. Chem. 119, 277(1925).
2. R. Becker and W. Doring, Ann. Phys. 24, 719(1935).
3. J. Frenkel, **Kinetic Theory of Liquids**, Oxford University Press, 1946; Dover, New York, 1955; **Statistical Physics**, Akadamiya Nauk, Moscow, 1948; **Introduction to the Theory of Metals**, Nauka, Leningrad, 1972.
4. J. Cahn and J.E. Hilliard, J. Chem. Phys. 28, 258(1958); 31, 688(1959); Acta Met. 9, 795(1961); 10, 179(1962).

Problems

1. In the heterogeneous nucleation of a stable phase from a supersaturated or supercooled parent phase what energies act as barriers to the nucleation process?
2. Why does the radius of the critical spherical nucleus depend upon temperature?
3. How does the number of embryos of a given radius depend upon temperature?
4. On the assumption of isotropic elastic properties and an incoherent interface between nucleus and matrix explain why certain nuclei are plate shaped and others are spherically shaped.
5. Do microscopic precipitates necessarily have the same shape as the nuclei from which they formed? Explain your answer.
6. What is the chemical potential of a vacancy at equilibrium? Is there a real chemical potential? Explain your answer.
7. Why are grain boundaries preferred sites of nucleation in solids?

8. In the catalysis of nucleation in supersaturated or supercooled matrices why are certain catalyst particles more effective than others?
9. If at the temperature of precipitation, the curvature of the free energy–composition function of the supersaturated matrix phase at its corresponding composition is positive will there be a barrier to the development of composition fluctuations? Explain your answer.
10. Give the analytic definition of the spinodal.
11. In a miscibility gap binary system, why does the minimum possible thickness of the interface between the two terminal phases depend upon the maximum difference between the free energy of the supersaturated solution and that given by the common tangent to this free energy–composition curve representing the free energy of the mixture of equilibrium compositions?
12. What is the basic assumption concerning the free energy of a non-homogeneous phase that allows an analytic description of the energy of a diffuse interface?
13. If we define the generalized spinodal to represent the temperature–composition function in a phase diagram that separates the region where a composition fluctuation in the supersaturated solid solution increases the free energy of the system from that in which such a fluctuation decreases this free energy how will the temperature corresponding to a given composition at the generalized spinodal boundary behave: (a) as the wavelength of the composition fluctuation decreases and (b) as the linear expansion per unit increase in composition decreases?
14. How will elastic anisotropy of a supersaturated solid solution affect its spinodal decomposition?
15. Using the material parameters in the example illustrated on Problem 142, at what degree of supercooling would more than one solid nucleus exist at equilibrium in a metastable liquid sample of 1 cm^3?
16. Show on a plot of free energy per atom for a terminal phase alpha and a precipitate phase beta in a two component system, A + B, the driving force for nucleation of the beta phase in a solution at a concentration X_o.

CHAPTER VI

Thermodynamics of Defects

Introduction

Defects in solids exert significant effects in kinetic phenomena, as well as on the thermodynamic properties of semiconductors and ionic materials. We consider the thermodynamics of defects in this chapter. Applications involving defects in kinetic phenomena are considered in later chapters.

1. Monatomic solids

1.1. Point defects

The defects that we consider in this section are vacancies and interstitials. Figure 6.1 defines these defects and also makes the point that a vacancy is sometimes called a Schottky defect while the combination of the two defects (Figure 6.1(c)) is called a Frenkel defect. Let us determine the concentration of vacancies that may be expected in a monatomic solid containing N atoms at equilibrium. By the definition of equilibrium, the Gibbs free energy must be at a minimum with respect to a variation in the concentration of vacancies. Hence, we formulate the Gibbs free energy of the monatomic solid containing vacancies. If the increase in free energy of the solid on the introduction of one vacancy into some particular site in the lattice by transfer of the atom at this site to a surface site* is g_v, then the free

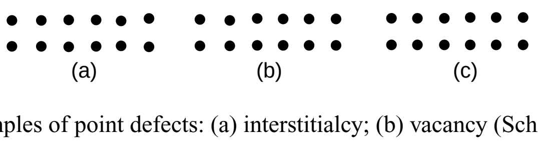


Figure 6.1. Examples of point defects: (a) interstitialcy; (b) vacancy (Schottky defect) and (c) displacement (Frenkel defect).

---

* The surface site must be at an inexhaustable source or sink for vacancies and yields the lowest value of g_v for such sites.

energy of the system containing N_v vacancies is

$$G = N_v g_v + Ng - Tk \ln W \tag{6.1}$$

where g is the free energy per atom in the absence of vacancies and W is the number of different complexions of N_v vacancies and N atoms (see equation 1.2b and note the change in symbol to prevent misinterpretation of the symbols).

Now, for a similar situation we have shown that

$$W = (N + N_v)!/[N_v!N!] \tag{6.1a}$$

Since we are concerned with the equilibrium situation we must minimize the free energy of the system of atoms and vacancies, noting that vacancies are formed at constant N. Hence

$$\partial G/\partial N_v|_N = 0 = g_v + kT \ln[N_v/(N + N_v)] \tag{6.2}$$

In obtaining this result we have made use of Stirling's approximation for the factorial of a large number, substituted the result into equation 6.1 and then carried out the differentiation. The reader's attention is called to the fact that the left-hand side of equation 6.2 is just the definition of the chemical potential of a vacancy and hence it has just been shown that the latter is equal to zero at equilibrium.* Since $N_v \ll N$ then equation 6.2 yields the following result for the number of vacancies at equilibrium in a monatomic solid

$$N_v = N \exp(-g_v/kT) \tag{6.3}$$

The work to form a single vacancy has been measured for many monatomic solids. Table 6.1 lists many of these values. Substitution of these values and the corresponding melting temperatures yields the result that the equilibrium concentration of vacancies in monatomic solids at the melting point is between 0.0001 and 0.001.

A relation equivalent to equation 6.3 holds also for the number of interstitialcies in monatomic solids. However, it is changed slightly from equation 6.3 because the number of sites that interstitialcies can occupy is not equal to the number of atoms. In general there will be greater than one interstitialcy site per atom. Let this number be p. Thus, equation 6.1a must be replaced by

$$W = (pN)!/[n!(pN - n)!]$$

where n is the number of interstitialcies. This leads to the following result for the number of interstitialcies at equilibrium

$$n = pN \exp[-g_i/kT] \tag{6.4}$$

---

* This statement is not strictly true as will be shown later in this section.

Table 6.1. Vacancy formation enthalpies in metals (eV).

Metal	**Enthalpy (±0.05, except where indicated)**
Al(fcc)	0.71
Cu	1.22
Au	0.95
Ni	1.67 ± 0.12
Pb	0.57 ± 0.07
Pt	1.32 ± 0.17
Ag	1.14
Fe(bcc)	1.5 ± 0.1
Mo	3.1 ± 0.1
Nb	2.65
Ta	2.95 ± 0.15
W	3.5 ± 0.5
Cd(hcp)	0.44
Co	1.34
In	0.49 ± 0.1
Mg	0.58
Zn	0.52 ± 0.02

There are fewer of these defects than vacancies for the simple reason that the work to form interstitialcies g_i, by exchange with an atom at a surface site is larger than that to form vacancies in these materials.

The presence of vacancies affects not only the free energy but also all the other thermodynamic quantities. It would require a much more detailed discussion than is warranted by the scope of this book to describe all these relationships. A thorough discussion of them can be found, however, in Reference 1. Divacancy “molecules” may also be an equilibrium entity in monatomic solids. Indeed, it is believed that divacancies contribute measurably to diffusion in the vicinity of the melting point. We may derive their concentration at equilibrium using equation 6.1. However, the expression given above for W must be altered in the present case because a given divacancy has several possible distinguishable orientations of one vacancy about the other fixed vacancy member of the divacancy pair. Thus, for the divacancy case

$$W = (zN/2 + N_{2v})!/[(N_{2v})!(zN/2)!]$$

where z is the coordination number. This result then yields for the number of divacancy pairs, to a good approximation

$$N_{2v} = zN/2 \exp(-g_d/kT) \tag{6.5}$$

at equilibrium in a monatomic solid (i.e. $N_{2v} \ll zN/2$). It should be noted that the atoms involved in the formation of a divacancy exchange with vacancies at sites on the surface.

In the above derivation we have not allowed the divacancies to occupy any of the lattice sites already occupied by single vacancies. Thus, the total crystal contains $N + N_v + 2N_{2v}$ total number of sites and equations 6.3–6.5 are in error to the extent that the extra number of sites in the crystal differs with respect to the total number of atoms. We shall neglect this error because it is small. Nevertheless, this physical situation suggests that we must reexamine our definition of chemical potential of a vacancy. It should now be apparent that the definition of a chemical potential of a species requires the variation of the free energy with respect to the number of that species at constant number of other species and other independent variables, such as T and P. However, the species in question must be an independent variable. Since, vacancies and clusters of vacancies can all exist in a crystal specimen of N atoms, they are no longer independent variables (i.e. the presence of one diminishes the number of sites available to the other by the number of sites that are nearest neighbors to the former defects). Nevertheless, it is useful to consider these partial derivatives with respect to a defect species as virtual chemical potentials because they obey certain relations characteristic of chemical potentials. In particular, the virtual chemical potentials of the vacancy and vacancy cluster defects are equal to zero at equilibrium.

It is possible for two vacancies to become nearest neighbors and hence form a divacancy pair. At equilibrium, this reaction obeys the mass action law, according to which for dilute concentrations of defects

$$(N_{2v}/N)/(N_v/N)^2 = K$$

where K is an equilibrium constant for the reaction. Substitution from equations 6.3 and 6.5 yields

$$K = (z/2)\exp[-(g_d - 2g_v)/kT]$$

where the expression in parentheses in the exponent represents the binding free energy of the divacancy.

It should be noted that in the above the vacancies and interstitials were formed by exchange with sites on surfaces. For such sites the free energy of formation of the defects represents the minimum possible *for an inexhaustable type of surface site* and hence these sites are likely to control the actual number of such defects present in crystals. These defects could have been formed by exchange with sites (atoms) in the vapor state. It is apparent that the work to form the defects for the latter case is larger per atom in the defect by the free energy of sublimation (i.e. the free energy required to remove an atom at a kink site along a ledge on the surface and transfer it to the vapor).

Many other defects and defect clusters can exist at equilibrium in monatomic solids. For example, rather than there being a well-defined interstitialcy, relaxation can lead to the formation of a split interstitial, which is a defect that distributes the distortion due to an interstitialcy over two atoms centered about a lattice site. In solutions, there will be association of solute atoms and defects. In general, the numbers of such defects will be given by an equation of the form of equation 6.4 or 6.5, where N is multiplied by some geometric factor.

1.2. Electronic defects

1.2.1. Intrinsic semiconductors

Semiconductors and covalent bonded monatomic solids may also, in addition to the defects noted above, contain charged entities as defects, at equilibrium. In this case, there exists the constraint of overall electroneutrality for the solid. In a pure semiconductor, the intrinsic number of conduction electrons and holes must therefore balance. The number of conduction electrons at equilibrium is affected by the Pauli exclusion principle, in that the energy levels are quantized and no more than two electrons (of opposite spin) can have the same energy. Thus, it is no longer possible to use equation 1.2b to evaluate the entropy contribution due to the conduction electrons, but instead it is necessary to use equation 1.2a. However, it is also necessary to make use of band theory to aid in solving equation 1.2a. In particular, it is necessary to know how many energy states (N(E)dE) there are between an energy E and E + dE. This number corresponds to the number of boxes available to be filled by electrons having this energy. However, each energy level may be filled by no more than two electrons of opposite spin.

It is convenient to consider the problem of determining the configurational entropy of the conduction electrons using the electrons of one spin quantum number only. The total number of ways of mixing such electrons so that there is no more than one electron per level is given by

$$W = \Pi_i N_i!/[n_i!(N_i - n_i)!]$$

where $N_i = N(E_i)dE$ and n_i is the number of electrons of one type spin in the ith increment of dE at E_i.

To obtain the equilibrium distribution of electrons we now maximize the entropy at constant volume V, energy of the system U and total number of electrons n, i.e.

$$dS = k d \ln W = 0 = k\Sigma \ln(N_i/n_i - 1)dn_i$$

where we have made use of Stirling's approximation. But, from thermodynamics

$$0 = dS = \partial S/\partial U|_{V,n} dU + \partial S/\partial n|_{U,V} dn = (1/T)dU - (\mu/T)dn$$

where μ is the chemical potential of the electrons in this case. Hence, we may write

$$0 = dS = k\Sigma \ln(N_i/n_i - 1)dn_i = (1/T)dU - (\mu/T)dn$$

or

$$k\Sigma \ln(N_i/n_i - 1)dn_i - (1/T)dU + (\mu/T)dn = 0$$

But, since $U = \Sigma n_i E_i$ and $n = \Sigma n_i$ we have $dU = \Sigma E_i dn_i$ and $dn = \Sigma dn_i$. Substituting, we then obtain the result that

$$k\Sigma[\ln(N_i/n_i - 1) - E_i/T + \mu/T]dn_i = 0$$

But, this can only be satisfied for all variations of n_i by

$$n(E)dE = N(E)dE/[1 + \exp(E - \mu/kT)] \qquad (6.6)$$

where n(E)dE has been put in place of n_i and N(E)dE has been substituted for N_i (i.e. the subscript i has been removed). Here n(E)dE is the number of electrons of one type spin that have an energy in the range between E and E + dE. For $(E_c - \mu) \gg kT$ (see Figure 6.2 for definition of E_c) then the total number of electrons (of both types of spin) in the conduction band can be shown to equal

$$n = N_c \exp[-(E_c - \mu)/kT] \qquad (6.7)$$

where $N_c = 2(2\pi m_c kT/h^2)^{3/2}$, m_c is the effective mass of the electron in the conduction band and h is Planck's constant. If $\exp((E_c - \mu)/kT)$ is on the order of unity, then the 1 in the denominator of equation 6.6 cannot be neglected and equation 6.7 must be modified. In the latter case, the electron distribution is said to be degenerate.

An equivalent relation can be derived for the number of holes n_+, by recalling that a hole corresponds to an empty state and hence that the probability that a state is empty is equal to 1 − F(E), where F(E) is given by

$$F(E) = 1/[1 + \exp(E - \mu/kT)]$$

(Reference to equation 6.6 indicates that n(E)dE = F(E)N(E)dE.) Thus, replacing F(E) by 1 − F(E) in the latter relation yields a relation similar to equation 6.7

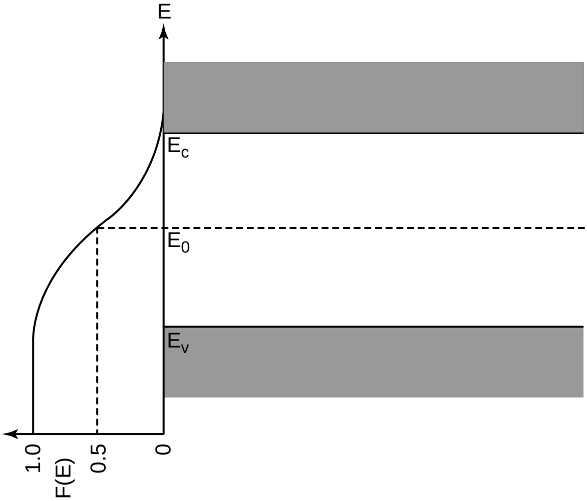


Figure 6.2. Illustration of energy levels in intrinsic semiconductor. From L.V. Azaroff and J.J. Brophy, **Electronic Processes in Materials**, McGraw-Hill, New York, 1963 with permission.

except that the term in the parenthesis in the exponent is replaced by $(\mu - E_v)$ and the effective mass of the electron in the conduction band is replaced by that in the valence band. We may now relate the chemical potential of the electrons to the Fermi level by noting that the latter is defined to equal the energy E when F(E) = 0.5. Substitution in F(E) then yields that the Fermi energy $E_f = \mu$ (see Figure 6.2).

To determine the actual position of the Fermi level we equate the total number of electrons to the total number of holes. This yields

$$E_f = (E_v + E_c)/2 = (3\,kT/4)\ln(m_v/m_c)$$

If the two effective masses are equal then the Fermi level is in the middle of the gap.

1.2.2. Extrinsic semiconductors

In extrinsic semiconductors it is of interest to know how the Fermi level depends upon the donor and/or acceptor concentrations. We will find that the limit of solubility of a solute is dependent upon the position of the Fermi level. We assume that classical statistics apply so that we can make use of the law of mass action. In effect, this assumption corresponds to the case where the difference in energy between the donor or acceptor levels and the Fermi energy is much greater than zero. In this case, the relations are much simplified. In any case we have the requirement of electrical neutrality, according to which the number of electrons plus ionized acceptors equals the sum of the number of holes and ionized donors. Now, it can be shown that the number of ionized species is given by

$$n_{s*} = n_s/[2\exp\{(E_s - E_f)/kT\} + 1] \tag{6.8}$$

where E_s is the energy level of the donor or acceptor, as s may represent.

We can consider three limiting cases, as follows:

1. Both acceptor and donor impurities present, but weakly ionized. In this case, setting the electroneutrality condition yields

$$E_f = (E_d + E_a)/2 + (kT/2)\ln(m_v/m_c) + [(E_c - E_d) - (E_a - E_v)]/2 \tag{6.9}$$

For the assumed case in which classical statistics applies the third term on the right-hand side is small compared to the first term and if the effective masses are equal then the Fermi level is near the center of the gap as in an intrinsic semiconductor.

2. Only donor solute present: In this case, for the assumed conditions, the number of electrons is given almost wholly by the number of ionized donors, i.e. the contribution of intrinsic electrons from valence band excitation will be negligible. Thus, setting equation 6.7 equal to equation 6.8 and solving for the Fermi level yields

$$E_f = E_c + kT\ln(n_d/\{2[2\pi m_c kT/h^2]^{3/2}\}) \tag{6.10}$$

where n_d is the concentration of donors per unit volume. Thus, the Fermi level is close to the conduction level provided that the temperature is not too high.

3. Only acceptor type solute present: Proceeding in a manner analogous to the case just described, the following result for the Fermi level is obtained

$$E_f = E_v - kT\ln(n_a/\{2[2\pi m_v kT/h^2]^{3/2}\}) \tag{6.11}$$

where n_a is the concentration of acceptors per unit volume. Hence, in this case the Fermi level is close to the valence level, if the temperature is not too high. Figure 6.3 shows how the Fermi level varies with temperature in n- and p-type semiconductors.

Suppose that we have a semiconductor that is equilibrated with respect to a reservoir of solute that maintains the chemical potential of the latter constant. At equilibrium, the change in free energy of the system must be equal to zero. Hence, for transfer of dn_s solute to the semiconductor, which produces for the case of a donor solute the same number of ionized solute atoms and electrons, we have

$$\mu_s^o = \mu_s^+ + \mu_{el}$$

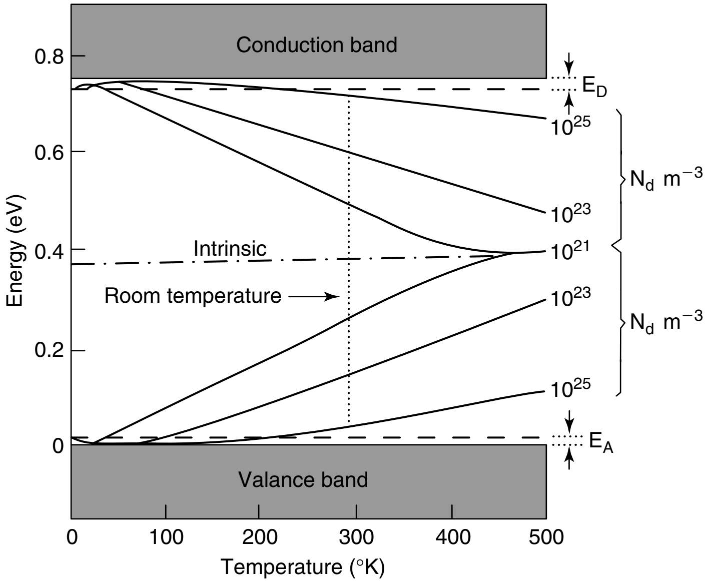


Figure 6.3. Illustrating the temperature dependence of the Fermi energy in an extrinsic semiconductor. From L. Azaroff and J.J. Brophy, **Electronic Processes in Materials**, McGraw-Hill, New York, 1963 with permission.

Now, μ_s^o is maintained constant. Consequently, any change in μ_{el}, which is equal to the Fermi level must bring about the opposite sign change in μ_s^+. Since the latter is monotonically related to the concentration of the solute in solution, the latter must be similarly changed by the change in Fermi level, i.e. an increase in the Fermi level will decrease the concentration of the solute that is in equilibrium with an external reservoir and vice versa. In particular, adding a donor solute to a solid solution that already contains acceptors at the solid solubility level will raise the Fermi energy and decrease the solid solubility of the acceptor type solute. On the other hand, adding an acceptor type solute to a solution that already contains donor type solute at its solid solubility level will decrease the Fermi energy and thereby raise the solid solubility of the donor type solute. It can be shown that the ratio of the solubility of a donor solute in the presence of ternary solutes X_d, to that in the absence of ternary solute X_d^o, is given by

$$X_d/X_d^o = \exp[(E_F^o - E_F)/kT]$$

Similarly, for an acceptor solute, this relation becomes

$$X_a/X_a^o = \exp[-(E_F^o - E_F)/kT]$$

Only a slight change in the Fermi level can exert orders of magnitude change on the solubility.

We do not present here a complete description of the thermodynamics of the electrical defects in semiconductors for it involves matters outside the scope of this book. The interested reader is referred to Reference 1, the book by Swalin and the articles by Reiss cited in the Bibliography.

2. Compounds

2.1. Defects in stoichiometric compounds

In ionic compounds there are two constraints governing the defects that may form in such materials. One is that the ratio of the number of cation to anion lattice sites is a constant and equal to the stoichiometric ratio of cations to anions, even for non-stoichiometric compositions. Thus, for the compound MX, the ratio of cation to anion lattice sites is unity even for off-stoichiometric compositions. The second constraint is that the solid must be electrically neutral. To achieve electroneutrality the total number of positive charges must equal the total number of negative charges.

In ionic compounds the equivalent of the vacancy defect is the Schottky defect, which consists of a combination of cation and anion vacancies obeying the electroneutrality condition. This configuration obeys the electroneutrality requirement because the effective positive charge on the anion vacancies is balanced by the effective negative charge on the cation vacancies. The effective charge on a vacancy is due to the absence of the charge on the lattice site that would normally be present if an ion occupied that site. Thus, the sign of the effective charge of an anion vacancy is positive. Schottky defects can form, as in metals, by exchange of a lattice sited cation and anion with equivalent vacant surface lattice sites.

The equivalent of the interstitialcy is the Frenkel defect. This defect consists of a vacancy–interstitial pair, which can be formed by the jump of an ion from a lattice site into a neighboring interstitial site. This defect is obviously electrically neutral.

Let us consider the number of Schottky pair defects and Frenkel pair defects that may exist at equilibrium in a stoichiometric crystal of formula MX. For the case of the Schottky defect consider that there are a total of 2N atoms and that the Schottky defect is produced by exchange of interior ions with surface sites. Thus, mixing of vacancies having a given sign with their equivalent ions on the sites belonging to these ions contributes

$$\Delta S = k \ln[(N + n)!/(n!N!)] \tag{6.12}$$

to the entropy of the system. Equation 6.12 also represents the contribution of mixing the vacancies having the opposite sign on their sites. Hence, the total change in entropy for the Schottky defects is twice that given by equation 6.12. If g_p is the work to form one Schottky pair of vacancies then the change in free energy of the system due to the production of n Schottky pairs randomly distributed in the lattice is

$$\Delta G = ng_p - T2\,\Delta S$$

where ΔS is given by equation 6.12. Minimizing the free energy with respect to n and using Stirling's approximation yields

$$n = (N + n)\exp(-g_p/2kT) \tag{6.13}$$

for the equilibrium number of Schottky pairs.

For the case of Frenkel defects, the mixing entropy involves a somewhat different relation compared to that for the Schottky defect. Again assume that there are N ions, some of which will move to interstitial sites to form n Frenkel defect vacancy–interstitial pairs. In this case, the contribution to the mixing entropy due to the vacancies is

$$S_v = k\ln\{N!/[(N - n)!n!]\} \tag{6.14}$$

If there are p interstitial sites per Frenkel ion site then the contribution to the mixing entropy due to the interstitial ions is

$$S_i = k\ln\{(pN)!/[(pN - n)!n!]\} \tag{6.15}$$

Proceeding as for the case of the Schottky defect leads to the following result for the number of Frenkel defects at equilibrium

$$n = [(N - n)(pN - n)]^{1/2}\exp(-g_F/2kT) \tag{6.16}$$

where g_F is the work to form one particular Frenkel defect pair.

The Schottky and Frenkel defects are the one usually found in stoichiometric ionic compounds. However, other types of defects can also occur in such materials. For example, there can be an exchange of ions such that cations sit on anion sites and vice versa, or if one type of ion sits on a wrong site it is compensated by a vacancy on its own site or by the other type of ion in an interstitial site, etc. Indeed, additional types of complex defects have been found and with time and continuing research it is likely that new defects will be discovered.

Table 6.2. Enthalpies of defect formation in ionic crystals.[a]

Crystal	ΔH (kJ/g atom)	Crystal	ΔH (kJ/g atom)
Schottky defects			
LiF	225–258	KI	154
LiCl	212	CsCl	179
LiBr	173	CsBr	193
LiI	129	CsI	183
NaCl	210–229	TlCl	125
NaBr	166	$PbCl_2$	150
KCl	218–222	$PbBr_2$	135
Frenkel defects (cation sublattice)			
AgCl	119, 139	AgBr	102
Frenkel defects (anion sublattice)			
CaF_2	220–270	BaF_2	180
SrF_2	220		

[a] From tabulation by L.W. Barr and A.B. Lidiard, in **Physical Chemistry, An Advanced Treatise**, Vol. X, Academic Press, New York, 1970.

Some measure of the likelihood of the existence of a given type of defect in some stoichiometric compounds can be obtained from the data given for the enthalpies of formation of the various defects, which are listed in Table 6.2. If the enthalpies of a given defect have been measured for some compound it is reasonable to assume that they are the majority defect in that compound.

Electronic defects will be present in compound as well as monatomic insulators or semiconductors. The relations governing their concentration have been derived in a previous section. It is of interest to note for the case of small concentrations of electrons and holes, that from equation 6.7 and the equivalent one for holes, the product of the electron and hole concentrations will be exponentially dependent upon the energy gap. Table 6.3 lists the energy gap values for some pure stoichiometric compounds and the corresponding concentrations of electrons and holes.

2.2. Non-stoichiometric compounds

Non-stoichiometry can be achieved by having an excess (or deficiency) of some component relative to the stoichiometric composition. Because the stoichiometric ratio of lattice sites is conserved, non-stoichiometry is equivalent to the presence of point defects. Electroneutrality in such a case is preserved by the formation of complementary electronic defects. For example, an anion deficient compound can have either anion vacancies or cation interstitials maintain the ratio of lattice sites for the compound. If anion vacancies are formed then their effective positive charge

Table 6.3. Band gap[a] and approximate concentrations of electrons and holes in pure, stoichiometric solids.

		$n \approx 10^{19} \exp\left[-\frac{E_g}{2kT}\right]$ electron/cm^3			
Crystal	**E_g (eV)**	**Room temperature**	**1000°K**	**Melting point**	**Temperature (°K)**
KCl	7	10^{-40}	20	150	1049
NaCl	7.3	10^{-43}	4	70	1074
CaF_2	10	10^{-66}	10^{-6}	10^{3}	1633
UO_2	5.2	10^{-25}	10^{6}	10^{15}	3150
NiO	4.2	10^{-16}	10^{8}	10^{13}	1980
Al_2O_3	7.4	10^{-44}	2.0	10^{11}	2302
MgO	8	10^{-49}	0.01	10^{12}	3173
SiO_2	8	10^{-49}	0.01	10^{8}	1943
AgBr	2.8	10^{-5}	10^{12}	10^{9}	705
CdS*	2.8	10^{-5}	10^{12}	10^{15}	1773
CdO*	2.1	20	10^{13}	10^{16}	1750
ZnO*	3.2	10^{-8}	10^{11}	10^{14}	1750
Ga_2O_3	4.6	10^{-20}	10^{7}	10^{13}	2000
LiF	12		10^{-11}	10^{-8}	1143
Fe_2O_3*	3.1	10^{-7}	10^{11}	10^{14}	1733
Si	1.1	10^{10}	10^{16}	10^{17}	1693

[a] Most of the data are based on the optical band gap which may be larger than the electronic band gap.
* Sublimes or decomposes.

is balanced by extra electrons associated with the cations. However, if cation interstitials are formed, then the extra electrons, necessary to balance the effective positive charge of the cation interstitials, are associated with the lattice sited cations. In such a case, the number of point defects is not determined solely by thermal equilibrium of the defects, but by the non-stoichiometry as well. Let us consider one example to illustrate the thermodynamics of one such situation.

We consider the case of an anion deficient compound for which the anion vacancy is the most probable associated defect. If the non-stoichiometry is accomplished by equilibrium of the compound with some partial pressure of the anion vapor and if the compound has the formula MX and the molecules of X in the vapor phase are normally diatomic (i.e. X_2), then the transfer of an anion from a lattice site to the vapor produces an anion vacancy. The equation for this reaction can be written as follows

$$X_X <=> V_X + 1/2X_2 \tag{6.17}$$

In this equation the anion vacancy is neutral. Depending on the temperature, the trapped electrons associated with the vacancy may be excited and freed from the vacancy, with the result that the following equilibrium reactions are obeyed

$$V_X <=> V_{X^{\cdot}} + e' \tag{6.18}$$

$$V_{X^{\cdot}} <=> V_{X^{\cdot\cdot}} + e' \tag{6.19}$$

(The absence or presence of dots (..) associated with the defect indicates the absence or presence of effective positive charge. Thus, with the absence of effective positive charge the defect is electrically neutral. Also, the presence of a prime (′) indicates the presence of an effective negative charge.) The free electrons indicated in these equations are associated with the cations on their normal sites and hence these equations may be rewritten as

$$M_M + V_X <=> V_{X^{\cdot}} + M_{M'} \tag{6.20}$$

$$M_M + V_{X^{\cdot}} <=> V_{X^{\cdot\cdot}} + M_{M'} \tag{6.21}$$

In the latter equations M-type cations are changed to M′ cations with one effective negative charge. The valence of the M cations is changed from +2 to +1 in these reactions.

In addition to the above equations, another is involved in the intrinsic formation of electronic defects, namely, the formation of electrons and holes by excitation of electrons from the valence band to the conduction band of the solid, which we have considered in the previous section.

All of the above reactions are governed by equilibrium constants and the mass action law. The latter corresponds to equations of the type illustrated by equations 6.13 and 6.16, whereas the equilibrium constants are given by the reciprocal of the exponential terms in these equations. Thus, the concentration of the various defects denoted by the terms in brackets [ ] are

$$[V_X](p(X_2))^{1/2} = K_1[X_X] \tag{6.22}$$

$$[V_{X^{\cdot}}]n = K_2[V_X] \tag{6.23}$$

$$[V_{X^{\cdot\cdot}}]n = K_3[V_{X^{\cdot}}] \tag{6.24}$$

where n equals the concentration of electrons. If, as assumed, anion vacancies and the complementary electrons are the most numerous defects then electroneutrality requires

$$n = [V_{X^{\cdot}}] + 2[V_{X^{\cdot\cdot}}] \tag{6.25}$$

Substitution into the above relations and solution for separate variables yields for the electron concentration equation

$$n^3 = K_1K_2(p(X_2))^{-1/2}(2K_3 + n) \qquad (6.26)$$

There are two limiting solutions for n as follows

$$\text{for } n \gg 2K_3 \quad n = (K_1K_2)^{1/2}[p(X_2)]^{-1/4} \qquad (6.27)$$

$$\text{and for } n \ll 2K_3 \quad n = (2K_1K_2K_3)^{1/3}[p(X_2)]^{-1/6} \qquad (6.28)$$

Similarly, there are several limiting solutions for the total number of anion vacancies depending upon which of the vacancies is the predominant species, i.e. doubly charged, singly charged or neutral ones. For the case that neutral vacancies predominate then the total anion vacancy concentration is given by $K_1[p(X_2)]^{-1/2}$. If singly charged vacancies predominate then the total vacancy concentration is equal to $(K_1K_2)^{1/2}[p(X_2)]^{-1/4}$. If, on the other hand, doubly charged anion vacancies predominate then the total vacancy concentration is given by $[(1/4)K_1K_2K_3]^{1/3}$ $[p(X_2)]^{-1/6}$. Examination of these results shows that both the electron concentration and the anion vacancy concentration are dependent on the partial pressure of the anion vapor in equilibrium with the compound.

It is possible to consider other limiting conditions, such as the case of an anion deficient compound in which the predominant defect is the Frenkel defect, or the case of a cation deficient compound in which one or the other defect species predominates, etc. It is obvious that there are many possible sets of conditions that can exist in any particular material. Table 6.4 lists the types of defects that can be found in oxides. Table 6.5 gives the equilibria describing the formation of typical point defects in oxides based on the criteria of electroneutrality, conservation of mass and maintenance of the proportion between cation and anion sites.

Table 6.4. Defects in non-stoichiometric oxides.

Defects	Defect type	Compensating species	Oxide system
Point defects			
Oxygen vacancies	V_O^X, V_O^*, V_O^{**}	$M_M^{'}$	MO_{2-x}
Metal vacancies	V_M^X, $V_M^{'}$, $V_M^{''}$	M_M^*	$M_{1-y}O$
Interstitial oxygen	O_i^X, $O_i^{'}$, $O_i^{''}$	M_M^*	MO_{2+x}
Interstitial metal	M_i^X, M_i^*, M_i^{**}	$M_M^{'}$,	$M_{1+y}O$
Substitutional disorder	$M_M^{''}$	V_O^{**}	
Extended defects			
Clusters	Combination of point defects		$Fe_{1-x}O$ UO_{2+x}
Shear planes	Elimination of point defects		TiO_{2-x} WO_{3-x}

Table 6.5. Defect equilibria.

	Defect reaction	Oxide
$V_O^{\cdot\cdot}$	$2M_M + O_O \rightarrow V_O^{\cdot\cdot} + 2M_M' + \frac{1}{2}O_2$	Any
V_M''	$2M_M + \frac{1}{2}O_2 \rightarrow V_M'' + 2M_M^{\cdot} + O_O$	MO
O_i''	$2M_M + \frac{1}{2}O_2 \rightarrow O_i'' + 2M_M^{\cdot}$	Any
$M_i^{\cdot}$	$M_M + O_O \rightarrow M_i^X + \frac{1}{2}O_2$ $M_i^X + M_M \rightarrow M_i^{\cdot} + M_M'$	MO
	$2M_M + O_O \rightarrow M_i^{\cdot} + M_M' + \frac{1}{2}O_2$	
$M_i^{\cdot\cdot}$	$M_M + 2O_O \rightarrow M_i^X + O_2$ $3M_M + M_i^X \rightarrow M_i^{\cdot\cdot\cdot} + 3M_M'$	MO_2
	$4M_M + 2O_O \rightarrow M_i^{\cdot\cdot\cdot} + 3M_M' + O_2$	

Table 6.6. Defect formation: $p(O_2)$ dependence.

$M_i^X, V_O^X \propto p(O_2)^{-1/2}$
$M_i^{\cdot}, V_O^{\cdot} \propto p(O_2)^{-1/4}$
$M_i^{\cdot\cdot}, V_O^{\cdot\cdot} \propto p(O_2)^{-1/6}$

$O_i^X, V_M^X \propto p(O_2)^{1/2}$
$O_i', V_M' \propto p(O_2)^{1/4}$
$O_i'', V_M'' \propto p(O_2)^{1/6}$

For the case where the mass action law can be applied to the defect reactions, the partial pressure of oxygen in the non-stoichiometric range of composition becomes proportional to $p(O_2)^{-1/n}$. Table 6.6 summarizes this dependence for the various defects. In order to simplify the situation diagrams, called Brouwer or Kroger–Vink plots, have been prepared showing the behavior to be expected in a compound that adheres to the conditions assumed. Figure 6.4(a)–(d) illustrates several such diagrams. It is to be emphasized, however, that these diagrams apply only for the assumed conditions. If there is association between defects or there are impurities present or there are multiply charged defects, etc., the diagrams will be affected.[2]

The tendency for the existence of a range of solubility about the stoichiometric composition is affected by the ionization potential of the cation. For cations with low ionization potential there will be a tendency for an extensive region of solubility. For cations with high ionization potential the range of solubility about the stoichiometric composition will be limited and non-stoichiometry in these materials is often related to impurity content.

The range of defect possibilities in compounds is enormous. We have alluded to only a few of them. Defects can be ordered, clustered, complexed, sheared, etc.[3] It is beyond the scope of the present simplified discussion to consider them in detail. We have not considered molecular solids in this section, in which another

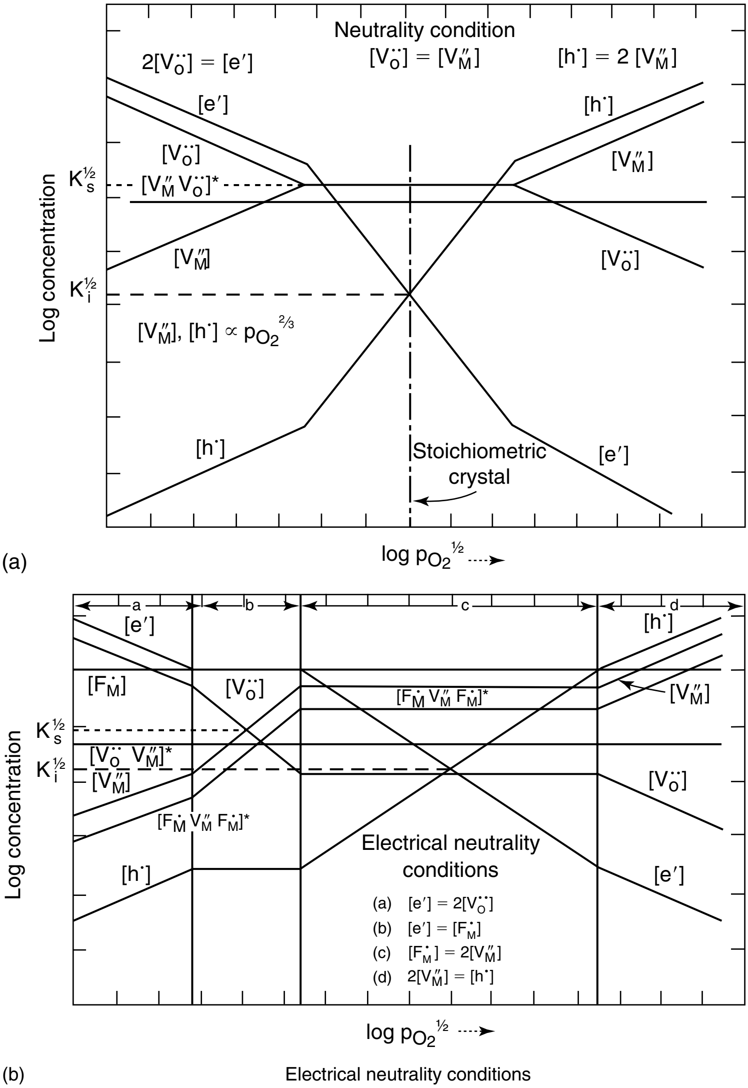


Figure 6.4. Schematic representation of defect concentrations as a function of oxygen pressure for: (a) a pure oxide which forms predominantly Schotitky defects at the stoichiometric composition. (b) An oxide which forms Schottky defects but contains cation impurities $[F_M^{\cdot}] > K^{1/2}$. From W.D. Kingery, H.K. Bowen and D.R. Uhlmann, **Introduction to Ceramics**, John. Wiley & Sons, New York, 1976 with permission.

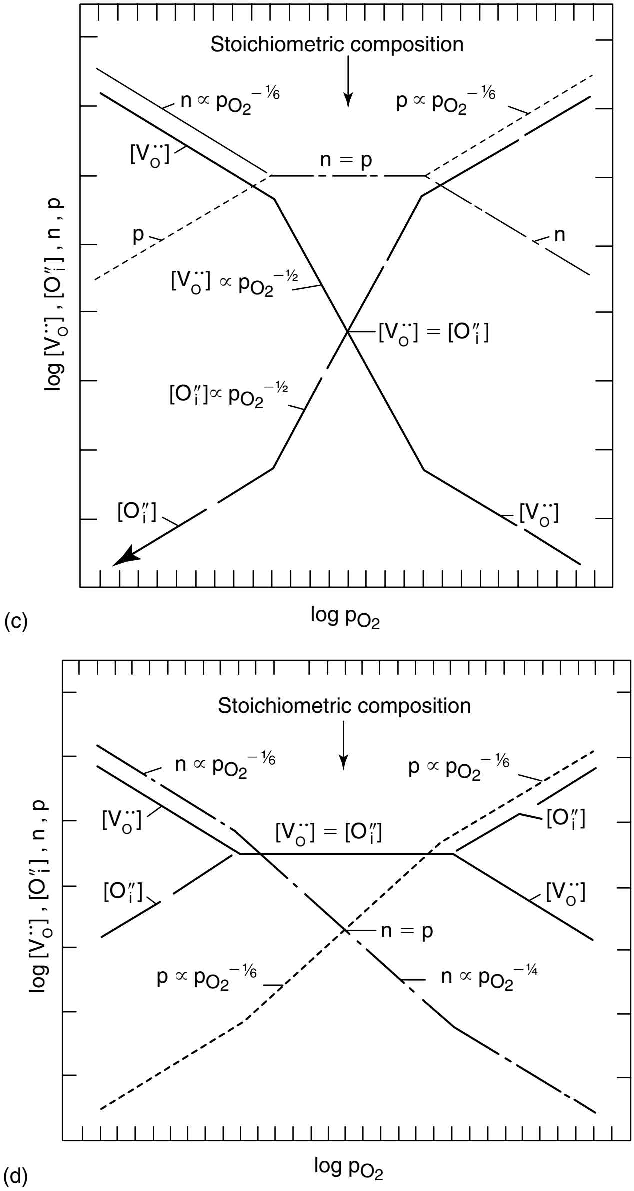


Figure 6.4. Brouwer plot for: (c) $V_O^{\bullet\bullet}/O_i''$ system with intrinsic ionization dominating near the stoichiometric composition; (d) $V_O^{\bullet\bullet}$ system with atomistic disorder dominating near the stoichiometric composition.

From P. Kofstad, **Non-Stoichiometry, Diffusion and Electrical Conductivity in Binary Metal Oxides**, Wiley-Interscience, New York, 1972 with permission.

type of defect involving rotational disorder of molecular units can occur, as, for example, is found in polymers.[4] Nor have we considered those solids in which orientational disorder can occur, as, for example, is found to take place in KCN and other salts that contain polyatomic ions.[5] The interested reader must be prepared to devote sufficient time to develop expertise in the subject of defects in solids.

3. Comparison of defects in metals, semiconductors and ionic crystals

Metals are characterized by having one predominant equilibrium point defect – the vacancy. This situation exists because the work to form a single vacancy is much smaller than that for any other defect or defect complex. In some metals, the work to form a divacancy pair is not much larger than that for the single vacancy and divacancies then complement the single vacancies at high temperature. In semiconductors, the situation is more complicated in that the possible equilibrium point defects can exist in both an electrically neutral and charged condition and in that the lattice is more open than for the other solids. Thus, not only can charged and neutral vacancies exist but because of the open nature of the lattice the work to form an interstitial type defect may not be significantly higher and may actually be lower than that for a vacancy type defect. Finally, in ionic solids the complexity of the equilibrium point defect spectrum increases significantly.

The previous discussion suggests that some quantitative knowledge of the work to form point defects can be helpful in ascertaining the equilibrium population of the competing defects. Indeed, although much effort has been expended to distinguish between the defects experimentally, most of the observations cannot be unambiguously interpreted. Thus, a systematic attack on the problem of modeling the point defects energetically has been begun. To date, the most success has been achieved in modeling the defects in ionic solids.[6] Unfortunately, for the case of metals, calculations based on rigorous interatomic potentials have been limited to a few metals to which pseudopotential theory could be applied. Also, use of pair-functional methods, such as the embedded atom method, allows point defect energies to be calculated, but only for non-transition elements. Nevertheless, the use of empirical interatomic potentials has been helpful in the discovery of new modes of impurity-point defect complexing and the like.[7]

4. Average free energy of defects

Although, as we have noted, the work to form one defect in the lattice is greater than zero, nevertheless the total Gibbs free energy of the solid is reduced by the entropic contribution of the defects. The amount by which this total Gibbs free energy is reduced has been evaluated for elements and compounds.[8–10] Not surprisingly, it turns out that the average Gibbs free energy contributed per thermally excited defect is −kT.

References

1. L.A. Girifalco, **Statistical Physics of Materials**, John Wiley, New York, 1973.
2. F.A. Kroger and H.J. Vink, **Solid State Physics** 3, 317(1956); F.A. Kroger, **The Chemistry of Imperfect Crystals**, North Holland, New York, 1956.
3. O.T. Sorensen in **Mass Transport in Solids**, eds. F. Beniere and C.R.A. Catlow, Plenum Press, New York, 1981; ibid, C.R.A. Catlow.
4. P. Flory, **Statistical Mechanics of Chain Molecules**, Hanser, New York, 1989.
5. N.G. Parsonage and L.A.K. Staveley, **Disorder in Crystals**, Clarendon Press, Oxford, 1978.
6. C.R.A. Catlow and W.C. Mackrodt, eds., **Computer Simulation of Solids**, Lect. Notes Phys. 166, Springer-Verlag, Berlin, 1982.
7. P.C. Gehlen, J.R. Beeler Jr. and R.I. Jaffee, **Interatomic Potentials and Simulation of Lattice Defects**, Battelle Institute of Materials Science Colloquia, Plenum, New York, 1971.
8. R.A. Johnson, Phys. Rev. B. 50, 13799(1994).
9. P.T. Landsberg and S.G. Canagaratna, Phys. Rev. B. 55, 5531(1997).
10. M. Fahnle, Phys. Stat. Sol. B 206, R1(1998).

Bibliography

1. R.A. Swalin, **Thermodynamics of Solids**, John Wiley, New York, 1962.
2. F.A. Kroger, F.H. Stieltjes and H.J. Vink, Philips Res. Rep. 14, 557(1959).
3. P. Kofstad, **Nonstoichiometry, Diffusion and Electrical Conductivity in Binary Metal Oxides**, Wiley-Interscience, New York, 1972.
4. H. Reiss, C.S. Fuller and F.J. Morin, Bell System Tech. J. 35, 535(1956) and later publications.

Problems

1. Suppose that the work to form a vacancy by transfer of an atom from a particular lattice site to a kink site along an edge dislocation line is more (less) than that corresponding to a surface site. What would you expect to occur to the dislocations if the specimen is equilibrated at elevated temperature?
2. Evaluate the binding free energy to form a divacancy if the concentration of divacancies is to be the same as that for single vacancies at the melting point of aluminum (660°C).
3. Why is the concentration of vacancies at equilibrium at 300°C in aluminum likely to be different from that of divacancies even if the assumption of problem 2 regarding the equality of their concentrations at the melting point is valid?

4. Derive the expression for the equilibrium concentration of solute–vacancy pairs in terms of the work to form one such pair.
5. Suppose that the substitutional solute B present is not mobile at temperature T while the interstitial solute C is mobile at this temperature. Also suppose that the binding energy of a B–C pair is E_b. Calculate the number of such pairs that will exist at this temperature if the atomic fractions of B and C per substitutional atoms present are X_B and X_C, respectively, in a lattice containing N total substitutional atoms.
6. Suppose that the addition of a ternary solute increases the Fermi energy by 0.1 eV. What is the ratio of the new solubility to the solubility at 300°K before the addition of the ternary solute of an acceptor solute equilibrated with respect to a constant chemical potential external reservoir?
7. For the case of an oxide in which oxygen Frenkel defects occur, the oxygen concentration varies over a range of stoichiometry, and the electron and hole concentrations are appreciable, determine the oxygen partial pressure dependence of the various defects under the assumption that (a) the concentration of electronic defects is substantially greater than that of Frenkel defects and (b) vice versa.
8. Relate the mass action law to the "quasi-chemical" approximation considered in Chapter II. Indeed, derive this law for the $2V \rightarrow V_2$ reaction using the statistics involved in the Q–C approximation. If quantum statistics are applicable, can the mass action law be used?
9. Suppose the addition of a ternary acceptor type solute changes the Fermi energy by 0.03 eV. Evaluate the change in solubility of a donor solute at room temperature (1 eV/molecule = 23,050 cal/mol).

CHAPTER VII

Concepts in Kinetics in Solids

Introduction

In this chapter we begin our study of kinetic phenomena in materials. We should become knowledgeable about concepts common to all kinetic phenomena in condensed matter. In gases, molecules come into contact by collision processes. In condensed matter, the atoms or molecules oscillate or rotate about a stable position or orientation, at any finite temperature. Even at 0°K they perform a zero point motion. These oscillations or rotations make possible the existence of kinetic phenomena in condensed matter. Each such phenomenon involves the motions of atoms or molecules in some particular way, which we will describe as a motion along a reaction path. If the energy of the system passes through a maximum as these particles move along the reaction path, then we describe this maximum energy as an "activation energy", which the system must exceed for it to proceed from its initial state to a different state along this reaction path. Since there are many reaction paths from reactant state to product state, the path with the highest probability of being followed is that involving the least activation free energy, which is usually called the "saddle point" energy. We shall explore the deeper significance of the activation energy in this chapter.

Kinetic phenomena that occur not too far from equilibrium conditions can be treated using linear approximations to rate relations. The linear realm including concepts from reaction rate theory and the thermodynamic theory of irreversible processes comprises the scope of this chapter. The non-linear regime applicable to far from equilibrium conditions and the production of spatial patterns is discussed in later chapters and Chapter XIII. We do not consider homogeneous chemical reaction kinetics, a subject outside the scope of this chapter – the kinetics of processes in condensed phases.

1. Activation energy

Let us consider a process taking place in a condensed phase in which reactants existing in some metastable state proceed to a more stable product state. Atom or molecule arrangements define the reactant and product states. The transition between the initial

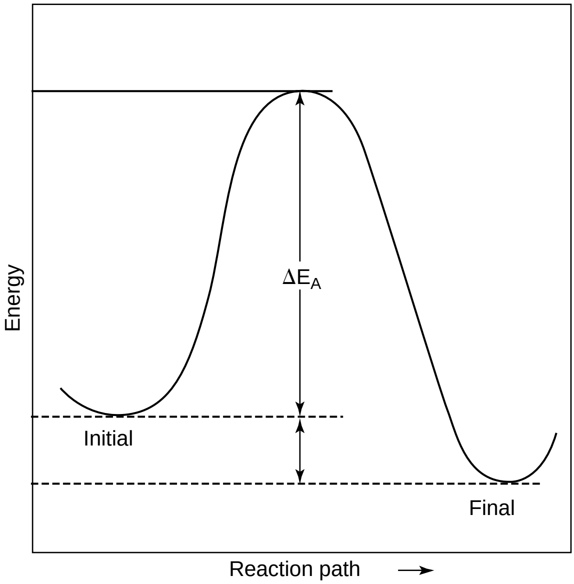


Figure 7.1. Energy of system as a function of reaction path from initial to final state.

and final states involves rearrangements of atoms or molecules and is called a reaction path. Since the initial and final states represent either metastable or stable states, the energy of the system must increase along any reaction path between them, as illustrated schematically in Figure 7.1. The assumption has been made in the foregoing that any atomic or molecular arrangement characterisitic of a position along a reaction path, the equivalent of an excited state, has a unique energy associated with it. However, the energy is not the only measure of the state of the system that is important in the problem of describing the transition from the reactant state to the product state. It is also necessary that the set of momentum vectors associated with the atoms or molecules obey some criteria, to be defined, if the transition from the reactant state to the product state is to be possible. As will be shown, the free energy is the quantity that describes both these factors.

The maximum energy along the reaction path relative to the energy of the reactant state is called the activation energy (see Figure 7.1). Thus, if the system of reactant atoms is to proceed from the reactant to the product state it must somehow be excited at least to the energy corresponding to the activation energy. The problem of obtaining an analytic relation for the rate at which this system can overcome the activation energy can be approached via several levels of sophistication. At the lowest level, the assumption can be made that the energy states of the system obey classical statistical mechanics. We may then calculate the probability that the system can have an energy that exceeds the activation energy. To obtain the rate at which the system procedes from the reactant to the product state, we multiply this probability by the frequency with which the system makes attempts to move along the reaction path from the reactant to the product state. It may be a surprising fact, but this crude concept is usually adequate. Higher levels of sophistication in treating this problem act to confirm the result given by the simple concept that the rate is proportional to the exponential term, $\exp(-E_A/kT)$ and to yield a quantitative expression for the frequency factor, which in most cases cannot be experimentally validated due to the lack of knowledge of some of the factors involved.

Let us therefore make use of the simplest method of obtaining an analytic formulation for the rate at which a system can overcome an activation energy barrier. We assume the case where the energy levels of the system are equally spaced such as in an harmonic oscillator, for which the ith level has the energy $(i + 1/2)h\nu$, where ν is the frequency of the oscillator. The partition function for the oscillator, in analogy with that developed for a set of classical states in Appendix 1 of Chapter I is

$$Z = \Sigma_i \exp[-(i + 1/2)h\nu/kT] = \exp(-h\nu/2\,kT)\Sigma_i \exp[-ih\nu/kT] \tag{7.1}$$

Let the activation energy E_A, correspond to say the nth level, i.e. $E_A = nh\nu$. Then the fraction of the number of systems having energy higher than or equal to $nh\nu$ is thus

$$p_i = Z^{-1}\Sigma_{i=n} \exp[-(i + 1/2)h\nu/kT] = \exp(-nh\nu/kT) = \exp(-E_A/kT) \tag{7.2a}$$

The fractional rate of a reaction having an activation energy E_A is then the product of the frequency with which the reactants tend to overcome the activation energy barrier ν, by the fraction of these attempts that have the correct momentum vectors $P \exp(-E_A/kT)$, where P represents the probability that these vectors are directed within some solid angle favorable for the reaction to take place or

$$R/N = \nu P \exp(-E_A/kT) \tag{7.2b}$$

Wiener[1] has given a rigorous treatment of the development of a rate equation for the case of impurity diffusion in a single degree of freedom situation. It is somewhat more physical than the previous treatment and will be repeated below. Consider a collection of impurity atoms. The phase space applicable involves specification of the positions and momenta of these atoms. The positions need only be specified relative to the equilibrium positions of the atoms. The positions $x = \pm b$ represent the points where the energy of the oscillating atoms relative to that in the equilibrium position equals the activation energy. The coordinates of the phase space are thus as shown in Figure 7.2. Now, atoms that have momentum between p and dp and will cross the line at $x = +b$ in a time dt are represented by points that occupy the shaded region in this diagram. We wish now to determine the fraction of impurity atoms that have coordinates within this region. To accomplish this we first define a

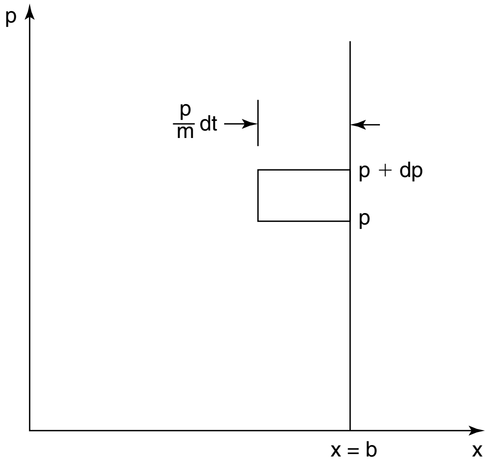


Figure 7.2. Phase space of momenta versus positions. After Wiener.[1]

distribution function f such that $\int_R f(x, p)\, dx\, dp$ equals the fraction of the impurity atoms that have coordinates in an arbitrary region R of phase space. The function f must satisfy the normalization condition that

$$\int_{-b}^{b} dx \int_{-\infty}^{\infty} df\, f = 1 \qquad (7.3)$$

It is assumed that the distribution function is not dependent upon time. Indeed, if the impurity atoms do not interact with each other and are in thermal equilibrium with a thermal bath at temperature T, then f is given by the canonical distribution function

$$f = C \exp(-E/kT) = C \exp(-p^2/2m + Kx^2/2) \qquad (7.4)$$

where the latter quadratic term is an approximation to the potential energy of an harmonic oscillator, m is the mass of the impurity atom oscillator and C satisfies the normalization condition of equation 7.3.

Now the fraction of impurity atoms having coordinates within the rectangular box of Figure 7.2 is given by

$$f(b, p)p\, dt\, dp/m$$

Division by the time interval dt and integration over all positive velocities, yields the desired fractional rate R/N, with which impurity atoms cross the activation energy barrier at x = b

$$R/N = (1/m)\int_0^{\infty} pf(b, p)dp \tag{7.5}$$

Substituting and solving yields $C = \nu_o/kT$, where $\nu_o = (K/m)^{1/2}/2\pi$ is the natural frequency of oscillation, and

$$R/N = \nu_o \exp(-E_A/kT) \tag{7.6a}$$

Because the above relation was derived for a single degree of freedom the factor P in equation 7.2 equals 1 in this case.

In the theory of absolute reaction rates[2] the fractional rate is given by

$$R/N = (kT/h)\exp[\Delta S_A]\exp[-\Delta H_A/kT] \tag{7.6b}$$

This theory has been generalized to include the case of many degrees of freedom,[3] with the result that the coefficient is no longer proportional to temperature, but has the form of equation 7.2b, with the coefficient having the significance of a ratio of the product of N normal frequencies of the entire system at the starting point of the transition over the product of the N − 1 normal frequencies of the system constrained in the saddle point configuration. The fractional rate then becomes this ratio of products ν^*, multiplied by $\exp[-E_A/kT]$. In the above, the rate to proceed along a particular reaction path between reactant and product states was evaluated. There are usually many reaction paths possible between these states. From the above derivations, it is apparent that the path with the smallest value of the activation free energy will correspond to the path with the highest reaction rate and therefore the one observed experimentally. The activation energy corresponding to this path is called the "saddle point" energy, to provide a pictorial description of its status as the minimax condition.

The net rate of a reaction proceeding in a given direction is the difference between the forward reaction rate and the reverse reaction rate. These latter two rates can be different if the activation energy in the forward direction differs from that in the reverse direction, Figure 7.1 illustrates this case. Suppose the difference in the activation free energy in these directions equals ΔG, the driving free energy. In this case, the net rate for the reaction to proceed will have the form

$$R_{net}/N = \nu_o \exp(-\Delta H_A/kT)[1 - \exp(-\Delta G/kT)] \tag{7.7}$$

For the case where the driving free energy ΔG is small compared to the thermal energy kT, the exponential term can be expanded with the result that the net rate is proportional to the driving free energy.

2. Computer assisted methods in kinetics

It is not usually possible to know the reaction path(s) along which reactants proceed to produce products. Simulation procedures have been developed to aid in this endeavor. Among these procedures are those denoted as the static lattice, molecular dynamic, Monte Carlo and phase field simulation methods. In addition, the computer has been used to assist with a solution to the chemical rate process relations for reactions involving many dependent processes with common reactant or product molecules or condensed phase entities (defects, etc.).

The objective of static lattice calculations is to evaluate the saddle point energy (the minimum of the activation energies involved in traversing reaction paths between the reactant state and the product state). To accomplish this objective an interatomic potential model, which allows the energy of the solid to be calculated as a function of the nuclear coordinates of each atom in the solid, is necessary. Also, an energy minimization method, such as the conjugate gradient,[4] is required in order to generate the lowest energy configuration consistent with the applicable constraints. Often, where defects are involved, a technique is necessary to arrive at the relaxed configuration and its concomitant energy and the Mott–Littleton[5] two region strategy is incorporated to achieve this goal. For the lattice static method to be useful it is necessary to explore the possible reaction paths with thorough evaluation of the energies along these paths. The results are usually sensitive to the choice of the interatomic potential and, hence, care must be taken to use one that provides agreement with various energies (lattice, defect, etc.) and structural parameters. Harding[6] provides a discussion of the use of lattice static defect calculations to evaluate activation energies and study defect mobilities.

Molecular dynamic simulations in kinetic problems also require interatomic potentials or functionals that enable the forces acting on each atom of the condensed phase to be calculated. Use is then made of these forces and Newton's equations of motion via an integration algorithm, such as the Verlet,[7] to describe the positions and velocities of the atoms as a function of time given an original set of these parameters at the start of the simulation process. (The velocities are chosen to correspond to a Maxwell–Boltzmann distribution for a chosen temperature.) Because the computational load is limited by the computer, the sample is usually limited to no more than about several thousand atoms (with periodic boundary conditions) and a total simulation time interval of 10^{-9} seconds. Repetition of the algorithm yielding the positions and velocities each time step of about 10^{-15} to 10^{-14} seconds and storage of the results provides the trajectories of each particle and, thus, allows the microscopic reaction pathways to be revealed. Among the computer codes used in molecular dynamic simulations are those of Parrinello–Rahman,[8] the DL_POLY code,[9] the Discover3 and Solids_Diffusion codes of Molecular Simulations Inc.

The use made of Monte Carlo simulation procedures in kinetic processes depends on the kinetic processes themselves. In general, this simulation procedure controls which of the possible parallel kinetic events will occur obeying the constraints of the applicable statistics for a sequential series of such set of events. Thus, for example, in a diffusion problem the Monte Carlo procedure may focus on the jumps of a vacancy, while maintaining in storage a record of its trajectory and of the sequence of atomic configurations in the condensed phase and from these data deduce various diffusivities, correlation factors, etc. or in a growth process the Monte Carlo procedure may determine which of possible parallel reactions occurs at a site on the surface of a growing phase, adjust for the change in surface due to the reaction and repeat the process for other sites along the surface chosen at random, while maintaining a record of the results of the growth process, etc.

In the Monte Carlo simulation method as applied to diffusion problems, one needs to determine the probability of a unit process involving a jump over a reaction path. The latter is determined as follows. Suppose that the unit process under consideration can occur by N different paths (e.g. if the unit process is the exchange of a vacancy with one of its nearest-neighbor atoms then N is the coordination number of the vacancy). The jump frequency for a particular path is given by

$$\nu_i = \nu_\iota^0 \exp(-\Delta E_i/kT) \tag{7.8}$$

where ν_ι^0 is the attempt frequency for the particular unit process and ΔE_i is its activation energy. The probability for the unit process to occur per unit time is then

$$\Gamma = \Sigma_i \nu_i = 1/\tau(n) \tag{7.9}$$

where n is the nth unit process. Now choose a random number R on the interval [0, 1]. The unit process that will occur is such that $\Sigma_1^{i^*-1}\nu_i < R/\tau \leqslant \Sigma_1^{i^*}\nu_i$ where i* denotes the particular reaction path followed to produce the new configuration. The process is then repeated for another time interval of τ. A Monte Carlo algorithm applicable to a change in configuration of an ensemble of A, B, C … type atoms on a lattice of N sites is described in Reference 10. Another algorithm applicable to a growth process involving reactions at a surface is given in Reference 11.

Other classes of computer assisted solutions to problems in kinetics have also come into existence. One of these classes comprises reaction–diffusion phenomena. Complex, coupled and non-linear processes often exist in this realm. Non-linear equations arise when conditions are not close to equilibrium conditions or when mobilities or activation energies become dependent upon concentration. Highly non-linear, far from equilibrium processes are involved in the chaos phenomenon, which yield spatiotemporal patterns. This class has only recently had applications in materials science where the production of periodic patterns of quantum dots is a desired objective. Another class of computer assisted solutions involves the free boundary problem. There are many phenomena in materials science, such as solidification, solid state transformations involving transport of atoms to and from interfaces (e.g. eutectoid transformation, precipitation), in which a free boundary moves while transforming a parent to a product phase. The phase field method has evolved to provide computer based solutions to this problem which has evaded rigorous solution by analytical means.

We will give results of examples of each simulation method in the chapters to follow so as to provide a better understanding of the applicability of each technique. There has been an exponential growth in the use of these methods in the past 20 years and they have contributed greatly to our understanding of the mechanisms involved in the kinetic processes of interest to materials science.

3. Competing processes

Suppose that independent processes for proceeding from an initial state to a final state may compete in the sense that they are physically possible processes and are not coupled to other irreversible processes. If these processes can occur in parallel, are equally probable in choice (except for the rate) and do not require one to occur before another, then it is reasonable to expect that the fastest process will be the one to produce the product phase and hence will be responsible for the observed transformation kinetics. This statement should not be taken to mean that any irreversible process occurs at its fastest possible rate. The process may be coupled to other irreversible processes that do not have maxima in their rates at the same values of the boundary conditions. For example, during solidification via the growth of dendrites it is known that the dendrites do not grow at their maximum possible velocity.

To illustrate the concept of competing independent processes consider the phenomenon known as sintering. In the sintering of spherical particles, there are four possible paths for atomic transport: via diffusion through the volume of each particle, or along the grain boundaries between particles, or along the surface of each particle, or via vapor phase transport involving evaporation and condensation steps. The transport path or combination of paths that provides the most rapid rate of sintering produces the sintered compact that is observed experimentally. This is one manifestation of the significance of independent processes in competition.

The same principle can be applied to competing independent mechanisms proposed as explanations for some irreversible process. Assuming all mechanisms do not violate thermodynamics or other physical laws then nature will operate according to that mechanism that provides the fastest means of executing the irreversible process.

The same concept underlies a new development in the kinetics of processes that occur in matter that is not at equilibrium. This development involves the application of a variational method to describe the *time dependence of approaches to the equilibrium state.*[12] This method is based on a proper choice of a dot product of vectors that gives a kinetic measure of the time it would take for a system to evolve from one configuration to another and a gradient operator such that operation of the gradient operator on the dot product produces the fastest rate of decrease of the free energy. From this description of this method it is apparent that the *method is applicable to systems that are varying their state parameters in an approach to equilibrium.* It does not apply to systems undergoing steady-state irreversible processes *in which the state parameters are invariant in time*.

However, often we may have a condition where there are steady-state fluxes, e.g. the flow of heat into a closed system equals that out of the closed system and is independent of time. In the situation of the previous paragraph the free energy of the system attempts to approach equilibrium at the fastest rate while changing its state parameters. In the second there is no change in the free energy of the system because its state parameters in each differential volume are maintained constant with respect to time. However, entropy is produced in the latter case because the heat flux out of the system occurs at a lower temperature than the same heat flux into the system. Thus, even though the entropy of the system itself is constant there is a production of entropy because the flux into a differential volume is at a higher temperature than that out of it. We shall in the next section provide relations for the entropy production rate and for those between the coefficients in the relations between forces and fluxes.

If the approach to equilibrium involves a series sequence of processes, then it is the slowest process that controls the overall rate. Usually, in this circumstance, the boundary conditions are not fixed, but are dependent upon the processes themselves. Hence, the processes involved in such a series sequence are not strictly independent of each other, i.e. they are coupled. In principle, coupled reactions should not be treated independently in a description of the kinetics of some overall process dependent upon them. However, it is possible to obtain useful approximations when the contribution of one process to the rate of decrease of the free energy overwhelms all the others to which it is coupled.

4. Thermodynamic theory of irreversible processes

The thermodynamic based consideration of irreversible processes dates back to 1854 when Thomson (Lord Kelvin) analyzed various thermoelectric phenomena.[13] It wasn't until 1931 when Onsager derived his "reciprocal relations" that the modern theory of irreversible processes had its rebirth and it has had an active development to date beginning with the systematic contributions of Prigogine,[14] DeGroot[15] and others.

We must first understand that the theory applies only to a limited range of irreversible phenomena that can be described in terms of the existence of linear relations between fluxes and forces. Although efforts are being made to apply the theory to the non-linear regime it is too early at this time to consider these results. Thus, the theory cannot be used to mandate that the relation between force and flux must be linear. Instead, only where this relation is linear can the theory be applied. We describe the generalized linear relation by the following equation

$$J_i = \Sigma\, L_{ik} X_k \quad \text{for } i = 1 \text{ to } n \tag{7.10}$$

where the L_{ik} are coefficients not dependent on the forces X_k. The fluxes and forces in these linear relations have the same tensor rank, which is a statement of the Curie principle that in an isotropic system, forces and fluxes of different tensor rank do not couple.

The fluxes and forces in equation 7.10 have not been defined. Whatever their definition they must yield the result that the sum over the product of the conjugate forces and fluxes equals the rate of production of entropy. Here we have used a concept that we have not yet defined, i.e. the rate of production of entropy. Let us consider the meaning in "the change in entropy of a system". We need first to define the system. For an open system that can exchange mass and heat with its surroundings we split the change in entropy in a process in the system into two parts: the change in entropy due to flow of matter and energy from or into the system ΔS_e, and change of entropy due to production inside the system ΔS_i. Since the change in entropy for the system can be evaluated from a knowledge of its state variables before and after the process, and the entropy associated with the exchange of matter and energy with the surroundings can also be evaluated, then the entropy change due to entropy production inside the system can also be evaluated. In general, the latter will be positive definite. Let the entropy production in the system be designated by $\Delta_i S$. Then our previous statement leads to the relation that the proper fluxes and forces in equation 7.10 satisfy

$$\Delta_i S/dt = \Sigma\, J_i X_i \tag{7.11}$$

Fluxes and forces that satisfy equation 7.11 will then have coefficients in equation 7.10 that satisfy Onsager's reciprocal relations, i.e.

$$L_{ik} = L_{ki} \tag{7.12}$$

The matrix of phenomenological coefficients in equation 7.10 is thus symmetric. Onsager's reciprocal relations follow from time reversal invariance of the microscopic equations of motion of the system of particles, which is a statement of the principle of "microscopic reversibility", and responsible for the detailed balancing of chemical reactions. Let us consider the latter for the simple case of reactions in which the rates of change of concentrations of the reactants is proportional to the concentrations of all the reactants. The reactions considered are schematically illustrated below.

Then

$$\begin{aligned} dc_A/dt &= -(k_1 + k_3')c_A + k_1'c_B + k_3c_C \\ dc_B/dt &= k_1c_A - (k_1' + k_2)c_B + k_2'c_C \\ dc_C/dt &= k_3'c_A + k_2c_B - (k_3 + k_2')c_C \end{aligned} \tag{7.13}$$

At equilibrium, the time derivatives equal zero and the equations can be rearranged to give

$$c_{Ae}/c_{Be} = [k_1'(k_3 + k_2') + k_2k_3]/[k_1(k_3 + k_2') + k_2'k_3']$$

$$c_{Be}/c_{Ce} = [k_2'(k_1 + k_3') + k_1k_3]/[k_2(k_1 + k_3') + k_1'k_3']$$

But, we can consider each separate reaction as well and for these separate reactions at equilibrium we would find

$$\begin{aligned} k_1c_{Ae} &= k_1'c_{Be} \\ k_2c_{Be} &= k_2'c_{Ce} \\ k_3c_{Ce} &= k_3'c_{Ae} \end{aligned} \tag{7.14}$$

Now, set $y_A = c_A - c_{Ae}$ and, hence $dy_A/dt = dc_A/dt$. Substituting, the latter relation in the first of the equations 7.13 with the corresponding expressions for y_B and y_C we obtain

$$dy_A/dt = -(k_1 + k_3')y_A + k_1'y_B + k_3y_C - (k_1 + k_3')c_{Ae} + k_1'c_{Be} + k_3c_{Ce}$$

The sum of the last three terms in the latter equation is equal to dc_{Ae}/dt and hence equals zero. Thus,

$$dy_A/dt = -(k_1 + k_3')y_A + k_1'y_B + k_3y_C \tag{7.15}$$

Now the chemical potential of A at concentration c_A can be related to that at c_{Ae} by $\mu_A - \mu_{Ae} = RT \log(c_a/c_{Ae}) = RT \log(1 + y_A/c_{Ae})$. For small deviations from equilibrium the logarithmic term is simply replaced by y_A/c_{Ae}. Thus, rearranging we find

$$y_A = -(\mu_A - \mu_{Ae})c_{Ae}/RT$$

Carrying out similar calculations to obtain expressions for y_B and y_C and substituting for these quantities in equation 7.15 and in the equivalent expressions for dy_B/dt and dy_C/dt then

$$\begin{aligned} dy_A/dt = \{&[(k_1 + k_3')c_{Ae}/RT][-(\mu_A - \mu_{Ae}) + [-k_1'c_{Be}/RT][-(\mu_B - \mu_{Be})] \\ &+ [-k_3c_{Ce}/RT][-(\mu_C - \mu_{Ce})]\} \end{aligned}$$

$$\begin{aligned} dy_B/dt = \{&[-k_1c_{Ae}/RT][-(\mu_A - \mu_{Ae})] + [(k_1' + k_2)c_{Be}/RT][-(\mu_B - \mu_{Be})] \\ &+ [-k_2'c_{Ce}/RT][-(\mu_C - \mu_{Ce})]\} \end{aligned}$$

$$\begin{aligned} dy_C/dt = \{&[-k_3'c_{Ae}/RT][-(\mu_a - \mu_{Ae})] + [-k_2c_{Be}/RT][-(\mu_B - \mu_{Be})] \\ &+ [(k_3 + k_2')c_{Ce}/RT][-(\mu_C - \mu_{Ce})]\} \end{aligned}$$

These relations can be compared to the linear relations consistent with the thermodynamic theory of irreversible processes

$$dy_A/dt = L_{11}X_1 + L_{12}X_2 + L_{13}X_3$$

$$dy_B/dt = L_{21}X_1 + L_{22}X_2 + L_{23}X_3$$

$$dy_C/dt = L_{31}X_1 + L_{32}X_2 + L_{33}X_3$$

Thus, $L_{12} = -k_1'c_{Be}/RT$ and $L_{21} = -k_1c_{Ae}/RT$, etc. But, at equilibrium, by detailed balancing of the individual reactions, as indicated by equation 7.14: $k_1c_{Ae} = k_1'c_{Be}$, etc. Hence, we find that $L_{12} = L_{21}$, etc. The Onsager relations hold for the linearized chemical reactions.

When the boundary conditions on a system are maintained constant with time the system will develop a steady-state condition in which the state variables at a point do not vary with time. For example, when a bar is maintained at temperature T_1 at one end and temperature T_2 at the other end, a steady-state distribution of temperature will develop along the bar. A consequence of the thermodynamic theory of irreversible processes (linear relations between forces and fluxes, Onsager relations) with constant values for the phenomenological coefficients L_{ik} is that this configuration-invariant steady state corresponds to the state of minimum entropy production rate. There has been a tendency in the literature to claim that all irreversible processes occur such that the rate of production of entropy is a minimum. This statement cannot be true because there are many irreversible processes that do not involve linear relations between fluxes and forces, nor have constant L_{ik} even if linear relations are applicable. Also these processes usually do not involve time invariant boundary conditions. (See References 14 and 15.)

Linear relations between fluxes and forces are applicable to some of the irreversible processes we will consider in the remainder of this book. The prime example of such a process is that of diffusion. Some growth situations obey linear laws, some do not. We will examine both cases and show where the thermodynamic theory can be of use and where not.

5. Computer simulation methods

A short description of each of the most used simulation methods is provided so that the reader may better understand their uses and limitations in the simulation of phenomenon involving time-dependent quantities.

5.1. Monte Carlo

The Monte Carlo method[16] is applicable to time-dependent phenomena as well as to materials in thermodynamic equilibrium. In the dynamic application the time scale is normalized so that N/τ_o single particle transitions are attempted in unit time, where N is the number of particles in the sample undergoing the dynamic process in the simulation. This time unit is called 1 MCS (Monte Carlo step per particle). It is necessary to be able to calculate the potential energy for all configurations of the system. If the potential energy is finite for each configuration then from the finiteness of the system it is ergodic. The probability that a configuration X occurs at the time t, P(X, t), is given by a rate equation as follows

$$dP(X, t)/dt = \Sigma_{X'} W(X \rightarrow X')P(X, t) + \Sigma_{X'} W(X' \rightarrow X)P(X', t)$$

where W is the transition probability and one choice that satisfies detailed balance is

$$w(X \rightarrow X') = \begin{cases} \tau_o^{-1} & \Delta H < 0 \\ \tau_o^{-1}\exp(-\Delta H/kT) & \Delta H > 0 \end{cases}$$

where ΔH is the difference in potential energy between the two configurations. The algorithm one uses to evaluate the rate of a process depends upon the process being simulated. For example, for an Ising model or its equivalent where S_i represents the spin at site i an algorithm governing the flipping of spins (the change in order at the site) is as follows (this algorithm is reproduced precisely from Reference 17):

1. Select one lattice site i at which the S_i is considered for flipping ($S_i \rightarrow -S_i$).

2. Compute the energy change ΔH associated with that flip.

3. Calculate the transition probability $\tau_{oo}W$ for that flip.

4. Draw a random number p uniformly distributed between zero and unity.

5. If $p < \tau_{oo}W$ *flip the spin, otherwise do not flip it. In any case, the configuration of the spins obtained in this way at the end of step 5 is counted as a "new configuration".*

6. Analyze the resulting configuration as desired, store its properties to calculate the necessary averages. For example, if we are just interested in the (unnormalized) magnetization M_{tot} *and its distribution* $P(M_{tot})$*, we may update* M_{tot} *by replacing* M_{tot} *by* $M_{tot} + 2S_i$*, and then replacing* $P(M_{tot})$ *by* $P(M_{tot}) + 1$ *(appropriate initial values before the process starts are set to* $M_{tot} = L^3$*;* $P(M') = 0$*;* $P(M')$ *being an array where* M' *can take integer values from* $-L^3$ *to* $+L^3$*). Here L is one dimension of the box in integer value.*

Another algorithm used in determining the rate of advance of a grain boundary is shown in Figure 7.3. This algorithm biases the time at a boundary site according to

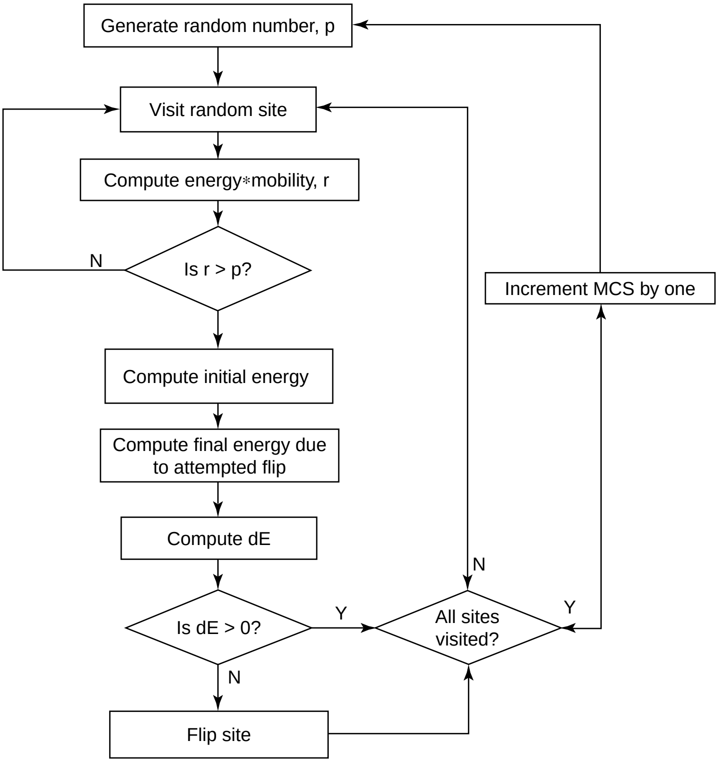

Generate random number, p
Visit random site
Compute energy∗mobility, r
N
Is r > p?
Increment MCS by one
Compute initial energy
Compute final energy due to attempted flip
Compute dE
N
Y
Is dE > 0?
Y
All sites visited?
N
Flip site

Figure 7.3. A proposed algorithm for use in grain boundary migration Monte Carlo simulation. From Modell. Simul. Mater. Sci. Eng. 10, 227(2002) with permission. © 2002 Institute of Physics Publishing, London.

the value of the product of the mobility by the boundary energy at the site r. It was developed to provide an alternative to an algorithm that is based on each boundary site being visited for each MCS with the probability of a flip at a specific site depending upon the product of the mobility and the boundary energy.[18]

It is apparent from the above that in order to compare results with experiment, realistic values of the energies of the configurations and for the time period for a MCS are needed. However, useful results can also be obtained without needing to quantize the results with respect to experiment. It is possible to determine the effect of altering the values of parameters on the rates, one parameter at a time. This knowledge itself is useful to determine which of the parameters are the most significant ones, a result that requires much effort to determine experimentally, especially when many parameters affect the rate process.

The Monte Carlo process is useful for simulating slow processes relative to that for an atomic jump whereas the molecular dynamics method is better suited to processes that occur in times less than about 10^{-10} seconds.

5.2. Molecular dynamics

Molecular dynamics[19] solves Newton's equations of motion applied to an interacting many-body system. Hence, it is necessary to have a knowledge of the force each particle experiences in this many-body system. This requires in turn a knowledge of the interparticle potentials and the distances between particles. Different relations apply to different classes of applications. One can quickly determine which relations to use for any class by quizzing Google. In less than 1 minute I determined that the following force fields are available to use in molecular dynamic problems involving molecules: MM2, MM3, AMBER, UU, MMFF, OPLSA, REAX FF, CHARM. Also, the MS-Q force field has been developed for metal oxides. Interatomic potentials developed from ab initio studies are available for metallic and other materials from which force fields may be derived. Molecular dynamics is suited to evaluate mobilities of processes involving individual atom jumps but the rates of reactions are normally too slow to simulate directly via molecular dynamics. Equilibrium properties of very small units, e.g. nanoparticles, may also be usefully studied via molecular dynamics.

5.3. Phase field

The phase field method[20] is based on the use of two relations describing the time dependence of conservative and non-conservative parameters. These relations are the coupled Cahn-Hilliard non-linear diffusion equation for a conserved field c given by

$$\partial c/\partial t = \nabla \cdot \{M\nabla(\delta F/\delta c)\}$$

and the time-dependent Ginzburg–Landau equation for a non-conserved field φ_ι given by

$$\partial\varphi_\iota/\partial t = -L_i(\delta F/\delta\varphi_\iota)$$

where L_i is the relaxation constant associated with the order parameter φ_ι, M is an atomic mobility and F is the total free energy. Numerical solutions to these equations are provided by the computer.

There are many phase field models. However, all of them when properly used give precisely equivalent results.[21] In application to sharp interface problems it is necessary that the asymptotic approach of the diffuse interface to the sharp interface converges adequately for the results to be correct.[21]

Briefly, when used in a free boundary problem (e.g. phase transformation involving motion of an interface between product and parent phases) the free energy may contain a double well potential function of order and composition, interface energies represented by gradient energies, and strain energy terms if applicable, and each minimum in these wells representing the parent and product phase free energy densities, which may have different order parameters, e.g. -1, 1, and compositions.[20] The interface between the phases is diffuse with the order and composition parameters varying between the value for one phase to that for the other phase within the interface. An analysis of the asymptotic convergence of the diffuse to a sharp interface is provided by the work of Karma and Rappel.[21]

The phase field concept at this writing is being made more efficient and will undoubtedly gain in power and applicability with time.

5.4. Level set

This method is preferred for sharp interface problems where it is desired to know details about this interface. A description of the method and its use in the dendrite solidification problem has been provided recently.[22] The uniqueness of this method resides in the use of a procedure for the second order accurate symmetric discretization of the Poisson equation. A particularly clear exposition of the method as applied to solidification is given by Kim et al.[22]

References

1. J.H. Wiener, Proceedings of Sixth U.S. National Congress of Applied Mechanics, A.S.M.E., New York, 1970, pp. 62–77.
2. S. Glasstone, K.J. Laidler and H. Eyring, **The Theory of Rate Processes**, McGraw-Hill, New York, 1941.
3. G.H. Vineyard, J. Phys. Chem. Solid. 3, 121(1957).

4. G. Golub and C. Van Loan, **Matrix Computations**, The Johns Hopkins University Press, Baltimore, 1989.
5. N.F. Mott and M.J. Littleton, Trans. Faraday Soc. 34, 485(1938).
6. J.H. Harding, Rep. Prog. Phys. 53, 1403(1990).
7. M.P. Allen and D.J. Tildesley, **Computer Simulation of Liquids**, 1st edition, Oxford University Press, Oxford, 1989.
8. M. Parrinello and A. Rahman, Phys. Rev. Lett. 45, 1196(1980).
9. W. Smith and T.R. Forester, J. Mol. Graphics 14, 136(1996).
10. M. Athenes, P. Bellon, G. Martin and F. Haider, Acta Mater. 44, 4739(1996).
11. D.T. Gillespie, J. Comp. Phys. 22, 403(1976).
12. W.C. Carter, J.E. Taylor and J.W. Cahn, J. Metals 49 (12), 30(1997).
13. W. Thomson, Proc. Roy. Soc. Edin. 3, 225(1854).
14. I. Prigogine, **Etude Thermodynamique Des Phenomenes Irreversibles**, Dunod, Paris and Desoer, Liege, 1947.
15. S.R. DeGroot, **Thermodynamics of Irreversible Processes**, North-Holland, Amsterdam, 1951.
16. K. Binder and A. Baumgaertner, **The Monte Carlo Method in Condensed Matter Physics**, Springer-Verlag, New York, 1992.
17. K. Binder, Rep. Prog. Phys. 60, 487(1997).
18. B. Radhakrishnan and T. Zacharia, Model. Simul. Mater. Sci. Eng. 10, 227(2002).
19. D.C. Rapaport, **The Art of Molecular Dynamics Simulation**, Cambridge University Press, New York, 1996.
20. L.Q. Chen, Ann. Rev. Mater. Res. 32, 113(2002).
21. A. Karma and W.-J. Rappel, Phys. Rev. E 53, 3017(1995); E 57, 4323(1998).
22. F. Gibou et al., J. Sci. Comput. 19, 183(2003). Y.-T. Kim, N. Goldenfeld, and J. Dantzig, Phys. Rev. E 62, 2471(2000).

Bibliography

1. J.W. Christian, **Theory of Transformations in Metals and Alloys**, Pergamon Press, New York, 1965, pp. 16–22, 80–94.
2. J.H. Wiener, **Statistical Mechanics of Elasticity**, John Wiley, New York, 1983, Chapter 7.
3. S.R. DeGroot and P. Mazur, **Non-Equilibrium Thermodynamics**, North-Holland, Amsterdam, 1962.
4. Y.L. Yao, **Irreversible Thermodynamics**, Science Press, Beijing, 1981.
5. K. Binder, **Monte Carlo Methods in Statistical Physics**, Springer, Berlin, 1986.
6. D. Frenkel and B. Smit, **Understanding Molecular Simulation: From Algorithms to Applications**, Academic, New York, 1996.

Problems

1. Given an activation energy of 1 eV. How much slower would the reaction be at 500ºK for an activation energy 10% larger?

2. Consider the relaxation process of recovery after plastic deformation which proceeds at a rate proportional to the stored excess energy due to the plastic deformation. Comment upon the existence or lack of existence of a coupling of this process to that of heat flow along a temperature gradient.
3. Comment on the existence of a coupling between the flux of electric charge and a flux of matter in a metallic conductor.

CHAPTER VIII

Diffusion

Introduction

Most kinetic phenomena in solids involve diffusion, or a unit step very similar to that operating in diffusion. It is appropriate therefore to start our study of kinetic phenomena with a study of the fundamentals involved in diffusion in various materials. We cannot provide a complete description of diffusion because, as with nearly every other chapter in this book, such a complete description requires a book in itself. This chapter begins with a phenomenological treatment of linear processes that is based on the thermodynamic theory of irreversible processes. This very general approach is made specific by application to diffusion in a binary alloy. Diffusion in ionic crystals and semiconductors follows on the basis of the relations developed for metals and the thermodynamics of defects in these materials. Then diffusion along high diffusivity regions, such as grain boundaries, is discussed. Finally, the current computer assisted strategies for extending the basic concepts described in this chapter to problems involving multicomponent diffusion in technological applications are discussed. Also, diffusion of polymer strands is treated briefly. Diffusion in liquids is not considered in this chapter.

1. Phenomenological basis

1.1. Intrinsic diffusivities

It is often possible to relate fluxes of matter, heat and charge to corresponding forces linearly, as shown in equation 8.1

$$J_i = \Sigma_k L_{ik} X_k \quad i = 1, \ldots n \tag{8.1}$$

where J_i is the number or quantity of i passing unit area per unit time, L_{ik} is a conductance relating the kth generalized force X_k to the flux J_i. In the case of matter flow, the generalized force is the negative gradient of the chemical potential (i.e. $X_k = -\nabla\mu_k$). (In a network solid, $-J_i = \Sigma_j B_{ij} X_{jk}$ $(i = 1, \ldots n)$ with $X = \nabla M$, where M is the diffusion potential.) Onsager has shown that the matrix of conductance coefficients in equation 8.1 is symmetric

$$L_{ik} = L_{ki} \tag{8.2}$$

We choose a binary alloy to illustrate the use of these relations to describe matter transport in the remainder of this section. In these equations, we assume that vacancies are the species denoted by the subscript 3, and the other two components are the atom species in a binary system

$$\begin{aligned} {}^*J_1 &= {}^*L_{11}X_1 + {}^*L_{12}X_2 + {}^*L_{13}X_3 \\ {}^*J_2 &= {}^*L_{21}X_1 + {}^*L_{22}X_2 + {}^*L_{23}X_3 \\ {}^*J_3 &= {}^*L_{31}X_1 + {}^*L_{32}X_2 + {}^*L_{33}X_3 \end{aligned} \tag{8.3}$$

We now act, by setting $\Sigma J_i = 0$, to define a coordinate system relative to which the J's are measured. Using this definition, then

$$\begin{aligned} {}^*L_{11} + {}^*L_{21} + {}^*L_{31} &= 0 \\ {}^*L_{12} + {}^*L_{22} + {}^*L_{32} &= 0 \\ {}^*L_{13} + {}^*L_{23} + {}^*L_{33} &= 0 \end{aligned}$$

by the independence of the X_i's. Also, from equation 8.2 and substitution into equation 8.1 yields

$$\begin{aligned} {}^*J_1 &= {}^*L_{11}(X_1 - X_3) + {}^*L_{12}(X_2 - X_3) \\ {}^*J_2 &= {}^*L_{21}(X_1 - X_3) + {}^*L_{22}(X_2 - X_3) \\ {}^*J_3 &= {}^*L_{31}(X_1 - X_3) + {}^*L_{32}(X_2 - X_3) \end{aligned}$$

But, $X_1 = -\nabla\mu_1$, $X_2 = -\nabla\mu_2$ and we make the assumption that there are sufficient sources and sinks for vacancies to maintain $X_3 = 0$, i.e. $X_3 = -\nabla\mu_{\text{vacancy}}$. Further, we assume constant molar volume, i.e. $C_1 + C_2 = C$ is independent of position. Then, if we consider a variation of composition only in one direction x, we obtain

$$\begin{aligned} {}^*J_1 &= [{}^*L_{11}\, d\mu_1/dC_1 - {}^*L_{12}\, d\mu_2/dC_2](-dC_1/dx) = D_1(-dC_1/dx) \\ {}^*J_2 &= [{}^*L_{22}\, d\mu_2/dC_2 - {}^*L_{12}\, d\mu_1/dC_1](-dC_2/dx) = D_2(-dC_2/dx) \\ {}^*J_3 &= [{}^*L_{31}\, d\mu_1/dC_1 - {}^*L_{32}\, d\mu_2/dC_2](-dC_1/dx) = -({}^*J_1 + {}^*J_2) = Cv \end{aligned} \tag{8.4}$$

This relation in which the flux is proportional to the negative gradient of the concentration is known as Fick's first law of diffusion and was an empirical discovery long before the thermodynamic theory of irreversible processes was developed. We shall now interpret these results via independent considerations.

Let there be a diffusing system in which there are inert markers (i.e. markers that do not interact with the matrix in which they are embedded). Let us now measure the flux of components 1 and 2 relative to these markers and denote these by ${}^*J_1$ and ${}^*J_2$. Now, these markers may move with respect to a fixed laboratory coordinate system with a velocity v. We conserve matter, assuming that $C_1 + C_2 = C$, is a constant, independent of the distance x. Suppose that there is a concentration gradient in the diffusing system, that a marker at the distance x is moving with a velocity v to the right, and the instantaneous concentration of species at this point is C_1, as shown in Figure 8.1. If we now consider a differential volume of unit area normal to x and thickness dx, that is fixed in the laboratory coordinate system at x, and consider the change in the concentration of 1 that takes place in the time dt, then we obtain

$$(\partial C_1/\partial t)dx\,dt = [{}^*J_1(x) - ({}^*J_1(x) + (\partial {}^*J_1/\partial x)dx)]dt - [\partial(vC_1)/\partial x]dx\,dt$$

Substitution of ${}^*J_1 = -D_1(\partial C_1/\partial x)$ yields

$$(\partial C_1/\partial t) = \partial[D_1(\partial C_1/\partial x) - vC_1]/\partial x \tag{8.5a}$$

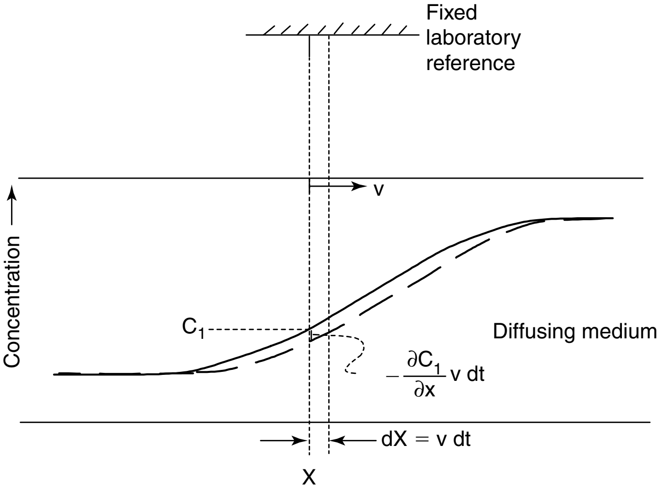


Figure 8.1. Illustration of change in composition, as observed in fixed laboratory frame of reference, due to motion of a diffusing medium.

Proceeding in a similar manner for component 2, we can obtain

$$(\partial C_2/\partial t) = \partial[D_2(\partial C_2/\partial x) - vC_2]/\partial x \tag{8.5b}$$

Summing equations 8.5a and 8.5b yields

$$\partial(C_1 + C_2)/\partial t = 0 = \partial[-({}^*J_1 + {}^*J_2) - vC]/\partial x$$

This result holds for all x and hence the integration constant must equal zero (i.e. as $x \rightarrow \pm\infty$, ${}^*J_1$, ${}^*J_2$, $v \rightarrow 0$). Hence

$$vC = -({}^*J_1 + {}^*J_2)$$

Comparison with equation 8.4 yields the interpretation that the coordinate system relative to which the ${}^*J$'s are measured at a given position, in equations 8.1–8.4, is fixed relative to the inert marker at that position, which itself moves with respect to the laboratory coordinate system! In this case, note that $D_1 \neq D_2$. These diffusivities, D_1 and D_2, are called intrinsic diffusivities. Further, note that when $D_1 \neq D_2$, there is a net flux of vacancies in the same direction as the marker motion, i.e. ${}^*J_3 = vC$.

The effect that has just been described may be given a physically based interpretation as follows. Vacancies are generated on one side of the marker and they annihilate on the other side, to which they move. To conserve molar volume, the marker moves in the direction of the vacancy annihilation, i.e. atoms move in the opposite direction.* The vacancy sinks and sources may be considered to be edge dislocation jogs, so that the annihilation of vacancies causes extra half planes of dislocations to disappear, while their generation causes such half planes to grow, corresponding to the process of dislocation climb. The disappearance of half planes on one side of the diffusion couple and growth on the other side results in a bulk movement of the matter in between in the direction of the side in which the half planes disappear. This phenomenon, which was discovered by E. Kirkendall has been called the Kirkendall Effect.

It is worth calling attention to the fact that in the above, although the concentration of vacancies was assumed to be the equilibrium concentration everywhere (i.e. $X_3 = 0$), nevertheless there exists a non-zero value of the vacancy flux ${}^*J_3$. The existence of vacancy sources and sinks, distributed throughout the diffusion couple, is implicitly assumed in the above analysis. Without such distributed sources and sinks, it would not be possible to maintain the vacancy concentration at the equilibrium value, and the above equations would have to be modified to take this fact into account. Usually, any development of a supersaturation in the vacancy concentration will lead to a tendency to nucleate voids, except in very clean and dislocation free single crystals. Such voids can be found in actual diffusion couples when the difference in the intrinsic diffusivities leads to the pumping of vacancies into a region at a rate faster than they can be annihilated at dislocation sinks. The detailed equations for vacancy migration must therefore include divergence terms to account for the generation and annihilation of the vacancies. In the above, we have avoided this problem by use of the assumption of constant molar volume, and consideration of the diffusion of the matter species only.

---

* The diffusion couple is fixed relative to the laboratory, far from the interface, where the concentration gradients equal zero.

1.2. Chemical diffusivity

By measuring the flux of the species 1 and 2 relative to a fixed laboratory coordinate system, instead of with respect to the markers, then new diffusivities are defined. To obtain these new diffusivities let us proceed as follows

$$^0D_1 = -J_1/(dC_1/dx) \quad ^0D_2 = -J_2/(dC_2/dx)$$

where the J's are measured relative to the fixed laboratory coordinate system, and the ends of the diffusion couple are fixed relative to the laboratory coordinate system (as they were in the previous case as well, in the regions where the gradients in the composition were equal to zero.) Note, that to distinguish these fluxes from the previous ones they do not have asterisks. In the present case, $J_1 + J_2 = 0$, the net number of 1 type atoms moving to one side past a plane fixed relative to the laboratory coordinate system equals the net number of 2 type atoms moving past this plane in the opposite direction. We continue the assumption of constant molar volume. Thus, since $C_1 + C_2 = C =$ constant then $(\partial C_1/\partial x) = -(\partial C_2/\partial x)$. Since the sum of the fluxes of 1 and 2 equals zero, then by the above we obtain $^0D_1 = {^0D_2} = {^0D}$. Consider now the change in C_1 in a region of volume of unit area normal to x, the direction in which diffusion occurs, and distance dx in the time dt

$$(\partial C_1/\partial t)dx\,dt = [J_1(x) - (J_1(x) + (\partial J_1/\partial x)dx]dt \quad \text{or}$$
$$(\partial C_1/\partial t) = \partial(^0D(\partial C_1/\partial x))/\partial x \tag{8.6}$$

We shall make use of this relation in a little while. First, let us substitute in equation 8.5a the following relation for the marker velocity v

$$\begin{aligned} v = -(^*J_1 + {^*J_2})/C &= -[D_1(-\partial C_1/\partial x) + D_2(-\partial C_2/\partial x)]/C \\ &= [D_1 - D_2](\partial C_1/\partial x)/C \end{aligned}$$

to obtain

$$(\partial C_1/\partial t) = \partial[D_1(\partial C_1/\partial x) - C_1 v]/\partial x$$

$$(\partial C_1/\partial t) = \partial[D_1(1 - C_1/C)(\partial C_1/\partial x) - (C_1/C)D_2(\partial C_2/\partial x)]/\partial x$$

But, $C_1/C = N_1$, the mole fraction of component 1, and $\partial C_1/\partial x = -\partial C_2/\partial x$. Hence

$$\partial C_1/\partial t = \partial[(N_2 D_1 + N_1 D_2)(\partial C_1/\partial x)]/\partial x \tag{8.7}$$

Comparison of equations 8.6 and 8.7 yields the result

$$^0D = N_2 D_1 + N_1 D_2 \tag{8.8}$$

The diffusivity ^0D is called the chemical diffusivity and equation 8.8 relates it to the intrinsic diffusivities.

Let us consider how these diffusivities are measured in an alloy system. First, let us consider the measurement of ^0D. Fick's second law of diffusion, as given in equation 8.6, which is a second order partial differential equation, can be solved by using the Boltzmann integration factor $\lambda(C) = x/t^{1/2}$. Note that

$$(\partial C_1/\partial t) = [dC_1/d\lambda](\partial\lambda/\partial t) = [dC_1/d\lambda](-\lambda/2t)$$

$$(\partial/\partial x) = [d/d\lambda](\partial\lambda/\partial x) = t^{-1/2}[d/d\lambda]$$

Substitution yields

$$(-\lambda/2)[dC_1/d\lambda] = [d/d\lambda](^0D[dC_1/d\lambda]),$$

a total differential equation in λ and C_1. We now integrate to obtain

$$-(1/2)\int_{C_{1o}}^{C_{1f}} \lambda\, dC_1 = {}^0D\left\{[dC_1/d\lambda]_{C_{1f}} - [dC_1/d\lambda]_{C_{1o}}\right\}$$

For t constant, we replace λ to obtain

$$-(1/2t)\int_{C_{1o}}^{C_{1f}} x\, dC_1 = {}^0D\left\{[\partial C_1/\partial x]_{C_{1f}} - [\partial C_1/\partial x]_{C_{1o}}\right\} \tag{8.9}$$

The significance of this equation can be appreciated on referring to Figure 8.2. To define the zero of x, we take the integration limits to equal zero, i.e.

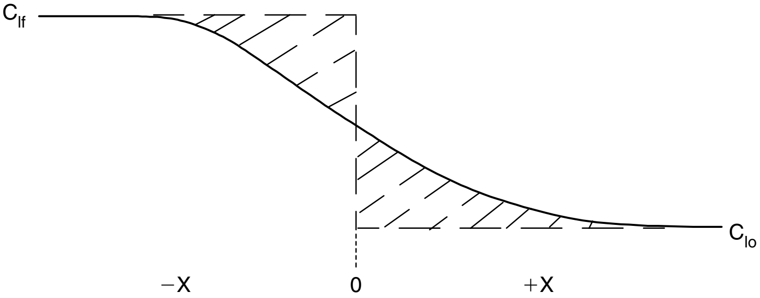


Figure 8.2. Illustrating the Matano interface at x = 0. The cross-hatched areas are equal in area in the plot of concentration versus distance.

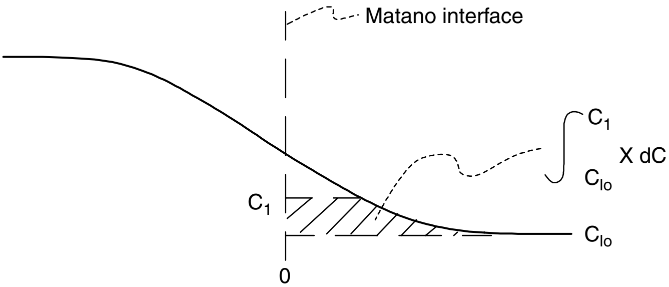


Figure 8.3. Illustrating equation 8.10.

$[\partial C_1/\partial x]_{C_{1f}} = [\partial C_1/\partial x]_{C_{1o}} = 0$. This serves to set the cross-hatched area to the left of x = 0 equal to the cross-hatched area to the right of this interface, which is called the Matano interface. ^{0}D can now be evaluated using equation 8.9, except now the upper integration limit is allowed to float between C_{1o} and C_{1f}, so that, as shown in Figure 8.3 and equation 8.10, all that is needed to define ^{0}D(C) is a measurement of the cross-hatched area in Figure 8.3, the slope ($\partial C_1/\partial x$) at the composition C_1, and the diffusion time required to obtain the concentration–penetration profile shown in Figure 8.3

$$-(1/2t)\int_{C_{1o}}^{C_1} x\mathrm{d}C_1 = {}^0D\left\{[\partial C_1/\partial x]_{C_1} - [\partial C_1/\partial x]_{C_{1o}}\right\} = {}^0D[\partial C_1/\partial x]_{C_1} \quad (8.10)$$

With ^{0}D measured, all that is necessary to evaluate the intrinsic diffusivities is a measurement of the marker velocity corresponding to the concentration C_1. This can be obtained by measuring the distances between the various markers in a diffusion couple out to the ends, where the concentration gradients remain equal to zero throughout the experiment, as a function of time. In the above we have provided the theoretical bases for measurements of the intrinsic and chemical diffusivities in a binary alloy system that obeys the conditions explicitly stated in the assumptions. Next we shall relate the self-diffusivity of a pure metal to that in an alloy and later we shall provide the bases for measurement of the diffusivities of the ionic species in an ionic crystal.

1.3. Darken's relations

Let us extract from equations 8.4 the following expressions for the intrinsic diffusivities

$$\begin{aligned} D_1 &= [{}^*L_{11}\, d\mu_1/dC_1 - {}^*L_{12}\, d\mu_2/dC_2] \\ D_2 &= [{}^*L_{22}\, d\mu_2/dC_2 - {}^*L_{12}\, d\mu_1/dC_1] \end{aligned} \qquad (8.11)$$

Since $\mu_i = \mu_i^o + kT \ln \gamma_i N_i$ then

$$\begin{aligned} d\mu_i/dC_i = C^{-1}\, d\mu_i/dN_i &= C^{-1}kT\{1/N_i + (1/N_i)d \ln \gamma_i/d \ln N_i\} \\ &= (kT/C_i)[1 + d \ln \gamma_i/d \ln N_i] \end{aligned}$$

But, $d \ln \gamma_1/d \ln N_1 = d \ln \gamma_2/d \ln N_2$ and setting $\varphi = 1 + d \ln \gamma_1/d \ln N_1$ then

$$\begin{aligned} D_1 &= kT\varphi[{}^*L_{11}/C_1 - {}^*L_{12}/C_2] \\ D_2 &= kT\varphi[{}^*L_{22}/C_2 - {}^*L_{12}/C_1] \end{aligned} \qquad (8.12)$$

At this point we make the assumption that the cross-terms can be neglected, i.e. we assume $L_{ij} = 0$. Further, we make use of the relation between flux and average velocity: $J_i = C_i\langle v_i\rangle = L_{ii}X_i$. We now note that the mobility of a species is defined as the average velocity per driving force acting to move the species, i.e.

$$M_i = \langle v_i\rangle/X_i = L_{ii}/C_i$$

Thus, we can set down the intrinsic diffusivities in terms of the mobilities as

$$D_1 = M_1kT\varphi$$

But, the mobility of a species is independent of the driving force. Hence, $M_1 = M_{1*}$, where the latter is the mobility of the radioactive tracer of component 1. Since, the tracer diffusivity is determined in an alloy of constant composition C_1, using only dilute amounts of the tracer species, then the corresponding $\varphi = 1$ and $D_{1*} = M_1kT$. Hence, the result is $D_1 = D_{1*}\varphi$. Similarly, for component 2. In general, the Darken's relations are

$$D_i = D_{i*}\varphi \tag{8.13}$$

It can also be shown that the second derivative of the molar free energy equals $(d\mu_A/dC_A)/C_B$. Hence, at inflection points in the molar free energy–composition dependence, the spinodal compositions, where the second derivative of the molar free energy equals zero, the Darken thermodynamic factor will equal zero. By equation 8.8 and the above expression for the intrinsic diffusivity in terms of the tracer diffusivity this result will lead to a zero value for the chemical diffusivity! Volkl and Alefeld[1] have found an experimental situation illustrating this effect.

If the cross-terms are not neglected Manning[2] has shown that in a random solid solution containing A and B atoms that an additional factor multiplies the right-hand side of equation 8.8 given by

$$1 + [2C_AC_B(D_{A*} - D_{B*})^2]/[M_o(C_AD_{A*} + C_BD_{B*})(C_AD_{B*} + C_BD_{A*})] \tag{8.14}$$

where M_o is defined by the self-diffusion correlation factor $f_o = M_o/(M_o + 2)$ for a vacancy mechanism. This multiplicative term is called a vacancy wind term since it reflects the coupling between the transport of components A and B through the vacancy flux.

1.4. Additional driving forces

In the above we have considered only the negative gradient of the chemical potential as a driving force for diffusion. However, other gradients can interact with the diffusing species. In particular, the negative gradients of the temperature and the electric potential provide driving forces for the flux of matter. The general phenomenological relations in these cases can be written as

$$\begin{aligned}
J_1 &= L_{11}[-\nabla(\mu_1/T)] + L_{1q}[-\nabla(1/T)] + L_{1\phi}[-\nabla(\Phi\,/T)] \\
J_q &= L_{1q}[-\nabla(\mu_1/T)] + L_{qq}[-\nabla(1/T)] + L_{q\phi}[-\nabla(\Phi/T)] \qquad (8.15) \\
J_e &= L_{1\phi}[-\nabla(\mu_1/T)] + L_{q\phi}[-\nabla(1/T)] + L_{\phi\phi}[-\nabla(\Phi/T)]
\end{aligned}$$

More detailed development of these relations can be found in the literature.[3]

A variety of phenomena are associated with different combinations of constraints on the fluxes and forces as illustrated in Table 8.1. We shall discuss the effect

Table 8.1. Cross effects.

Constraints	Phenomena
$\nabla(\Phi/T) = \nabla(1/T) = 0$	$J_q/J_1 = L_{1q}/L_{11}$ = Energy of transfer
$\nabla(\Phi/T) = \nabla(\mu_1/T) = 0$	$J_1/J_q = L_{1q}/L_{qq}$ = Phonon drag
$\nabla(\Phi/T) = \nabla(1/T) = 0$	$J_1/J_e = L_{1\Phi}/L_{\Phi\Phi}$ = Electron drag
$\nabla(\Phi/T) = \nabla(\mu_1/T) = 0$	$J_e/J_q = L_{q\Phi}/L_{qq}$ = Thomson coefficient
$\nabla(\mu_1/T) = \nabla(1/T) = 0$	$J_q/J_e = L_{q\Phi}/L_{\Phi\Phi}$ = Thomson heat
$J_e = 0$	$\lvert\nabla(\Phi/T)/\nabla(1/T)\rvert = -L_{q\Phi}/L_{\Phi\Phi}$ = Seebeck effect

Seebeck effect is negative of Thomson heat (or Peltier heat) for a thermocouple.

Table 8.2. Relations between diffusivities and phenomenological coefficients for ionic crystals.

For $q_A = q_B$

$$D_B = [(L_{AA}L_{BB} - L_{AB}L_{BA})kT]/[NC_B(L_{AA} + L_{AB} + L_{BA} + L_{BB})] \quad (1)$$

When $L_{AA} \gg L_{AB}$ and L_{BB} then $D_B = kT\, L_{BB}/C_B$

For $2q_A = q_B$

$$D_B = \frac{2kT(L_{AA}L_{BB} - L_{AB}L_{BA})(X_v' - X_k)}{N(L_{AA} + 2L_{AB} + 2L_{BA} + 4L_{BB})(2X_vC_B - X_k^2 - C_B^2)} \quad (2)$$

When $L_{AA} \gg L_{AB}$ and L_{BB} and $C_B \ll X_v$ then (2) simplifies to (1). When in other limit, $C_B \gg X_v$ then $D_B = 2kT\, L_{BB}/C_B(1 + p)$, where p is the fraction of impurity ions associated with vacancies, i.e. $p = X_k/C_B$. Here q_A and q_B are the charges on the ions, X_v is the vacancy atom fraction, X_k is the concentration of impurity–vacancy pairs and C_B is the concentration of the impurity B.

of electron drag and phonon drag in the production of failures in integrated circuits later. Thus, these cross-effects can be significant in many practical situations.

1.5. Applications of phenomenological equations of irreversible thermodynamics to ionic crystals

Howard and Lidiard[4] have given a description of diffusion in ionic crystals in terms of relations (8.1) and (8.2). Although the diffusivities applicable to a lattice depend on the nature of the associated defects, the general result yielding Fick's law description is valid for ionic crystals. Table 8.2 collects some of Howard and Lidiard's results.

2. Mechanisms of diffusion

In the above we have implicitly assumed that the mechanism of diffusion is via the exchange of atoms with thermal vacancies. Of course, this is not the only possible

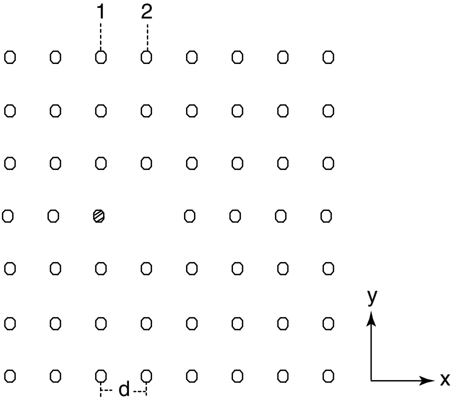


Figure 8.4. Schematic lattice defining atomic planes, interplanar spacing d and directions.

mechanism of diffusion. Among the others are direct exchange of neighboring atoms, cyclic exchange along a closed loop of nearest-neighbor atoms, interstitialcy movement in which an interstitial sited atom bumps a neighboring substitutional sited atom into an interstitial position with the stable arrangement of the interstitial defect being either a split interstitial (two atoms centered about a substitutional site) or a crowdion (n atoms along a close-packed row containing $n - 1$ substitutional sites), vacancy aggregates which aid diffusion of atoms much as does the isolated vacancy by exchange of a component vacancy with a neighboring atom. In most metals we need only consider the vacancy and the divacancy as the major defects that control diffusion. However, in some metals that exhibit abnormally fast solute diffusion, some solute may diffuse interstitially or via some combination of interstitial and vacancy mechanisms. In ionic solids and semiconductors, we need to consider both vacancy and interstitial diffusion mechanisms.

2.1. Metals

2.1.1. Pure metals

We consider the mechanisms of diffusion in order to provide detailed expressions for the diffusivity that cannot be obtained from the phenomenological description of diffusion. The case of self diffusion in a pure metal is particularly instructive. Consider a crystal as shown in Figure 8.4 and diffusion in the x direction. The interatomic plane spacing in the x direction is d, the atomic concentration of radioactive solute A* in the plane corresponding to position x is X_{A^*}. We make the assumption that the concentration of vacancies is everywhere that given by thermal

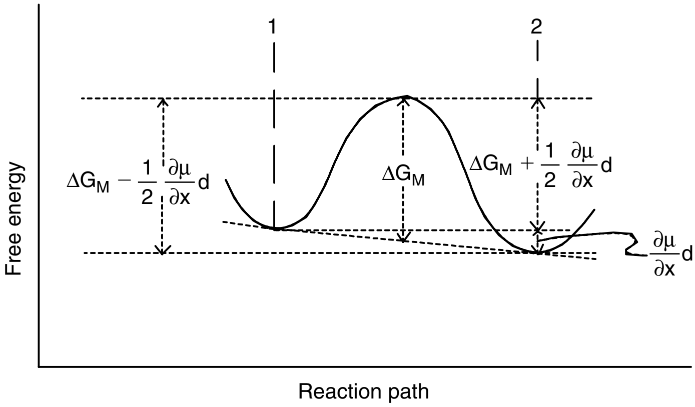

1
2
$\Delta G_M - \frac{1}{2}\frac{\partial\mu}{\partial x}d$
ΔG_M
$\Delta G_M + \frac{1}{2}\frac{\partial\mu}{\partial x}d$
$\frac{\partial\mu}{\partial x}d$
Free energy
Reaction path

Figure 8.5. Defining free energy as a function of reaction path and activation energies for forward and backward jumps.

equilibrium and that there is a gradient of A* along the x direction. Figure 8.5 illustrates the dependence of the free energy in the course of exchange of the A* atom in plane (1) with a nearest-neighbor vacancy in plane (2). From this figure the effective activation free energy for this exchange is

$$\Delta G_m + (1/2)(\partial\mu/\partial x)d \tag{8.16}$$

The flux of A* atoms originating in plane (1) that traverse the dividing plane between the atomic planes (1) and (2) in moving to the right is given by

$$J_{A*\rightarrow} = C_{A*}\, dm\, X_v \nu \exp[-(\Delta G_m + (1/2)(\partial\mu/\partial x)d)/kT]$$

where C_{A*} is the concentration of A* atoms per unit volume at the position x, $C_{A*}d$ is then the number of A* atoms per unit area of plane (1), m is the number of nearest neighbors in plane (2) to an atom in plane (1), X_v is the equilibrium atom fraction of vacancies everywhere. In the above relation, $C_{A*}\, dm\, X_v$ represents the number of A* − V pairs, where the A* atom is in plane (1) and the vacancy in plane (2), per unit area of plane, ν is the frequency of attempts of such pairs to interchange sites and the exponential term is the fraction of such attempts which are successful. In a similar way, the flux of A* atoms originating in plane (1) which exchange places with vacancies that lie in the plane to the left of plane (1) is given by

$$J_{A*\leftarrow} = C_{A*}\, dm\, X_v \nu \exp[-(\Delta G_m - (1/2)(\partial\mu/\partial x)d)/kT]$$

The net flux of A* atoms moving to the right out of plane (1) is then

$$J_{A^*,\ net} = J_{A^*\rightarrow} - J_{A^*\leftarrow}$$

Collecting terms and neglecting second-order effects then it can be shown that

$$\begin{aligned} J_{A^*,net} &= [d^2 m \nu X_v C_{A^*}/kT](\partial \mu_{A^*}/\partial C_{A^*})]\exp[-\Delta G_m/kT](-\partial C_{A^*}/\partial x) \\ &= D_{A^*}(-\partial C_{A^*}/\partial x) \end{aligned} \tag{8.17}$$

Thus, we have obtained an explicit relation for the intrinsic diffusivity D_{A^*}, namely

$$D_{A^*} = (C_{A^*}/kT)(\partial \mu_{A^*}/\partial C_{A^*})]d^2 m \nu X_v \exp[-\Delta G_m/kT]$$

For the case of metals we can make use of the result we have derived in Chapter VI for the temperature dependence of the equilibrium vacancy concentration, $X_v = \exp[-\Delta G_f/kT]$, where ΔG_f is the free energy of formation of a vacancy to obtain

$$D_{A^*} = (C_{A^*}/kT)(\partial \mu_{A^*}/\partial C_{A^*})]d^2 m\nu \exp[-\Delta G^*/kT] \tag{8.18}$$

where $\Delta G^* = \Delta G_f + \Delta G_m$. Also, since we are considering the diffusion of A* tracer atoms in dilute solution in a matrix of pure A, then $\partial \mu_A^*/\partial C_A^* = kT/C_A^*$. Further, by using the relation $\Delta G^* = \Delta H^* - T\Delta S^*$, we obtain

$$D_A^* = D_o \exp(-Q/RT) \tag{8.19}$$

where $D_o = m\, d^2 \nu \exp(S^*/k)$ and $Q = N\,\Delta H^* = N(\Delta H_m + \Delta H_f)$.

In the above derivation we have assumed that each atomic jump is independent of all jumps that preceed it. Usually successive jumps are not independent and our result must be corrected for this neglect. The correction factor is known as the correlation factor f, and has been extensively investigated.[5,6] The correlation factor multiplies the diffusivity to give a corrected diffusivity. Table 8.3 gives a collection of these factors for self-diffusion. Thus, to be strictly correct, D_0 should be multiplied by the correlation factor f. In the literature, the factor m is sometimes incorporated into the entropy factor S*. Substituting reasonable values yields that the frequency factor D_0 should be on the order of $10^{-5}\,m^2/s$ for the case of diffusion in metals. In fact, it is found to lie between 10^{-6} to $10^{-4}\,m^2/s$.

Some metals exhibit a curvature in the Arrhenius plots of log diffusivity versus reciprocal temperature. If the activation energy is independent of temperature, as expected, then this observation may imply the existence of a contribution to atom transport of other than single vacancies. Indeed, this is currently the favorite

Table 8.3. Correlation factors for self-diffusion.

Crystal structure	**Correlation factor**
Vacancy Mechanism, 2D lattices	
Honeycomb lattice	0.5
Square lattice	0.46694
Hexagonal lattice	0.56006
Vacancy Mechanism, 3D crystal structures	
Diamond	0.5
Simple cubic	0.65311
Body-centered cubic	0.72722
Face-centered cubic	0.78146
Hexagonal close-packed (with all jump frequencies equal)	0.78121 normal to c-axis 0.78146 parallel to c-axis
Interstitialcy Mechanism (θ = angle between the displacement vectors of the atoms participating in the jump)	
NaCl, collinear jumps ($\theta = 0$)	0.666
NaCl, non-collinear jumps with $\cos\theta = 1/3$	0.9697
NaCl, non-collinear jumps with $\cos\theta = -1/3$	0.9643
Ca in CaF_2, collinear jumps ($\theta = 0$)	0.80
Ca in CaF_2, non-collinear jumps with $\theta = 90°$	1.0

After J.R. Manning, **Diffusion Kinetics for Atoms in Crystals**, Chapter 3, D. Van Nostrand, 1968.

explanation and the additional contribution to atom transport is believed to be provided by divacancy clusters. However, other explanations for the curvature in the Arrhenius plot have been offered. For example, it has been suggested that a strong thermal expansion coefficient for the vacancy leads to a temperature dependence of both D_0 and Q. Also, it has been suggested that dynamical correlation between successive vacancy jumps can contribute to this curvature. Despite these modifications there is no doubt that in normal metals the primary mechanism of atom transport is by vacancy–atom exchange jumps.

The approach we have taken in this section can be generalized to provide a relation for the 1D diffusion of any species in terms of a jump frequency Γ, as follows

$$D^* = (1/6)d^2\Gamma \tag{8.20}$$

where Γ is the jump frequency (number of jumps in all directions/time). When correlation is taken into account then

$$D^* = (1/6)d^2\Gamma f \tag{8.21}$$

Another way of arranging these quantities is useful when application is made to diffusion in ionic crystals and semiconductors. The relation of interest is

$$D^* = C_d D_d f \tag{8.22}$$

where C_d is the concentration of the defect denoted by the letter d, D_d is the diffusivity of this defect, which for the above case of vacancy diffusion is given by

$$D_v = m\, d^2 \nu \exp(-\Delta G_m/kT) \tag{8.23}$$

2.1.2. Dilute alloys

In the case of dilute alloys attention must be paid to the diffusivity of the solute atom. When the vacancy mechanism operates the relation for the diffusivity of the solute atom is

$$D_2 = f_2 d^2 \omega_2 C_v \exp(-\Delta G_b/kT) \tag{8.24}$$

where f_2 is the correlation factor for solute atom diffusion, ω_2 is the vacancy–solute atom exchange frequency, C_v is the concentration of vacancies and ΔG_b is the Gibbs free energy of binding of the vacancy–solute atom complex. But, $C_v = \exp(-\Delta G_f/kT)$, where ΔG_f is the Gibbs free energy of formation of the vacancy and $\omega_2 = \nu \exp(-\Delta G_{2m}/kT)$, where ΔG_{2m} is the Gibbs free energy of activation for the vacancy–solute atom exchange. Hence, the activation energy of diffusion for the solute atom is of the form

$$Q_2 = \Delta H_f + \Delta H_{2m} + \Delta H_b - C \tag{8.25}$$

where C accounts for the temperature dependence of the correlation factor f_2.

The significant factor in the relation for the solute atom activation energy is the binding enthalpy of the vacancy–solute atom complex. If this term is negative, which in the sense used in this book implies an attraction between vacancy and solute atom, then generally Q_2 will be smaller than Q for self diffusion, and vice versa for the case of repulsion of the vacancy by the solute atom. Thus, it is useful to have an understanding of the vacancy–solute atom interaction. For simple metal solvents and solutes the electrostatic model[7] appears to provide an adequate explanation of this interaction. This model is based on the interaction between charged entities, i.e. the vacancy acts as an effective negatively charged entity in the free-electron sea, while the solute atoms are either positively charged or negatively charged entities depending upon whether they donate extra electrons or absorb electrons from the sea of electrons, respectively. Another approach to this interaction is via thermodynamics. We note that the motion of a vacancy adjacent to a solute atom to a position away from the latter introduces a solute–solvent atom bond at the expense of a solvent–solvent atom bond. Thus, if the former bond is stronger than the latter one then this bond strength difference contributes to a repulsion of the vacancy from the solute atom and to a decrease in the activation energy for diffusion of the solute atom relative to that for self diffusion. However, as shown by equation 8.25 other terms contribute to the activation energy of diffusion besides ΔH_b. In particular, ΔH_{2m} is also a function of the solute identity and if the solute–solvent bond is stronger than that between solvent atoms then the effect of ΔH_{2m} on Q is just opposite to that of ΔH_b due to this difference in binding energy and is the stronger effect. We must add, however, that the real situation is much more complicated than that deduced on the basis of presumed nearest-neighbor bond energies. It is just for this type of evaluation that molecular static calculations based on applicable potential functions are useful. Although molecular static calculations have been made for ionic materials their use for dilute alloys is sparse or absent. Part of the problem is that adequate potential functions for transition elements have only recently been developed and the success of the electrostatic model for simple metals provided a lack of urgency for the use of potentials, such as the embedded atom one, for the latter case.

2.1.3. Ordered alloys/intermetallic compounds

Inasmuch as the predominant defect involved in diffusion in metals and alloys is the vacancy it is reasonable to expect that the vacancy will also play a role in diffusion in ordered alloys and intermetallic compounds. However, the vacancy mechanism may no longer be the simple one applicable to metals and dilute alloys. We need to consider each crystal structure separately. In one class of structures a site on one sublattice has as nearest neighbors only sites on another or other sublattices. Let us consider the case where each sublattice has only one type of component at its sites, such as in the B2 (CsCl) structure. For this structure several mechanisms have been suggested, as follows: six jump cycle mechanism, antistructural bridge mechanism, triple defect mechanism, intrasublattice vacancy jump mechanism, and others. The six jump mechanism involves the coordinated jumps of vacancies from one sublattice to the other. Antisite defects are produced in the first three jumps and the lattice is progressively reordered in the next three jumps. In the antistructural bridge mechanism the assumption is made that antisite defects exist at equilibrium and that vacancy jumps can be made from one sublattice to the other without changing the degree of order. A minimum concentration of antisite defects is required to accomplish such vacancy jumps without affecting the order parameter. If an antistructural atom has two vacant sites on the neighboring sublattice then it is possible for the antisite atom to diffuse by next nearest neighbor jumps and the other component to diffuse by nearest neighbor jumps.

Experiments and first-principle calculations that aid in distinguishing between the competing mechanisms are not many. A few results will be cited. For the case of Fe–Al, quasielastic Mossbauer experiments are able to distinguish the jump vectors of Fe atoms or of a sequence of jump vectors with residence times in intermediate configurations that are much less than the lifetime of an excited Mossbauer level. These experiments reveal that the jump vector for the Fe atom is a[100] and to a lesser extent a[110]. Jumps along [111] are not observed negating the applicability of the antistructural bridge mechanism for Fe diffusion. Also, there are measured values of the migration energies, which for the Fe atom is 1.7 eV. In addition, first-principle calculations of migration and formation energies were made for various paths and defects. The vacancy formation energy on the Fe sublattice is much smaller than that on the Al sublattice by a factor of at least three. The only path and mechanism that gave agreement with the observed migration energy and jump vector involves a correlated two-atom jump as follows. An Al atom jumps to a neighboring iron vacancy and, simultaneously, a neighboring Fe atom moves into the germinating, but unborn, Al vacancy. In making this tentative conclusion it was assumed that a correction to the effective defect energies to account for spin polarization on the Fe atoms (evaluated on a supercell with 32 sites) of -0.35 eV would bring about quantitative agreement between the migration energy calculated for a 54 site supercell, 1.98 eV and the observed migration energy of 1.7 eV.[8] Thus, this recent study resulted in a mechanism not described in the above list for the B2 structure. In metallic B2 ordered phases it is known that both components have roughly similar values of diffusivity,[9] a fact with which this new mechanism of a correlated two-atom jump is consistent.

The experimental technique of perturbed angular correlation of gamma rays (PAC) measures the vacancy fraction of the sites next to the probe atom. Measurements made at high temperature reveal the equilibrium value of this vacancy fraction. For the case of the B2 phase of NiAl it was concluded that the predominant defect is the triple defect ($2V_{Ni} + Ni_{Al}$).[10] This result suggests that diffusion in a particular metallic ordered alloy crystal structure need not be characterized by the operation of only one mechanism.

Another structure in which diffusion mechanisms have been studied is the DO_3 structure. There are two types of A sublattices: the α and γ sublattices. The A atoms on the α sublattice have as nearest-neighbors 4 A atoms on the γ sublattice and 4 B atoms on the β sublattice. The A(B) atoms on the $\gamma(\beta)$ sublattice have only A atom nearest neighbors on the α sublattice. Now in DO_3 type Fe_3Al, Fe vacancies occur mainly on the α sublattice and not the γ sublattice, whereas the Al(Fe) anti-structure atoms appear readily on the $\gamma(\beta)$ sublattice. There are also Al vacancies on the β sublattice.[8] Al diffusivity is less than a factor 10 higher than that of the Fe atom diffusivity.[11] However, in DO_3 type Fe_3Si, Fe diffusion is much faster than Si diffusion and the Fe diffusivity is much higher than the Fe diffusivity in Fe_3Al.[11] First-principle calculations similar to those performed for Fe_3Al have not yet been carried out for Fe_3Si. However, it is likely that the fast diffusion of Fe in Fe_3Si indicates that the vacancy formation energy on the γ sublattice must not be much higher than that for formation on the α sublattice, contrary to the case for Fe_3Al, and allowing for nearest-neighbor Fe atom–vacancy exchanges on Fe sublattices. For DO_3 type Ni_3Sb the Ni atoms have the fastest self-diffusivity measured for any metallic system.[11] There are a very large concentration of thermal vacancies on the α sublattice and a smaller concentration on the γ sublattice with the vacancy formation energies being much smaller than found in other DO_3 or B2 metallic phases.[8] Hence, Ni can diffuse readily by vacancy–Ni exchange on nearest neighbor atom sites. The predominant defect identified in Ni_3Sb consists of 4 Ni vacancies on α sublattice sites adjacent to 1 Ni antistructure site atom on the β sublattice.[8]

In the B20 FeSi ordered phase the diffusivity of Fe is much smaller than that for Si and both are much smaller than that for Si in DO_3 Fe_3Si.[11] The explanation suggested for this result is that in the B20 FeSi the bonds are predominantly covalent type and, consequently, the formation energies of intrinsic defects are higher than in comparable metallic phases. Indeed, the diffusivities at the melting point are on the order of $10^{-14}\,m^2/s$, compared to $10^{-15}\,m^2/s$ for covalent Si and Ge and to 10^{-13} to $10^{-12}\,m^2/s$ for metallic alloys.

Thus, ordered alloys can exhibit diffusivities that are both larger and smaller than those in comparable metals depending upon the relative ease of forming intrinsic defects in these materials.

2.1.4. Amorphous metals

The mechanism of diffusion in amorphous metals appears to involve other than a vacancy mechanism. In particular, molecular dynamics simulations suggest that groups of atoms (containing on the order of 10 atoms) associated with low-frequency localized vibrational modes execute a cooperative jump which can be either reversible or not.[12] In the latter case a structural change takes place with a relative displacement of the atoms involved and a diffusivity can be evaluated from the continued displacement of specified atoms. After a structural change new regions appear with low-frequency vibrational modes to repeat the cycle.

2.2. Ionic crystals

2.2.1. Stoichiometric-cation vacancy predominant defect

For the case of a pure stoichiometric ionic solid (a rare species), if the major mechanism involved diffusion of vacancies on one type of site and the Schottky defect predominated, then the vacancy concentration corresponding to equation 6.13 and the vacancy diffusivity from equation 8.23 would be substituted into equation 8.22 to obtain the diffusion coefficient. The subscript m corresponds to the vacancy that has the least migration free energy, in the present case it is assumed to be the cation vacancy.

If soluble impurities were present in such a stoichiometric ionic solid to a larger concentration than the intrinsic concentration of the controlling vacancies, then it would be likely that the impurity concentration would determine the vacancy concentration. The resulting expression for the vacancy concentration would depend upon the valence of the impurity and its compensating defect. In this case, the vacancy concentration need not be temperature dependent and, in this event, the activation energy for diffusion will be that for migration of the vacancy.

2.2.2. General case

Given a knowledge of the predominant defect and the identity of the most mobile defect, then equation 8.22 can be applied along with the relations developed in Chapter VI, for the case of a simple MO type oxide, to evaluate the appropriate diffusivity. Normally, it is desired to obtain the dependence of the diffusivity on the partial pressure of O_2 for an oxide. In this case, the data in Table 6.6 can be used to obtain the partial pressure dependence of the defect concentration, which is the same as for the diffusivity. For example, for an hypothetical pure oxide of formula MO, in which the predominant mobile defect is the singly charged cation vacancy, the diffusivity will vary as the 1/4th power of the partial pressure of O_2. The order of the latter is usually reversed, i.e. the experimental partial pressure dependence of the diffusivity is used to obtain one limiting condition on the identity of the predominant defect. A real example illustrating this behavior is CoO, in which at 1200°C diffusion is by singly charged cation vacancies at high oxygen activity and doubly charged cation vacancies at low oxygen activity. At high oxygen pressure the cobalt tracer diffusivity is proportional to the 1/4th power of the oxygen activity.[13]

To obtain the relation for the partial pressure dependence of the predominant defect for the general case of an M_xO_y type oxide recourse should be made to the literature cited in the bibliography. It should come as no surprise, considering the number of alternative possibilities that exist in ionic crystals (e.g. defect type, charged impurities and their charge compensating defects, non-stoichiometry, contribution to diffusion from both cation and anion species, association of vacancies and solute, precipitation of solute, etc.), that the mechanism of diffusion in ionic crystals is much more complicated than in metals. Indeed, one cannot speak of a mechanism for diffusion in ionic crystals. There are many possible mechanisms and some have been reliably identified only for a few ionic systems. The latter statement is especially applicable to oxides.

Another instructive example is that of diffusion in magnetite Fe_3O_4, where more than one diffusion mechanism is operative. Magnetite has the inverse spinel structure, with Fe^{2+} on the octahedral sites and half of the Fe^{3+} on the octahedral and the other half on the tetrahedral sites at room temperature. At higher temperatures the distribution of these two cations is randomized over the two types of sites. There are four octahedral and eight tetrahedral sites/three cations. At low oxygen activities there is a cation excess with iron ions on interstitial sites as the dominant defect. At high oxygen activities the oxide is metal deficient with cation vacancies as the dominant defect. The relations governing the species in magnetite are

$$3\mathrm{Fe}_{\mathrm{Fe}} + 4\mathrm{O}_{\mathrm{O}} = 3\mathrm{Fe}_{\mathrm{i}} + 2\mathrm{O}_2 \quad \text{(low oxygen activity)}$$

(electroneutrality is maintained by change in the valence of the appropriate cation)

$$8\mathrm{Fe}^{2+}_{\mathrm{Fe}} + 2\mathrm{O}_2 = 8\mathrm{Fe}^{3+}_{\mathrm{Fe}} + 3\mathrm{V}_{\mathrm{Fe}} + 4\mathrm{O}_{\mathrm{O}} \quad \text{(high oxygen activity)}$$

(In both reactions the stoichiometric ratio of lattice sites of the two species is conserved.) Hence, the diffusivity should vary as the minus 2/3rd power of the oxygen partial pressure at low oxygen activity and as the 2/3rd power of the partial pressure at high oxygen activity. The results shown in Figure 8.6 are consistent with these relations.

The methods of determining ionic diffusion mechanisms have been reviewed recently.[14] Atomistic simulation procedures are extremely useful in this regard and also these have been reviewed recently.[15]

Diffusion in ionic crystals is particularly technologically significant in the process of oxidation and in superionic materials, as used in fuel cells and the

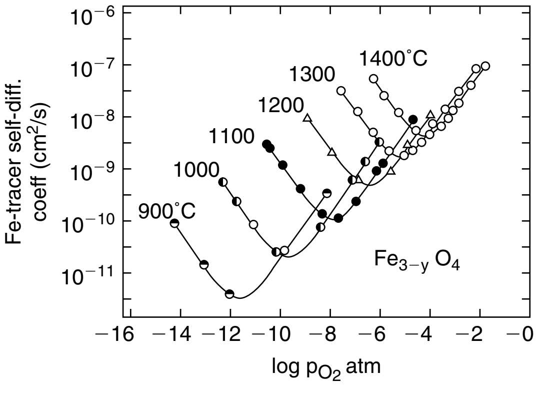


Figure 8.6. The iron tracer self-diffusion coefficient in magnetite as a function of oxygen activity. From Dieckmann and Schmalzried, Ber. Bunsenges. Phys. Chem. 81, 344(1977) with permission.

like. Of course, it is also important in the control of the processing of ceramics and, hence, of their properties.

2.3. Semiconductors

The mechanisms of diffusion in semiconductors are more varied than found in metals. In particular, self-diffusion via interstitial motion occurs in silicon at high temperatures above about 1270°C. Also, the ability to vary vacancy and self-interstitial concentrations by the addition of donors and acceptors can materially affect self-diffusivity. The open lattice of covalently bonded materials makes interstitial diffusion of solute much more common.

As for ionic crystals, the applicable diffusivity can be expressed by equation 8.22 and the defect concentration by the appropriate mass action relations. It has been found that the effect of solute on the tracer diffusivity of the host elemental semiconductor atoms can be expressed as the following linear relation

$$D(C_B)/C(0) = 1 + bC_B$$

where b is an enhancement factor that lies between -20 and $+100$.

Figure 8.7 shows solute diffusivities in silicon as a function of temperature. The more rapid diffusion of solutes compared to the self-diffusion of silicon is obvious. It appears that the Group III and V solutes are incorporated solely on regular lattice sites and diffuse via the same mechanisms as self atoms. The much higher diffusivities of the metal solutes is likely to be due to the existence of these solute as interstitial atoms. A detailed explanation of these relative diffusivities is given in the article by Frank et al. referenced in Figure 8.7.

3. Nernst–Einstein relation

To derive the Nernst–Einstein relation we note that the equations for the flux of atoms given equations 8.16 and 8.17 can be written as

$$J_{i\rightarrow} = \Gamma_o \exp(F_i d/2kT) \quad \text{and} \quad J_{i\leftarrow} = \Gamma_o \exp(-F_i d/2kT)$$

yielding for the net flux

$$J_{net} = \Gamma_o F_i d/kT = \langle v_i \rangle C_i$$

where $\Gamma_o = C_i d m \nu \exp[-\Delta G^*/kT]$. But, the intrinsic diffusion coefficient is

$$D_i = \varphi d^2 m \nu \exp[-\Delta G^*/kT]$$

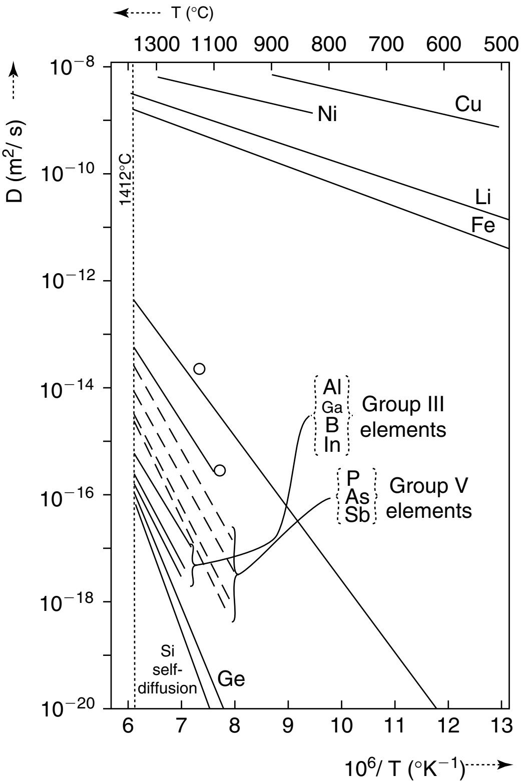


Figure 8.7. Diffusivities of foreign atoms in silicon. From Frank et al., in Diffusion in Crystalline Solids, eds. Murch and Nowick, Academic Press, New York, 1984 with permission.

Hence, $\Gamma_o = D_iC_i/\varphi d$ and, by substitution, $<v_i> = D_iF_i/\varphi kT = D_{i*}F_i/kT$. This equation is the Nernst–Einstein relation. Note that the diffusivity in this relation is the tracer diffusivity. The alert student will be aware that this relation was used in deriving the Darken relations.

It is convenient to consider the various driving forces for diffusion which are

$F_i = d\mu_i/dx$ for a concentration gradient

$= d\Phi/dx = Z_i^*eE$ where E is the electric field for the latter as a driving force and Z_i^* is the effective valence

$= <Q_{ia}> d\ln T/dx$ for a temperature gradient, where $<Q_{ia}>$ is the effective heat of transport.[3]

4. Empirical rules

Certain generalizations regarding diffusion have been discovered which are useful as follows:

- *Metals*

1. The activation energy for self-diffusion obeys the relation $Q = 34T_m$(cal/g atom/°C), where T_m is the melting temperature.

2. The activation entropy for self diffusion obeys the relation $\Delta S = Z\beta Q/T_m$, where Z is a constant that depends on the lattice and equals 0.55 for fcc and 1 for bcc, $\beta = -d(\mu/\mu_o)/d(T/T_m)$, μ is the shear modulus.

3. The activation volume for self diffusion obeys the relation $\Delta V = 4\chi Q$, where χ is the compressibility.

- *Alloys*

1. The diffusion coefficient D, at the solidus temperature is roughly a constant for a given structure and bond type, i.e. $D_{fcc}(T_{solidus}) = 5 \times 10^{-9}$ cm²/second; $D_{bcc}(T_{solidus}) = 3 \times 10^{-8}$ cm²/second, except for the bcc alkali and actinide metals.

2. $Q/RT_{solidus} = 18(\pm 2)$.

3. Pre-exponential is roughly constant for a given structure and bond type. $D_o(fcc) = 0.3$ cm²/second; $D_o(bcc) = 1.6$ cm²/second with exceptions as noted in 1.

4. In dilute alloys, addition of a faster(slower) diffusing species as an impurity causes an increase (decrease) in the rate of diffusion of both solute and solvent atoms; an increase in concentration of a solute that lowers (raises) the melting point increases (decreases) the diffusion rate of this solute at temperature T.

5. In concentrated alloys, the tracer diffusion coefficients of the different components do not differ by more than an order of magnitude if the diffusion mechanisms of the species is the same.

6. No generalizations exist concerning the diffusion of interstitial solute. Hence, some data are collected in Appendix 1 of this chapter.

- *Ionic solids*

The possibilities for not achieving reproducible results and the multitudes of possible mechanisms have worked against the discovery of similar reliable empirical rules in these materials.

5. Solutions to diffusion equations

Interdiffusion of component species is often involved in the processing of materials. For control of the processing, and consequently the properties of a material, it is desired to know the concentration of a species C, as a function of position, time and temperature, i.e. know C(x, t, T). For diffusion in 3D the applicable relation is Fick's second law in the form

$$\partial C/\partial t = D[\partial^2 C/\partial x^2 + \partial^2 C/\partial y^2 + \partial^2 C/\partial z^2]$$

for D independent of composition. Solutions to this equation for a variety of boundary conditions are given in J. Crank, The Mathematics of Diffusion, Oxford, 1956 and Ghez's book (see Bibliography).

For example, the solution for the 1D diffusion into a plate from a source maintaining a surface concentration C_s into a solid in which the concentration of the diffusing species is initially at the concentration C_o is

$$C = C_o + [(C_o - C_s)/2]\mathrm{erfc}(x/2[Dt]^{1/2})$$

where $\mathrm{erfc}(u) = 2/\pi^{1/2}\int_0^u \exp(-w^2)dw$

A very useful approximation stems from the integrating factor of Fick's second law given by $\lambda = x/\sqrt{t}$. When in the form $\lambda' = \langle x\rangle/\sqrt{Dt}$, then $\lambda'$is a dimensionless function of the concentration having a value on the order of unity for various boundary conditions in 1D diffusion. Thus, $\langle x\rangle$can be regarded as the penetration distance of some average concentration between that of the source and that of the host into which the species diffuses. The relation between the penetration distance, time and temperature has many applications in materials science and is probably the most widely used relation in the practical application of diffusion concepts, which may account for it being on the cover of Borg and Dienes's book.

6. High diffusivity regions in solids

It is a fact that polycrystals and even single crystals contain regions (dislocations and grain boundaries), where the diffusivity is higher than in the perfect crystal. We should like to answer two questions: (1) Can the intrinsic diffusivity of these regions be measured? (2) Can such measurements help define the nature of these imperfect regions?

The answer to the first question is that if the diffusivity of the imperfect region, at a constant temperature and pressure, is assumed to be constant over some volume associated with the imperfection, then it is a relatively simple matter to measure this "defined intrinsic diffusivity." The answer to the second question is still unknown, although it appears more likely that the reverse – that a knowledge of the imperfect regions will help define the mechanism of diffusion in them-will occur sooner.

At low temperatures the diffusivity in these imperfections is generally many orders of magnitude greater than that in the perfect lattice. When the product of the volume fraction by the diffusivity for each type region is of the same order of magnitude then it is important to consider the contribution of each to the measured transport of matter.

There are mainly two techniques for measuring the intrinsic self-diffusivities of defect regions. One involves plating a radioactive layer on the surface and then measuring the concentration of the radioactive species of sections parallel to the surface after a diffusion anneal. The radioactive material transported to each section by diffusion through perfect and imperfect material is then calculated on the basis of a model and measured values of the volume fraction of each region. The other technique requires the specimen to be a thin sheet in which the imperfect regions are mainly normal to the sheet surface so as to provide "short-circuit" paths from one surface of the sheet to the other. Then a radioactive layer is plated on a portion of one side of the sheet and the amount reaching the other side by "short-circuit" diffusion is measured after a diffusion anneal at a sufficiently low temperature. The former is called the sectioning technique and the latter the surface accumulation technique. In both of these cases, other sensitive techniques of measuring the concentrations of the diffusing species can be substituted for the techniques of using a radioactive species, i.e. in the surface accumulation technique AES may be used.

6.1. Models for evaluating the grain boundary diffusivity

Fisher's approximate method[16]

According to the model of Fisher for the sectioning technique, consider a grain boundary normal to the surface, of thickness d, of a semi-infinite crystal. The boundary conditions for the diffusion are

$$C = C_o \quad \text{for } y = 0 \text{ and } t \geqslant 0$$

$$C = 0 \quad \text{for } y > 0 \text{ and } t = 0$$

Fisher assumes the concentration in the grain boundary to be independent of x, the distance in the direction normal to the grain boundary plane. Further, in going out of the grain boundary into the bulk in the x direction, in the case of self-diffusion in a pure element, there is no discontinuity in concentration. Conservation of matter in the grain boundary region of thickness δ and length dy then yields

$$\begin{aligned}\partial C/\partial t &= [1 \cdot dy \cdot \delta]^{-1}\{\delta(J_y - [J_y + (\partial J_y/\partial y)dy]) - 2\,dyJ_X\}\\ &= -(\partial J_y/\partial y) - (2/\delta)J_X\end{aligned}$$

But $J_y = -D_b\,(\partial C/\partial y)$; $J_x = -D_l[\partial C/\partial x]_{\delta/2}$. Therefore

$$\partial C/\partial t = D_b(\partial^2 C/\partial y^2) + (2/\delta)D_l[\partial C/\partial x]_{\delta/2}$$

Outside the grain boundary $(\partial C/\partial t) = D_l \nabla^2 C$.
Fisher assumes:

1. C_b is near to its final value during all of the time t of diffusion;

2. J_x is normal to the grain boundary, there is no J_y in the bulk lattice except that which initially diffused from the surface through the bulk.

These assumptions allow the problem to be divided into two parts: matter transported to a section at y either through the lattice or through the grain boundary and then along the x direction. For the latter we have the following solution

$$C(x, y, t) = C_b(y)[1 - \mathrm{erfc}\{x/2(D_l t)^{1/2}\}]$$

where

$$C_b(y) = C_o \exp[-(2y)^{1/2}/((\pi D_l t)^{1/4}(\delta D_b/D_l)^{1/2})]$$

Experimentally, the total amount of solute in a section dy thick is measured, i.e. $2\int_0 C(x, y, t)\, dx$ plus that diffused in from the surface through the bulk. The latter contribution overwhelms the former until $D_b/D_l \gg 5 \times 10^4$ when the reverse occurs.

The net effect of enhanced diffusion along grain boundaries in polycrystalline samples is to produce a deviation in the plot of $\log D_{\mathrm{apparent}}$ versus 1/T at a temperature below which $D_b/D_l \gg 5 \times 10^4$. The actual ratio at this point depends upon grain size.

The exact solution for the semi-infinite bicrystal and for the same boundary conditions has been given by Whipple[17] and by Suzuoka[18] for the case of an instantaneous source instead of a constant source.

The above analysis holds for the case of tracer diffusion where there is no bias for segregation of the tracer atoms to the grain boundary. In the case of solute diffusion the probability of solute segregation at the grain boundary must be taken into account. Harrison[19] has pointed out that there are several regions of kinetic behavior for solute diffusion along grain boundaries. In Harrison's type B regime the product δD_b in the equation for the grain boundary concentration is replaced by the product $s\delta D_b$, where s is the segregation coefficient and $s = s_o \exp[-\Delta H_s/kT]$, where ΔH_s is the binding enthalpy* of the solute to the grain boundary. Thus, in this type B regime additional data concerning s and its temperature dependence is necessary to evaluate the boundary diffusivity. At much lower temperatures, in Harrison's type C regime, the leakage of solute to the lattice from the grain boundaries is negligible and D_b can be evaluated directly from the penetration profile of the radioactive solute. It is only in this low temperature type C regime that the surface accumulation technique can yield interpretable results in that loss of solute to the lattice is negligible once steady state concentration profiles are established along the grain boundaries.

* The binding enthaply is negative when solute tends to concentrate at the grain boundary.

6.2. Mechanism of diffusion along grain boundaries

Balluffi[20] has considered the possible atomic mechanisms for diffusion along grain boundaries and has come to the conclusion that vacancies in metals and alloys are most likely to be the defect responsible for the fast diffusion. Grain boundary dislocations are likely to be the sources and sinks for vacancies with the energy of formation of vacancies at grain boundary sites being somewhat larger than the migration energy of such vacancies along the boundaries, but with both being smaller than the corresponding values in the bulk lattice. These results are based on molecular dynamic simulation of grain boundary diffusion, which may or may not be sensitive to assumptions concerning the interatomic potential used in the calculations. Hence, it is desirable to have more evidence concerning the mechanism of grain boundary diffusion. Among such evidence that support the vacancy mechanism are measured values of activation volume that equal the atomic volume.

6.3. Empirical results

6.3.1. Metals

1. For self-diffusion the apparent activation energy for diffusion along grain boundaries is from 0.4 to 0.6 that for bulk diffusion.

2. For solute diffusion the apparent activation energy includes the energy of segregation of the solute to the boundary.

3. In low angle boundaries the diffusion is anisotropic, being much faster along the direction of the tilt axis for tilt boundaries than perpendicular to it. For example, the ratio of activation energy for grain boundary diffusion to that for bulk lattice diffusion is 0.4 when the grain boundary diffusion is parallel to the <100>tilt axes, 0.6 when the grain boundary diffusion is parallel to the <211>tilt axis, and 0.7 when the grain boundary diffusion is parallel to the <111>twist axis, for both nickel and silver.

4. Diffusion is more rapid parallel to undissociated dislocation cores than to dissociated dislocation cores.

5. The activation energies in fcc metals for surface, grain boundary and dissociated dislocations are $13T_m$, $17T_m$ and $25T_m$, respectively, which are to be compared to that for lattice diffusion of $34T_m$.*

6.3.2. Ionic crystals

In principle, grain boundaries are also high-diffusivity regions in ionic solids. However, because of the excess charge associated with grain boundaries there appear to be limitations on the species that can diffuse rapidly along them. This aspect of the high-diffusivity problem has not yet been resolved. Nevertheless, there are many reported cases involving high-diffusivity paths along grain boundaries in ionic crystals.

---

*Measurements of activation energies for grain boundary diffusion in silicides yield values from $25T_m$ to $39T_m$, which although higher than the value listed for grain boundary diffusion in fcc metals in still smaller than that for volume diffusion in the same silicides of $36T_m$ to $46T_m$, respectively.[33]

Grain boundary diffusion occurs in oxides and is believed to be responsible for the observed oxidation rates of many metals at temperatures less than about $0.5T_m$. This is an important result because many oxidation resistant metals are used in this temperature range. Among the oxides that reveal grain boundary diffusion at temperatures less than $0.5T_m$ are NiO, Fe_3O_4 and ZrO_2.

6.4. Electromigration along grain boundaries

An important and sometimes limiting phenomenon found to occur in thin films used in integrated circuits is that of failure induced by electromigration. The failure is a consequence of the existence of divergences in the transport of matter along grain boundaries. For example, such divergences are present whenever three grain boundaries between differently oriented grains meet in a line because the grain boundary diffusivities of such boundaries are not likely to be equal! Further, neither will the forces be exactly equal because of the different angles made by the normals to such boundaries and the applied potential gradient. The latter can lead to very high current densities due to the small dimensions of the conducting lines in the integrated circuits, which can then lead to significant values of atom transport via the coupling between the two, i.e. the electron drag. The attempts to ameliorate this problem have focussed on means of decreasing the grain boundary diffusivities of the diffusing species and this was the driving force of the initial research carried out to solve the problem. Recently, improvements in the electromigration time to failure have been achieved empirically, but not by affecting the grain boundary diffusivity of the host species. With the continued drive for miniaturization, electromigration failure remains a concern to integrated circuit manufacturers.

7. Computer assistance in solving diffusion problems

As the number of components increases so does the difficulty in obtaining mobility and activation energy parameters from diffusion data. One computer program that is helpful in this regard is DICTRA.[21] This computer code has several applications. One of them is to provide numerical solutions to the continuity equation

$$\frac{\partial C_k}{\partial f} = -\mathrm{div}(J_k)$$

that yield concentration profiles in multicomponent diffusion problems while allowing for the possibility that the cross-terms in the matrix of conductance coefficients are significant, i.e. yield composition dependent diffusivities. One example, which illustrates the importance of having sufficient thermodynamic and kinetic data from binary systems in order to take advantage of DICTRA to predict values for multicomponent systems, is described in Reference 22. Other methods exist that are described in a recent text,[23] but which assume concentration independent diffusivities, an assumption that is rarely followed by Nature. There is still room for improvement in the computer code to enable the simulations to more nearly simulate reality without exacting an unacceptable computer burden. Also, mathematical tools for analytically solving the non-linear relations involved in diffusion problems continue to be discovered.

A database enabling the calculation of solute diffusion in multicomponent Ni base alloys is provided in Reference 24. Application of this database to the calculation of the solute profiles of a join in the γ phase of Ni based alloys of the Rene type using the DICTRA program[25] yielded the results shown in Figure 8.8 which are compared there to the experimental profiles. The largest error evaluated for these data is for the case of Re and equals 5.8% for the average deviation of the experimental Re concentration from its corresponding calculated value. The state of the art has reached the point where useful results may be obtained by computer simulation of the diffusion process in multicomponent alloys.

The above results were carried out for a multicomponent alloy involving a diffusion couple between alloys having the same matrix phase. In the following the couple is between alloys having the same two phases but different compositions within the two phase region. This system was treated using a phase-field approach.[26] It was a generic study to determine the change in the relative volume of the two phases induced by the diffusion for various values of the diffusivities. The initial compositions were 19 at.%B, 23 at.%C and 52 at.%B, 27 at.%C in the ternary system having the phase diagram shown in Figure 8.9. After a diffusion anneal based on an assumed set of parameters the results shown in Figure 8.10 were provided by the phase-field simulation. These results differ from those provided by an error function solution.[27] In particular, the latter provides a diffusion path given by the zig-zag line in Figure 8.9 while the phase-field solution reveals a single phase region at the join after some diffusion period as shown in Figure 8.10. This result is brought about by a difference in the mobilities of two diffusing components.

The point that computer simulation provides a means of obtaining information about problems in kinetics that are too complex to yield to analytical solutions is well illustrated by the above examples.

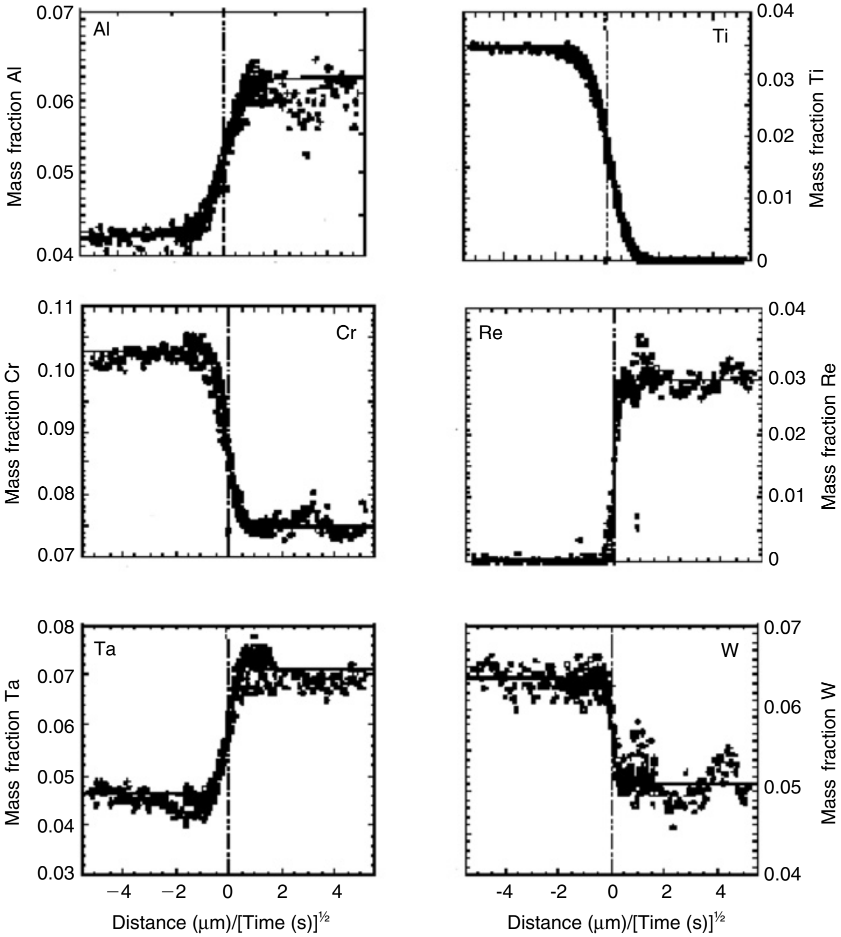


Figure 8.8. Composition profiles after a diffusion anneal of a join between Rene N4 and Rene N5 alloys heat treated to yield γ phases in both alloys at the join. Abscissa is diffusion distance divided by the square root of time. From C.E. Campbell et al., IAC-3 Conference Proceedings 2003 with permission. © 2003

8. Polymer diffusion

Although the phenomenological relations (equation 8.1) apply to polymer diffusion the mechanisms of diffusion in polymers differ from those in alloys and other crystalline materials. The subject of diffusion in polymers is large enough to

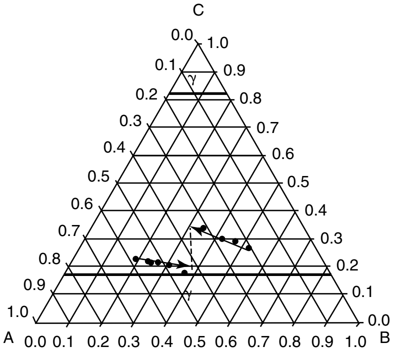


Figure 8.9. Ternary diagram indicating the initial compositions of a diffusion join between alloys within the two-phase field. The arrows indicate the diffusion path resulting from an error function prediction to the diffusion problem. From Acta Mater. 49, 3401(2001) with permission. © 2001 Elsevier.

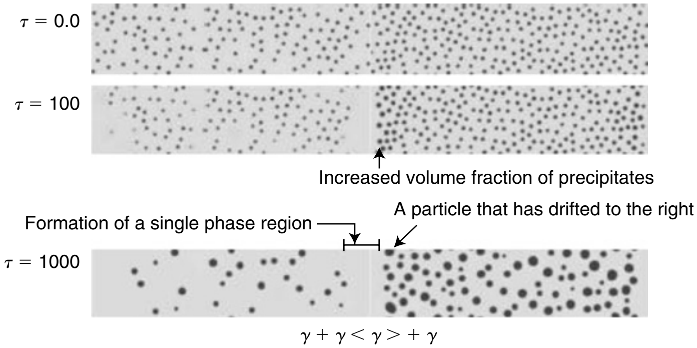


Figure 8.10. Phase field predictions for the course of diffusion between alloys within a two-phase field showing that a single phase region is produced on one side of the original join. From Acta Mater. 49, 3401(2001) with permission. © 2001 Elsevier.

occupy a book's length description. However, a brief summary here will attempt to provide the basis for an informed search for more information. Rouse in 1953[28] suggested a model for the diffusion of polymer strands which are not entangled. deGennes in 1971[29] suggested the reptation mechanism for polymer strands in an entangled environment. By entanglement is meant the effect of molecules in the environment which affect (limit) the random walk of mers along a polymer strand. In the Rouse model the polymer chains are considered to be random coils of monomers connected by entropic springs following a Brownian dynamics under the influence of these springs and an external Gaussian white noise. Hence, this model is applicable to polymers in dilute solution in some solvent. Polymer strands in their own melt are likely to be entangled even for short chain lengths.

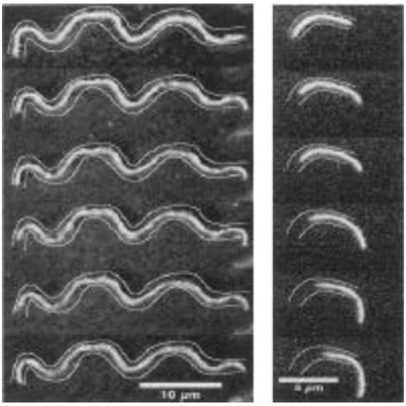


Figure 8.11. Snapshots revealing (left) back and forth motion of polymer chain within tube of entanglement; (right) effect of length on reptation. From Biophys. J. 70, 615(1996) with permission.

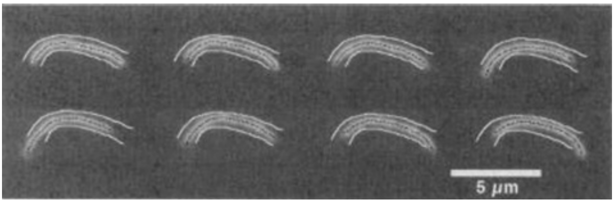


Figure 8.12. Showing back and forth motion of oligomer. From Biophys. J. 70, 615(1996) with permission.

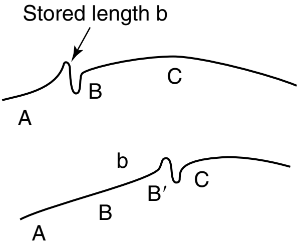


Figure 8.13. Schematic illustration of translation of point B by b due to reptation of loop having stored length b.

Reptation has been directly observed.[30] Figures 8.11 and 8.12 reveal characteristics of such reptation. One such characteristic is the sidewise motion within the limits defined by the entanglement with the environment (Figures 8.11 and 8.12). Another is the back and forth motion along the tube defined by the environment (Figure 8.12). As can be imagined the analytic description of diffusion of polymer chains is very complex. Numerous molecular dynamics and Monte Carlo studies have been conducted in attempts to clarify the mechanisms of polymer chain diffusion as well as to evaluate the scales applicable to each mechanism or combination of mechanisms. The Rouse Gaussian chain and deGennes reptation mechanisms do not exhaust the list of suggested mechanisms of polymer diffusion. The mechanism of the back and forth motion is not obvious from the above or from Figure 8.12. However, it is made clear in Figure 8.13. It is the diffusion of the loop that results in a motion of the chain itself. It is believed that this type of motion may be involved in the motion of nucleosomes bound by a DNA chain.[31]

For much more detail regarding the dynamics of polymer solutions see the book by Teraoka referenced in the bibliography. An instructive Monte Carlo study of the dynamics of polymer solutions and melts is given in Reference 32 which also contains references to older books and key references on the subject.

Appendix 1

Tables of data giving diffusivities and activation energies for interstitial diffusion of interstitially sited solute atoms in metals

Bcc metals

Metal	**Diffusing element**	$\frac{d}{a}$	D_{oi} (10^{-6} m^2/s)	Q_i (kJ/mol)	$\frac{Q_i}{Q_{SD}}$	$\frac{S_M}{R}$	$\frac{Q_i}{RT_m}$
Ca	C	0.35	0.27	97.6		0	10
β-Th	C	0.37	2.2	113	0.38	2.0	6
	N	0.34	0.23	71	0.24	0.05	4
	O	0.29	0.13	46	0.15	−0.2	2
γ-U	C	0.44	21.8	123	1.07	4.4	10
β-Ti	C	0.47	0.6	94.6	0.38	1.0	5
	N	0.43	3.5	141.5	0.56	2.6	8
	O	0.36	8.3	131	0.52	3.6	8
β-Zr	C	0.43	24.5	150.3	0.55	4.4	8
	N	0.39	1.5	128.5	0.47	0.5	7
	O	0.33	98	172	0.63	5.9	9
β-Hf	C	0.44	80	211		5.4	10
	N	0.40	0.8	124		1.2	6
	O	0.34	32	171		4.8	8
V	C	0.51	0.88	97.6	0.32	1.5	5
	N	0.47	5.02	151	0.49	3.1	8
	O	0.40	2.66	125	0.41	2.6	6
Nb	C	0.47	1.0	142	0.35	1.3	6
	N	0.43	2.6	152.3	0.38	2.3	6
	O	0.36	0.53	109.5	0.27	1.0	4
Ta	C	0.47	0.67	162	0.39	0.9	5
	N	0.43	0.52	158.5	0.38	0.7	5
	O	0.40	1.05	110.4	0.27	1.6	4
Cr	C	0.53	0.87	111	0.36	1.4	6
	N	0.49	1.6	115	0.37	2.1	6
	O	0.42	0.01	155	0.50	−3.0	8

(*Continued*)

Metal	Diffusing element	$\frac{d}{a}$	D_{oi} $(10^{-6}\,m^2/s)$	Q_i (kJ/mol)	$\frac{Q_i}{Q_{SD}}$	$\frac{S_M}{R}$	$\frac{Q_i}{RT_m}$
Mo	C	0.49	3.4	172	0.37	2.5	7
	N	0.45	0.43	109	0.24	0.73	5
	O	0.38	3.0	130	0.28	2.7	6
W	C	0.51	1.2	188	0.32	1.4	6
	N	0.47	2.4	119	0.20	2.4	4
	O	0.40	0.013	100	0.17	−2.6	3

Fcc metals

Metal	Diffusing element	$\frac{d}{a}$	D_{oi} $(10^{-6}\,m^2/s)$	Q_i (kJ/mol)	$\frac{Q_i}{Q_{SD}}$	$\frac{S_M}{R}$	$\frac{Q_i}{RT_m}$
Ag	O	0.294	2.72	46	0.25	1.4	4.8
Co	C	0.434	50	159	0.56	3.8	10.8
Cu	O	0.332	1.7	67	0.33	0.8	5.9
γ-Fe	C	0.42	74	159	0.59	4.0	10.5
	N	0.39	91	169	0.62	4.3	11.2
	O	0.33	575	169	0.62	6.2	11.3
Ni	C	0.44	5	146	0.51	1.4	10.2
	O	0.34	1.82×10^6	301	1.06	14.0	21
Pt	O	0.31	930	326	1.14	10.9	19.2
α-Th	C	0.3	2.7	159	0.53	0.4	9.5
	N	0.28	2.1	94	0.31	−1.8	5.6
	O	0.24	1.3	205	0.68	−0.3	12.2

Hcp metals

Metal	Diffusing element	$\frac{d}{a}$	Doi $(10^{-6}\,m^2/s)$	Qi (kJ/mol)	$\frac{S_M}{R}$	$\frac{Q_i}{RT_m}$
Y	C	0.42	20	123	3.2	2.3
	N	0.39	0.1	251	−2.4	4.6
Re	C	0.56	10	222	2.5	7.7
	N	0.51	14	154	5.2	5.4
α-Ti	C	0.52	506	182	6.4	11.3
	N	0.48	1.2	189	0.5	11.7
	O	0.41	80	201	4.6	12.4

(*Continued*)

Metal	**Diffusing element**	$\frac{\mathbf{d}}{\mathbf{a}}$	**Doi (10^{-6} m²/s)**	**Qi (kJ/mol)**	$\frac{S_M}{R}$	$\frac{Q_i}{RT_m}$
α-Zr	C	0.48	0.2	151.6	−1.4	8.6
	N	0.44	30	238.6	3.5	13.5
	O	0.37	235	213	5.8	12.0
α-Hf	C	0.48	7.4×10^3	310	8.7	14.9
	O	0.38	66	213	4.4	10.3

References

1. J. Volkl and G. Alefeld, in **Hydrogen in Metals**, eds. G. Alefeld and J. Volkl, Springer, Berlin, 1978.
2. J.R. Manning, **Diffusion Kinetics for Atoms in Crystals**, Chapter 1, Van Nostrand, Princeton, 1968.
3. G. Bre'bec, in **Defauts Ponctuels Dans Les Solides**, Les Editions de Physique, Orsay, 1978, p. 181.
4. R.E. Howard and A.B. Lidiard, Rep. Prog. Phys. 27, 161(1964).
5. Reference 2, Chapter 3.
6. A.D. LeClaire, in **Physical Chemistry, An Advanced Treatise**, Vol. 10, Chapter 5, eds. H. Eyring, D. Henderson and W. Jost, Academic, NY, 1970.
7. A.D. Le Claire, Phil. Mag. 7, 141(1962); J. Nucl. Metal. 69, 70, 70(1978); G. Neumann and W. Hirschwald, Phys. Stat. Sol. B 55, 99(1973).
8. M. Fahnle, B. Meyer, J. Mayer, J.S. Oehrens and G. Bester, MRS Symp. Proc. 527, 23(1998).
9. H. Mehrer, Mat. Trans. JIM 37, 1259(1996).
10. B.B. Jiawen Fan and G.S. Collins, MRS Symp. Proc. 527, 203(1998).
11. H. Mehrer, T. Zumkley, M. Eggersmann, R. Galler and M. Salamon, MRS Symp. Proc. 527, 3(1998).
12. B.D. Wirth and G.R. Odette, MRS Symp. Proc. 481, 151(1998).
13. N.L Peterson and W.K. Chen, J. Phys, Colloq. C-6, 319(1980); R. Dieckmann, Z. Phys. Chem. Neue Folge 107, 189(1977).
14. J. Corish, MRS Symp. Proc. 527, 443(1998).
15. C.R.A. Catlow in **New Trends in Materials Chemistry**, eds. C.R.A. Catlow and A. Cheetham, Kluwer Academic Publishers, Amsterdam, 1997, p. 141.
16. J.C. Fisher, J. Appl. Phys. 22, 74(1951).
17. R.T.P. Whipple, Phil. Mag. 45, 1225(1954).
18. T. Suzuoka, Trans. Jap. Inst. Metal. 2, 25(1961); J. Phys. Soc. Japan 19, 839(1964).
19. L.G. Harrison, Trans. Farad. Soc. 57, 1191(1961).
20. R.W. Balluffi, in **Diffusion in Crystalline Solids**, eds. G. Murch and A.S. Nowick, Academic Press, New York, 1984.
21. A. Borgenstam, A. Engstrom, L. Hoglund, J. Agren, J. Phase Equilib. 21, 269(2000).
22. C.E. Campbell, J-C. Zhao, M.F. Henry, J. Phase Equilib. Diff. 25, 6(2004).
23. M.E. Glicksman, **Diffusion in Solids**, John Wiley & Sons, New York, 2000.

24. E. Campbell, W.J. Boettinger, U.R. Kattner, Acta Mater. 50, 775(2002).
25. C.E. Campbell, W.J. Boettinger, T. Hansen, P. Merewether, B.A. Mueller, **Complex Inorganic Solids: Structural, Stability, and Magnetic Properties of Alloys**, eds. P.E.A. Turchi, A. Gonis, K. Rajan, A. Meike, Springer, New York, 2005.
26. K. Wu, J.E. Morral and Y. Wang, Acta Mater. 49, 3401(2001).
27. W.D. Hopfe and J.E. Morral, Acta Mater. 42, 3887(1994).
28. P.E. Rouse, J. Chem. Phys. 21, 1273(1953).
29. P.G. deGennes, J. Chem. Phys. 55, 572(1971).
30. J. Kas, H. Strey and E. Sackmann, Nature 368, 226(1994).
31. H. Schiessel et al., Phys. Rev. Lett. 86, 4414(2001).
32. W. Paul et al., J. Chem. Phys. 95, 7726(1991).
33. P. Gas and F. D'Heurle, MRS Symp. Proc. 402, 39(1996).

Bibliography

1. J.P. Stark, **Solid State Diffusion**, John Wiley, New York, 1976.
2. For diffusion data see Diffusion and Defect Data, eds. F.H. Wohlbier and D.J. Fisher, Trans. Tech. Publications, Aedermannsdorf, Switzerland.
3. G. Martin and B. Perrhaillon, in **Grain Boundary Structure and Kinetics**, ASM, Metals Park, Ohio, 1979.
4. R.J. Borg and G.J. Dienes, **An Introduction to Solid State Diffusion**, Academic, NY, 1988.
5. I. Kaur and W. Gust, **Fundamentals of Grain and Interphase Boundary Diffusion**, Zeigler Press, Stuttgart, 1989.
6. I. Kaur, W. Gust and L. Kozma, **Handbook of Grain and Interphase Boundary Diffusion Data**. Zeigler Press, Stuttgart, 1989.
7. R. Ghez, **A Primer of Diffusion Problems**, John Wiley & Sons, New York, 1988.
8. J. Philibert, **Atom Movements: Diffusion and Mass Transport in Solids**, Les Ulis, France, 1991.

Metals

J.L. Bocquet, G. Brebec and Y. Limoge, in Physical Metallurgy, 3rd Edition, eds. R.W. Cahn and P. Haasen, North-Holland, Amsterdam, 1983.

Ordered alloys and intermetallic compounds

H. Nakajima, W. Sprengel and K. Nonaka, Intermetallics 4, 17(1996).

M. Koiwa, H. Numakura and I. Ishioka, Defect Diffus. Forum 143–147, 209(1997).

Ionic crystals

P. Kofstad, **Nonstoichiometry, Diffusion and Electrical Conductivity in Binary Metal Oxides**, John Wiley, New York, 1972.

F. Beniere and C.R.A. Catlow, **Mass Transport in Solids**, Plenum Press, New York, 1981.

A. Atkinson, Rev. Mod. Phys. 57, 437(1985).

P. Kofstad, **High Temperature Corrosion**, Elsevier Applied Science, NY, 1988.

Polymers

I. Teraoka, **Polymer Solutions**: **An Introduction to Physical Properties**. John Wiley & Sons, Inc., 2002.
M. Doi and S.F. Edwards, **Theory of Polymer Dynamics**, John Wiley, New York, 1987.

Problems

1. Distinguish between how intrinsic diffusivities are measured and how chemical diffusivities are measured.
2. Can the chemical diffusivities of two species in a binary alloy differ? Explain your answer.
3. Will there be a Kirkendall effect if the intrinsic diffusivities in a binary diffusion couple are not equal?
4. In which direction will an inert marker move if the intrinsic diffusivity of component B is larger than that of component A? In this case, in which direction will there be a net drift of vacancies?
5. What is the solution for C(r, t) for diffusion into a cylindrical bar from a thin plate on the surface of thickness ΔR, having a concentration C_i, when at time t_o the concentration everywhere in the bar is C_o?
6. What is the significance of the term phonon drag?
7. Why is it possible for voids to form in a diffusion couple in the absence of heat or current flow?
8. Give one reason for enhanced atom transport in a region under ion bombardment other than the formation of short circuit defects.
9. If one component of a solid solution is normally interstitially sited why would it be expected that its activation energy for diffusion will be less than that for the substitutionally sited component?
10. Why should you expect that the temperature required to change from a random solid solution to a short-range ordered solid solution to be less than that for the homogenization of a spinodally decomposed solid solution having a wave-like composition profile?
11. What would happen to the measured activation energy for diffusion if the vacancy concentration was maintained constant at some supersaturated level?
12. Within what range of values is the expected value for the pre-exponential coefficient of the lattice diffusivity? If the apparent value for the latter quantity falls outside this expected range how would you explain this result?
13. Suppose there is strong association between an impurity atom and a vacancy and that the frequency of vacancy–host atom interchange is greater than impurity–vacancy interchange, what value does the correlation factor have? If this inequality is reversed what does the correlation factor become and what frequency then controls the diffusion of the impurity?
14. The diffusion coefficient of the host atom at the temperature and composition of the solidus is roughly a constant value (e.g. for a fcc crystal it is about 5×10^{-9} cm^2/seconds). Provide an explanation of this empirical result.

15. Grain boundaries in metals vary in their associated values of grain boundary diffusivity from values even smaller than that for lattice diffusivity to those orders of magnitude larger than the latter. Explain this observation.
16. One consequence of the observation in the previous question is that at a grain boundary triple junction it is likely that div(J_I)=/0, where J_i represents a vector flux in the ith grain boundary, for a constant external driving force. What do you expect will occur at the triple junction in this event?
17. Describe several possible reasons why the gradient of concentration along a grain boundary will not equal a constant in the surface accumulation technique of measuring the grain boundary diffusivity.
18. How would a measurement of the activation volume for diffusion along grain boundaries help interpret whether the mechanism for grain boundary diffusion involves vacancy migration or not?
19. (a) Estimate the tracer diffusion coefficient for nickel at 900°C.
    (b) What would this diffusion coefficient equal in a 90 at.%Ni–10 at.%Al alloy at 900°C?
    (c) Estimate the self-diffusion coefficient at a high angle grain boundary in nickel at 900°C.
20. Given that the activity of component B increases with concentration of component C in a ternary alloy of A, B and C. Suppose a diffusion couple is constructed in which the concentration of component B is the same initially in both halves of the diffusion couple. Let the left half of the couple initially have C(C) = 5%, while the right half has C(C) = 10%. Further suppose that the diffusivity of C much lesser than that of B. Show the concentrations of B and C along the diffusion couple assuming local equilibrium is reached at the original join, but that equilibrium is not attained away from this join upon heating for a finite time at some elevated temperature sufficient to achieve transport of B.

CHAPTER IX

Nucleation and Growth Kinetics

Introduction

We have developed in Chapters V and VIII the knowledge needed to derive rates of nucleation and spinodal decomposition. The steady-state nucleation rate is proportional to the product of the equilibrium distribution of nuclei and the rate of attachment of molecules to the nuclei, which for nucleation in condensed phases involves diffusion or diffusion-like jumps across an interface. The time dependence of the nucleation rate follows from a consideration of the mode of growth and shrinkage of embryos. Experimental verification of the theory is very difficult and to the extent that it has been carried out does not contradict the theory.

The rate of spinodal decomposition depends on the thermodynamic driving force and diffusion over a penetration distance corresponding to a quarter wavelength of a composition wave and can be shown to have a maximum at a particular wavelength of the periodic composition wave developed in this mode of decomposition of a metastable host.

Growth of nuclei to produce precipitates and other product phases needs to consider depletion of the solute in the parent metastable phase for the case of partitioning modes of decomposition of the parent phase and impingement of the product phases. This problem is considered from its original classical analytical treatment to more modern treatments based on the availability of computers to simulate processes as well as to solve non-linear coupled equations numerically.

1. Rate of heterophase nucleation

1.1. Homogeneous nucleation theory and its experimental verification

One approach to obtaining a relation for the steady-state rate of nucleation assumes that an equilibrium distribution of nuclei in the untransformed volume can be maintained despite the draining off of nuclei by their growth. In this approach the steady-state nucleation rate becomes this equilibrium number multiplied by the rate at which nuclei accrete molecules, i.e.

$$J^*_{eq} = \beta^* N \exp[-\Delta G^*/kT] \tag{9.1}$$

where N $\exp[-\Delta G^*/kT]$ is the equilibrium number of nuclei per untransformed volume and β^* is the number of atoms or molecules within a jump distance of an attachment site about the critical embryo multiplied by the jump frequency toward the embryo. The latter frequency will depend upon the nature of the embryo. For example, in the case of nucleation of a pure solid from a pure liquid the jump frequency will have the form

$$\nu = \nu_o\, e^{-Q/kT} \tag{9.2}$$

where Q is the activation energy for the jump that leads to the attachment of one atom in the liquid to the solid embryo.* In the case of formation of a solute rich precipitate Q will be the activation energy for the jump of a solute atom from a position in the lattice to a position on the surface of the critical precipitate embryo. If the latter is coherent with the matrix, then Q is the activation energy for diffusion of the solute in the matrix. Only in the case of nucleation of a condensed phase from a vapor phase will the relation for β^* not contain an activation energy term. In the latter situation, β^* is obtained from the kinetic theory of gases as the collision rate of molecules with the condensed embryo.

Thus, the steady-state rate of heterophase nucleation in condensed phases in the above approach is temperature dependent and of the form

$$J^* = J^o \exp[-(Q + \Delta G^*)/kT] \tag{9.3}$$

This relation only attempts to describe the steady-state rate of nucleation. However, it is known that the rate of nucleation can be time dependent subsequent to a change in the boundary conditions (e.g. temperature) and neither equation 9.3 nor its derivation provides an expected theoretical time dependence of the nucleation rate. Thus, an alternate approach to the derivation of equation 9.3 is required to attempt to describe this time dependence of the nucleation rate.

The classical alternate approach, originally due to Becker and Doring,[1] focuses on the reactions involved in the growth and shrinkage in size of the embryos having n molecules per embryo. This approach yields a system of coupled differential equations of the form

$$[dN_{n,t}/dt] = k^+_{n-1}N_{n-1,t} - [k^-_n N_{n,t} + k^+_n N_{n,t}] + k^-_{n+1}N_{n+1,t}\cdots \tag{9.4}$$

---

* Another possible accretion mode is via a cooperative motion that adds several atoms to the nucleus, a concept first suggested by Smoluchowski.

where k_n^+ is the rate of addition of molecules to an embryo of size n, and k_n^- is the rate of loss. The time-dependent nucleation rate is given by

$$J_n(t) = k_n^+ N_{n,t} - k_{n+1}^- N_{n+1,t}$$

The solution for the time-dependent nucleation rate upon substitution and appropriate approximation is

$$J^*(t) \cong J^* \, e^{-\tau/t} \tag{9.5}$$

where $\tau \cong -4kT/[\beta^*(\partial^2(\Delta G^o)/\partial n^2)|_{n^*}]$, β^* is the rate at which single atoms impinge upon the critical embryo, ΔG^o is the Gibbs free energy of formation of an embryo containing n atoms. The steady-state rate of nucleation is given by

$$J^* = Z\beta^* N \exp[-\Delta G^*/kT] \tag{9.6}$$

where Z is the Zeldovich[2] factor given by

$$Z = [-(1/(2\,kT))(\partial^2(\Delta G^o)/\partial n^2)|_{n^*}]^{1/2}$$

The product of the exponential term in equation 9.6 by N is simply the equilibrium number of heterophase embryos of critical size (see equation 5.6) per unit volume. The Zeldovich factor Z, times this number is the actual number of critically sized embryos per unit volume. The Zeldovich factor has a value of about 1/20. Solutions for Z and β^* have been given[3] for various nuclei shape.

Comparison of equations 9.1 and 9.6 for the steady-state nucleation rate reveals that they differ only by the inclusion of the Zeldovich factor in the latter. The practical significance of this difference is difficult to demonstrate. However, the time dependence of the nucleation rate can be experimentally observed on rapidly changing from one equilibrium distribution of embryos to another by rapid change of the temperature. The time dependence in equation 9.5 is found to be valid in the region near to the approach to the steady state.

There have been many tests of the validity of the theory of nucleation described above and which is now called the classical nucleation theory (CNT) in the literature. There still exists questions concerning the range of application of CNT. Auer and Frenkel,[4,5] in part to answer these questions, proposed that the ideal hard sphere system for which the thermodynamics are known and can be calculated precisely is an ideal system for studying nucleation. They were able to simulate the formation of a statistical distribution of embryos in a metastable liquid of the hard sphere system using a Monte Carlo technique. From the resulting numbers of embryos as a function of the embryo size they were able to calculate the work to form embryos as a function of the number of hard spheres in the embryo using equation 5.6.

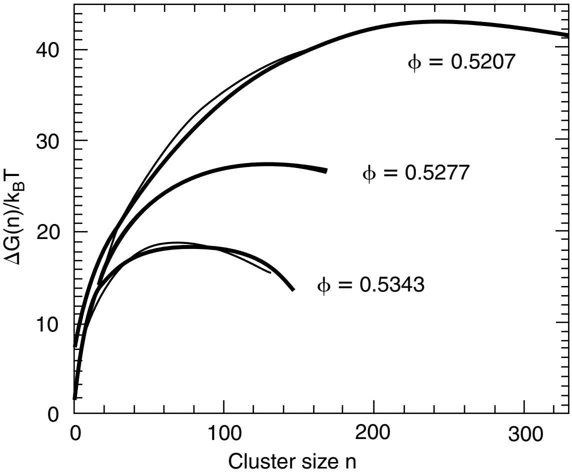


Figure 9.1a. Free energy to form embryo as a function of embryo size n as deduced from counts of embryo population is Monte Carlo simulation of hard sphere fluid.* From J. Chem. Phys. 120, 3015(2004) with permission. © 2004 American Institute of Physics.

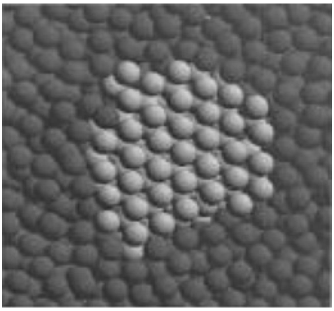


Figure 9.1b. Snapshot of a critical nucleus of a hard sphere crystal. From J. Chem. Phys. 120, 3015(2004) with permission. © 2004 American Institute of Physics.

Their results are the light lines in Figure 9.1(a). The three curves correspond to three different parent metastable hosts having different volume fractions of hard spheres. The result shown in Figure 9.1(a) is the first quantitative test of equations 5.6 and 5.7 that has been reported to the date of this writing and the results are consistent with these relations. The other of their results that agrees with the assumptions of CNT is that the embryos were found to have a definite crystal structure, the order of which did not vary over an appreciable fraction of the embryo volume, i.e. the transition from crystal to liquid is fairly sharp (see Figure 9.1(b)).

The interface energies corresponding to these hosts were evaluated using equation 5.7. These calculated interface energies correspond to planar interfaces between the nucleus and the metastable host liquid. Values of the planar interface energies have been independently calculated as a function of metastable liquid density using several methods including density functional theory (DFT).[6] These values are shown in Figure 9.1(c). The

---

* Light lines deduced from measurements of number of embryos at equilibrium using equation 5.6. Dark lines fitted using equations 5.7–5.10 to obtain values of interface energy.

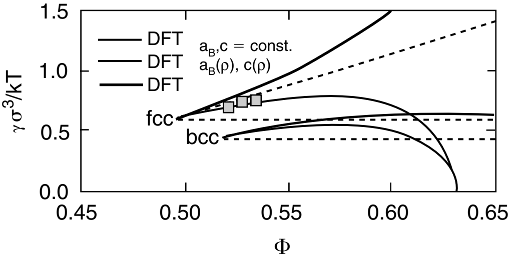


Figure 9.1c. Lines represent theory deduced values of reduced interface energy as a function of volume fraction of hard spheres. From J. Chem. Phys. 117, 10121(2002) with permission. © 2002 American Institute of Physics.

values obtained by Auer and Frenkel[4,5] are shown as square points in this figure. Hence, there appears to be agreement between the planar interface energies deduced from the Monte Carlo simulation and CNT with the values predicted by two of the methods. For one of these methods, the center of the nucleus is assumed to be at the same conditions as is applicable in CNT. Hence, it appears that the interface energy deduced by Auer and Frenkel using the relations of CNT agrees with one calculated on the basis of a diffuse interface between the conditions of a nucleus phase that is coexistent with the host phase at the center of the nucleus and a metastable host. Now, this result is in agreement with the relations of CNT but modified to be in agreement with Gibbs original concept of the interface between the nucleus and metastable host. Although Gibbs is just as classical as those who assumed this interface to have the value between coexisting phases, Cahn and Hilliard who, I believe, were the first to call this assumption that of CNT, reiterated Gibbs formulation of this interface as that between stable nucleus and metastable host and also pointed out that for mild supersaturations the formulations of equations 5.6 and 5.7 would hold, which is what the data in Figures 9.1(a) and (c) affirm. So we may conclude that the results of Auer and Frenkel's simulation are consistent with Gibbs nucleation theory (GNT), if not of those who made the assumption that the interface is between coexisting phases at stable equilibrium. (GNT has the same formulations as equations 5.6–5.9 but the interface involved in these relations is that between the nucleus phase and the metastable host phase with the molecules in the nucleus phase having the same chemical potential as their correspondents in the metastable host phase.)

Auer and Frenkel[4,5] then used the work of nucleation deduced from equation 5.8 to evaluate nucleation rates which they then compared to experimental rates of nucleation. Those based on GNT were low compared to the experimental rates. However, there are several possible reasons for this discrepancy which have nothing to do with the validity of GNT.

One of the possible contributions to this discrepancy is the intervention of a metastable equilibrium between the metastable host state and the stable equilibrium state. When this occurs there can be an appreciable effect on the operating work of nucleation. Talanquer and Oxtoby[7] have investigated the effect of a metastable fluid→fluid 1 + fluid 2 transition intervening in a stable fluid→ solid transition. They found that the nucleation rate of the stable product was strongly enhanced at conditions close to a metastable critical point. Let us explore the reasons for this result.

In their model it is apparent that, for the conditions assumed, the work of nucleation is at a minimum at the metastable critical point. Figure 9.2 shows the curve evaluated for one set of assumed conditions in the transition considered by Talanquer and Oxtoby.[7] Apparently, the main effect of the metastable critical point in this case is on the interface energy. However, the sequence of phase transitions in the critical embryo varies from above the critical point to below it. Above, the first step in the transition is the formation of liquid-like clusters which above a critical size transform rapidly to crystals. Below the critical point, solid-like clusters are formed which are wetted by a liquid-like layer. However, the results are quite sensitive to the assumed conditions. Nevertheless, one possible reason for the discrepancy between predicted and experimental rates of nucleation for ideal hard sphere systems is that there may be a metastable critical point in the experimental cases where the systems may not really be *ideal* hard sphere systems. The qualitative results just described were predicted by Ostwald[8] many years ago. The concept is now known as Ostwald's step rule, which is that the presence of intermediate metastable equilibria between the host state and the stable state will influence heterophase nucleation by introducing intermediate stages of nucleation in the sequence of least stable to more stable. However, this phenomenon is not really a contradiction to CNT or GNT since each stage of nucleation may proceed via CNT or GNT. Whether CNT or GNT is applicable or not to each stage depends on the conditions of the host state relative to those for the critical points and spinodals of the metastable states. As we will learn later,

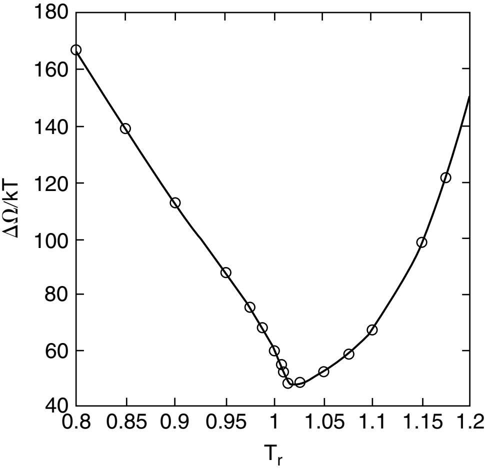


Figure 9.2. Work of nucleation as a function of temperature relative to that for the metastable critical point at constant classical work of nucleation.* From J. Chem. Phys. 109, 223(1998) with permission. © 1998 American Institute of Physics.

---

* In Figure 1c by "classical work of nucleation" is meant the work of nucleation in the absence of a metastable critical point, i.e. the critical nucleus has the configuration of the stable solid phase and is surrounded by the metastable parent phase.

contrary to experiment, CNT usually predicts a finite work of nucleation at the spinodal, whereas this work approaches zero at the spinodal as it does in GNT as well.

1.1.1. Temperature dependence of the steady-state nucleation rate

One of the qualitative behaviors predicted by CNT has been amply verified. This behavior is the temperature dependence of the nucleation rate. From Chapter V we may obtain the knowledge that ΔG^* is itself temperature dependent and for the case of heterophase nucleation from a metastable parent phase, where the latter is the stable phase above the transition temperature and nucleation is occurring below the transition temperature, ΔG^* decreases as the temperature decreases below the transition temperature. For example, for the case of homogeneous nucleation of spherical precipitates, $\Delta G^* = (16/3)\sigma^3/(g_h - g_l)^2$ and to a first approximation $g_h - g_l = (s_h - s_l)(T_e - T)$. Since both the numerator and the denominator in the exponent of equation 9.6 decrease with decreasing temperature, it is to be expected that J* will reveal a maximum and in fact it does. If we define a time t′as the time required to nucleate a given number N′, of product phase particles per unit volume, then t′ = N′/J*. Hence, if J* reveals a maximum as a function of temperature then t′ will reveal a corresponding minimum, as illustrated in Figure 9.3. As shown, this dependence of t′is the same as that found in studies of the rates of transformation, where the time to achieve a given fractional volume of the product phase t_f, exhibits the so-called C curve behavior in a plot of transformation temperature versus time. Although the latter curves have a more complex dependence of t_f on T than of t′on T, due to the effect of impingement of individual growing nuclei, so

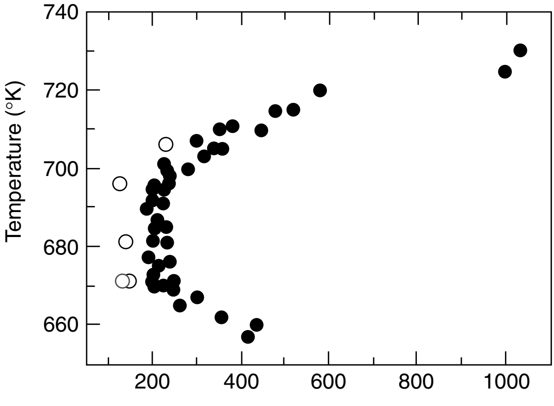


Figure 9.3. A temperature–time transformation curve for $Pd_{43}Ni_{10}Cu_{27}P_{20}$ on fluxed (●) and unfluxed (○) material. From Appl. Phys. Lett. 77, 1158(2000) with permission. © 2000 Amercian Institute of Physics.

that $t_f \neq t'$, the presence of a minimum in t_f may or may not be due to the minimum in t'. We shall have more to say on this matter later.

The temperature corresponding to the maximum in J* or the minimum in t' can be obtained by differentiating J* with respect to T and setting the derivative equal to zero. The result is that this temperature T* is defined by

$$d[Q + \Delta G^*(T^*)]/dT = [Q + \Delta G^*(T^*)]/T^*$$

The existence of a minimum in the time to achieve a certain transformed volume cannot be taken as proof that this minimum corresponds to a maximum in the nucleation rate. Another process upon which the transformed volume depends, such as the growth rate, may also exhibit a turning point. For example, suppose that the growth rate can be expressed as a product of the driving force for growth and the mobility for the growth process. Also, suppose that the host phase is stable at high temperature while the product phase is stable at low temperature. In this case, as the temperature decreases below that corresponding to equilibrium between the two phases the driving force for growth increases while the mobility for growth decreases. Thus, the growth rate of the stable phase into a metastable host will exhibit a maximum as the temperature decreases below the corresponding equilibrium temperature between the two phases. In general, the temperatures corresponding to the maximum nucleation and growth rates will differ, as illustrated schematically in Figure 9.4. Such separation of nucleation and growth rates has been observed in glass systems.[9] The ease of forming glasses is enhanced when the temperature corresponding to the maximum rate of nucleation is less than that corresponding to the maximum in the growth rate.

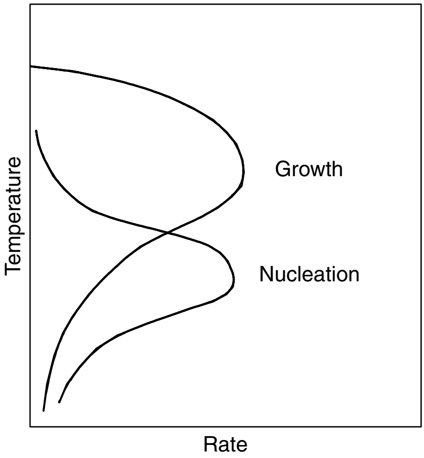


Figure 9.4. Illustrating the different temperatures for the maximum rates of growth and nucleation.

The expressions given in Chapter V for the work of nucleation apply to sufficiently large embryos where the surface energy and specific free energy of the embryo does not depend on its size. Nucleation of phase transitions is a general phenomenon, not limited to any material and the relations given in this section are applicable to all materials.

1.2. Heterogeneous nucleation

The overwhelming incidence of nucleation in practise is heterogeneous or catalyzed nucleation. Such heterogeneous nucleation occurs at defects within crystals, at interfaces and at any other center of heterogeneity in a parent phase. The equations applicable to homogeneous nucleation for the rate of nucleation are also applicable to heterogeneous nucleation with corrections for the work of nucleation and, where applicable, for the accretion rate.

Where the effect of the defect or interface can be described in terms of a wetting angle then the equations derived in Chapter V (i.e. equations 5.18–5.22) are probably applicable. However, experiments reveal that grain boundaries are not homogeneous and any grain boundary may contain sites that have more excess energy than that corresponding to the average energy of the boundary. These sites of excess energy seem to be the sites at which heterogeneous nucleation along such boundaries occurs. Further, at least one of the interphase interfaces of heterogeneous nuclei at grain boundaries may be a coherent boundary so that the shape of the nucleus is not that of a lens as assumed in the derivation of the factor $f(\theta)$ in equation 5.18.

The existence of anisotropic interface energies is another factor complicating the quantitative application of the nucleation rate equations. However, the factors mentioned do not affect the principles of the theory of nucleation, but rather do affect the practise of it, or the interpretation of experimental nucleation kinetics in terms of it.

This section has considered one extreme of the spectrum of nucleation – that of CNT and of the GNT that is applicable at low supersaturations. In the next section we will consider the other extreme of this spectrum – that associated with the transformation of an unstable phase to a more stable equilibrium.

2. Spinodal decomposition

It is possible to show that the quantity φ of equation 8.13 is equal to $C_1C_2(d^2g/dC_2^2)/kT$ and hence by equation 8.8 that the chemical diffusivity also equals

$$
\begin{aligned}
{}^{o}D &= (N_2D_1^* + N_1D_2^*)C_1C_2(d^2g/dC_2^2)/kT \\
{}^{o}D &= M\, d^2g/dC_2^2
\end{aligned}
$$

where g is the molar free energy of the solution and the D_i^* are tracer diffusivities. Thus, the chemical diffusivity has the form of a product of a mobility term M and a driving force d^2g/dC_2^2. It is possible for the latter to have a negative value in a metastable solution. This fact is responsible for the so-called “uphill diffusion” with a negative diffusivity[10] that occurs in the spinodal decomposition of a metastable phase.

In Chapter V we noted that the Helmholtz free energy F for a solid solution containing gradients of composition is given by

$$F = \int_V [f'(C) + \eta^2 Y(C - C_o)^2/(1 - \nu) + K(\nabla C)^2]dV$$

The change in this free energy due to a change in the composition profile is obtained by taking the differential

$$\delta F = \int_V \{[(\partial f'/\partial C) + 2\eta^2 Y(C - C_o)/(1 - \nu) + (\partial K/\partial C)(\nabla C)^2]\delta C \\ + 2K\nabla C\delta(\nabla C)\}dV$$

Integrating the latter term in the right-hand side of this equation by parts yields

$$\delta F = \int_V [(\partial f'/\partial C) + 2\eta^2 Y(C - C_o)/(1 - \nu) - (\partial K/\partial C)(\nabla C)^2 \\ - 2K\nabla^2 C]\delta C\, dV$$

The term in the square bracket times δC is the change in free energy due to a local change of composition δC. The driving force for change of the composition profile is then just the gradient of the term in the square bracket. We can thus set the flux of component C to be equal to

$$J = -M\, \mathrm{grad}[(\partial f'/\partial C) + 2\eta^2 Y(C - C_o)/(1 - v) - (\partial K/\partial C)(\partial C)^2 - 2K\, \nabla^2 C] \\ \equiv -M\, \mathrm{grad}[A]$$

where M is the mobility coefficient. We now note that

$$-\mathrm{div}J = \partial C/\partial t = \mathrm{div}\{M\, \mathrm{grad}[A]\}$$

Considering only the initial stages of decomposition and neglecting all terms not linear in C yields

$$\partial C/\partial t = M\{(\partial^2 f'/\partial C^2) + 2\eta^2 Y/(1 - \nu)\}\nabla^2 C - 2MK\, \nabla^4 C$$

The solution of the latter equation is

$$C - C_o = B(\beta, t)\cos(\beta d)$$

where $B(\beta, t)$ obeys

$$\partial B/\partial t = -M\{(\partial^2 f'/\partial C^2) + 2\eta^2 Y/(1-\nu)\}\beta^2 B - 2MK\beta^4 B$$

according to which

$$B(\beta, t) = B(\beta, 0)\exp[R(\beta)t]$$

where

$$R(\beta) = -M\beta^2[(\partial^2 f'/\partial C^2) + 2K\beta^2 + 2\eta^2 Y/(1-\nu)]$$

The characteristics of R, called the amplification factor, are that when the term [ ] defined by the term in the square brackets in the latter equation obeys $[\ ] > 0$ then $R < 0$ and the solution is stable with respect to a fluctuation of composition; $[\ ] = 0$ then $R = 0$; $[\ ] < 0$ then $R > 0$ and the solution is unstable with respect to a fluctuation of composition.

In the above it should be recalled that $[\ ] > 0$ when $\beta > \beta_c$ or $\lambda < \lambda_c$ and vice versa for reversal of the inequalities. Now for $[\ ] < 0$, the larger is R, the larger is B and the larger is the change in the composition achieved in a given time t at a given position. Hence, that fluctuation that yields the maximum value of R will grow fastest. Solving for the maximum R with respect to β yields that

$$\beta_{max} = \beta_c/\sqrt{2}$$

and

$$R(\beta_{max}) = 2MK(\beta_{max})^4 = MK(\beta_c)^4$$

Since $(\beta_c)^2 = -(1/2K)[(\partial^2 f'/\partial C^2) + 2\eta^2 Y/(1-\nu)]$, while the second derivative term in the parenthesis is negative and the second term is positive, then the smaller is the elastic modulus Y, the larger is β_c and R. This dependence on Y has been substantiated experimentally using observations of the orientation dependence of Y.

Physically, a maximum in the spinodal decomposition rate as a function of the wavelength may be explained as follows. As the wavelength increases from the critical wavelength value the driving force for spinodal decomposition increases, as shown in Figure 9.5. However, in the same variation the distance over which the components have to diffuse increases and hence it will take longer to achieve a given local change in composition. The spinodal decomposition rate is proportional to the quotient of the driving force over the relaxation time associated with a given local composition change. Hence, there is a maximum in the decomposition rate as may be deduced from Figure 9.5.

The above classical treatment of spinodal decomposition due to Cahn and Hilliard[11] is applicable to the linear stage of the decomposition process, which occurs early in the decomposition process. The later non-linear stage has been treated by other authors, see Bibliography. Experimentally, the kinetics of the linear stage have been confirmed for the conditions where the linear theory is applicable.

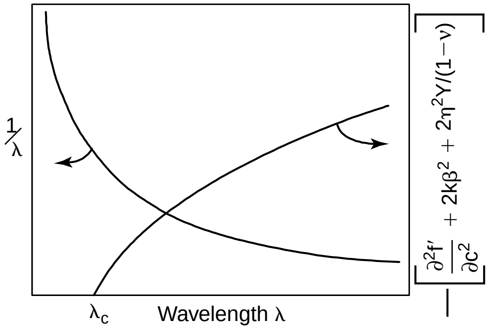


Figure 9.5. Illustrating the dependence on wavelength of the inverse diffusion length and the driving force for the diffusion in spinodal decompostion.

3. Nucleation at high supersaturation but not beyond spinodal

Having considered the two extreme initial conditions of the phase hosting nucleation it is now appropriate to consider the conditions of nucleation in between these extremes. The approaches to this problem may be classified into two groups. One group attempts to modify the CNT model. The other group applies either the extension of the concept of diffuseness (inhomogeneity) of properties to all properties including density (DFT since density was the first property this extension considered) or develops entirely new methods of treating this problem.

The basis for the attempt to extend the CNT model is to use Gibbs thermodynamics of capillarity, as applied by Hill, to treat the Helmholtz free energy of a heterogeneous system by setting up two homogeneous systems separated by an interface (dividing surface) and to maintain the conditions defining the dividing surface constant during changes in the physical state of the system. The nucleus, which is now a construct of the actual diffuse fluctuation, has then the properties of a bulk phase with temperature and chemical potentials the same as those of the host phase. The dividing surface Gibbs called the surface of tension is defined by the relation

$$2\gamma_T/R_T = p_n - p_h \equiv \Delta p$$

where p_n is the pressure of the bulk nucleus phase having the temperature and chemical potentials of the host phase while p_h is the pressure of the host phase. Also, g is the specific interface energy of the surface of tension and R is its radius. In this relation the pressures on the right-hand side are independent of the location of the dividing surface while both γ and R depend upon it. Hence, this relation acts to define both γ_T and R_T of the surface of tension. For the surface of tension the work of nucleation can be written as

$$\Delta G^* = C\gamma_T^3/\Delta p^2$$

where the geometrical factor $C = 16\pi/3$ for a spherically shaped nucleus. Now, Gibbs showed that $\gamma_T = 0$ at the spinodal and hence ΔG^* also equals zero at this condition. Thus, with this expression for W all that is necessary to obtain values for it for conditions involving diffuse fluctuations is to determine the corresponding values of Δp and either γ_T or R_T. In principle, a knowledge of the thermodynamics of the system allows Δp to be evaluated. However, it is much more difficult to obtain γ_T or R_T values. One way that has been used to overcome this barrier is to determine a scaling function that satisfies the thermodynamic constraints and yields the values known to be valid for the two extremes of small supersaturation and spinodal conditions. Kashchiev has found one such approximate function. The corresponding work of nucleation is shown in Figure 9.6. It is shown there normalized with respect to the CNT value of the work of nucleation as a function of the ratio of the overpressure corresponding to the conditions of the metastable host to that at the spinodal. It is not known at this writing how well this scaling function agrees with experiment.

Reguera and Reiss have found for the case of the liquid–vapor system that modeling a diffuse fluctuation again as a system of homogeneous embryo within another homogeneous host that has a physically defined volume yields results that are in remarkable agreement with simulation results and experiment. The definition of the volume of this system is that it is the volume at which the rate of accretion equals the rate of evaporation – the volume of the critical nucleus –

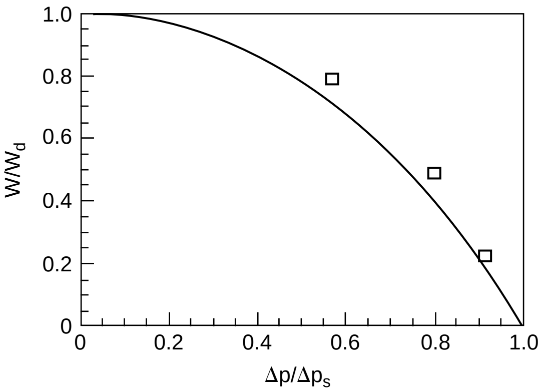


Figure 9.6. Normalized work of nucleation versus normalized overpressure. From J. Chem. Phys. 120, 3749(2004) with permission. © 2004 American Institute of Physics.

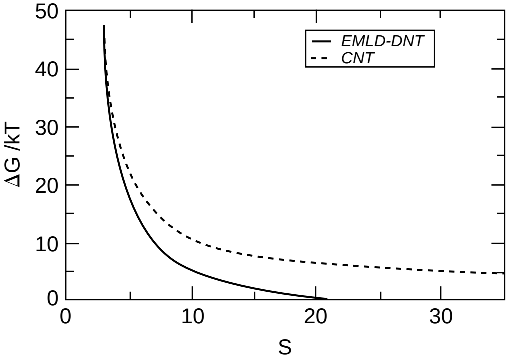


Figure 9.7. Work of nucleation as a function of saturation ratio for Argon. From Phys. Rev Lett. 93, 165701(2004) with permission. © 2004 American Physical Society.

as well as it being the volume at a minimum in the P–V relation for the system. Using this model they were able to calculate the work of nucleation shown in Figure 9.7. We have used these data to calculate the corresponding values for Figure 9.6, which are shown as square points in the figure. We made use of the equation of state for argon from Reference 12. It is apparent that the scaling function for the work of nucleation obtained from the model of Reguera and Reiss yield slightly higher values of the work of nucleation than those of Kashchiev. The applicability of the scaling function needs much more study.

The primary other approach to study the realm of metastable phase decomposition in the regime from medium saturation to the spinodal saturation has been DFT. DFT has been applied as a method of solving the nucleation problem for such situations mainly by Oxtoby,[13] his students and collaborators. Also, there have been additional contributions involving dynamical nucleation theory,[14] scaling relationships,[15] the concept of an extended modified liquid drop[16] and combinations of these models.[17] Further, there have been simulations which contributed new knowledge about nucleation. Most of the developments have been related to the work of nucleation. Some concern the kinetic aspects of nucleation.

3.1. DFT and nucleation

DFT came about as an extension of the concept pioneered by Cahn and Hilliard[18] that the interface between a nucleus and host may be diffuse. Since the first application was to the vapor–liquid transition this led to the consideration of use of density as the order parameter transitional between these two phases. Just as the Helmholtz free energy exhibits the functional form as a function of composition shown in Figure 5.8 it will exhibit a similar form as a function of density between the densities that define the two phases.

The objective of DFT in nucleation theory is to provide a basis for constructing a free energy functional for a non-homogeneous system or phase. Before we discuss the details of DFT it is worthwhile to consider some thermodynamic relations which will enable us to make a smooth transition between CNT and DFT. We note that in CNT we make use of the Gibbs free energy G, in evaluating the work of nucleation. In DFT it is more convenient to use the grand potential Ω, and the grand canonical ensemble to evaluate the work of nucleation. Let us consider the relations between these potentials in evaluating work of nucleation. Consider the case of CNT. In a single component system with constant μ, V and T, the grand potential for a two-phase system is

$$\Omega = -p_l V_l - p_v V_v + 4\pi R^2 \gamma$$

where $V_l = (4/3)\pi R^3$, $V_v = V - (4/3)\pi R^3$, and γ is the energy per unit area for a flat interface separating the v and l phases. Substituting for V_l and V_v and setting the derivative of Ω with respect to R equal to zero yields

$$p_l - p_v = 2\gamma/R_o$$

to define the radius of the droplet phase that is mechanically stable under the difference in pressures across the interface. Now for a system, the Legendre relations between the thermodynamic potentials are $G = F + pV$, $\Omega = F - \mu N$. Hence, we may write $G = \Omega + \mu N + pV$. Since the pressure on the system is $p = p_v$, and the corresponding chemical potential is $\mu = \mu(\rho_v)$ then the work to form the droplet is $\Delta G = \Omega + \mu(\rho_v)N$. But, the original system consisting only of vapor in the volume V has the grand potential $\Omega_o = -p_v V$. Hence, the difference in the grand potential for the system containing the drop less the uniform vapor throughout the volume V is $\Delta\Omega = \Omega - \Omega_o = \Omega + p_v V$. Substituting in the relation for G, then $G = \Delta\Omega + \mu(\rho_v)N$. However, $\Delta G = G - G_o = G - \mu(\rho_v)N$. Hence, $\Delta G = \Delta\Omega$.

In DFT the work of nucleation is the same in that $\Delta\Omega$ equals the grand potential for the system of nucleus plus host phase in the volume V at the nucleation temperature less the grand potential for the metastable host phase in volume V in the metastable vapor state. In DFT, the density at the center of the nucleus need not be that of the new bulk phase, nor must the surface behave like a planar interface. Instead, the critical nucleus density profile $\rho^*(r)$ is the saddle point in functional space separating “reactants” (small clusters that tend to shrink) from “products” (large clusters that tend to grow). At the saddle point, the functional derivative $\delta\Omega/\delta\rho$ vanishes, giving a mathematical equation for the critical nucleus profile. This result is a non-linear integrodifferential equation that can be solved for the density as a function of position relative to the center of the nucleus. The reader is advised to read the literature to learn the details of the several procedures used to obtain numerical solutions of this mathematical problem. These procedures vary depending in part upon the number of order parameters involved for the system being explored. For the case of the single order parameter density, an iteration procedure has been used to arrive at a solution. In this case, the Helmholtz free energy is taken to be a functional of the density having the form

$$F[\rho(r)] = \int d\mathbf{r} f_h[\rho(\mathbf{r})] + (1/2)\int\int d\mathbf{r}\, d\mathbf{r}'\rho(\mathbf{r})\rho(\mathbf{r}')\, w_2(|\mathbf{r} - \mathbf{r}'|)$$

where $f_h[\rho]$ is the Helmholtz free energy per unit volume of a uniform hard sphere fluid of density ρ, equal to

$$f_h(\rho) = \rho\mu_h(\rho) - p_h(\rho)$$

The hard sphere chemical potential μ_h and pressure p_h are evaluated using the Carnahan–Starling[19] relations. The interatomic potential was taken to have the Yukawa form. The grand potential then has the functional form

$$\Omega[\rho(r)] = \int d\mathbf{r} f_h[\rho(\mathbf{r})] - \mu\int d\mathbf{r}\rho(\mathbf{r}) + (1/2)\int d\mathbf{r}\, d\mathbf{r}'\rho(\mathbf{r})\rho(\mathbf{r}')w_2(|\mathbf{r} - \mathbf{r}'|)$$

The result of setting $\delta\Omega/\delta\rho = 0$ is $\mu_h[\rho(\mathbf{r})] = \mu - \int d\mathbf{r}'\rho(\mathbf{r}')w_2(|\mathbf{r} - \mathbf{r}'|)$. The latter equation was then solved by iteration to achieve a convergent solution. Figure 9.8 shows one such solution for particular value of the Yukawa parameter λ, temperature and driving force. In this figure, the CNT radius is shown as well as the density of the liquid at coexistence and also at the μ and T of the grand canonical system. Note that the diffuse interface extends over 2/3 of the droplet and that the density at the center is not that of the stable phase at coexistence. Despite the diffuseness of the interface and because the conditions corresponding to the figure are removed from the spinodal it is likely that the apparent difference between CNT and DFT values of the work of nucleation for these conditions corresponds to the $4\pi k_s$ term due to the curvature dependence of the interface energy.

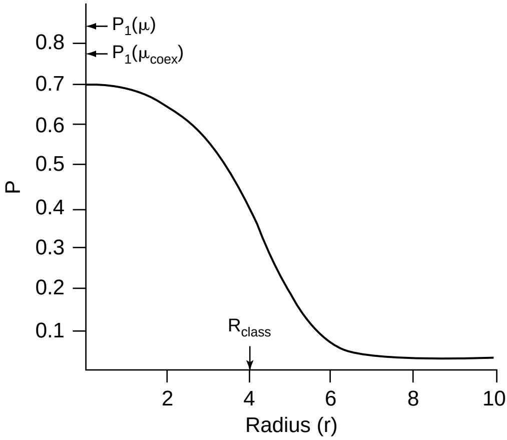


Figure 9.8. Predicted density versus distance from center of nucleus. From J. Chem. Phys. 89, 7521(1988) with permission. © 1988 American Institute of Physics.

In more complex situations where the density is not the only order parameter the fundamental variational principle in DFT that the functional derivative of the grand potential with respect to the order parameters is zero at equilibrium still holds. As mentioned above the result is a non-linear integrodifferential equation that can be solved for the order parameters as a function of position relative to the center of the nucleus. The reader is again advised to read the literature to learn the details of the several procedures used to obtain numerical solutions.

For example, in the case of a crystal the density itself may be written as a Fourier series

$$\rho(r) = \rho_o[1 + m_o + \Sigma m_i \exp(ik_i r)]$$

where the sum is over the lattice reciprocal vectors k_i of the crystal, m_i are the order parameters of the liquid to crystal transition. The parameter m_o gives the density change on solidification, while the higher order parameter m_i determines its crystallinity. The higher order parameters are constrained by the relation

$$\ln m_i = (k_i/k_1)^2 \ln m_1$$

The density and order parameters are considered variables. Equilibrium states are found by minimizing a free energy functional. To study nucleation a canonical system is used in which the total number of particles is constant and maintained within a spherical box. The free energy functional may be of the Landau–Ginzburg type, i.e. $\Omega(m, \rho) = d\mathbf{r}[\omega(m(\rho)) + K|\nabla m(\rho)|^2]$, where $\omega = f(m(\rho)) - \mu\rho$ is the local grand potential. To obtain an approximation to the free energy functional an interatomic potential may be assumed or the local part may be expanded as a power series and the coefficients fitted to experiment. Contrary to the dependence of the order parameters in CNT where inside the nucleus their values are uniform and their variation is limited to a sharp interface in the DFT result these order parameters may vary.

One new feature brought to light by these studies is that the variation of the order parameters across the liquid–crystal interface varies to different extent as a function of distance. The order parameter measuring crystallinity increases at a greater rate than that of the density in moving from the liquid to the crystal across the interface. This dependence is revealed in Figure 9.9 where the crystallinity parameter first begins to increase at constant density in the liquid as the interface is traversed from liquid to solid (left to right). The interface is much thicker than is indicated by the density profile. As might be expected, the diffuseness of the interface increases the closer the conditions are to the spinodal conditions.

DFT has also been useful in determining the distribution of components in a multicomponent system in a spherical nucleus. Figure 9.10 reveals a result of DFT for the distribution of the three components of a ternary system in the nucleus. It is apparent that one of the components preferred to localize in the

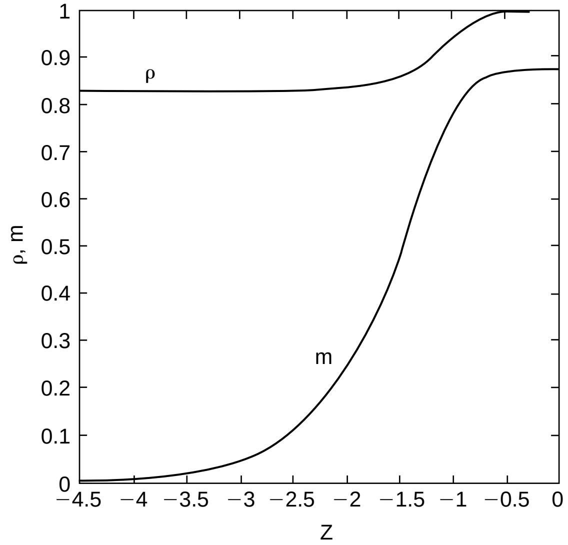


Figure 9.9. Density functional theory solution for the dependence of the order parameter m, measuring degree of crystallinity, and density ρ in a traverse from liquid to crystal from left to right. From J. Chem. Phys. 104, 4233(1996) with permission. © 1966 American Institute of Physics.

region of the interface closest to the vapor phase. CNT would not reveal such interface segregation since the composition of the nucleus would be determined by the phase boundary composition corresponding to the composition of the system and the temperature.

An important contribution governing the composition of the nucleus was made by Oxtoby and Kashchiev,[20] now known as the nucleation theorem. According to this theorem

$$(\partial \Delta\Omega^*/\partial \mu_{o.i})_T = \Delta n_i^*$$

where $\mu_{o.i}$ is the chemical potential of component i in the host phase and Δn_i^* is the number of molecules of i in the critical nucleus. The latter quantity can be obtained from measurements of the dependence of the rate of nucleation on the partial pressure of a component.

Nucleation theory is a living discipline. Contributions continue to be made to it. For example, dynamic nucleation theory treats nucleation via variational transition state theory.[21] It defines the nucleus in the gas–liquid transition as a collection of N molecules in a volume V that minimizes the evaporation rate from the liquid cluster within the volume.

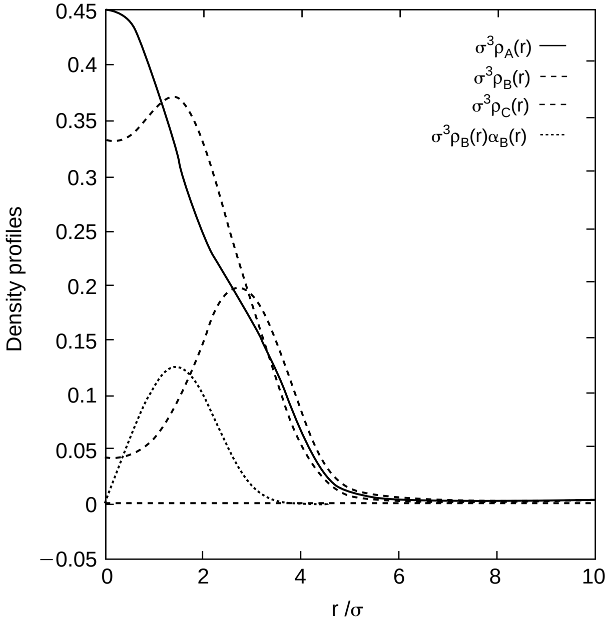


Figure 9.10. Density profiles for the components of a ternary in the nucleus. Solid line represents the water-like component, the long dashed line butanol, and the short dashed line nonane. The dotted line represents the orientation of the hydroxyl group in the butanol, positive values pointing towards the water. From Accounts of Chemical Res. 31, 97(1998) with permission. © 1998 The American Chemical Society.

4. Growth from nucleation onwards

Computers have made it possible to solve coupled non-linear equations and to simulate processes, both of which capabilities have been exploited to provide us with more complete descriptions of the possible events that can occur in the process of decomposition of a metastable parent phase into the more stable products characteristic of the equilibrium phase diagram and also of the resulting microstructure. Historically, the first descriptions of the growth process coupled to the nucleation process, were due to Kolmogorov,[22] Johnson–Mehl[23] and Avrami[24] (KJAM) and provided an exactly solvable model of a first-order phase transition and yielded an equation for the volume fraction transformed as a function of time, the number of grains (equal to the number of nuclei), the unimpinged and impinged surface area, the length of triple junctions and the number of point junctions of four grains in the product structure.

In an abridged version these models can be described by the following equations. The extended volume produced at time t by the unrestricted growth of nuclei nucleated at time τ in the absence of impingement is given by

$$V^*(t) = (4\pi V/3)\int \lambda^3 J^*(t - \tau)^3 d\tau$$

where λ is the growth rate (dr/dt), J* is the steady-state nucleation rate, V is the total volume of the parent metastable phase. Now, the essential assumption made to account for impingement and the consequent cessation of growth is that the change of the extended volume is proportional to the total volume, while the change of the actual transformed volume is proportional to the untransformed volume and that the constants of proportionality in these relations are equal to each other. Hence

$$dV^*/V = dV^t/(V - V^t)$$

or

$$1 - (V^t/V) = \exp(-V^*/V)$$

is the assumed relation, where V^t is the transformed volume. From these relations and with a further assumption that $\tau << t$ then

$$\ln[1/(1 - y)] = (k^*t)^4$$

where $y = V^t/V$. This relation yields a sigmoidal curve for y as a function of t as shown in Figure 9.11. Now, it is obvious that this relation is not applicable to nucleation following spinodal decomposition, nor is it applicable to situations where the growth rate is not a constant (e.g. time-dependent supersaturation), nor is it applicable to the case of time-dependent nucleation, although the latter dependencies can be factored into the integral. Thus, there have been many investigations that attempt to provide a more accurate relation for the fraction transformed, the number of precipitates, etc., as a function of time. It should be noted in Figure 9.11 that curves for values of the exponent of t other than 4 are included. The reason for this is that the KJAM relation for the volume fraction transformed as a function of time has been used in the literature with the exponent of time determined empirically to describe experimental results for various transformations. Table 9.1 provides a summary of the exponent values obtained in various kinetic processes. As indicated, many processes deviate from the behavior predicted in the KJAM treatment of the nucleation and growth problem.

Among the first of the follow-up treatments of the growth problem was the mean-field approach of Langer and Schwartz.[25] Their treatment allowed the precipitate growth rate to depend upon the supersaturation and the precipitate size and was consistent with the basic equation used by Lifshitz and Slyozov[26] to study

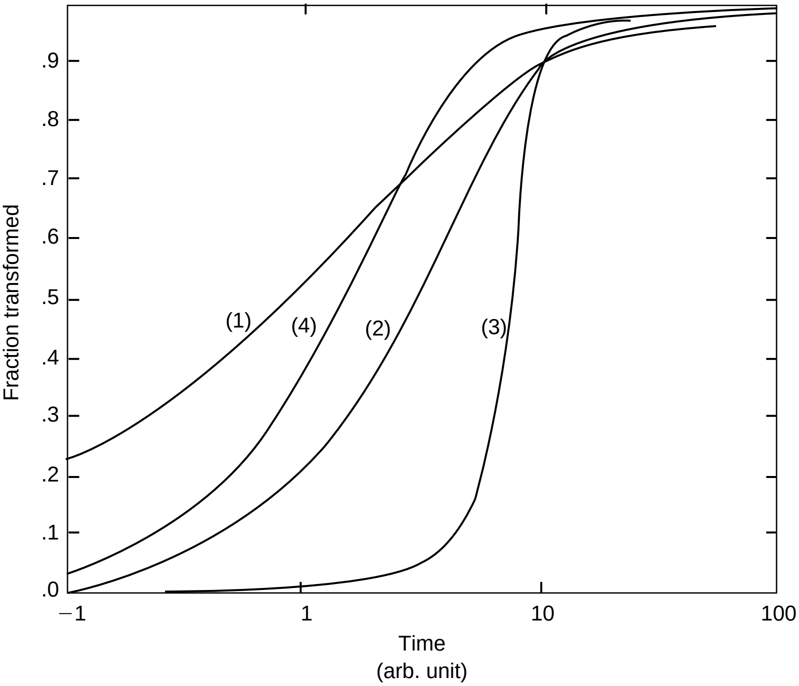


Figure 9.11. Reaction curves corresponding to Johnson–Mehl–Avrami equation. Curves (1), (2) and (3) have the same value of k and n = 0.5, 1 and 4, respectively. Curve (4) has n = 1 and k half the value of the other curves. After J. Burke, The Kinetics of Phase Transformations in Metals, Pergamon Press, Oxford, 1965.

coarsening. Theories involving more mathematically sophisticated concepts have since been developed to accomplish the same and additional objectives. Among these are the time-cone methods of Ohta et al.[27] and Cahn,[28] the survival probability method of Yu et al.[29] the phase field method,[30] various mean-field developments[31] and Monte Carlo based procedures,[32] and the time-dependent Ginzburg–Landau model.[33] Many of these models allow the prediction of microstructure and are well worth study by those intending to specialize in this field. In this writer's opinion they represent a significant advance in the ability to solve problems in kinetics involving condensed phases over the capability existing prior to about 1990. These mathematical techniques do not add new physical concepts. The latter are incorporated into the equations governing the kinetics. Some of these physical concepts will be discussed in the following chapters.

We should note that Cahn used the time-cone method to prove that the KJAM theory is an exact statistical solution for the expected fraction transformed

Table 9.1. Values of the exponent of time in the KJAM equation for the volume fraction transformed.

Conditions	**n**
Polymorphic changes, discontinuous precipitation, eutectoid reactions, interface controlled growth, etc.	
Increasing nucleation rate	>4
Constant nucleation rate	4
Decreasing nucleation rate	3–4
Zero nucleation rate (saturation of point sites)	3
Grain edge nucleation after saturation	2
Grain boundary nucleation after saturation	1
Diffusion controlled growth	
All shapes growing from small dimensions, increasing nucleation rate	$>2\frac{1}{2}$
All shapes growing from small dimensions, constant nucleation rate	$2\frac{1}{2}$
All shapes growing from small dimensions, decreasing nucleation rate	$1\frac{1}{2}$–$2\frac{1}{2}$
All shapes growing from small dimensions, zero nucleation rate	$1\frac{1}{2}$
Growth of particles of appreciable initial volume	1–$1\frac{1}{2}$
Needles and plates of finite long dimensions, small in comparison with their separation	1
Thickening of long cylinders (needles) (e.g. after complete end impingement)	1
Thickening of very large plates (e.g. after complete edge impingement)	$\frac{1}{2}$
Precipitation on dislocations (very early stages)	$\sim\frac{1}{2}$

in a nucleation and growth reaction in an infinite specimen, when nucleation is random in the untransformed volume and the radial growth rate after nucleation is constant until impingement. This result contradicts the contrary implication of Yu et al.[34] in this regard.

5. Summary

We have considered the steady-state rate of heterophase nucleation in a homogeneous medium and found that it is proportional to a product of the equilibrium number of nuclei (critically sized embryos) and the nucleus accretion rate for molecules. A consideration of the rates of growth and shrinkage of embryos, a molecule at a time leads to a system of coupled differential equations. An approximate solution to these equations yields an expression for the proportionality constant between the steady-state rate of nucleation and the product of the equilibrium number of nuclei and the nucleus accretion rate. This solution also gives an expression for the time dependence of the nucleation rate. Experiment confirms this classical model of nucleation.

We have also considered heterogeneous (catalyzed) nucleation and emphasized the need to take into account the anisotropy of the interface energy for many cases of nucleation in the solid state. In addition, we considered the rate of spinodal decomposition and found that this rate is maximized at a wavelength equal to $\sqrt{2}$ times the critical wave length in the linear theory. Experiment verifies the linear theory in its range of application, but not the non-linear models for the later stage of spinodal decomposition.

CNT was found to neglect an important effect – that of the dependence of the interface energy on curvature. When this effect is taken into account the CNT derived work of nucleation is in agreement with experiment for conditions far from the spinodal. DFT describes the nucleation problem well at and near the spinodal for conditions between the binodal and spinodal. The nucleation theorem relates the number of molecules i in the nucleus to the variation of the work of nucleation with respect to the chemical potential of species i.

References

1. R. Becker and W. Doring, Ann. Physik. 24, 719(1935).
2. J.B. Zeldovich, Acta Physicochim URSS 18, 1(1943).
3. K.S. Chan, J.K. Lee, G.F. Shiflet, K.C. Russell and H. Aaronson, Met. Trans. AIME 9A, 1016(1978).
4. S. Auer and D. Frenkel, Nature 409, 1020(2001).
5. S. Auer and D. Frenkel, J. Chem. Phys. 120, 3015(2004).
6. L. Granasy and T. Pusztai, J. Chem. Phys. 117, 10121(2002)
7. V. Talanquer and D.W. Oxtoby, J. Chem. Phys. 109, 223(1998).
8. W. Ostwald, Z. Phys. Chem., Stoechiom. Verwandtschaftsl. 22, 289(1897).
9. A.H. Ramsden and P.J.F. James, J. Mat. Sci. 19, 2844(1984).
10. P.G. Shewmon, **Diffusion in Solids**, McGraw-Hill, NY, 1963.
11. J. Cahn and J.E. Hilliard, J. Chem. Phys. 28, 258(1958); 31, 688(1959); Acta Met. 9, 795(1961); 10, 179(1962).
12. C. Tegeler, R. Span and W. Wagner, J. Phys. Chem. 28, 779(1999).
13. D.W. Oxtoby, Ann. Rev. Mat. Res. 32, 39(2002).
14. G.K. Schenter, S.M. Kathmann and B.C. Garrett, Phys. Rev. Lett. 82, 3484(1999).
15. R. McGraw and A. Laaksonen, Phys. Rev. Lett. 76, 2754(1996).
16. D. Reguera et al., J. Chem. Phys. 118, 340(2003).
17. D. Reguera and H. Reiss, Phys. Rev. Lett. 93, 165701(2004).
18. J.W. Cahn and J.E. Hilliard, J. Chem. Phys. 28, 258(1958).
19. N.F. Carnahan and K.E Starling, J. Chem. Phys. 51, 635(1969).
20. D.W. Oxtoby and D. Kashchiev, J. Chem. Phys. 100, 8665(1994).
21. G.K. Schenter, S.M. Kathmann and B.C. Garret, J. Chem. Phys. 110, 7951(1999).
22. A.N. Kolmogorov, Izv. Akad. Nauk SSSR, Ser. Matem. 3, 355(1937).

23. W. Johnson and R.F. Mehl, Trans. AIME 135, 416(1939).
24. M. Avrami, J. Chem. Phys. 7, 1103(1939); 8, 212(1940); 9, 177(1941).
25. J.S. Langer and A.J. Schwartz, Phys. Rev. A2, 948(1980).
26. I.M. Lifshitz and V.V. Slyozov, J. Phys. Chem. Solid. 19, 35(1961).
27. S. Ohta, T. Ohta and K. Kawasaki, Physica 140A, 478(1987).
28. J. Cahn, MRS Symp. Proc. 398, 425(1996).
29. G. Yu and J.K.L. Lai, J. Appl. Phys. 78, 5865(1995); 79, 3504(1996); G. Yu, S.T. Lee, J.K.L. Lai and L. Ngai, J. Appl. Phys. 81, 89(1997); G. Yu, J.K.L. Lai and W. Zhang, J. Appl. Phys. 82, 4270(1997).
30. R. Kobayashi, Physica D 63, 410(1993); W.J. Boettinger, A.A. Wheeler, B.T. Murray and G.B. McFadden, Mater. Sci. Eng. A 178, 217(1994); J.A. Warren and W.J. Boettinger, Acta Metall. Mater. 43, 689(1995); H.-J. Jou and M.T. Lusk, Phys. Rev. 55, 8114(1997); J.P. Simmons, C. Shen and Y. Wang, Scripta Mater. 43, 935(2000).
31. A.G. Khachaturyan, **Theory of Structural Transitions in Solids**, John Wiley, New York, 1983; F. Ducastelle, **Order and Phase Stability in Alloys**, North-Holland, Amsterdam, 1991; G. Martin, Phys. Rev. B 41, 2279(1990); K.D. Belashchenko and V.G. Vaks, J. Phys. Condens. Matter 10, 1965(1998); V. Yu. Dobretsov et al., Europhys. Lett. 31, 417(1995); Phys. Rev. B 54, 3227(1996-I); J. Phys., Condens. Matter 10, 2275(1998).
32. G. Martin, M. Athenes, C. Desgranges, M. Nastar and F. Soisson, MRS Symp. Proc. 527, 35(1998); B.D. Wirth and G.R. Odette, MRS Symp. Proc. 481, 151(1998).
33. K. Binder, Colloid Polymer Sci. 265, 273(1987); S. Nambu and A. Sato, J. Am. Ceram. Soc. 76, 1978(1993); Y. Wang, H. Wang, L.-Q. Chen and A. Khachaturyan, ibid 78, 657(1995); J.D. Gunton, M. San Miguel and P.S. Sahni in **Phase Transitions and Critical Phenomena**, Vol. 8, eds. C. Domb and J.L. Lebowitz, 8, Academic Press, London, 1983.
34. G. Yu, S.T. Lee and J.K.L. Lai, MRS Symp. Proc. 481, 97(1998).

Bibliography

Summaries of linear spinodal decomposition stage

1. J.W. Cahn, Trans. AIME 242, 166(1968).
2. J.E. Hilliard, in **Phase Transformations**, ASM, Metals Park, Ohio, 1970, p. 497.

Later spinodal decomposition stage

1. H.E. Cook, Acta Met. 18, 297(1970).
2. J.S. Langer in **Fluctuations, Instabilities and Phase Transitions**, ed. T. Riste, Plenum Press, New York, 1975, p. 19.
3. B. Ditchek and L.H. Schwartz, Acta Met. 28, 807(1980).
4. R.D. Doherty in **Physical Metallurgy**, 3rd edition, eds. R.W. Cahn and P. Haasen, North-Holland, New York, 1983, p. 934.

Nucleation

D. Kashchiev, **Nucleation: Basic Theory with Applications**, Butterworth-Heinemann, Oxford, 2000 UK.

Problems

1. Why does the nucleation rate exhibit a maximum as the temperature decreases below the temperature corresponding to equilibrium between the parent and product phases?
2. Why does the rate of spinodal decomposition exhibit a maximum as the wavelength of the composition wave increases from the critical value corresponding to the generalized spinodal condition?
3. In the technological process of strengthening an alloy by precipitation in a supersaturated matrix phase, the procedure of quenching the matrix phase from the solutioning temperature to a temperature T_1 and holding for a limited time prior to raising the temperature to T_2 and completing the precipitation process is used. Can you provide a theoretical justification for this procedure? *Hint*: The strength of the alloy increases with the population density of the precipitates.
4. If the condensed product phase were to nucleate from a vapor parent phase would the rate of nucleation exhibit a maximum as a function of temperature? Explain your answer.
5. If you had a measurement of the time for complete spinodal decomposition at one temperature T_1, a knowledge of the average wavelengths for the decomposed product at two temperatures T_1 and T_2, and a knowledge of the activation energy for diffusion, how would you estimate the time for complete spinodal decomposition at the temperature T_2?
6. Design an experiment to yield the interface energy between a product and parent phase from measurements of nucleation rates. Discuss the problems associated with this method.
7. Assume homogeneous nucleation of crystal solids from liquids of the same composition. We compare two cases: nickel versus SiO_2. Assume crystal–melt interface energies of 25 erg/cm^2 in both cases and that the driving force for the phase change is given by $\Delta g = \Delta s(T_M - T)$. For simplicity, we will take $\Delta s = 2$ cal\mol\°C and $T_M = 1900$°K for both materials. (Use a handbook to obtain the required specific volumes.) The frequency ν for liquid metals deduced from their viscosity is about 10^9 T/second. For glasses, including silica, this jump frequency is activation energy dependent with an average activation energy about 14,000 cal/mol over the temperature range of interest, with ν_0 about 10^{12}/second. Calculate the number of crystal nuclei produced in 1 cm^3 of each material when the corresponding liquid is quenched from just above 1900°K to 300°K at the rate of 106°C/second. *Hint*: Use a computer to solve the integral. Assume a nucleus grows with a velocity equal to the product of a lattice parameter, v and $\Delta g/kT$, where the latter is dimensionless. Calculate the size of the largest crystal particle produced after the quench in each material.
8. In the spinodal decomposition of a supersaturated solution the composition–distance function normal to some crystal plane obeys the relation $C = \delta(t)\sin \beta x$, where $\delta(t)$ is a time-dependent amplitude that increases with time and the other variables are defined in the text. If you plot this function for two successive values of the time of annealing you will note that solute is being transported from a low concentration to a region of higher concentration. Provide the basis for this “uphill diffusion” where solute moves up a concentration gradient.

CHAPTER X

Solid–Solid Interface Migration Kinetics

Introduction

Primary recrystallization, secondary recrystallization, grain growth and A/α (crystalline/amorphous) solid phase epitaxy are some of the phenomena that involve the motion of an interface between solid phases that have the same composition. We shall consider each of these phenomena from a common viewpoint. However, the mechanism of each phenomenon differs, and needs to be considered separately.

1. Driving forces for recrystallization and grain growth

We consider two classes of rate processes in this chapter. One involves the migration of an interface between crystals of the same phase which, except for the excess energy associated with defects, are stable. The other involves migration of an interface between a crystal of a material and the amorphous phase of the same composition. Both, in phenomenological terms, can be described by the same relation in which the interface velocity is proportional to the product of the driving force and a mobility. However, as will be noted later in the section on mechanisms, there is little to be gained by considering the second process, known as A/α solid phase epitaxy, from this viewpoint, since the driving force does not appear to be a parameter that can be varied and its effect tested, at least at this writing.

The first class comprises the processes that go under the names of recrystallization and grain growth. It is worthwhile to discuss recrystallization and grain growth kinetics from the separate viewpoints of the driving force and mobility. Because of the generally low values of the driving force relative to the thermal energy, kT, in these phenomena, it will usually be possible to express the rate of any particular one of these processes in terms of a product of a driving force and a mobility. Let us consider the various driving forces first.

For the sake of illustration of the various phenomena associated with the motion of an interface between identical phases, but of possibly different energy, consider the spherical monocrystal shown in Figure 10.1, that is surrounded by deformed crystals of the same phase in Figure 10.1(a), by smaller, but undeformed

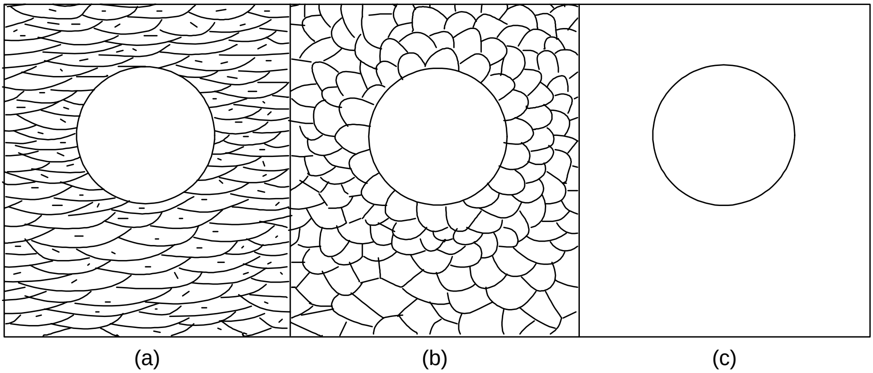


Figure 10.1. Schematic illustration of the initial conditions at the start of (a) primary recrystallization, (b) secondary recrystallization and (c) grain growth. In (a) the matrix is deformed. The host of (b) is a fine-grained polycrystal in which the boundaries are pinned and cannot migrate. The diagram in (c) makes the point that during grain growth the driving force for boundary migration is reduction in boundary area.

crystals of the same phase in Figure 10.1(b), and by an undeformed crystal of the same phase in Figure 10.1(c). Let the free energy per unit volume in the spherical monocrystal and in the crystals of the surrounding grains in Figures 10.1(b) and (c) be g_o, whereas that in the surrounding area of Figure 10.1(a) is g_e. If g_e is the average free energy per unit volume including the energy of grain boundaries in the surrounding volume, then the driving force for the migration of the boundary of the spherical monocrystal into the deformed region of Figure 10.1(a) is $g_e - g_o - 2\sigma/R$. This is the driving force for the stage of primary recrystallization. In general, $g_e - g_o$ is much larger than $2\sigma/R$ and is time and temperature dependent. It is made up of the excess energy due to dislocations, grain boundaries and point defects and clusters of same. During the stage of recovery, point defects and dislocations can move so as to either decrease their densities or decrease the total free energy without annihilation. Sufficient energy remains after recovery of sufficiently cold-worked samples as to drive their primary recrystallization into product crystals that have the ground state free energy per unit volume g_o. In this process, the excess free energy due to the excess defects acts to move the interfaces between deformed and undeformed crystals so as to annihilate the former and produce more of the latter.

In the stage corresponding to grain growth, Figure 10.1(c), if the radius of the spherical monocrystal increases by dR, the energy associated with the grain boundary increases by $d(4\pi R^2)\sigma = 8\pi R\sigma dR$, which when divided by the volume of the atoms transported across the boundary in its motion by dR, $(4\pi R^2 dR)$, yields $2\sigma/R$ as the free energy increase per unit volume of "transformed" material. Hence, the driving force is $-2\sigma/R$ since the latter is the free energy of the material before

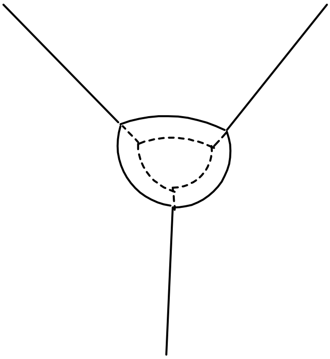


Figure 10.2. Schematic showing decrease in grain boundary area as boundaries move toward their center of curvature.

"transformation" minus that after "transformation". The same conclusion can be made by using a result derived in Chapter IV, which gives the difference in chemical potential between material in the spherical monocrystal at the interface and material in its surroundings, but at the interface, directly as 2σ/R, with that in the monocrystal being higher than in the surroundings. Thus, the driving force for grain growth is such as to tend to move the grain boundary toward its center of curvature, i.e. to decrease the diameter of the spherical monocrystal. There seems to be a contradiction between the term "grain growth" and the direction in which the grain boundary moves so as to achieve that grain growth. This contradiction can be resolved on considering that in grain growth it is the average grain size that increases, but this average grain size increase is accomplished by the decrease in grain boundary area per unit volume. This decrease in grain boundary area per unit volume corresponds in detail to the motion of individual grain boundary segments toward their centers of curvature. This process is schematically illustrated in Figure 10.2.

The remaining situation of Figure 10.1(b) corresponds to the process called secondary recrystallization or exaggerated grain growth or abnormal or discontinuous grain growth. In this case, the driving force is the excess energy associated with the grain boundaries of the smaller grains that surround the spherical monocrystal. If the average diameter of these grains is D*, then the grain boundary area per unit volume associated with these grains is C/D*, where C is a constant whose value depends upon the shape of the grains. For cubic-shaped grains of the same size $C = 3$; for grains that are equal tetra-kadecahedra, the grain boundary area equals 3.35/L, where L is the separation of square faces. Thus, the driving force for secondary recrystallization is $C\sigma^*/D^* - 2\sigma/R$. Secondary recrystallization will occur when $D^* < C\sigma^*R/2\sigma$. This condition has to be satisfied during the entire stage of secondary recrystallization. If the average grain size of the surrounding grains can increase with time it becomes necessary for R to increase at a greater rate. Because this condition is impossible to satisfy for a grain assembly with the distribution corresponding to normal grain growth,[1] secondary recrystallization has only been found when the surrounding grain boundaries are prevented from moving. This restraint on boundary motion can be achieved via the use of particles or via preferred orientation for the assembly of grains. Thin films with grain sizes corresponding to the limit conditioned by the thickness of the film also exhibit such restraint to boundary motion. Let us for the moment accept that these factors can allow us to maintain a constant average grain size in the region surrounding the spherical monocrystal. In the next section we will consider the growth laws for the three stages of growth described above.

In thin films, where the grains extend completely from one surface to the other, there is an additional driving force that can act to cause secondary recrystallization. This driving force is the difference in surface energy between the differently oriented grains. When the surface energy is highly anisotropic then a significant driving force can exist to grow those grains having surfaces with the lowest surface energy at the expense of the other differently oriented grains. The process of producing cube texture oriented Fe–Si transformer sheet takes advantage of this effect.

Most of the data that have been obtained in an attempt to elucidate the phenomenological relations governing recrystallization and grain growth have come from studies on metals. However, observations of these phenomena in ceramics and semiconductors have been found to obey the same relations.

A listing of the magnitudes of the driving forces for some of the various modes of boundary migration follows:

A/α solid phase epitaxy	10^9 N/m^2 ($\approx$2500 cal/mole)
Primary recrystallization	10^8 N/m^2 ($\approx$250 cal/mole)
Grain growth	10^4 N/m^2 ($\approx$0.025 cal/mole)
Secondary recrystallization	10^3 N/m^2 ($\approx$0.0025 cal/mole)
Electromigration	10^2 N/m^2 ($\approx$0.00025 cal/mole)

Note:According to this list $\Delta G/RT \ll 1$ only for the ultimate three processes!

2. Growth laws for pure materials

2.1. Primary recrystallization

On the assumption that the growth rate dR/dt is given by

$$dR/dt = M(g_e - g_o) = M\Delta g$$

and that the driving force changes with time during the recovery process according to simple second-order kinetics, i.e.

$$\Delta g^{-1} - \Delta g_o^{-1} = k_r t$$

where Δg_o is the driving force corresponding to the start of isothermal annealing at $t = 0$, k_r is a rate constant for the recovery process and t is the time at the annealing

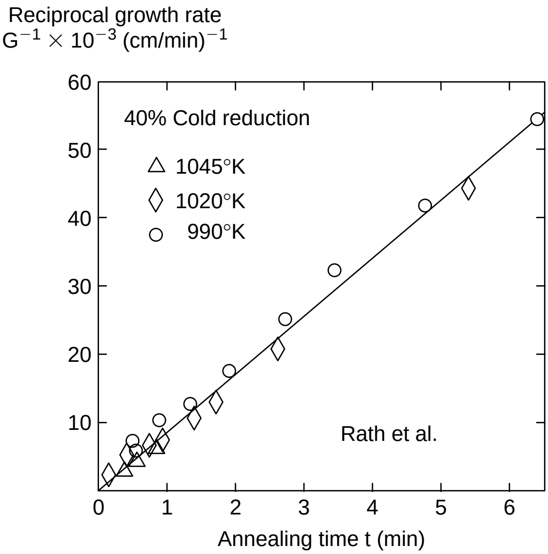


Figure 10.3a. Reciprocal growth rate during primary recrystallization as a function of time for beta titanium. From Rath et al., Met. Trans. 10A, 1013(1979) with permission.

temperature, Li[2] has shown that the growth for the primary recrystallized grains obeys

$$(dR/dt)^{-1} = M^{-1}(\Delta g_o^{-1} + k_r t)$$

Figure 10.3(a) illustrates a case where this law is obeyed experimentally. In this case, RT ⩾ 2000 cal/mole and the use of a proportionality between growth rate and driving force is barely justified. Incidentally, the fact that the points corresponding to different annealing temperatures are superimposed suggests that the activation energies for the recovery process and the growth of the recrystallized grains are nearly equal in this case.

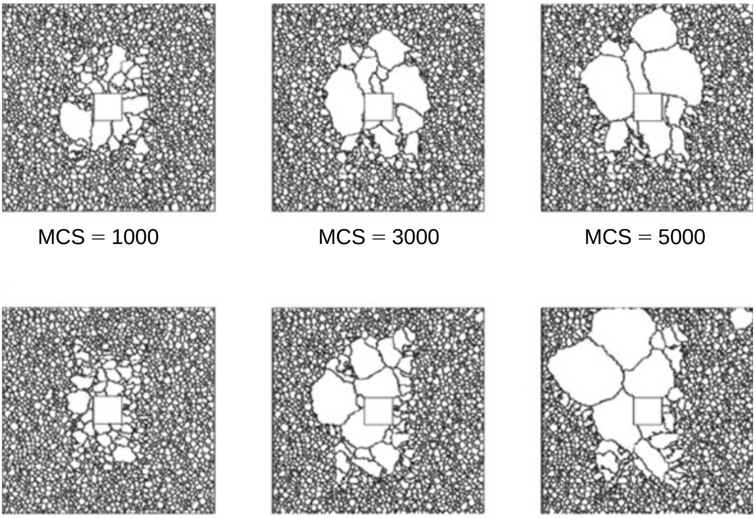
The phenomenological law of recrystallization is given by the following relation

$$X_v = 1 - \exp(-Bt^k)$$

where X_v is the volume fraction recrystallized, t is time and B and k are constants. Deviations from this law were believed to be due to the simultaneous effect of recovery on the rate of growth, as discussed in the previous paragraph. However, later experiments and computer simulations lead to the conclusion that these deviations are due to a non-homogeneous distribution of the stored excess energy produced by the deformation.[3]

One of the questions we need to address is "How can computer simulation procedures help improve our knowledge or control of the primary recrystallization process?" From the results to date of computer simulations of primary recrystallization the answer seems to be – not much. Nothing new has been added by these simulations to our knowledge of primary recrystallization. The main effort in these studies has been to adjust the simulation procedures so that they better mimic experiment. This comment applies to phase field, Monte Carlo, and cellular automata schemes that have been investigated.[4] One advantage that computer simulation has demonstrated is the ability to provide data for complex experimental situations that involve less effort to obtain in the simulation mode than in the experimental mode.[5]

Computer simulation can also easily provide detailed views of physical phenomena that are difficult (but, not impossible) to obtain experimentally. Figure 10.3(b) illustrates this point. The start of primary recrystallization about a particle contained in a deformed matrix, with the stored deformation localized



MCS = 1000 MCS = 3000 MCS = 5000

MCS = 500 MCS = 2500 MCS = 6500

Figure 10.3b. Recrystallization simulation in a deformed matrix containing a particle about which the plastic strain is higher than in the matrix itself. Bottom figure shows the results of 33% higher initial strain compared to the top figure. MCS represents Monte Carlo steps. Reproduced with permission from B. Radhakrishnan, G. Sarma, H. Weiland and P. Baggethun, Modell. Simul. Mater. Sci. Eng. 8, 737(2000). © 2000 US Government.

mainly about the particle, is illustrated in these figures. The two figures correspond to two deformation strains, one (bottom) 33% higher than the other (top). The figures show the recrystallized grains about the square particle. An isolated recrystallized and growing grain surrounded only by high-angle boundaries exists only for the more deformed case. The simulations also emphasize an already known result that it is the grains with an orientation, relative to the average texture, that promotes high boundary mobility that grow at the expense of the slower growing grains.

The mechanism of nucleation of the primary recrystallized grains has been explored experimentally with the conclusion that nucleation occurs by subgrain coalescence.[6] Contrariwise, a Monte Carlo simulation study[7] of the abnormal growth of subgrains yielded the result that certain subgrains produced in the plastic straining of crystalline materials often have the characteristics needed to grow abnormally and act as recrystallization nuclei. These subgrains are at the outer extremity of the subgrain orientation distribution. The controlling characteristic is that the boundaries of abnormal growing subgrains that form recrystallized grains have mobilities much higher than those of the surrounding subgrains prior to impingement with other abnormal growing subgrains. There must also be such subgrains in the original strained matrix that are larger than critical nucleus size for growth to occur. Still another simulation study based on vertex dynamics[8] reveals that nucleation appears to occur preferentially at subgrains adjacent to grain boundaries or grain boundary vertices by motion of the grain boundaries or vertices, i.e. grain boundary bulging. This mechanism was also among those investigated previously.[6] The mechanism of supercritical subgrain growth seems the most reasonable as an explanation for primary recrystallization in as much as supercritical subgrains have been found experimentally[9] and capable of growing abnormally in computer simulations and because their orientation differ markedly from the deformation texture corresponding to the majority of the grains and subgrains in agreement with experiment.

2.2. Secondary recrystallization or abnormal grain growth

There is an incubation period for the onset of secondary recrystallization. After this incubation period the growth rate is independent of time. This is just the behavior expected from the dependence of the growth rate on the driving force discussed above, i.e.

$$dR/dt = M_b[C\sigma^*/2R^* - 2\sigma/R]$$

According to this relation, the incubation period corresponds to the case where R is sufficiently small that the two terms in the bracket are similar in magnitude. When R has grown to the point that the second term in the bracket is small compared to the first term and the first term is constant, then the successive growth rate will be independent of time. Although this relation provides a qualitative explanation of the observations, it has not yet been proven that it is in quantitative agreement with experiment. The above relation assumes stationary grains about the growing grain and the geometry of Figure 10.1(b) where there is no local equilibrium at the grain boundary vertices with the growing grain. This is not necessarily always the experimental state. Rollett and Mullins[10] have refined this relation to bring it into better agreement with the likely experimental state which includes the possibility of mobile boundaries about the abnormal grain.

The above formulation applies only when the driving force for the abnormal grain growth stems from the grain boundaries alone. In a thin film, the driving force for secondary recrystallization is due to a difference in surface energy between the abnormal grain and the normal grains. In this case, the abnormal grain boundary velocity becomes

$$v = M_b(2\Delta\gamma/h + C\sigma^*/2R^* - 2\sigma/R)$$

where $\Delta\gamma$ is the difference in surface energy between abnormal and normal grains, σ^* and R^* are the energy and radius of the normal grain boundaries, respectively, and σ and R are the energy and radius of the grain boundary between abnormal and normal grains, respectively.

For secondary recrystallization to be possible the normal grains must be prevented from moving or they must move at a much slower velocity than the boundaries of the abnormal growing grain. In bulk materials the presence of particles that act to pin the migration of most of the grain boundaries (Zener pinning) and texture that produces a majority of grains having low mobility favor abnormal grain growth. In addition in thin films thermal grooving at the intersection of grain boundaries with the surface of the thin film can act to pin grain boundary migration and thereby favor abnormal grain growth. Conditions can be developed to produce grain or subgrain boundaries having much higher mobility than the average and the consequent potential for abnormal grain or subgrain growth of grains or subgrains possessing such boundaries.[11]

2.3. Grain growth

A law of grain growth can be derived from the relation that

$$dR/dt = M_g[-2\sigma/R]$$

as follows. The rate of change in the total grain boundary area per unit volume

$$d[C/D]/dt = \{d[C/D]/d\langle R\rangle\}(d\langle R\rangle/dt)$$

or

$$dD/dt = \{dD/d\langle R\rangle\}d\langle R\rangle/dt$$

where D is the average grain diameter and $\langle R\rangle$ is the average radius of the moving boundaries. Now, it is known that the grain size distribution function is time invariant. This fact implies that the average radius $\langle R\rangle$, of the moving boundaries is proportional to D, i.e. $D \sim \langle R\rangle$. Another giant step is to assume that the velocity corresponding to $d\langle R\rangle/dt$ is the same as that of a grain having the radius of curvature $\langle R\rangle$, but of opposite sign. With these assumptions we arrive at the result

$$dD/dt = (k/D)$$

which integrates to give

$$D^2 - D_o^2 = 2k(t - t_o)$$

This law is obeyed sometimes. The phenomenological law of grain growth is $D = K_1t^n$, where n can vary from 1/3 to 1/2.

There have been many simulations of grain growth under the assumption of isotropic properties for the grain boundary energy and mobility.[12] The relation obtained from all of them

$$D^m - D_o^m = 2k(t - t_o) \quad m = 2$$

is in agreement with the result derived above. However, there are differences between simulation results themselves[12] and between experiment for the grain size distribution function. The latter is lognormal experimentally. Also, there are analytic models for these functions.[13,14]

Computer simulation has clarified another effect on the time dependence of the average grain diameter during grain growth. Anisotropy of grain boundary energies introduces an effect on the time exponent n in the empirical relation $D = Kt^n$. Increasing anisotropy decreases n until it reaches the value 0.25.[15a] Thus, although the law governing the motion of an individual boundary may be governed by curvature, other factors have to be taken into account in describing the dependence of average grain diameter on time for an assembly of boundaries. That anisotropy of the grain boundary energy has the most important effect on the growth exponent and on the topology of the product grain structure is substantiated by another simulation involving both the Potts Monte Carlo model and the phase field model which yielded the same results.[15b] Another important effect on grain growth occurs at small grain sizes, low temperature and near high symmetry misorientations where the mobility of triple junctions is much smaller than that for the grain boundaries. In this situation, the triple junction provides a significant drag on grain growth kinetics.[16a]

With grain growth and the concomitant destruction of grain boundary area the excess volume along these grain boundaries must be accommodated in some manner. One way this accommodation takes place has been found in a molecular dynamics simulation[17] – lattice vacancies are injected into the lattice behind moving grain boundaries. This phenomenon appears to have little effect on grain growth kinetics.

2.4. A/α solid phase epitaxy

Since the driving force in solid phase epitaxy for the motion of the amorphous/crystal (α/A) interface between amorphous and crystalline phases having the same composition is independent of time it is expected that the velocity of the interface will be independent of time, as well. This expectation is in agreement with the observations, but this result is not informative of the mechanism applicable in this phenomenon. Other observations will be more useful in this regard and will be discussed in a later chapter.

3. Effect of dispersed particles on grain boundary migration

Consider a grain boundary, tending to migrate in the negative y direction, which meets a spherical particle of radius r, as is illustrated in Figure 10.4. The maximum drag force resolved in the y direction due to the particle is $\pi r\sigma$. Assume a random distribution of n particles per unit volume. It follows then that a boundary of unit area will intersect a total of 2nr particles (i.e. there are 2nr particle centers in each region r thick about unit area of boundary). But the volume fraction of the particles is $f = 4\pi r^3 n/3$ or $n = 3f/(4\pi r^3)$. Substituting, the number of particles intersecting unit area of grain boundary is then $3f/(2\pi r^2)$. If the boundary is migrating

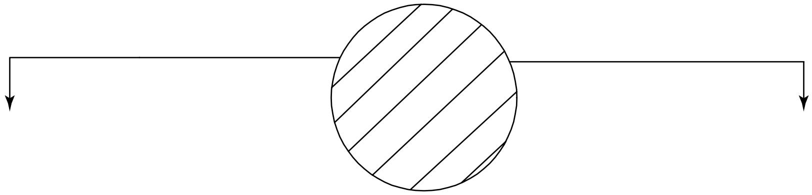


Figure 10.4. Schematic illustration of the initial decrease followed by an increase in grain boundary area as the latter moves past a particle.

under the influence of its own surface tension then a balance of its driving force $2\sigma/R$, where R is the radius of curvature of the boundary, and the total drag force $(\pi r\sigma)(3f/[2\pi r^2])$ yields that

$$R^* = 4r/3f$$

where R* represents the largest radius of curvature of the moving boundary that can traverse the particles, i.e. there is a limiting grain size above which grain growth in a matrix containing particles is not possible. The precise value of R* found above, a result originally due to Zener,[18] has been questioned, because it appears to be in disagreement with experimental values of R*. More recent models[14,19] yield improved agreement, but in view of the new insights that computer simulation have provided in the grain growth problem it would be foolhardy to suggest that we have heard the last word in this matter. Nevertheless, it is likely that Zener's concept, for the origin of a critical radius beyond which normal grain growth ceases, is likely to be valid. Thus, one of the means of maintaining an immobile grain boundary array during secondary recrystallization is the use of inert particles deliberately added to the matrix, such as AlN in steels, or via particles precipitated out of a supersaturated matrix at a temperature above that at which the secondary recrystallization is carried out.

A serious examination of the assumptions in the Monte Carlo Potts model simulation of the interaction between moving grain boundaries and particles revealed that certain artifacts of the model yielded wrong results.[20] Correction for the effects of these artifacts then yielded results validating Zener's model with the exception that the coefficient 4/3 was found to equal 0.73 for monosized, spherical particles.[20] The value of this coefficient was found to depend upon particle shape and size distribution. Incidentally, to obtain these results a large investment in computer time was required (1 month of computer time using 128 processors of an Intel Paragon supercomputer).

4. Grain boundary mobility

Any discussion of grain boundary mobility is faced with the problem, because of the gamut of possible grain boundaries and concomitant grain boundary structures, that there may be a variety of grain boundary mobilities and of their dependences upon variables, such as driving force, solute concentration and the like. One may conceive of two extremes of grain boundary behavior. In one extreme, atoms may be added with equal probability anywhere to the surface of a growing grain across a grain boundary. In the other, atoms can only be added along kinks at steps on the surface of a growing grain across the grain boundary, where it is likely that these boundary steps are associated with grain boundary dislocations. Indeed, both extremes of behavior have been observed as will be noted later. However, there are other possible modes of grain boundary migration. Another that has been recognized is grain boundary migration constrained by the simultaneous glide and climb of grain boundary dislocations. With multiple mechanisms of migration possible for a single grain boundary (i.e. the mechanism may change with the orientation of the boundary for a given misorientation between the bounding grains) it is not difficult to understand that precise rules governing the mobilities of grain boundaries have yet to be developed.

When conditions are such as to bring about a diversity of grain boundary mobilities, then the boundaries with the highest mobility will by their migration tend to produce the largest volume fraction of product. Such product may be either the strain-free grains resulting from primary recrystallization or the abnormally large grains that grow during secondary recrystallization. This phenomenon may be responsible for the textures that are produced by these processes. For example, a very common recrystallization texture in fcc metals and alloys is a rotation about a <111> axis by 30–40° relative to the deformation texture. As shown in Figure 10.5 the boundaries of grains having a rotation of 40° about <111> relative to the deformation texture move at the maximum rate in deformed aluminum and this is just the rotation of the primary recrystallization texture relative to the deformation texture found for aluminum, as shown in Table 10.1. This rotation yields a cube texture. The rapidly moving grain boundaries tend to choke off the growth of the grains with slower moving boundaries and thus those grains having the appropriate relative orientation to the parent matrix to confer maximum mobility to their boundaries act to consume the parent matrix. However, the process is not as simple as just described as made evident in computer simulations. First, we may note that there is some uncertainty concerning the sharpness of the mobility–orientation function shown in Figure 10.5. A molecular dynamics simulation[21] reveals a much sharper function than that in Figure 10.5. An examination of the effect of the mobility peak height relative to the mobility for high-angle grain boundaries away from special orientations

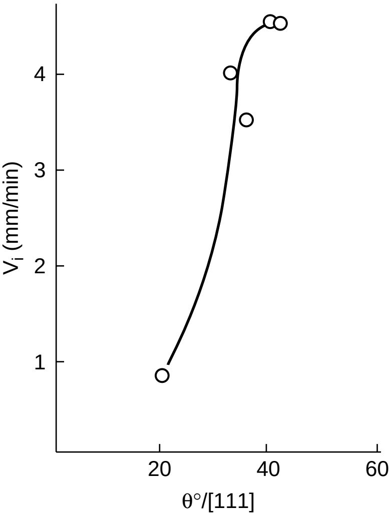


Figure 10.5. Growth rates of new grains having the orientation defined by the abscissa (rotation about [111]) into plastically deformed aluminum crystal at 615°C. The [111] axis is always the same one in the deformed crystal. From Trans. AIME 209, 427(1956) with permission.

Table 10.1. Rotations relating recrystallization texture to deformation texture

Material	**Rotation angle and axis**	**Neighboring coincidence orientation**
Cu	30° [111]	$\Sigma^* = 13$
Al	40° [111]	$\Sigma = 7$ 38.2° [111]
Zn	30° [0001]	$\Sigma = 13$ [0001]
Fe–Si	27° [110]	$\Sigma = 19$ 26 [110]
Al 0.5% Mn	43° [661][a]	?

* Σ^{-1} is the fraction of sites belonging to the two lattices of the crystals that bound the grain boundary that are coincident. $\Sigma = \delta(h^2 + k^2 + l^2)$, where $\delta = 1$ for $(h^2 + k^2 + l^2)$ odd and $\delta = 1/2$ for $(h^2 + k^2 + l^2)$ even.

[a] Disorientation corresponding to 153° [335].

yielded the result that the sharper is this peak the greater is the volume of cube texture yielded upon recrystallization. However, if the grain boundary energy is isotropic then no cube texture is produced by the recrystallization. The literature has tended to contrast oriented nucleation and oriented growth as competing models acting to control primary recrystallization. The results of computer simulations suggest that both are necessary to explain primary recrystallization data.

The physical origin of the high mobility of the grain boundaries corresponding to the Kronberg–Wilson rotation is not known. Originally, significance was placed on the fact that the orientations of the recrystallization textures, such as listed in Table 10.1, are near those for a high degree of coincidence of the atom sites for the two crystal orientations. But there are relative grain orientations that yield higher degrees of coincidence (i.e. a smaller value of Σ than that for the Kronberg–Wilson orientation (7)) as well as lower degrees of coincidence. No one has noted anything special about the $\Sigma = 7<111>$ grain boundaries relative to the other equivalent coincidence boundaries. Further, it appears that a small amount of impurity content in the matrix, but not too much, is necessary for the cube texture to be observed. From the work of Aust and his coworkers, it appears that in very pure materials there are many high-angle grain boundaries that have mobilities equal to those for the Kronberg–Wilson case but that the effect of a small concentration of impurities is to markedly diminish the mobility of the random high-angle boundaries relative to that for the Kronberg–Wilson orientation (see Figure 10.6). Thus, one possible explanation of the fact that the Kronberg–Wilson boundaries have higher mobility than the other boundaries is that dilute solute acts to decrease the mobility of all boundaries by a different mechanism than that applicable to the Kronberg–Wilson case. However, molecular dynamics simulation of grain boundary migration in pure material does exhibit a marked maximum in boundary mobility at $\Sigma = 7$ orientation.[21] Thus, there must be an intrinsic, still unknown, origin for this high mobility.

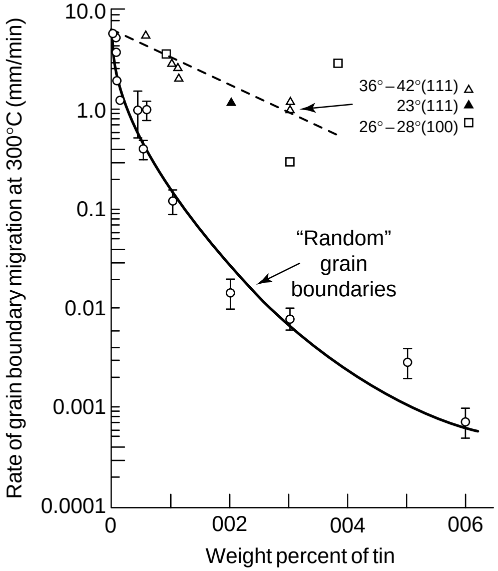


Figure 10.6. Effects of solute concentration on grain boundary migration velocity of “special” and “random” grain boundaries. From Aust and Rutter, Trans. AIME 215 (119), 608(1959) with permission.

The mobility of grain boundaries in silicon exhibits an effect not found in metals. In particular, certain dopants, such as P and As, enhance the grain boundary mobility relative to that in pure silicon. It is believed that this effect is a consequence of the effects of these dopants on the concentrations of certain grain boundary dislocation kinks and jogs. (Donors enhance this concentration and acceptors decrease it.[22,23]) In polysilicon, twins comprise the majority of grain boundaries. The migration of twins should be dependent on the climb and glide of grain boundary dislocations containing these kinks and jogs. Thus, because grain boundaries in semiconductors are constrained to minimize the number of dangling bonds and because the electronic states of defects at grain boundaries can be affected by dopants it should be expected that the mobilities of grain boundaries in semiconductors should differ qualitatively from those in metals.

Large-angle grain boundaries in ceramics appear to approximate those in metals. There does not appear to be a strong constraint on the types of grain boundaries that can appear in ceramics, except for some particular boundary planes where the geometry forces like ions to be nearest neighbors. Some light on the nature of grain boundaries that appear in polycrystalline materials may be shed by a comparison of observed grain boundary mobilities at a given fraction of the absolute

Table 10.2. Mobilities for grain boundary motion during grain growth and A/α solid phase epitaxy.

Material	T/T_M	Mobility (m^3/Ns)	Activation energy Q_{bm} (kJ/mol)
Fe	0.62	1.3×10^{-10}	
Al	0.64	3.2×10^{-11}	54
MgO	0.63	8×10^{-14}	275
Si	0.64	1.8×10^{-18}	230
Si A/α	0.64	3×10^{-16}	260
Si	0.7	1.6×10^{-16}	230
Si A/α	0.7	6×10^{-15}	260
LiF	0.69	7.5×10^{-13}	150

melting point. Table 10.2 compares observed grain boundary mobilities in metals, silicon and ionic crystals.

5. Mechanisms of interface migration

5.1. Grain boundaries

5.1.1. Pure materials

As indicated above there are two extremes associated with the motion of grain boundaries. In one mode of boundary migration any site along the boundary is as likely to move from one of the bounding crystals to the other. In the other extreme mode of boundary migration, the jump of an atom from one of the bounding crystals to the other can occur only at certain sites along the grain boundary. Let us consider the first mode at this point. The treatment that follows is that originally given by Turnbull.[24] For an atom to move from a lattice site on one bounding crystal to an adjacent empty lattice site of the neighboring crystal, the unit step is equivalent to that of exchange of an atom with a vacancy in lattice diffusion. Since, every atom along the boundary is assumed to jump at the same rate it is only necessary to relate the boundary velocity to the jump rate of one atom at the boundary. Let the boundary move by the distance d for such a jump. The boundary velocity will then be given by

$$V = d\nu \exp(-\Delta G_{bm}/kT)\Delta G/kT$$

where ν is the frequency of vibration of the atoms at the boundary sites, ΔG_{bm} is the free energy of activation for the jump of an atom from one lattice to the other at the boundary in the absence of a driving force and ΔG is the difference in free energy of the system between that before the atom jump and that after the jump.

Measured values of some boundary migration velocities in pure metals yield values for the activation energy and pre-exponential term that vary markedly with orientation.[25] Indeed, it has been suggested that different grain boundaries move by different migration mechanisms.

It is recognized that the model on which the above equation for the boundary velocity is based is unrealistic in the sense that the kinetics of attachment/disattachment is assumed to be the same for every atom at the boundary. Turnbull's model is more likely to be applicable to the A/α solid phase epitaxy interface migration. Indeed, reference to Table 10.2 shows that in silicon the mobilities of the A/α boundaries are much higher than the grain boundary mobilities. We must conclude that since the crystalline nature of the grains surrounding the boundary has an effect on mobility of the boundary that Turnbull's model is too simple and not generally applicable to grain boundary migration.

Other models have since been proposed in an attempt to make more specific the actual motions of the atoms at the boundary in the migration process. In particular, Gleiter[26] has proposed a model based on the assumption that migration occurs by the motion of atoms from kinked lattice steps of close-packed planes in the shrinking crystal, that end at the boundary, to nearby similar steps associated with the growing grain. This model predicts a dependence of the mobility both on grain misorientation and for a fixed grain misorientation on boundary orientation. Some aspects of these predictions are in agreement with experiment (small misorientation) and other aspects disagree with experiment (the dependence near critical misorientations).

More recently, Smith and coworkers[27] have noted that boundary migration can occur by the migration of DSC* lattice dislocations, where such migration is usually limited by diffusion due to the presence of a dislocation climb component during this migration, and by thermally activated atomic shuffles or shears. They have noted that if the sum of the Burgers vectors of these migrating vectors does not equal zero, then there must be a net boundary sliding component accompanying the boundary migration. Evidence that such boundary sliding can occur during boundary migration has been found. However, if the sum of the Burgers vectors equals zero then the boundary must correspond precisely to a low-energy structure. The difference between the two situations, other than the absence of sliding in the latter case, is the need for long-range diffusion of vacancies in the former and its absence in the latter during boundary migration. It is possible that the difference in mobility of the $\Sigma = 7$ boundary and those of other boundaries is related to the differences just noted for the migration of coincidence and other boundaries.

---

*Displacement Shift Complete.

There are certain orientations for which the boundary migrates by a step mechanism, as illustrated in Figure 10.7. In this case, it is obvious that the boundary velocity should not conform to the previous equation. It is likely that in this case, spirals, that are formed when a crystal lattice dislocation meets a grain boundary with its Burgers vector having a component normal to the boundary plane, will provide the means of such stepwise growth, much as they do in the growth of crystals from the vapor phase. Accordingly, the equation for the migration velocity becomes

$$V = nv_s h$$

where n the length of step per unit area equals $\Delta G/(20\,h\sigma d^2)$ in analogy with the crystal growth case,[28] v_s is the step velocity and equals $(d\nu\Delta G/kT)\exp(-\Delta G^*/kT)$, h is the step height, σ is the grain boundary energy, d is the atomic diameter. Substitution then yields

$$V = [(\Delta G)^2\nu\,/(20\sigma dkT)]\exp(-\Delta G^*/kT)$$

for the boundary migration velocity. Comparison of the two equations for the boundary velocity shows that the main difference is in the pre-exponential term, where the latter provides that the boundary velocity depends on the square of the driving force while boundary migration in the absence of spiral steps varies linearly with the driving force. There is evidence for a dependence of boundary velocity on the square of the driving force in grain growth. However, as noted previously, anisotropy of grain boundary energy can also provide the equivalent of a square dependence on the driving force by yielding $n = 0.25$ as the exponent in the empirical growth law.

There are many experiments that support the concept that generation[16b] and diffusion of vacancies is involved in the migration of some grain boundaries. Lucke and Estrin[29] have considered the effects of vacancies on the boundary migration process theoretically, stemming from Lucke's work on the effect of solutes on boundary migration, which will be considered in the next section. Recent experiments involving the effect of ion bombardment on grain growth lead to the conclusion that vacancies generated at or near grain boundaries are required for grain boundary migration, at least in Au, Ge and Si.[30] However, comparison of the activation energy for boundary migration and that for vacancy formation yield values of the former that are both smaller and larger than the latter. Much more experimentation will be required to unravel the mechanisms of grain boundary migration, which incidentally, need not be the same in all materials, nor the same for different grain boundaries in the same material.

Molecular dynamic simulations of flat high-angle twist boundaries in gold[31a] and in copper[31b] have shown that the migration mechanism involves collective reshuffling, either of correlated, in-plane rotational displacements or of local

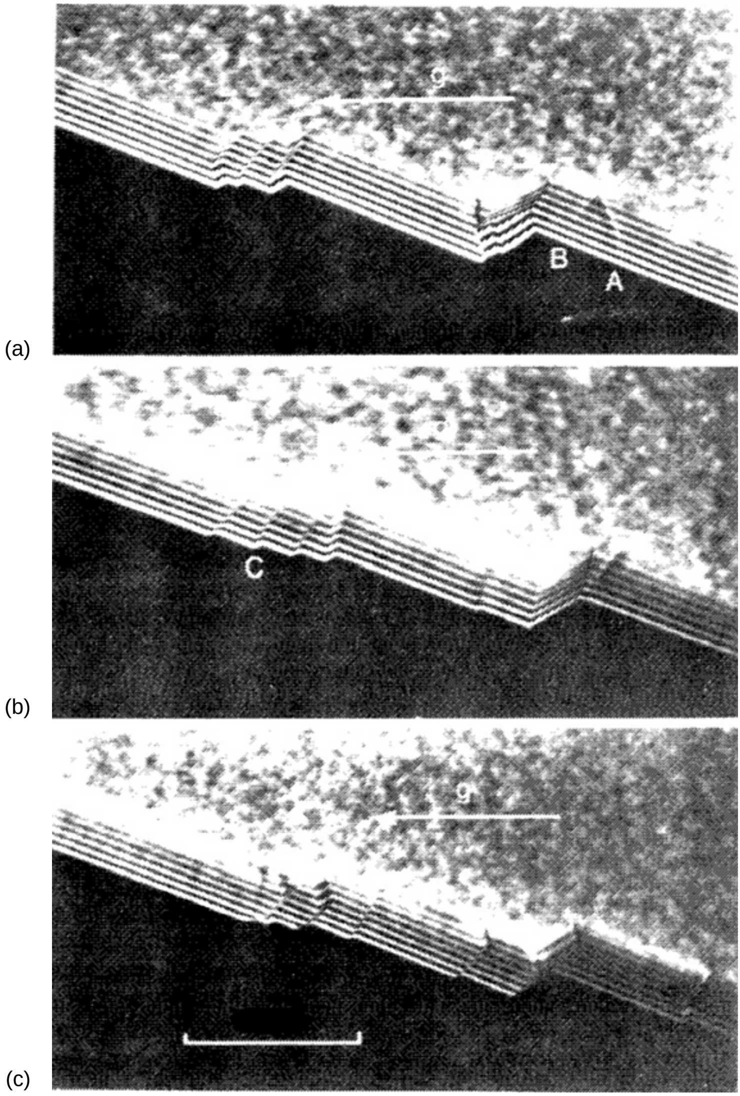


Figure 10.7. Contribution of step motion to growth of an annealing twin in aluminum; the marker represents 100 nm. From **Grain Boundary Structure and Kinetics**, ASM, Metals Park, Ohio, 1980, p. 337 with permission.

disordering (“melting”) of small groups of atoms and subsequent resolidification onto the other crystal.

Perhaps the most exciting information concerning the mechanism of grain boundary migration has been furnished by Nature via experiments involving

HRTEM (high-resolution transmission electron microscope) observations.[32] Movies have been made of the atomic motions at grain boundaries which show that reversible fluctuations involving collective motions of on the order of 100 atoms occur at grain boundaries of various types! The crystalline nature of the arrays in motion were maintained so that if any "melting" occurred it took place over a time interval less than that separating two frames of the movies, i.e. <0.03 second. A more likely analogy for the collective motion is that associated with a martensitic transformation.* This is a significant step forward in our knowledge of the detailed mechanism associated with grain boundary migration and perhaps, if followed up, it can provide a link between modern heterophase fluctuation theory[33] and experiment. Indeed, the concept that heterophase fluctuations were observed in the experiments of Merkle et al.[32] makes sense. The experiments of Seidman,[34] referred to in Chapter IV, demonstrate beyond any doubt that there are a multitude of grain boundary structures (phases) at a constant tilt or twist angle separated by very slight differences in free energy. One impression I received on viewing the video frames is that localized excess free volume along the boundary provided one condition allowing for the collective motion along with atomic shuffles of a group of atoms and another impression is that the passage of the excess free volume through the group of atoms during their collective motion facilitated this motion.

5.1.2. Homogeneous alloys

Still less is known about the mechanism of boundary migration in homogeneous alloys as compared to pure materials. The tendency for equilibrium segregation at a grain boundary must affect the migration kinetics. This phenomenon is known as the solute drag effect and its treatment does not involve any exploration of an effect on the mechanism of the boundary migration, i.e. on the mobility of the migrating interface. Rather, the theory of this effect is concentrated on a determination of the relationship between the solute–grain boundary interaction and the driving force. We shall consider this effect in detail in a later section.

5.2. Boundaries between different phases

5.2.1. Amorphous/crystal interface migration

Many of the studies of the A/α interface migration are consistent with a model based on an interface situated rate limiting process. A candidate for the latter is that of the breaking of bonds to form dangling bonds and the elimination of dangling bonds elsewhere along the interface.[35] Also, in the amorphous/crystalline interface migration[36] that results in the crystalline phase of rutile in TiO_2, the activation energies measured are not consistent with any bulk diffusion process in this material, but are consistent with interface diffusion steps.

---

* See Figure 5.13.

According to the list of interface migration driving forces at the end of Section 1 of this chapter, the driving force for the amorphous/crystal interface migration is not small compared to the thermal energy RT. This situation also holds for interface migration during solidification. Under this circumstance, the approximation that yielded a proportionality between interface velocity and driving force is no longer valid and we need to determine the interface velocity as the difference between a forward and its reverse rate process. Consequently, the interface velocity is given by the following dependence on the driving free energy ΔG

$$v = v_o(1 - \exp[-\Delta G/RT])$$

Pulsed laser melting followed by rapid alloy solidification has enabled measurements to be made of interface velocity verifying the applicability of this equation. These measurements have also confirmed that the continuous growth model of Aziz and Kaplan[37] is consistent with experiment[38] and that solute drag[39] is not active in this process. We consider this model of Aziz and Kaplan in detail in the next chapter.

5.2.2. Partitionless transformations

A metastable phase in a multicomponent material can transform to a more stable phase via a partitionless transformation by the motion of an interface, i.e. the product phase has the same composition as the parent phase. However, there is a difference in the configurations of the atoms on both sides of the interface. The rearrangements that are needed must occur within the interface and these rearrangements determine the mobility of the interface. These transformations are to be distinguished from the diffusionless "martensitic" type transformations in that the interface motion in the latter transformations does not depend upon thermal activation as it does in the former type of partitionless transformations. The class of "massive" transformations is a member of the set of partitionless transformations. Of course, in pure materials, the phase transformations that occur by the thermally activated migration of interfaces between parent and product phases, such as the growth of Widmanstatten ferrite plates from metastable austenite also obey the same kinetic laws applicable to the migration of interfaces in the absence of coupled processes. The relation between migration velocity and driving force for the interface separating product from parent phase is the same as for grain boundary migration.

There is one situation where another irreversible process couples with the interface migration process and does not involve long-range diffusion or heat flow. This situation involves an interaction between a moving interface and solute atoms that results in a drag resistance that effectively decreases the driving force acting to move the interface. This solute drag phenomenon is discussed in a later section.

5.3. Effect of interface roughness on migration mechanism

Since singular interfaces (facets) can migrate only as a consequence of the lateral motion of steps, whereas a rough interface may migrate via a continuous mode (as per Turnbull) or via a stepwise growth, it is desirable to have a criterion which can distinguish between these modes of growth. Two procedures have been used to devise such a criterion. One is based on thermodynamic concepts and is due to Cahn.[40] The other is based on statistical concepts and is due to Jackson.[41] Although the former was developed for solid state transformations and the latter for liquid–solid interfaces there is no reason why they should not apply to incoherent interfaces between phases having constant compositions, regardless of the nature of the phases.

According to Cahn, if the interface can attain a metastable equilibrium configuration during growth, in presence of a driving force, then stepwise growth is required when the driving force is less than some critical value. On the other hand, if the local displacements of an interface in its motion lead to no local minima in the free energy of the interface then the interface configuration can change in the course of growth. This condition is fulfilled for all interfaces above a critical value of the driving force. The need for consideration of the magnitude of the driving force in evaluating the operating mode of growth (i.e. stepwise or continuous) is obvious for the case of a singular interface near a cusp orientation. In this case, unless the interface can be reoriented away from the cusp orientation the interface will remain singular and grow by step propagation. The driving force needs to be sufficiently large to accomplish this reorientation for continuous growth to be possible. On the other hand, it is not obvious why non-singular interfaces cannot grow in a continuous manner. Yet, according to Cahn, if a lateral growth step mechanism is required whenever the interface is able to attain a metastable condition in the presence of a driving force then the growth will be non-continuous.

According to Jackson,[41a] the interface can be taken to correspond to a mixture of empty and filled sites at the interface. With N the total number of sites per unit area, the fraction of filled sites is taken to equal N_A/N and $(1 - N_A/N)$ is the number of unfilled sites. The change in free energy associated with the filling of such sites is

$$\begin{aligned}\Delta G/NkT_e = {} & \alpha N_A(N - N_A)/N^2 + (N_A/N)\ln(N_A/N) \\ & + [(N - N_A)/N]\ln[(N - N_A)/N]\end{aligned}$$

where $\alpha = L/kT_e$, L is the heat of the transformation and T_e is the equilibrium temperature of the transformation. This equation predicts that for $\alpha < 2.76$, the interface is rough and growth is continuous, while for $\alpha > 2.76$, the interface is smooth and growth occurs in a stepwise mode.* Regardless of the validity of the derivation, there is evidence to support this separation between the two modes of growth in the liquid–solid transition at about $\alpha = 2$.

---

* When $\alpha > 2.76$ the free energy as a function of the fraction of filled sites exhibits two stable values of the latter at the minimum in the free energy, whereas for $\alpha < 2.76$ only one value of the fraction of filled sites is at the minimum in the free energy.

Jackson[41b] has shown that α plays a role in the rate of migration of the solid–liquid interface in a pure material. He derives a net growth rate, given by

$$V = v' J \alpha f(h, k)(\Delta T/T) \exp(-Q/RT)$$

where v' is the atomic volume, $\alpha = L/kT_e$ as before, f(h, k) is a geometric factor that is dependent on α and the Miller indices, $\Delta T = T - T_e$, Q is the activation energy for diffusion in the liquid and J is a frequency factor. Jackson found that for the orientation of the fastest growing edges (faces), f(h, k) is independent of α. Also, he found that for small α, f(h, k) is nearly independent of the crystal indices. However, at high α, where there is a large difference between the f(h, k) values for different orientations, the fastest growing edges (faces) disappear early in the growth leaving behind the slowest growing edges (faces). The implication is that at small values of α, growth is almost isotropic (non-faceted), whereas, for high α, growth is faceted; a conclusion that is in agreement with Jackson's first model.

5.4. Effect of solute on boundary mobility

It is well known that boundary mobility can be very sensitive to the presence of solute. Indeed, only a few parts per million of solute can change the boundary mobility by orders of magnitude, as shown in Figure 10.6. The theory of the effect of solute on grain boundary mobility has been explored by Lucke and coworkers[42] and by Cahn.[43] Their theories based on the existence of a distance-dependent interaction energy between the solute and the grain boundary E(x), yield different relations between the boundary velocity and the driving force. For the low velocity limit, this relation takes the form

$$v = P/[M^{-1} + \alpha C]$$

where M is the mobility, P is the driving force, C is the bulk solute concentration and

$$\alpha = 4N_v kT \int_{-\infty} \{[\sinh^2(E(x)/2kT]/D(x)\} dx$$

where D(x) is the solute diffusivity, which is a function of the distance x, between solute and boundary. Rearrangement of the terms in the above relation for the boundary velocity makes it apparent that the solute drag resistance is given by $\alpha C v$ in this low velocity limit.

At the high velocity limit, the boundary velocity becomes

$$v = MP(1 - \alpha C/(\beta^2 P^2 M))$$

where $$\alpha/\beta^2 = (N_v/kT)\int_{-\infty}^{\infty} (\partial E/\partial x)^2 D(x)dx$$

The relation applicable to both regimes is

$$P = M^{-1}v + \alpha Cv/(1 + \beta^2 v^2)$$

An experimental dependence of boundary velocity on temperature in gold containing 20 ppm of iron, which is qualitatively consistent with this theory is presented in Figure 10.8. As shown, at low boundary velocities, the apparent activation energy approaches a constant value, which is determined by the function D(x). It is found experimentally that the activation energy for boundary migration in this low velocity regime approximates that for solute diffusion in the bulk. This result according to the theory suggests that D(x) in this regime approximates a step function,

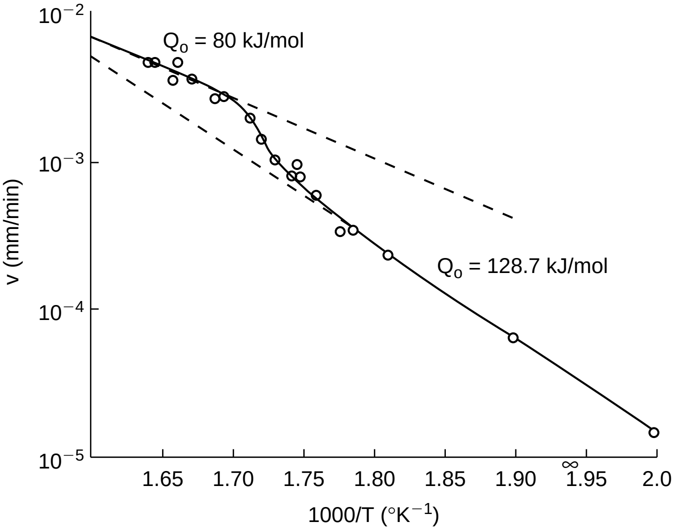


Figure 10.8. The effect of temperature on the migration velocity of a 30° <111> tilt boundary in gold containing 20 ppm iron. From W. Grunwald and F. Haessner, Acta Metall. 18, 217(1970).

i.e. D_{sb} within the grain boundary of thickness δ, and D_{sl} outside the grain boundary. Further, in the high boundary velocity regime, it would appear that the solute drag can contribute little resistance to the boundary migration and in this case the temperature dependence of the boundary velocity is determined mainly by the grain boundary diffusivity rather than the bulk diffusivity. A recent result on the effect of Ga in Al on the mobility of a $<111>$ tilt boundary with a tilt angle close to $\Sigma = 7$ to the effect that a few parts per million of Ga increases the mobility[44] significantly emphasizes the relative significance of the boundary diffusivity because it is known in this system that a liquid layer of grain boundary thickness is formed at the grain boundaries at these temperatures.[45]

There are data which support some of the concepts associated with these theories. In particular, in the low-velocity region it is predicted that the reciprocal of the boundary velocity should depend linearly on the solute concentration and this functional relationship has been found in the work of Gordon and Vandermeer[46] in dilute aluminum alloys. Further, the relation between velocity and driving force given by the last equation has been verified qualitatively in the experiments of Sun and Bauer[47] and Drolet and Galibois.[48] However, some aspects of the results of the latter workers are not in agreement with the theory, as noted by Simpson et al.[49] and the relationship between boundary velocity and solute concentration shown in Figure 10.6 does not obey this relationship. Thus, there is some uncertainty regarding the limits of application of this theory. Recently, in a review of the solute drag phenomenon, Mendelev and Srolovitz[50] have noted additional inadequacies of the theories of Cahn and Lucke and collaborators. In particular, they note that the theory needs to be modified to account for non-ideal behavior of the solutes, for the simultaneous presence of more than one type of impurity atom, and for the possibility of grain boundary impurity saturation. They provide such corrections. However, they believe that more drastic revisions of the theory are required. First, they suggest that boundary migration involves the formation and migration of kinks at the boundary and that the Cahn, Lucke, (Detert), Stuwe theory neglect this possibility. Inasmuch as these theories are not transparent to experimental verification Mendelev and Srolovitz developed a Monte Carlo model of grain boundary migration influenced by solute drag. Figure 10.9 taken from their work provides a comparison between the CLS* theory and the predictions of the simulation. It is apparent that there is a marked disagreement between CLS and the simulation results for the parameters corresponding to the figure. The CLS model does yield the breakaway phenomenon apparent in Figure 10.9, as shown in Figure 10.10, which represents a plot of a solution to the CLS model for another set of parameters. The conclusion from these figures is that the CLS model does not yield the breakaway as easily as does the simulation. One possible reason for this result is that the boundary in the Monte Carlo simulation can undergo local fluctuation in its curvature which can hasten breakaway as compared to the constant curvature

---

* CLS = Cahn, Lucke, Stuwe

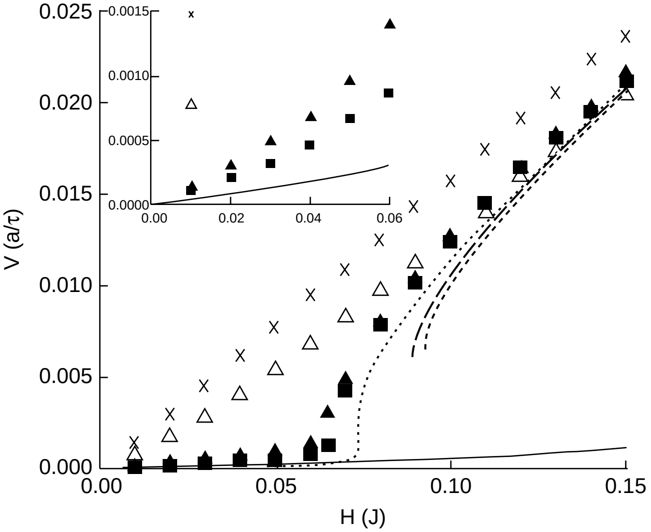


Figure 10.9. Simulation results for boundary velocity V, versus driving force H. The x points represent the pure system where V is proportional to H. The filled and open points correspond to attractive and repulsive impurity interactions with the grain boundary for one set of parameters. The triangles and squares correspond to impurites which do not interact with each other and do interact, respectively. The solid line represents the CLS model with a solute–grain boundary attractive interaction energy 1.5 times that for the dashed lines with solute boundary concentration equal to 0.25 for the solid line and long-dashed line and 1 for the short-dashed line. The dotted line corresponds to a repulsive interaction of the same strength as the dashed lines and a solute boundary concentration of 0.25. Dotted and dashed lines correspond to a discrete model.[50] The inset show a magnification of the data at small values of the driving force. Reproduced with permission from Interfac Sci. 10, 91–98 © 2002 Kluwer.

boundary of the CLS model. Another is that boundary migration in the simulation involves the formation and migration of kinks along the boundary which interact differently with solute atoms than as described in the CLS model.

The theories of Cahn and Lucke and Stuwe yield complicated relations which, in practise, are normally simplified, as in the above discussion, to exhibit the behavior characteristic of the low and high boundary velocity regimes. These are regimes in which, respectively, the solute diffuses along at the same velocity as the boundary and in which it does not diffuse along with the boundary but does exert a drag on it. We can provide simple physical concepts for these two regimes as follows.

For the low-velocity regime, the velocity of a diffusing solute is given by the Nernst–Einstein relation as $<v_i> = D_iF_i/kT$, where D_i is the tracer diffusivity

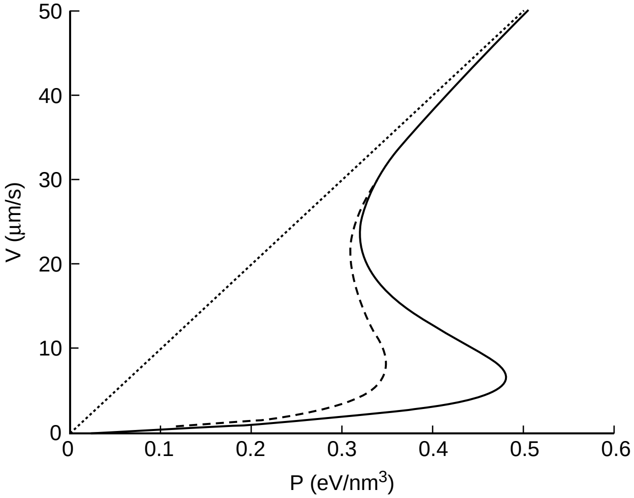


Figure 10.10. Solution to the CLS equations for a particular set of parameters. From Modell. Simul. Mater. Sci. Eng. 10, R79(2002) with permission. © 2002 IOP Publishing Ltd.

for the solute and F_i is the force acting on the solute atom. Now, to a good approximation, F_i equals the force acting on unit area of the moving grain boundary P, divided by the number of solute segregated to unit area of the boundary. For a slow moving boundary the latter number should be equal to the solute concentration in the bulk C, multiplied by the segregation coefficient s. Thus, for sufficient solute for v to equal $<v_i>$, then $v = D_iP/\delta skTC$, where δ is the thickness of the grain boundary containing the segregated solute. For C sufficiently small for $v \neq <v_i>$, then $v = MP$. In the intermediate regime $v = P/[M^{-1} + \delta skTCD_i^{-1}]$, the same relation derived from the more accurate theory.

When the solute atoms remain stationary in the high-velocity regime, each atom exerts a drag force corresponding to the maximum interaction force between boundary and solute atom that produces a reverse curvature of the grain boundary, i.e. one that opposes the motion of the grain boundary. If the maximum drag force that the solute atom can exert on the grain boundary is f_{max} then $v = M(P - C\delta f_{max}/2)$, where δ is the grain boundary thickness. When the solute atoms move, but at a lower velocity than the grain boundary, and the force acting on the boundary is larger than the sum of the solute drag forces, then the drag force exerted depends upon the relative velocities of the solute atoms and the grain boundary. We know that when the boundary and solute atoms move at the same velocity that the kinetic segregation coefficient must equal the equilibrium value s. Also, when the boundary moves much faster than the solute atoms the kinetic segregation coefficient approximates unity corresponding to stationary solute. Thus, allowing the kinetic segregation coefficient to depend linearly on the normalized difference in velocity of the grain boundary and solute atoms, both acted upon by the same average drag force $f_{max}/2$, we obtain that $v = M[P - (C\delta f_{max}/2)\{(1 - D_i f_{max}/2kTv)(1 - s) + s\}]$. This relation approximates to that for the more accurate theory for the regime in which the kinetic segregation coefficient is less than the equilibrium value in that the higher is the grain boundary velocity relative to that of the solute atoms the smaller is the drag force due to the solute atoms.

A phase field model has also been used to simulate the effect of solute drag on grain growth kinetics.[4] In the low velocity regime that was studied in this simulation the dependence of velocity on driving force was found to be non-linear. A consequence of this result is that the grain growth exponent can vary from 2 to 3 dependent upon the ratio of lattice diffusion to grain boundary mobility.

The models of solute drag have also been applied to a study of the effect of solute on boundary migration in non-metallic substances.[51] One such fascinating study involves ice in the Antarctica.[52]

6. Summary

We have shown that the phenomena of A/α solid phase epitaxy, primary recrystallization, secondary recrystallization, grain growth, thermally activated phase transformation in pure materials and partitionless transformations can be described in common terms, i.e. interface migration velocity equals the product of a mobility and a driving force. The driving forces for those processes obeying this relation involve an energy that is small compared to the thermal energy. The laws governing grain growth in each of these regimes were related to the corresponding driving forces. We found that a distribution of insoluble particles and that a strong crystallographic texture can lead to the cessation of growth. The mobilities of the grain boundaries were then discussed and mechanisms for their migration were considered and found to be wanting relative to the elegant experimental discovery that cooperative motion of about 100 atoms is involved in the mechanism of boundary migration. The effect of roughness on the mode of interface migration – ledge or continuous – was then discussed. Then the large effect of solute on the grain boundary migration velocity was described.

References

1. D.J. Srolovitz, G.S. Grest and M.P. Anderson, Acta Metall. 33, 2233(1985).
2. J.C.M. Li, in **Recrystallization, Grain Growth and Textures**, ed. H. Margolin, ASM, Metals Park, Ohio, 1966, p. 45.

3. W.B. Hutchinson, S. Jonsson and L. Ryde, Scripta Metall. 23, 671(1989); L. Ryde, W.B. Hutchinson and S. Jonsson, in **Recrystallization** '90, ed. Chandra, TMS, Warrendale, PA, 1990, p. 313; T. Furu, K. Marthenson and E. Nes, Mat. Sci. Tech. 6, 1093(1990).
4. D. Raabe, Adv. Eng. Mater. 4, 255(2002); D. Fan et al., J. Mater. Res. 14, 1113(1999).
5. X. Song, M. Rettenmayr and G. Liu, Image Anal. Stereo. 22, 163(2003).
6. R.W. Cahn, **Physical Metallurgy**, eds. R.W. Cahn and P. Haasen, North-Holland, Amsterdam, 1983.
7. E.A. Holm, M.A. Miodownik and A.D. Rollett, Acta Mater. 51, 2701(2003).
8. D. Weygand Y. Brechet and J. Lepinoux, Interface Sci. 9, 311(2001).
9. D.A. Hughes, D.C. Chrzan, Q. Liu and N. Hansen, Phys. Rev. Lett. 81, 4664(1998); A. Godfrey and D.A. Hughes, Acta Mater. 48, 1897(2000); D.A. Hughes, Scripta Mater. 47, 697(2002); D.A. Hughes and N. Hansen, Acta Mater. 48, 2985(2000).
10. A.D. Rollett W.W. Mullins, Scripta Mat. 36, 975(1997).
11. S.B. Lee, N.M. Hwang, D.Y. Yoon and M.F. Henry, Metall. Mater. Trans. A 31A, 985(2000); J.B. Koo, D.Y. Yoon and M.F. Henry, Metall. Mater. Trans. A 33A, 3803(2002); G.S. Rohrer, C.S. Rohrer and W.W. Mullins, JACS 84, 2099(2001).
12. M.P. Anderson, D.J. Srolovitz, G.S. Grest and P.S. Sahni, Acta Metall. 32, 783(1984); C.E. Krill, III and L.-Q. Chen, Acta Mater. 50, 3057(2002); A. Miyake, Contrib. Mineral Petr. 130, 121(1998); X. Song, G.J. Liu, J. Univ. Sci. Technol. Beijing 5, 129(1998); K. Fuchizaki, T. Kusaba and K. Kawasaki, Phil. Mag. B 71, 333(1995); D. Weygand, Y. Brechet, J. Lepinoux, W. Gust. Phil. Mag. B 79, 703(1999); F. Wakai, N. Enomoto and H. Ogawa, Acta Mater. 48, 1297(2000); F. Wakai, J. Mater. Res. 16, 2136(2001); J.A. Glazier, Phys. Rev. Lett. 70, 2170(1993); Y. Saito, ISIJ Intern. 38, 559(1998); X. Xue, F. Righetti, H. Telley and T.M. Liebling, Phil. Mag. B 75, 567(1997).
13. N.P. Louat, Acta Metall. 22, 721(1974).
14. M. Hillert, Acta Metall. 13, 227(1965).
15. (a) D. Moldavan et al., Phil. Mag. A 82, 1271(2002); (b) M. Upmanyu et al., Interface Sci. 10, 201(2002); A. Kazaryan, Y. Wang, S.A. Dregia and B.R. Patton, Acta Mater. 50, 2491(2002).
16. (a) M. Upmanyu, D.J. Srolovitz, L.S. Shvindlerman and G. Gottstein, Acta Mater. 50, 1405(2002); (b) M. Upmanyu, R.W. Smith and D.J. Srolovitz, Interface Sci. 6, 41(1998).
17. M. Upmanyu, D.J. Srolovitz, L.S. Shvindlerman and G. Gottstein, Interface Sci. 6, 287(1998).
18. C. Zener, in C.S. Smith, Trans. AIME 175, 47(1948).
19. P.R. Rios, Acta Metall. 35, 2805(1987).
20. M. Miodownik, E.A. Holm and G.N. Hassold, Scripta Mater. 42, 1173(2000).
21. M. Upmanyu, D.J. Srolovitz, L.S. Shvindlerman and G. Gottstein, Acta Mater. 47, 3901(1999).
22. P.B. Hirsch, J. Phys. Colloq. C 6, 117(1979).
23. D.A. Smith, T.Y. Tan and C. Fontaine, in **Advances in Ceramics**, vol. 6, eds. M.F. Yan and A.H. Heuer, American Ceramics Society, 1983.
24. D. Turnbull, Trans. Metall. Soc. AIME 191, 661(1951).
25. M. Winning, G. Gottstein and L.S. Shvindlerman, Acta Mater. 50, 353(2002).
26. H. Gleiter, Acta Metall. 17 (565), 853(1969).
27. D.A. Smith, C.M.F. Rae and C.R.M. Grovenor, in **Grain Boundary Structure and Kinetics**, ASM, Metals Park, Ohio, 1980, p. 337.

28. W.K. Burton, N. Cabrera and F.C. Frank, Phil. Trans. Roy. Soc. A 243, 299(1950–1951).
29. Y. Estrin and K. Lucke, Acta Metall. 30, 983(1982).
30. H.A. Atwater, C.V. Thompson and H.I. Smith, J. Appl. Phys. 64, 2337(1988).
31. (a) R.-J. Jhan and P.D. Bristowe, Scripta Metall. Mater. 24, 1313(1990); (b) B. Schonfelder et al., Interface Sci. 5, 245(1997).
32. K.L. Merkle, L.J. Thompson and F. Phillipp, Phys. Rev. Lett. 88, 225501(2002); Interface Sci. 12, 277(2004).
33. V.I. Yukalov, Phys. Rep. 208, 395(1991).
34. D. Seidman, Ann. Rev. Mater. Res. 32, 235(2002).
35. K.A. Jackson, J. Mater. Res. 3, 1218(1988).
36. M.D. Wiggins, M.C. Nelson and C.R. Aita, MRS Symp. Proc. 398, 381(1996).
37. M.J. Aziz and T. Kaplan, Acta Metall. 36, 1335(1988).
38. M.J. Aziz, J.Y. Tsao, M.O. Thompson, P.S. Peercy and C.W. White, Phys. Rev. Lett. 56, 2489(1986); P.M. Smith and M.J. Aziz, Acta Metall. Mater. 42, 3515(1994); J.A. Kittl, M.J. Aziz, D.P. Brunco and M.O. Thompson, J. Cryst. Growth 148, 172(1995).
39. J.A. Kittl, M.J. Aziz, D.P. Brunco and M.O. Thompson, MRS Symp. Proc. 398, 119(1996).
40. J.W. Cahn, Acta Metall. 8, 556(1960).
41. (a) K.A. Jackson, in **Liquid Metals and Solidification**, American Society for Metals, Metal Park, OH, 1958; (b) J. Cryst. Growth 3/4, 507(1968).
42. K. Lucke and K. Detert, Acta Metall. 5, 628(1957); K. Lucke and H.P. Stuwe in **Recovery and Recrystallization of Metals**, ed. L. Himmel, Wiley-Interscience, NY, 1962, p. 171; K. Lucke and H.P. Stuwe, Acta Metall. 19, 1087(1971).
43. J.W. Cahn, Acta Metall. 10, 789(1962).
44. D. Weygand, Y. Brechet, E. Rabkin, B. Straumal and W. Gust, Phil. Mag. Lett. 76, 133(1997).
45. B. Straumal, T. Muschik, W. Gust and B. Predel, Acta Metall. Mater. 40, 939(1992).
46. P. Gordon and R.A. Vandermeer, Trans. Metall. Soc. AIME 215, 577(1959); 224, 917(1962); in **Recrystallization, Grain Growth and Textures**, ed. H. Margolin, ASM, Metals Park, Ohio, 1966, p. 205.
47. R.C. Sun and C.L. Bauer, Acta Metall. 18 (635), 639(1970).
48. J.P. Drolet and A. Galibois, Metall. Trans 2, 53(1971).
49. C.J. Simpson, W.C. Winegard and K.T. Aust, in **Grain Boundary Structure and Properties**, eds. G.A. Chadwick and D.A. Smith, Academic Press, New York, 1976, p. 201.
50. M.I. Mendelev and D.J. Srolovitz, Modell. Simul. Mater. Sci. Eng. 10, R79(2002).
51. A.M. Glaeser, H.K. Bowen and R.M. Cannon, J. Am. Cer. Soc. 69, 119(1986).
52. J. Weiss et al., Ann. Glaciol. 35, 552(2002).

Bibliography

1. R.W. Cahn, in **Physical Metallurgy**, vol. 2, eds. R.W. Cahn and P. Haasen, North-Holland, New York, 1983, p. 1595.
2. H. Hu, in **Metallurgical Treatises**, eds. J.K. Tien and J.F. Elliott, TMS-AIME, Warrendale, PA, 1981, p. 385.

3. E. Nes, N. Ryum and O. Hunderi, Acta Metall. 33, 11(1985). This paper is a good summary of theories on the effect of particles on boundary migration.
4. F.J. Humphreys and M. Hatherly, **Recrystallization and Related Annealing Phenomena**, Elsevier Science, New York, 1995.

Problems

1. Distinguish in terms of driving force between primary recrystallization, secondary recrystallization and grain growth.
2. Why do grain boundaries migrate toward their center of curvature in grain growth and not in the recrystallization processes?
3. Why is it necessary to be able to achieve secondary recrystallization to immobilize the as primary recrystallized grain boundaries?
4. Why do low-angle boundaries have low mobilities for migration?
5. Provide a reason for the fact that the activation energy for high-angle boundary migration is less than that for diffusion along these boundaries.
6. What would you expect the grain boundary diffusivity for a coincidence boundary to be as compared to that of a boundary inbetween coincidence orientations? Substitute the words "migration mobility" in place of "diffusivity" in the above and answer the question again. *Hint*: The answer to the latter question depends upon the excess volume at the grain boundary.
7. Why does the presence of solute exert such a large effect upon the velocity of grain boundary migration?
8. Why is the effect of impurities on grain boundary velocity small at high values of the velocity or driving force?
9. Why does the impurity drag on grain boundaries exhibit a maximum as the velocity of grain boundary migration increases? *Hint*: At high velocity how does the concentration of solute at the grain boundary compare to that at low velocity?
10. Prove that in the absence of any restriction to grain growth that secondary recrystallization is not possible.

CHAPTER XI

Growth of Phases: Diffusion or Interface Reaction Control

Introduction

Most of the processes studied in this chapter involve transport of matter across an interface while a phase transformation also takes place at this interface, all under conditions of morphological stability and in the solid state. The interaction between the diffusion process and the interface reaction is of great interest and we develop the coupling between the two processes. Also, we consider processes that involve matter transport to an interface with a divergence of this transport at the interface. We do not explore the latest results of studies of each phenomenon considered in this chapter. To do so would involve book length reviews of each and is obviously not in the scope of this introductory text. Rather, we limit ourselves to a development of the coupling between the various transport phenomena in order to emphasize their significant physical aspects. An important factor in this coupling of atom transport to and/or from an interface and the interface reaction is the stress developed ahead of the moving interface due to a gradient of composition. Most theories of moving interface processes have not included stress as a factor. In many of these processes, non-equilibrium thermodynamics contributes a formalism that is often useful in defining the relations between parameters.

1. Diffusion couple (alpha/beta), no intermediate phase

1.1. Components have equal molar volumes

Suppose that a semi-infinite element A is placed into intimate contact with another semi-infinite element B and then is brought to a temperature T at which diffusion occurs for a finite time. If the corresponding binary phase diagram is the type illustrated in Figure 11.1(a), then a plot of the composition along the diffusion couple would appear as shown in Figure 11.1(b).

It is often asked why does not a two phase region appear between the two terminal solid solutions. The answer to this question is simple. Suppose that a two-phase region was produced between the two terminal solid solutions having the respective equilibrium compositions of the corresponding terminal solid solutions.

Thus, in phase alpha, the concentration of A(B) is constant at that equilibrium value given by the phase diagram. Similarly, for phase beta. Since the concentration gradients in the two phase region equal zero, there is no transport of A or B across the two-phase region! Consequently, only at the interfaces of the two-phase region with the terminal solid solutions will the transport of atoms across the interface be possible. In particular, at the interface between the B poor solid solution (alpha) and the two-phase region, B atoms diffuse from the interface into the alpha solution. These B atoms cannot be provided by diffusion across the two-phase region, as just demonstrated. Hence, they can only be provided by the transformation of B rich beta phase at the interface to the B poor alpha phase, thereby causing the alpha/two-phase interface to move in the direction of consumption of the two-phase region. Similarly, at the A poor beta/two-phase interface, the diffusion of A into the beta phase requires the conversion of the A rich alpha phase to the A poor beta phase and the motion of the beta/two-phase interface in the direction of consumption of the two-phase region. Consequently, the two-phase region will disappear – it is not stable under the boundary conditions.

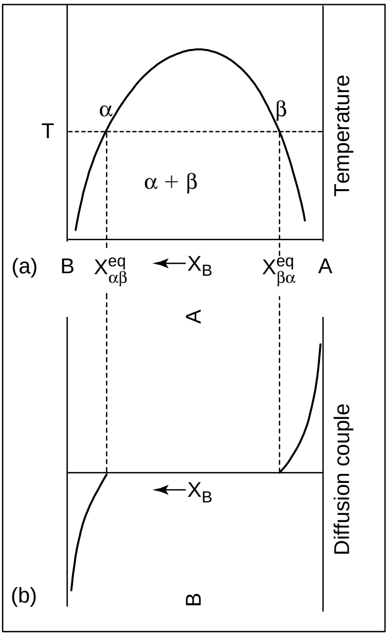


Figure 11.1. Illustrating compositions at interface of diffusion couple relative to phase boundary compositions in phase diagram.

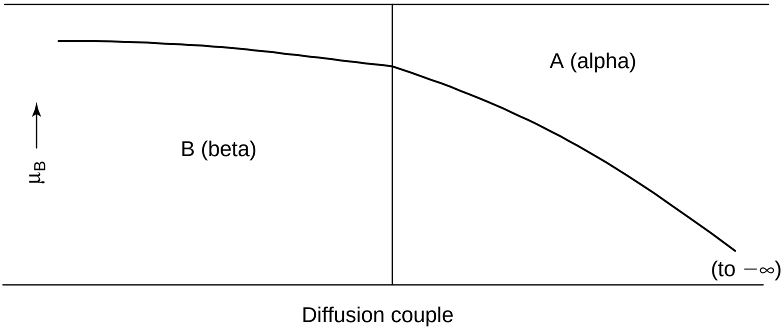


Figure 11.2. Illustrating how the chemical potential of species B varies continuously along the diffusion couple in the absence of interface migration.

The point so laboriously conveyed above can be made more simply using a plot of the chemical potentials of the components in the diffusion couple, as shown in Figure 11.2. Contrary to the concentrations, which exhibit a discontinuity at the alpha/beta interface, the chemical potentials are continuous, as shown. If a two-phase region were to exist between the terminal solid solutions, it is now obvious that the chemical potentials would have to be constant in this two-phase region. *Thus, there is no driving force for transport of matter across the two-phase region.*

1.2. Components have unequal partial molar volumes

The situation described in the previous paragraph corresponds to the case that local equilibrium holds at the alpha/beta interface, that the latter is incoherent, and

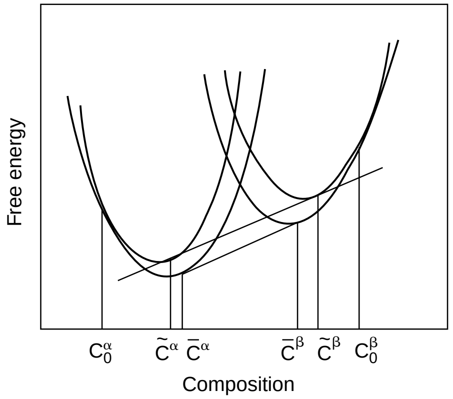

Free energy
C_0^α
$\tilde{C}^\alpha$
$\bar{C}^\alpha$
$\bar{C}^\beta$
$\tilde{C}^\beta$
C_0^β
Composition

Figure 11.3. The unstressed free energies are shifted by f_{el} upwards. The common tangent gives the coexisting compositions. From Acta Metall. 33, 331(1985) with permission. © 1985 Elsevier.

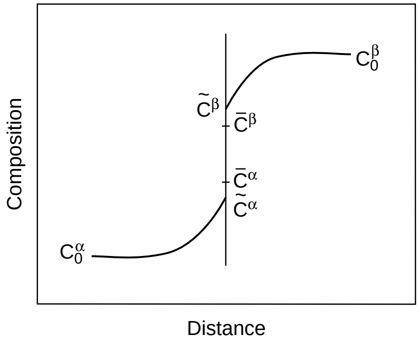

Composition
C_0^β
$\tilde{C}^\beta$
$\bar{C}^\beta$
$\bar{C}^\alpha$
$\tilde{C}^\alpha$
C_0^α
Distance

Figure 11.4. Compositions in a diffusion couple at incoherent interface. See Figure 11.3 for definition of compositions. From Acta Metall. 33, 331(1985) with permission. © 1985 Elsevier.

that the partial molar volumes of the diffusing species are equal. In the event that the partial molar volumes of the diffusing species differ then the gradients in composition will develop gradients in stress, which will affect the compositions along the diffusion couple. In this case, the diffusion potential replaces the chemical potential in the phenomenological relations for diffusion. Also, it is necessary to incorporate the effect of stress gradients on the diffusive fluxes. This problem has been considered by Larche and Cahn,[1] who have shown, in the absence of a stress normal to the interface, that interface compositions can be determined using free energy–composition curves. Figure 11.3 illustrates the construction and results. The shifted Helmholtz free energy for the phase is obtained by adding to that corresponding to the absence of stress the elastic energy per mole, which is just a function of the local composition and is given by

$$f_{el} = V_o E \eta^2 (C^i - C^i_o)^2 / (1 - \nu)$$

where V_o is the molar volume at composition C_o far from the interface, E is Young's modulus, $\eta = (V_1 - V_2)/(3V_o)$ and ν is Poisson's ratio. The interface compositions for $\eta \neq 0$ are shown relative to those for $\eta = 0$ in Figure 11.4.

2. Diffusion couple (alpha/beta) with intermediate phase, incoherent interfaces and at constant molar volume

2.1. One intermediate phase

Suppose now that the phase diagram for the A–B system exhibits one stable intermediate phase at the temperature T, as shown in Figure 11.5. What will occur after a diffusion anneal of the diffusion couple formed from A and B? The situation is more complex than that discussed above. Instead of only diffusion there is now an additional reaction – that of the transformation of crystal structures, which occurs at two interfaces: that of the interface between alpha and gamma and that of the interface between beta and gamma. This transformation is called an interface reaction. Normally, for a kinetic reaction to take place, the driving force for that reaction must be finite. For example, diffusion requires a finite value of the chemical potential gradient to proceed. In the case of the interface reaction, there must be a finite value of the difference in free energy between reactants and products to cause it to move and thereby transform one crystal structure to another of a different composition. If the chemical potentials of the species were continuous across the interfaces, as shown in Figure 11.5, then it is possible for species to be

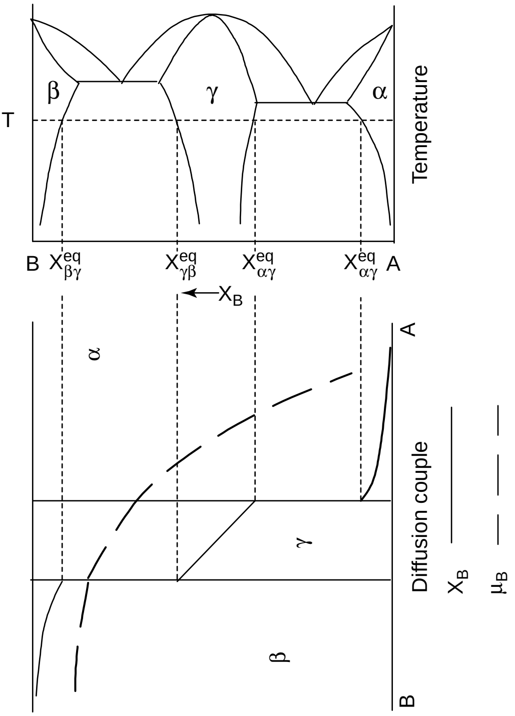


Figure 11.5. Interface compositions and solute chemical potential in a diffusion couple with intermediate phase under diffusion controlled conditions.

transported across these interfaces, but the interfaces themselves would not move, i.e. there would be no transformations and the thickness of the gamma phase would be constant. For transformations and interface motion to be possible, there must be a discontinuity in the chemical potentials at the interface, as shown in Figure 11.6, so that the difference in free energy between reactants and products is finite. In non-equilibrium thermodynamics, this driving force for the chemical reaction at the interface is called the chemical affinity.

2.1.1. Growth of gamma phase under condition that terminal solubilities are very small

In this section we will assume: either that the terminal phases are saturated; or that the terminal phases alpha and beta have such small solubilities for the respective minor components that diffusion in alpha and beta can be neglected with respect to diffusion in gamma. We will also assume that the difference in specific free energy between the pure component and that for the terminal phase, at the composition in

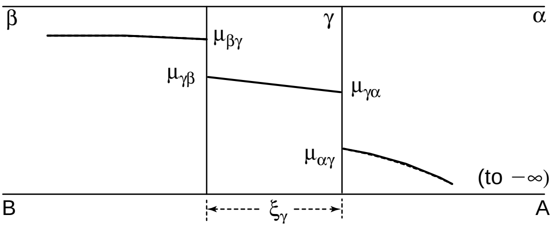

β
γ
α
$\mu_{\beta\gamma}$
$\mu_{\gamma\beta}$
$\mu_{\gamma\alpha}$
$\mu_{\alpha\gamma}$
(to −∞)
B
ξ_γ
A

Figure 11.6. Chemical potential of solute as a function of distance along diffusion couple with moving boundaries.

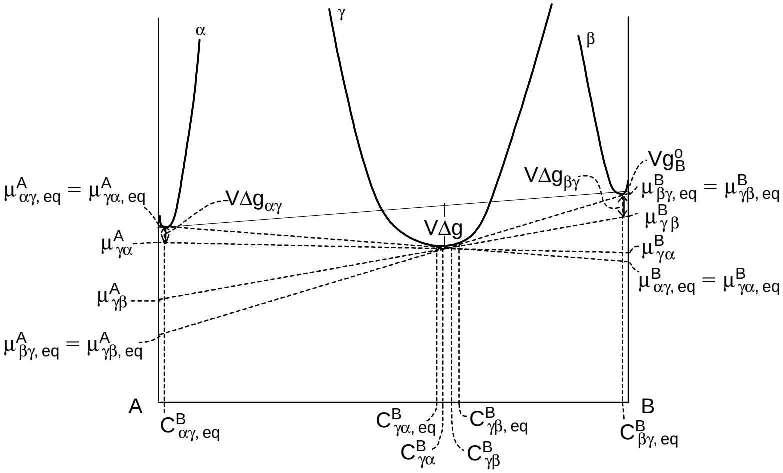

γ
α
β
Vg^o_B
$V\Delta g_{\beta\gamma}$
$\mu^B_{\beta\gamma,\,eq} = \mu^B_{\gamma\beta,\,eq}$
$\mu^A_{\alpha\gamma,\,eq} = \mu^A_{\gamma\alpha,\,eq}$
$V\Delta g_{\alpha\gamma}$
$\mu^B_{\gamma\beta}$
$V\Delta g$
$\mu^A_{\gamma\alpha}$
$\mu^B_{\gamma\alpha}$
$\mu^B_{\alpha\gamma,\,eq} = \mu^B_{\gamma\alpha,\,eq}$
$\mu^A_{\gamma\beta}$
$\mu^A_{\beta\gamma,\,eq} = \mu^A_{\gamma\beta,\,eq}$
A
B
$C^B_{\alpha\gamma,\,eq}$
$C^B_{\gamma\alpha,\,eq}$
$C^B_{\gamma\beta,\,eq}$
$C^B_{\beta\gamma,\,eq}$
$C^B_{\gamma\alpha}$
$C^B_{\gamma\beta}$

Figure 11.7. Defining chemical potentials, compositions and driving forces for growth of the intermediate gamma phase at the expense of the terminal alpha and beta phases. Note that the deviations from the local equilibrium compositions occur in the gamma phase and not in the terminal phases.

equilibrium with the intermediate phase, is negligible, as shown in Figure 11.7. Further, we will assume that in gamma the diffusivity of one component is much faster than that of the other component. Finally, we will assume that the molar volumes of the components are equal and independent of composition. These assumptions are primarily for the purpose of simplifying the algebra.

Conservation of B at the gamma/beta interface yields the following equation

$$(C^B_{\beta\gamma} - C^B_{\gamma\beta})(d\xi/dt)_{\beta\gamma} = D^B_{\gamma}(dC^B_{\gamma}/dy)_{\gamma\beta} \tag{11.1}$$

C = X/V, where X is the atom fraction and V the molar volume and we have already assumed that all three phases have the same molar volume. Similarly, conservation of B at the alpha/gamma interface yields

$$(C^B_{\gamma\alpha} - C^B_{\alpha\gamma})(d\xi/dt)_{\alpha\gamma} = -D^B_\gamma(dC^B_\gamma/dy)_{\alpha\gamma} \tag{11.2}$$

Let us assume that within gamma the gradient of composition of B is independent of distance.* Thus

$$(dC^B_\gamma/dy)_{\gamma\beta} = (dC^B_\gamma/dy)_{\alpha\gamma} = J_B/(-D^B_\gamma) = (C^B_{\gamma\beta} - C^B_{\gamma\alpha})/\xi_\gamma \tag{11.3}$$

which states that the flux of B is everywhere constant in gamma including at the interfaces.

But, if the reactions at the interfaces are homogeneous along the interface (i.e. the interface does not move via a step or ledge mechanism), then they can be taken to be proportional to the difference in free energy between reactants and products. Hence

$$(d\xi/dt)_{\beta\gamma} = L_{\beta\gamma}(\Delta g_{\beta\gamma}) \qquad -(d\xi/dt)_{\alpha\gamma} = L_{\alpha\gamma}(\Delta g_{\alpha\gamma})$$

The assumption we have made, that diffusion can be neglected in alpha and beta, makes the calculation of the difference in free energy between reactants and products particularly simple. We note that in a motion of the interface into either terminal phase, the constituents transported across unit area of interface correspond simply to the product of the composition of the terminal phase, at the interface, by the distance moved by the interface. Thus, for unit motion of the interface, the difference in free energy across the beta/gamma interface will be

$$\Delta g_{\beta\gamma} = (X^B_{\beta\gamma}/V)(\mu^B_{\beta\gamma} - \mu^B_{\gamma\beta}) + [(1 - X^B_{\beta\gamma})/V](\mu^A_{\beta\gamma} - \mu^A_{\gamma\beta})$$

Also, that across the alpha/gamma interface will be

$$\Delta g_{\alpha\gamma} = (X^B_{\alpha\gamma}/V)(\mu^B_{\alpha\gamma} - \mu^B_{\gamma\alpha}) + [(1 - X^B_{\alpha\gamma})/V](\mu^A_{\alpha\gamma} - \mu^A_{\gamma\alpha})$$

In both cases, we make the further reasonable assumptions: in the gamma phase the interface concentrations are not those corresponding to equilibrium, whereas in the terminal phases the interface compositions do correspond to the respective equilibrium concentrations; and the deviations of the chemical potentials from the equilibrium values at the interfaces are small enough that we need only expand to first order in the difference in interface compositions from the respective equilibrium values. Figure 11.7 illustrates the various terms that contribute to the driving forces for interface motion. It shows that small deviations from the equilibrium interface compositions in gamma can yield large positive driving forces for motion of the interfaces into the terminal phases, whereas with equilibrium interface compositions in gamma and off equilibrium composition at the interface in the terminal phases, the possible driving force for motion of the interfaces into the terminal phases are negative. Hence, for the approximations made the driving forces become

---

* This assumption is equivalent to the assumption that the diffusivity is independent of composition.

$$\Delta g_{\beta\gamma} = g^o_\beta - (1/V)(X^B_{\beta\gamma}\mu^B_{\gamma\beta} + X^A_{\beta\gamma}\mu^A_{\gamma\beta})$$

and

$$\Delta g_{\alpha\gamma} = g^o_\alpha - (1/V)(X^B_{\alpha\gamma}\mu^B_{\gamma\alpha} + X^A_{\alpha\gamma}\mu^A_{\gamma\alpha})$$

We now expand the chemical potentials in the gamma phase about their interface equilibrium values to obtain

$$\mu^B_{\gamma\beta} = \mu^B_{\gamma\beta,eq} + (\partial\mu^B_\gamma/\partial X^B_\gamma)(X^B_{\gamma\beta} - X^B_{\gamma\beta,eq}) + \cdots$$

$$\mu^A_{\gamma\alpha} = \mu^A_{\gamma\alpha,eq} + (\partial\mu^A_\gamma/\partial X^A_\gamma)(X^A_{\gamma\alpha} - X^A_{\gamma\alpha,eq}) + \cdots$$

Since

$$\mu^B_{\beta\gamma,eq} = \mu^B_{\gamma\beta,eq} \quad \mu^A_{\alpha\gamma,eq} = \mu^A_{\gamma\alpha,eq} \quad Vg^o_\beta \approx X^B_{\beta\gamma}\,\mu^B_{\beta\gamma,eq} + X^A_{\beta\gamma}\,\mu^A_{\beta\gamma,eq}.$$

and $$V g^o_\alpha \approx X^B_{\alpha\gamma}\,\mu^B_{\alpha\gamma,eq} + X^A_{\alpha\gamma}\,\mu^A_{\alpha\gamma,eq}$$

and by the Gibbs–Duhem relation $(\partial\mu^A/\partial X^A) = (X^B/X^A)\,(\partial\mu^B/\partial X^B)$ then, substitution yields

$$\begin{aligned}(d\xi/dt)_{\beta\gamma} &= L_{\beta\gamma}(\partial\mu^B_\gamma/\partial X^B_\gamma)\,[X^B_{\beta\gamma} - (X^A_{\beta\gamma}X^B_{\gamma\beta}/X^A_{\gamma\beta})](C^B_{\gamma\beta,eq} - C^B_{\gamma\beta}) \qquad (11.4)\\ &= M_{\beta\gamma}(C^B_{\gamma\beta,eq} - C^B_{\gamma\beta})\end{aligned}$$

$$(d\xi/dt)_{\alpha\gamma} = L_{\alpha\gamma}(\partial\mu^A_\gamma/\partial X^A_\gamma)\,[X^A_{\alpha\gamma} - (X^B_{\alpha\gamma}X^B_{\gamma\alpha}/X^B_{\gamma\alpha})](C^B_{\gamma\alpha,eq} - C^B_{\gamma\alpha}) \tag{11.5}$$

$$= M_{\alpha\gamma}(C^B_{\gamma\alpha,eq} - C^B_{\gamma\alpha})$$

Let us now solve for $C^B_{\gamma\beta}$ and $C^B_{\gamma\alpha}$ and substitute into equation (11.1)

$$C^B_{\gamma\beta} = C^B_{\gamma\beta,eq} - (d\xi/dt)_{\beta\gamma}/M_{\beta\gamma} \qquad C^B_{\gamma\alpha} = C^B_{\gamma\alpha,eq} - (d\xi/dt)_{\alpha\gamma}/M_{\alpha\gamma}$$

$$J_B = -(D_\gamma/\xi_\gamma)[C^B_{\gamma\beta,eq} - C^B_{\gamma\alpha,eq} - (d\xi/dt)_{\beta\gamma}/M_{\beta\gamma} + (d\xi/dt)_{\alpha\gamma}/M_{\alpha\gamma}]$$

Now replace $(d\xi/dt)_{\beta\gamma}$ from equation 1 and $(d\xi/dt)_{\alpha\gamma}$ from equation 2 to obtain

$$J_B = -(D_\gamma/\xi_\gamma)[\Delta C^B_{\gamma,eq} + J_B/\{(C^B_{\beta\gamma} - C^B_{\gamma\beta})M_{\beta\gamma}\} + J_B/\{(C^B_{\gamma\alpha} - C^B_{\alpha\gamma})M_{\alpha\gamma}\}]$$

Solving for J_B yields

$$J_B = -\Delta C^B_{\gamma,eq}/[(\xi_\gamma/D^B_\gamma) + 1/\{(C^B_{\beta\gamma} - C^B_{\gamma\beta})M_{\beta\gamma}\} + 1/\{(C^B_{\gamma\alpha} - C^B_{\alpha\gamma})M_{\alpha\gamma}\}]$$

We now relate the growth rate of the intermediate phase gamma to J_B

$$(d\xi/dt)_\gamma = (d\xi/dt)_{\beta\gamma} - (d\xi/dt)_{\alpha\gamma} = -J_B/(C^B_{\beta\gamma} - C^B_{\gamma\beta}) - J_B/(C^B_{\gamma\alpha} - C^B_{\alpha\gamma})$$

$$= -J_B\{1/(C^B_{\beta\gamma} - C^B_{\gamma\beta}) + 1/(C^B_{\gamma\alpha} - C^B_{\alpha\gamma})\}$$

Therefore

$$(d\xi/dt)_\gamma = \frac{\Delta C^B_{\gamma,eq}\{1/(C^B_{\beta\gamma} - C^B_{\gamma\beta}) + 1/(C^B_{\gamma\alpha} - C^B_{\alpha\gamma})\}}{(\xi_\gamma/D^B_\gamma) + 1/\{(C^B_{\beta\gamma} - C^B_{\gamma\beta})M_{\beta\gamma}\} + 1/\{(C^B_{\gamma\alpha} - C^B_{\alpha\gamma})M_{\alpha\gamma}\}} \tag{11.6}$$

Let us now examine the behavior of equation 11.6. When the thickness of the gamma phase ξ_γ is small near the start of the growth process, then the last two terms in the denominator can exceed the term containing the thickness ξ_γ. One of these two remaining terms usually will be much larger than the other, but in any case in this regime of behavior it is apparent that the growth rate will be nearly independent of the time. The slowest process in this regime is obviously that for interface migration and, hence, the regime is called interface or reaction controlled. When ξ_γ becomes larger than a critical value given by

$$\xi_\gamma^* = D_\gamma^B[1/\{(C_{\beta\gamma}^B - C_{\gamma\beta}^B)M_{\beta\gamma}\} + 1/\{(C_{\gamma\alpha}^B - C_{\alpha\gamma}^B)M_{\alpha\gamma}\}]$$

then the growth rate of the gamma phase depends upon the thickness of the gamma phase. Because the term containing the thickness is the term depending upon diffusion of the most rapid species in the gamma phase, this regime of behavior, in which the thickness depends parabolically on time, is termed diffusion controlled. These two regimes are illustrated in Figure 11.8. The literature abounds with examples of reaction controlled growth of intermediate phases. For example, SiO_2 first grows linearly then parabolically with time in the oxidation of silicon.* Many thin film silicides grown by reaction between a thin metal film and a silicon substrate exhibit the same behavior. Indeed, this mode of growth is common in non-metallic systems.

It is easy to show that for the case of diffusion controlled growth of the intermediate phase A_nB_m in the type of system considered above, that the rate constant k in the relation $\xi^2 = 2kt$ is nearly given by

$$k = (n + m)^2(L_A/n^2 + L_B/m^2)\Delta g.$$

(It is merely necessary to show that for the case where the terminal solutions are dilute at the phase boundaries and for the intermediate phase A_nB_m, that the irreversible free energy dissipation rate for the diffusion process in the intermediate phase is given by

$$\Delta g(d\xi/dt) = \int_o^\xi \Sigma J_i X_i \, dx = L_B(\mu_{\beta\gamma}^B - \mu_{\gamma\alpha}^B)^2/\xi + L_A(\mu_{\beta\gamma}^A - \mu_{\gamma\alpha}^A)^2/\xi$$

---

* Doremus (J. Appl. Phys. 66, 4441(1989)) has argued that the linear regime in the growth of SiO_2 is not a manifestation of reaction controlled growth.

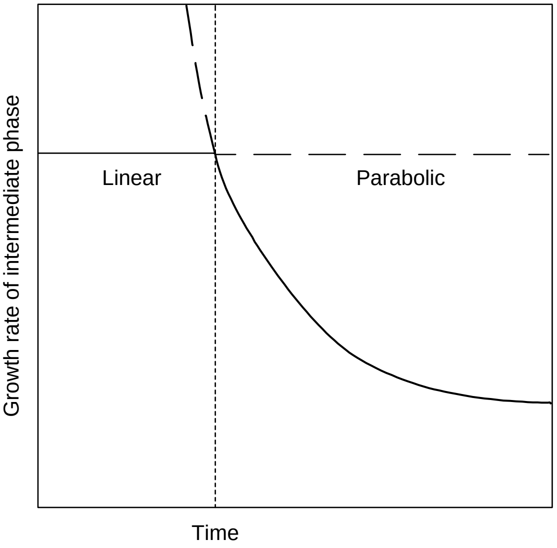


Figure 11.8. Illustration of the behavior of equation 11.6 in this chapter. The initial stage yields a linear dependence of the intermediate phase thickness ξ as a function of the time, whereas the second stage yields a parabolic dependence of ξ on time. Notice that the two processes are coupled in series and hence that the slower controls the observed rate of growth.

where the free energy and chemical potentials are defined in Figure 11.7, ξ is the thickness of the intermediate phase and L_i are the phenomenological coefficients in the relations between flux and force, i.e. $J_i = L_iX_i$ and $VX_i = \nabla\mu$. Now by the assumption that the terminal phases have only slight solid solubility, $\mu^B_{\alpha\beta} \approx \mu^B_{\beta\gamma}$ and $\mu^A_{\alpha\beta} \approx \mu^A_{\alpha\gamma}$. Hence, by Figure 11.7 and similar triangles $(\mu^B_{\beta\gamma} - \mu^B_{\gamma\alpha})m/(n + m) = V\Delta g$ and $(\mu^A_{\beta\gamma} - \mu^A_{\gamma\alpha})n/(n + m) = V\Delta g$. Substituting above and integrating then

$$\xi^2 = 2V\,\Delta g(n + m)^2[L_A/n^2 + L_B/m^2]t$$

A more extended analysis of the growth rate for the intermediate phase under *diffusion controlled conditions* has been given by Kidson.[2] See also the articles in the Bibliography for additional treatments of diffusion controlled intermediate phase growth or what is sometimes called “reaction diffusion”.

2.2. Many possible intermediate phases

Of course, when many intermediate phases are stable at the diffusion temperature then a description of the kinetics is more complicated than given in the previous section.[2] In semi-infinite bulk diffusion couples all the intermediate phases will ultimately form and grow. However, in thin film/bulk couples, as in the formation of silicides by reaction between a deposited thin metal film and a silicon substrate, it is usually possible for only one intermediate phase to form during the time period it takes to exhaust the thin film source. A further reaction between this intermediate phase and the bulk substrate can occur, but, generally, the kinetics involved in the various possible phase transformations vary so greatly that it is necessary to raise the temperature to initiate a new reaction in a reasonable time period.

One question that arises in a consideration of the thin film/bulk couples is what condition governs the choice of the intermediate phase that appears first on increasing the diffusion temperature. One "answer" is that the phase that appears first as the diffusion temperature is raised is the one that consumes the thin film material the fastest. The reader may recall our discussion of competing rate processes in Chapter VII. However, this "answer" does not allow us to predict which possible phase will appear first, without a detailed knowledge of the various diffusivities, nucleation kinetics, and reaction rates for all the possible reactions. Nevertheless, it is possible to investigate some of the factors that affect the relative kinetics as follows.

We can ask the question, "Suppose that somehow two intermediate phases formed with planar interfaces between two terminal phases, one a thin film and the other a substrate, that composed the original diffusion couple. What are the conditions that govern whether these two phases can grow or one will shrink at the expense of the other's growth?" To answer this question, under the assumption that diffusion controlled conditions apply to both intermediate phases, we set up the equations for conservation of matter at the interfaces and for diffusion through the intermediate phases. We assume, for simplicity, that the solubilities for the corresponding solute in the terminal phases are negligible, that the diffusivity of one species overwhelms that for the other species in both intermediate phases ($D_2 >> D_1$), and that the solubility ranges of the intermediate phases are very small. Figure 11.9 illustrates both the diffusion couple and the free energy composition diagram for this system and defines the various quantities to be used. We focus on each interface and set down the matter conservation relations that act to define the velocities of the interfaces. Thus, for the interface between the film and the phase α we can write with C, corresponding to the concentration per unit volume of component 2, and μ, the chemical potential of this component

$$(c^M - c^\alpha)v_1 = -(c^\alpha/RT)(\mu^{\alpha M} - \mu^{\alpha\beta})D^\alpha/h_\alpha$$

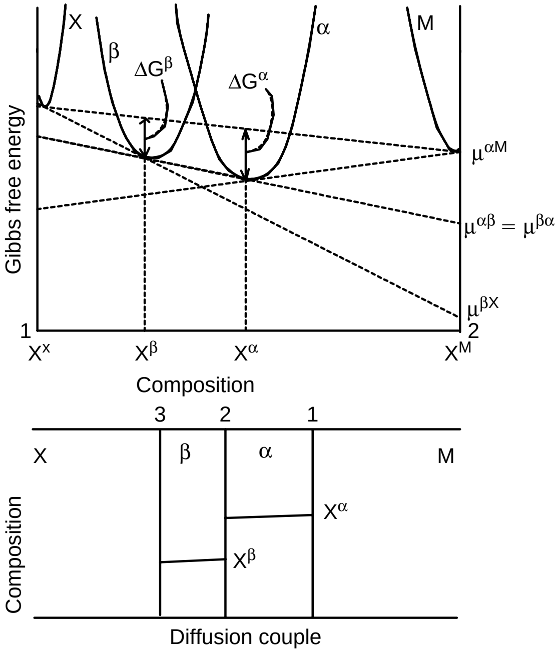

X
β
ΔG^β
ΔG^α
α
M
Gibbs free energy
μ^αM
μ^αβ = μ^βα
μ^βX
1
2
X^X
X^β
X^α
X^M
Composition
3
2
1
X
β
α
M
Composition
X^α
X^β
Diffusion couple

Figure 11.9. Defining chemical potentials and average compositions of phases for analysis of relative growth or disappearance of intermediate phases.

Similarly, for the interface between the substrate and phase β

$$(c^\beta - c^X)v_3 = (c^\beta/RT)(\mu^{\beta\alpha} - \mu^{\beta X})D^\beta/h_\beta$$

and for that between phases α and β

$$\begin{aligned}(c^\alpha - c^\beta)v_2 &= (c^\alpha/RT)(\mu^{\alpha\beta} - \mu^{\alpha\beta})D^\alpha/h_\alpha - (c^\beta/RT)(\mu^{\beta\alpha} - \mu^{\beta X})D^\beta/h_\beta \\ &= -(c^M - c^\alpha)v_1 - (c^\beta - c^X)v_3\end{aligned}$$

Now, the criterion for growth of any intermediate phase is $v_{n+1} - v_n > 0$. Thus, to determine whether the phase alpha grows or shrinks we need to evaluate whether $(v_{n+1} - v_n)$ is greater or smaller than zero, respectively. Substituting and collecting terms we arrive at the relations governing growth or shrinkage given by

$$dh_\alpha/dt = A/h_\alpha - B/h_\beta$$

$$dh_\beta/dt = C/h_\beta - D/h_\alpha$$

where

$$A = D^{\alpha}(C^{\alpha}/RT)(C^{M} - C^{\beta})\frac{(1 - X^{\beta})\Delta G^{\alpha} - (1 - X^{\alpha})\Delta G^{\beta}}{(C^{\alpha} - C^{\beta})(C^{M} - C^{\beta})(X^{\alpha} - X^{\beta})}$$

$$B = D^{\beta}(C^{\beta}/RT)\frac{(1 - X^{\beta})(X^{\alpha}\Delta G^{\beta} - X^{\beta}\Delta G^{\alpha})}{(C^{\alpha} - C^{\beta})X^{\beta}(X^{\alpha} - X^{\beta})}$$

$$C = D^{\beta}(C^{\beta}/RT)\frac{(1 - X^{\beta})[X^{\alpha}\Delta G^{\beta}/X^{\beta} - \Delta G^{\alpha}](C^{\alpha} - C^{X})}{(C^{\beta} - C^{X})(X^{\alpha} - X^{\beta})(C^{\alpha} - C^{\beta})}$$

and

$$D = D^{\alpha}(C^{\alpha}/RT)\frac{(1 - X^{\beta})\Delta G^{\alpha} - (1 - X^{\alpha})\Delta G^{\beta}}{(C^{\alpha} - C^{\beta})(X^{\alpha} - X^{\beta})}$$

Thus, depending upon the values of the constants, these equations can lead to either growth of both intermediate phases, or to the growth of one and the shrinkage of the other. Figure 11.10 describes the possible behavior in the h_a, h_b space.

For silicides we simply do not have enough information to evaluate A, B, C and D. Nor do we have enough data to determine whether the analysis provides

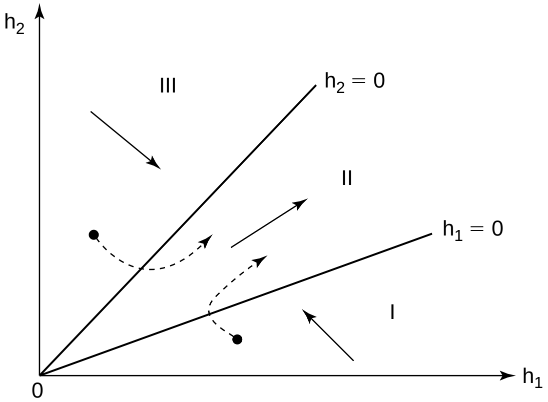


Figure 11.10. In domain I phase α regresses and phase β grows. In domain II phase α grows and phase β grows. In domain III phase α grows and phase β regresses.

an answer to the question concerning the sequential nature of appearance of the intermediate phases in thin film couples.

In all of the above we have implicitly, or explicitly, assumed that the interfaces were incoherent so that vacancies and interstitials could be created or destroyed at the interface. If the interface is coherent or partially coherent then we must be concerned with this fact. A divergence of vacancy (interstitial) flux at the interfaces will develop stresses or voids or generate climb type dislocations.

2.3. Amorphous intermediate phase

One of the recent interesting discoveries is that metastable amorphous phases can form as intermediate phases in some diffusion couples.[3] It is noted that the systems that exhibit this behavior are characterized by having large negative free energies of formation for the stable intermediate phases and an appreciable difference in atomic radii between the components. In this case, it is conceivable that an amorphous phase at the intermediate composition will be stable relative to the stable mixture of the terminal phases, although it will be metastable relative to the stable intermediate phase. Thus, following the expectation that the fastest reaction between the two terminal phases will be the one to appear first and the possibility that diffusion through the amorphous phase can be rapid for the species with least atomic radius, if its atoms are small enough to diffuse through the interstitial space in the amorphous phase, then the observed reaction of two crystalline terminal phases to yield an intermediate amorphous phase is at least plausible. Indeed, it is known that one component, the smaller one, diffuses anomalously fast through the amorphous phase. Also, it is known that nucleation barriers to the formation of the amorphous phase (i.e. a monocrystalline thin film terminal phase would produce such a nucleation barrier) slow down the reaction sufficiently that the amorphous phase does not form in such systems whereas it does form when the reactant thin films are polycrystalline.

3. Growth involving interface migration

3.1. Continuous precipitation

3.1.1. Early stage of growth

3.1.1.1. Analytic evaluation

Another mode of growth that involves both diffusion and a chemical reaction in the kinetics of formation of the growing phase is that of precipitation from a supersaturated solution. An example of a process that uses this mode of synthesis to produce a

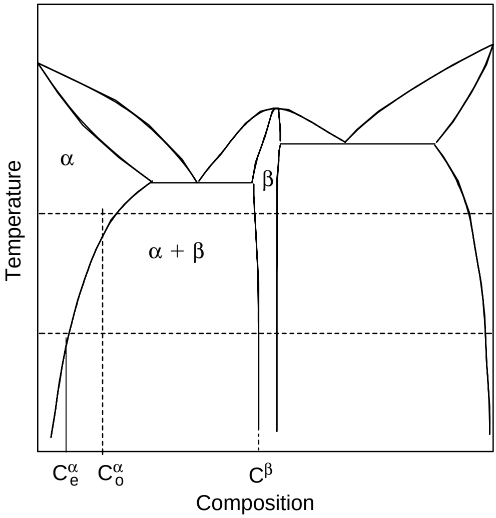


Figure 11.11. Phase diagram illustrating the compositions for local equilibrium at the interface and that in the supersaturated solution.

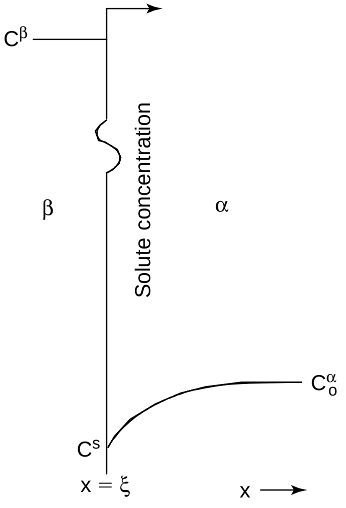


Figure 11.12a. Showing the interface between the precipitate β and the supersaturated α phase and the concentration profile in the latter.

desired material is liquid phase epitaxy. This process also occurs in the precipitation of phases in the aging of supersaturated alloy solid solutions to produce distributions of precipitates in alloy systems. As shown in Figure 11.11 the supersaturated solution is generally produced by cooling a solution from a temperature at which it is stable to another at which it is supersaturated. We will assume that no diffusion occurs in the precipitated phase, beta, but does occur in the parent phase alpha and that the composition of beta is that in equilibrium with alpha. Also, we will consider the case of growth by the motion of a planar interface separating the two phases. Hence, the plot of concentration of the solute species against distance is schematically as illustrated in Figure 11.12a. We will also assume that the solute species is the more mobile of the two species and that the temperature distribution does not affect the kinetics.* Further, we assume constant molar volume to simplify the relations because we wish merely to distinguish between diffusion and reaction control in this example.

The change in free energy due to the transport of matter across the interphase interface per unit area of interface per unit motion of interface is given by

$$\Delta F_{int} = v_m[\{c^{\beta}\mu_B^{\alpha}(c^S) + (c - c^{\beta})\mu_A^{\alpha}(c^S)\} - \{c^{\beta}\mu_B^{\beta}(c^{\beta}) + (c - c^{\beta})\mu_A^{\beta}(c^{\beta})\}] \tag{11.7}$$

where v_m is the molar volume, the concentrations are defined in Figures 11.12a and 11.12b, $c = (v_m)^{-1}$ and $\mu_i^{\varphi}(c)$ are the chemical potentials of component i in phase φ at composition c. In Figure 11.12b c^{β} is chosen to maximize ΔF_{int}.

Sometimes equation 11.7 is written as

$$\Delta F_{int} = v_m[\{c^{\beta}\{\mu_B^{\alpha}(c^S) - \mu_B^{\beta}(c^{\beta})\} + (c - c^{\beta})\{\mu_A^{\alpha}(c^S) - \mu_A^{\beta}(c^{\beta})\}] \tag{11.8}$$

or when at a maximum then $\Delta F_{int} = v_m\{\mu_B^{\alpha}(c^S) - \mu_B^{\beta}(c^{\beta})\} = v_m\{\mu_A^{\alpha}(c^S) - \mu_A^{\beta}(c^{\beta})\}$.

We shall assume linear kinetics (the interface is rough and incoherent and growth does not involve lateral motion of ledges), i.e.

$$d\xi/dt = M_{\alpha\beta}\,\Delta F_{int} \tag{11.9}$$

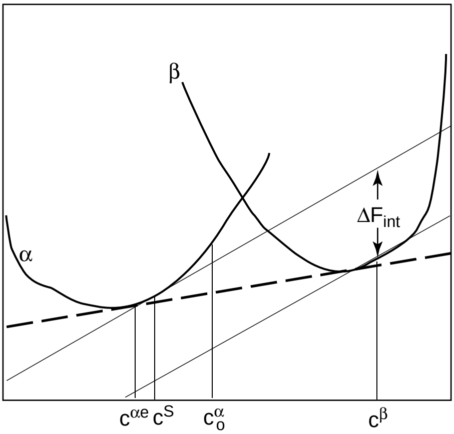


Figure 11.12b. Free energy–composition diagram defining the compositions and free energies.

---

* The situation where only one component diffuses in a multicomponent system is denoted paraequilibrium in the literature. A typical example is that of C in Fe–C.

Because it has been assumed that only the solute component is diffusing, conservation of this component at the interphase interface yields

$$(c^{\beta} - c^{S})(d\xi/dt) = D^{\alpha}(\partial c^{\alpha}/\partial x)_{x=\xi} \qquad (11.10)$$

Now, diffusion in the supersaturated phase alpha must satisfy Fick's second law and the boundary conditions or

$$\partial c^{\alpha}/\partial t = D^{\alpha}\, \partial^2 c^{\alpha}/\partial x^2 \qquad (11.11)$$

with the boundary conditions being

$$c^{\alpha} = c^{\alpha}_{o} \quad \text{at } x = \infty$$

$$(\partial c^{\alpha}/\partial x)_{x=\xi} = 0$$

and

$$c^{\alpha} = c^{S} \quad \text{at } x = \xi \text{ for all } t > 0. \qquad (11.12)$$

Equations 11.11 and 11.12 are satisfied by

$$c^{\alpha} = A + B\, \mathrm{erfc}\{x/[2(D^{\alpha}t)^{1/2}]\}$$

Substitution of this solution in equation 11.12 yields

$$c^{\alpha}_{o} = A + B\, \mathrm{erfc}(\infty) = A + B \quad c^{S} = A + B\, \mathrm{erfc}\{\xi/[2(D^{\alpha}t)^{1/2}]\}$$

Hence $c^{S} = c^{\alpha}{}_{o} - B + B\, \mathrm{erfc}\{\xi/[2(D^{\alpha}t)^{1/2}]\}$

or $c^{\alpha}_{o} - c^{S} = B(1 - \mathrm{erfc}\{\xi/[2(D^{\alpha}t)^{1/2}]\}$

But $\partial c^{\alpha}/\partial x = B \exp(-x^2/4D^{\alpha}t)/[\pi\,(D^{\alpha}\,t)]^{1/2}$

(because $(\partial/\partial x)(\int_{o}^{u} \exp(-\eta^2)d\eta) = (du/dx)\exp(-x^2/4D^{\alpha}t)$, where $u = x/[2(D^{\alpha}t)^{1/2}]$). Substitution for B yields

$$\partial c^{\alpha}/\partial x = \exp(-x^2/4D^{\alpha}t)\frac{(c^{\alpha}_{o} - c^{S})/[\pi(D^{\alpha}t)]^{1/2}}{(1 - \mathrm{erfc}\{\xi/[2(D^{\alpha}t)^{1/2}]\})}$$

We now substitute this relation into equation 11.10 to obtain

$$(c^{\beta} - c^{S})(d\xi/dt) = D^{\alpha}\exp(-\xi^{2}/4D^{\alpha}t)\frac{(c_{o}^{\alpha} - c^{S})/[\pi(D^{\alpha}t)]^{1/2}}{(1 - \text{erfc}\{\xi/[2(D^{\alpha}t)^{1/2}]\})} \tag{11.13}$$

Now, let $\xi = j(4D^{\alpha}t)^{1/2}$. Hence, $d\xi/dt = (j/2)(D^{\alpha}/t)^{1/2}$. We substitute for $d\xi/dt$ in equation 11.13 to obtain the relation for j, as follows

$$(c^{\beta} - c^{S})j = 2(c_{o}^{\alpha} - c^{S})\exp(-j^{2})/\left[\sqrt{\pi}(1 - \text{erfc}(j))\right] \tag{11.14}$$

For $j << 1$, this relation can be approximated by

$$j = 2(c_{o}^{\alpha} - c^{S})/\sqrt{\pi}(c^{\beta} - c^{S}) \tag{11.15}$$

We equate the expressions for $d\xi/dt$ in equations 11.9 and 11.10 and substitute for j in $(\partial c^{\alpha}/\partial x)_{x=\xi}$ and for the latter in equation 11.10 to obtain the following relation for small values of j

$$d\xi/dt = M_{\alpha\beta}\,\Delta_{Fint} = D^{\alpha}(c_{o}^{\alpha} - c^{S})^{2}/[(c^{\beta} - c^{S})^{2}\xi] \tag{11.16}$$

It is apparent from Figure 11.12b that ΔF_{int} monotonously increases with $(c^{S} - c^{\alpha e})$. To a first approximation they are proportional at least for small values of $(c^{S} - c^{\alpha e})$. Thus, at the start of growth with ξ a value close to zero, since the only other variable in equation 11.16 is c^{S} then $(c^{S} - c_{o}^{\alpha})$ is also close to zero. As growth proceeds then this difference in concentration must also increase and hence $c^{S} \rightarrow c^{\alpha e}$. Hence, diffusion control holds at large values of the thickness ξ, whereas for small values of the thickness ξ interface reaction control holds. A similar result was obtained in Section 2.1.1 for the mode of growth of an intermediate phase between two terminal solid solutions. Interface reaction control is more common in non-metallic systems than in alloys.

3.1.1.2. Computer simulations

Phase field simulation provides a means of obtaining a solution to the free boundary problem involved in the growth of stable phases from a metastable supersaturated parent phase. Essentially, the same equations given in the above sub-section apply in the phase field treatment. In particular, equation 11.9 corresponds to the equation in the phase field method derived from the Ginzburg–Landau equation for the non-conserved order field

$$\partial\varphi/\partial t = -L\,\partial F/\partial\varphi$$

which for the reaction being considered can be written as

$$V = M\ \Delta F_{int}$$

For the liquid–solid transformation the applicable free energy diagram is shown in Figure 11.12c with the compositions corresponding to those defined above. The phase field simulation for a constant velocity (constant ΔF_{int} and consequently c^S) is shown in Figure 11.12d. As shown, the value of c^S is constant after the velocity itself is constant (beyond the first profile).

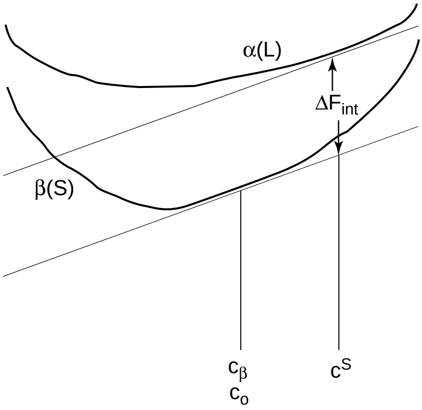


Figure 11.12c. Free energy–composition diagrams for liquid and solid phases showing compositions corresponding to Figure 11.12d.

Figures 11.12c and d also apply to the case of a massive transformation, where α is the parent phase and β is the product phase and diffusion occurs only in the α phase. For diffusion in both phases the traveling wave yielded by α phase field simulation is shown in Figure 11.12e.

In the event that the composition of the alloy and temperature of transition is in the two-phase field at equilibrium then the free energy–composition diagram is as shown in Figure 11.12f and for decomposition of the parent α phase at alloy composition c_o the equations given in the previous sub-section hold. As shown by equation 11.16, the thickness of the product β phase should increase as the square root of time. The phase field simulation result for this situation is shown in Figure 11.12g. As shown, the thickness varies as the square root of time at the early stage of the decomposition process. The slowing down of the process is due to increase in composition of the remaining α phase due to its finite size and consequent impingement, as indicated in Figure 11.12h, a result of the phase field simulation.

Equation 11.16 is not a complete solution of the problem of defining the motion of the planar interface between parent and product phases, since there are two variables in the equation: ξ and c^S. However, there is another relation that has not been used, which is that the total available free energy consists of two terms in this problem: ΔF_{int} and that dissipated in diffusion in the α (parent) phase. The total free energy available is as indicated in Figure 11.12f, by the vertical broad bar. The phase field simulation takes this relation into account and thus defines c^S, which we have not done analytically. This relation for this problem involving the diffusion of only one component is given by

$$\Delta F_{diff} = v_m \int JX\, dx$$

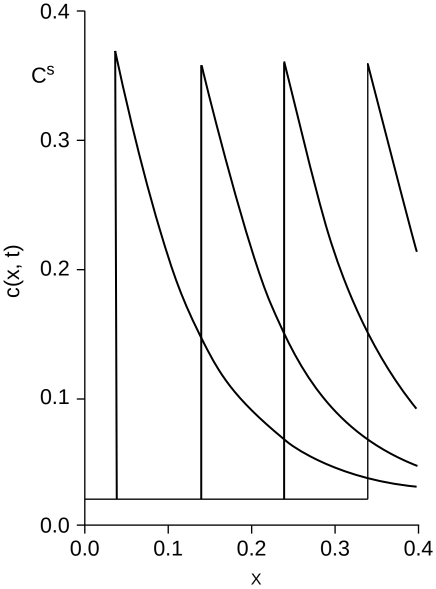


Figure 11.12d. Concentration profiles at succeeding times of constant intervals provided by a phase field simulation of solidification along a planar front. From Phys. Rev. A 45, 7424(1992) with permission.

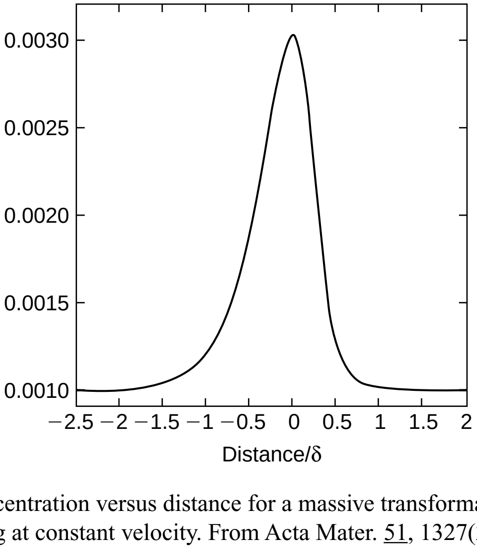


Figure 11.12e. Concentration versus distance for a massive transformation with α/β interface moving at constant velocity. From Acta Mater. 51, 1327(2003) with permission. © 2003 Elsevier.

where $X = \partial\mu/\partial x = (\partial\mu/\partial c)(\partial c/\partial x)$ and J is the flux of the diffusing component at position x and is defined in terms of X by equation 8.1.

The situation corresponding to Figure 11.12c involving precipitation of cementite in the Fe–C system has been treated using the DICTRA software.[4]

3.1.2. Later stage-Ostwald ripening

Decomposition of a supersaturated solution via the nucleation and growth of precipitates of the product phase yields, at the point where the average concentration in the parent solution is close to equilibrium, a distribution of precipitates that vary in size. We have not treated the dependence during diffusion limited growth of the radius of the precipitate, if it is a sphere, on time. This problem and that of the growth when the diffusion fields of neighboring particles overlap is of interest, but these problems are problems in the solution of the diffusion equations for various boundary conditions, which we will not consider here. In this section, we will treat the interface energy driven growth of large particles at the expense of small particles – the problem known as Ostwald ripening.

Ostwald ripening involves both diffusion and interface reaction and we shall give solutions corresponding to control by one or the other process. For diffusion control the driving force for the reaction is given by the difference in chemical potentials corresponding to equilibrium with particles of radius r and r′. Since the chemical potential $\mu = 2\sigma V/r$ and $\mu' = 2\sigma V/r'$, we can express the driving force as $2\sigma V(1/r - 1/r')$. For simplicity, let us assume that the solution is dilute so that the difference in chemical potentials equals $RT(C - C')/C_{eq}$, where C is the concentration of the solute species in equilibrium with the particle of radius r, etc., and C_{eq} is that for an infinite radius (a planar surface).

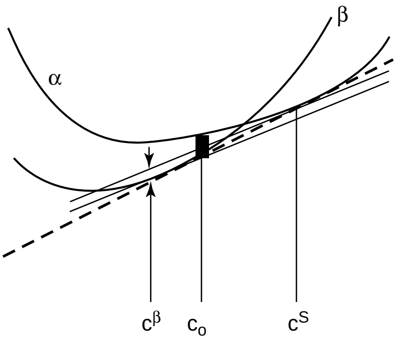


Figure 11.12f. Schematic free energy-composition diagram applicable to decomposition of α phase at composition c_o with interface compositions of c^β on β side and c^S on α side of interface. Note that free energy driving interface motion ΔF_{int}, the vertical interval between arrows, is smaller than the total free energy available for the reaction at c_o, the thick vertical bar.

A solution to the diffusion problem[5,6] involves an analysis of how the initial distribution function of particle radii varies with time, if at all. In the case that it is valid to assume that the distribution function does not depend on time and that transport is via volume diffusion then the solution to the problem yields for the average particle radius <r>

$$\langle r\rangle = [8\sigma V^2 C_{eq} D/(9RT)]^{1/3} t^{1/3}$$

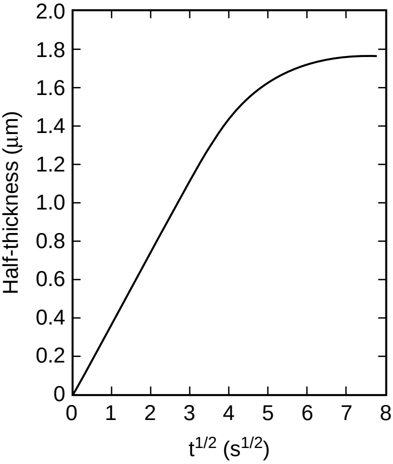


Figure 11.12g. Phase field simulation result showing thickness of product plate versus the square root of time. From Acta Mater. 51, 1327(2003) with permission. © 2003 Elsevier.

It is possible to illustrate the derivation of this result using an oversimplified model. If the distribution function of particle radii is independent of time then we can focus on any particle and consider its radius as a function of time. Accordingly, consider the flux between two neighboring particles of radii r and r′. Now the driving force for diffusion between these particles is the chemical potential gradient between them for the solute. We have already shown that this gradient will approximately equal $2\sigma V(r' - r)/(rr'L)$, where L is the distance between the particles. Thus, the flux at the surface of the larger particle will be given by $J = D(C_{eq}/RT)[2\sigma V(r' - r)/(rr'L)]$. Roughly $rr' = \langle r\rangle^2$. Also, the rate of change of the average radius will equal that of the larger particle by the time independence of the particle distribution curve. Thus, we can write

$$\begin{aligned}(d\langle r\rangle/dt)/V &= J = D(C_{eq}/RT)[2\sigma V(r' - r)/(rr'L)]\\ &= [2\sigma VD(C_{eq}/RT)(r' - r)/L]/\langle r\rangle^2\end{aligned}$$

This equation integrates to give an equation having the same time dependence for r as the rigorous one given before, although the coefficient differs. (The difference $(r' - r)$ may be considered to be independent of $\langle r\rangle$ since the distribution function is independent of time. It represents an average driving force, independent of time. Also, the average diffusion distance between particles is independent of time. Hence, $(r' - r)/L$ is some number.)

Experiment is not quite in accord with this result. Nevertheless, it is worthwhile to generalize this result for the case that either transport occurs via other paths or that Ostwald ripening is reaction controlled. The general result is

$$d^n = d_o^n + Kt$$

where the values of n for the various rate controlling processes are listed below:

$n = 1$ for viscous flow as the rate controlling process
$n = 2$ for interfacial reaction control
$n = 3$ for volume diffusion in all phases
$n = 4$ for interfacial diffusion as the rate controlling process
$n = 5$ for dislocation pipe diffusion control.

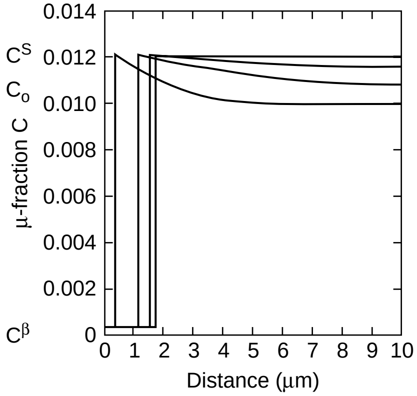


Figure 11.12h. Composition profile in α and β phases for constant time intervals. From Acta Mater. 51, 1327(2003) with permission. © 2003 Elsevier.

The values of K vary with the process; however, for all diffusion controlled processes K is proportional to the product of σ, D and C, as in the above equation.

Ostwald ripening is of technological importance in the development of high temperature resistant superalloys, which depend upon stable particles to provide time independent strength. Thus, it is not a surprise that the interfaces between the particles in nickel based superalloys can be tailored to have negligible interface energy.

This area is still an active one for research despite much attention in the past (see Bibliography). In particular the main disagreement has to do with the form of the particle distribution function after Ostwald ripening. It is expected that this problem will succumb to the computational expertise now being devoted to it. One phase field study of Ostwald ripening[7] verifies the $n = 3$ growth although the kinetic coefficient (K) depends upon the volume fraction of precipitate.*

3.1.3. Coarsening with convection in mushy zone

The phenomenon of solidification from a mushy liquid, in which atom transport through the liquid in the mushy zone occurs by diffusion and convection, is of high technological interest. Although it is possible to develop analytic relations that approximate the phenomena that take place in such a complex situation the complexity of the phenomena usually do not succumb to the simplicity of the analytic approximations. Hence, simulation of the phenomena is a reasonable alternative more likely to describe their complex behavior. At this writing one phase field simulation of solidification in a mushy zone[8] affected both by diffusion and convection has been carried out and now will be described briefly.

The equations used in the phase field simulation are: the evolution equation for the phase field φ , the species conservation equations, and the mass and momentum conservation equations. In this simulation a stationary front profile of the form $\varphi(x) = 1/2[1 - \tanh(x/2\delta)]$ is assumed where the thickness of the diffuse interface is 6δ. The evolution equation for the phase field becomes then

---

* See Section 1 in Chapter XIII.

$$\partial\varphi/\partial t = \mu\Gamma(\nabla^2\varphi - \varphi(1-\varphi)(1-2\varphi)/\delta^2) + (\mu/\delta)\varphi(1-\varphi)(T - T_m - m_l C_l)$$

where $\Gamma = \sigma/\Delta S_f$ and σ is the surface tension and ΔS_f is the entropy of fusion, m_l is the liquidus slope from the equilibrium phase diagram, and C_l is the solute concentration in the liquid. For derivation of these equations see References 9, 10.

The species concentration equation is written in terms of a phase mixture as follows

$$C = (1-\varphi)C_l + \varphi C_s$$

Equilibrium partitioning in the diffuse interface is assumed so that $C_s = kC_l$. so that the solute balance is given by

$$\begin{aligned}&\partial C/\partial t + \nabla\cdot\{[(1-\varphi)v_l C]/(1-\varphi+k\varphi)\}\\ &\quad = \nabla\cdot D\{\nabla C + [(1-k)C/(1-\varphi+k\varphi)]\nabla\varphi\}\end{aligned}$$

where $D = [(1-\varphi)D_l + k\varphi D_s]/(1-\varphi+k\varphi)$ where D_s and D_l are the mass diffusivities in the solid and liquid and v_l is the intrinsic velocity of the liquid. Assuming equal densities in the solid and liquid phases and a stationary solid the mass and momentum conservation relations are

$$\nabla\cdot(1-\varphi)v_l = 0$$

and

$$\begin{aligned}&\partial/\partial t[(1-\varphi)v_l] + \nabla\cdot[v_l(1-\varphi)v_l]\\ &\quad = -\rho^{-1}(1-\varphi)\nabla p_l + \nu\cdot\nabla^2[(1-\varphi)v_l] - h\nu\varphi^2(1-\varphi)v_l/\delta^2\end{aligned}$$

where p_l is the pressure, ν is the kinematic viscosity of the liquid, and ρ is the density of each of solid and liquid phases. The constant h is chosen so that for plane flow past the interface the velocity profile outside the smeared interface matches the one for a sharp interface with a no-slip condition at $\varphi = 0.5$.

Appropriately setting initial conditions, boundary conditions and using now standard procedure for obtaining numerical solutions to the transport equations the phase field simulations yielded results, some of which are described below.

Representative 2D simulations with and without flow in the mushy zone are given in Figure 11.13. The top four snapshots were made in the absence of flow. The bottom four snapshots show the results of coarsening in the presence of

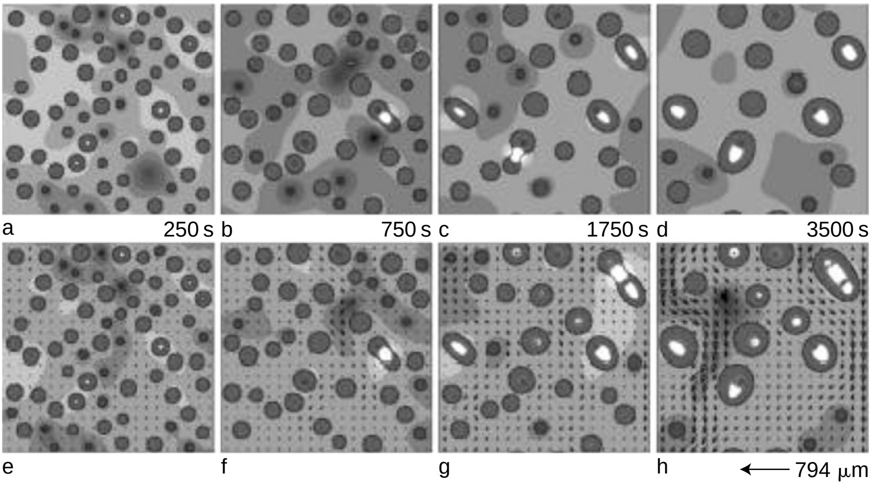


Figure 11.13. Snapshots showing concentrations (gray scale), phases and liquid velocities (arrows in bottom row) in absence of convection (top row) and its presence (bottom row) as a function of time. From Acta Mater. 47, 3663(1999) with permission. © 1999 Elsevier Science Ltd.

convection of the liquid. Concentrations are in six gradations of the gray scale for the liquid phase and four for the solid phase. Also shown are arrows in the liquid phase, indicating the liquid flow velocity, and the two phases. It is apparent that coarsening kinetics differ greatly for solidification in the absence and presence of convective flow in the mushy zone. In the absence of flow the coarsening kinetics were found to obey that expected for LSW Ostwald ripening, i.e. $R^3 - R^3_o = Kt$. In the presence of flow the exponent found was 2 instead of 3. Although this behavior was predicted by the model of mushy coarsening in a convective environment of Ratke and Thieringer[11] the dependence of the kinetic constant on volume fraction of solid phase found in this simulation differed greatly from that of the latter, see Figure 11.14a. A similar dependence of kinetic constant in the absence of convection is found in agreement with previous studies, see Figure 11.14b. In this figure the line corresponds to a theoretical derivation by Marqusee[12] and the circles to results of numerical simulation by Aikawa and Meiron[13]. The triangles are the results of Diepers et al.[8] from which paper the figure is taken.

3.2. Diffusion-induced grain boundary migration

At temperatures where lattice diffusion is negligible (i.e. diffusion penetration distance is less than a few Angstroms) it is found that solute diffusion along grain

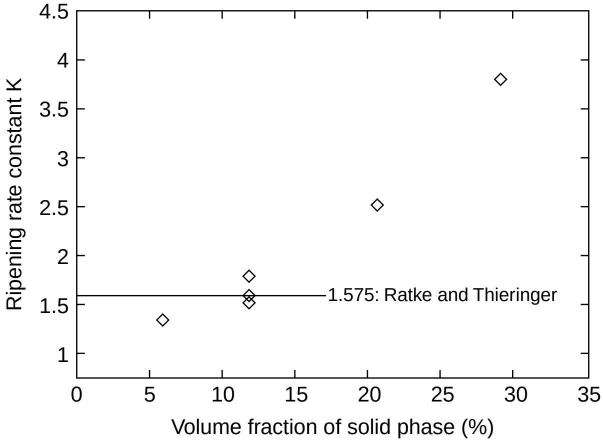


Figure 11.14a. Ripening rate constant in the presence of convection versus volume fraction of solid phase. Points are values deduced from phase field simulation. From Acta Mater. 47, 3663(1999) with permission. © 1999 Elsevier Science Ltd.

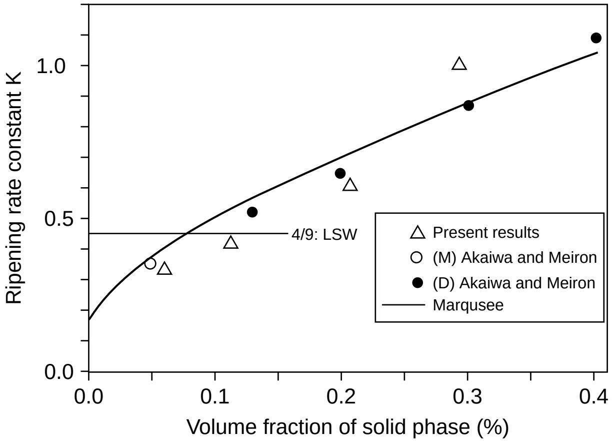


Figure 11.14b. Ripening rate constant in the absence of convection versus volume fraction of solid phase. From Acta Mater. 47, 3663(1999) with permission. © 1999 Elsevier Science Ltd.

boundaries can induce grain boundary migration.[14–18] As the boundary moves it leaves a higher concentration of the diffusing solute behind it. This phenomenon occurs in both metals and non-metallic materials. A driving force for the process of diffusion along the grain boundaries exists because the solute has a lower chemical potential in the grain boundaries than at the triple junction of the grain boundary with the interface where the solute concentration is high. It is observed at the low temperatures at which lattice diffusion is negligible, because diffusion along the grain boundary can still occur. The solute atoms enter the lattice at the grain boundaries, not by lateral lattice diffusion into the surrounding grains, but as a consequence of trapping at lattice sites as the boundary migrates. Hence, enrichment of the lattice in solute occurs only at the trailing edge of the migrating grain boundary.

The model generally accepted assumes that at the onset diffusion occurs outwards from the now stationary boundary to produce a layer containing solute adjacent to the solute on both sides of the boundary. (In this case diffusion is from a higher chemical potential to a lower one.) The coherency stress is induced on both sides of the boundary by the difference in lattice parameter due to the difference in composition between the layer and the solute poor parent phase adjacent to it. Now it is assumed that this stress is relieved first on one side of the boundary setting up a *difference in strain energy* across the boundary. This difference in strain energy becomes the driving force for migration of the boundary that is maintained since the moving boundary deposits solute behind it as it moves in such a way that coherency strain does not develop on the trailing side of the moving boundary. This scenario, in which a strain energy difference across an interface acts as the driving force for interface migration, was originally conceived for the diffusion-induced grain boundary migration (DIGM) process. Purdy[19] has provided one description of DIGM consistent with the above.

A phase field study of DIGM[20] has verified the concept that DIGM results when there exists a difference in strain energy due to different local compositions on the sides of the grain boundary, that DIGM will not occur due to forces arising only from the free energy of mixing, and that DIGM may or may not occur due to solute adsorption forces.[21] However, an experimental study of DIGM in Au–Ag alloys implies that other than a strain energy difference may drive DIGM since strain energy is likely to be small in this alloy system. Identification of this other source remains to be secured. A difference in local compositions between the two sides of the grain boundary can engender energy difference other than due to strain. This difference must involve a cross-term between composition and the order parameter defining the difference between the adjoining grains.

4. Solidification

4.1. Pure materials

In pure materials the extraction of the heat of fusion usually controls the rate of solidification, except for those materials which have extremely slow crystallization rates.

Examples of the latter materials are SiO_2, P_2O_5 and GeO_2, covalent materials and many polymers. Doremus[22] has shown how the crystallization velocity in these materials may be estimated on the assumption that the free energy dissipated per unit jump of a molecule across the solid–liquid interface in the solidification step equals the work to move the molecule through a viscous liquid at the solidification velocity. Accordingly, the crystallization velocity is then

$$v_C = \Delta H_f(T_M - T)/(3\pi\eta T_M\lambda^2)$$

where ΔH_f is the heat of fusion, T the temperature at the interface on the liquid side of the interface, T_M the melting point, η is the viscosity, and λ is the jump distance at the solid–liquid interface associated with transition of a molecule from the liquid state to the solid state. When T is the temperature of the liquid phase then v_C represents the velocity corresponding to the case where no heat is rejected into the liquid phase at the liquid–solid interface and all of it goes towards increasing the enthalpy of the solidifying solid phase by the amount $C_p(T_M - T) = \Delta H_f$. The corresponding value of v_C may be considered to be the maximum velocity of solidification possible for a given undercooling of the liquid with no heat rejected to the liquid.

The shape of the solid–liquid interface is controlled by the factor α described in Section 5.3 of Chapter X (i.e., $\alpha = \Delta S_f/R$). For α values less than 2 the interface tends to be rough and non-faceted. Metals have α values < 2. Most organic compounds have α values > 2 and grow facets, while polymers have large α values and grow either faceted or as spherulites. The development of dendrites during growth will be considered below.

Although not apparent from the above relation between crystallization velocity and undercooling, it is known empirically that the coefficient is a function of the interface orientation for faceted crystallization. Thus, this relation must be considered to be a rough approximation only.

4.2. Multicomponent systems

4.2.1. Local equilibrium at solid–liquid interface

In the process of solidification of a melt, having a phase diagram, such as shown in Figure 11.15a, with a planar solidification front, the concentrations in the melt and solid proceed with time as illustrated in Figure 11.15b, on the assumption of no convection, no diffusion in the solid phase, local equilibrium at the solid–liquid interface and a constant velocity for the solidification front. Contrary to the cases treated in Sections 2 and 3, the driving force for interface motion during solidification in metals does not derive solely from chemical potential differences across the interface, but does depend primarily on the dissipation of the heat of fusion via the appropriate temperature gradient at the interface. When steady state is reached the concentration in the liquid with respect to the solidification front is given by

$$c_L/c_o = 1 + [(1 - k)/k]\exp(-Rx/D)$$

where k is the distribution coefficient (c_S/c_L), R is the solid–liquid interface velocity and x is distance measured from the solid–liquid interface. We can now calculate a liquidus temperature–distance curve corresponding to the composition $c_L(x)$ function. If the liquidus and solidus phase boundaries are straight lines then the equation for $T_L(x)$ is

$$T_L = T_o - mc_o\{1 + [(1 - k)/k]\exp(-Rx/D)\}$$

where m is the slope of the liquidus (dT_L/dc). A plot of a $T_L(x)$ function is shown in Figure 11.15c. Suppose now that the actual temperature in the liquid is as shown by the dashed line in this figure. In this case there is a zone ahead of the solid–liquid interface in which the actual temperature is less than the liquidus temperature corresponding to the concentration in the

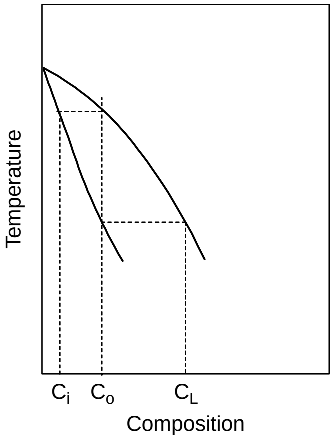


Figure 11.15a. Solidus–liquidus boundaries in phase diagram defining compositions.

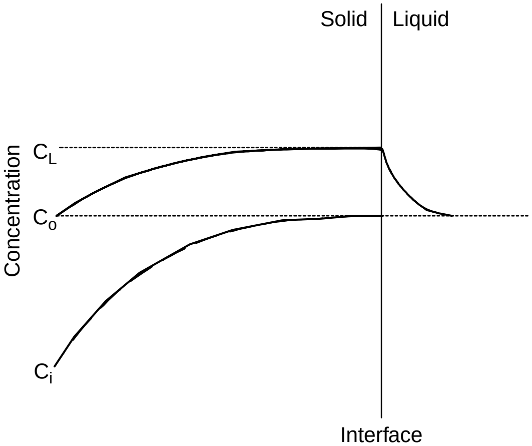


Figure 11.15b. Composition–distance plots. Interface moving at constant velocity. Refer to Figure 11.15a for definitions of compositions relative to phase diagram.

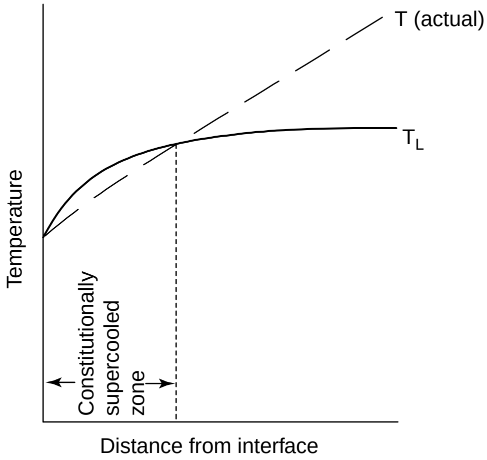


Figure 11.15c. Showing the actual temperature in the liquid phase ahead of the solid–liquid interface (dashed line). Also, the liquidus temperature corresponding to the liquidus composition–distance plot in Figure 11.15a is shown as the solid line. The region where the actual temperature is less than the liquidus temperature is the constitutionally supercooled zone.

liquid at that position. This zone has been called the constitutionally supercooled zone. We will show in the next chapter that any supercooled zone leads to a tendency towards morphological instability of the interface, because any part of the interface that moves ahead of the planar front can get rid of heat and solute laterally and thus move at a faster velocity than the planar front.

At this point it is appropriate to honor Pfann[23] who helped initiate the revolution that brought materials science to its present state of development through his recognition of the role of the liquid–solid solute distribution coefficient in his invention of the process of zone melting. He recognized that the purification of the solid relative to the liquid in the first part to freeze, as illustrated in Figure 11.15a, could be used to achieve the purification of semiconductors needed to make transistors a reality. His scheme was to cause a liquid zone, convectively mixed, to travel through a solid bar. Simple, but effective. The need of the semiconductor industry initiated with his work has spurred the further interest in solid–liquid distribution coefficients that still exists.

4.2.2. Non-equilibrium liquid–solid interface conditions

It is possible with modern methods of heating very thin surface layers and with rapid solidification techniques to achieve, in semiconductors and even in metals, crystallization rates that are not limited by the rate of heat removal. Under these conditions, it is no longer valid to assume that local equilibrium exists at the solid–liquid interface. If we use the viscosity to estimate a maximum crystallization velocity for metals we arrive at a magnitude of about 1 m/second. It is possible for the solid–liquid interface velocity to exceed this value experimentally. Thus, we need a new concept to determine the maximum crystallization velocity. Turnbull[24] has suggested that the solid/liquid interface can move by essentially a diffusionless mode in these materials, limited in velocity only by the velocity of sound. With this concept it is then easy to comprehend how solute may be trapped at concentrations greatly exceeding those corresponding to local equilibrium at the interface.

Aziz[25] has treated the problem of calculating the effective distribution coefficient in rapid solidification. He assumes that there are two processes occurring at the interface: solvent atoms joining onto the solid phase via diffusionless shifting of positions and dragging along with them solute atoms in this process and concurrently the loss of solute atoms in the solid, at the solid–liquid interface, by thermally activated jumps into the adjoining liquid. This model leads to the following expression for the distribution coefficient when the mode of interface migration involves the lateral motion of steps

$$k(v, T) = k_e(T) + [1 - k_e(T)]\exp(-v_D/v)$$

and when the process is continuous, as at a rough interface, the distribution coefficient is given by

$$k = [(v/v_D) + k_e(T)]/[(v/v_D) + 1]$$

where $k_e(T)$ is the equilibrium distribution coefficient at the temperature T of the interface, and v_D is the maximum crystallization velocity corresponding to diffusion limited motion of atoms at the interface given by either the relation in Section 4.1 or by D_i/λ, where D_i is the diffusivity for the jump across the interface and λ is the interatomic distance. Thus, for large v/v_D, k can approach unity. The trapping of supersaturations of solute from the liquid is important technologically in the semiconductor area where it is sometimes desired to develop large supersaturations of desired dopants in electrically active substitutional lattice sites.

Ahmad et al.[26] have investigated a phase field model of rapid solidification with one result in agreement with the rough interface equation for the dependence of k on v. Their result comparing the two relations is shown in Figure 11.16a. The phase field simulation also yielded an approximate result for the value of v_D

$$V_D = 0.375(D_I/l_A)\{\ln(1/k_e)/(1 - k_e)\}$$

where D_I is the interface diffusivity given by the weighted mean of those for the liquid and solid phases, l_A is the interface thickness for component A (solvent) and k_e is the equilibrium distribution coefficient. Comparison to experiment for this deduction is shown in Figure 11.16b. Ahmad et al.[26] concluded that this showed qualitative agreement of their result for v_D with experiment.

Jackson[27] has provided another model of solute trapping based on a modification of the physical concepts governing the crystallization velocity. His relation for the distribution coefficient dependence upon velocity for a rough interface is

$$k = k_e^{1/(1+A'v)}$$

(for a stepped interface v is replaced by $\sqrt{v}$), which is also shown to be consistent with experiment.[27]

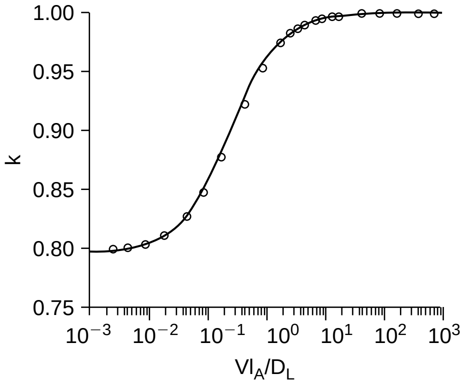


Figure 11.16a. Distribution coefficient as a function of solidification front velocity V. Points are generated by phase field model. Line corresponds to $k = [(v/v_D) + k_e(T)]/[(v/v_D) + 1]$. From Phys. Rev. E 58, 3436(1998) with permission.

Caroli et al.,[28] using non-equilibrium thermodynamics, have considered, for linear interface kinetics (rough and not stepped interfaces), the general conditions which can and cannot yield solute trapping in the solid. Further, they show, contrary to previous assertion,[29] that the Onsager reciprocal relations are applicable to all the representations of the linear interface kinetics. (Interface motion requiring nucleation, or lateral motions of steps emanating from screw dislocations does not correspond to linearizable kinetics.) One result is that solute trapping is possible only in systems with positive values of the off-diagonal (cross) conductance coefficients in the linear relations between generalized forces and fluxes. They generalized this result to far from equilibrium solidification and suggest that solute trapping may result from interactions between unlike atoms such that the freezing in of one constituent leads to the freezing-in of interacting constituents.

The non-equilibrium thermodynamic treatment of this problem by Caroli et al.[28] is so instructive that it is reproduced below. The mass currents are defined by

$$j_i = \rho_i^{L,S}(\mathbf{v}_\mathbf{i}^{L,S} - \mathbf{v}_\mathbf{s})\mathbf{n}$$

where **n** is the unit vector normal to the interface pointing into the liquid, $\mathbf{v}_\mathbf{s}$ is the interface velocity, and $\mathbf{v}_\mathbf{i}^{L,S}$ the velocity of species i and ρ_i the mass per unit volume in the L (or S) phase. The currents j_i thus correspond to the flux of i through the interface. The linear relations between these currents and their corresponding driving forces are for a binary system

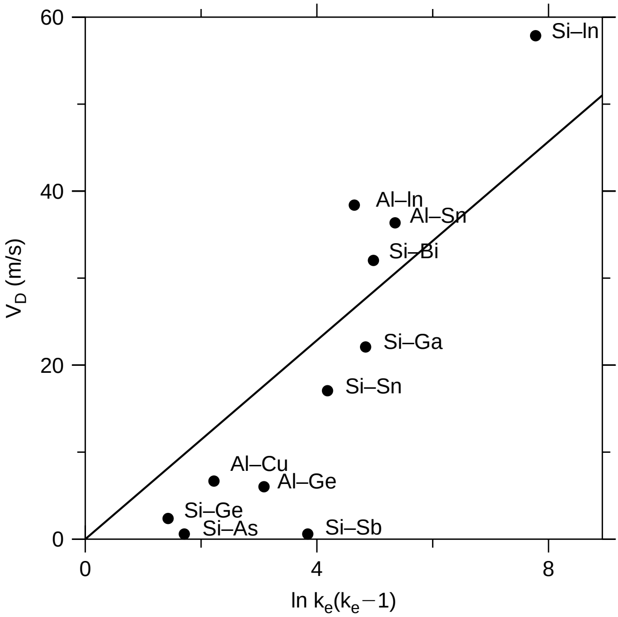


Figure 11.16b. Experimental values for v_D versus ln $k_e/(k_e-1)$. From Phys. Rev. E 58, 3436(1998) with permission.

$$j_A = L_{AA}\delta\mu_A + L_{AB}\delta\mu_B$$

$$j_B = L_{BA}\delta\mu_A + L_{BB}\delta\mu_B$$

where $\delta\mu = \mu^S - \mu^L$. These linear kinetic coefficients satisfy the Casimir–Onsager relations

$$L_{AB} = L_{BA}$$

$$L_{ii} \geqslant 0 \quad L_{AA}L_{BB} - L_{AB}^2 \geqslant 0.$$

Other currents, which are linear combinations of j_A and j_B can be devised, which are useful. In particular, the total mass current

$$J = j_A + j_B$$

and the diffusion current in either of the two phases e = L, S

$$j_c^e = (1 - C_e)j_B - C_e j_A$$

where

$$C_e = \rho_B^e(\rho_A^e + \rho_B^e)$$

is the mass concentration of the solute species B in the e phase.

Caroli et al. showed that the linearized flux relations based on these currents that yield the same interface entropy production also obey the Onsager conditions, contrary to the previous assertion of Baker and Cahn.[29]

For the case that diffusion in the solid can be neglected, $J_c^S = 0$ provides a constraint on the relative values of μ_A and μ_B given by

$$\delta\mu_B = -\frac{(1 - C_S)L_{AB} - C_S L_{AA}}{(1 - C_S)L_{BB} - C_S L_{AB}}\,\delta\mu_A$$

Also, this constraint leads to

$$j_A = (1 - C_S)J \quad j_B = C_S\, J$$
$$j_L = (C_S - C_L)\, J$$

Also, if $L_{AB} = 0$, then

$$J = j_A/(1 - C_S) = L_{AA}\delta\mu_A/(1 - C_S) = j_B/C_S = L_{BB}\delta\mu_B/C_S$$

Since $L_{AA} > 0$ and $L_{BB} > 0$, and for the case of solidification, for which $J < 0$, then $\delta\mu_A < 0$ and $\delta\mu_B < 0$. This result implies that solute trapping is impossible because for solute trapping $\delta\mu_B > 0$.

Consequently, we have arrived at the conclusion that solute trapping requires $L_{AB} \neq 0$ and, as we show below, it requires that $L_{AB} > 0$. Indeed, the condition for solute trapping is $\delta\mu_B/J < 0$. But,

$$\begin{aligned} J &= j_B/C_S = L_{AB}\,\delta\mu_A + L_{BB}\,\delta\mu_B \\ &= -\{(L_{AA}L_{BB} - L_{AB}{}^2)C_S/[(1 - C_S)L_{AB} - C_S L_{AA}] \end{aligned}$$

Thus, $\delta\mu_B/J < 0$ requires that

$$[C_S L_{AA} - (1 - C_S)L_{AB}]/[(L_{AA}L_{BB} - L_{AB}{}^2)C_S] < 0.$$

Since $(L_{AA}L_{BB} - L_{AB}{}^2) \geqslant 0$ and $L_{AA} \geqslant 0$, by the Casimir–Onsager conditions then

$$L_{AB} > [C_S/(1 - C_S)]L_{AA} > 0$$

is the condition that satisfies solute trapping.

We hasten to add that the above model assumes near equilibrium conditions, where the interface velocity does not depend upon the flux of heat. It provides an analysis of the solute trapping effect, however, that is of help in analyzing solute trapping models applicable to far from equilibrium conditions.

It is clear that the treatment of the solidification problem by Caroli et al. is applicable to other growth processes involving a reaction at an interface and diffusion to or from it. We have neglected cross-coupling effects in such processes examined in this chapter. The formal treatment is given in the second paper by these authors listed in Reference 28.

5. Physical vapor deposition

At the outset we need to distinguish between two regimes of physical vapor deposition (PVD). In one, the growth rate is given by the impingement rate and occurs at high supersaturation ratios and low substrate temperatures, where the evaporation rate is much smaller than the impingement rate. This is the situation normally encountered in PVD. In the other regime, the supersaturation is low and deposition is accomplished at near equilibrium conditions. This regime is sometimes used in the production of defect-free, monocrystalline, epitaxial silicon thin films. We will explore the kinetics associated with the latter regime.

Many stages are involved before a molecule incident upon a surface is incorporated into the lattice of the surface material. The possible stages involved in this process are as follows:

1. *Adsorption onto the surface.*
2. *Diffusion along the surface.*
3. *Adsorption along a surface step (see Figure 11.17 for definition of a surface step and kink along the step).*
4. *Diffusion along the step.*
5. *Incorporation into the lattice at a kink.*

Let us define certain concepts necessary to arrive at the growth rate. First, since along vicinal surfaces atoms can only be incorporated into the lattice at kink sites, which exist along surface steps, and such incorporation leads to the lateral motion of such steps, the growth rate along vicinal surfaces is given by

$$G = hv/d_S$$

where h is the height of the surface step, v is the velocity of the step normal to itself along the surface and d_S is the distance between such parallel steps. We will assume that the surface consists of parallel steps. (Indeed, in the epitaxial growth of silicon films the substrate is chosen to have an off-axis orientation to produce

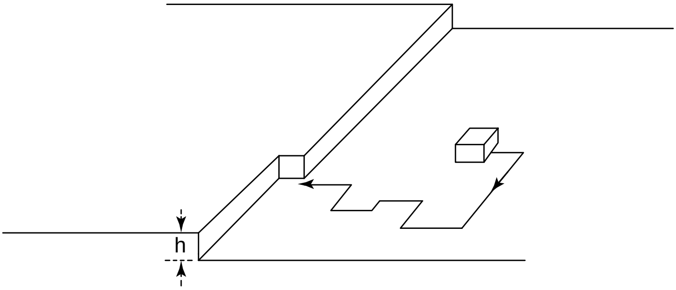


Figure 11.17. Showing a kink at a step on a crystal surface, diffusion of a condensed molecule along the surface to the kink where it is incorporated into the crystal.

an arrangement where the surface consists of parallel surface steps, i.e. $d_s = h/\theta$, where θ is the angle between the surface normal and the low index plane normal.)

To get at the lateral velocity of these steps let us note that the impingement rate I given by gas kinetic theory is

$$I = P/(2\pi mkT)^{1/2}$$

where m is the mass of the molecule, P is the pressure and T is the temperature of the gas. At equilibrium, the impingement rate is equal to the evaporation rate. The latter, far from the steps, is given by

$$E = -dn_S/dt = n_S\nu \exp(-g_{des}/kT) = n_S/\tau$$

where ν is the frequency of attempts at the surface for a surface admolecule to evaporate and g_{des} is the difference in free energy between an admolecule and an evaporated one. Also, it is apparent that

$$\tau = \exp(g_{des}/kT)/\nu.$$

Thus,

$$n_S = P\tau/(2\pi mkT)^{1/2}$$

Equilibrium exists also near the kink sites for the reaction between surface molecules and incorporation of same at kink sites. This reaction yields for the surface concentration near the kink sites

$$n_{Se} = n_o \exp(-g_k/kT)$$

where g_k is the difference in free energy between an admolecule and one at a kink site and n_o is the number of molecular sites per unit area on the surface where adsorption can take place. Thus, a surface concentration profile of admolecules is set up between the region near the steps and far from them.

Let us consider now the diffusion of molecules from the surface to the kinks which are sinks for these molecules. Fick's law gives

$$J_S = -D_S(\partial n_S(x)/\partial x)$$

Conservation of matter requires that

$$\begin{aligned} dJ_S/dx &= I - E = (n_S - n_S(x))/\tau \\ \text{or} \quad -D_S(\partial^2 n_S/\partial x^2) &= (n_S - n_S(x))/\tau \end{aligned}$$

We define a diffusion distance x_S by

$$x_S^2 = D_S\tau$$

Hence

$$x_S^2(\partial^2 n_S/\partial x^2) = n_S(x) - n_S.$$

For the boundary conditions, $n_S(x) = n_{Se}$ at $x = 0$ at the step and $n_S(x) = n_S$ far from it, this equation has the solution

$$n_S(x) = n_S - (n_S - n_{Se})\exp(-x/x_S)$$

where x is considered positive on both sides of the step. Now, the step velocity is related to the flux by

$$vn_o = -2J_S|_{x=0}$$

and the latter is given by

$$\begin{aligned} J_S|_{x=0} &= -D_S(\partial n_s(x)/\partial x)_{x=0} = -D_S(n_S - n_{Se})/x_S \\ &= -D_S(P - P_e)\tau/[x_S(2\pi mkT)^{1/2}] = -x_S(P - P_e)/(2\pi mkT)^{1/2} \end{aligned}$$

Thus, $v = 2x_S(P - P_e)/[n_o(2\pi mkT)^{1/2}] = 2x_S I_e(S - 1)/n_o$, where S is the supersaturation ratio, i.e. $S = P/P_e$. We have neglected the effect of step velocity on the concentration $n_S(x)$. (See Voigtlander et al., Appl. Phys. A 38, (1985) for a treatment that takes this effect into account.)

We have assumed implictly in the above derivation that the concentration at the steps is given by the equilibrium concentration of molecules, and that the diffusion distance x_S is much greater than the distance between kinks along a step. If the distance between steps is less than the diffusion distance then another solution is found. In this case the boundary condition is that at the midpoint between the steps $dn_S(x)/dx = 0$. The solution is

$$(n_S - n_S(x))/(n_S - n_{Se}) = \cosh[(d_S - 2x)/2x_S]/\cosh(d_S/2x_S)$$

Taking the derivative of $n_S(x)$ and substituting into Fick's law we have

$$J_S(x = 0) = -x_S(n_S - n_{Se})\tanh(d_S/2x_S)/(2x_S)$$

The step velocity is then

$$v = x_S(S - 1)I_e \tanh(d_S/2x_S)/n_o$$

where S is the supersaturation ratio, equals $n_S/n_{Se} = P/P_e$. The growth rate is related to the step velocity by

$$G = hv/d_S$$

Thus, the growth rate of a surface having parallel steps will be linearly related to the supersaturation ratio. A more complicated step arrangement on the surface, such as a series of spiral steps about an emerging screw dislocation, will yield a more complicated relation between the growth rate and the supersaturation ratio.

There have been many simulations of PVD (see Bibliography). The commercial applications of this process demand the achievement of control of the process and the resulting deposit and hence the process has been studied extensively. Among the desired products are a strictly ordered array of quantum dots. One of the simulations that provides information relative to this objective is a phase field simulation which takes into account the effects of anisotropic strain fields and surface steps (mesas) on the onset of traveling wave instabilities in the deposition product.[30] Another analysis,[31] which includes a significant rate of adatom desorption in the treatment of the kinetics of deposition, yields an ordered hexagonal array of deposited units for certain values of the kinetic parameters. The latter analysis stems from a discipline that is concerned with the origin of patterns, and is related to the study of chaos in reaction–diffusion systems (see Chapter 13).

6. Chemical vapor deposition

We shall limit our discussion of this complex topic to concepts that are applicable to all the various types of Chemical vapor deposition (CVD). Usually, at least for other than low pressure CVD, there are two series processes, diffusion and interface reaction, that can limit the rates of growth (interface migration) in CVD. Diffusion is in the vapor phase. Since there are gaseous reactants and products the diffusion may be controlled by the slower of the reactants or the slower of the products. Usually, diffusion is through a laminar layer set up by convective flow arrangements in the CVD reactor. Often, at high temperatures the growth rate is limited by diffusion through the gas phase, while at low temperatures the growth rate is limited by some interface reaction. Figure 11.18 illustrates growth rate versus reciprocal temperature for the CVD of silicon. Since diffusion through the gas is only weakly temperature dependent, whereas the interface reaction is strongly temperature dependent, these two stages can be easily delineated.

Very little is known about the specific steps in the interface reaction or even the identity of the limiting step in the reaction, for any CVD process. Silicon

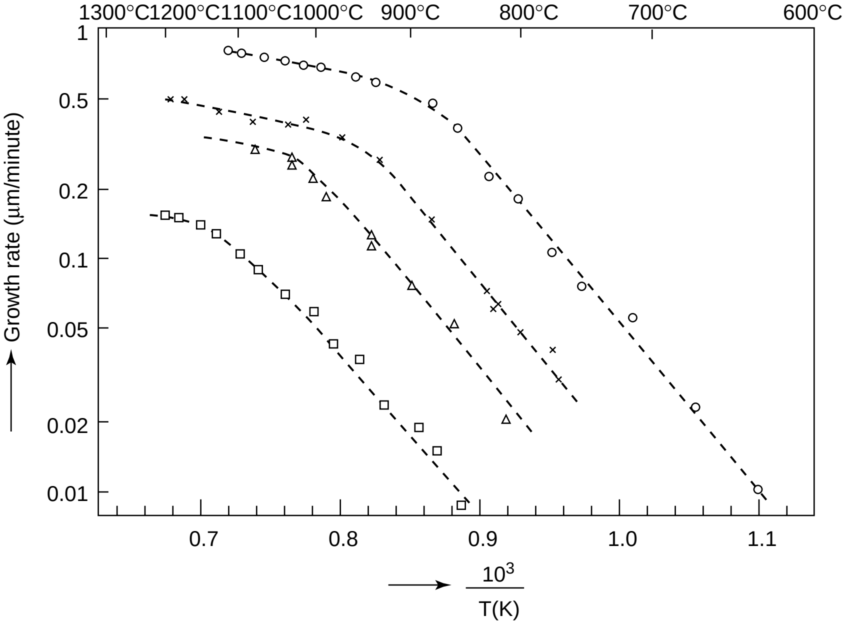


Figure 11.18. Temperature dependence of CVD growth rates of silicon from various source gases. After F.C. Eversteijn, Philips Res. Report 29, 45(1974).

deposition has been studied extensively and still this process is not defined in terms of the identity of the limiting step. Another complication is that reactions can occur in the gas phase that yield intermediate precursors. Thus, often the identity of the precursor responsible for the reaction that donates the substrate atom or molecule is not known. A survey of the status of ignorance and knowledge of the mechanism of CVD of silicon up to 1987 is given in Reference 32. Despite this ignorance, it is still possible to control and deposit acceptable films via CVD.

Among the surface steps that can control the interface reaction are: decomposition of the gaseous precursor yielding the monomer; diffusion of the monomer over the surface; attachment of the monomer to a kink site at the surface; desorption of a product species from the surface. There is no rule that nature must be simple and CVD is certainly one of the less simple processes to define mechanistically.

We shall learn in the next chapter that under diffusion controlled conditions a planar interface between parent and product phases is morphologically unstable, whereas it is stable under interface controlled conditions. A goal of CVD is to deposit smooth surfaces, which are certainly easier to produce when growth is reaction controlled. Thus, there is interest in defining the transition between these two regimes. This has led to the definition of the CVD number, deduced from equating the rates of growth given by diffusion through the gas laminar layer to that given by the reaction at the surface. This number is

$$\mathrm{CVD} = [k_d \delta T_o^2 \ln(T_s/T_b)]/[D_o(T_s - T_b)T_s]$$

where k_d is the mass transfer coefficient at the deposit's surface, cm/second; δ is the boundary layer thickness, cm; $T_{s,\,b}$ the temperature of the surface and "bulk" gas, i.e. the temperatures at the laminar layer interfaces with the deposit and bulk gas stream, respectively; D_o the diffusion coefficient at the temperature T_o.

CVD is one of the complex kinetic processes that is likely to benefit from the application of computer simulation procedures and such studies have begun to appear in the literature.[33] A review of the state of the art in 2002 of simulating CVD[34] reveals the paucity of knowledge required to use simulation methods effectively. Apparently, only the CVD of diamond has benefitted from simulation.

7. Sintering

Sintering is another process that is interface energy driven. It is a process in which the interstices between contacting particles shrink as a consequence of the tendency to decrease the overall interface energy. There are three stages to the sintering process. In the first stage, necks grow between contacting particles, i.e. the area of contact between two contacting particles increases. This process is

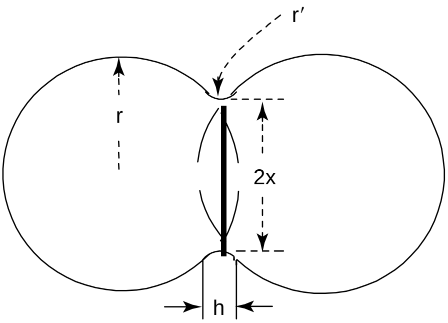


Figure 11.19. Illustrating geometry of neck formed between two spheres in process of sintering.

illustrated in Figure 11.19. These contact areas between crystalline particles are perforce along grain boundaries since adjacent particles are differently oriented. At the end of the first stage, the growth of these necks leads to the formation of spherical pores connected by grain boundaries. The shrinkage of these pores is the process occurring during the second stage of sintering. The grain boundaries, which had been anchored by the pores, now begin to move and their interaction with the remaining few volume percent of pores leading to the removal of the latter comprises the third stage. Incoherent grain boundaries hasten the sintering process because they are sinks for vacancies. Although surfaces and dislocations are vacancy sinks, as well, they are less important, either because they are far removed from the source of vacancies, or they are too few in number, respectively.

There are a variety of paths for matter transport: vapor transport, volume diffusion, grain boundary diffusion and plastic flow. This fact appears to have been appreciated first by Kuczynski,[35] who provided a quantitative theory of sintering that has stood the test of time in its major aspects. It appears that volume diffusion is the significant path for many experimental observations of the sintering process. There are two pathways for volume diffusion of matter: from surface of curvature 2/r to neck surface of curvature $(1/x - 1/r')$ and from grain boundary between the particles to the same neck surface. The latter pathway appears to be the predominant one for $r' << x << r$.

Let us derive a relation for the rate of increase of x. A simplified model is as follows. Vacancies diffuse from just below the neck surface to the grain boundary, where they annihilate. Their virtual chemical potential at the neck surface is $\sigma\Omega(1/r' - 1/x)$, whereas at the grain boundary it equals zero. The corresponding vacancy concentrations are C and C_o, respectively, where $kT(C - C_o)/C_o = \alpha\Omega(1/r' - 1/x)$, $C_o = \exp(-\Delta G_f/kT)$ and ΔG_f is the work to form a vacant site in the lattice by exchange of a subsurface atom with a vacant site on the surface. The diffusion geometry for the vacancies is complicated in that it is radially inwards from the lateral surface of a squat cylinder and axial to one of the two end surfaces. A typical approximation is to assume 1D diffusion through a cross-sectional area πx^2 over a diffusion distance L. The value of L used by various investigators varies, but in all cases it is taken to be proportional to x. Suppose $L = Kx$. Then, by equating the flux from one side of the grain boundary to the rate of increase of half the neck volume, we obtain, with the approximation that half the thickness of the neck equals $r'/2$

$$(1/\Omega)(d/dt)(\pi x^2 r'/2) = (\pi x^2)(D^*\sigma/kTKxr')$$

where $r' \ll x$, $D^* = D_V C_o$ and D_V is the vacancy diffusion coefficient.

But, by the Pythagorean theorem and $r' \ll x \ll r$, $r' = x^2/2r$. Substituting and solving for x as a function of time yields

$$x^5 = (10\sigma\Omega D^* r^2/KkT)t$$

The values of K vary from 1/8 to 1/2 in the literature. (Compare the values in the books of Borg and Dienes, Kingery et al. and the paper of Ashby.[36])

It is possible to derive relations between x and t for the other modes of atom transport with the result that the exponent n in the relation $x^n = At$ equals 2, 3, 5 and 7 when the mechanisms are viscous flow, evaporation–condensation, bulk diffusion and surface diffusion, respectively.

Sintering is even more a complicated process than is Ostwald ripening, considering the gross changes that occur in the geometry and microstructure. A simple model for the second stage[37] yields a relation in which the pore volume depends on the logarithm of time. This relation seems to be obeyed by the data despite the simplicity of the model on which it is based. Sintering has also been simulated in the phase field approach.[38]

8. Summary

We have considered the motion of a planar interface, either between two terminal phases of a binary system having or not having stable intermediate phases in the phase diagram for the system or between a supersaturated or supercooled parent phase and a product phase. For a multicomponent system this motion involves the coupling between an interface reaction and diffusion. Since these processes are in series, the slower controls. We have obtained relations for the interface velocity and defined the transition between these two controlling regimes. We have also briefly considered the processes of PVD, CVD, Ostwald ripening and sintering.

References

1. F.C. Larche and J.W. Cahn, Acta Met. 33, 331(1985).
2. G.V. Kidson, J. Nuclear Mater. 3, 21(1961).
3. X.L. Yeh, K. Samwer and W.L. Johnson, Appl. Phys. Lett. 42, 242(1983).
4. G. Ghosh and G.B. Olson, Metall. and Mater. Trans. A 32, 455(2001).
5. I.M. Lifshitz and V.V. Slyozov, J. Phys. Chem. Sol. 19, 35(1961).

6. C. Wagner, Z. Elektrochem. 65, 581(1961).
7. D. Fan et al., Acta Mater. 50, 1895(2002).
8. H.J. Diepers, C. Beckermann and I. Steinbach, Acta Mater. 47, 3663(1999).
9. H.-J. Diepers, C. Beckermann and I. **Steinbach in Solidification Processing**, eds. J. Beech and H. Jones, Ranmoor House, Sheffield, 1997.
10. C. Beckermann, H.-J. Diepers, I. Steinbach, A. Karma, and X. Tong, J. Comp. Phys. 154, 468(1999).
11. L. Ratke and W.K. Thieringer, Acta metall. 33, 1793(1985).
12. J. A. Marqusee, J. Chem. Phys. 81, 976 (1984).
13. N. Akaiwa and D.I. Meiron, Phys. Rev. 51E, 5408(1995).
14. M. Hillert and G.R. Purdy, Acta Met. 26, 333(1978).
15. P.G. Shewmon and G. Meyrick, in **Diffusion in Solids, Recent Developments,** eds. Dayananda and G. Murch, AIME, Warrendale, PA, 1985.
16. D.B. Butrymowicz, J.W. Cahn, J.R. Manning and D.E. Newbury, "Advances in Ceramics", Vol. 6, **Grain Boundaries and Interfaces in Ceramics**, Ceramics, American society 1983, p. 202.
17. R.W. Baluffi and J.W. Cahn, Acta Met. 29, 493(1981).
18. R.S. Hay and B. Evans, Acta Met. 35, 2049(1987).
19. K. Tashiro and G.R. Purdy, Scripta Mettal. 17, 455(1983).
20. J.W. Cahn, P. Fife and O. Penrose, Acta mater. 45, 4397(1997).
21. G. Meyrick, Scripta Metall. 10, 649(1976).
22. R.H. Doremus, **Rates of Phase Transformations**, Academic Press, New York, 1985, p. 116.
23. W.G. Pfann, Zone Melting, John Wiley, New York, 1957; Kruger, New York, 1978.
24. D. Turnbull, J. Physical Chem. 66, 609(1962).
25. M.J. Aziz, J. Appl. Phys. 53, 1158(1982); Apply. Phys. Lett. 43, 552(1983).
26. N.A. Ahmad et al., Phys. Rev. E 58, 3436(1998).
27. K.A. Jackson, Interface Sci. 10, 159(2002).
28. B. Caroli, C. Caroli and B. Roulet, Acta Met. 34, 1867(1986); J. Cryst. Growth 66, 575(1984).
29. J.C. Baker and J.W. Cahn, in **Solidification**, ASM Seminar, ASM, Metals Park, Ohio, 1971.
30. J.J. Eggleston and P.W. Voorhees, Appl. Phys. Lett. 80, 306(2002).
31. D. Walgraef, Physica 15, 33(2002).
32. J.M. Jasinski, B.S. Meyerson and B.A. Scott, Ann. Rev. Phys. Chem. 38, 109(1987).
33. A. Netto and M. Frenklach, MRS Symp. Proc. 527, 383(1998).
34. C.C. Battaile and D.J. Srolovitz, Ann. Rev. Mat. Res. 32, 297(2002).
35. G. Kuczynski, Trans. AIME 185, 169(1949).
36. M.F. Ashby, Acta Met. 22, 275(1974).
37. R.L. Coble, J. Appl. Phys. 32, 787(1961).
38. A. Kazaryan, Y. Wang and B.R. Patton, Scripta Mater. 41, 487(1999).

Bibliography

Growth of intermediate phases

1. V.I. Dybkov, J. Mater. Sci. 21, 3078, 3085(1986); 22, 4233(1987); J. Phys. Chem. Sol. 47, 735(1988).
2. C. Wagner, Acta Met. 7, 99(1969).
3. S.R. Shatynski, J.P. Hirth and R.A. Rapp, Acta Met. 24, 1071(1975).
4. D.S. Williams, R.A. Rapp and J.P. Hirth, Met. Trans. 12A, 639(1981).
5. Guan-Xing Li and G.W. Powell, Acta Met. 33, 23(1985).

Vapor deposition

1. W.K. Burton, N. Cabrera and F.C. Frank, Phil. Trans(London) A 243, 299(1951).
2. E.S. Machlin, **Materials Science in Microelectronics: The Relationships Between Thin Film Processing and Structure**, Elsevier, Oxford, 2005.

Sintering

1. G.C. Kuczynski, **Materials Science Research**, Vol. 13, Sintering Processes, Plenum Press, New York, 1980.
2. W.D. Kingery, H.K. Bowen and D.R. Uhlmann, **Introduction to Ceramics**, 2nd edition, John Wiley, New York, 1976. This book has a good chapter on sintering processes.
3. M.F. Ashby, Acta Metall. 22, 275(1974).
4. F.B. Swinkles and M.F. Ashby, Acta Metall. 29, 259(1981).
5. C.H. Hsueh, A.G. Evans, and R.L. Coble, Acta Metall. 30, 1269(1982).
6. M.A. Spears and A.G. Evans, Acta Metall. 30, 1281(1982).

CHAPTER XII

Morphological Instability and Growth of Phases

Introduction

We will explore the subject of morphological instability in this chapter. There are many manifestations of morphological instability: in solidification, where a planar growth front can degenerate into a dendritic growth morphology; in thin film deposition where the thin film can decompose into an array of globules; in chemical vapor deposition (CVD), where a planar front can degenerate into one containing whiskers, etc. Historically, the first quantitative treatment of morphological instability was due to Mullins and Sekerka[1] and their work has since inspired a flood of other studies involving further mathematical development of their linear analysis into the non-linear realm and still other types of morphological instability phenomena.

The instability of a morphology can occur in the presence or absence of a reaction front. At a reaction front, instability is induced by composition or temperature gradients ahead of the interface between parent and product phases and, normally, capillarity forces tend to restore stability in these situations. The morphological instability at a planar reaction front between a parent and a single product phase is that treated by Mullins and Sekerka. The morphological instability at a reaction front between a single metastable parent phase and a duplex product is denoted as an Eckhaus instability. Both these morphological instabilities will be discussed in this chapter. The morphological instability of a dendrite during its growth, treated first by Langer and Muller-Krumbhaar,[2] belongs to the first category. Transitions in morphology can occur during growth which are also a consequence of morphological instability, i.e. from a cellular pattern to a dendritic pattern. These transitions will be discussed in this chapter as well. Thus, in this chapter we will first consider the morphological stability of product interfaces in contact with either supersaturated or supercooled parent phases. Also, we will consider the morphological instability of various product morphologies.

1. Morphological instability

1.1. Basis for morphological instability at a reaction front

Whenever growth of a stable phase occurs by migration of an interface between the stable and metastable phases, the interface is subject to morphological instability

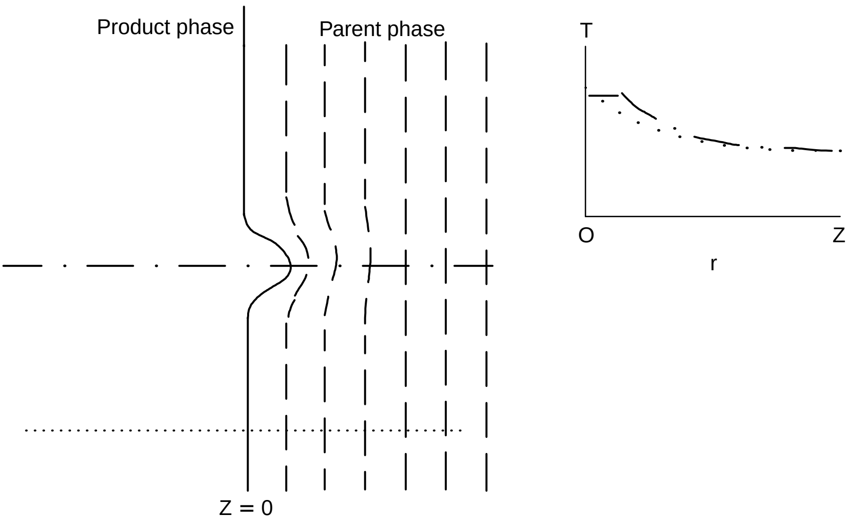


Figure 12.1. Showing the crowding of isotherms (dotted lines) about a bump on a flat interface between an unstable parent phase and a growing product phase. Inset shows plots of temperature versus distance along corresponding lines in main figure.

when the region adjacent to the interface in the metastable phase is either supersaturated or supercooled. This tendency toward morphological instability can be understood qualitatively with the aid of Figure 12.1. This figure shows iso-temperature profiles in a supercooled melt (or iso-concentration profiles in a supersaturated solution) adjacent to a spherical bump on a planar interface.

It is apparent that because of the crowding of the isotherms about the bump that the temperature gradient in the vicinity of the bump is larger in magnitude than it is adjacent to the planar region. Hence, the bump can reject the heat of fusion per unit extension into the melt faster than can the planar section, and consequently solidify faster than the planar portion of the interface. This qualitative analysis neglects the stabilizing influence of capillarity due to the resistance to increase of the interface area. The influence of capillarity is such that bumps having radii smaller than some critical radius will tend to disappear, whereas bumps larger than this critical radius will grow. Typical products of morphological instabilities are dendrites and whiskers.

These concepts describe the physical nature of the morphological instability problem, at least for the solidification of pure materials into a supercooled liquid. In principle, the same tendency for the morphological instability of a planar interface exists for growth into a supersaturated or supercooled phase. Morphological instabilities can arise from other sources as well. Internal stress induced in thin films that are coherent with their substrates can be a source of morphological instability as can surface tension. Also, there are conditions where the reaction front may undergo an oscillation superimposed on its steady state propagation and thereby induce morphological inhomogeneities. The mathematical treatment of all these problems is complex. In the following we reproduce Mullins and Sekerka's original derivation of the morphological instability of a liquid–solid interface during solidification.

1.2. Liquid–solid interface (solidification)

Consider a solid–liquid interface which originally was planar, but has been perturbed to have the form of a periodic wave, as illustrated in Figure 12.2. This wave has an amplitude δ and a wave number $2\pi/\lambda$, where λ is the wavelength. The interface moves with a constant velocity V in the z direction, where $z = 0$ at the interface. We assume a binary system in which m is the slope of the liquidus and k is the distribution coefficient. Also, the concentration gradient ahead of the interface is G_C, and the temperature gradients in the liquid and solid are G_L and G_S. If $(1/\delta)(d\delta/dt) > 0$, then the interface is morphologically unstable. If, on the other hand, this ratio is negative then the periodic perturbation in the planar interface will decay to restore the planar interface.

This problem was solved originally by Mullins and Sekerka.[1] The solution is obtained by setting down the partial differential equations that govern the transport of heat in the liquid and the solid and the transport of solute in the liquid. It is assumed that no diffusion occurs in the solid phase. Also, conservation of matter and heat balance at the interface are obeyed. (There is of course a divergence of heat at the interface due to the heat of fusion.) Further, the effect of capillarity on the temperatures and concentrations at the interface is taken into account. The far field boundary conditions are assumed equal to those for a planar interface. The local interface velocities calculated from heat flow relations must equal those evaluated from diffusion considerations. At the same time they also must equal $V + (d\delta/dt)\sin\omega X$, where V is the velocity of the planar interface. The partial differential equations are then solved subject to the boundary conditions and the other conditions.

Consider a planar interface, moving with a velocity V relative to the bounding phases which are fixed with respect to the laboratory coordinate system,

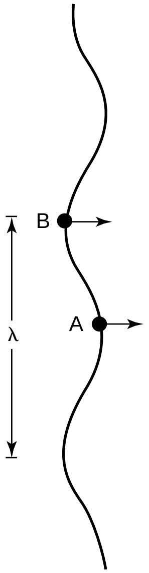


Figure 12.2. Illustrating a perturbed interface with a wave length λ. A and B are the highest and lowest points on the interface, respectively.

and another coordinate system, to which we will refer certain distances, that moves with the interface. Let the z axis point toward the liquid and the plane $z = 0$ coincide with the interface. Suppose a sinusoidal fluctuation in the shape of the interface described by

$$z = \delta(t)\sin\omega x$$

is introduced. The object of this analysis is to obtain an expression for $d\delta/dt$ to determine whether the perturbation in the shape of the interface grows or decays. This analysis requires the calculation of the local velocity V(x), where x denotes the distance along the interface from some origin, in terms of the values of the local thermal and diffusion gradients. The applicable diffusion equations are

$$\frac{\partial u}{\partial t} = \nabla^2 u + (V/D_u)(\partial u/\partial z)$$

where u is a field of either concentration (in the liquid phase) or temperature (in both the liquid and solid phases) and D_u represents the corresponding diffusivity. These equations have to satisfy the far field boundary conditions (the concentration and temperatures that would exist several wavelengths away from the interface for the unperturbed interface). Also, at the interface local equilibrium is assumed and hence the local temperature at the interface is given by

$$T^* = mc^* + T_M(1 + \Gamma K) = mc^* + T_M(1 + \Gamma\delta\omega^2\sin\omega x)$$

Here K is the average curvature at a point on the interface (positive when concave toward the liquid), $\Gamma = \sigma/L$, where σ is the specific interface energy and L the latent heat of the solvent per unit volume.

The local interface velocity calculated from heat flow considerations must agree with that calculated from diffusion conditions, i.e.

$$V(x) = (1/L)[K_S(\partial T'/\partial z) - K_L(\partial T/\partial z)] = [D/(c^*(k - 1))](\partial c/\partial z)^*$$

where the asterisk denotes interface values, k is the distribution coefficient (ratio of equilibrium solute concentration in solid to that in liquid), and the denominator $c^*(k - 1)$ gives the difference in concentration between the solid and liquid sides of the interface.

By setting

$$T^* = T_o + a\delta(t)\sin\omega x$$

$$c^* = c_o + b\delta(t)\sin\omega x$$

where the subscript $_o$ denotes values for a flat interface. The solutions that satisfy the diffusion equations and the boundary conditions are then on the liquid side

$$c(x, z) - c_o = (G_c D/V)(1 - \exp[-Vz/D]) + \delta(b - G_c)\sin \omega x\, e^{-\omega^* z}$$

$$T(x, z) - T_o = (G'' D''_{th}/V)(1 - \exp[-Vz/D''_{th}]) + \delta(a - G'')\sin \omega x\, e^{-\omega'' z}$$

and on the solid side

$$T(x, z) - T_o = (G' D'_{th}/V)(1 - \exp[-Vz/D'_{th}]) + \delta(a - G')\sin \omega x\, e^{-\omega' z}$$

where G'' and G' are the thermal gradients at the unperturbed flat interface in the liquid and solid, respectively, G_c is the concentration gradient in the liquid at this interface and where

$$\omega^* = (V/2D) + [(V/2D)^2 + \omega^2]^{1/2}$$

$$\omega'' = (V/2D''_{th}) + [(V/2D''_{th})^2 + \omega^2]^{1/2}$$

$$\omega' = (V/2D'_{th}) - [(V/2D'_{th})^2 + \omega^2]^{1/2}$$

The parameters a and b can be obtained by appropriate substitution and are

$$b = \frac{2G_c T_M \Gamma \omega^3 + \omega G_c (H' + H) + G_c[\omega^* - (V/D)](H' - H)}{2\omega m G_c + (H' + H)[\omega^* - (V/D)(1 - k)]}$$

$$a = mb - T_M \Gamma \omega^2$$

where $H = (K_L/\bar{K})G$; $H' = (K_S/\bar{K})$ and $\bar{K} = 0.5(K_S + K_L)$.

To determine $d\delta/dt$, we substitute for the temperature gradients in the equation for the local velocity to obtain

$$v(x) = V + (d\delta/dt)\sin \omega x = (\bar{K}/L)(H' + H) + (\bar{K}/L)\omega\{2a - (H' + H)\}\delta \sin \omega x$$

Equating like Fourier components yields

$$V = (\bar{K}/L)(H' + H)$$
$$(d\delta/dt) = (\bar{K}/L)\omega\{2a - (H' + H)\}\delta$$

Substituting for a we then obtain the desired result

$$(1/\delta)(d\delta/dt) = \frac{V\omega\{-[\omega^* - (V/D)(1-k)][2T_M\Gamma\omega^2 + (H' + H)] + 2mG_c[\omega^* - (V/D)]\}}{(H' + H)[\omega^* - (V/D)(1-k)] + 2\omega mG_c}$$

Above we repeated the classical perturbation analysis of Mullins and Sekerka.[1] However, we present here the more general results for the stability criterion due to Trivedi and Kurz.[3] Their result for the stability criterion is

$$-\Gamma\omega^2 - \bar{K}_L G_L\left[\frac{\omega_L - (V/a_L)}{\bar{K}_S\omega_S + \bar{K}_L\omega_L}\right] - \bar{K}_S G_S\left[\frac{\omega_S + (V/a_S)}{\bar{K}_S\omega_S + \bar{K}_L\omega_L}\right]$$

$$+ mG_C\left[\frac{\omega_C - (V/D)}{\omega_C - (V/D)(1-k)}\right] \quad \begin{array}{l} >0 \text{ (Morphological instability)} \\ <0 \text{ (Morphological stability)} \end{array}$$

where $\Gamma = \sigma/\Delta S$, σ is the solid–liquid interface energy, ΔS is the entropy of fusion per unit volume, $\omega = 2\pi/\lambda$, where λ is the wave length of the sinusoidal perturbation in shape of the planar interface. Also

$$\bar{K}_L = K_L/(K_S + K_L)$$

and

$$\bar{K}_S = K_S/(K_S + K_L)$$

where $K_{S,L}$ are the thermal conductivities in liquid (L) and solid (S). Further, $a_{L,S}$ are the thermal diffusivities, V is the velocity of the interface and D is the solute diffusion coefficient in the liquid.

Let us consider this stability criterion for various conditions. We obtain the stability conditions of Mullins and Sekerka by setting $\omega_C = \omega_S = \omega_L = \omega$ in the above relations. In this approximation, the stability criterion becomes

$$\Gamma\omega^2 > -\bar{K}_L G_L - \bar{K}_S G_S + mG_C\frac{\omega_C - (V/D)}{\omega_C - (V/D)(1-k)}$$

At low velocities, the ratio in the last term approaches unity and, if the capillarity term is neglected, is the same as the modified constitutional supercooling criterion.

At high velocities or large undercoolings, where $G_S = 0$ is an acceptable approximation and where we also assume $K_S = K_L$ and $a_S = a_L$, then the criterion for stability becomes

$$V^2 > (mG_CD^2/k - G_La_L^2)/\Gamma$$

By use of the conservation relations at the interface, i.e.

$$DG_C = C_L(1 - k)V \quad \text{and} \quad K_LG_L = -V\,\Delta H$$

and $a_LC_p = K_L$, and by substitution we obtain

$$V > mC_S(1 - k)D/(k^2\Gamma) + a_L\Delta H/(C_p\Gamma)$$

for the criterion for stability at high interface velocities. The first term on the right-hand side corresponds to the result of the Mullins and Sekerka analysis. The second term is a correction due to the Trivedi and Kurz analysis. For the Mullins and Sekerka conditions

$$V < (\Gamma\omega^2 + \bar{K}_LG_L + \bar{K}_SG_S)D/[mC_S(1 - k)/k]$$

for the criterion for stability at low velocities. In the Trivedi and Kurz analysis the assumptions $\omega_C = \omega_S = \omega_L = \omega$ are not made.

A plot of log solute concentration versus interface velocity is shown in Figure 12.3. As shown, the stable regions at low velocity and high velocity are separated by an unstable region. The morphological stability at low velocity is due to the lack of undercooling, whereas that at high velocity depends upon the relative values of the capillary length ($d_o = \Gamma/[mC_S(1 - k)/k] = \Gamma/\Delta T_o$), the solute diffusion length ($l_S = D/V$), the thermal length ($l_T = \Delta T_o/G$) and the thermal diffusion length (a_L/V). The microstructural consequences of morphological instability in directional solidification of a binary alloy depend upon the

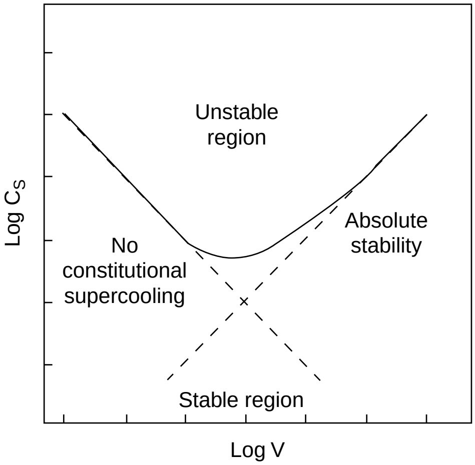


Figure 12.3. Interface stability as revealed in a plot of log solute concentration as a function of log interface velocity.

relative magnitudes of these lengths, i.e. $\nu = l_T/l_S$ and $A = kd_{oi}/l_S$. The latter is the Sekerka capillary parameter and the former gives the level of morphological instability. The limits of planar interface stability are defined by $\nu = 1$ and $A = 1$. Incidentally, the instability criterion can be written in terms of these lengths as follows: $\lambda > 2\pi[d_o l_S \, \nu/(\nu - 1)]^{1/2}$, where $\omega = 2\pi/\lambda$. This description of the instability criterion allows the following statement. Instability of a planar front solidifying in a binary alloy sets in when the wavelength of the perturbation exceeds the mean square root of the product of the capillary length by the solutal length, at least for large values of the ratio of the thermal length to the solutal length.

This result is applicable in a general sense to other processes. Whenever, the perturbation wavelength (divided by 2π) exceeds a mean square root of the capillary length by another controlling length of the process then we may expect morphological instability. A thorough review of the morphological stability problem is given in Reference 4.

One of the questions one may ask at this point is: *Since a band of wavelengths of periodic perturbations can produce morphological instability in a plane front during solidification in a binary alloy, which wavelength does nature choose?* One would think that the answer is the wavelength with the fastest growing amplitude. However, we have not succeeded in obtaining measurements to evaluate this answer. We do have measurements of the steady-state values of the apparent primary spacing between cells and of that between dendrites, but there is no basis for concluding that these spacings are related to the wavelengths yielding morphological instability. Indeed, as will be shown in the next chapter these spacings correspond to growth within the regime of morphological stability of these products of morphological instability.

The transitions of the microstructural manifestations of the modes of growth, planar to cellular to dendritic and back to planar, are discussed in detail in Reference 5 and will be considered later in this chapter.

1.3. Vapor–solid interface

1.3.1. Chemical vapor deposition

A stability analysis for CVD has been carried out by Brekel[6] that is analogous to that for solidification carried out by Mullins and Sekerka. The main difference is that in CVD, diffusion and heat flow occur through a laminar layer of constant thickness h_g, that is large compared to the amplitude of the sinusoidal perturbation in shape of the interface. The temperature and composition at the outer boundary of the laminar layer are maintained constant. Thus, the solutions to the diffusion equations differ to some extent, but not importantly. In the notation used in the previous section, the result for the instability criterion under conditions where the wavelength of the perturbation λ, satisfies $\lambda \ll h_g$ and $\lambda \ll d$, where d is the thickness of the substrate, becomes

$$(1/\delta)(d\delta/dt) = V\omega\left[\frac{2G_C - 2C_e^{\infty}\Gamma\omega^2 + z(G_g + G')}{2(1 + \omega h_g/Nu)G_C + z(G - G' - G'')}\right] > 0$$

where $$z = C_e^{\infty}(\Delta H/(RT_o^2)) + (C_o - C_e^{\infty})(\Delta E/RT_o^2)$$

$$G = 2K_gG_g/K;\ G' = 2K_SG_S/K;\ G'' = 2a\sigma T_o^4/K$$

$$Nu = k_Dh_g/D;\ K = K_S + K_g + K_r;\ K_r = 4aT_o^3/\omega$$

Also, ΔH is the heterogeneous reaction molar enthalpy ($\Delta H < 0$ for an exothermal reaction) and ΔE is the activation energy for the heterogeneous decomposition reaction

$$AB_g = A_S + B_g$$

Further, C_e^{∞} is the equilibrium concentration of AB at a planar interface at temperature T_o, C_o is the actual concentration at the planar interface at temperature T_o, K_i is the thermal conductivity in the i phase, G_C is the concentration gradient, G_g the thermal gradient in the gas phase, G_S is the thermal gradient in the solid phase at the unperturbed interface, a is the emission coefficient, $\sigma = 1.35 \times 10^{-12}\,\text{cal/K}^4\,\text{s}\,\text{cm}^2$, the Stefan–Boltzmann radiation constant, Γ is the capillarity constant ($=2\gamma\Omega/kT_o$), where γ is the interface energy and Ω is the atomic volume of the solid, k Boltzmann's constant), ω is the wave number of the sinusoidal morphological perturbation on the planar interface ($\omega = 2\pi/\lambda$, where λ is the wavelength of this perturbation), k_D is the mass transfer coefficient or rate constant of the heterogeneous decomposition reaction as given by

$$k_D = k_o\exp(-\Delta E/RT)$$

and D is the diffusivity in the gas phase of the limiting reactant species and, finally, V is the vapor/solid interface velocity.

For infinite Nusselt number (Nu) the above result for the instability criterion is very similar to that obtained by Mullins and Sekerka and the one we have derived in the previous section. In particular, the parameters 1/z and C_e/z play the same roles as the parameters m and T_M play in the solidification instability analysis.

In the expression for $(1/\delta)(d\delta/dt)$ the denominator is positive (if it were negative it would result in the prediction that the capillarity term is not stabilizing, which is contrary to physical sense); in the numerator the term with G_C is positive since the concentration gradient is positive and favors instability; the term containing Γ is negative and represents the stabilizing effect of capillarity; and the term proportional to z(G + G′) can be either positive or negative. When the substrate is hot compared to the gas phase (the term (G + G′) is negative) and the protrusions are slightly cooler than the valleys of the sinusoidal perturbed surface. In this case and when z is positive then the thermal gradient term is stabilizing against propagation of the morphological perturbation. For diffusion controlled growth ($Nu > 1$ and $C_o = C_e$) a positive value of z implies a positive value of ΔH (i.e. an endothermic reaction for the decomposition, $AB_g \rightarrow A_S + B_g$). This result can also be understood using the relation for the equilibrium concentration at a curved interface given by

$$C_e = C_e^{\infty}[1 - (\Delta H/RT_o^2)(T - T_o)]$$

Hence, the equilibrium concentration is slightly lower in the valley as compared to a protrusion and consequently the supersaturation is higher in the valley than above the protrusion and the latter will grow at a slower rate than the valley, i.e. a stabilizing condition. A change in sign of ΔH will reverse this situation because the relative supersaturation also reverses. With a change in sign of the thermal gradients, the valleys become cooler than the protrusions and now endothermal reactions ($\Delta H > 0$) will destabilize the morphology while exothermal ($\Delta H < 0$) reactions stabilize the morphology.

In the event the deposition process is reaction controlled then, because $\Delta E > 0$, the sign of z is also positive for $\Delta H > 0$. For $\Delta H < 0$, the sign of z depends upon the supersaturation, the relative values of C_o and C_e. Hence, no general conclusions on stability can be drawn in this case.

To summarize, morphological instability during CVD is encouraged by a high supersaturation leading to a large positive value of G_c. It is also encouraged when the substrate is hot relative to the gas by an exothermal decomposition reaction, or when the substrate is cold relative to the gas by an endothermal decomposition reaction.

Empirically, morphological instability is quite common in CVD leading to the growth of whiskers. The onset of whisker growth, however, may depend upon other special factors, such as the existence of a special property at the tip of the whisker not existing elsewhere on the surface of the film. For example, in the vapor–liquid–solid mode of whisker formation in CVD, the tip of the whisker has a lower melting point than elsewhere resulting in the formation of a liquid drop on the tip and thus yielding a higher accommodation coefficient at the tip and its more rapid growth, i.e. supersaturation is not the driving force for morphological instability in this case.

1.3.2. Physical vapor deposition

There are at least two sources of morphological instability in physical vapor deposition (PVD). Shadowing of lower asperities by higher protrusions tends to yield a higher deposition rate on the highest protrusions and their growth at the expense of smaller ones. The finite size of the atoms results in a higher projected area at the surface through the centers of the atoms over protrusions than over valleys for the same area on the surface.[7] Figure 12.4(a) illustrates this effect, which occurs after sufficient deposition in a non-planar interface having protrusions, such as those shown in Figure 12.4(b). Srolovitz et al.[7] and Karunasiri et al.[8] have applied perturbation theory to study the morphological stability in PVD. In this case, diffusion is a stabilizing influence, tending to smooth out the asperities provoked by the two effects described above. Srolovitz et al. predict a transition between Zones I and II and the growth of columnar grains in the latter zone. Karunasiri et al. also find that it is possible to grow a flat surface up to certain critical film thickness. The protrusions illustrated in Figure 12.4(b) are a consequence of an instability with respect to surface curvature in both models. This instability is revealed using a linear stability analysis as per the dendrite stability analysis in the latter model.

Internal stress at a surface can be the origin of morphological instability leading to the development of a cusps into the stressed solid.[9] This instability can be understood physically as follows. For sufficiently small wavelength of periodic cusps the material above the troughs will be stress free although the surface energy will be increased by the formation of additional surface area. Hence, the film becomes unstable with respect to such surface morphology perturbation when the wavelengths of the perturbations are smaller than a critical wavelength that scales inversely as the square of the internal stress. One possible result of such an instability is a distribution of unstressed isolated islands. It is probable that this source

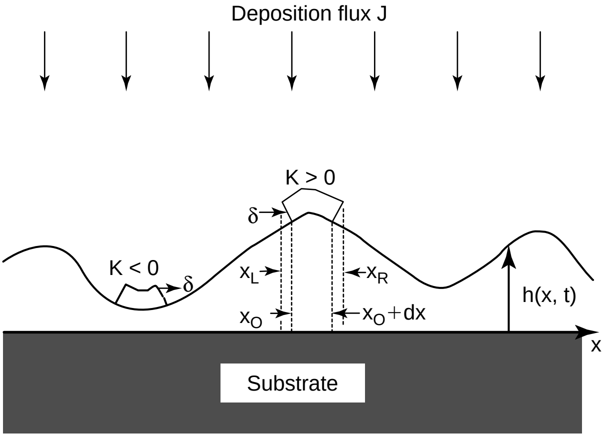


Figure 12.4a. Illustrating the difference in projected length due to surface curvature. K is the surface curvature and δ is the atomic radius. From J. Vac. Sci. Tech. A 6, 2371(1988) with permission. © 1988 American Institute of Physics.

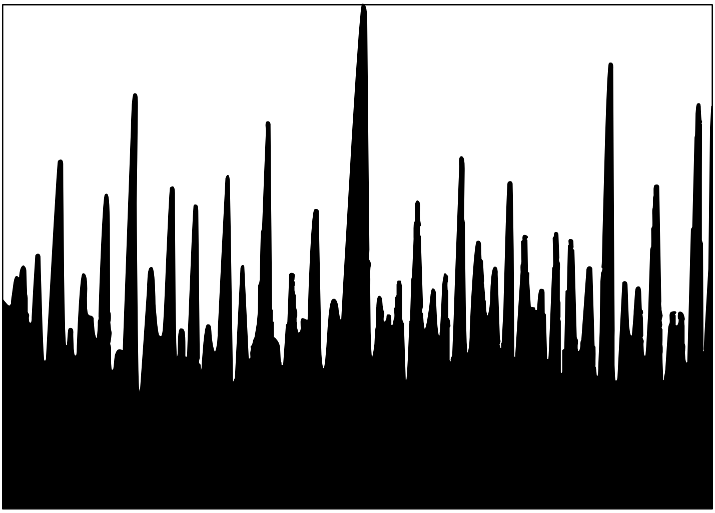


Figure 12.4b. Illustrating the calculated profile developed by deposition in the absence of surface diffusion. From Phys. Rev. Lett. 62, 788(1989) with permission. © American Physical Society.

of instability operates in most of the other growth processes, as well as in vapor deposition, and it may well be that the onset of Stranski–Krastanov growth in which islands form on a monocrystalline layer is due to this phenomenon. A review of this and other instabilities in crystal growth problems can be found in Phys. Rep. 324, 271(2000).

1.4. Solid–solid interface

Mullins and Sekerka[10] treated the morphological stability of a growing sphere in a supersaturated field, describing the infinitesimal perturbation in the spherical shape in terms of an expansion in spherical harmonics. The result for the absolute instability criterion is

$$(1/\delta)(d\delta/dt) = \left\{\frac{D(n-1)}{(C_\beta - C_\alpha(r))r}\right\}\left[\frac{C_o - C_\alpha(r)}{r} - \frac{C_o\sigma v_\beta(n+1)(n+2)}{kTr^2}\right]$$

where

$$C_\alpha(r) = C_\alpha\{1 + [(1 - C_\alpha)/(C_\beta - C_\alpha)][(2\sigma v_\beta)/(kTr\zeta)]\}$$

and

$$\zeta = 1 + d\ln f_\alpha/d\ln C_\alpha$$

where f_α is the activity coefficient and C_α is the concentration of solute in alpha in equilibrium with a planar precipitate of beta. Also, r is the radius of the spherical precipitate, D is the diffusivity of the solute in alpha, C_o is the far field concentration in alpha, σ is the interfacial energy, v_β is the molar volume of the beta phase, k is Boltzmann's constant and T the absolute temperature.

Thus, since instability will occur when $(1/\delta)(d\delta/dt) > 0$, it is required that $n \geqslant 2$ and $r > [\sigma v_\beta/(kT(C_o - C_\alpha(r)))][C_o(n + 1)(n + 2)]$ for the absolute instability. However, a shape change will not occur unless $(1/\delta)(d\delta/dt) > (1/r)(dr/dt)$. The latter inequality requires that $n \geqslant 3$. If we define

$$r^* = [C_\alpha/(C_o - C_\alpha)][2\sigma v_\beta/kT]$$

and for $n = 3$ we obtain that the growing spherical precipitate will be morphologically unstable when $r > 21r^*$. In this problem, the supersaturation is the driving force for instability and capillarity provides the stabilizing force.

It is known that the growth of solid phases from supersaturated solids does not often lead to the development of the morphological instability that exists. One reason for the absence of morphological unstable growth forms in this situation is that the growth may be controlled by an interface reaction rather than by diffusion. Other possible explanations are: that rapid lateral diffusion of the solute along the interface between parent and product phases decreases the differences in growth velocity between bump and hollow in a planar interface; that anisotropy of the interface energy, which is more prevalent between solid phases, increases the contribution of the capillarity term to the stabilization of shape perturbations. Still another possible reason may be related to the dependence of the elastic energy induced by the solid state phase transformation on the shape of the product phase.

2. Dendritic solidification

2.1. Pure materials

Up to this point in this book it has been possible to provide analytic relations that yield solutions without the *need* for computer simulations to obtain answers to questions, although in the previous chapters we found that computer simulation helped to provide understanding of the simplest problems involving the motion of free boundaries. However, an increase in the number of variables affecting the boundary migration induces non-linearity in the relations and a marked increase in the difficulty of providing analytic solutions for the free interface migration problem in the presence of morphological instability. One of the main difficulties has been that of obtaining a unique solution. It is believed, at this writing, that conceptually the problem has been solved, although the details of the solution may be questioned. The long road followed to this point, with many incorrect suggestions for the conditions defining the unique solution, may be found described in the previous editions of this book and in certain references.[11] The conceptual solution to the simplest free-boundary problem of solidification in a one-component system involves what is called the solvability theory. A short description of the problem and solution for the case of solidification in a one-component system follows.

From the morphological instability that exists at a sufficiently undercooled liquid we may expect that the initial shape of the growing (stable) phase will approximate a needle. Suffice it to say that for the case of a steady-state growth of the needle, Ivantsov[12] found in his investigation of this transformation of state, which assumed that the only rate-limiting process is the diffusion of the latent heat released at the moving interface into the undercooled liquid and which neglected capillary and kinetic effects at the interface, that paraboloids of revolution are the shape of the needle/liquid interface consistent with solutions to the thermal diffusion equation

$$\partial T/\partial t = D\nabla^2 T$$

and heat conservation at the growing interface

$$LV_n = c_P D\mathbf{n} \cdot (\nabla T\,|_S - \nabla T\,|_L)$$

where T is the temperature, t the time, D the thermal diffusivity, L the latent heat of fusion per unit volume, c_P the specific heat at constant pressure per unit volume, V_n the interface velocity normal to the interface, $\mathbf{n} \cdot \nabla T$ the normal gradient of the temperature, and S, L the solid and liquid sides of the interface. Ivantsov also derived the relationship $P = \rho V/(2D)$, where P is the Peclet number, and ρ the tip radius. In this solution the Peclet number is defined but the individual values of V and ρ are not.

Many attempts were made to discover the conditions that provide defined values for ρ and V and these were found in disagreement with experiment. The one procedure that until now has not been rejected is known as solvability theory. *Study of the effects of capillarity and kinetics revealed that the effect of capillarity was to perturb the Ivantsov continuous family of needle solutions into a discreet set of solutions.*[13] *Further, for solutions to exist (solvability) either or both capillary and kinetic properties must be anisotropic.*[13] *Then, of the existing solutions, the only one that is linearly stable is that corresponding to the maximum velocity.*[14]

The theoretical and mathematical sophistication and effort expended to arrive at this point was a significant fraction of this type of activity carried on in the world and this short review does not do the effort justice.

With the advent of computer power we now have additional means of evaluating the hypothesis that solvability requirements determine the solution to the dendrite formation problem in pure materials. Although experiment is one means of evaluating the predictions of this hypothesis, we do not at this time have enough information to perform a valid experimentally based evaluation. In particular, we do not know the effect of a small quantity of impurities upon the relationships involved so that any experiment, even in a purified material, involves sufficient impurities to possibly affect the results. We do know that the segregation of impurities to grain boundaries is extremely sensitive to variation of the relative orientation of the grains. We do not know anything about this dependence for solid/melt interfaces. Since, the solvability results are sensitive to anisotropy of surface tension and interface kinetics we remain in the land of "ambiguity" in experimental attempts to evaluate the solvability hypothesis. However, computer simulations can be carried out on strictly pure materials. Hence, computer simulations provide, in principle, means of evaluating solvability solutions. But, the simulation procedures need themselves to be validated since assumptions are involved in their construction.

The phase field method makes use of a diffuse interface between phases. (In this sense, the method is an outgrowth of the Cahn–Hilliard[15] and Allen–Cahn[16] procedures.) This fact introduces some ambiguity relative to the relation between the results of phase field simulation and reality for the cases where this interface is sharp. The phase field method for a single component in addition to the two equations governing the heat flow given at the beginning of this section makes use of a relation which stems from the Allen–Cahn equation for the time dependence of the order parameter on the change in free energy with a change in the order parameter. This relation is

$$V_{\mathbf{n}} = M(\mathbf{n})\{(T_M - T_I)C_p/L - (T_M C_p/L^2)\gamma(\mathbf{n})\kappa\}$$

where V_n is the interface velocity along the vector **n** which is normal to the interface at the point the equation is being applied, M is the local mobility of the interface (kinetic coefficient), T_M the melting temperature, T_I the local interface temperature, C_p the specific heat, L the heat of fusion, γ the local interface energy and κ the local interface curvature. Setting aside considerations of achieving a solution to these three independent equations, we note that there are three independent parameters, V, T_I and κ (or ρ). Hence, in principle it should be possible that a solution exists for them. The difficulty over the years has been to develop this solution in part because the boundary conditions are contained in the solution. But perseverance prevailed and, as already noted, answers are provided by solvability theory and computer simulation procedures, such as the phase field and level set methods.

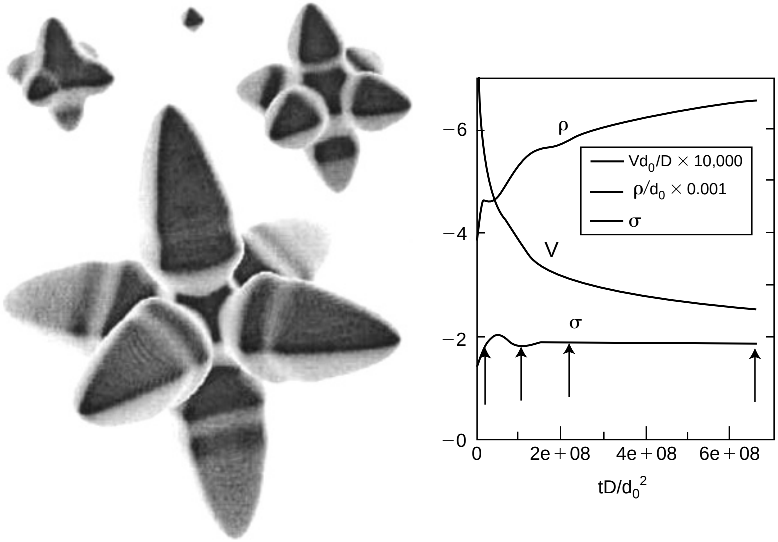


Figure 12.5. Three-dimensional dendritic growth phase field simulation. The snapshots of the dendrites correspond to the arrows in the right-hand plot of dimensionless tip velocity and tip radius versus dimensionless time. From Ann. Rev. Mater. Res. 32, 163(2002) with permission. © 2002 Annual Reviews.

These simulation methods provide the great advantage of yielding 3D snapshots of stages of product growth. For the case of the dendrite in pure materials an example of a phase field result is shown in Figure 12.5. The 3D simulation is for a dimensionless undercooling of 0.05 and a 2.5% surface tension anisotropy. The snapshots correspond to the times denoted by arrows in the right-hand plot. As noted in the figure caption at the time this simulation was performed it was at the limit of computer usage feasibility. A similar 3D result was obtained much more efficiently using the level set method. This result is illustrated in Figure 12.6 for a larger undercooling and different surface tension anisotropy relative to those used for the phase field simulation of Figure 12.5. The

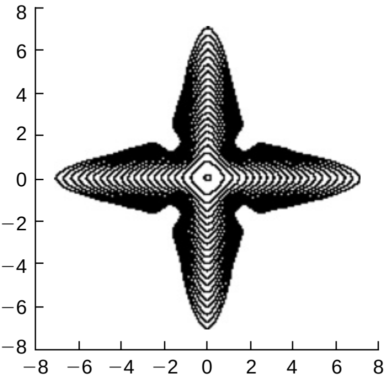


Figure 12.6. Contours of dendrite growth produced in level set simulation of solidification of a pure material. From J. Sci. Comput. 19, 183(2003) with permission. © 2003 Plenum Publishing Co.

attainment of the detailed information regarding the growth front shown in these two figures *requires* the use of computer simulation. It is apparent that free boundary problems are best treated using computer simulation.

The proper use of these models requires that an asymptotic analysis be performed in order to obtain a mapping between the parameters of the phase-field equations and the sharp-interface equations. Such an asymptotic analysis has been described by Karma and Rappel.[17] When the ratio of the interface width to the thermal diffusion length is sufficiently small then it has been shown that such asymptotic analysis yields the same results for all the phase field methods that were examined.[18] By "sufficiently small" in the previous sentence is meant convergence to the sharp interface limit. Further, in this case of adequate convergence the steady state achieved *is that predicted by solvability theory*. Also, the transient dynamics approach the steady state uniquely. Thus, all phase field models that purport to evaluate solvability theory need to demonstrate that the solutions converge to the sharp interface limit of an asymptotic expansion of the phase field equations. Not all of these phase field studies of dendrite growth investigate this convergence requirement. Indeed, for certain studies convergence may require computational expense beyond that acceptable. For this reason some investigators have begun studies involving the application of the level set method[19] to sharp interface free boundary problems. This is the state of the art in the development of computer simulation models at this writing.

We noted above that the applicability of solvability theory can be tested by use of either the phase field method with a proper asymptotic analysis or by the level set method. One appropriate test by a phase field method[20] yielded the results shown in Figure 12.7 where it is shown that both phase field and solvability theory yielded the same values for the velocity as a function of the undercooling. This same study revealed that the velocity is not sensitive to anisotropy of the interfacial energy but is sensitive to anisotropy of a kinetic coefficient ε_k on which the mobility M depends, as shown in Figure 12.8.

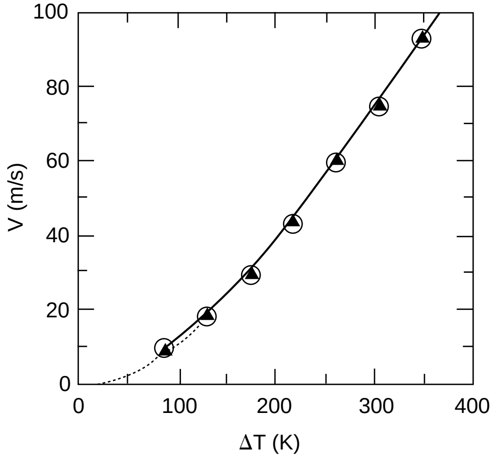


Figure 12.7. Dendrite velocity versus undercooling for the case of isotropic interfacial energy and anisotropic kinetic coefficient. Phase field results given by filled triangles and solid line, solvability results given by dashed line for same values of kinetic anisotropy and isotropic interfacial energy. From Interface Sci. 10, 121(2002). © 2002 Kluwer Academic Publishers.

A test using the level set method[18] resulted in the data shown in Figure 12.9, where the value of the tip velocity given by both the level set and phase field methods is shown to asymptotically approach the value given by solvability theory for the same conditions. A further test of solvability theory using the dependence of the product of tip radius squared and velocity on the ratio of the diffusivities in solid and liquid states is shown in Figure 12.10. Again, solvability theory is in good agreement with the level set simulation results.

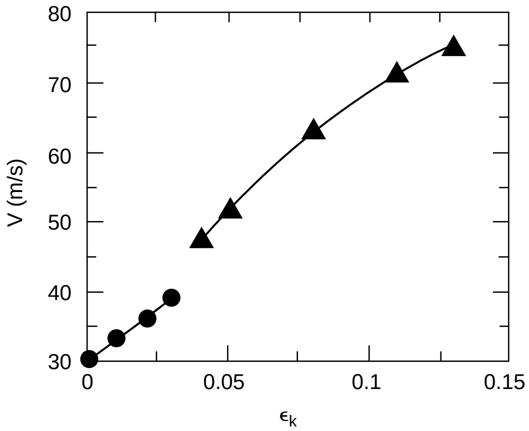


Figure 12.8. Interface velocity (filled triangles) and average velocity of the envelope of the interface (filled circles) versus kinetic anisotropy coefficient. From Interface Sci. 10, 121(2002). © 2002 Kluwer Academic Publishers.

From the above we may state that both phase field and level set simulations confirm the results of solvability theory. Kim[18] has also shown that when the phase field method does not asymptotically approach the sharp interface requirements then it does not yield agreement with solvability theory and when care is taken to make the method properly approach the

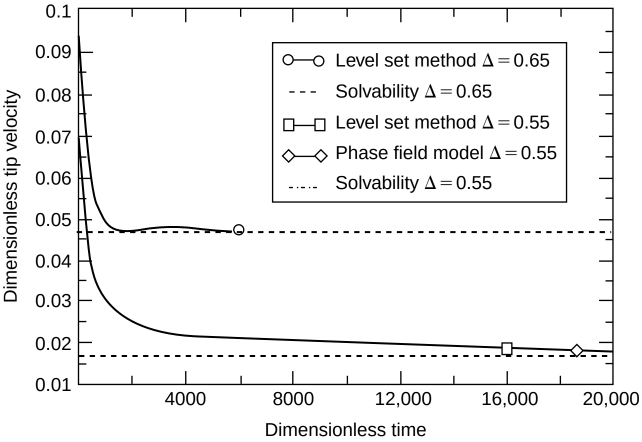


Figure 12.9. Time evolution of the tip velocity for phase field and level set simulations compared to the values predicted by solvability theory. From Phys. Rev. E 62, 2471(2000) with permission.

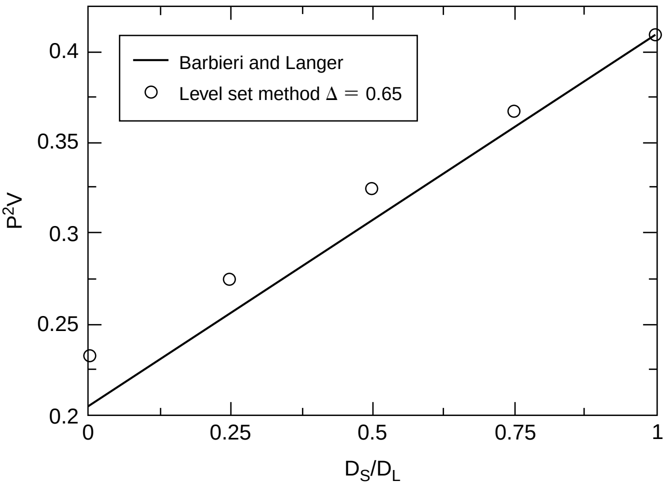


Figure 12.10. Comparison of level set simulated values of the product (tip radius)^2 velocity versus the ratio of diffusivities in solid and liquid states with those given by the solvability theory treatment of Barbieri and Langer.[22] From Phys. Rev. E 62, 2471(2000) with permission.

sharp interface requirement then phase field simulation is consistent with solvability theory.

The origin of the side branching from the needle to form the dendrite shape is morphological instability. However, other than this general statement the particulars of this instability are still being debated. Undoubtedly, anisotropy still plays a role in that the side branches have crystallographic orientation. Kessler and Levine stated that their results[14] "mean that side branching must be understood via the amplification of *finite* noise as the disturbance moves away from the tip." Langer in his analysis of side branching[21] came to the conclusion "that noise of some kind does seem to be a plausible source of the side branching observed in dendritic growth." Brener and Temkin[21] showed theoretically that in a non-axisymmetric dendrite needle thermal noise is sufficient to induce the growth of side branches. Bisang and Bilgram's experiments[21] verified this theory and concluded that side branches originated from thermal fluctuations. It is known that the inclusion of a term for noise in the Ginzburg–Landau equations facilitates the observation of side branching in phase field simulations.[22] Indeed, without a noise term in these equations side branching is not observed in a phase field simulation that ignores kinetic anisotropy and which has a refined computational mesh.[23] With a coarse computational mesh, noise is introduced via discretation errors and side branching is observed even in the absence of deliberately added noise. However, it is not known at this writing whether thermal noise is the sole cause of side branching or whether tip oscillational behavior, which has been experimentally observed[24] in some cases, induces side branching as well. There is a suggestion based on remarks in Kim's thesis[25] that side branching may occur for a large enough kinetic anisotropy in the absence of a noise term in the phase field simulation procedure.

2.2. Multicomponent materials

Phase field simulation involves a diffuse interface and when it is applied to a sharp interface phenomenon it requires additional modification. To make phase field simulation quantitative for one-component dendritic solidification, as mentioned above, it is necessary to modify the relations so as to attain adequate asymptotic approximation to the sharp interface condition. For multicomponent solidification to be simulated by the phase field method additional modifications are necessary. One modification that has been found efficacious is to introduce a non-variational term in the phase field equations to correct for surface diffusion, interface stretching and chemical potential discontinuity that appear in its absence and which prevent quantitative application of the phase field method to alloy solidification and the like.[26] This non-variational term defines an anti-trapping current. With this term modified phase field equations have been investigated for their ability to simulate Mullins– Sekerka morphological instability and the instability that gives rise to the production of concentration cells with results showing that for reasonable values of parameters the modified phase field simulates both instabilities.[27] We may expect many more phase field studies exploiting this capability in the future.

The onset of concentration cells, as mentioned, is a consequence of an instability related to the transport of solute. One of the results of studies of instabilities is that not only does morphological instability disappear at low and high velocities of a planar interface in alloy solidification (see Figure 12.3), but its existence is also dependent upon the wavelength of the perturbation as revealed in Figure 12.11. Stability exists either because the undercooling is insufficient or capillarity is too strong.

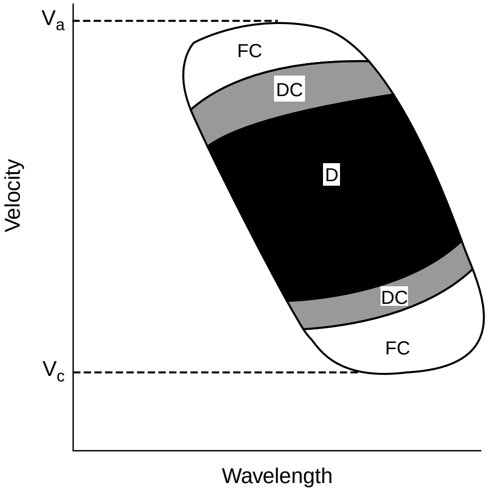


Figure 12.11. Schematic representation of solidification regimes. FC: finite cellular; DC: deep cellular; D: dendritic. From J. Cryst. Growth 106, 410(1990) with permission.

Non-linear stability analysis reveals that below the lower critical velocity V_C a planar interface may become unstable for finite amplitude perturbation. (The critical velocity V_C results from the M-S analysis of the stability of an infinitesimal perturbation.) This instability is denoted to be a subcritical bifurcation and represents a transition from a planar to a non-planar interface. Figure 12.12 reveals such a subcritical bifurcation. The type of bifurcation that can exist depends upon the distribution coefficient, k. For $k < 0.45$ a subcritical bifurcation occurs, whereas for $k > 0.45$ a supercritical bifurcation is predicted. Also, only a supercritical bifurcation is predicted at high velocities.[28]

Within the instability regime in the least unstable zone next to the boundary the tendency is for concentration cells to develop. Figure 12.13 shows for one type of non-linear instability the asymmetric wave front of a cell and the concentration of solute along the interface for various values of a parameter that depends upon distribution coefficient and morphology as deduced from a theory of the non-linear instability.[29]

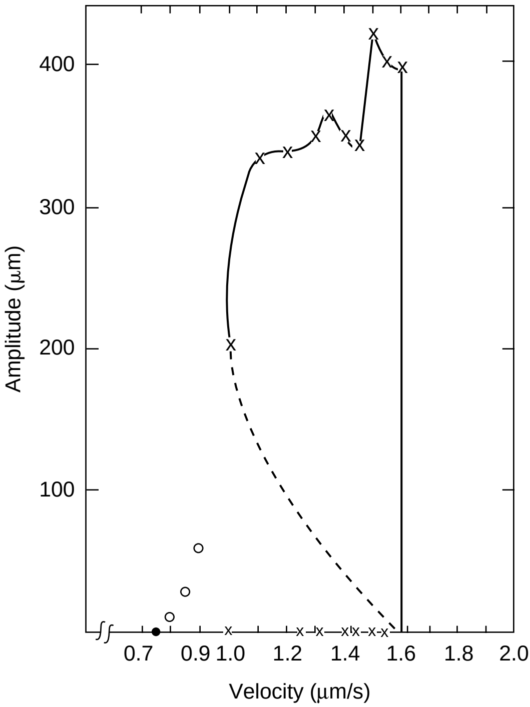


Figure 12.12. Subcritical bifurcation in the succinonitrile-0.10 wt.% acetone system. G = 38.2°C/cm. Amplitude of the interface microstructure versus velocity. Open circle denotes smooth interface shapes, without cusps, which were unstable. From Acta Metall. 35, 2443(1987) with permission.

The onset of cellular and dendritic patterns in solidification occurs as noted in Figure 12.11. We distinguish here between the onset of these patterns and their steady-state parameters. The onset of these patterns occurs with successive degrees of morphological instability relative to a planar front. The literature on this subject is immense. Some reviews are cited in the Bibliography.

There are two types of cellular patterns denoted by the terms finite cellular and deep cellular. These patterns are distinguished not only by their morphological difference (i.e. with deep cellular patterns the depth of their grooves is much larger than the spacing between cells while with finite cellular

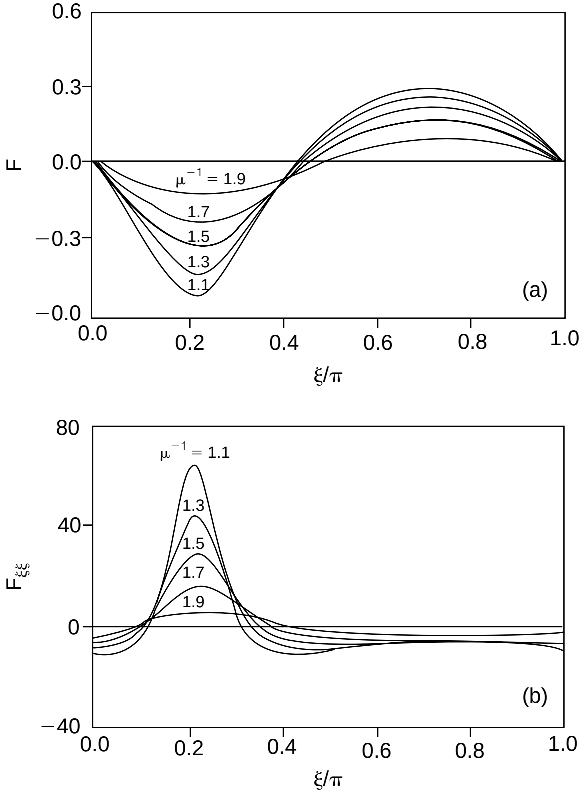


Figure 12.13. (a) One wavelength of a periodic front of the stable solid–liquid interface produced some time after onset of instability; (b) Corresponding distribution of solute. From Phys. Rev. B 38, 11452(1988) with permission.

patterns the groove depth is on the order of their spacing), but also by the dependence of their spacing on velocity. For finite cells the spacing decreases with increase in the velocity whereas for deep cells the spacing increases with increase in the velocity. The deep cellular pattern is a transition pattern between the finite cellular one and the dendritic one. A microstructure map showing the transitions between these patterns has been calculated by Billia et al.,[30] based on the Sekerka stability diagram and is shown in Figure 12.14. The parameter ν is given by the ratio of the thermal length l_T to the solutal length l_S, where $l_T = \Delta T_o/G$ and $\Delta T_o = T_L - T_S$, the liquidus minus the solidus temperatures with G the temperature gradient ahead of the solid–liquid interface and where the solutal length equals the ratio of the diffusivity D, to the velocity V. The parameter $A = kd_o/l_S$, where k is the distribution coefficient, $d_o = \gamma/\Delta S\Delta T_o \cdot \gamma$ is the solid–liquid interfacial energy and ΔS is the entropy of fusion.

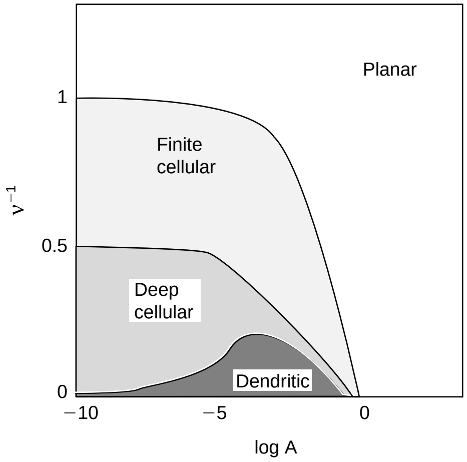


Figure 12.14. A map showing the transitions between planar, cellular and dendritic microstructures in a plot of two dimensionless parameters, ν^{-1} and A. See text for definition of these parameters. From J. Cryst. Growth 106, 244(1990) with permission.

3. Eutectic solidification

In eutectic solidification, the corresponding phase diagram is similar to the two right-hand diagrams of Figure 3.5. In this case, a liquid with composition near or at the eutectic composition solidifies to produce two solid phases. The metastable parent liquid phase produces these two more stable solid product phases at an interface that we will call either the growth front or the reaction front. Normally, the major transport of the component species defining the binary system takes place in the liquid phase as a consequence of the higher diffusivities of these components in the liquid than in either solid phase. However, in some systems solid state transport can contribute to the partitioning of the solute and solvent species.

First, we present the solutions to the transport equations due to Jackson and Hunt[41] for the case of steady-state eutectic solidification under the influence of a temperature gradient that is moving with a constant velocity (forced-velocity condition) for the case where the interfaces are non-faceted. These investigators were the first to obtain a solution that provided agreement with many aspects of the eutectic solidification of non-faceted interfaces. The basic assumption of this work is that the temperature at the growth front is constant along the interface, i.e. the reaction front is planar.

In the above class of eutectic solidification the transport equations have a solution of the form

$$\Delta T(\lambda) = AV\lambda + B/\lambda$$

where

$$A = mP(1 + \zeta)^2 C_o/(\zeta D)$$

$$P = \sum_{1}^{\infty}(n\pi)^{-3}\sin^2(n\pi S_\alpha/(S_\alpha + S_\beta))$$
$$B = 2(1 + \zeta)[a_\alpha^L/m_\alpha + a_\beta^L/(\zeta m_\beta)]$$
$$a_\alpha^L = (T_E/L)_\alpha \sigma_\alpha^L \sin\theta_\alpha^L$$
$$a_\beta^L = (T_E/L)_\beta \sigma_\beta^L \sin\theta_\beta^L$$
$$1/m = 1/m_\alpha + 1/m_\beta$$

and V is the reaction front velocity, $\zeta = S_\beta/S_\alpha$ the ratio of the lamellar plate thicknesses in phases β and α, m_i is the slope of the liquidus in equilibrium with the i phase, C_o is the concentration difference at the eutectic temperature T_E, between the concentrations of the solid phases in equilibrium with the liquid phase at the eutectic composition, D is the diffusivity in the liquid phase, L is the latent heat for the eutectic reaction: liquid $\rightarrow \alpha + \beta$, σ is the interfacial energy of the interface between the phases denoted by the subscript and superscript, θ_β^L is the angle between the tangent to the β/L interface and the reaction front at the α/β/L triple junction, and finally $\lambda = 2(S_\alpha + S_\beta)$, i.e. the S values correspond to half the respective plate thickness.

The above result is for the steady-state solidification achieved by having the reaction front move at the same velocity as a moving temperature gradient G, as occurs in directional solidification. A derivation of the Jackson–Hunt result that is simpler than theirs follows.

By conservation

$$J_\alpha = -V(1 - k_\alpha)C_e$$
$$J_\beta = V(1 - k_\beta)(1 - C_e)$$

where J_i is the flux to or from the i phase in the z direction of reaction front motion, V is the velocity of this front, k_i is the distribution coefficient for the i phase and C_e is the eutectic composition. (These relations make the same assumption made by Jackson and Hunt that the concentration at the interface is everywhere C_e.)

The free energy dissipated in the diffusion process is now evaluated. We note that the total flux emanating from half of the alpha phase is $J_\alpha\lambda_\alpha/2$ while that to the beta phase is $J_\beta\lambda_\beta/2$. These two total flux values must be equal since the liquid phase suffers no change in its composition at the steady-state condition. We now define an envelope $r(\theta)$ such that $J(\theta)r(\theta) = J_\alpha\lambda_\alpha/2 = J_\beta\lambda_\beta/2$ where $J(\theta)$ is the mean flux normal to the plane at the angle θ contained within the region from $r = 0$ to $r(\theta)$. These quantities are defined in Figure 12.15. The free energy dissipated per unit of time per unit area of the parent/product interface is

$$(2/\lambda)\int_0^\pi (J^2(\theta)r^2(\theta)/L)d\theta$$

Substituting for $J(\theta)r(\theta)$ and integrating we obtain

$$\lambda_\alpha^2 V^2[(1 - k_\alpha)C_e]^2\pi/2L\lambda$$

where $L = DC_e/RT$, and $\lambda = \lambda_\alpha + \lambda_\beta$. We can simplify still further by noting that the diffusivity in the liquid phase, D, is related to the viscosity η, by the relation, $D = RT/\eta d$, where d is the average atomic diameter of the liquid metal solution and that

$$\lambda_\alpha = A\lambda$$

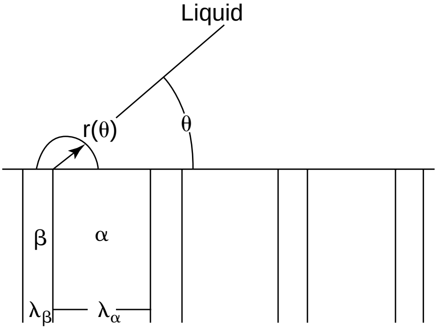


Figure 12.15. Illustrating the envelope about the zone of transverse transport of solute between lamellae.

where $A = [(C_\beta - C_e)/(C_\beta - C_\alpha)]$. Substituting and setting the rate of free energy dissipation devoted to diffusion, attachment kinetics and formation of the interfaces between the lamellae to $V\Delta G$ we obtain

$$V\Delta G = V^2\lambda A^2[(1 - k_\alpha)C_e]^2\pi\eta d/2C_e + 2\sigma V/\lambda + V^2/M$$

We divide both sides by V, set $\Delta G = \Delta S\Delta T$ and obtain the Jackson–Hunt result when attachment kinetics is ignored in accord with their assumption, i.e.

$$\Delta T = \mathbf{A}V\lambda + \mathbf{B}/\lambda$$

But now $\mathbf{A} = A^2[(1 - k_\alpha)C_e]^2\pi\eta d/2C_e\Delta S$ and $\mathbf{B} = 2\sigma/\Delta S$. Substitution of values for the parameters then yields the same numerical results as obtained from the Jackson–Hunt relation within the uncertainty in the values of these parameters. The relation for ΔT as a function of λ is shown in Figure 12.16 for the Pb–Sn eutectic system. Also shown is the range of experimental values of the interlamellar spacing, which is revealed by the limits to the horizontal lines in this figure. We now need to explore the possible explanations for these limits. But before we do so it is worthwhile noting that the microscopic solvability condition which yields a unique solution for the case of steady-state dendrite growth does not do so in this case of steady-state eutectic solidification.

Let us consider first the minimum limit to the interlamellar spacing. Many authors[32,33] have shown that the minimum in the undercooling as a function of the interlamellar spacing corresponds to a limit of the morphological stability of the eutectic structure. However, as shown later, this conclusion, as one encompassing all conditions, is in error.[34] Experiments on a transparent organic alloy showed the

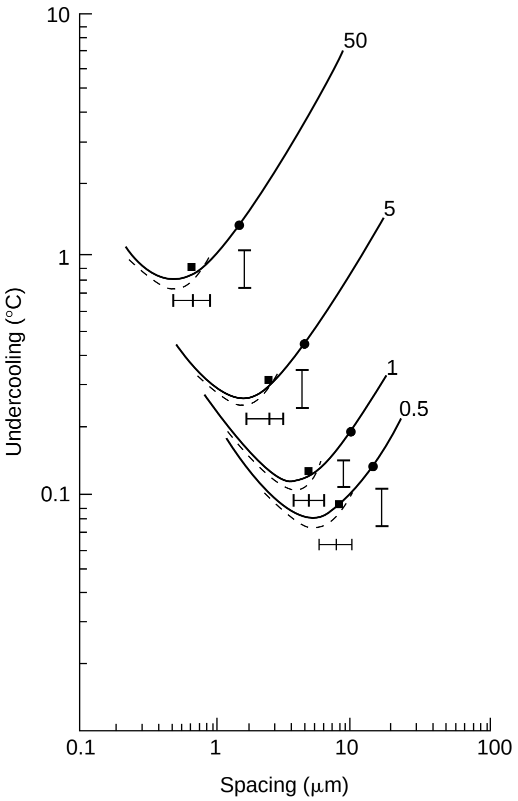


Figure 12.16. Undercooling versus interlamellar spacing for Pb–Sn. From Acta Metall Mater. 43, 3301(1995) with permission.

growth of eutectics with spacing as small as 0.8 of that corresponding to the minimum undercooling. The data which yielded the horizontal lines in Figure 12.16 on the other hand are consistent with the expectation from the earlier theories that the minimum interlamellar spacing corresponds to that at the minimum undercooling. According to Akamatsu et al.[34] the effect of triple junctions on this spacing accounts for the discrepancy between these two observations. They found that the triple junctions have a force tending to make them slip sideways that depends upon the local gradient of the interlamellar spacing. In metals this force becomes appreciable for front velocities below about 1 μm/s at temperature gradients in the 100°K/cm range.

Langer[32] provided an analysis of the marginal stability condition for this mode of solidification based on mainly dimensional considerations, which were then justified by a more detailed linear stability analysis by Datye and Langer.[35] We shall reproduce the main arguments of the Langer analysis.

Let the position of the interface be described by the function η(x, t), which measures the distance of the local interface at x (as defined in Figure 12.17) from its undeformed position relative to the frame of reference moving with the velocity V. The moving temperature gradient has the value G and thus the local undercooling at the displaced interface at position x will be given by

$$G\eta(x, t) = -\Delta T(\lambda)$$

Langer assumed that η(x, t) and the vertical displacement y (as defined in Figure 12.17) are coupled by the condition that each lamella must grow in a direction

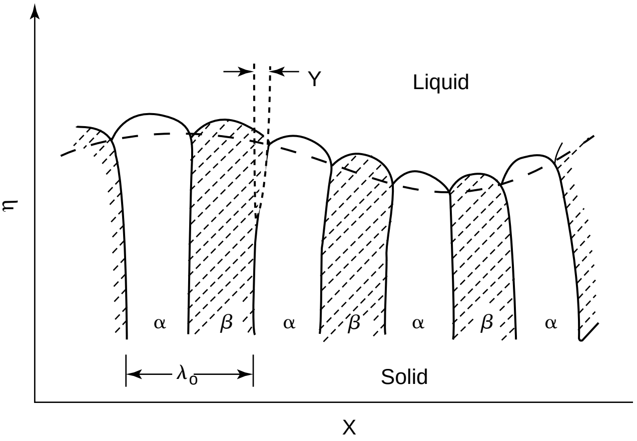


Figure 12.17. Figure defining the coordinates η, y, x at the solidification front. After Langer, Phys. Rev. Lett. 44, 1023(1980).

which is locally perpendicular to the solidification front after the original assumption of Cahn.[36] Thus

$$\partial y/\partial t = -V\,\partial\eta/\partial x$$

Also, the local lamellar spacing and y are related through

$$\lambda(x, t) = \lambda_o(1 + \partial y/\partial x)$$

Taking the appropriate derivatives we obtain

$$\partial(\lambda/\lambda_o)/\partial t = \partial^2 y/\partial t\,\partial x = -V(\partial^2\eta/\partial x^2) = (V/G)(\partial^2[\Delta T(\lambda)]/\partial x^2)$$

The latter equation can be rewritten into the form

$$\partial\Lambda/\partial\tau = \{\partial/\partial x\}\,[D(\Lambda)\{\partial\Lambda/\partial x\}]$$

where $\Lambda = \lambda/\lambda_C$ and $D(\Lambda)$ plays the role of a Λ dependent diffusion constant

$$D(\Lambda) = (V/G)(d/d\lambda)\Delta T(\lambda) = [1 - (1/\Lambda^2)](AV^2/G)$$

and λ_C is defined by the relation $(B/AV)^{1/2}$ (i.e. setting the derivative of $\Delta T(\lambda)$ with respect to λ equal to zero and equating the value of λ to λ_C at this condition).

When $\Lambda < 1$, $D(\Lambda) < 0$. Langer noted that when Λ is slightly larger than unity and a fluctuation has caused Λ to drift to slightly subcritical values in some

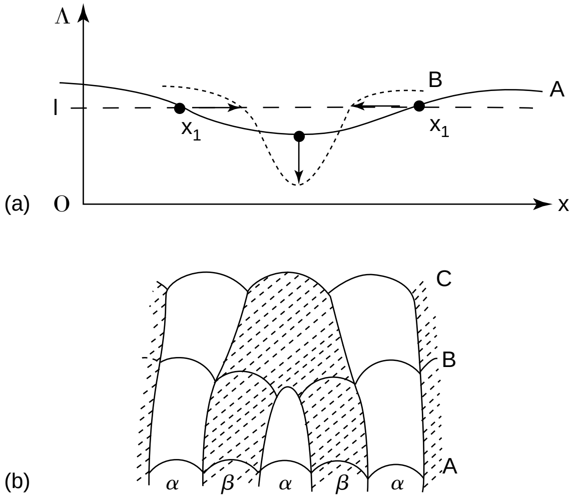


Figure 12.18. From Phys. Rev. Lett. 44, 1023(1980) with permission.

finite region, as illustrated in Figure 12.18, then at the minimum in this curve $\partial\Lambda/\partial x = 0$, $\partial^2\Lambda/\partial x^2 > 0$ and $D < 0$: thus Λ decreases. Also, at $\Lambda(x_1, t) = 1$, it can be shown by differentiating this relation that

$$dx_1/dt = -(\partial\Lambda/\partial t)/(\partial\Lambda/\partial x) = -2(\partial\Lambda/\partial x)_{X=X1}$$

Thus, the pair of points labelled x_1 in the figure approach each other. The resulting behavior is indicated by the arrows and dashed curve in Figure 12.18(a). All the intensity of an initially diffuse and shallow fluctuation is concentrated at a point. When Λ touches zero, the lamella at that point disappears, and the equation of motion loses its validity. The physical system presumably reverts to a state with fewer lamellae and larger average Λ, as illustrated in Figure 12.18(b), from which configuration the entire process must start all over again.

Akamatsu et al.[34] have corrected this analysis noting that in addition to their being a D value for motion perpendicular to the solidification front $D_\perp$, there is a contribution for motion of the triple junctions parallel to the solidification front $D_{||}$. Thus, a term is added to the previous relation for $\partial y/\partial t$ to yield

$$\partial y/\partial t = -V\ \partial\eta/\partial x + D_{||}\partial\lambda/\partial x/\lambda_o$$

where λ_o is the interlamellar spacing at the front being perturbed.

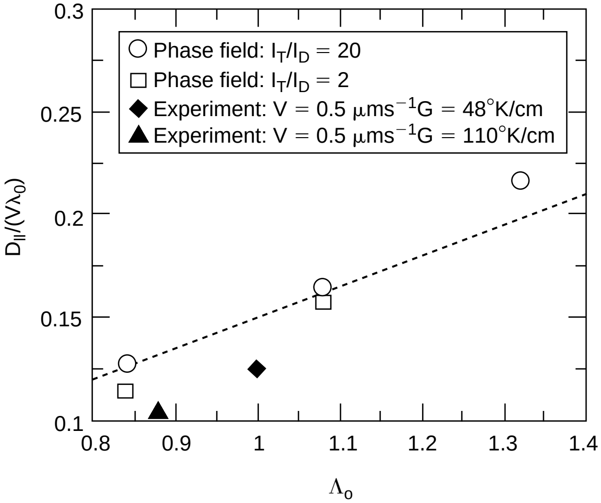


Figure 12.19. Results from phase field simulation and experiment showing dependence of $D_{||}$. ($\Lambda_0 = \lambda/\lambda_{in}$). From Phys. Rev. E 66, 030501(2002).

Akamatsu et al. showed that this relation then yielded the same relation for Λ as before, i.e.

$$\partial\Lambda/\partial\tau = \{\partial/\partial x\}[D(\Lambda)\{\partial\Lambda/\partial x\}]$$

with now $D = D_\perp + D_{||}$. Further, their study of $D_{||}$ yielded the dependence shown in Figure 12.19. They also evaluated a minimum value for the interlamellar spacing corresponding to a zero value for the modified $D = D_\perp + D_{||}$. This value of $\lambda_{min} = 0.7\lambda_C$ for their experimental situation.

Many suggestions have been made as to the factor that limits the maximum interlamellar spacing.[37] It is known that the morphological stability studies prior to 1990 yield a maximum spacing much larger than the experimental maximum limit. Similarly, the Jackson–Hunt ad hoc definition for the maximum spacing as that corresponding to an infinite slope of the interface between one of the phases and the liquid yields too large a value as compared to the experimental value of this limit. The closest agreement between a proposed condition for this limit and experimental values reported in the literature prior to 1995 has been obtained by Liu and Elliott[38] who numerically solved the Laplace equation for diffusion in the liquid state for non-isothermal and non-isoconcentration conditions along the parent/product interface under forced-velocity external constraint. They found that they could not obtain self-consistent solutions* for interlamellar spacings above some maximum value. They suggested that this maximum value defines the upper limit of interlamellar spacings obtained experimentally in steady-state forced-velocity eutectic solidification. The maximum limit they predicted for the Pb–Sn system is revealed in Figure 12.16 as the maximum interlamellar spacing shown for the dashed curve, which corresponds to their calculated curve for $\Delta T(\lambda)$, i.e. their calculated $\Delta T(\lambda)$ curve ends at this maximum value as λ increases. The solid circles in the figure correspond to Jackson–Hunt predicted maximum interlamellar spacings. The filled square points correspond to the mean experimental values of the interlamellar spacing. They obtained similar good agreement between their calculated interlamellar spacings (minimum, mean, maximum) for other eutectic systems also (Al/Al_2Cu and CBr_4/C_2Cl_6). Thus, it would appear that their maximum limit corresponds to a solvability criterion, albeit not one that results from an analytical derivation. Inasmuch as Brattkus has been reported[39] to have solved the non-isothermal coupled diffusion problem analytically and has not found any solvability type limits to the range of interlamellar spacings the question arises as to whether the inability of the Liu–Elliott method to obtain solutions above a maximum interlamellar spacing is real or an artifact of their method. We suggest that what Liu–Elliott discovered represents a real phenomenon. For forced velocity conditions we will demonstrate for a simple model that beyond a maximum interlamellar spacing it is not possible to have sufficient atom transport to maintain a steady state with the production of a supersaturated alpha phase. This model assumes that the fastest mode of atom transport is along the liquid–solid interface in a zone δ thick. The conservation relation for solute along this zone ahead of the alpha–liquid interface is

$$\partial J/\partial y = (V/\delta)(C_e - C_\alpha) = -(D\partial^2 C(y)/\partial y^2)$$

where C_e is the eutectic composition, C_α is the composition in the alpha phase, V the forced velocity of the liquid–solid interface, y is the distance along the alpha/liquid interface normal to the lamellae with the origin at the center of the alpha lamella and D the solute diffusivity in the interface zone of thickness δ. The relation for $C_i(y)$ that satisfies the boundary conditions is of the form $C_i = K_1 y^2 + K_2$. The boundary conditions are at $y = 0$, that $\partial C/\partial y = 0$ and at $y = \lambda_\alpha/2$, that $C = C_e$. Thus, substituting we find $K_1 = -(V/2D\delta)(C_e - C_\alpha)$ and $K_2 = C_e(1 + (V\lambda_\alpha^2/8D\delta)) - (V\lambda_\alpha^2/8D\delta)C_\alpha$. We thus find that at $y = 0$ that the interface solute concentration C_0 is given by

$$(C_0 - C_e)/(C_e - C_\alpha) = (V\lambda_\alpha^2/8D\delta)$$

---

* Self-consistency is defined by these authors to exist when it is possible to obtain a convergent solution by an iterative method.

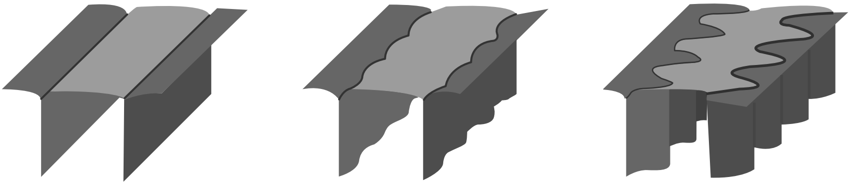


Figure 12.20. A series of snapshots from a 3D phase field simulation illustrating the transverse instability that limits steady-state eutectic growth at large spacings. From J. Met., p. 28, April 2004 with permission.

Since C_e and C_α are assumed fixed by the equilibrium diagram then the previous equation shows that C_0 increases as λ_α increases. But, C_o cannot increase above the value corresponding to pure solute (i.e. $1/\Omega$). Hence, λ_α and λ have maximum values for the assumption that solidification yields near equilibrium concentrations in the solid products. At a given set of externally controlled conditions then the interlamellar spacing cannot increase indefinitely without the production of supersaturated alpha phase. We suggest that the physical basis for a maximum in the interlamellar spacing is that beyond this maximum interlamellar spacing atom transport is not able to maintain steady-state solidification. It is interesting to note that under free boundary conditions where only ΔT is constrained in the liquid far from the interface then according to the above considerations there should be no solvability type upper limit to the interlamellar spacing because the velocity can decrease to accommodate the increase in the interlamellar spacing.

Recent phase field simulations of forced-field eutectic solidification reveal an instability along the lamellae, as shown in Figure 12.20 for a 3D simulation. Studies[40] using the boundary integral method revealed a period preserving oscillatory instability (illustrated in this figure) for compositions close to the eutectic composition at twice the minimum interlamellar spacing for this mode of solidification. Reference to Figure 12.16 reveals that experimentally the maximum interlamellar spacing observed is about twice the minimum and that Liu and Elliott's calculated maximum interlamellar spacing is also about twice the observed minimum. It is plausible then that this period preserving instability arises from the inability of the system to maintain the required atom transport for a planar lamellar geometry as the interlamellar spacing increases beyond some maximum.* By manipulating the line corresponding to the maximum diffusion distance for a planar lamella (the vertical line in Figure 12.21) along the wavelike contour in Figure 12.21, as illustrated by the tilted line, one will notice that the

---

*A different view is held by J.D. Hunt in **Solidification and Casting**, Chapter 10, eds. B. Cantor and K. O'Reilly, Institute of Physics, Bristol, 2003.

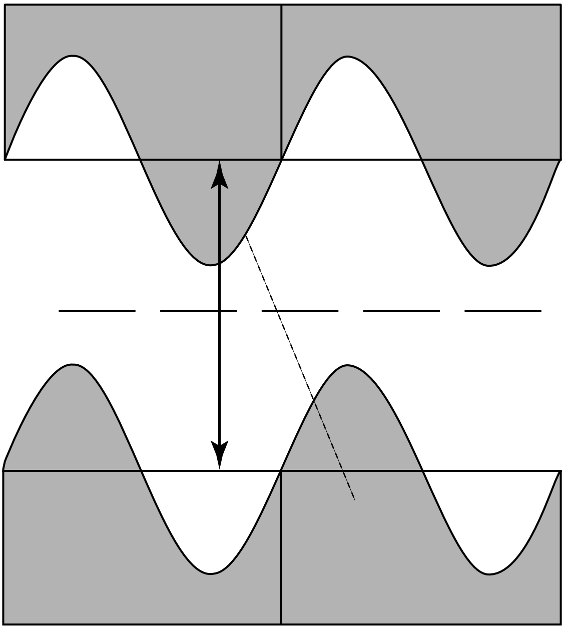


Figure 12.21. Schematic illustration of distances for diffusion to or from gray phase to white phase. Vertical bar terminated by arrows shows maximum diffusion distance in front interface when lamellae are planar. It is apparent that this distance is shorter when contour of lamellae approximates a periodic wave. Hence, under forced velocity condition the latter contour can supply the atom transport as lamellar thickness increases beyond that thickness at which planar lamellae can supply needed transport.

average maximum diffusion distance to or from one phase to the other is less when the interlamellar interface is wavelike than when it is planar. Hence, the instability that sets in when the interlamellar spacing for strictly planar interlamellar interfaces increases beyond that at which the needed atom transport under forced velocity conditions can no longer be supplied will give rise to the wavelike interlamellar contour so that the supply of the needed atom transport for that interlamellar spacing continues.

4. Summary

Summarizing, morphological instability in the shape of an interface between a growing phase and its environment can exist for a variety of reasons. If the environment consists of a supersaturated solution or a supercooled solution then such instability occurs whenever, the diffusion length or the thermal diffusion length exceeds the capillarity length. Also, with a chemical reaction at the interface, as may occur in CVD, the effect of the latter on the instability depends upon the sign of the heat of the reaction. In PVD and at low temperature, surface curvature drives morphological instability and surface diffusion acts to prevent it. Also, internal stress can drive morphological instability in thin films. This is not a complete catalogue of all the sources of morphological instability as a complete reading of the applicable literature will demonstrate. The tendency to morphological instability in solid phase transitions from metastable to stable phases is resisted by several factors, among which are lateral diffusion along an interface, interface immobility and strong interfacial energy anisotropy. Other than planar front morphological instability there is also morphological stability for a cellular pattern at a planar front, termed an Eckhaus instability, which reverses the dependence on perturbation wavelength relative to that for the Mullins–Sekerka case. The onset of transitions between microstructural patterns is also controlled by morphological instability considerations.

References

1. W.W. Mullins and R.F. Sekerka, J. Appl. Phys. 35, 444(1964).
2. J.S. Langer and H. Muller-Krumbhaar, Acta Metall. 26, 1681, 1689, 1697(1978).
3. R. Trivedi and W. Kurz, Acta Metall. 34, 1663(1986).
4. S.R. Coriell and G.B. McFadden in ibid, p. 785.
5. B. Billia and R. Trivedi in ibid, p. 899.
6. C.H.J. van den Brekel, Philips J. Res. 33, 20(1978).
7. D.J. Srolovitz, A. Mazor and B.G. Bukiet, J. Vac. Sci. Tech. A 6, 2371(1988).
8. R.T.U. Karunasari, R. Bruinsma and J. Rudnick, Phys. Rev. Lett. 62, 788(1989).
9. R. Asaro and W. Tiller, Metall. Trans. 3, 1789(1972); M. Grinfeld, Sov. Phys. Dokl. 31, 831(1986); D.J. Srolovitz, Acta Metall. 37, 621(1989).
10. W.W. Mullins and R.F. Sekerka, J. Appl. Phys. 34, 323(1963).
11. R. Trivedi and A. Karma, Encyclopedia Appl. Phys. 23, 441(1998);
12. G.P. Ivantsov, Doklady Akad. Nauk. SSSR 58, 567(1947).
13. D.A. Kessler, J. Koplik and H. Levine, Phys. Rev. A 31, 1712(1985).
14. D.A. Kessler and H. Levine, Phys. Rev. Lett. 57, 3069 (1986).
15. J.W. Cahn and J.E. Hilliard, J. Chem. Phys. 28, 258–267(1958).
16. S.M. Allen and J.W. Cahn, Acta Metall. 27, 1085(1979).
17. A. Karma and W.J. Rappel, Phys. Rev. E. 53, R3017(1996); Phys. Rev. E 57, 4323(1998).
18. Y.-T. Kim et al., Phys. Rev. E 59, R2546(1999); E 62, 2471(2000).
19. Frédéric Gibou et al., J. Sci. Comput. 19, 183(2003).
20. J. Bragard, A. Karma, Y.-H. Lee and M. Plapp, Interface Sci. 10, 121(2002).
21. J. S. Langer, Phys. Rev. A 36, 3350(1987); E. Brener and D. Temkin, Phys. Rev. E 51, 351(1995); U. Bisang and J.H. Bilgram, Phys. Rev. Lett. 75, 3898(1995); U. Bisang and J.H. Bilgram, Phys. Rev. E 54, 5309(1996).

22. A. Barbieri and J.S. Langer, Phys. Rev. A 39, 5314(1989).
23. A. Karma, Y.H. Lee and Mathis Plapp, Phys. Rev. E 61, 3996(2000).
24. J.C. LaCombe, M.B. Koss, J.E. Frei, C. Giummarra, A.O. Lupulescu and M.E. Glicksman, Phys. Rev. E 65, 031604(2002).
25. Y.-T. Kim, PhD Dissertation, Physics, University of Illinois, Urbana, 2003.
26. A. Karma, Phys. Rev. Lett. 87, 115701(2001).
27. B. Echebarria et al., Phys. Rev. E 70, 061604(2004).
28. D.J. Wolkind and L.A. Segel, Phil. Trans. Roy. Soc. 268A, 351(1970); B. Caroli, C. Caroli and B.J. Roulet, J. Phys. 43, 1767(1982).
29. K. Brattkus and S.H. Davis, Phys. Rev. B 38, 11452(1988).
30. B. Billia, H. Jamgotchian and R. Trivedi, J. Cryst. Growth 106, 244(1990).
31. K.A. Jackson and J.D. Hunt, Trans. Metall. Soc. AIME 236, 1129(1966).
32. J.S. Langer, Phys. Rev. Lett. 44, 1023(1980).
33. B.E. Sundquist, Acta Metall. 16, 1413(1968).
34. S Akamatsu et al., Phys. Rev. E 66, 030501(2002).
35. V. Datye and J.S. Langer, Phys. Rev. B 24, 4155(1981).
36. J.W. Cahn referred to in Reference 31.
37. W. Eckhaus, **Studies in Nonlinear Stability Theory**, Springer, Berlin, 1965; J.S. Langer and L.A. Turski, Acta Metall. 25, 1113(1977); G. Dee and R. Mathur, Phys. Rev. B 27, 7073(1983).
38. J. Liu and R. Elliott, Acta Metall. Mater. 43, 3301(1995).
39. K. Brattkus and C. Misbah, Phys. Rev. Lett. 64, 935(1990).
40. K. Kassner and C. Misbah, Phys. Rev. A 44, 6533(1991); A. Karma and A. Sarkissian, Metall. Mater. Trans. A 27, 635(1996).

Bibliography

1. W. Kurz and D.J. Fisher, **Fundamentals of Solidification**, Trans. Tech Publications, Switzerland, 1986.
2. H. Eugene Stanley and N. Ostrowsky, eds., Random Fluctuations and Pattern Growth, Proceedings of NATO Advanced Study Institute, NATO Advanced Science Institute Series E, Applied Sciences 157, Kluwer Academic Publishers, Hingham, MA, 1988.

CHAPTER XIII

Thermodynamics, Kinetics and Patterns

We have already in Chapter XII become acquainted with several types of instability-induced patterns: dendrites, concentration cells, eutectic lamellae. However, we have not provided any reason for expecting such patterns other than they are a manifestation of morphological instability.

Patterns exist in natural phenomena. Thermodynamics, via stability considerations, governs one group of such patterns. However, as noted in the previous chapter, other patterns are found which have no origin in thermodynamics since they are associated with kinetic processes. Still other patterns depend upon both thermodynamics and kinetics. We will explore the bases of these classes of patterns in this chapter. Our discussion of patterns will be brief and omit many classes of patterns due to the limitation of space. Nevertheless, we will provide an understanding of the origins of the class of spatial periodic patterns in Nature.

1. Spatial periodic patterns having a thermodynamic origin

For systems of particles at thermodynamic equilibrium, periodicity occurs because the periodic array of the particles is the array having the minimum free energy. Crystals represent one class of matter for which the ground states of minimum free energy are periodic arrays. That periodic arrays of minimum free energy can exist when the interaction between units is of repulsive character only is not obvious. A simple proof can be given which shows that a periodic array is the minimum energy configuration for the case of a concave upwards repulsive potential between particles. This is carried out in Appendix 1 of this chapter. Figure 13.1 is believed to be an example of

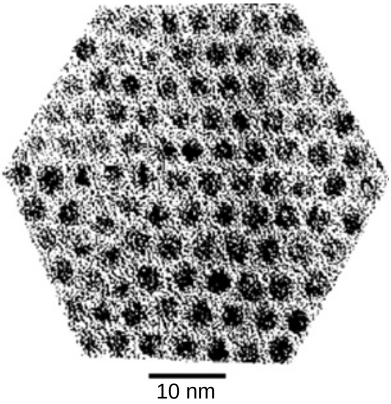


Figure 13.1. TEM image of compressed film containing 27A diameter Ag particles. From Science 277, 1978(1997) with permission. © 1997 Science Magazine.

such an array due to repulsive forces alone and an inward pressure, or to the equivalent attractive and repulsive interactions.

1.1. Dipole patterns

One class of periodic arrays may be described as of dipole origin. Seul and Andelman[1] were the first to point attention to the fact that patterns consisting of periodic stripes or hexagons were common to a large variety of different systems. In a two-component system in which the heat of solution is positive, we found in Chapter II that the corresponding phase diagram was Figure 2.7. Now suppose that for one of the components in such a system a long-range repulsion exists between the units of this component. The energy associated with a repulsive interaction can be decreased by subdivision of the domains of this component into smaller units and separation of these units. This procedure, however, introduces new interfaces between the two phases, which acts to increase the free energy. The overall free energy of the system at the minimum then involves a balance of the decrease in repulsive interaction energy and the increase in interface energy. The phase diagram corresponding to the minimum free energy of this system is that shown in Figure 13.2(a). This phase diagram modifies that of Figure 2.7 to show the stable regions of the periodic patterns of the two phases within the original miscibility gap.

A theory describing the free energy of these periodic patterns on the assumption that one of the components has a long-range repulsive potential equivalent to that between dipoles has been given by Keller et al.[2] Here we will provide a graphic illustration of the decrease in the electrostatic field energy that accompanies a division

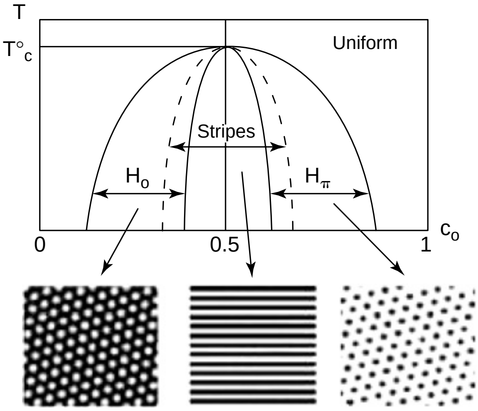


Figure 13.2a. Phase diagram expected for a system of dipoles corresponding to one of the two terminal components. The other component is a fluid.

of a patch of dipoles into two patches maintaining the same total number of dipoles and dipole density. As shown in Figure 13.2b the volume of the electrostatic field decreases upon such a division of a patch of dipoles into two patches. The electrostatic interaction energy is proportional to this volume.

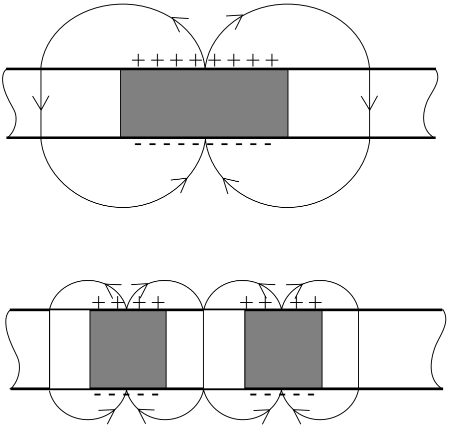


Figure 13.2b. Schematic diagram showing effect of subdividing dipole patch on volume of electrostatic field.

The explanation just given accounts for the tendency for dipole patches to subdivide into smaller dipole patches. It does not account for the periodicity of these dipole units. Appendix 1 accounts for the periodicity in that it proves that a repulsive interaction between equal units will produce a periodic arrangement at the minimum free energy. One can appreciate this result upon considering the forces between equal repelling units. Mechanical stability requires that the forces exerted on a unit be balanced.

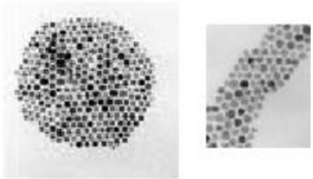


Figure 13.2c. Magnified images of spot and stripe revealing individual nanoparticle dipoles within each unit.

One may well ask how is it possible for the phase diagram to be that for a miscibility gap system when there is a repulsion between dipoles. An answer is that the miscibility gap will be produced when the average of the like interaction energies ($E_{11} + E_{22}$) is more negative than that for the unlike interaction E_{12}. Either there is a strong 1–1 attractive interaction, where the component 1 is not the dipole, or the interaction between component 1 and the dipole (2) is more strongly repulsive than that between dipoles. In any case, the interactions must be such as to yield a positive value for the heat of solution. Figure 13.2(c) show high magnification images of a stripe and a spot belonging to periodic arrays of dipoles. The close packing of the dipoles implies either a strong external pressure forcing them together or a short-range attraction between the individual dipoles. How the latter may occur and still exhibit a long-range repulsion between dipole domains is a mystery. The driving force for domain refinement must be either an effective repulsive force between like domain elements or an effective negative interface energy between unlike phases. In the former case, the force opposing the driving force for domain refinement would be the increase in positive interface area and energy. In the latter case, for the interface energy to be negative, E_{12} would have to more negative than $0.5(E_{11} + E_{22})$, a condition that violates the condition for a miscibility gap, i.e. $E_{12} > 0.5(E_{11} + E_{22})$.

Figure 13.3(a) shows a typical hexagonal periodic pattern in a system containing dipoles. Figure 13.3(b) shows a serpentine stripe pattern that is, nevertheless, periodic on a small scale, as revealed in this figure right side. A serpentine pattern is usually formed in an isotropic system as a result of near simultaneous initiation of

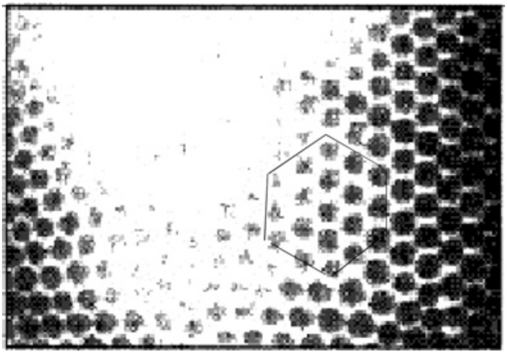


Figure 13.3a. Hexagonal bubble phase in Langmuir film composed of amphiphilic molecules having permanent electric dipole moments. Distorted hexagonal cell outlined. From M. Losche and H. Mohwald, Eur. Biophys. J. 11, 35(1984), Figure 4 with permission.

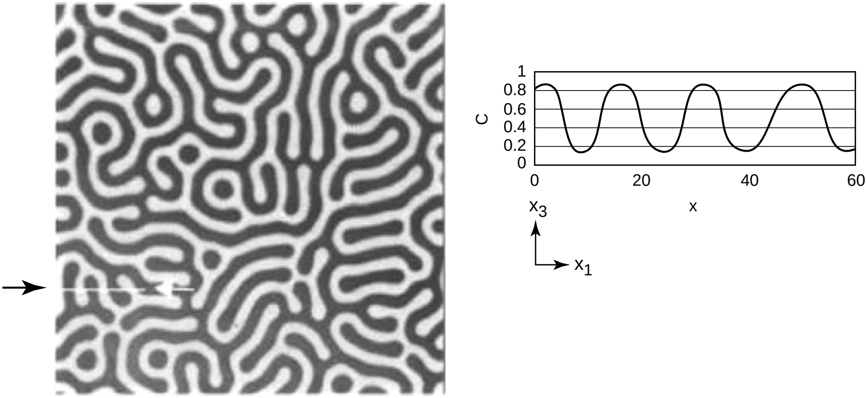


Figure 13.3b. Stripe phase showing periodic fluctuation of composition (right) along line between arrows in (left). From Phys. Rev. B 65, 085401(2002) with permission.
© 2002 American Physical Society.

the periodic pattern at many sites. When the phase transition yielding the pattern belongs to the class of spinodal decomposition then the final pattern will be periodic only if this periodic pattern represents a minimum free energy configuration.

This mechanism of forming periodic patterns is believed to operate in a number of systems. The following table lists a few of such systems.

Film system	**Order parameter**	**Non-local interaction**
Langmuir[1]	Dipole orientation	Electrostatic
Magnetic garnets[3]	Spin orientation	Demagnetizing energy
Type 1 superconducting[4]	Resistivity	Demagnetizing field

Suo and Lu[5] have provided a theoretical treatment of thermodynamic origin for pattern formation in constrained monolayer multicomponent films. Pattern formation in such films occurs when there is a thermodynamic force for phase separation, a positive interface energy driving phase coarsening and elastic strain energy driving phase refining. Physically, the phase refining derives from two sources. One is the strain energy in the solution phase which is a driving force for phase separation in the absence of coherency of the film to the substrate. The other stems from the coherency of the film and substrate interface which introduces stresses into the film's phases and the substrate. The volume in the substrate subjected to these coherency stresses decreases the smaller are the domains in the film inducing these substrate stresses. Consequently, the overall positive strain energy decreases by domain refinement. Effectively, these surface (film–substrate) stresses act as dipoles in much the same manner as the electric dipoles in the Langmuir film, or magnetic dipoles, and these act to refine the phases.

The serpentine-like stripe pattern of alternate phases illustrated in Figure 13.3(b) is a consequence of multiple points of initiation of the periodic pattern and the absence of a symmetry breaking force in the *homogeneous system prior to the pattern formation.* However, if symmetry of the homogeneous system is broken in some way, as for example by an external uniaxial stress, then the pattern can assume some order, as shown in Figure 13.4. This order can be increased by control over the symmetry breaking mode. The group of Suo have produced a series of papers on this subject which are worthwhile

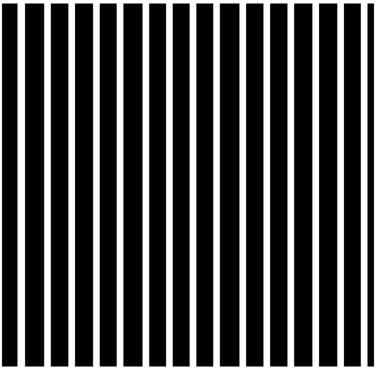


Figure 13.4. Stripe pattern on isotropic surface with uniaxial stress at equiatomic composition. From Acta Materialia 50, 2297(2002) with permission.

reading.[5] Their work may provide explanations for the observations of patterns in monolayer films of various sorts.[6]

Constrained films with internal stress are produced not only by deposition alone of films having slightly different lattice parameters as compared to those of the substrates, but also by phase transformations of films deposited at high temperature and then cooled below the transition temperature. Ferroelectric films so produced also yield a periodic array of twins.[7] Here again elastic stress is the domain refining force while domain boundary energy is the coarsening one. Further, the minimum free energy array is the periodic one.

One way of understanding the patterns produced by dipole domains is to consider the dipole domain as a source of a field of energy in a volume that is related to the area of the dipole array. For example if the dipole domain occupied a length L and a width w, with $w < L$, then the volume enclosing the space in which the specific energy exceeded some value, say E*, roughly equals Lw^2. For the same dipole density and dividing the width w into n domains of width w/n, the volume enclosing the space in which the specific energy exceeded E* would be $L(w/n)^2n$ or n times smaller. Hence, refining the domain size of dipole type domains lowers the energy due to the dipoles stored in the field.

1.2. Modulated surface patterns

Periodic ripple patterns also appear in Langmuir films of two components when there is a coupling between the composition of the components and the film curvature, as might occur when one component is hydrophilic and the other hydrophobic. The free energy of the film will contain the interaction energy between particles in the film, the energy of interfaces between phases in the film, both in the free energy F_X, and also one that includes the energy due to the surface curvature, the stiffness of the surface and the coupling term between composition and curvature, namely

$$F_c = \int d^2r[0.5\sigma|\nabla h(\mathbf{r})|^2 + 0.5\kappa|\nabla^2 h(\mathbf{r})|^2 + \Lambda\Phi(\mathbf{r})\nabla^2 h(\mathbf{r})]$$

where $h(\mathbf{r})$ is the height profile of the film relative to a flat surface, σ is its surface tension and κ is its bending modulus; Λ is the term that measures the strength of the coupling between the local composition Φ and the local curvature $\nabla^2 h(\mathbf{r})$. Minimization of the total free energy, $F_X + F_c$, with respect to the film shape $|h(\mathbf{r})|$ yields an effective free energy that depends only on $\Phi(\mathbf{r})$

$$F - F_x = \int d^2r[0.5b'|\nabla\Phi|^2 + 0.5\Lambda^2\kappa(\nabla^2\Phi)^2/\sigma^2] = \int d^2q\Phi_{-q}G(q)\Phi_q$$

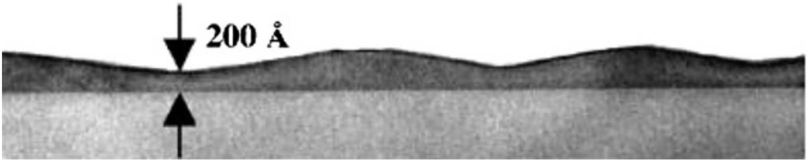


Figure 13.5. Cross-sectional TEM of a $Si_{0.8}Ge_{0.2}$/Si layer deposited at 630°C. From Phys. Rev. B 59, 1990(1999) with permission. © 1999 American Physical Society.

where $b' = b - \Lambda^2/\sigma$ and $G(q) = 0.5b'q^2 + (\Lambda^2\kappa/2\sigma^2)q^4$ is the 2D Fourier transform. A negative b' signals the onset of a curvature instability of the film. The characteristic domain size d is given by $[(\Lambda^2\kappa/\sigma^2)/|b'|]^{1/2}$.

This concept is applicable to the explanation of patterns in two component biological (bilayer) membranes and amphiphilic monolayers.[8] Another theoretical treatment of this system can be found in Reference 9.

Stress in a surface film that is coherent with the substrate can be decreased by a surface modulation that increases the surface area at constant volume of the substrate–film system. This decrease in stress occurs because there is an increase in free surface at which stresses normal to the surface are zero. See Figure 1.16 in Chapter I and Figure 13.5. Coupling of composition with such modulations also has been found such that when the film stress is compressive then constituents having smaller atomic diameter than the mean in the film concentrate in the hollows and those atoms having diameter larger than the mean concentrate along the hills. These and other driving forces to increase surface area that result from a concomitant decrease in free energy are opposed by an increase in free energy due to the stiffness of the surface. The latter increase in free energy is proportional to the square of the surface curvature.

There are many shapes the modulation can take and satisfy the constraint of increase in surface area at constant volume. Why take the periodic wave as the equilibrium shape? The answer is that periodicity minimizes the energy at constant surface area and volume. The periodic waveform that minimizes the positive energy due to the curvature at constant volume and surface area is the shape Nature chooses. Thus, we have again a situation where the periodic pattern provides the minimum energy configuration for the system at thermodynamic equilibrium.

1.3. Short-range repulsion

As already noted a repulsive interaction between particles can bring about a periodic distribution of them. The effects of a single step short-range repulsive potential on pattern development has been explored in a simulation.[10] This single step potential consists of an infinite repulsive potential below a radius r_o and a constant finite

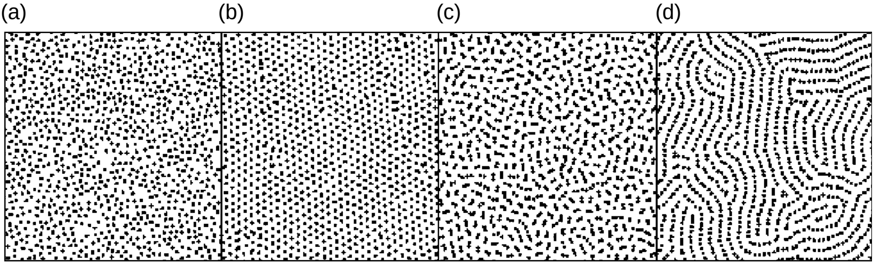


Figure 13.6. Snapshots of particle distributions at various particle densities: (a) 0.1, (b) 0.15, (c) 0.227 and (d) 0.291. Density is in units $1/(2r_o)^2$. From Nature Materials 2, 97(2003) with permission. © 2003 Nature Publishing Group.

repulsive potential between r_o and r^* and a zero value for $r > r^*$. Figure 13.6 shows the patterns found on minimizing the energy for increasing numbers of particles per unit area. As shown, when the particle density is less than $(2r^*)^{-2}$ the particle array is random (Figure 13.6(a)). (In this range of particle separation there is no repulsive interaction between particles.) When the particle density is such that each particle is about a distance 2r* of another particle the lowest-energy pattern is a periodic one based on a triangular lattice (Figure 13.6(b)). (In this case only a periodic pattern allows no interparticle distance smaller than 2r*.) For higher density, clusters of particles that are separated from other clusters by more than 2r* represent the low-energy configuration. This trend is continued with higher particle density forming stripe-like patterns of close-packed particles where the stripes are separated from neighboring stripes by more than 2r*. (Within the stripes the distance between particles can vary between r_o and r* at constant energy.) Comparison of this result to that for the dipole suggests that the pattern may be sensitive to the interparticle potential.

1.4. Bulk matter

Diblock copolymers exhibit patterns in which strands of one polymer per lamella alternate normal to the interface between lamellae.[11] In this system the domain refining force is the configurational entropy decrease associated with the formation of the domains, i.e. in the mixed copolymer the strands are randomly oriented whereas in the ordered lamellar structure the strands are aligned normal to the lamellar interfaces. This decrease in configurational entropy increases with increase in the interlamellar spacing. The domain coarsening force is again the lamellar interface energy which varies inversely as the interlamellar spacing. Although this model provides a minimum in the free energy for a configuration consisting of lamellae at some interlamellar spacing it does not provide a configuration that is more stable than a copolymer solution. There must also be an interaction energy which favors like over unlike nearest-neighbor strand contacts. However, this energy which stabilizes copolymer strand unmixing is independent of the interlamellar spacing and hence has no effect on the pattern.

The above discussion by no means exhausts the thermodynamic origins of patterns. For example, hard sphere colloids form fcc crystals where the thermodynamic origin of this periodic array of hard spheres is entropic and not energetic. We have examined this phenomenon in Section 2.5 of Chapter I. There it was found that above some critical volume fraction of hard spheres that a fcc crystalline array had a larger entropy than a random array. This contrarian notion is a consequence of the fact that the increase in translational entropy overcomes the decrease in configurational entropy in the transition from a random array to the fcc array. The latter has the larger free volume which the centers of the hard spheres are able to explore. As noted there the examples of such entropy-induced patterns can fill a book. Most of these examples are in the biological sphere.

Here we will consider another entropic driving force for pattern production, which is the "depletion force" due to a decrease in the excluded volume. Such reduction in excluded volume may occur in mixtures of colloidal particles having different shape or size. The Helmholtz free energy decreases by $\pi\Delta V$, where π is the osmotic pressure and ΔV is the reduction in excluded volume. The excluded volume is defined in Figure 13.6 where the mixture consists of spheres of radius a and polymer coils having a radius of gyration R_g. The white area around the spheres corresponds to the volume excluding polymer centers while the dark region between spheres where the excluded volumes of the two spheres overlap corresponds to the reduction in the excluded volume. The osmotic pressure π tends to force the spheres together to decrease the excluded volume (increase the free volume available to the polymer coils). This effect stems from the entropy increment of the polymer coils due to an increase in the free volume available to them. Alternately, one may describe the effect of the increase in entropy due to the reduction in excluded volume in terms of a depletion potential acting between the two large spheres. The depletion potential has been measured using an optical tweezer device.[12] Figure 13.7(a) and (b) shows one such potential between two large hard spheres induced by a small volume fraction of smaller hard spheres, for two volume fractions. The

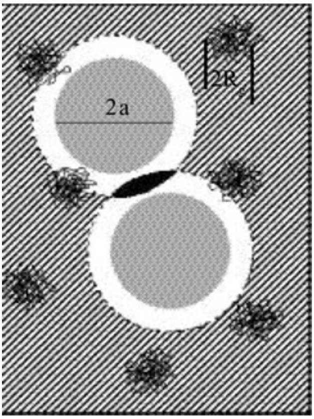


Figure 13.7a. Schematic diagram showing two large colloidal particles. The white region about each is the volume excluded to the centers of the small polymer particles having a radius of gyration R_g. The black region between the large spheres is the overlap region.

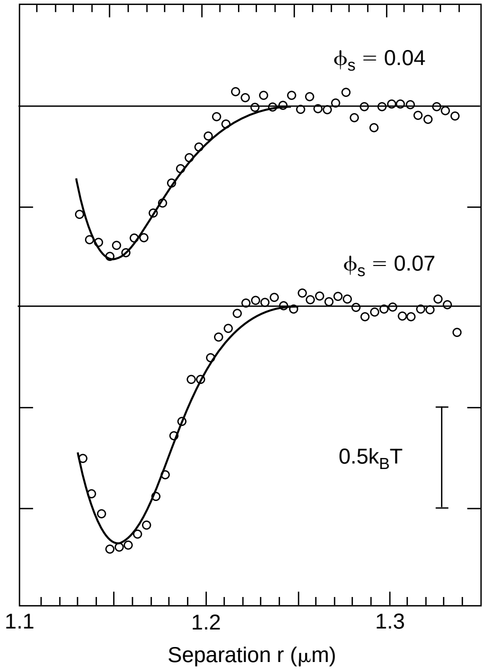


Figure 13.7b. Depletion potential measured between two hard spheres for two volume fractions of small spheres. From Phil. Trans. Roy. Soc. London A 359, 921(2001) with permission. © 2001 The Royal Society.

lines through the points correspond to the much earlier predictions of Asakura and Oosawa.[13]

One might enquire as to whether the depletion potential can bring about phase separation of a fluid into a fluid and solid mixture in a bidisperse system. This particular question has been addressed in the work of Reference 14 with the answer that indeed the entropy increment of the small spheres due to reduction of their exclusion volume can bring about the stabilization of solid phases in the bidisperse hard sphere system. In particular, Xiao et al.[14] have calculated the critical volume fraction above which the packing together of large spheres to form a solid increases the entropy of the system for values of the size ratio $\alpha \leqslant 0.33$. This critical volume fraction is shown in Figure 13.8.

We now have encountered two entropic sources for the formation of solids. One is the increment in entropy of translation in an ordered solid as compared to a fluid of the same particle density which becomes appreciable at high particle densities. The other is the one just described in which the entropy increment due to the reduction of excluded volume induces a depletion potential tending to

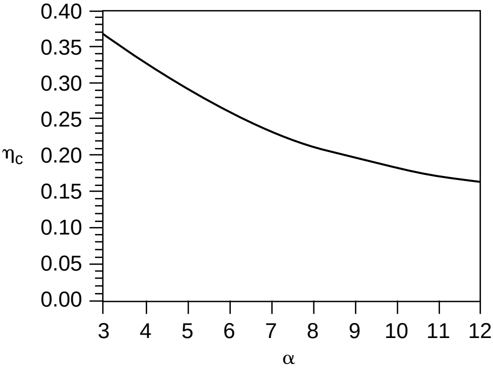


Figure 13.8. The predicted critical volume fraction versus the ratio of large to small particle radius at 300°K. From Phys. Rev. E 64, 011402(2001) with permission. © 2001 American Physical Society.

produce a solid of higher particle density in which the reduction of excluded volume is maximized. Although there is some ambiguity in the literature concerning the separate identities of these entropic increments it appears to this author that the two entropic increments are due to physically different phenomena and will give rise to different solid phases. In particular, as already suggested the depletion potential acts to bring about the largest reduction in excluded volume and acts as a short-range attractive force between large spheres, whereas the increment in entropy of translation of the large spheres tends to provide an ordered solid with the largest free volume consistent with the external constraints and is also equivalent to a short-range attractive force. Indeed, one may conceive of a situation in an ordered solid where the reduction of excluded volume for the small spheres is accompanied by a reduction in the free volume available for translation in the solid of the large spheres. Thus, the solid phases that appear in a bidisperse system at equilibrium may well involve an optimization of the two entropy factors.

The depletion force is involved in the stability of the AlB_2, $NaZn_{13}$, NaCl and NiAs structures found in bidisperse hard sphere systems. It is also involved in the various periodic patterns produced in rod–sphere systems. Among the latter is one shown in Figure 13.9.

Much of the discussion in this section has concentrated upon the effect of entropy in *stabilizing* periodic patterns. Usually, entropy exerts the opposite effect of destabilizing periodic patterns as temperature increases. This entropic destabilization also corresponds to a Newtonian force that acts in the direction opposite to that due to interparticle potentials and external pressures that tend to decrease the distance between particles. At the melting point these two forces are equal and opposite in direction. We note that opposing forces acting on elements of the pattern

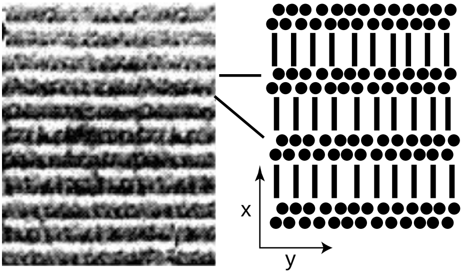


Figure 13.9. Periodic pattern produced in a hard rod–hard sphere mixture. The pattern on the left shows lamellae which when magnified appear as shown schematically in the pattern on the right.

are present whenever periodic patterns form. Further, we note that whenever a transition yields a periodicity in the resultant pattern then there has been a spontaneous break-in symmetry. (Such a break-in symmetry is not present in the gas to liquid transition, whereas the competing forces are acting in this phase transition.)

Incidentally, the proof that the fcc crystal pattern is the ground state pattern for conditions where spherical particles attempt to achieve the closest possible packing was only achieved recently by Hales.[15] He solved what is known as the Kepler conjecture that the fcc array is the closest possible packing of like size spheres. It has been shown in Chapter I that entropic and energetic influences tend to produce the closest packing of the spheres and hence the lowest free energy for the fcc structure – the closest packing array. This principle applies in many situations involving the patterns of spherical particles.

Copolymers and micelles form similar phase diagrams containing similar phases in 3D. The phases are all periodic arrays – crystalline arrays. These phases in the order of increasing surface area per unit volume (geometric shape) are: bcc (spheres), hexagonal (cylinders), gyroid (bicontinuous) and lamellar (planes). This is the same order in terms of concentration of micelles or copolymers. Also, one may notice a similarity to the 2D phases of Figure 13.2(a) – the hexagonal and stripe phases, which are the 2D patterns in the 3D arrays listed above. From the above we detect a correlation between the surface area per unit volume of the array and the concentration of the pattern units in the two-component system.

Summarizing, the periodic patterns found at thermodynamic equilibrium are, as expected, the minimum free energy patterns for the respective systems.

2. Spatial periodic patterns far from equilibrium

2.1. Dissipative crystals

Certain dissipative systems form stationary periodic patterns at steady-state conditions. These patterns can be obtained as possible solutions to non-linear equations describing the dynamics of these systems in that these solutions have certain

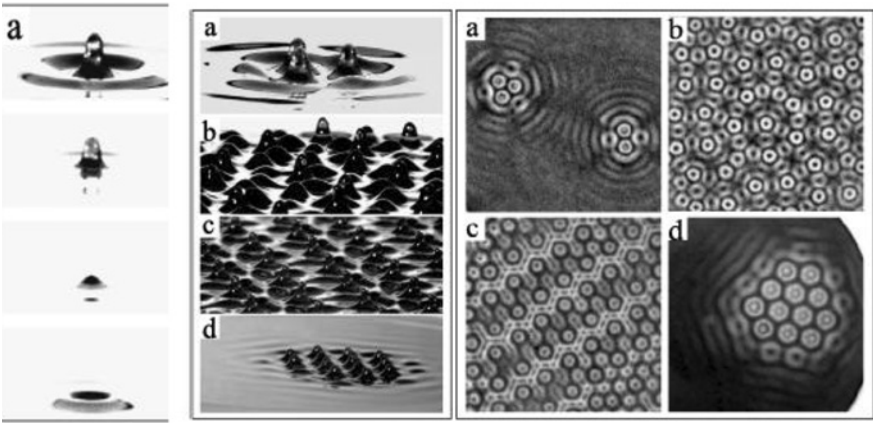


Figure 13.10. Left: (a) Photograph of a single oscillon in a fluid. Right: Vertical (a)–(d) are side views of various arrays of oscillons in a fluid. Corresponding top views are to the right. From Phys. Rev. Lett. 85(4), 756(2000) with permission. © 2000 American Physical Society.

symmetry properties also applicable to the patterns. The patterns appear experimentally, when the system under the influence of a change in some external control parameter undergoes an instability and transforms to a different state. It has been shown that for control parameters in the vicinity of the transition that the pattern found corresponds to a minimum of a free energy functional derivable from the dynamic equations. This result and another observation due to Swinney and his collaborators[16] suggest a new (really old) way of deriving phase diagrams for these systems which will be described below in the hope that it will inspire research in this direction. The normal way of treating the subject of patterns in dissipative systems involving the solution of non-linear equations describing the dynamics of these systems will be treated very briefly here for several reasons. One is that this would require too lengthy descriptions. Another is that there are already several monographs on the subject.[17–19]

For certain boundary conditions and at conditions close to the onset of the instability transition the patterns obtained are homogeneous over most of the system. These are patterns of dissipative units, which will be called dynons, in a crystalline array, which will be called a dissipative crystal. These dissipative crystals can be 2D or 3D.

Let us now consider the specific results that prove that one such dissipative crystal has positive values of shear elastic constants. This dissipative crystal consists of what Swinney has called “oscillons.” An image of an oscillon and arrays of oscillons found in a vertically vibrated fluid are shown in Figure 13.10. Oscillons form into periodic patterns. One such pattern is shown in Figure 13.11. What is more important is the fact that this dissipative crystal acts as does a harmonically coupled crystal. Swinney and collaborators evaluated the dispersion relation between the lattice frequency and phonon wave number for the (1, 1) normal mode from measured images over a sequence in time. They found that the measured relationship corresponded to the same one that would result from a calculation for a harmonically coupled lattice of particles. Further, they were able to make evident via images the normal modes of vibration of a square lattice of oscillons. One such mode is revealed in Figure 13.12. These modes are normal to the particle motions in the oscillons. As further support for the crystal aspect of these

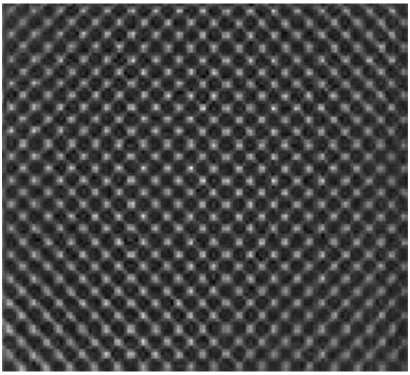


Figure 13.11. Square lattice of oscillating (25 Hz) bronze spheres 0.165 mm diameter. From **The Physics of Complex Systems**, eds. F. Mallamace and H.E. Stanley, IOS Press, Amsterdam, 2004 with permission.

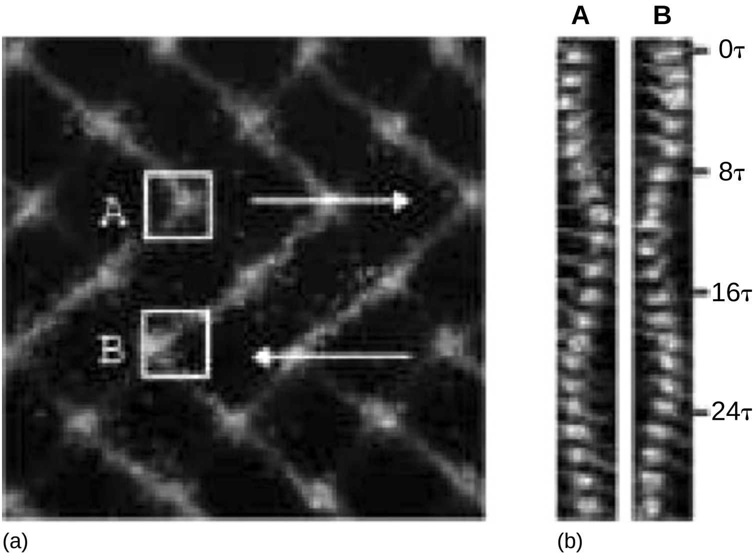


Figure 13.12. (a) Close-up shot of a granular square lattice ($\Gamma = 2.9$, f = 25 Hz, h = 4d). (b) Time evolution of the peaks of the boxes A and B as shown in image (a). The peaks oscillate out of phase with a frequency about 20 times smaller than f. From Phys. Rev. Lett. 90, 104302(2003) with permission. © 2003 American Physical Society.

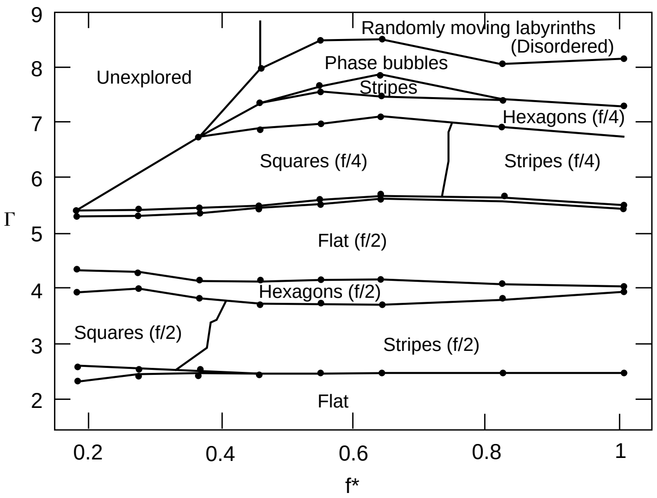


Figure 13.13. Phase diagram for patterns observed in a vibrating particle system. Bronze spheres d = 0.165 mm, layer depth 5d, container diameter 770d. Coordinates: dimensionless acceleration versus dimensionless frequency. From Phys. Rev. E 65, 011301(2002) with permission. © 2002 American Physical Society.

lattices of oscillons is the finding that the lattice could be made to melt upon decreasing the frictional resistance opposing the relative motion of particles and thereby increasing the amplitude of the normal modes, i.e. the entropy of vibration. This result suggests that the entropy of dissipative crystals may be an important factor in determining the relative stability of competing periodic patterns.

Figure 13.13 shows a type of phase diagram for the vibrating particle system with coordinates of dimensionless acceleration versus dimensionless frequency. In this system the control parameter for which changes lead to instabilities and pattern transitions is primarily the dimensionless acceleration with the dimensionless frequency only affecting the transition between a square pattern and a stripe pattern. The square pattern may be viewed as a pattern of perpendicular stripes such that the easiest horizontal vibration mode is the shear mode parallel to the stripes. For the stripe pattern this mode is also easy as is one involving waviness of the stripes. However, the hexagonal pattern appears to offer stiffer vibrational modes than for either of the square or stripe patterns, but, because of the closer packing of the oscillons, a greater cohesive energy for this pattern relative to the other two patterns. If the energy that can be diverted into horizontal vibrations decreases with increase in the vertical acceleration, perhaps the explanation of the stability of these patterns can be explained. This conjecture remains to be tested.

A dissipative crystal system similar to the vibrating particle system, but different from it to the extent that it involves a continuous media and consequently,

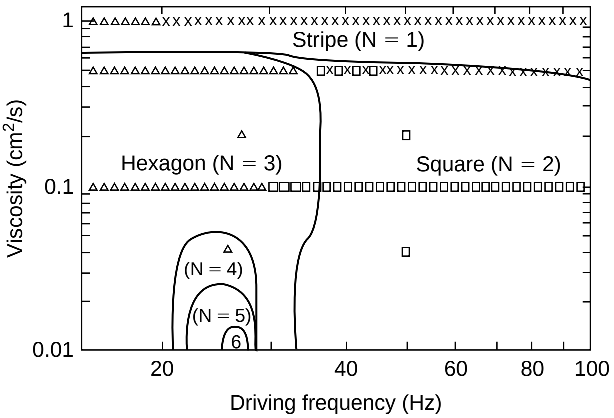


Figure 13.14. Predicted and observed patterns in Faraday wave systems in a map of viscosity versus driving frequency. □: square; Δ: hexagon; x: stripe experimental points; lines: theoretical. From Phys. Rev. Lett. 79, 2670(1997) with permission. © 1997 American Physical Society.

that a free energy function (Lyupanov function) can be written for it is the Faraday wave system.* The minima free energies of this function may be obtained and consequently, the corresponding patterns may be predicted.[20] This procedure has been carried out with the result shown in Figure 13.14. It should be recognized that the Lyupanov function is not in the usual form of an energy minus a temperature–entropy product. Rather it stems from the dynamical equations for the system. Hence, it is difficult to extract the energy and entropy contributions to the Lyupanov function from the data provided in the paper. Nevertheless, the patterns shown represent the minimum free energy (Lyupanov) arrays in their region of stability. These results contradict part of the previous interpretation regarding the vibrational qualities of the various patterns. First, since the stripe phase is stable at the highest viscosity one may conclude that vibrational entropy does not enter into its stability which is primarily determined by energetic considerations. Second, the stability of the square pattern at high frequencies where capillary waves are the driving oscillations suggests that the square pattern has the higher entropic contribution to its stability than the hexagonal pattern. This conclusion is the same as previously drawn. What is needed for both the vibrating particle and Faraday systems is a knowledge of the inter-oscillon potential. Here is another research problem.

The Rayleigh–Benard system provided Ciliberto and Caponeri[21] the possibility of calculating the system entropy as a function of the displacement from the transition temperature, which is shown in Figure 13.15 along with the entropy production as measured by the Nusselt number. It is apparent that the entropy of the system is constant in the region-denoted convection in Figure 13.15(a). This

---

* A vertically vibrating water in a pan system.

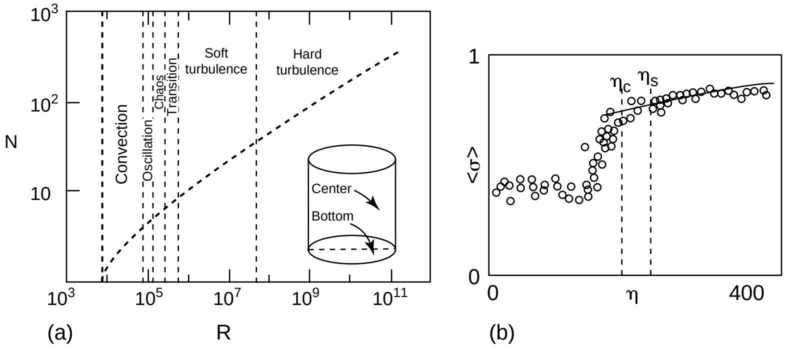


Figure 13.15. (a) Nusselt number vs Rayleigh number in gaseous He. From Phys. Rev. A 36, 5870(1987) with permission. © 1987 American Physical Society. (b) Measure of entropy of Rayleigh–Benard system in silicone oil. Transition to spatiotemporal intermittency at $\eta = 248$. From Phys. Rev. Lett. 64, 2775(1990) with permission. © 1990 American Physical Society.

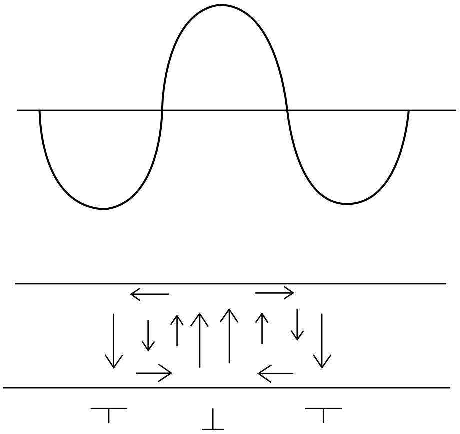


Figure 13.16. Top is plot of vertical velocity versus horizontal position. Bottom shows vectors of velocity in one cycle.

result is consistent with the fact that the pattern in this region for this calculation is the roll (stripe) pattern, such as shown in Figure 13.16. Thus, the roll or stripe pattern is a stationary stable pattern in this region. A transition occurs at $n = 248$ to spatiotemporal intermittency or turbulence in Figure 13.15(a). Defects begin to be generated and move at an earlier stage denoted the oscillation region in Figure 13.15(a). The doubling of the entropy over a small temperature interval may be due both to an increase in dislocation density and to enhanced motion of the dislocations. This correlation has not been explored in the literature.

The dissipative crystals most studied in the literature belong to the Rayleigh–Benard system.* One may find a book **The Dynamics of Patterns**[18] which examines the properties of crystals in this system from the viewpoint of solutions to the free energy functionals as obtained from dynamics. Here we are suggesting an alternate route to a description of dissipative crystals, which is at this time merely a potential route in that no one has traversed it as yet. However, since a free energy function that is minimized is involved in both concepts it may be possible to transform knowledge gained from the dynamics realm to that of the thermodynamic realm applicable to patterns at thermodynamic equilibrium. For this reason, the usual analysis of patterns in dynamical systems will be briefly described below.

The Ginzburg–Landau equation is often used in the analysis of patterns in both static and dynamical systems. The form the Ginzburg–Landau equation takes in this application is

$$d\varphi_j/dt = 2\mu\varphi_j - \sum_{l=1}^{n}\beta_{jl}|\varphi_l|^2\varphi_j \quad j = 1, \ldots, n$$

where φ is the amplitude of an order parameter, $w = 2\mathrm{Real}\,[\Sigma_{j-1}^{n}\,\varphi_j\,\exp(ik_jx)]$. A dynamical system that under the influence of some control parameter undergoes a transition and a spontaneous break-in symmetry from a homogeneous system to one with less symmetry will be described by the Ginzburg–Landau equation in the vicinity of the transition and the dynamics will obey a gradient flow where $d\varphi/dt = -dF/d\varphi$. It is this relation that yields the form of the Ginzburg–Landau relation given above when the free energy F is given by

$$F = -\mu\sum_{j=1}^{n}|\varphi_j|^2 + (1/2)\sum_{j,l=1}^{n}\beta_{j,l}|\varphi_j|^2|\varphi_l|^2$$

At a minimum in the free energy F solutions to the above equation will have the form $\varphi = \varphi_o e^{ikx}$, where $k = k_c$. Depending upon experimental conditions that define the symmetry and the parameter values there are several possible solutions. When there is a horizontal symmetry plane in the Rayleigh–Benard system then the lowest free energy corresponds to $\varphi_1 = \varphi_o e^{ikx}$, $\varphi_2 = \varphi_3 = \cdots = 0$, a roll pattern. With a quadratic term in the Ginzburg–Landau equation in the Rayleigh–Benard system the lowest free energy yields a hexagonal solution $\varphi_1 = \varphi_2 = \varphi_3 \neq 0$. Other dissipative systems may have other solutions at the lowest free energy corresponding to some combination of the various possible waves. Thus, the origin of the periodic patterns found in the vicinity of transitions from a parent homogeneous state in a dynamic system stems from the applicability of the gradient flow condition, which then yields the Ginzburg–Landau equation given above, and to have solutions which correspond to periodic patterns.

---

* The Rayleigh–Benard system consists of a thin layer of a fluid between two solid surfaces in which one surface is heated and the other cooled.

One may note that the above considerations provide no clue as to the wave numbers that will be found. These wave numbers are determined primarily by stability considerations in that only a certain range of wave numbers and Rayleigh numbers yield stable patterns at least in Rayleigh–Benard systems.

The remaining dissipative system that produces periodic patterns similar to crystalline ones is the reaction–diffusion system. The latter differs from the other three dissipative systems in that it involves a conservative variable of the system, namely the composition. The vibrating particle, Faraday wave and Rayleigh–Benard systems are dependent upon non-conservative variables, and hence can be described by Ginzburg–Landau type relations. The reaction–diffusion system requires a different equation, such as the Cahn–Hilliard equation, for its modeling. These two classes of systems differ in another aspect. The dynons of the order–disorder group are well-defined graphically. The oscillon is shown in Figure 13.10. The convecton is defined schematically in Figure 13.16. As a hexagonal convecton it appears as one of the white spots shown in Figure 13.17. In this example the flow is up through the center and down along the sides. As a roll convecton, it appears as one of the stripes shown in Figure 13.18, the rotation of the rolls alternate clockwise–anti-clockwise.

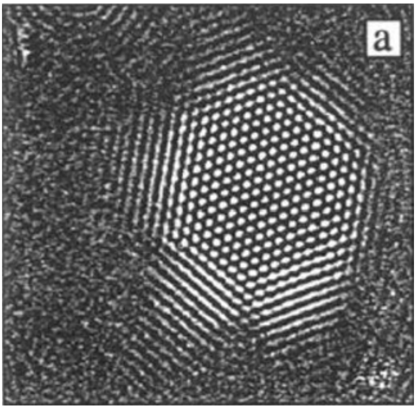


Figure 13.17. Snapshot of surface soon after nucleation of Benard pattern. From Phys. Rev. Lett. 67, 3078(1991) with permission. © 1991 American Physical Society.

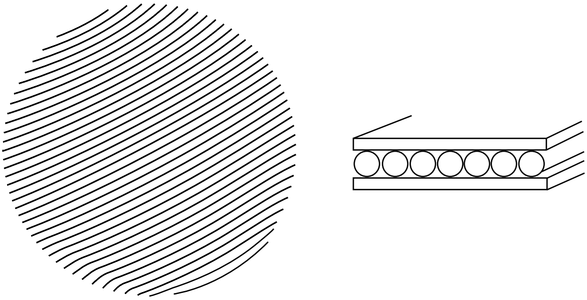


Figure 13.18. Roll pattern produced in a Rayleigh–Benard cell. Left pattern is looking down on rolls. Right pattern shows the rolls relative to the plates of the cell diagramatically.

For both the oscillon and convecton forces exist between neighboring units and each unit itself has a stiffness, so that the pattern of units can deform to satisfy boundary constraints, such as fitting an integral number of half wavelengths between two confining sides. Further, as for crystals, these dissipative crystals can host defects, such as dislocations. The physical dissipative unit for a reaction–diffusion system is more difficult to define and visualize. One way of accomplishing this visualization is to consider the elements involved in a reaction–diffusion system. The reactions of this system provide sources of one or more reactants and sinks for one or more of the products. The diffusion action of this system is that of diffusion of some component between the effective source and sink. When the rate of input of a reactant at a source equals the rate of its transport away to neighboring sinks and the rate of its conversion to a product and removal at these sinks then there is a possibility of developing a steady-state dissipative process involving a stable stationary array of sources, sinks and diffusing species. This stable array can be a crystalline array of sinks and/or sources. This general description applies to a range of reaction–diffusion systems including those describable as Turing[22] systems. The description just given is not that generally used in treatments of reaction–diffusion problems. These treatments are generally mathematical. However, by attempting to relate physical processes to the mathematics one arrives at the model just described. The sinks and sources need not be arrays of physical units. They can be merely regions where the actions of sink and source take place. For example, consider the following equations which represent one specific example of a reaction–diffusion system

$$\frac{\partial u}{\partial t} = 0.6u - v - u^3 + 0.0002\frac{\partial^2 u}{\partial x^2}$$

$$\frac{\partial v}{\partial t} = 1.5u - 2v + 0.01\frac{\partial^2 v}{\partial x^2}$$

The last terms in each equation represent the action of the diffusion process since they represent the divergence of the fluxes at the positions x. The other terms represent the effects of the concentrations of u and v on the reactions that supply and remove these components from the system. Starting with a random noise distribution of concentrations u and v in the system it evolves as shown in Figure 13.19. The concentrations u and v interact with reactants external to the system to enable u and v to be supplied to or removed from the system at any point x. The equations govern these reactions so that a steady state is achieved in time. At this steady state the situation produced is one that can be described in terms of sources of u and v at the peaks in the concentrations and sinks at the valleys in their concentrations. Actually, there can be a homogeneous distribution of the reactants that supply and remove u and v, but the action at steady state is such that the source action is mainly at the peaks of concentrations and the sink action is mainly at the valleys in these concentrations.

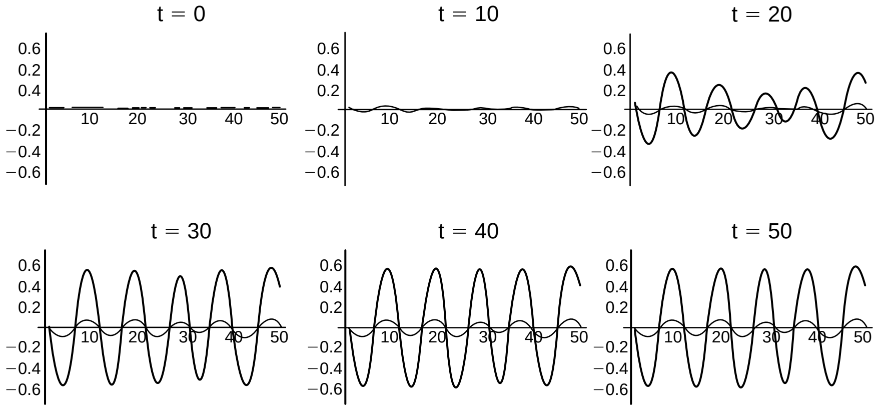


Figure 13.19. Values of u and v (see text) as a function of time. From Anatomic. Sci. Int. 79, 112(2004) with permission.

A reaction–diffusion system that exhibits the Turing pattern obeys the following general relations

$$\partial X/\partial t = D_X \nabla^2 X + f(X, Y)$$

$$\partial Y/\partial t = D_Y \nabla^2 Y + g(X, Y)$$

where f and g are two non-linear functions that describe the reaction kinetics. These equations have been investigated for a variety of choices of f and g.

There are several conditions that need to be satisfied for a Turing pattern to evolve:

- *The reaction needs to involve:*
  - *an activator (X) that acts in an autocatalytic manner to enhance its own production or exerts a positive feedback and*
  - *an inhibitor (Y) that induces a diminution of the activation process or acts to yield negative feedback.*
- *The inhibitor needs to diffuse more rapidly than the activator ($D_Y > D_X$).*

A general exposition of the Turing problem has been provided in terms of bifurcation theory by Judd and Silber[23] as follows. Transform the composition variables to: $u = X - X_0$, $v = Y - Y_0$ and set $D_Y/D_X = K > 1$ so that we can write the above equations as

$$\partial u/\partial t = \nabla^2 u + f(u, v)$$

$$\partial v/\partial t = K\nabla^2 v + g(u, v)$$

where for a 2D problem $\partial^2 = \partial^2/\partial x^2 + \partial^2/\partial y^2$. The Turing instability first occurs as a symmetry breaking steady-state bifurcation of a spatially uniform state.

Associated with the instability is a critical wave number $q_c \neq 0$, which is determined in the following.

The spatially uniform state is taken to be $u = v = 0$. The functions $f(u, v)$ and $g(u, v)$ are expanded about this state to yield

$$f(u, v) = au + bv + F_2(u, v) + F_3(u, v)$$

$$g(u, v) = cu + dv + G_2(u, v) + G_3(u, v)$$

where $F(G)_{2,3}$ are quadratic, cubic terms in the Taylor expansions. It can be shown that at the Turing bifurcation where the homogeneous state becomes unstable that the parameters ($a = a_c$, $b = b_c$, $c = c_c$, $d = d_c$, $q = q_c$, $K = K_c$) satisfy

$$q_c^2 = (K_c a_c + d_c)/(2K_c) > 0$$

$$(K_c a_c - d_c)^2 + 4K_c b_c c_c = 0$$

$$K_c a_c + d_c > 0$$

$$b_c c_c < 0$$

$$a_c + d_c < 0$$

$$a_c d_c - b_c c_c > 0$$

All Fourier modes with wave vector $|q| = q_c$ on the critical circle are neutrally stable spatially. It can be further shown upon consideration of the non-linear terms in the reaction–diffusion equations that stable solutions for which $q = q_c$ yield lattices based on either squares, hexagons or rhombs. The stable periodic patterns that result are either stripes, squares, rhombs, super squares or anti-squares on the square lattice, or stripes, simple hexagons, rhombs, super hexagons, or super triangles on the hexagonal lattice, which occur within limited values of coefficients associated with the non-linear terms in the reaction–diffusion equations. It is emphasized here that it is the non-linear terms in the reaction–diffusion equations that are responsible for the stationary patterns that may appear. It has been found that cubic interactions favor the formation of stripes (lamellae) whereas quadratic interactions favor the formation of spots (spheres) in a system of Turing equations in which $F_2(u, v) = -\nu Cuv$, $F_3(u, v) = -\nu uv^2$, $G_2(u, v) = -F_2(u, v)$, $G_3(u, v) = -F_3(u, v)$. The parameter C monitors the transition between patterns.

Figure 13.20 shows some typical stationary Turing patterns formed in a continuously fed open spatial reactor system. The wavelength of the Turing patterns

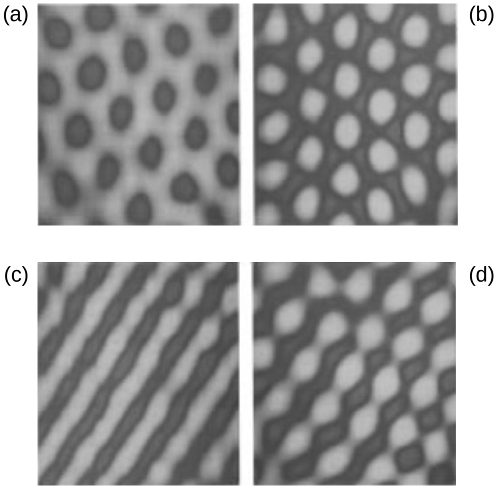


Figure 13.20. Stationary Turing patterns formed in a continuously fed open spatial reactor: a, transient honeycomb; b, hexagons; c, stripes; d, mixed. From Chaos 1(4), 411(1991) with permission. © 1991 American Institute of Physics.

is determined by physical properties of the system and not by stability considerations. Since the pattern depends upon diffusion then by the relation $\lambda = 2(D\tau)^{1/2}$ and the fact that the diffusion of two components and two reactions are involved we use $D = (D_1D_2)^{1/2}$ and $\tau = (k_1k_2)^{-1/2}$, where the k values represent the reaction constants in the equations for the reaction rates and the D values are the diffusivities of the activator and inhibitor components.

Stationary patterns appear not only at thermodynamic equilibrium or in dissipative systems that involve dissipative crystals but also in dissipative systems in which the patterns may be products of dissipative processes but themselves are not sources of dissipation of energy. For example, there are cellular patterns of solidification processes, a host of ripple patterns, such as sand ripples, ripples on a sputter eroded surface, etc. Each of these processes have their own set of equations governing the kinetics and morphology of their appearance and all appear upon some instability that involves a break-in symmetry.

Nature hosts temporal patterns, as well as the spatial ones considered above. There are several books devoted to this subject.[19] Consequently, the interested reader is advised to go to these sources for enlightenment in this area. Generally, the nonlinear equations involved are similar to those considered above but may be more comprehensive, e.g. involve the complex Ginzburg–Landau equation. A current thrust of research in spatiotemporal pattern formation is in its application to living matter. As an introduction to this field the next chapter considers the thermodynamics of micelles, a unit found in living matter.

Appendix 1

Consider a repulsive potential that is concave upward as shown in Figure A1.1. Further, consider a periodic particle array either in a fcc, bcc or hcp pattern. We draw attention to the changes in interatomic distances between a particle and its neighbors as the particle is moved from its position on the periodic lattice. To illustrate the point we wish to make we will make use of a 2D lattice, such as a triangular lattice illustrated in Figure A1.2. For any motion of a particle from its periodic position a consequence of this motion is that for any increase in distance between this particle and one of its neighbors (denoted here as the first

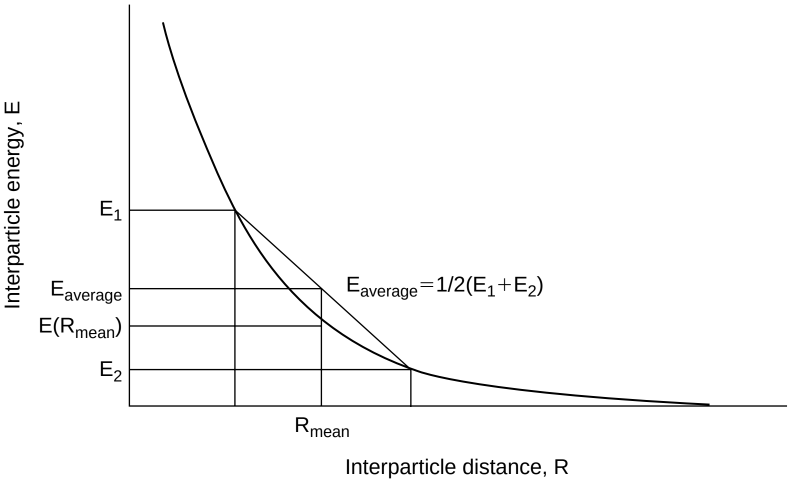


Figure A1.1. Interparticle energy (repulsive interaction) as a function of interparticle distance.

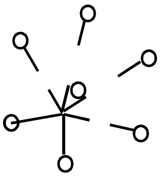


Figure A1.2. Showing that any motion of a particle in a periodic lattice (here a 2D triangular one) results in a situation where every interatomic distance increase is coupled to decrease in another interatomic distance.

neighbor) there is a corresponding decrease in distance between this particle and another neighbor that was originally at the same distance from it as the first neighbor. This situation is a consequence of symmetry of the periodic array and the fact that each neighbor shell about any particle contains an even number of particles. Another consequence is that any motion of a particle from its position of the periodic lattice increases the energy since the energy of a configuration corresponding to a displacement that increases the interparticle distance which is coupled with one that decreases it by Figure A1.1 must have higher energy than that for the undistorted interparticle distance.

References

1. M. Seul and D. Andelman, Science 267, 476(1995).
2. D.J. Keller, H.M. McConnell and V.T. Moy, J. Phys. Chem. 90, 2311(1986); A. Fischer et al. J. Phys. Lett. Paris 45, L785(1984); H. Mowald, Ann. Rev. Phys. Chem. 41, 441(1990); D. Andelman et al., J. Chem. Phys. 86, 3673(1987); D. Andelman et al., C.R. Acad. Sci. Paris 301, 675(1985).
3. T. Garel and S. Doniach, Phys. Rev. B 226, 325(1982).
4. R.P. Huebener, **Magnetic Flux Structures in Superconductors**, Springer-Verlag, Berline, 1979; F. Haenssler and L. Rinderer, Helv. Phys. Acta 40, 659(1967); T.E. Faber, Proc. Roy. Soc. London A 248, 460(1958).
5. Z. Suo and W. Lu, J. Nanopart. Res. 2, 333(2000); W. Lu and Z. Suo, Phys. Rev. B 65, 085401(2002); ibid, 205418(2002); Y.F. Gao, W. Lu and Z. Suo, Acta Mater. 50, 2297(2002); Y.F. Gao and Z. Suo, J. Mech. Phys. Sol. 51, 147(2003).
6. W. Kern et al., Phys. Rev. Lett. 67, 855(1991).
7. B.S. Kwak et al., Phys. Rev. Lett. 68, 3733(1992).
8. S. Leibler and D. Andelman, J. Phys. France 48, 2013(1987); D. Andelman et al., Europhys. Lett. 19, 57(1992); T. Kawakatsu et al., J. Phys. II France 3, 971(1993); T. Taniguchi et al., ibid 4, 1333(1994); M. Marder et al., Proc. Natl. Acad. Sci. 81, 6559(1984).
9. X. Ren and J. Wei, Physica D 188, 277(2004).
10. L. Leibler, Macromolecules 13, 1602(1980); T. Ohta and K. Kawasaki, ibid 19, 2621(1986); A.N. Semenov, Sov. Phys. JETP 61, 733(1985); F.S. Bates and G.H. Fredrikson, Ann. Rev. Phys. Chem. 41, 525(1990).
11. G. Malescio and G. Pellicane, Nature Mater. 2, 97(2003).
12. A.G. Yodh et al., Phil. Trans. Roy. Soc. London A 359, 921(2001).
13. S. Asakura and F. Oosawa, J. Chem. Phys. 22, 1255(1954).
14. C.-M. Xiao et al., Phys. Rev. E 64, 011402(2001).
15. T.C. Hales, Ann. Math. 162(3), 1065(2005).
16. H.L. Swinney and E.C. Rericha, in **The Physics of Complex Systems**, eds. F. Mallamace and H.E. Stanley, IOS Press, Amsterdam, 2004; P.B. Umbanhowar, F. Melo and H.L. Swinney, Nature 382, 793(1996); Q. Ouyang and H.L. Swinney, Nature 352, 610(1991).
17. M.C. Cross and P.C. Hohenberg, Rev. Mod. Phys. 65, 851(1993).
18. M.I. Rabinovich, A.B. Ezersky and P.D. Weidman, **The Dynamics of Patterns**, World Scientific Publishing, Singapore, 2000.

19. A.T. Winfree, **The Geometry of Biological Time**, Springer-Verlag, New York, 2001, 2nd edition. D. Walgraef, **Spatio-Temporal Pattern Formation**, Springer-Verlag, New York, 1997.
20. P. Chen and J. Vinals, Phys. Rev. Lett. 79, 2670(1997).
21. S. Ciliberto and M. Caponeri, Phys. Rev. Lett. 64, 2775(1990).
22. A.M. Turing, Phil. Trans. Roy. Soc. London B 237, 37(1952).
23. S.L. Judd and M. Silber, Physica D 136, 45(2000).

CHAPTER XIV

Thermodynamics of Micelles

1. Micelles

1.1. Nucleation of micelles

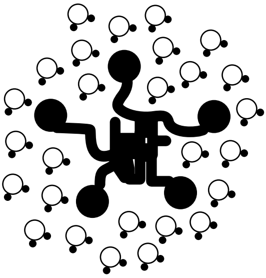


Figure 14.1a. A micelle.

The aggregates formed upon an increase in amphiphile* concentration in water beyond some "critical" concentration are called micelles. Figure 14.1(a) illustrates a micelle schematically. The open circles with two attached dots represent water molecules. The hydrophilic head groups of the amphiphiles are the filled circles. The hydrophobic tails are the dark lines attached to the filled circles. As shown, the hydrophobic groups are inside this roughly spherical micelle where they are shielded from the water molecules. Figure 14.1(b) shows a plot of free (unbonded) amphiphile concentration

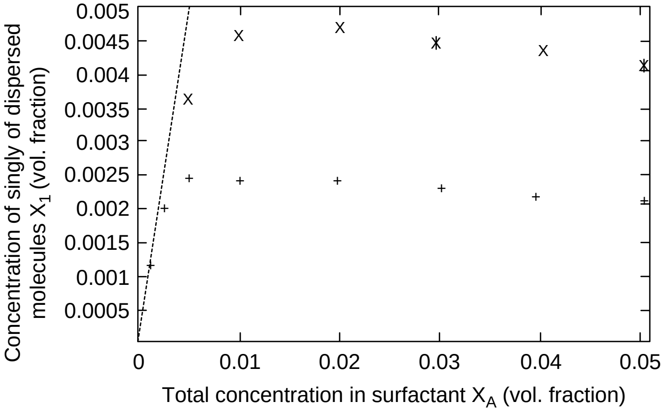


Figure 14.1b. Plot of free amphiphilic concentration versus total amphiphilic concentration for two different amphiphiles indicating two different critical micelle compositions. A typical result of a Monte Carlo simulation.

---

* A molecule which contains a solvophobic group on one side and a solvophilic group on the opposite side of the molecule.

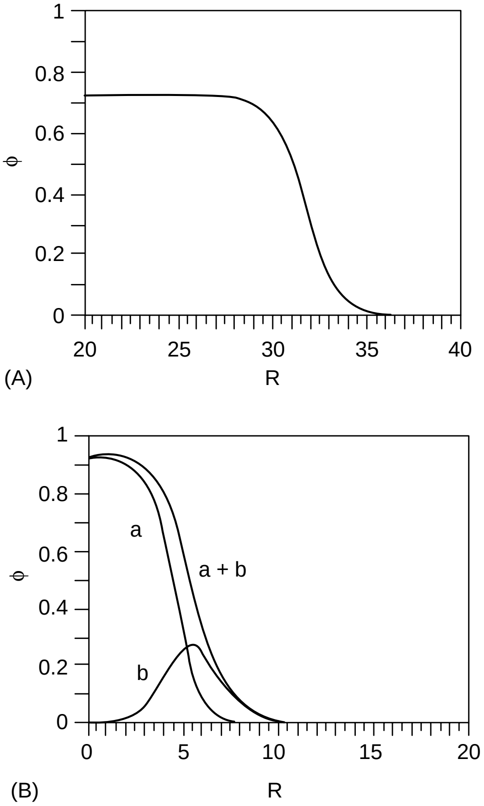


Figure 14.2. Schematic diagrams showing the radial volume fractions of aggregates of: (A) non-amphiphilic molecules; (B) amphiphilic molecules showing positions of hydrophobic ends, a, and of hydrophilic ends, b.

versus total amphiphile concentration of a host solution. As indicated in the figure the free amphiphile concentration remains approximately constant beyond the critical micelle concentration (CMC). The CMC is not a phase boundary. Micelles do not represent a thermodynamic phase. These statements will be explained later in this chapter. Before we proceed to this explanation it is instructive to consider the nature of the product formed at concentrations in excess of the CMC. This product differs in a significant way from that formed in a metastable solution having non-amphiphilic solute in the transition to the equilibrium state. The latter occurs either by classical nucleation and growth, or by spinodal decomposition, or by some combination of both processes. Consider the case of a metastable solution at a composition adjacent to and slightly above the binodal composition. The cores of the embryos (clusters of the stable phase) have a constant composition with a more or less sharp interface between core of the aggregate (embryo or nucleus) and solution parent phase, much as illustrated in Figure 14.2(A). The diameter or thickness of the core may be many (more than two) times the largest dimension of the non-amphiphilic molecule. In amphiphilic systems the smallest dimension of the micelle may be no more than about two times the extended length of the amphiphilic molecule with the core region rich in the hydrophobic portion of the molecule and the surface region rich in the hydrophilic component, much as shown in Figure 14.2(B).

The work to form an aggregate at a fixed position, constant T, μ and p, also differs as a function of aggregate size between these two different types of molecules. For the non-amphiphilic molecules the volume of and work to form a critical nucleus approaches infinity as the binodal composition is approached from above, and the work to produce a composition fluctuation goes to zero at the appropriate spinodal composition. For the amphiphilic molecules both the size and work to form a micelle are finite and small from the CMC (below the binodal composition for strong amphiphiles) through the binodal composition and, as for the non-amphiphilic molecules, approaches zero at the spinodal composition. The latter behavior is illustrated in Figure 14.3, where the curve denoted AB is the critical work to form (work to nucleate) micelles as a function of the volume fraction of free amphiphilic molecules in the bulk parent solution.[1]

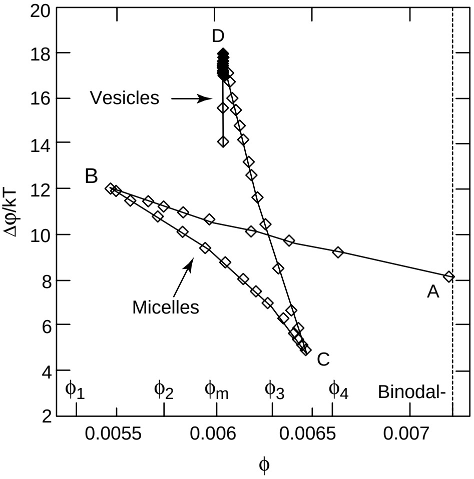


Figure 14.3. Curve AB is the work to form the critical nucleus as a function of mole fraction of amphiphiles in parent solution. Curve BC shows the minima in the work to form aggregates versus free amphiphile composition and, hence, the energies of the metastable micelles. Curve CD represents the loci of maxima in the work to form aggregates beyond the size of the metastable micelles and, hence, the work to nucleate vesicles. From J. Chem. Phys. 113(16), 7013(2000) with permission. © 2000 American Institute of Physics.

The work to form micelles as a function of size differs from that to form stable phases. In the latter case the work to form an aggregate of the stable phase in a metastable host decreases once the aggregate size becomes larger than that for the nucleus. Hence, beyond the critical nucleus aggregates of the stable phase grow spontaneously. For the amphiphilic molecules and within the range of host composition between the cusps B and C in Figure 14.3, the work to form spherical aggregates as a function of aggregate size first exhibits a maximum at the size of the critical nucleus, then exhibits a minimum at some larger size (see Figure 14.4(a)). The curve BC presents this minimum free energy as a function of size of the aggregate.

Cylindrical aggregates start life as spherical micelles. For free micelle concentration beyond the CMC cylindrical micelles grow at essentially constant free micelle concentration and excess free energy per micelle. Hence, they do not grow at total micelle concentrations less than that equivalent to ϕ_4 in Figure 14.3. Indeed, for aggregates smaller than that corresponding to the minimum all the aggregates have a nearly spherical shape. It is only for sizes larger than this minimum in the work to form a spherical micelle will this work differ significantly between the various shapes. For the amphiphiles that produce spherical aggregates, the work to form the micelle will then increase with size and for host amphiphile compositions beyond a critical free amphiphile concentration (ϕ_m, the coexistence composition) exhibit another maximum and then decrease with increasing size. In this case the spherical micelle transforms to a bilayer disk or vesicle. This behavior is shown schematically in Figure 14.4(a). For amphiphiles that tend to form the cylindrical shaped micelle, at and beyond the minimum in the work to form a spherical micelle, the spherical micelle will elongate to a spherocylindrical shape as the work to form the aggregate continually decreases with increase in the aggregate size. (Which shape of aggregate will be formed depends on the shape of the amphiphile molecule, a subject to be considered later in this chapter.)

Thus, one contribution to the free energy *of the system* consisting of a distribution of micelles of various sizes must involve a sum over all work to form these micelles, i.e. a sum over the excess free energy of formation of each micelle in the system. Another must come from the entropy of mixing the micelles in the solution, which is a driving force for micelle formation. The question arises as to whether the

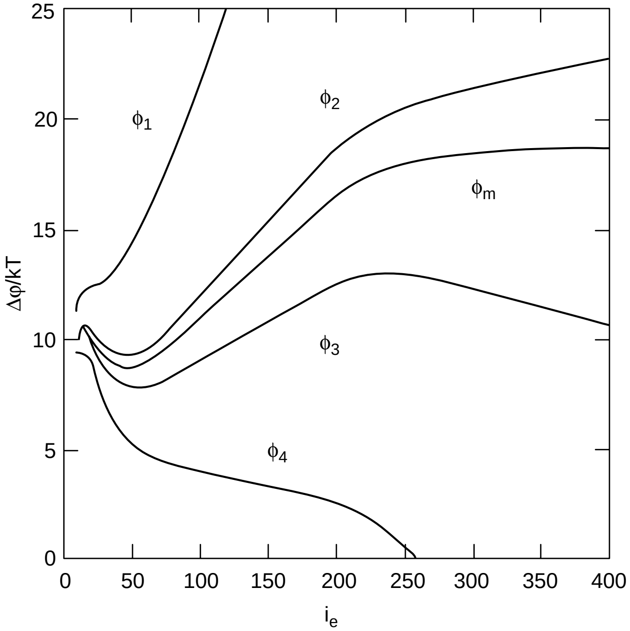


Figure 14.4a. The work to form aggregates as a function of number of amphiphile molecules in aggregate for various compositions of the metastable parent solution. From J. Chem. Phys. 113(16), 7013(2000) with permission. © 2000 American Institute of Physics.

free energy of the system containing micelles is less than that for that in which all the amphiphiles are in solution. Thus, we need to develop a relation for the free energy of the system of water and amphiphiles. We will later make use of that of Christopher and Oxtoby to ascertain that indeed the system of micelles and solution is stable relative to a supersaturated solution containing all the amphiphiles and no micelles. Values of the excess free energy of micelle formation calculated from the micelle populations revealed at equilibrium in the simulation (dashed line), in a solution maintained at constant chemical potential, are shown in Figure 14.4(b). As shown, the simulation results are consistent with those deduced by Talanquer and Oxtoby shown in Figure 14.4(a).

As already noted, the composition φ_m, at which the work to form a membrane or a vesicle (a closed aggregate of amphiphilic molecules surrounded by solvent but also having solvent molecules in its interior) from a solution of amphiphiles changes from an increasing function of size to one that reveals a maximum as a function of size, is denoted the coexistence composition. Above this composition the work to form vesicles from amphiphiles in the host solution becomes negative for sufficiently large vesicles, i.e. sufficiently large vesicles are stable in the host solution. For host composition beyond the coexistence composition vesicles can also form from micelles and grow spontaneously upon a fluctuation that exceeds the barrier to its nucleation. The work to nucleate vesicles from proper micelles* is given by the difference in height between the curves CD and BC in Figure 14.3. Between the coexistence and binodal compositions the vesicle, once nucleated, can grow without limit providing that the composition of the host solution is maintained constant. For the free amphiphile composition less than the

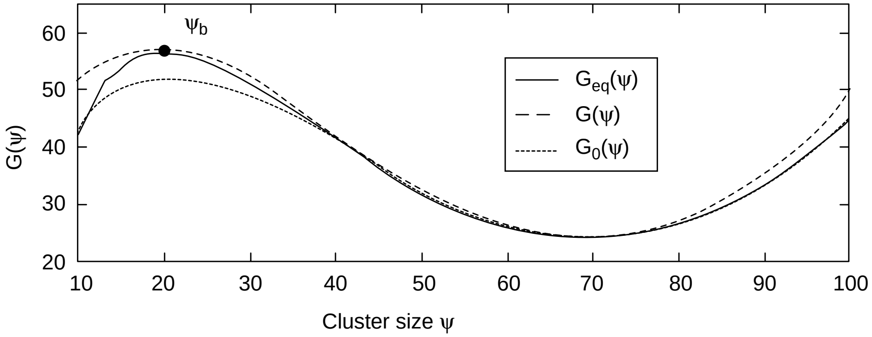


Figure 14.4b. Free energy of solution containing micelle clusters as a function of cluster size as deduced by three different methods based on a Monte Carlo simulation. From J. Chem. Phys. 122, 044908(2005) with permission. © 2005 American Institute of Physics.

---

* We call the micelle having the minimum work of formation the proper micelle.

coexistence composition, vesicles and membranes spontaneously shrink. For free amphiphilic concentrations less than that corresponding to cusp B (the CMC), already formed micelles and membranes become unstable and breakdown.

Thus, as contrasted to the case of non-amphiphilic molecules for which classical nucleation yields a stable product via a first-order transformation only at compositions in excess of the binodal composition, micelles in statistical equilibrium with the host solution are nucleated from amphiphilic molecules at host compositions less than the coexistence composition without a transformation to another type of aggregate. When the free amphiphile composition of the solution is larger than the coexistence composition then micelles may transform to bilayer membranes or vesicles during growth. This growth involves a first-order transition. Cylindrical micelles can continue to grow as cylinders without any transition for free amphiphile concentrations higher than that equivalent to ϕ_4 provided that the chemical potential is maintained constant. The characteristics of micelles described above are among the main results of theoretical studies of micelles using a variety of methods.[1–5]

This property of amphiphiles in solution – to form micelles – is critical to the phenomenon of life as we know it.

2. Characteristics of micelles

2.1. Total free energy per unit volume of system

We will make use of the work of Christopher and Oxtoby[6] to describe the free energy of the system of amphiphile solution plus micelles. Accordingly, the total free energy per unit volume of the system consisting of amphiphile solution (water plus monomeric amphiphiles) and micelles in dimensionless units is given by

$$\beta F = \beta f + \Sigma_s Z_s \beta\, \Delta F_s + \Sigma_s Z_s(\ln Z_s - 1) \tag{14.1}$$

where $\beta = 1/kT$, f is the free energy per unit volume of the amphiphile solution (water plus monomeric amphiphiles), $Z_s = n_s/V$, where n_s is the number of micelles of type s containing $N_0^{(s)}$ water molecules and s amphiphiles in the volume V of the system, and ΔF_s is the excess free energy of formation of a micelle of type s.

2.1.1. Free energy per unit volume of amphiphile solution f

Christopher and Oxtoby[6] then use a density functional model[7] to derive the free energy per unit volume of amphiphile solution f. This model derives f in terms of the bulk density of solvent m_o, and amphiphile m_a. Here $m_X = N_x^b/V_b$, where N_X^b

is the number of molecules of type X in the bulk phase b, consisting of water plus monomeric amphiphiles (what we called the amphiphile solution above) and V_b is the volume of the bulk phase in the system of volume V. Then f is given by

$$\begin{aligned}\beta f = {} & m_o(\ln m_o - 1) + m_a(\ln m_a - 1) + m^*\psi(\eta) \\ & + (\beta/2)\Sigma_{ij}(-16\pi/3)\varepsilon_{ij}m_i m_j \qquad (14.2)\end{aligned}$$

The first two terms in this equation represent the ideal entropy of the bulk solution. The third term accounts for the hard sphere repulsions in the solution via the Carnahan–Starling[8] free energy excess where

$$\psi(\eta) = \frac{\eta(4 - 3\eta)}{(1 - \eta)^2}$$

m* is defined by

$$m^* = \frac{\Sigma_{i=(o,a)}\nu_i m_i}{\Sigma_{i=(o,a)}\nu_i}$$

and where ν_i is the molecular volume of species i and the packing fraction η is

$$\eta = \Sigma_{i=(o,a)}\nu_i m_i$$

The last term in the equation for the free energy per unit volume of the bulk phase takes intermolecular interactions into account. The attractive interaction is assumed to be of the form $-4\varepsilon_{ij}/|r - r'|^6$, which when integrated over the volume yields the last term in the equation for f.

Setting $dF(r_o, r_a, Z_s) = 0$, where $r_x = N_x^b/V$, the equilibrium values of r_o, r_a, Z_s and F were determined. Thus, the proper micelles represent an equilibrium phenomenon.

2.2. Are micelles phases?

For micelles to be phases the saturated solution containing monomers and micelles at coexistence must be in equilibrium with an infinite aggregate. The latter corresponds, thermodynamically, to a phase. Now, the chemical potential of an amphiphile in the infinite aggregate is μ_∞^o. For coexistence between this aggregate and monomers and finite micelles in the saturated solution then

$$\begin{aligned}\mu_\infty^o &= \mu_1 = \mu_1^o + kT\ln X_1^* \\ &= \mu_s = \mu_s^o + (kT/s)\ln(X_s^*/s) \quad \text{for } s \geqslant p \qquad (14.3)\end{aligned}$$

where the asterisk denotes the values at saturation, the underline denotes the value corresponding to a single amphiphile, the superscript ^{o} represents the packing or standard value, p is the number of amphiphiles in the spherical end caps of the cylinder and s is the total number of amphiphiles in the cylindrical micelle.

Noting that for a finite cylindrical micelle that

$$\mu_s^o = s\underline{\mu}_{cyl}^o + p(\underline{\mu}_{sph}^o - \underline{\mu}_{cyl}^o) \tag{14.4}$$

where we rewrite the equation for $\underline{\mu}_s$ to yield

$$X_s^* = s[\exp\{s[\underline{\mu}_\infty^o - \underline{\mu}_{cyl}^o] - p[\underline{\mu}_{sph}^o - \underline{\mu}_{cyl}^o]\}/kT] \tag{14.5}$$

Since $\underline{\mu}_\infty^o - \underline{\mu}_{cyl}^o = 0$ and p is a constant then X_s^* increases with increase in s and X_{tot}^* does not converge to a finite total concentration at saturation.

For the case of a disk-like micelle, p in the above equation becomes proportional to $s^{1/2}$. The same equation applies with the subscript disk replacing cyl. Since X_s^* converges to zero as s increases then the total concentration X_{tot}^* at saturation is a finite value and the disk geometry of the micelle can undergo a first-order phase transition to an infinite aggregate.

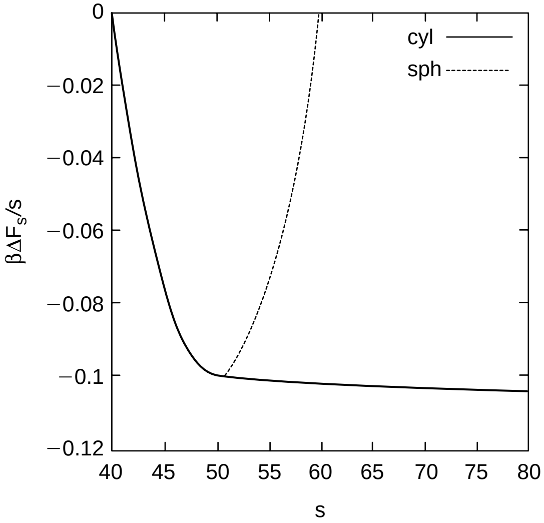
The above derivation is consistent with the theorem of Landau and Lifshitz (L. Landau and E.M. Lifshitz, **Statistical Physics**, Pergamon, 1968) on "the impossibility of the existence of phases in a 1D dimensional system".

2.3. The critical micelle concentration

We now have a basis for discussing the origin of the CMC and the approximate invariance of the free amphiphile concentration as the total amphiphile composition of the system increases. Let us consider the case first of cylindrical micelles. Since for cylindrical micelles X_s^* becomes proportional to s, then $\underline{\mu}_s$ in the equation

$$\underline{\mu}_s = \underline{\mu}_s^o + (kT/s)\ln(X_s^*/s) \quad \text{for } s \geqslant p \tag{14.6a}$$

approaches the value of the packing chemical potential $\underline{\mu}_s^o$. Hence, $\mu_1 = \underline{\mu}_s$ also approaches the value of $\underline{\mu}_s^o$. Now the value of the latter quantity is a near constant value because every amphiphile added to the micelle experiences the same environment as every other amphiphile in the micelle that is not in the spherical caps. This feature is also illustrated in Figure 14.5, a result of the calculations of Christopher and Oxtoby, where it is shown that beyond the minimum value for the sphere the cylinder has a near constant value of the excess Helmholtz free energy of the micelle per amphiphile. Neglecting the water molecules in the micelle we can approximate $(\Delta F_s/s)$ by $[(\mu_1 - \mu_s^o) - \gamma v_s/R]$, where γ is the surface tension, v_s the volume per amphiphile, and R the radius of the cylindrical micelle. Hence, $(\Delta F_s/s) \rightarrow -\gamma v_s/R$. Thus, γ and X_1 approach a constant value since $\gamma(X_1)$ is a monotonic function and by Figure 14.5 $(\Delta F_s/s) \rightarrow$ a constant. In words, the chemical potential of the amphiphile in the water solution approaches a constant value beyond the CMC as cylindrical micelles grow beyond the size corresponding to the aggregation number of the minimum packing chemical potential for the sphere. Thus, beyond the CMC the constancy of the free amphiphile chemical potential and corresponding free amphiphile concentration exists because growth of cylindrical micelles occurs at constant values of these parameters. The cylindrical micelles act as sponges for amphiphiles in excess of the CMC free amphiphile concentration!



Figures 14.5. Showing the excess free energy of a micelle per amphiphile in the micelle as a function of the number of amphiphiles in the micelle. From J. Chem. Phys. 118, 5665(2003) with permission. © 2003 American Institute of Physics.

For the case of amphiphiles that do not form cylindrical micelles, at amphiphile compositions beyond ϕ_m in Figure 14.4(a) the free amphiphile concentration is determined by the first-order transformation to a membrane or vesicle and is maintained constant at ϕ_m by this coexistence condition.

Although we have provided an explanation for the near invariance of the free amphiphile concentration in the water solution for compositions beyond the CMC we have not provided an explanation for the sharpness of the transition from a dependence of the free amphiphile concentration on the total amphiphile composition to a disappearance of this dependence. The explanation stems from the dependence of the micelle population on ΔF_s. In the work of Christopher and Oxtoby it can be shown that the number of micelles of size s per unit volume Z_s is given by

$$Z_s = M \exp(-\beta \Delta F_s) \qquad (14.6b)$$

where $M = \exp\{\beta\mu_o(r - m_o\nu_s) + \beta\mu_1(s - m_1\nu_s)\}$, r is the number of water molecules in a micelle having a total of s amphiphiles and $\beta\mu_X = \partial\beta f/\partial m_X$ for X = 0, 1, standing for water and amphiphile, respectively. Now, all the terms in M are independent of s with the exception of the term containing s. The latter merely skews the dependence of ΔF_s on s and since the latter exhibits a minimum value then Z_s reveals a maximum value as shown in Figure 14.6. Now, as shown in Figure 14.4(a) it is only at amphiphile compositions higher than the CMC that ΔF_s exhibits a minimum. At concentrations less than the CMC this quantity increases as s increases. Hence, it is only at free amphiphile concentrations higher than the CMC that micelles having the aggregation number corresponding to the minimum in ΔF_s and maximum in Z_s will appear in rapidly increasing number, i.e increasing by many orders of magnitude with only a small increase in the free amphiphile concentration in the host solution. The appearance of numerous micelles then limits the free amphiphile concentration as noted above. Hence, the rapid leveling off of the free amphiphile concentration with increase in total amphiphile composition is a consequence of the change in dependence of the work to form a micelle on micelle size from a monotonously increasing function to one that exhibits a minimum, which itself decreases in positive value with increase in the free amphiphile concentration.

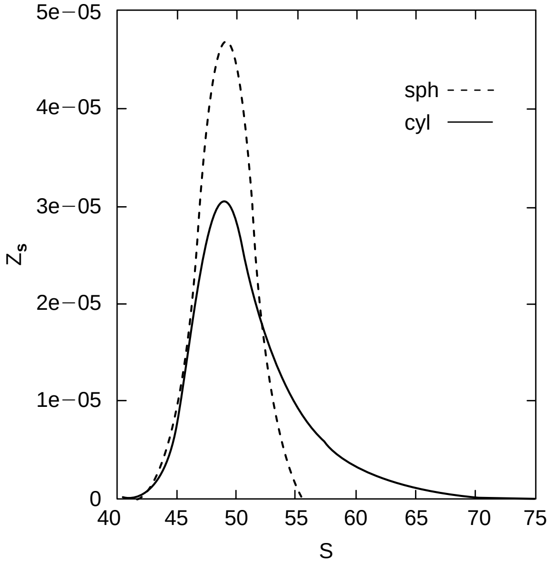


Figure 14.6. Population density of micelles as a function of number of amphiphiles in micelle. From J. Chem. Phys. 118, 5665(2003) with permission. © 2003 American Institute of Physics.

2.4. Contrasting nucleation of heterophases and micelles

First we may note that the form of the population distribution of heterophase embryos and micelles differs. That for heterophase embryos is a monotonously decreasing function as embryo size increases. Figure 14.6 shows that for micelles. These distributions follow from the dependence of the work to form these entities on their size which also differ as shown by the comparison of equation 5.7 where this work increases monotonously with the radius of the embryo and Figure 14.4 where it is apparent that the work to form a micelle reveals a minimum for free amphiphile concentrations larger than the CMC. Indeed, nucleation of stable phases can only occur in regions where they are stable. Nucleation of micelles to yield distributions of micelles in statistical equilibrium with the host solution occurs above the CMC. These micelles may or may not later form the embryos that nucleate stable phases depending on whether they are one dimensional or higher dimensional entities. We will need to explore how these micelles act as embryos for the nucleation of other types of entities containing amphiphiles.

2.5. What factor(s) control the shape of micelles?

With respect to the shape of micelles most of the knowledge that existed until recently relative to this question was empirical. Israelachvili[9] has suggested the use of a packing parameter to delineate the regions unique to each shape. This packing parameter is $P = v/la$, where v is the volume of the amphiphile, l is the length of the tail and a is the surface area of the head group. According to this empirical "rule" for $0 < P < 1/3$ the sphere predominates, for $1/3 < P < 1/2$ the cylinder will be found and for $1/2 < P < 1$ the bilayer or disk holds sway. It is likely that all micelles at their smallest size are spherical in shape. A cylinder proceeds from a sphere by growth in 1D. A disk grows out from a sphere in two dimensions. The sphere itself cannot continue to grow in 3D because a high-energy void would be formed in its center. Any form other than a sphere, with the head groups of the amphiphiles at the circumference, would leave the tails unshielded from the solvent and consequently would have higher work of nucleation than that for the spherical shape. This expectation seems to be borne out by simulation studies which will be discussed shortly.

Let us now consider some phenomenological approaches to micellar growth that also affect micellar shape. Both Tanford[10] and the other phenomenological model due to Israelachvili, Mitchell and Ninham (IMN)[11] express the free energy per amphiphile (the packing chemical potential per amphiphile μ_s^o) as a function of the head surface area a, as follows

$$\mu_s^o(a) = \gamma a + \kappa/a + f_t \tag{14.7}$$

where γa is the interfacial free energy associated with hydrocarbon–water contact, κ/a represents the leading term in a repulsive potential due to electrostatic or excluded volume interactions, or both, between head groups, and f_t is a constant tail contribution to free energy per amphiphile in the micelle. This relation yields a minimum in $\mu_s^o(a)$ at a particular value of the head area equal to $a_o = \sqrt{(\kappa/\gamma)}$. This result is independent of the geometry of the micelle. Another assumption is codified in the packing constraint (i.e. the minimum radius $R < l$, the length of the amphiphile molecule) that must be obeyed by all micelles. This constraint can be shown to be equivalent to the following constraints on the areas per head group

$$a \geqslant i^* v/l \tag{14.8}$$

where $i^* = 1$, 2 and 3 for spheres, cylinders and bilayers, respectively and v is the volume V of the micelle divided by the number of amphiphile molecules in the micelle. Thus, if $a < 3v/l$ then spherical micelles will be unstable. Similarly, the other geometries must obey their packing constraint to be stable geometries. (The concept just described is the basis for the empirical rule given at the beginning of this section.) If all the geometries obey the packing constraint then in the IMN view the amphiphiles will preferentially organize in spherical micelles, because for a given total concentration, this arrangement minimizes the volume of the micelle containing a given number of molecules and thereby maximizes the "mixing" entropy of the system. Certainly, this concept may partially explain the shapes assumed by micelles, but as we will show later other factors also play a role. Tanford's model differs from the IMN model in assuming that in all micelle geometries the amphiphile molecule has the same mean chain length 0.75l, and he uses this value to determine a and the corresponding $\mu_s^o(a)$ for all the geometries. The minimum value of the latter is then in his model the stable geometry for the micelle. In his model the area a is not a variable that can be minimized. It is fixed by l and v.

The above description of one phenomenological model should be sufficient to convince the reader that the above analysis does not enable us to know in detail what controls the geometry of the micelle other than the general statement that the geometry that minimizes the free energy of the system is the one observed. Unfortunately, we do not know enough to predict the latter. However, as will be shown later, there are more sophisticated numerical models that enable one to evaluate the packing free energy per amphiphile for assumed conditions. At this writing, some large scale molecular dynamics simulations[12,13] are able to provide evidence that for given numbers of water and amphiphile molecules there is an equilibrium state that can be reached either from completely dissolved or already assembled spherical or cylindrical micelles. This equilibrium state for the case examined consisted of spherical micelles of about the same size. At this writing, no direct self-assembly of cylindrical micelles from solution (without the intermediate spherical micelle stage) has been achieved in a molecular dynamics simulation, except as an artifact of periodic boundary conditions which made a cylindrical micelle more stable than the spherical micelle for the solution mole fraction. The mechanism of self-assembly of the spherical micelle in the molecular dynamics simulation, however, provides information about the processes involved.

The following quote is taken from Reference 14. "It appears that in the first stage (between 0 and 100 ps), the surfactant molecules approach one another, forming aggregates without any well-defined organization, both from a translational and rotational point of view. The size of these disordered micelles is small ($\approx$10 molecules). In the second step these random aggregates rearrange to form spherical micelles. The driving force for this rearrangement appears to be the minimization of the repulsive interactions between head groups *together with enhancement of the hydrophobic attractions between the hydrocarbon tails*. During this structural rearrangement, the aggregation number of the micellar clusters increases via the addition of small clusters of surfactant (typically comprising 2 or 3 molecules). Finally, by the end of the simulation, both micelles exhibit a spherical shape, containing 15 and 17 surfactant molecules respectively." There are several points of interest in this description of the self-assembly process. One is that the process does not occur by the addition of one molecule at a time, but by the addition of small clusters. Second is that *rearrangement* of the amphiphile molecules in the clusters is the mechanism of producing the spherical geometry of the micelle. Third is that two interactions contribute to the rearrangement of the amphiphiles in the clusters: minimization of the repulsive interactions between head groups and enhancement of the hydrophobic attractions between the hydrocarbon tails. The latter result suggests the situation that should be apparent to any researcher in this field. Namely, that the equilibrium shape is determined by a complex of interactions between the components of the system, which can vary with variation in amphiphile, temperature, solvent and other species in the solvent as well as the composition of the solution.

There are a series of recent studies in the literature which provide insight into the factors at play in the determination of the micelle geometry in non-ionic surfactant solutions. However, before we consider them let us dwell a moment and consider the effects of the two interactions known to be at play. As noted above, these two interactions are the repulsive interactions between head groups and the attractive interactions between the hydrocarbon tails. It is obvious under these circumstances that the tails tend to approach each other as closely as possible while the head groups tend to be as far apart as possible, consistent with the requirement that they be close enough to prevent contact between water molecules and hydrocarbon tails. The geometry that best satisfies these tendencies is the sphere with tails in its center and head groups on its surface. Hence, it comes as no surprise that at low concentrations of non-ionic surfactants, micelles are always nearly spherical. Although these two factors are explicitly contained in the IMN model equation 14.8 for the free energy per amphiphile of a micelle there is nothing in this equation that takes into account the fact that the attractive interaction is between the tail portions of amphiphiles and the repulsive interaction is between the head groups of the amphiphiles. May et al.[14] explicitly take this spatial separation of interactions into account as follows.

Their formulation of the IMN free energy (the packing chemical potential per amphiphile in the micelle μ_s^o) is

$$\mu_s^o = \gamma a + B/a_h \tag{14.9}$$

But now, a is the area per head group at the hydrophobic–aqueous micellar interface, γ is the surface tension of the latter and a_h is the average head group area at the surface of head group repulsions. These areas are related to geometry via

$$a_h = a[1 + (\kappa_1 + \kappa_2)l_h + \kappa_1\kappa_2 l_h^2] \tag{14.10}$$

where κ_i is a principal curvature (i.e. $1/R_i$) and l_h is the average distance between the hydrophobic–aqueous micellar interface and the surface of head group repulsions. These curvature corrections are ignored in the IMN model but may become important in situations where curvature is involved in the local shape of the micelle.

A more important correction to the IMN model is that involved with the chain conformational free energy, the term f_t in equation 14.7. According to May et al.[14] this term is not independent of geometry. They evaluated this free energy using a mean-field approximation. A result for this free energy for the different micelle geometries and for a choice of parameters given by ($\gamma = 50$ dyne/cm, $l_h = 1$ Å, $l_c = 18$ Å, corresponding to $(CH_2)_{13}CH_3$(C14) alkyl chains) is shown in Figure 14.7 from their work. In this figure $f \equiv \mu_s^o$, and the factor b in the abscissa is the thickness of the hydrophobic core (i.e. the radius of the spherical and cylindrical micelles or the half-thickness of a planar bilayer). ($b = 18$ Å corresponds to the fully extended hydrocarbon chain.) The increase in the free energy once b is larger than 18 Å occurs because of the creation of a "hole" in the center of the hydrophobic region. The minima in the chain conformational free energies occur at three different values of b and may affect the optimal packing geometry. To get at the latter, these data were added to those given by equations 14.9 and 14.10, and for each value of the head group repulsion parameter B, f was calculated as a function of the hydrophobic thickness b. Then, for any given B, the value of b for which f is minimal was determined to provide the equilibrium free energy per molecule for that particular B. The results represented as a difference between the equilibrium free energies for the sphere and bilayer with respect to that for the cylinder are

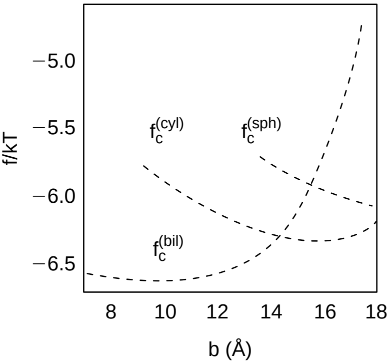


Figure 14.7. The chain conformational free energy as a function of radius or half-thickness of micelle for three geometries. From J. Phys. Chem. B 101, 8648(1997) with permission. © 1997 American Chemical Society.

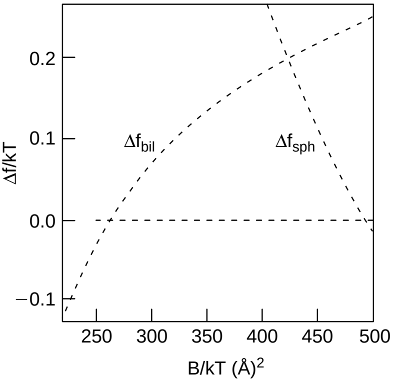


Figure 14.8. Difference in free energy per molecule between bilayer Δf_{bil}, or between sphere Δf_{sph}, and cylinder as a function of the repulsive parameter B. From J. Phys. Chem. B 101, 8648(1997) with permission. © 1997 American Chemical Society.

shown in Figure 14.8. (In this figure $\Delta f \equiv \Delta\mu_s^o$). From this figure it is apparent that for small values of the repulsion parameter B the bilayer is the most stable geometry, while for very large values of B the sphere is the most stable geometry and for intermediate values of B the cylinder shape is most stable. May et al. also calculated the b value corresponding to the minimum in free energy and these values are shown as a function of the repulsive parameter B for the three shapes in Figure 14.9. The dashed lines corresponds to taking f_t in equation 14.7 equal to zero while the full line to the values taking the chain conformational free energy into account. This work thus focuses on the effect of the repulsion parameter on the stable geometry of the micelle and indirectly on the effect of the chain conformational free energy for solutions containing one, single-tail, surfactant species. Further, it should be noted that the free energies per amphiphile that have been calculated above are those corresponding to the uniform sections of each shape, i.e. regions away from end caps for the bilayer and cylinder. The free energy of micelles with caps has already been treated above in relation to equation 14.7. We repeat the results below.

Suppose we wish to describe the packing chemical potential or packing free energy of a spherocylindrical micelle. Further suppose that there are N amphiphiles in the micelle of which M are in the spherical caps and (N − M) are

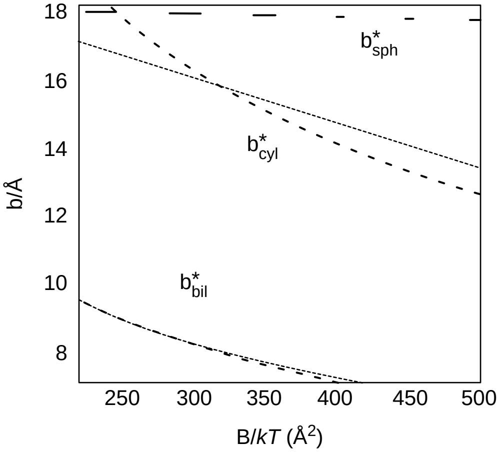


Figure 14.9 Thickness of the hydrocarbon core at the minimum free energy as a function of the repulsive parameter B for the three micelle shapes. From J. Phys. Chem. B 101, 8648(1997) with permission. © 1997 American Chemical Society.

in the cylindrical region. Then the packing chemical potential is given by equation 14.4 or in this notation

$$\mu_s^o(N) = (N - M)\mu_{s\,cyl}^o + M\mu_{s\,sph}^o = N\mu_{s\,cyl}^o + 2\delta$$

where $\delta = (M/2)(\mu_{s\,sph}^o - \mu_{s\,cyl}^o)$ is the excess energy of an end cap relative to the cylindrical body. From this relation it is apparent that a spherical micelle of a given radius would have a lower packing free energy than a capped cylindrical or bilayer because of a smaller total number of amphiphiles in the spherical micelle. Hence, the spherical micelle is the first type of micelle to form on exceeding the CMC. The question is whether the growth of the spherical to either the cylindrical or bilayer can proceed continuously or does it require a nucleation step. Another question is whether spherical and other geometries of micelles can coexist. According to the relation above, the cylindrical micelle can grow continuously. However, there is evidence that the growth is discontinuous, at least for some amphiphiles.

2.6. Equilibrium size distribution of micelles and their coexistence

As noted previously Christopher and Oxtoby obtained an equation for the equilibrium size distribution by setting the derivative of their relation for the free energy of the system with respect to Z_s equal to zero. The equation they obtained is as follows

$$Z_s = M'\exp(-\beta[\Delta F_s - \mu_1 s])$$

where $M' = \exp\{\beta\mu_o(r - m_o\nu_s) - \beta m_A\nu_s\}$, r is the number of water molecules in a micelle having a total of s amphiphiles and $\beta\mu_X = \partial\beta f/\partial m_X$ for X = 0, 1, standing for water and amphiphile, respectively. This relation provided the data shown in Figure 14.6. Thus, there is no doubt that micelles exist in statistical equilibrium at total amphiphile compositions of the system less than the binodal composition.

A similar result yielding a size distribution for micelles is obtained from a phenomenological approach due to Israelachvili[9] as follows.

We assume that the amphiphiles are packed in the micelle so as to have a fixed volume v. Further, it is assumed that the packing chemical potential per amphiphile in the micelle is given by

$$\underline{\mu}_s^o \equiv \mu_s^o/s = (\gamma/a)(a - a_o)^2 + \underline{\mu}^o \tag{14.11}$$

where γ is the interfacial tension, a is the surface area per molecule, a_o is the optimal head group area and $\underline{\mu}^o$ is the packing chemical potential at the optimum surface area per molecule a_o. In the case of spherical micelles, the surface area per molecule is given by $a = (36\pi v^2)^{1/3} N^{-1/3}$. But the packing chemical potential equation 14.11 has a minimum at the surface area per molecule a_o. The corresponding aggregation number $M^* = (36\pi v^2)/a_o^3$. The free energy can be expanded around this minimum to yield

$$\underline{\mu}_s^o = \text{const.} + (\gamma a_o/9M^{*2})(s - M^*)^2 \tag{14.12}$$

as a function of the deviation in size from M. Inserting this into equation 14.3 one obtains a size distribution for spherical micelles which is approximately Gaussian

$$X_s \propto \exp[-(s - M^*)^2/2\sigma^2] \quad \text{with } \sigma = [9kTM^*)/2\gamma a_o]^{1/2} \tag{14.13}$$

Hence, in this phenomenological model, the sizes of spherical micelles are distributed around a most probable aggregation number M*, which depends only on molecular details of the surfactants in this simplest approximation. A Monte Carlo simulation reveals a Gaussian distribution about an aggregation number, i.e. the thin line in Figure 14.10. Also in this figure are the calculated radii of gyration, which show that the peak corresponds to spherical micelles while the *tail of the peak reveals a cylindrical symmetry.* This simulation of Nelson et al.[15] assumes amphiphiles having structure different from those used in the nucleation studies that gave rise to the results in Figures 14.3 and 14.4(a). These nucleation studies yielded no evidence for cylindrical micelles, but did reveal that proper micelles transformed via a first-order transition to bilayer vesicles or disks for the shape of amphiphiles used in these studies, i.e. nearly equal size of head and tail diameters in that both head and tail occupied sites on a lattice[1] or both head and tail particles

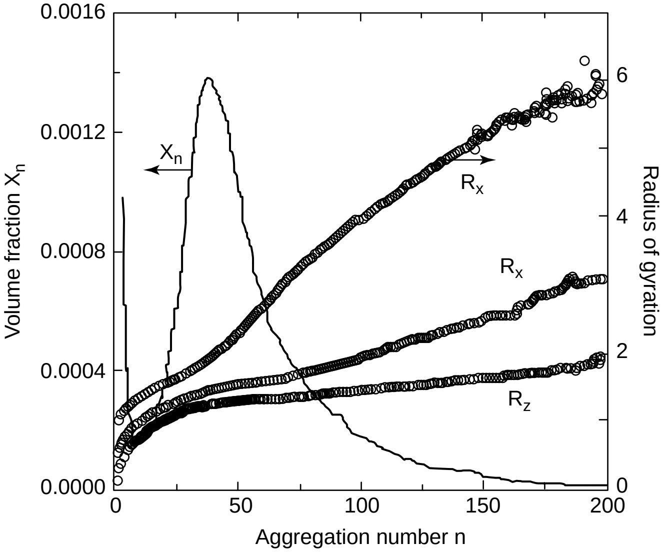


Figure 14.10. Results of a Monte Carlo simulation showing the equilibrium size distribution of micelles and the three radii of gyration of the micelle as a function of the number of amphiphile molecules in the micelle. From J. Chem. Phys. 107, 10777(1997) with permission. © 1997 American Institute of Physics.

had the same diameters.[1] The amphiphiles of Nelson et al.[15] have a head cross-sectional area larger than that for the tail and are likely to have a packing factor corresponding to the region where cylindrical micelles are stable.

The fact that the proper micelles which spawn the cylindrical micelles are spherical in shape is apparent from the equality of the radii of gyration at the peak in the size distribution of Figure 14.10. However, it is also apparent that the work to continue to grow in the cylindrical form must be smaller than that to form a bilayer disk or vesicle. How the transition from spherical to cylindrical shape occurs has been considered in some recent papers and will be discussed later.

The micellar size distribution obtained by Nelson et al.[15] and shown in Figure 14.10 reveals a tail in which cylindrical micelles predominate. As they noted the tail distribution fits an exponential distribution of the form $X_s \propto \exp(-\alpha s)$, where the prefactor interpolates smoothly between f(s) = 0 at s = M* and f(s) = 1 at s >> M*. Now, it can be shown that this is exactly what is expected from equation 14.3 (i.e. $\alpha = \ln X_1 - (\mu_1^o - \mu_s^o/s)/kT$). Hence, the size distribution of micelles appears to be explained satisfactorily.

Another aspect of the simulation results shown in Figure 14.10 is that both spherical and cylindrical micelles exist in the same solution. As was shown in

Section 2.3 of this chapter, the chemical potential per amphiphile in the cylindrical micelle, and hence that of the free amphiphile in solution, approaches the constant value of the packing chemical potential for the cylindrical micelle. Now, there will be a distribution of the spherical micelles such that the chemical potential per amphiphile in these micelles equals that for the cylindrical micelles. Since, the packing chemical potential per amphiphile in the proper spherical micelle does not differ much from that in the cylindrical micelle, the population densities of these micelles will not differ greatly. In effect, the micelles act to sponge up the excess amphiphiles in solution above the CMC as the total concentration increases and X_1 tends to approach a constant value. Once this constant value is approached then the total population distribution and population of the spherical micelles must remain constant. This is the origin of the second CMC. The second CMC is characterized by a constant X_{sph} for $X >$ CMC (second). This process is illustrated schematically in Figure 14.11. X_{sph} is the total amphiphile concentration contained in the spherical micelles. A derivation of the X_{sph} follows.

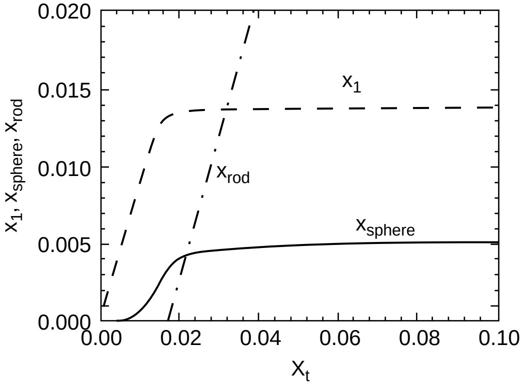


Figure 14.11. Showing first and second CMC.

From equation 14.6b we have $Z_s = M'\exp(-\beta[\Delta F_s - r\mu_o - s\mu_1])$. But, now we are at the condition where the cylindrical micelles have dimensions such that $\mu_s = \mu_s^o = \mu_1$. But, $\Delta F_s - r\mu_o - s\mu_1 = \Delta\Omega_s$ and $\mu_1 = \mu_1^o + kT\ln X_1$. Substituting in

$$X_{sph} = (1/N)\Sigma sZ_s$$

where the sum is over the spherical micelles, a fixed range of $s_{nucleus}$ to s_{proper}, we obtain

$$X_{sph} = (M'/N)\Sigma s\exp[-\beta\,\Delta\Omega_s]$$

But, as noted, μ_1 is a constant when sufficiently long cylindrical micelles are present. Hence, X_1 is constant and the right-hand side of this equation is constant from Figure 14.4(a), and consequently X_{sph} is a constant also.

Larson,[16] as a consequence of simulation studies, does report a phase diagram which shows a concentration region where cylindrical micelles and lamellar aggregates exist in a binary system of amphiphiles and water. The coexistence of micelles and bilayers has been observed experimentally. Bilayer aggregates and mixtures of such aggregates have been found to form in the absence of other micelles under equilibrium conditions. One such example is discussed in the next section. Under non-equilibrium conditions a sequence of formation of spherical micelles, bilayer oblate disks, and bilayer spherical vesicles has been reported in a simulation. This phenomenon is also discussed in the next section. Much remains to be studied about micelle equilibria.

3. Vesicles

In this section we will consider spontaneous transitions to the vesicle state, whether the vesicle is the state of thermodynamic equilibrium or not. At least, the transition will be in the direction of a decrease in the free energy of the system.

3.1. Bilayer disk to vesicle transition

Returning to Figures 14.3 and 14.4(a) we notice that for the curve corresponding to ϕ_4 the work to form an aggregate decreases with increasing size of the aggregate once the aggregate is nucleated. The free energy of the system decreases likewise. This aggregate has been found to be a spherical bilayer vesicle with solvent inside the vesicle. Hence, vesicle formation is spontaneous for the parameters applicable to this simulation. Unfortunately, at this writing there is no knowledge relating to how the spherical vesicles considered in Figures 14.3 and 14.4(a) are formed. It may be that a finite disk-like bilayer is formed first and then the spherical vesicle is formed by closing of the disk. Such a mechanism appears to operate at higher amphiphile concentration as revealed in another simulation for constant temperature, pressure and numbers of water and amphiphile molecules. The steps in aggregation from a random solution of amphiphiles in the latter simulation are shown in Figure 14.12(a). Here we see that the first aggregate formed is a spherical micelle. Upon the formation of many such micelles, collisions between them lead to the formation and growth of an oblate bilayer disk, which then fluctuates and forms the spherical vesicle. This mode of formation of the spherical vesicle occurred for several types of substitute amphiphiles in this simulation procedure. Yamamoto et al.[12] then showed that this spherical vesicle was the equilibrium product for the conditions of the simulation by starting the simulation run with an already formed bilayer disk and observing that the final state of the system was the same (a spherical bilayer vesicle) as that when the run is started from a random solution of the amphiphiles see Figure 14.12(b). (Incidentally, this is a beautiful demonstration of the phenomenon known as Ostwald's "staircase" where intermediate states of metastability are achieved in the descent from a high free energy state to the stable state.) This tendency for cluster–cluster coalescence and breakup

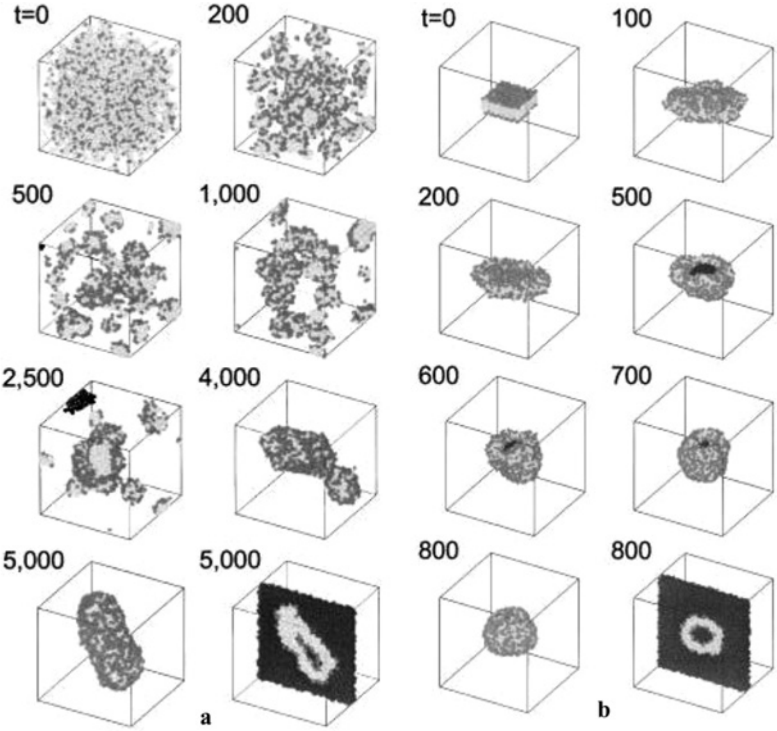


Figure 14.12. Monte Carlo simulation; (a) starting from a homogeneous supersaturated solution of amphiphilesl (b) starting from a bilayer disk. From J. Chem. Phys. 116, 5842(2002) with permission. © 2002 American Institute of Physics.

(Smoluchowski type) has been found to occur when conditions are far from equilibrium and the Becker–Doring type of stepwise addition or removal to occur closer to equilibrium.[13]

Several other studies[17] involving simulations, but using different models, found the same sequence was followed upon starting with a homogeneous supersaturated solution of amphiphiles, i.e. formation of spherical micelles with concomitant depletion of concentration of free amphiphiles in solution, collision and merging of spherical micelles to form finite planar bilayer aggregate with rim, and then transition of bilayer aggregate to vesicle enclosing solvent.

Let us consider what is known about the thermodynamic forces that drive or prevent the formation of vesicles. The feature of vesicles responsible for their relative thermodynamic stability in the simulations that produced Figures 14.3, 14.4, and 14.12 is the absence of a high-energy rim in their vesicle form, a rim that is present in the planar bilayer disks. However, the vesicle has a source of energy increment not present in the planar bilayer disk. The curvature of the vesicle form increases its energy relative to that for the planar bilayer. A model of the transition from bilayer disk to spherical vesicle due to Lipowsky[18] is based on the balance between these two sources of energy, as follows.

The following analysis applies primarily to non-ionic amphiphiles. It assumes that the total concentration of amphiphiles is more than enough to produce the situation where at least bilayer disks, if not membranes, are stable relative to a solution of amphiphiles of that concentration. The rim of a bilayer disk may be considered to be a line of positive surface tension even though the planar areas of the bilayer disk may have negative surface tension. See Figure 14.13 which shows the dependence of the calculated line tension ($\Gamma\sigma/kT$) in units of amphiphile diameter σ for typical surfactant parameters over a range of amphiphile concentration. (Extrapolation of the curve indicates that above about a volume fraction of amphiphile equal to $0.0014\pi/6$ the line tension would become negative.) The transition from bilayer to vesicle involves an invariant area. Consequently, the energy associated with this area does not contribute to the energies that change during the transition. (At the coexistence composition this specific surface energy is zero in the planar membrane.) There are only two terms in this relation for the work of nucleation of a vesicle from a finite bilayer disk: the positive energy associated with the line tension of the rim ($2\pi r\Gamma$) and the positive energy associated with the curvature of the

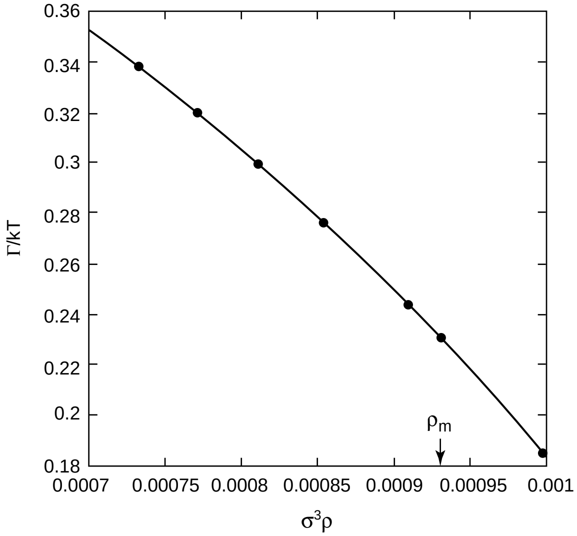


Figure 14.13. Line tension as a function of amphiphile concentration. From J. Chem. Phys. 118, 872(2003) with permission. © 2003 American Institute of Physics.

bilayer ($\pi r_o^2 K\kappa^2$). The curvature κ and the radius of the circular rim are related to each other by the condition that the area of the aggregate remains constant during the transition. The resulting relation for the disk energy [$E = 2\pi\Gamma r_o(1 - r_o^2\kappa^2/4)^{1/2} + \pi r_o^2 K\kappa^2$], where r_o is the radius of the parent bilayer disk, Γ is the line tension of the bilayer disk rim and K is a modified stiffness coefficient ($=2\underline{k} + k^*$, where $\underline{k}$ is the bending modulus and k* is the Gaussian modulus) in the expression for the curvature energy, yields a maximum as the curvature increases. The maximum energy less the initial energy of the bilayer disk corresponds to the work of nucleation of the vesicle starting from a bilayer disk. The value of the work of nucleation can be obtained from the above to equal

$$\Delta E^* = \pi\{\Gamma r_o[\Gamma r_o/(4K) - 2] + 4K\}$$

For $\Gamma r_o > 2K$ the energy of the planar disk is higher than the sphere formed from the disk, i.e. the spherical vesicle is stable relative to the disk. For $r_o > 4\,K/\Gamma$, ΔE^* is zero, i.e. the disk can spontaneously transform to the spherical vesicle given a thermal fluctuation of the curvature of wavelength larger than $4K/\Gamma$.

This result is consistent with the simulation shown in Figures 14.12. It is also desirable to confront the above theory against experiment as well. Insofar as the initial condition involves a homogeneous supersaturated distribution of amphiphile molecules in solution experiment seems to bear out the theory in that spontaneous vesiculation is observed when an amphiphile solution is quenched from a higher temperature to a lower temperature to produce the supersaturated solution of amphiphile molecules.[19a] Also, when two solutions of ionic amphiphiles, one cationic and the other anionic, are mixed producing a supersaturated solution of the mixed surfactants, spontaneous vesiculation occurs.[19b]

Bilayer disks may be produced by other methods not involving supersaturated solutions. For example, bilayer disks are produced by fragmenting already existing membranes, mechanically or otherwise. Spherical vesicles are observed subsequent to the fragmentation process. Spontaneous vesiculation of the bilayer fragments is the most likely origin of the observed spherical vesicles. In support of this suggestion we note that in the Moscho et al.[20] recipe for producing giant unilamellar vesicles (GUVs), schematically illustrated in Figure 14.14, the technique forms large fragments of bilayer disks that are not produced by fragmenting larger bilayer membranes and that undergo spontaneous vesiculation. The diameters of the fragments vary up to about 50 μm. Among the techniques for fragmentation of membranes is the use of an AC field that results in GUVs larger than 8000σ.[21] Indeed, most of the techniques of producing GUVs involve fragmentation of bilayer membranes.

Vesicles produced by the bilayer to disk transition may or may not be thermodynamically stable dependent upon the conditions of the aggregate

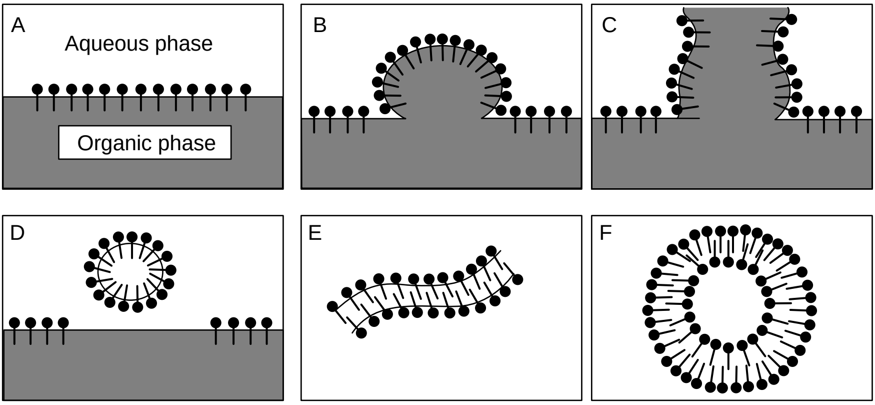


Figure 14.14. Proposed mechanism of formation of GUVs. Boiling in organic phase produces bubbles (B) that act to rupture monolayer of amphiphiles at water/organic interface (C). Fragments of monolayers join to form bilayer disk (E) which then spontaeously form GUV. From Proc. Nat. Acad. Sci. USA 93, 11443(1996) with permission.

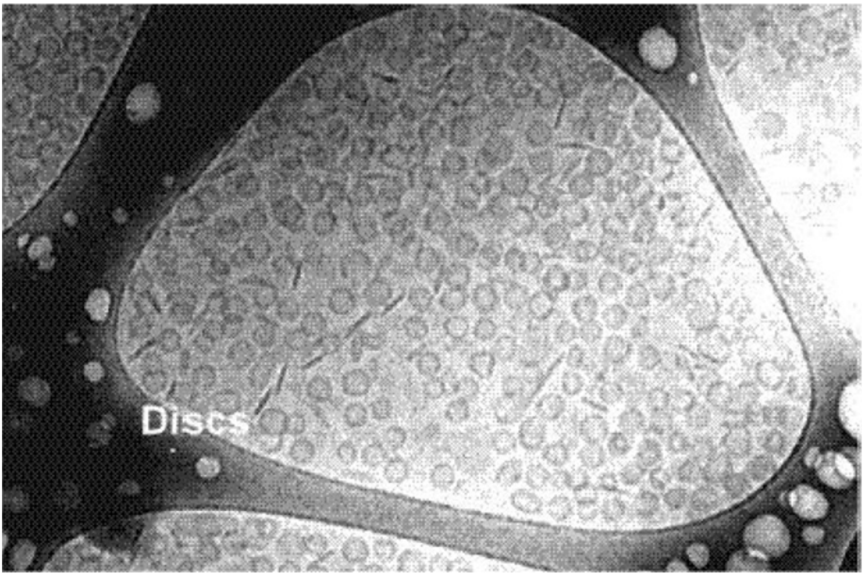


Figure 14.15. Electron micrograph showing disks (dark lines) and vesicles in a mixed amphiphile system. The disks have a radius of 35 ± 4 nm while the vesicles have a radius of 25 ± 2.9 nm, both independent of composition. From Proc. Nat. Acad. Sci. USA 99, 15318(2002) with permission.

environment. It is worthwhile to obtain some “ballpark” numbers for the range of r_o over which we may expect spherical unilamellar vesicles to be stable relative to planar unilamellar bilayer disks. From simulations[22] mimicking one-component amphiphile solutions that produce vesicles it is possible to obtain information concerning the rough values of K and Γ: K/kT has a value between about 1 and 100 while $\Gamma\sigma/kT$ varies with amphiphile concentration and bilayer surface tension and for these solutions varies between about 0.05 and 0.15, where σ is the diameter of an amphiphile in the plane of the bilayer. Thus, for these amphiphilic solutions metastable vesicles should be produced spontaneously from bilayer disks with radius larger than between about 10σ and 4000σ. Figure 14.15 shows electron micrographs of vesicles and planar bilayer disks that are present simultaneously.[23] The data suggest that the disk radii observed are smaller than $2\,K/\Gamma$while the vesicles have a radius with the corresponding disk radius larger than $2\,K/\Gamma$. K and Γ values can be chosen from the range of possible values to satisfy these results.

Summarizing, it is likely that the disk to vesicle transition will occur spontaneously without prior budding or fission in solutions originating as supersaturated solutions of amphiphiles or in systems where membranes have been fragmented to bilayer disks.

4. Phase equilibria of micellar aggregates

The aggregates of amphiphiles provide a rich phase equilibria. A schematic illustration of a real ternary phase diagram for the amphiphile–water–oil system is shown in Figure 14.16 with drawings of the phases corresponding to the various regions. For a binary system of amphiphiles and water a partial phase diagram deduced from a Monte Carlo simulation due to Larson[16] is shown in Figure 14.17 which acts to make the compositions of the phases more real. The phase designations are as follow: L_α is a lamellar phase consisting of stacks of lamellar bilayer membranes, H_1 is a hexagonal phase stacking of cylindrical micelles, V_1 the gyroid phase has cubic symmetry and is constituted by two intertwined networks, and L_3 is the sponge phase, a macroscopically isotropic but microscopically structured bicontinuous phase. The curvature of the elements of the gyroid phase is intermediate between the zero curvature of the lamellae in the lamellar phase and the larger curvature of the cylindrical vesicles in the hexagonal phase. This sequence may exist because the interaggregate space increases from the lamellar to the hexagonal phase and the water volume fraction also increases in this sequence. What may be more significant about this sequence is that the translational entropy (free volumes) of the ordered arrays of hexagonally stacked cylindrical micelles and the layered stacking of lamellar bilayers at the compositions at which these aggregates appear are larger than for disordered arrays of spherical or cylindrical or bilayered

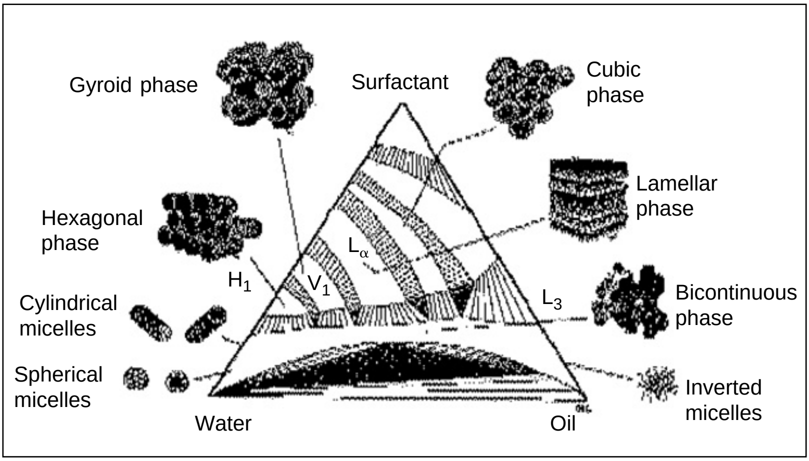


Figure 14.16. Real ternary phase diagram for amphiphile–water–oil system.

micellar aggregates. Indeed, there is a germ of truth in this statement that needs to be examined in greater detail.

Starting at the water end of a binary system of water and amphiphile we note that the first aggregate phase region consists of an isotropic array of spherical and cylindrical micelles. As the amphiphile concentration increases there is a phase change to a hexagonal array of columns of cylindrical micelles (except for case c) in Figure 14.17). The transition from spherical to cylindrical micellar shape in the solution phase adjacent to the water terminal comes about because micelles have the characteristic of being able to change shape at a constant amphiphile number in the aggregates, a characteristic which hard spheres do not share. Hence, it is possible for spherical micelles to change shape to cylindrical micelles at constant number of amphiphiles in micelles and it is conjectured to thereby decrease the free energy by increasing the free volume (i.e. the translational entropy) and the entropy associated with the rotational disorder of the cylindrical micelles. This conjecture is supported by model calculations by Taylor and Herzfeld[24] which show that rods are present at equilibrium in the isotropic solution near the coexistence composition between isotropic and hexagonal phases *despite the existence of a repulsive interaction between the spherical-like micelles tending to oppose their merger to form the rod-like micelles.* Then, as first suggested by Ben-Shaul and Gelbart,[25] as the free volume decreases with increase in total amphiphile concentration, the translational and rotational entropy of the disordered array of cylindrical micelles becomes smaller than that for the ordered hexagonal array of cylindrical micelles. Thus, the stable phase at the higher amphiphile concentration becomes the hexagonal phase of

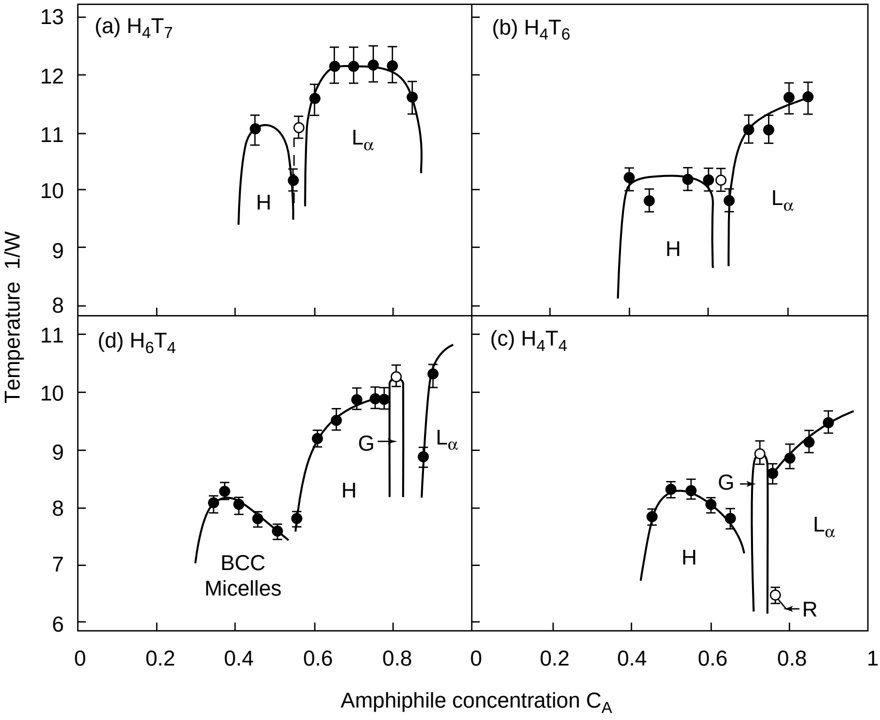


Figure 14.17. Calculated phase diagrams for the water-amphiphile system for various amphiphiles. See text for definitions of phases. From J. Phys. II France 6, 1441(1996) with permission.

cylindrical micelles. With similar reasoning bilayer disks in a lamellar array will, beyond some amphiphile concentration, have larger free volume than the same number of amphiphiles in cylindrical micelles when arrayed in a hexagonal phase. Qualitatively, this explains the sequence of phases along the water–amphiphile axis with the exception of the gyroid phase. However, demonstration of this conjecture in a rigorous manner is quite difficult. Nevertheless, there have been simulations and analytical studies that support the above conjecture.

We have already shown the results of Monte Carlo simulations based on a lattice model in Figure 14.17. The phase diagram of the rod-like model of Taylor and Herzfeld is shown in Figure 14.18. In this figure the higher the temperature the more significant is the role of entropy in the phase equilibria. The existence of rod-like micelles at the temperature d of Figure 14.18 is revealed by the size distribution shown in Figure 14.19, where the length of the rod relative to the diameter of the spherical micelle is given by n, the abscissa in the figure. The increase in rod

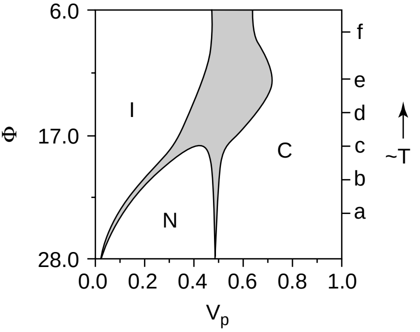


Figure 14.18. Calculated phase diagram for hard rods. From Langmuir 6, 911(1990) with permission. © 1990 American Chemical Society.

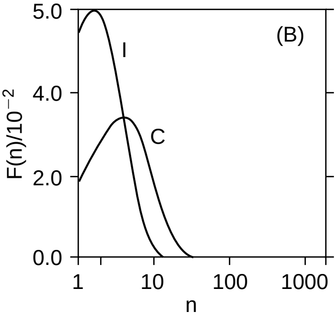


Figure 14.19. Size distribution of rods at temperature d of Figure 14.15 for isotropic I and columnar C phases. From Langmuir 6, 911(1990) with permission. © 1990 American Chemical Society.

length in the transition from isotropic to the columnar hexagonal phase has been considered by Ben-Shaul and Gelbart[25] and one may deduce from their analysis it occurs as a consequence of the increase in translational entropy which counters the loss of mixing entropy on the transition from the isotropic to the columnar morphology.

In view of a recently published study showing higher packing fraction for random disks than spheres it would be interesting to know whether disk-like micelles appear in the isotropic phase near the phase boundary with the lamellar phase (i.e. the high-temperature region of Figure 14.17 where the phase boundary of the lamellar phase is above 60 volume percent). Does entropy affect the micellar shape at volume fractions of micelles where intermicellar contact becomes appreciable, as is implied by the above analysis?

Some support for this concept is found in the Molecular dynamics simulation of Guo and Kremer[26] who modeled the amphiphile system for several amphiphiles in the absence of water. They found that in the amphiphile melt at high temperature that the melt consisted of randomly distributed small “sheets”. We take the sheets to correspond to disk-like micelles. On lowering the temperature the sheets ordered via a first-order transition into layers as shown in the sequence of snapshots in Figure 14.20 taken from their work.

Although the hexagonal and lamellar phases are believed to represent the phases with the highest entropy of translation in their range of stability, apparently, according to the work of Taylor and Herzfeld[24] and Cuesta and Sear,[27] repulsive interactions between the infinite micellar units also contribute to the stability of these structures. The origin of the stability of the gyroid phase, which exists in the

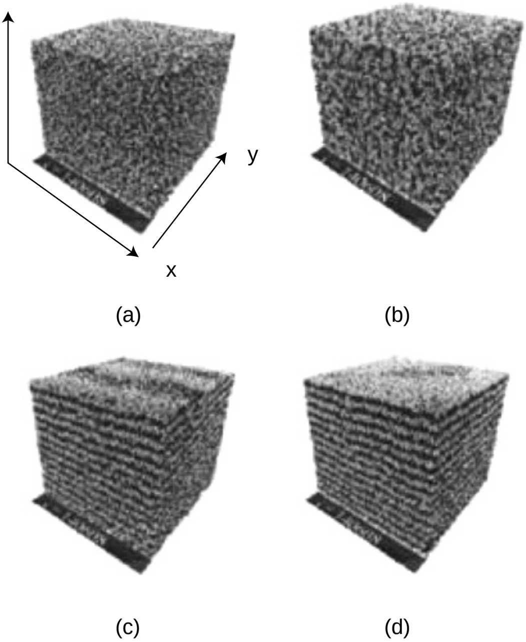


Figure 14.20. Snapshots during MD simulation. Temperature decreases along sequence (a)–(d). First-order transition occurs between (b) and (c). Order parameter increases form (c) to (d). From J.Chem. Phys. 119, 9308(2003) with permission. © 2003 American Institute of Physics.

composition range between that for the hexagonal and lamellar phases is not known. However, it appears as a stable phase in the Monte Carlo simulation of Larson, as shown in Figure 14.17.

References

1. V. Talanquer and D.W. Oxtoby, J. Chem. Phys. 113(16), 7013(2000).
2. M. Nilsson and S. Rasmussen, Discrete Mathematics and Theoretical Computer Science AB(DMTCS), 31(2003).
3. N.A.M. Besseling and M.A. Cohen Stuart, J. Chem. Phys. 110, 5432(1999).
4. I. Kusaka and D.W. Oxtoby, J. Chem. Phys. 115, 4883(2001).
5. I. Napari and A. Laaksonen, Phys. Rev. Lett. 84, 2184(2000).
6. P.S. Christopher and D.W. Oxtoby, J. Chem. Phys. 118, 5665(2003).
7. Talanquer and D.W. Oxtoby, J. Chem. Phys. 103, 3686(1995); P.S. Christopher and D.W. Oxtoby, J. Chem. Phys. 117, 9502(2002).

8. N.F. Carnahan and K.E. Starling, J. Chem. Phys. 51, 635(1969).
9. J.N. Israelachvili, **Intermolecular and Surface Forces**, Academic Press, London, 1992, p. 381.
10. C. Tanford, **The Hydrophobic Effect**, 2nd edition, John Wiley, New York, 1980.
11. J.N. Israelachvili, D.J. Mitchell, B.W. Ninham, J. Chem. Soc. Faraday Trans. 2, 72, 1525(1976).
12. S. Yamamoto, Y. Maruyama and Shi-aki Hyodo, J. Chem. Phys. 116, 5842(2002).
13. J.-B. Maillet, V. Lachetîa and P.V. Coveney, Phys. Chem. Chem. Phys. 1, 5277(1999).
14. S. May, Y. Bohbot and A. Ben-Shaul, J. Phys. Chem. B 101, 8648(1997).
15. P.H. Nelson, G.C. Rutledge, and T.A. Hattona, J. Chem. Phys. 107, 10777(1997).
16. R.G. Larson, J. Chem. Phys. 91, 2479(1989).
17. A.T. Bernardes, Langmuir 12, 5763(1996); J. Phys. 11, 6, 100(1996); J.M. Drouffe et al., Science, 254, 1353(1991); D. Brindle and C.M. Care, J. Chem Soc. Faraday Trans. 88, 2163(1992).
18. R. Lipowsky, Nature 349, 475(1991).
19. (a) T.D. Madden, C.P. Tilcock, K. Wong and P.R. Cullis, Biochemistry 27, 8724(1988); F.A. Nezil, S. Bayerl and M. Bloom, Biophys. J. 61, 1413(1992); H.G. Dobereiner, J. Kas, D. Noppl, I. Sprenger and E. Sackmann, Biophys. J. 65, 1396(1993); (b) E.W. Kaler et al., Science 245,1371(1989).
20. A. Moscho, O. Orwar, D.T. Chiu, B.P. Modi and R.N. Zare, Proc. Natl Acad. Sci. USA 93, 11443(1996).
21. M.I. Angelova and D.S. Dimitrov, Faraday Discuss. Chem. Soc. 81, 303.
22. V. Talanquer and D.W. Oxtoby, J. Chem Phys. 118, 872(2003).
23. H.T. Jung, S.Y. Lee, E.W. Kaler, B. Coldren and J.A. Zasadzinski, Proc. Nat. Acad. Sci. USA 99, 15318(2002).
24. M.P. Taylor and J. Herzfeld, Langmuir 6, 911(1990).
25. A. Ben-Shaul and W.M. Gelbart in **Micelles, Membranes, Microemulsions, and Monolayers,** eds. M. Gelbart, A. Ben Shaul, D. Roux, Springer-Verlag, New York, 1994.
26. H. Guo and K. Kremer, J. Chem. Phys. 119, 9308(2003).
27. J.A. Cuesta and R.P. Sear, Euro. Phys. J. B 8, 233(1999).

Index

Absolute reaction rate theory, 211
Acceptor type solute, 192
Activation energy, 207
  empirical rules for, 247
  for interstitial diffusion
    table of values, 257
Activity, 114
Adsorption
  Gibbs, 127
  Guggenheim, 129
  Langmuir–McLean, 133
Allotropy, 12
Amorphous phase
  at grain boundaries, 142
  metastable
    diffusion-induced growth of, 333
Amplification factor
  in spinodal decomposition, 273
Anisotropy of surface energy, 119
Arrhenius plot, 237
Artificial epitaxy, 119
Atom jump frequency, 238
Atom-vacancy interchange attempt frequency, 236
Average melting point, 93, 94

Band gaps, table of, 197
Becker and Doring
  theory of nucleation, 264
Bernal polyhedra, 140
Bifurcation
  subcritical, 385
Bilayer disks, 447
Bilayer to vesicle transition, 444
Boltzmann integration factor, 230
Bragg–Williams model, 27
Brouwer plot, 200

C curve of transformation temperature vs time, 269
Cahn, J.
  grain growth and solute diffusion, 310
  spinodal decomposition, 176, 271
Cahn–Hilliard equation, 272
CALPHAD, 18, 105, 108
Canonical ensemble and entropy, 50
Capillary length, 371
Carnahan–Starling equation of state, 21
Chemical affinity, 323
Chemical diffusivity, 271
Chemical potential
  in coherent equilibrium, 109
ChemSage, 108
Clapeyron equation, 15
Clapeyron relation, 110
Classical nucleation theory, 265
Coherent equilibrium, 109
Coherent interface, 143
Coil–globule transition, 34
Coil–lamellar crystal transition, 37
Coil–helix transition, 43
Communal entropy, 20
Competing processes
  independent processes, 214
Composition fluctuations, 174
Composition waves, 173
Compressibility, 15
Conduction band
  number of electrons in, 189
Conjugate gradient method, 212
Constitutional supercooling criterion, 370
Constitutionally supercooled zone, 348, 349
Consulate critical point, 110
Correlation factor, 237

Correlation parameter, 25
Coupled irreversible process, 214
Critical micelle concentration (CMC), 426, 432–434
Critical temperature
  regular solid solution, 71
Critically sized embryo, 163
Crowdion, 235
Crystallization velocity, 347, 350
Curie principle, 216
CVD number, 359
Cylindrical micelle, 430

Darken–Gurry plot, 78
Darken's relations, 232
Darken thermodynamics factor, 233
Debye solid, 8
Debye temperature, 54
  values, 8
Defects
  divacancy, 187
  electronic, 190
  energy of (table), 196
  Frenkel, 186
  interstitialcy, 186
  non-stoichiometric, 199
  point, 186
  Schottky, 186
  thermodynamics of, 186
  vacancy, 186
Detailed balancing of chemical reactions, 216
Diffuse interfaces
  width and energy of, 149
Diffusion
  in COO, 243
  cross terms in, 234
  in Fe_3O_4, 243
  in grain boundaries, 248
  mechanisms for lattice, 234
  multicomponent, 253
  phenomenological basis, 225
  in polymers, 256
  vacancy sources and sinks in, 228
Diffusion coefficients
  empirical relations for, 247
Diffusion couple
  between two phases, 319
  three phase, 322
Diffusion in ionic crystals, 234
Diffusion potential, 83, 322
  and coherent equilibrium, 109
  definition, 83
Diffusivity
  grain boundary
    chemical, 230
    intrinsic, 225
    in ionic crystals, 251
    in metals, 251
DIGM, 346
Dipole patterns, 400–404
Dissipative crystals, 410
  oscillons, 411
Distribution coefficient, 92, 348, 350
  in bulk/interface equilibrium, 131
  liquid/solid, 92
Donor solute, 192
Driving force, 290
  primary recrystallization, 290
  recrystallization, 290
  secondary recrystallization, 291
  for various processes, 292

Einstein characteristic temperature, 7
Einsteinian monatomic solid, 7
Electromigration, 252
Electron drag, 234
Electronegativity
  Pauling's, 78
Ellingham diagram, 112
Embryos
  coherent and non-coherent, 167
  critical size, 163
  effect of stress on shape of, 167
  heterophase, 160
  number at equilibrium, 161
  work to form, 162
Energies of diffuse
  interfaces, 148
Energies of materials
  uncertainties in calculated values, 4
Energy versus Bain strain, 5

Energy of formation
  uncertainty in values of, 4
  modes of calculating, 47
Enthalpy of mixing, 67
Entropy
  amorphous solids, 14
  and canonical ensemble, 6, 50
  configurational, 7
    binary random solution, 64
    polymer alloy, 81
  covalent solids, 13
  electronic, 10
  of hard sphere system, 19
  magnetic, 10
  of planar polymer, 52
  of polymer alloy, 81
  random binary solid solution, 63
  thermal, 7
Entropy of fusion, 86
Entropy of gases, 49
Entropy of transformation, 99
Entropy production rate, 216
  minimum, 218
Equilibrium
  chemical, 110
  distribution of electrons, 189
  distribution of holes, 190
  divacancy-vacancy, 188
Error function, 248
Eutectic phase diagram
  origin of, 96
Eutectic solidification
  forced-velocity, 387
    solvability limit to, 395
  marginal stability condition for, 390
Excess free energy, 63, 93
  composition dependence of, 106
Excess thermodynamic quantities, 87

Fermi energy, 190
Ferromagnetic-paramagnetic transformation, 13
Fick's first law of diffusion, 227
Fick's second law of diffusion, 230
First-order transformation, 12
Fisher's model for grain boundary diffusion, 249
Fluctuations
  composition, 174
  heterophase, 160
Free energy
  of ideal solutions, 63
  of mixtures, 66
  of regular solutions, 70
Free energy and phase transitions, 13, 15, 18
Free energy–composition curve
  of intermediate phases, 99

Gibbs' adsorption relation, 128
Gibbs-Duhem relation, 69, 128
Gibbs free energy, 3
  empirical relations for excess, 107
  of mixture of solid solutions, 67
  of random binary solution, 64
Gibbs nucleation theory, 267
Ginzburg–Landau
    equation, 222, 279, 337, 418
Gradient-energy ($K\nabla C^2$), 172
Grain boundary diffusion
  mechanisms of, 251
Grain boundary migration
  mechanisms of
    in homogeneous alloys, 307
    in pure materials, 303
Grain boundary mobility, 299
  vs relative orientation, 301
  table of values, 303
Graphoepitaxy, 119
Growth, 281, 328
  interface vs diffusion control, 328, 337
  metastable amorphous phase, 333
  of precipitates, 333
Growth rates
  in normal grain growth, 297
  in primary recrystallization, 292
  in secondary recrystallization, 296

Harmonic oscillator, 8, 53
Heat capacity, 53

Helmholtz free energy, 127
  dependence on composition
    gradients, 272
Henry's law, 81
Heterogeneous equilibrium, 110
Heterophase fluctuations
  homogeneous distribution of, 160
  images of, 179
Hume-Rothery's rule, 76

Ideal solution, 88
IMN model, 436
Interface energies
  of soft matter, 155
Interface migration
  between crystalline and amorphous
    phases, 308
  effect of interface roughness on,
    309
  mechanisms of, 303
  in partitionless phase transformations,
    308
  during primary recrystallization, 290
Interface reaction, 322
  vs diffusion control, 337
Interface stabilized bulk metastable
    phase, 154
Interlamellar spacing limits
  in forced velocity eutectic
    solidification, 391
Interphase interfaces, 142
Ionicity, 79
Irreversible processes, 215
  in approach to equilbrium, 215
Ising problem, 11

Jackson and Hunt
  eutectic solidification, 387
Jahn-Teller distortions, 178

Kauzmann temperature, 14
Kirkendall effect, 228
Kolmogorov, Johnson-Mehl and Avrami
    equations (KJA), 281
Kroger–Vink plot, 200
Kronberg–Wilson, 301

Landau model of phase transition, 177
Langmuir layers
  polymorphs of, 21
Langmuir–McLean isotherm, 133
Lattice stability, 106
Level set simulation, 222
Local equilibrium
  at grain boundary intersections, 144
Long-range order, 24

Marginal stability condition, 390
  for eutectic solidification, 390
Mass action law, 198
Matano interface, 231
Mechanisms of diffusion
  in alloys, 239
  in compounds, 240
  in metals, 235
  in oxides, 243
  in semiconductors, 245
Metallic solutions
  strain free energy, 75
Metastability, 102, 151
Metastable phase
  pseudomorphic stabilization of, 150
Micellar aggregates
  phase equilibria of, 449
Micelles, 425
  characteristics, 430
  critical micelle concentration, 432–434
  equilibrium size distribution and
    coexistence, 440–444
  free energy per unit volume, 430
    of amphiphile solution, 430–431
  nucleation, 425–430
  shape, factors controlling, 435–440
  transitions to bilayers, 430
Microscopic reversibility principle, 216
Microscopic solvability condition, 389
Microstructure map of cellular patterns,
    386
Miscibility gap compositions
  regular solid solution, 71
Mobility, 232
Modulated surface patterns, 404–405
Molar free energy, 15, 87, 91

Molecular dynamic simulation, 212
  computer codes for, 212, 221
Monte Carlo simulation, 212, 219
Morphological instability
  in CVD, 372
  in PVD, 375
  of solidification front, 367
Morphological stability criterion
  of Mullins and Sekerka, 370
  of Trivedi and Kurz, 370
Mott-Littleton two region strategy, 212
MTDATA, 108
Mullins and Sekerka, 371
  morphological stability criterion, 371

Nabarro shape function, 168
Nernst–Einstein relation, 245
Network solid, 83
Nucleation
  density functional theory and, 276
  heterogeneous or catalyzed, 270
  heterophase
    steady-state rate of, 263
    time-dependent rate of, 265
  homogeneous, 263
  homophase, 271
  of micelles, 425–430
    and heterophases, 435
  steady-state rate of, 263
  theorem, 280
  work of, 163
Nucleation rate
  temperature of maximum in, 270
  time dependent, 265

Onsager reciprocal relations, 215, 218
Order
  long-range, 27
  short-range, 25
Order–disorder transformation, 13
Oscillons, 411
Ostwald ripening, 124, 340
Ostwald's staircase, 444

Partial molar quantities, 87
Partial pressure, 111
Particle size
  and chemical potential, 123
  effect on
    pressure in particle, 121
    solvus composition, 124
Partition function, 51
  harmonic oscillator, 209
Pattern formation, 403
Periodic patterns
  dipole, 403
  in dissipative systems, 412
  in miscibility gap system, 403
  surface, 406
  of thermodynamic origin, 401
Phase diagrams
  ionically bonded solids, 96
  for regular solutions, 14, 104, 105
Phase field method, 221, 283
Phase refining, 403
Phase transition
  via Bain strain, 5
  Landau model of, 177
  order of, 12
Phonon drag, 234
Polymorphic structures
  relative stability of, 98
Polymorphic transition
  first-order, 9
Polymorphism
  effect on, of magnetic entropy, 18
  effect on, of pressure, 15
Precipitation
  early stage, 333
  late stage, 340
Premartensitic phenomena, 177
Pre-wetting along grain boundaries, 153

Rapid solidification, 353
Rayleigh–Benard system, 414
Reaction–diffusion system, 417, 418–419
Reaction path, 208
Recovery, 290
Recrystallization
  primary, 290
  secondary, 291

Recrystallization rates
  primary, 293
Recrystallization texture, 300
Redlich equation, 106
Regular solid solutions, 88, 105
Relative molar quantity, 87
Reptation, 256
Richard's approximation, 163
Richard's rule, 13

Sackur–Tetrode equation, 10
Saddle point energy, 211
Schottky defect, 185
Segregation ratio
  bulk/surface, 131
Sekerka capillary parameter, 372
Semiconductors
  extrinsic, 191
  intrinsic, 189
Serpentine stripe pattern, 403
Short-range order, 24
Short-range repulsion, 405–406
Simulation methods, 212
  static lattice, 212
Sintering, 359
Solid solubility
  effect of electronegativity
    difference, 78
  effect of ionicity, 78
Solid solutions, 63
Solidification
  cellular and dendritic patterns in, 385
  dendritic, 377
  effect of convection in, 342
  eutectic, 387
  multicomponent, 348
  rapid, 350
  slow crystallization rate, 347
Solute diffusion length, 371
Solute drag phenomenon, 310
Solute trapping, 353
Solvability theory, 378
Spatial periodic patterns, 399
  bulk matter, 406–410
  dipole patterns, 400–404
  dissipative crystals, 410
  modulated surface patterns, 404–405
  short-range repulsion, 405–406
Special grain boundaries, 141
Specific heat, 12, 53
Spherical micelle, 428
Spinodal, 172
  decomposition, 176, 271
Spontaneous spinodal decomposition, 176
Stirling's approximation, 64, 161
Stoichiometric solids
  approximate electron/hole
    concentrations in, 197
Strain embryos, 177
Strain entropy, 75
Stress and concentration gradients, 322
Superlattice, 24
Surface energy
  effect on melting point, 119
  Gibbs' definition of, 117
  model for, 118
Surface excess quantities, 127
Surface free energy values
  liquid/vapor (metals), 136
  liquid/vapor (salts, inorganics), 136
  solid/liquid, 138
  solid/solid, 142
  solid/vapor, 137
Surface phase transitions, 132
Surface reconstruction, 120
Surface tension, 134
Survival probability method, 283

Thermal diffusion length, 371
Thermal length, 371
Thermodynamic integration
  and free energy, 108
Thermodynamic potentials, 1, 3
Thermodynamic quantities
  definitions, 87
Thermophysical parameters
  relations between, 2
Thin film/substrate reactions, 330
Tilt and twist small angle boundaries, 138
Time-cone method, 283
Torque acting on grain boundary, 144
Truncated BET isotherm, 133

Uphill diffusion, 271

Vacancy
  formation enthalpies (table), 187
  virtual chemical potential of, 188
Vacancy–solute atom interaction, 239
Vapor deposition
  chemical, 358
  physical, 354
Vapor pressure
  temperature dependence of, 111
Vesicles, 444–449
Wetting transition, 152
Williams point
  and coherent equilibrium, 109
Wulff plot, 119
Wulff's relations, 122

Zeldovich factor, 265
Zener, C.
  iron phase transitions, 18
  particles and grain growth, 299
Zero creep method, 150
Zone melting, 349