Thermodynamics of Natural Systems

Greg Anderson

SECOND EDITION

CAMBRIDGE

CAMBRIDGE

more information - www.cambridge.org/9780521847728

Thermodynamics of Natural Systems

Second Edition

Thermodynamics deals with energy levels and the transfer of energy between states of matter, and is therefore fundamental to many branches of science. This new edition provides a relatively advanced treatment of the subject, specifically tailored to the interests of the Earth sciences.

The first four chapters explain all the necessary concepts of thermodynamics, using a simple graphical approach. Throughout the rest of the book the author emphasizes the use of thermodynamics to construct mathematical simulations of real systems. This helps to make the many abstract concepts accessible. Many computer programs are mentioned and used throughout the text, especially SUPCRT92, a widely used source of thermodynamic data. Links to useful information sites and computer programs as well as problem sets with detailed answers for instructors are available through http://www.cambridge.org/0521847729.

Building on the more elementary material in the first edition, this textbook will be ideal for advanced undergraduate and graduate students in geology, geochemistry, geophysics and environmental science.

GREG ANDERSON has been Professor of Geochemistry at the University of Toronto for 35 years and is the author of three textbooks on thermodynamics for Earth scientists: *Environmental Applications of Geochemical Modeling* (2002), *Thermodynamics in Geochemistry* (1993) and *Thermodynamics of Natural Systems* (1995). In 2000 he was awarded the Past President's Medal by the Mineralogical Association of Canada for contributions to geochemistry.

Thermodynamics of Natural Systems

Second Edition

G. M. Anderson
University of Toronto

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo

Cambridge University Press
The Edinburgh Building, Cambridge CB2 2RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge.org/9780521847728

First published in print format 2005

ISBN-13 978-0-511-12803-5 eBook (NetLibrary)
ISBN-10 0-511-12803-7 eBook (NetLibrary)

ISBN-13 978-0-521-84772-8 hardback
ISBN-10 0-521-84772-9 hardback

Contents

Appendices

1
What is thermodynamics?

1.1 Introduction

Thermodynamics is the branch of science that deals with relative energy levels and transfers of energy between systems and between different states of matter. Because these subjects arise in virtually every other branch of science, thermodynamics is one of the cornerstones of scientific training. Various scientific specialties place varying degrees of emphasis on the subject areas covered by thermodynamics – a text on thermodynamics for physicists can look quite different from one for chemists, or one for mechanical engineers. For chemists, biologists, geologists, and environmental scientists of various types, the thermodynamics of chemical reactions is of course a central concern, and that is the emphasis to be found in this book. Let us start by considering a few simple reactions and the questions that arise in doing this.

1.2 What is the problem?

1.2.1 Some simple chemical reactions

A chemical reaction involves the rearrangement of atoms from one structure or configuration to another, normally accompanied by an energy change. Let's consider some simple examples.

- Take an ice cube from the freezer of your refrigerator and place it in a cup on the counter. After a few minutes, the ice begins to melt, and it soon is completely changed to water. When the water has warmed up to room temperature, no further change can be observed, even if you watch for hours. If you put the water back in the freezer, it changes back to ice within a few minutes, and again there is no further change. Evidently, this substance (H_2O) has at least two different forms, and it will change spontaneously from one to the other depending on its surroundings.
- Take an egg from the refrigerator and fry it on the stove, then cool to room temperature. Again, all change seems now to have stopped – the reaction is complete. However, putting the fried egg back in the refrigerator will not change it back into a raw egg. This change seems not to be reversible. What is different in this case?

- Put a teaspoonful of salt into a cup of water. The salt, which is made up of a great many tiny fragments of the mineral halite (NaCl), quickly disappears into the water. It is still there, of course, in some dissolved form, because the water now tastes salty, but why did it dissolve? And is there any way to reverse this reaction?

Eventually, of course, we run out of experiments that can be performed in the kitchen. Consider two more reactions:

- On a museum shelf, you see a beautiful clear diamond and a piece of black graphite side by side. You know that these two specimens have exactly the same chemical composition (pure carbon, C), and that experiments at very high pressures and temperatures have succeeded in changing graphite into diamond. But how is it that these two different forms of carbon can exist side by side for years, while the two different forms of H_2O cannot?
- When a stick of dynamite explodes, a spectacular chemical reaction takes place. The solid material of the dynamite changes very rapidly into a mixture of gases, plus some leftover solids, and the sudden expansion of the gases gives the dynamite its destructive power. The reaction would seem to be nonreversible, but the fact that energy is obviously released may furnish a clue to understanding our other examples, where energy changes were not obvious.

These reactions illustrate many of the problems addressed by chemical thermodynamics. You may have used ice in your drinks for years without realizing that there was a problem, but it is actually a profound and very difficult one. It can be stated this way: What controls the changes (reactions) that we observe taking place in substances? Why do they occur? And why can some reactions go in the forward and backward directions (i.e., ice→water or water→ice) while others can only go in one direction (i.e., raw egg→fried egg)? Scientists puzzled over these questions during most of the nineteenth century before the answers became clear. Having the answers is important; they furnish the ability to control the power of chemical reactions for human uses, and thus form one of the cornerstones of modern science.

1.3 A mechanical analogy

Wondering why things happen the way they do goes back much further than the nineteenth century and includes many things other than chemical reactions. Some of these things are much simpler than chemical reactions, and we might look to these for analogies, or hints, as to how to explain what is happening.

A simple mechanical analogy would be a ball rolling in a valley, as in Figure 1.1. Balls have always been observed to roll down hills. In physical terms, this is "explained" by saying that mechanical systems have a tendency to change so as to reduce their *potential energy* to a minimum. In the case of the ball on the surface, the potential energy (for a ball of given mass) is determined by the height of the ball above the lowest valley, or some other

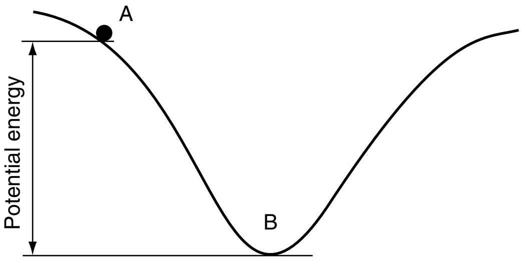


Figure 1.1 A mechanical analogy for a chemical system – a ball on a slope. The ball will spontaneously roll into the valley.

reference plane. It follows that the ball will spontaneously roll downhill, losing potential energy as it goes, to the lowest point it can reach. Thus it will always come to rest (equilibrium) at the bottom of a valley. However, if there is more than one valley, it may get stuck in a valley that is not the lowest available, as shown in Figure 1.2. This is discussed more fully in Chapter 2.

It was discovered quite early that most chemical reactions are accompanied by an energy transfer either to or from the reacting substances. In other words, chemical reactions usually either liberate heat or absorb heat. This is most easily seen in the case of the exploding dynamite, or when you strike a match, but in fact the freezing water is also a heat-liberating reaction. It was quite natural, then, by analogy with mechanical systems, to think that various substances contained various quantities of some kind of energy, and that reactions would occur if substances could rearrange themselves (react) so as to *lower* their energy content. According to this view, ice would have less of this energy (per gram, or per mole) than has water in the freezer, so water changes spontaneously to ice, and the salt in dissolved form would have less of this energy than solid salt, so salt dissolves in water. In the case of the diamond and graphite, perhaps the story is basically the same, but carbon is somehow "stuck" in the diamond structure.

Of course, chemical systems are not mechanical systems, and analogies can be misleading. You would be making a possibly fatal mistake if you believed that the energy of a stick of dynamite could be measured by how far above the ground it was. Nevertheless, the analogy is useful. Perhaps chemical systems will react such as to lower (in fact, minimize) their *chemical* energy, although sometimes, like diamond, they may get stuck in a valley higher than another

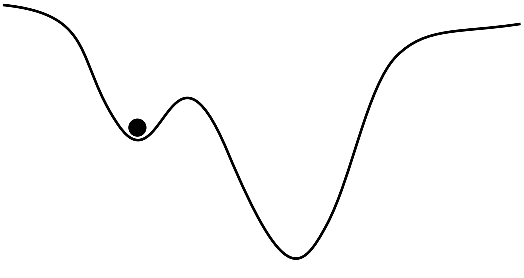


Figure 1.2 The ball has rolled into a valley, but there is a deeper valley.

Mechanics versus thermodynamics

Generally speaking, there are two main problems in learning thermodynamics. One, of course, is to learn the details of the specialized procedures in one's discipline, which in our case involves chemical reactions, activities, fugacities, equilibrium constants, and so on. The other, either more or less important depending on your point of view, is to understand thermodynamics as a whole – what is it, what do the variables mean, and how does it relate to other branches of science? In this book, the first four chapters deal mostly with this latter problem, and the rest of the book with the details.

At several points in these first few chapters we will make reference to classical mechanics, the science which deals with the motions of bodies and the forces causing those motions. This is an attempt to put thermodynamics into a broader perspective; to make analogies with possibly more familiar situations. Balls rolling in valleys and swinging pendulums are actually a special case in mechanics, known as dissipative systems. That is, dynamic systems in which energy is gradually dissipated (generally due to friction), and in which the moving body comes to rest in an equilibrium position.

In mechanics, the motion of the body and its trajectory in a three-dimensional valley might be considered, and the dissipated kinetic energy is simply lost from the system. The energy "loss" is dissipated as heat, but mechanics knows nothing about this. The potential energy change is independent of all this, as in Figure 1.1. Thermodynamics, on the other hand, knows nothing about kinetic energy and trajectories, but is concerned with energy changes between two different equilibrium states. In Figure 1.2 the two states are shown as valleys at different elevations, and the energy is potential energy; in thermodynamics the two states might be calcite and aragonite, and the energy is in a different form.

nearby valley. We will see that this is in fact the case. The analogy *is* useful. The problem lies in discovering just what kind of energy is being minimized. What is this *chemical* energy?

1.3.1 Chemical energy

We mentioned above that an early idea was that it is the *heat* energy content of systems that is minimized in chemical systems, that is, reactions will occur if heat is liberated. This is another way of saying that the heat content of the *products* is less than the heat content of the *reactants* of a reaction, so that the reaction liberates heat (Figure 1.3)

This view of things was common in the nineteenth century, and a great deal of effort was expended in measuring the flow of heat in chemical reactions.

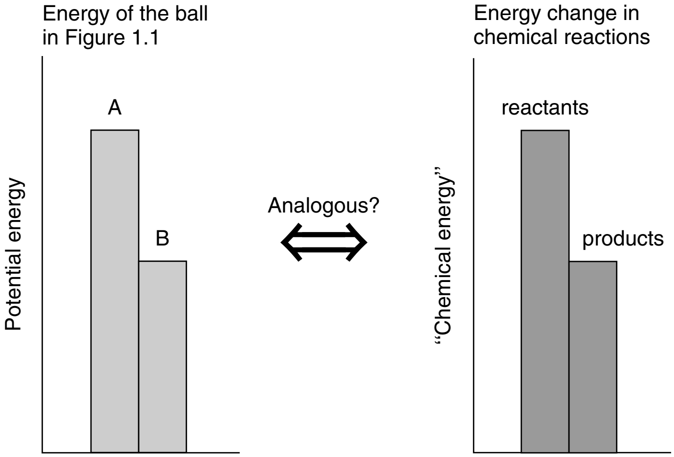


Figure 1.3 Mechanical processes always act so as to lower the potential energy content of the mechanical system. Perhaps, by analogy, chemical systems have some sort of "chemical energy" that is lowered during chemical reactions.

However, we don't even have to leave our kitchen to realize that this cannot be entirely correct. The melting of ice is obviously a reaction in which heat is *absorbed*, not liberated, which is why it is useful in cooling drinks. Therefore, despite the appealing simplicity of the "heat content" argument for explaining why chemical reactions occur, it cannot be the whole story. Nevertheless, the idea that some kind of "chemical energy" is liberated in reactions, or that "chemical energy" is minimized in systems at rest (equilibrium) is a powerful one. Perhaps heat is not the only energy involved. What other factors might there be? Not too many, we hope!

1.3.2 Plus something else?

Another very important clue we must pay attention to is the fact that some chemical reactions are able to take place with no energy change at all. For example, when gases mix together at low pressures, virtually no heat energy is liberated *or* absorbed. The situation is similar for a drop of ink spreading in a glass of water. These are spontaneous processes[1] characterized by a *mixing* process, rather than by a reorganization of atomic structures like graphite→diamond, or raw egg→fried egg. Our "chemical energy" term will have to take account of observations like these.

At this point, we might become discouraged, and conclude that our idea that some sort of chemical energy is being reduced in all reactions must be wrong – there seem to be too many exceptions. It certainly was a puzzle for a long time. But we have the benefit of hindsight, and because we now know that this concept of decreasing chemical energy of some kind is in fact the correct answer, we will continue to pursue this line of thought.

---

[1] We are using the terms *reaction* and *process* more or less synonymously here. Later on (§2.6) we will make a distinction.

1.4 Limitations of the thermodynamic model

This book outlines the essential elements of a first understanding of chemical thermodynamics, especially as applied to natural systems. However, it is useful at the start to have some idea of the scope of our objective – just how useful is this subject, and what are its limitations? It is at the same time very powerful and very limited. With the concepts described here, you can predict the equilibrium state for most chemical systems, and therefore the direction and amount of reaction that should occur, including the composition of all phases when reaction has stopped. The operative word here is "should." Our model consists of comparing equilibrium states, one with another, and determining which is more stable under the circumstances. We will not consider how fast the reaction will proceed, or how to tell if it will proceed at all. Many reactions that "should" occur do not occur, for various reasons. We will also say very little about what "actually" happens during these reactions – the specific interactions of ions and molecules that result in the new arrangements or structures that are more stable. In other words, our model will say virtually nothing about *why* one arrangement is more stable than another or has less "chemical energy," just that it does, and how to determine that it does.

These are serious limitations. Obviously, we will often need to know not only if a reaction *should* occur but *if* it occurs, and at what rate. A great deal of effort has also been directed toward understanding the structures of crystals and solutions, and of what happens during reactions, shedding much light on why things happen the way they do. However, these fields of study are not completely independent. The subject of this book is really a prerequisite for any more advanced understanding of chemical reactions, which is why every chemist, environmental scientist, biochemist, geochemist, soil scientist, and the like, must be familiar with it.

But in a sense, the limitations of our subject are also a source of its strength. The concepts and procedures described here are so firmly established partly because they are independent of our understanding of *why* they work. The laws of thermodynamics are distillations from our experience, not explanations, and that goes for all the deductions from these laws, such as are described in this book. As a scientist dealing with problems in the real world, you need to know the subject described here. You need to know other things as well, but this subject is so fundamental that virtually every scientist has it in some form in his tool kit.

1.5 Summary

The fundamental problem addressed here is why things (specifically, chemical reactions) happen the way they do. Why does ice melt and water freeze? Why does graphite turn into diamond, or vice versa? Taking a cue from the study of simple mechanical systems, such as a ball rolling in a valley, we propose that these reactions happen if some kind of energy is being reduced, much as the ball rolls in order to reduce its potential energy. However, we quickly find that this cannot be the whole story – some reactions occur with *no* decrease in energy. We also note that whatever kind of energy is being reduced (we call it "chemical energy"), it is not simply heat energy.

For a given ball and valley (Figure 1.1), we need to know only one parameter to determine the potential energy of the ball (its height above the base level, or bottom of the valley). In our "chemical energy" analogy, we know that there must be *at least* one other parameter, to take care of those reactions that have no energy change. Determining the parameters of our "chemical energy" analogy is at the heart of chemical thermodynamics.

2

Defining our terms

2.1 Something is missing

We mentioned in Chapter 1 that an early idea for understanding chemical reactions held that spontaneous reactions would always be accompanied by the loss of energy, because the reactants were at a higher energy level than the products, and they wanted to go “downhill.” This energy was usually thought to be in the form of heat, but this idea received a setback when it was found that some spontaneous reactions in fact absorb heat. Also, there are some reactions, such as the mixing of gases, where the energy change is virtually zero yet the processes proceed very strongly and are highly nonreversible. Obviously, something is missing. If the ball-in-valley analogy is right, that is, if reactions do proceed in the direction of decreasing chemical energy of some kind, something more than just heat is involved.

To learn more about chemical reactions, we have to become a bit more precise in our terminology and introduce some new concepts. In this chapter, we will define certain kinds of *systems*, because we need to be careful about what kinds of matter and energy transfers we are talking about; *equilibrium states*, the beginning and ending states for processes; *state variables*, the properties of systems that change during reactions; *processes*, the reactions themselves; and *phases*, the different types of matter within the systems. All these terms refer in fact to our models of natural systems, but they are also used to refer to things in real life. To be quite clear about thermodynamics, it is a good idea to keep the distinction in mind.

2.2 Systems

2.2.1 Real life systems

In real life, a *system* is any part of the universe that we wish to consider. If we are conducting an experiment in a beaker, then the contents of the beaker is our system. For an astronomer calculating the properties of the planet Pluto, the solar system might be the system. In considering geochemical, biological, or environmental problems here on Earth, the choice of system is usually fairly obvious, and depends on the kind of problem in which you are interested.

Figure 2.1 shows a seashore environment with three possible choices of natural system. At (a), we might be interested in the exchange of gases between the sea and the atmosphere (e.g., if the sea warms by one degree, how much CO_2 will be released to the atmosphere?). At (b), we might be interested in the dissolved material in the sea itself (e.g., the reactions between dissolved CO_2 and carbonate and bicarbonate ions). And at (c), we might be interested in reactions between the sediment and the water between the sediment particles (e.g., dissolution or precipitation of minerals in the sediment). The chosen systems are shown as boxes, but in most cases we are not concerned with the dimensions or shape of the box; we normally define the system in terms of the masses or moles of components in the system, as well as the nature of its contacts with whatever is outside the system (see §2.2.2).

These are examples of *inorganic* systems. Thermodynamics can also be applied to organic systems, including living organisms. A single bacterium could be our system, or a dish full of bacteria, or a single organelle within a bacterium. The choice depends on your particular interests and is obviously very wide. However, they are all similar in one respect. Because natural systems exist in the real world, whatever system you choose is bounded by (in contact with) other parts of the world and may exchange energy and matter (liquids, solids, gases) with these other parts of the world. Systems of this type are said to be "open." All living organisms are thus open systems because they take in nutrients, and get rid of waste products. All three systems in Figure 2.1 are obviously open, because water can flow in and out of (a) and (b), and even in (c), compaction of the sediments squeezes water out, and diffusion allows solutes to move in and out.

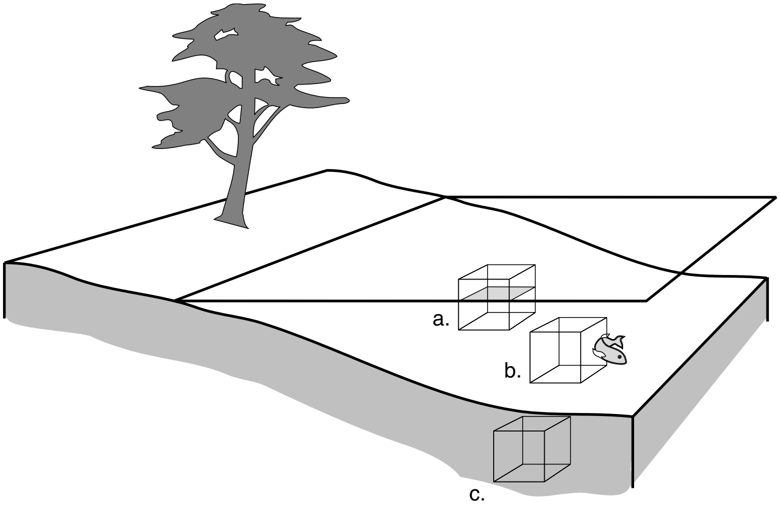


Figure 2.1 A seashore environment. The locations of three natural systems are shown.

Models

A model in the sense used here is an abstract object characterized or described by systems of equations, which attempt to represent the behavior of selected parts of the universe.

Thermodynamics deals with its subject matter (energy levels, energy changes) in an abstract way. The states and processes it describes are idealized; it does not describe or deal with any objects or processes in the real world, except to the extent that the variables in its equations are properties (e.g., volumes, energies) of real substances. Some processes in the real world are very similar to these idealized processes, and some are not. Where they are similar, thermodynamics is directly useful. Where they are not, we invent correction factors (e.g., "activity coefficients") to account for the differences.

The reason for mentioning this somewhat philosophical point is that many aspects of thermodynamics are abstract, or physically unrealistic. It helps to remember that we are using mathematics to simulate real systems.

Models are certainly used in other senses in the Earth sciences, such as the facies models of the sedimentologists.

2.2.2 Thermodynamic systems

Our goal is to understand the energy changes in natural systems. We will do this by mathematically simulating much simpler "models" of these systems, having variables that represent what we think are the essential elements of the natural systems. These models will not be material, but mathematical and conceptual. If we do it right, then the behavior of the model system will be very similar to (or will mimic) that of the real system. We will call this "understanding" the real system at the thermodynamic level.

Although most natural systems are open and are quite complex, our models of these systems can be much simpler and still be valuable. The kinds of thermodynamic or model systems that have been found to be useful in analyzing and understanding natural (real life) systems are as follows, and are illustrated in Figure 2.2. These thermodynamic systems are essentially defined by the types of walls they have. This is because we must be able to control (conceptually) the flow of matter and energy into and out of these systems.

- *Isolated systems* have walls or boundaries that are rigid (thus not permitting transfer of mechanical energy), perfectly insulating (thus preventing the flow of heat), and impermeable to matter. They therefore have a constant energy and mass content, since none can pass in or out. Perfectly insulating walls and the systems they enclose are called *adiabatic*. Isolated systems, of course, do not occur in nature, because there are no such impermeable and rigid boundaries. Nevertheless, this type of system has

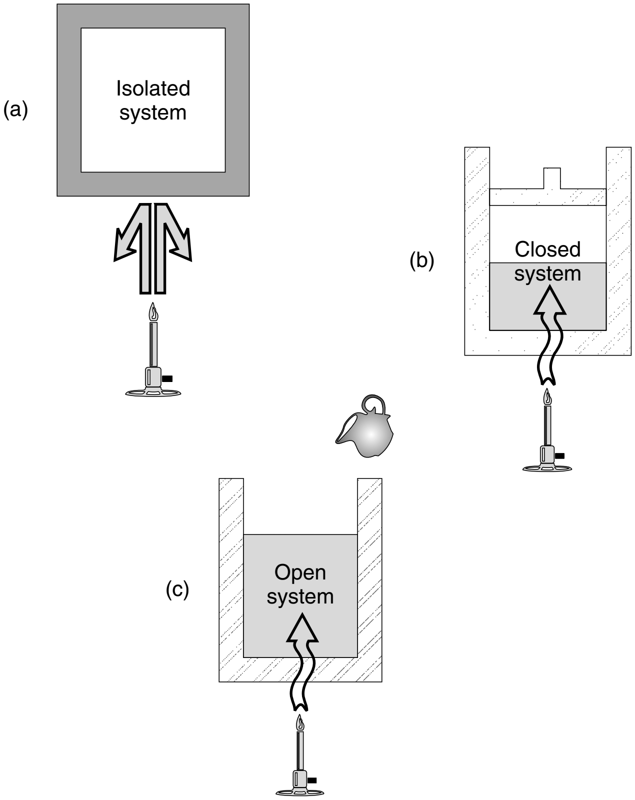


Figure 2.2 (a) Isolated system. Nothing can enter or leave the system (no energy, no matter). Whatever is inside the walls (which could be anything) will have a constant energy content and a constant composition. (b) Closed system. The closure is a piston to indicate that the pressure on the system is under our control. Energy can enter and leave the system, but matter cannot. The system here is shown as part liquid, part gas or vapor, but it could be anything. Both the liquid and the gas could also be considered as open systems, inside the closed system. Each may change composition, although the two together will have a constant composition. (c) Open system. Both matter and energy may enter and leave the system. The system may have a changing energy content and/or a changing composition. The pitcher shows one way of adding matter to the system.

great significance because reactions that occur (or could occur) in isolated systems are ones that *cannot* liberate or absorb heat or any other kind of energy. Therefore, if we can figure out what causes *these* reactions to go, we may have an important clue to the overall puzzle.

- *Closed systems* have walls that allow transfer of energy into or out of the system but are impervious to matter. They therefore have a fixed mass and composition but variable energy levels.
- *Open systems* have walls that allow transfer of both energy and matter to and from the system. The system may be open to only one chemical species or to several.

As mentioned above, most natural systems are open. However, it is possible and convenient to model them as closed systems; that is, to consider a fixed composition, and simply ignore any possible changes in total composition. If what happens because of changes in composition is important, it can often be handled by considering two or more closed systems of different compositions. Thus we will be dealing mostly with closed systems in our efforts to understand chemical reactions. Basically this means that we will be concerned mostly with individual chemical reactions, rather than with whole complex systems. In other words, even though a bacterium is an open system, it can be treated (modeled) as a closed system while considering many individual reactions within it. The reactants may need to be ingested and the products eliminated by the organism, but the reaction itself can be modeled independently of these processes. This greatly simplifies the task of understanding the biochemical reactions. The same is true of most geochemical and environmental systems.

The most common kind of open system in chemical thermodynamics is represented in Figure 2.2b, that is, two open subsystems within an overall closed system. There can be any number of these "open subsystems," and finding out how many there are and what their compositions are, given some physical conditions, is a common problem in the application of thermodynamics. We have a brief look at other kinds of open systems in Chapter 4.

It is one of the paradoxes of thermodynamics that isolated systems, that have no counterpart in the real world, are possibly the most important of all in terms of our understanding of chemical reactions. You will have to wait until Chapter 4 to see why.

2.3 Equilibrium

In studying chemical reactions, we obviously need to know when they start and when they have ended. To do this, we define the state of *equilibrium*, when no reactions at all are proceeding. Here we encounter a distinct difference between real and thermodynamic systems, because the state of equilibrium is defined differently in the two cases.

In thermodynamic systems, that is, in our models, equilibrium is defined in terms of *chemical potentials*, which we will get to in a later chapter. This state, as you might imagine, is one of *perfect* equilibrium, perfect rest, with absolutely no gradients or inhomogeneities of any kind. Real systems often approach this state more or less closely, but probably never attain it. When real systems do approach equilibrium, thermodynamics can be applied to them. Obviously, we need to have some way of telling whether real systems are "at equilibrium," or have closely approached equilibrium.

Equilibrium states in real systems have two attributes:

1. A real system at equilibrium has none of its properties changing with time, no matter how long it is observed.
2. A real system at equilibrium will return to that state after being disturbed, that is, after having one or more of its parameters slightly changed, then changed back to the original values.

This definition is framed so as to be "operational," that is, you can apply these criteria to real systems to determine whether they are at equilibrium. And in fact, many real systems do satisfy the definition. For example, a crystal of diamond sitting on a museum shelf obviously has exactly the same properties this year as last year (part 1 of the definition), and if we warm it slightly and then put it back on the shelf, it will gradually resume exactly the same temperature, dimensions, and so on that it had before we warmed it (part 2 of the definition). The same remarks hold for a crystal of graphite on the same shelf, so that the definition can apparently be satisfied for various forms of carbon. Many other natural systems just as obviously are not at equilibrium. Any system having temperature, pressure, or compositional gradients will tend to change so as to eliminate these gradients, and is not at equilibrium until that happens. A cup of hot coffee, for example, is not at equilibrium with the air around it until it cools down.

So if diamond and graphite are both at equilibrium, do we have two kinds of equilibrium? In our ball-in-valley analogy, the ball in any valley would fit our definition. What distinction do we make between the lowest valley and the others?

2.3.1 Stable and metastable equilibrium

In this section we use the simple mechanical analogy in §1.3 to distinguish between *stable* and *metastable* equilibrium. This explanation is satisfactory for an intuitive understanding, but we return to this subject for a better theoretical understanding in §4.9.1.

Stable and metastable are the terms used to describe the system in its lowest equilibrium energy state and any other equilibrium energy state, respectively. In Figure 2.3, we see a ball on a surface having two valleys, one higher than the other. At (a), the ball is in an equilibrium position, that fulfills both parts of our definition – it will stay there forever, and will return there if disturbed, as long as the disturbance is not too great. However, it has not achieved the lowest possible potential energy state, and therefore (a) is a *metastable equilibrium* position. If the ball is pushed past position (b), it will roll down to the lowest available energy state at (d), a *stable equilibrium* state. During the fall, for example, at position (c), the ball (system) is said to be *unstable*. In position (b), it is possible to imagine the ball balanced and unmoving, so that the first part

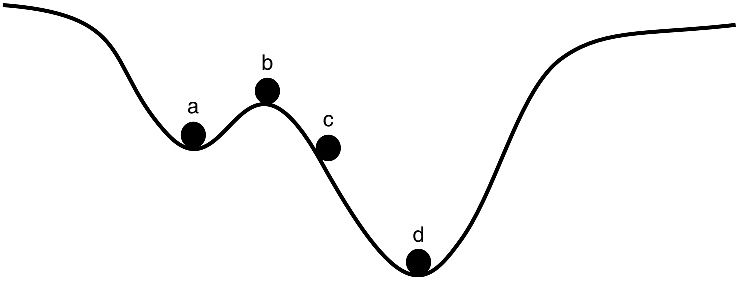


Figure 2.3 Four positions of a ball on a surface, to illustrate the concept of equilibrium. Position a – metastable equilibrium. Position b – unstable. Position c – unstable. Position d – stable equilibrium.

of the definition would be fulfilled, and this is sometimes referred to as a third type of equilibrium, admittedly a trivial case, called *unstable equilibrium*. However, it does not survive the second part of the definition, so we are left with only two types of equilibrium, stable and metastable.

Of course, we find that the stable form of substances is different under different conditions. For example, the stable form of H_2O is water at +5 °C, and ice at −5 °C (Figure 2.4). The freezing and melting of H_2O is normally fairly rapid, so we don't often see metastable ice above its melting temperature, or metastable water below its freezing temperature. But many such phase changes are not so rapid, in fact they may not happen at all, even though energy would be released if they did. These reactions, which get "stuck" in a high energy state are usually not melting/freezing reactions, but solid state reactions – that is, a reaction in which a mineral having one crystallographic structure changes to a mineral having the same composition but with a different structure.

A good example of this is the diamond/graphite reaction. We know now that the stable form of pure carbon at Earth surface conditions is the mineral graphite, but that at high temperatures and pressures, such as found deep in the Earth's mantle, graphite will spontaneously react to form diamond. However,

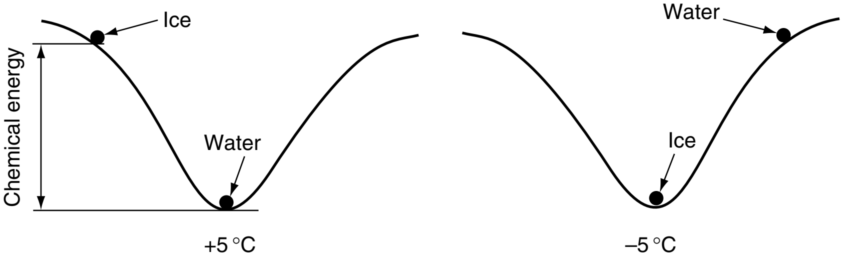


Figure 2.4 The mechanical analogy for H_2O at −5 °C and +5 °C and atmospheric pressure. At −5 °C, water is unstable and releases energy until it becomes ice at −5 °C. At +5 °C, ice is unstable and releases energy until it becomes water at +5 °C. The problem is, what kind of energy is being minimized?

when tectonic and igneous processes bring the diamond back to the surface, the diamond does not (fortunately) change back to graphite, so we say that diamond is a metastable form of carbon at Earth surface conditions (Figure 2.5). When we develop this subject further, we should be able to predict or calculate under what conditions it is the stable form of carbon.

2.3.2 Partial and local equilibrium

There are two other commonly used terms in connection with equilibrium states.

Partial equilibrium

"Partial equilibrium" is intended to indicate that part or parts of the system have reached equilibrium, but those parts have not reached equilibrium with each other. The usual example of partial equilibrium is that of a crystal in an aqueous solution. The crystal is actively dissolving, so the system as a whole is not in equilibrium, but the aqueous solutes re-equilibrate very quickly, so that the solution itself is very close to internal equilibrium. The system is then said to be in partial equilibrium.

That may be true for the real system, but there is no such thing as partial equilibrium in thermodynamics, or the systems that thermodynamics deals with. In thermodynamics equilibrium is defined as equality of potentials in every phase. Partial equilibrium in real systems is modeled in thermodynamics as a metastable equilibrium. This distinction may not be clear at this point, but will be further discussed in connection with metastable equilibrium (Chapter 4) and titration (Chapter 18).

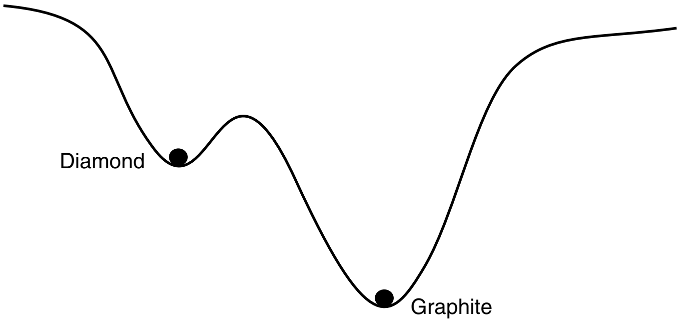


Figure 2.5 The mechanical analogy for carbon at Earth surface conditions. Graphite is the stable form of carbon because it has the lowest energy content of any form of carbon (under Earth surface conditions). Diamond has a higher energy content but is prevented from changing to graphite by an energy barrier.

Local equilibrium

Real world systems are in constant flux, and never really achieve thermodynamic equilibrium, but we want to apply thermodynamics to them anyway, so we have to choose parts of real systems which are reasonably close to thermodynamic equilibrium.

For example, you cannot apply thermodynamics to the ocean as a whole. Calcite is supersaturated at the surface, but undersaturated at 5 km depth (Chapter 16). Thermodynamics cannot be applied to a system which is both supersaturated and undersaturated. You can apply thermodynamics to volumes close to equilibrium at the surface or at depth, not both together, so we say we apply thermodynamics to areas of "local equilibrium." It is obviously important to apply thermodynamics appropriately, and generally we do this, but the point is that local equilibrium is not part of thermodynamics, it is a concept we need, a property that real systems must have, in order to apply thermodynamics.

Understanding thermodynamics does not depend in any way on local equilibrium, but applying it to natural systems does. The question then naturally arises as to how one distinguishes between places having local equilibrium from places that do not. This question does not have a good answer. Places having large gradients in temperature, pressure or composition can be ruled out, but how large is "large"? Quite often the practice is to apply thermodynamics and see how it works out. If it seems to work well, then local equilibrium is assumed. Obviously some better approach would be desirable. There have been several attempts at providing a quantitative criterion for local equilibrium. The most accessible for Earth scientists appears to be that of Knapp (1989), which is summarized in Zhu and Anderson (2002, Chapter 3), who also cite a number of other references on the subject.

Defining local equilibrium

The question of fluid – solid phase equilibrium arises in many subject areas, including environmental problems, studies of diagenesis, long range flow in sedimentary basins, ore genesis, magmatic – hydrothermal systems, regional metamorphism, and laboratory experimental systems. In each of these real systems, local equilibrium in theory requires that any disequilibrium condition relax instantaneously to an equilibrium state. In reality, this relaxation occurs over a finite time and, for a fluid-flow system, a finite distance. Knapp (1989) points out that each of these types of systems has a characteristic scale of interest, which is hundreds of meters or kilometers in studies of sedimentary basins, but perhaps microns in studies of surface processes. If the problem is defined on the kilometer scale, then disequilibrium over distances of centimeters is insignificant. The problem then is to determine, for a given system, the time required for a system in disequilibrium to reach equilibrium, and the distance the fluid has moved in that time period.

Knapp considers the problem in terms of a one-dimensional flow path in a quartz sandstone. The moving water is initially at equilibrium with quartz, then a pulse of pure water is introduced, and the time and distance required for the reattainment of equilibrium are calculated. Quite a few factors are involved, including concentrations (including pH), temperature, fluid velocity, diffusion and dispersion coefficients, and of course kinetics, including the surface area (m^2 of mineral per m^3 of fluid). The results, presented in terms of Damköhler and Peclet numbers,[1] show that there is a region where the time and distance to equilibrium is reaction dominated, and there is another region where they are transport or advection dominated. Local equilibrium can occur in both domains. Most natural environments with elevated temperatures fall in the reaction dominated domain, where the effects of dispersion and diffusion can safely be ignored, but local equilibrium would appear to be a questionable approximation in what Knapp (1989) refers to as “human controlled environments” due to characteristically large fluid velocities and low temperatures.

This analysis by Knapp is useful in defining and clarifying the local equilibrium problem in a quantitative way. Unfortunately, despite the rather drastic simplification, most of the parameters required to define the problem in real situations at the present time are poorly known. The quantitative results are then of questionable significance in any practical sense, but they are worth reflecting on. All applications of thermodynamics assume local equilibrium, but defining just what that is has proven difficult.

2.4 State variables

Systems at equilibrium have measurable properties. A property of a system is any quantity that has a fixed and invariable value in a system at equilibrium, such as temperature, density, or refractive index. Every system has dozens of properties. If the system changes from one equilibrium state to another, the properties therefore have changes that depend only on the two states chosen, and not on the manner in which the system changed from one to the other. This dependence of properties on equilibrium states and not on processes is reflected in the alternative name for them, *state variables*. Several important state variables (which we consider in later chapters) are not measurable in the absolute sense in a particular equilibrium state, though they do have fixed, finite values in these states. However, their changes between equilibrium states are measurable.

---

[1] The Damköhler number (Da) expresses the rate of reaction relative to the advection or fluid flow rate. A large Da value means that reaction is fast relative to transport and that aqueous concentrations may change rapidly in time and space. The Peclet number (Pe) expresses the importance of advection relative to dispersion in transporting aqueous compounds. A large Pe value means that advection dominates, which may result in large concentration gradients; a small Pe value suggests that dispersion dominates, which promotes mixing in the fluid phase.

Reference in the above definition to “equilibrium states” rather than “stable equilibrium states” is deliberate, since as long as metastable equilibrium states are truly unchanging they will have fixed values of the state variables. Thus both diamond and graphite have fixed properties. Metastable states are extremely common. For example, virtually all organic compounds are metastable in an oxidizing environment, such as the Earth’s atmosphere. We should be grateful for those “activation energy barriers” that prevent metastable states from spontaneously changing to stable states; otherwise we would not be here to discuss the matter.

2.4.1 Total versus molar properties

Many physical properties, such as the volume and various energy terms, come in two forms – the total quantity in the system and the quantity per mole or per gram of substance considered. We use different fonts for these total and molar properties. For example, water has a volume per mole (V) of about 18.0686 $\text{cm}^3\ \text{mol}^{-1}$, so if we have 30 moles of water in a beaker, its volume ($\mathbf{V}$) is 542.06 cm^3. This relationship for a pure substance such as H_2O is $Z = \mathbf{Z}/n_i$, where $\mathbf{Z}$ is any total property, Z is the corresponding molar property, and n_i is the number of moles of the substance. In our water example, above, $542.06/30 = 18.068$. In more complex systems where more than one substance is present, total and molar properties are related in the same way. A beaker containing, for example, a kilogram of water (55.51 moles H_2O) and 1 mole of NaCl occupies 1019.9 cm^3. The molar volume of the system is then $Z = \mathbf{Z}/\sum_i n_i$, or $1019.9/(1 + 55.51) = 18.05$ $\text{cm}^3\ \text{mol}^{-1}$.

These two types of state variables have been given names:

- *Extensive* variables are proportional to the quantity of matter being considered – for example, total volume ($\mathbf{V}$).
- *Intensive* variables are independent of the total size of the system and include concentration, viscosity, and density, as well as all the *molar* properties, such as the molar volume, V.

Scientific versus engineering units

In science, *molar* properties, such as molar volumes, molar energies, are most commonly used. In engineering on the other hand, *specific* properties are more common. Specific properties are mass-related rather than mole-related. Thus the specific volume of water at 25 °C is $1.0029\,\text{cm}^3\ \text{g}^{-1}$. Molar and specific properties are of course related by the molar mass (or so-called gram formulas weight, gfw) of the substance. That for water is 18.0153, so $1.0029\,\text{cm}^3\ \text{g}^{-1} \times 18.0153\,\text{g}\,\text{mol}^{-1} = 18.068\,\text{cm}^3\ \text{mol}^{-1}$.

Of course, many equations look much the same with total and molar properties because ratios are involved. That is, if $(\partial \mathbf{U}/\partial \mathbf{S})_{\mathbf{V}} = T$, then it is also true that $(\partial U/\partial S)_V = T$; or if $(\partial \mathbf{G}/\partial P)_T = \mathbf{V}$, then $(\partial G/\partial P)_T = V$, so that the distinction may seem to be unimportant. However, sometimes it *is* important, as we will see. In general terms, we use the total form of our variables (bold type) in some theoretical discussions, and the molar form (italic type) in most calculations.

Partial molar properties

In addition to total and molar properties, we have *partial* molar properties, which are a little trickier to understand. It's relatively easy to see that the volume (extensive variable) of a system depends on how much stuff you have in the system, but that its temperature or density (intensive variables) do not. This is true no matter how many different phases there are in the system, as long as you are considering the *whole* system, not just parts of it.

A problem arises, though, when you consider the properties of solutions, which can have variable concentrations of solutes. The volume per gram of halite, NaCl, is the same whether you consider 10 or 20 grams of it. But what is the volume per gram of 10 grams of NaCl dissolved in a liter of water? This property depends on the concentration of NaCl – the volume per gram or per mole of 20 dissolved grams is different from that of 10 dissolved grams. And what *is* the volume of something dissolved in something else? How is it defined, or measured? These are important questions, and will be discussed in Chapter 10.

The properties of dissolved substances is discussed in terms of *partial molar properties*, the formal definition of which is

$$\overline{Z}_i = \left(\frac{\partial \mathbf{Z}}{\partial n_i}\right)_{T,P,\hat{n}_i} \tag{2.1}$$

where $\mathbf{Z}$ is the total or extensive form of any thermodynamic parameter, $\overline{Z}$ the partial molar form, n_i is the number of moles of component i, and $\hat{n}_i$ is the number of moles of all components other than i in the same solution. Note particularly that the partial derivative is taken of the total quantity $\mathbf{Z}$, not the molar Z, and that the main constraints are T and P. However, the important thing to know about partial molar properties is not this differential equation, but that they are the properties per mole of substances at a particular concentration in a particular solution, as explained in Chapter 10. You think about partial molar properties in exactly the same way you think about molar properties. The only difference is that for a given substance, they are not fixed quantities at a given T and P, but vary with the concentration of the substance and the nature of the solution.

The differences between total, molar, and partial molar properties is also discussed in more mathematical terms in Appendix C.

2.5 Phases and components

We must also have terms for the various types of matter to be found within our thermodynamic systems. A *phase* is defined as a homogeneous body of matter, having distinct boundaries with adjacent phases, and so is mechanically separable from the other phases. The shape, orientation, and position of the phase with respect to other phases are irrelevant, so that a single phase may occur in many places in a system. Thus the quartz in a granite is a single phase, regardless of how many grains of quartz there are. A salt solution is a single phase, as is a mixture of gases. There are only three very common types of phases – solid, liquid, and gas or vapor. A system having only a single phase is said to be *homogeneous*, and multiphase systems are *heterogeneous*.

The term generally used to describe the chemical composition of a system is *component*. The components of a system are defined by the smallest set of chemical formulas required to describe the composition of all the phases in the system. This simple definition is sometimes surprisingly difficult to use. To take a simple example, consider a solution of salt (NaCl) in water (H_2O), in equilibrium with water vapor. This might look like Figure 2.2b. There are two phases, liquid and vapor, and two components, NaCl and H_2O. A chemical analysis could report the amounts or concentrations of Na, Cl, H, and O in the system, but only two chemical formulas are needed to describe the compositions of both phases.

Unfortunately, this does not nearly encompass all we need to say about components. We will have more to say in Chapter 11, but we should at least point out that the definition of components given above ("smallest set of chemical formulas…") is used for phases in our models, not in real systems. For example, analysis of any calcite crystal will reveal the presence of many elements besides those in the formulas $CaCO_3$. Nevertheless, component $CaCO_3$ is very often used to represent calcite, whatever its actual composition.

2.5.1 Real versus model systems

Equilibrium, phases, and components are terms that appear to apply to real systems, not just to the model systems that we said thermodynamics applies to, and in general conversation, they do. But real phases, especially solids, are never perfectly homogeneous. And real systems don't really have components, only our models of them do. Seawater, for example, has an incredibly complex composition, containing dozens of elements. But our thermodynamic models might model seawater as having two, three, or more components, depending on the application. As for equilibrium, real systems do often achieve equilibrium as we have defined it, but it is never a perfect equilibrium.

However, the fact that real phases are more or less homogeneous, and that real systems achieve an approximate equilibrium, is what makes thermodynamics useful. The model is perfect, but real life comes close enough in many respects so that the model is useful. In fact, the close similarity between reality and our models of reality, and the fact that we use the same terms to describe each, may lead to a certain degree of confusion as to what we are talking about. Usually no harm is done, and the distinction gets easier with practice.

2.6 Processes

Finally, we get to something that looks more interesting. *Processes* are what we are usually interested in – changes in the real world. In geology, these might be igneous, diagenetic, or metamorphic processes. In biology, they might be cellular processes. In the environmental world, they might be potentially harmful processes near waste disposal sites – the possibilities are endless. However, most of the processes of interest to us have one thing in common – they are extremely complicated. The only hope we have of understanding them is to break complex processes down into their simpler component parts, and to construct simplified models of them. We have already begun to do this by defining several types of simple *systems* that we can use; we will now define a *process* in a way that will help us model real processes.

A thermodynamic *process* is what happens when a system changes from one equilibrium state to another. Thus any two equilibrium states of the system may be connected by an infinite number of different processes because only the initial and final states are fixed; anything at all could happen during the act of changing between them. A chemical reaction is one kind of process, but there are others. For example, simply warming or cooling a system is a process according to our definition.

In spite of there seeming to be an endless number of kinds of processes in the world, we find that in thermodynamic models there are only two – reversible and irreversible.[2]

- The most important *irreversible processes* are those that begin in a metastable equilibrium state and lead to a more stable state, such as aragonite recrystallizing to calcite. Another kind would be a stable equilibrium state changing to a lower energy stable equilibrium state, such as when the weight on a piston is replaced by a smaller weight.
- Processes that begin in a stable equilibrium state and proceed to another stable equilibrium state, without ever leaving the state of equilibrium more than infinitesimally, are *reversible processes*.

---

[2] In some treatments of thermodynamics there is a third type – the virtual process. See Reiss (1965) for its use.

2.6.1 Irreversible processes

We have defined a metastable state of a system as a state that has more than the minimum energy for the given conditions, but is for some reason prevented from releasing that energy and reacting or changing to the stable state of minimum energy. An irreversible process is one that occurs when whatever constraint is holding the system in its high energy state is removed, and the system slides down the energy gradient to a lower energy state. We consider constraints in more detail in Chapter 4.

The only example we have given thus far of a metastable system is the mineral diamond, that could lower its energy content by changing into graphite but does not, because energy is required to break the carbon–carbon bonds in diamond (which are very strong) before the atoms can rearrange themselves into the graphite structure. There are many other similar examples of metastable minerals. We have also mentioned that most organic compounds, such as all the ones in living organisms, are metastable. When the life processes maintaining their existence cease, they quickly react (decompose) to form more stable compounds.

In most of the chemical reactions we will be considering, a combination of minerals, or minerals plus liquids or gases, reacts to form some different minerals under some given conditions. For example, the mineral corundum (Al_2O_3) is stable, considered by itself (i.e., there is no other form of Al_2O_3 that is more stable), but in the presence of water it reacts to form gibbsite ($Al_2O_3 \cdot 3H_2O$). The reaction is

$$\mathrm{Al_2O_3}(s) + 3\,\mathrm{H_2O}(l) = \mathrm{Al_2O_3} \cdot 3\mathrm{H_2O}(s) \tag{2.2}$$

and the energy relationships are shown in Figure 2.6. We will use (s), (l), (g), and (aq) after our formulas to indicate whether they are in the solid, liquid, gas, or aqueous (dissolved in water) state.

Do not confuse the metastability of diamond at Earth surface conditions with the metastability of corundum or water. Diamond is metastable because the same carbon atoms would have a lower energy in the crystal structure of graphite. But corundum by itself is not metastable, and neither is water, at 25 °C and atmospheric pressure. It is the *combination* of corundum and water that can be regarded as metastable, because their *combined* atoms would have a lower energy level in the form of gibbsite.

Another example is the dissolution of sugar in coffee (Figure 2.7), for which we cannot write a simple balanced reaction. Nevertheless, the assemblage of sugar lumps and a cup of coffee is a metastable assemblage in our usage. They are prevented from reacting (sugar dissolving) by the fact that they are separated, which constitutes a *constraint* on the system. When the constraint

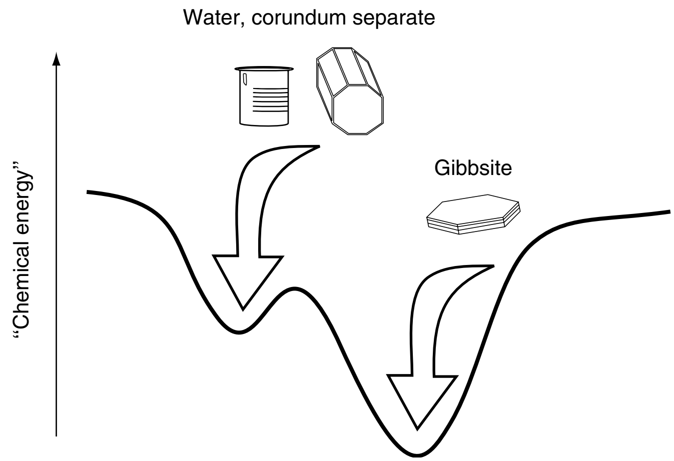


Figure 2.6 Water plus corundum can lower its energy content by reacting to form gibbsite.

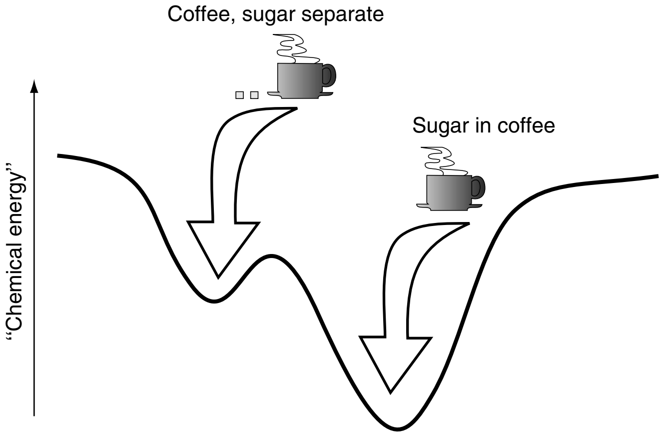


Figure 2.7 Sugar dissolves in coffee because the "chemical energy" of the dissolved state is less than that of the two coexisting separately.

is removed by putting the sugar in the coffee, the reaction occurs, because the "chemical energy" is lowered.[3]

The essence of most irreversible reactions is that *energy is released* during the change (exactly what kind of energy we have not yet discussed). Therefore, unless energy is *added* to the system, the reaction cannot go in the reverse direction under the given conditions. In other words, the reaction or change is *spontaneous* in one direction only. The ball will never roll uphill of its

---

[3] The corundum plus water example and the sugar plus coffee example are different in an instructive way. If you actually put a crystal of corundum in a beaker of water, nothing at all happens, except that the corundum gets wet, whereas when the sugar is put in the coffee, it dissolves immediately. Both assemblages are metastable but the constraints are different. Corundum is prevented from reacting with water by an *activation energy barrier*, meaning that the atoms in Al_2O_3 are too tightly bound to react, even though the system could lower its energy if they did. The sugar is prevented from dissolving in the coffee by a physical separation.

own accord. This does not mean that the reaction can never go in the opposite way. It may very well go in the opposite way *under different circumstances*. Thus the corundum plus water reacts spontaneously to form gibbsite at low temperatures, but at high temperatures gibbsite spontaneously decomposes to form corundum and water. Similarly, we said that ice$\rightarrow$water at 5 °C, but water$\rightarrow$ice at −5 °C. Spontaneous or irreversible (these terms are synonymous) refers to a single set of conditions, such as a given temperature, pressure, and composition. If the conditions are changed, the reaction may become spontaneous in the other direction.

Reactions involving organic compounds

Reactions involving organic compounds, whether in living organisms or not, are no different in principle from any other kind of reaction, such as those between minerals. The only difference is that for organic compounds, the reaction usually proceeds from one metastable state to another metastable state of lower energy, rather than from a metastable state to a stable state. Consider for example the reaction

$$\mathrm{C_8H_{16}N_2O_3}(aq) + \mathrm{H_2O}(l) = \mathrm{C_6H_{13}NO_2}(aq) + \mathrm{C_2H_5NO_2}(aq) \tag{2.3}$$

which represents the breaking of a peptide bond between two amino acids, one of the more fundamental processes in biochemistry. The (aq) here means that the compounds we are discussing are dissolved in water and, hence, the reaction takes place in water. If we use names rather than chemical formulas, this is

$$\text{leucylglycine} + \text{water} = \text{leucine} + \text{glycine} \tag{2.4}$$

This reaction occurs spontaneously, and the energy relations can be depicted exactly as for simpler compounds. The only difference is that rather than reacting to compounds in the lowest possible energy state, leucylglycine plus water reacts to form compounds in another metastable state (leucine plus glycine) of lower energy than the initial state, as shown in Figure 2.8. Virtually all organic compounds are metastable with respect to simple inorganic compounds and elements such as water, nitrogen, hydrogen, and graphite. Thus the reaction

$$\mathrm{C_6H_{13}NO_2}(aq) + \mathrm{C_2H_5NO_2}(aq) = 2\,\mathrm{H_2}(g) + 2\,\mathrm{NH_3}(g) + 4\,\mathrm{H_2O}(l) + 8\,\mathrm{C_{graphite}} \tag{2.5}$$

is also spontaneous, as shown in Figure 2.8.

Living organisms have developed mechanisms (involving enzymes) for overcoming the energy barriers separating products and reactants of reactions required for the life processes of the organisms. Obviously no enzymes have been developed to enable the breakdown of the organisms to the simple inorganic compounds of which they are composed, as this would be fatal.

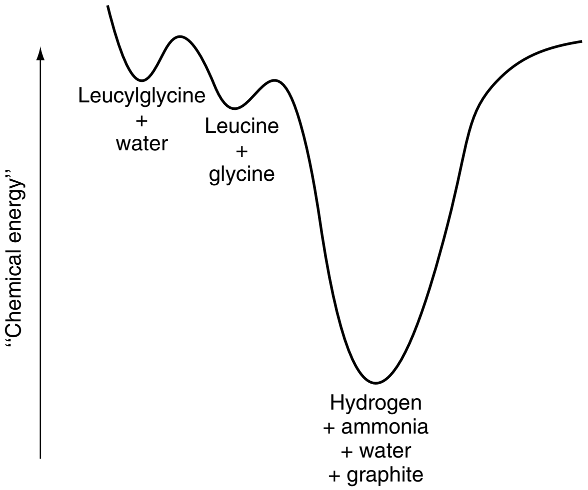


Figure 2.8
Energy relationships between organic compounds. Most organic compounds have much higher energy contents than do combinations of simple inorganic compounds of the same overall composition.

2.6.2 Reversible processes

In the real world, a process is reversible if it can go either way, like warming a crystal and later cooling it back down, or expanding a gas and later compressing it to its original state. But these processes can only be considered thermodynamically by integrating the relevant equations, and integration requires continuous functions. So "reversible" is another term which has different meanings in real and thermodynamic systems.

A reversible process in thermodynamics is one in which a system in a state of equilibrium changes to another state of equilibrium *without ever becoming out of equilibrium*. This type of process is not possible in the real world. For example, a crystal of diamond at 25 °C is warmed to 50 °C. What is so difficult about that? Although it is not difficult to warm a diamond to 50 °C, it is impossible to do it without leaving the state of equilibrium. To change the temperature of the crystal, heat must be applied to it. This sets up a temperature gradient between the inside and the outside of the crystal, and heat travels into the crystal, raising its temperature. But while a temperature gradient exists in the crystal, it is not at equilibrium (a system at equilibrium can have no gradients in temperature, pressure, or composition). In a real heating process, the crystal of diamond is at equilibrium at 25 °C, then it leaves the state of equilibrium for a time, and then it attains equilibrium later under its new conditions, 50 °C. However, in a reversible heating process, the crystal is at all times at equilibrium with its environment, or at least never more than infinitesimally different in temperature from its environment, and changes from 25 °C to 50 °C in a continuous state of equilibrium.[4]

---

[4] We consider a different reversible process in more detail in §3.4.1.

The reversible process as defined is impossible in the real world. However, it is quite simple in the thermodynamic model, because the temperature, volume, and all other properties of the diamond are just points on mathematical surfaces in the model, and there is nothing to prevent the point representing the temperature to move around on a surface representing the equilibrium values of various properties of the diamond.

Why in the world would we be interested in such a strange kind of impossible process? It's simple, really. The reason the reversible process (defined as a continuous succession of equilibrium states) is important in the thermodynamic model is that it is the only kind of process that our mathematical tools of differentiation and integration can be applied to – they only work on *continuous* functions. Once our crystal of diamond leaves its state of equilibrium at 25 °C, practically anything could happen to it, but as long as it settles back to equilibrium at 50 °C, all of its state variables have changed by fixed amounts from their values at 25 °C. We have equations to calculate these energy differences, but they refer to lines and surfaces in our model, and that means that they must refer to continuous equilibrium between the two states.

In other words, to *calculate* the energy difference between the two states, we must use a fictitious path (the reversible process) between the two states. The result is the real energy difference, no matter what actually happened to the system between the two states. The reversible process is another example of the difference between the real world and our models of the real world. Reversible processes are quite simple to carry out in our models, because the models are mathematical, not real.

2.6.3 Egg reactions

We have not discussed all the examples we used in Chapter 1. To conclude our discussion of various common chemical reactions (§1.2.1), we should discuss the thermodynamics of frying eggs. At a simple level, we could say that the egg in the refrigerator represents a metastable state, and that frying it promotes a reaction to a more stable state, analogous to the leucylglycine + water→leucine + glycine reaction in Figure 2.8. Even if this was the case, putting the fried egg back in the refrigerator would not suffice to reverse the reaction; going from a stable state to a metastable one requires a source of energy – it won't occur spontaneously. In the water/ice case, the water returns to ice in the fridge because ice is the stable form there.

Strictly speaking, however, we know that eggs in the refrigerator won't last indefinitely; they will eventually "go bad." This means that they are not in a truly metastable state in the refrigerator, but an unstable, slowly changing state. This means that because the raw egg occupies no "valley" for the egg components to roll into, it is very unlikely that we could restore the raw egg state, even if we had an energy source.

In studying natural systems, such as eggs, it is often quite difficult to distinguish stable, metastable, and unstable states from each other without a considerable amount of work and ingenuity, but it can be done. When you get numbers from tables, as we will be doing, all this work has been done for you, although you have to realize that because of the difficulties involved, some of the data may not be accurate and may be revised at some future date. A compound believed to be stable under given conditions may later be found to be metastable after more careful work is done.

Reactions in these complex systems are actually made up of a number of simpler reactions, and applying thermodynamics requires that the individual reactions be treated separately. The individual biochemical reactions in many organic systems still have not been figured out. Nevertheless, we are confident that any particular reaction, once defined, will follow the logic and the systematics described in this book.

2.6.4 Notation

Reaction deltas

We have now set up the general framework within which thermodynamics is able to deal with processes. Any given process or chemical reaction within a chosen system will proceed from an initial equilibrium state (normally a metastable equilibrium state) to another equilibrium state more stable than the first one. During this process or reaction the system is out of equilibrium. The system has a number of properties or state variables, such as volume and energy content, that have fixed values in equilibrium states and that therefore have fixed amounts of change between equilibrium states. These changes are always written using a delta notation, where the delta refers to the property in the final state minus the property in the initial state. For example, if the system undergoes a process during which its (molar) volume (V) changes from V_{initial} to V_{final}, we write

$$\Delta V = V_{\text{final}} - V_{\text{initial}} \tag{2.6}$$

If the process is a chemical reaction, a number of compounds may be involved. A generalized chemical reaction could be written as

$$a\,\mathrm{A} + b\,\mathrm{B} + \cdots = m\,\mathrm{M} + n\,\mathrm{N} + \cdots$$

An example is Equation (2.2), where A is Al_2O_3, B is H_2O, and M is $Al_2O_3 \cdot 3H_2O$ (there is no N); a and m are 1 and b is 3. The quantities A, B, M, and N are chemical formulas representing any compounds or elements we happen to be interested in, and each can be solid, liquid, gas, or a solute. One side of the reaction will usually be more stable than the other, and a reaction will tend to occur, unless there is an energy barrier preventing the reaction, or unless the compounds are all at equilibrium together. In this case, the volume change during the reaction is $\Delta_r V$ (we insert a subscript "*r*" to indicate a chemical reaction) and is equal to the sum of the volumes of the reaction products (the final state) minus the sum of the volumes of the reactants (the initial state). Thus

$$\Delta_r V = m\,V_{\mathrm{M}} + n\,V_{\mathrm{N}} + \cdots - a\,V_{\mathrm{A}} - b\,V_{\mathrm{B}} - \cdots$$

where V_{M} is the volume of a mole of compound M, and so on. For example, the change in volume for reaction (2.2) is

$$\Delta_r V = V_{\mathrm{Al_2O_3\cdot 3H_2O}} - V_{\mathrm{Al_2O_3}} - 3\,V_{\mathrm{H_2O}} \qquad (2.7)$$

Note that each volume must be multiplied by its corresponding stoichiometric coefficient in the reaction. Molar volumes are readily available for most pure substances.

Following this convention, the change in energy of the ball rolling down the hill in Figure 1.1 would be a negative quantity, as shown in Figure 1.3 (energy in state B minus energy in state A is negative). It follows, then, that the change in the "chemical energy" term we are looking for will always be a negative quantity in spontaneous reactions, as also shown in Figure 1.3 (energy of products minus energy of reactants).

Chemical equations

For the most part, when we write reactions such as (2.2) and (2.3), we use the = sign to indicate only that the reaction is "balanced," meaning that the same number and kinds of atoms appear on both sides, and that any electrical charges are also the same on both sides. If we want to emphasize that the reaction proceeds strongly or irreversibly we may use an arrow, as in $\mathrm{A} \rightarrow \mathrm{B}$, and if we want to emphasize that the two sides are in equilibrium, we might use $\mathrm{A} \rightleftharpoons \mathrm{B}$. However, the = sign includes these possibilities, and all others.

2.7 Summary

If you look around the physical world today, you realize that there is an incredible number of chemical and physical *processes* going on all around you, and as you look into these in more and more detail, as science has done, you find more and more complexity at all levels, right down to the atomic and subatomic levels. How can we systematize and understand these processes in such a way as to be able to control some of them for our own purposes?

Thermodynamics is the net result of our attempts to do this. It is not a description of any real process but a rather abstract *model* that can be used for all real processes. Processes in the real world are incredibly complex, but our models of them are quite simple, containing a number of carefully defined concepts. *Processes* (reactions, changes) involve energy and/or mass changes, and these must enter or leave the place where the process is occurring; so thermodynamics begins by defining several types of *systems*, depending on how the energy and/or mass is transferred. Processes must be defined by beginning and ending states, so thermodynamics defines *equilibrium* states, some having more energy (*metastable equilibrium* states) than others (*stable equilibrium* states), and processes or reactions that are able to go from higher energy states to lower energy states (*irreversible processes*), just like a ball rolling down a hill. Of course, a state of lower energy (stable) under one set of conditions

Volume change

The volume data in Appendix B are listed under $V°$, where superscript° means standard state conditions, which we will discuss later. In the corundum – gibbsite reaction, then,

$$\begin{aligned}\Delta_r V^\circ &= V^\circ_{\mathrm{Al_2O_3\cdot 3H_2O}} - V^\circ_{\mathrm{Al_2O_3}} - 3\,V^\circ_{\mathrm{H_2O}} \\ &= 63.912 - 25.575 - 3 \times 18.068 \\ &= -15.867\,\mathrm{cm^3\,mol^{-1}}\end{aligned} \tag{2.8}$$

There is therefore a net decrease in volume of $-15.867\,\mathrm{cm^3\,mol^{-1}}$ for the reaction *as written.* But you could equally well write

$$\begin{aligned}\Delta_r V^\circ &= \tfrac{1}{3}V^\circ_{\mathrm{Al_2O_3\cdot 3H_2O}} - \tfrac{1}{3}V^\circ_{\mathrm{Al_2O_3}} - V^\circ_{\mathrm{H_2O}} \\ &= -5.289\,\mathrm{cm^3\,mol^{-1}}\end{aligned} \tag{2.9}$$

Or you could write

$$\begin{aligned}\Delta_r V^\circ &= 2\,V^\circ_{\mathrm{Al_2O_3\cdot 3H_2O}} - 2\,V^\circ_{\mathrm{Al_2O_3}} - 6\,V^\circ_{\mathrm{H_2O}} \\ &= -31.734\,\mathrm{cm^3\,mol^{-1}}\end{aligned} \tag{2.10}$$

All these results are cm^3 *per mole*, so the question is, per mole of *what?*

Clearly, the meaning of $\Delta_r V°$ in this simple case is the volume change per mole of whatever species have a stoichiometric coefficient of 1.0. The volume change is $-15.867\,\mathrm{cm^3}$ per mole of Al_2O_3 consumed or $Al_2O_3 \cdot 3H_2O$ formed (Equation 2.2; 2.8), and $-5.289\,\mathrm{cm^3}$ per mole of H_2O consumed (Equation 2.9). However, a mole of Al_2O_3 or H_2O need not be consumed, conceptually or in reality. The $\mathrm{cm^3\,mol^{-1}}$ unit is actually a *rate* term, and as MacDonald (1990) points out, just as a car does not need to travel for an hour for its speed to be 100 km hour^{-1}, a mole of reaction need not occur for its $\Delta_r V°$ to be $-15.867\,\mathrm{cm^3\,mol^{-1}}$.

We make another small point about delta notation after introducing the affinity in Chapter 18 (page 567).

may be a state of higher energy (metastable) under other conditions (diamond is metastable at the Earth's surface, but stable deep in the mantle). Corundum and water are, by themselves, perfectly stable and unreactive, but together they have a higher energy state than does gibbsite.

The only thermodynamic difference between organic reactions (including those in living organisms) and inorganic reactions is that both the reactants and products of organic reactions are invariably metastable compounds; metastable, that is, with respect to simple inorganic compounds and elements. Inorganic reactions *may* involve metastable compounds, but more frequently they involve a metastable *assemblage* changing to a stable one (one having the lowest possible energy state).

Therefore, the determination of the energy states of substances and how they change under changing conditions is fundamental to understanding what processes are possible, and why they happen. The determination of the energy states of individual substances must be done by experiment and measurement, not by theoretical calculation, and the results are available in tables of data like those at the end of this book. Calculation of the change of these energy terms with changing conditions can be carried out only for hypothetical *reversible processes*, that are not possible in reality but are quite simple in the thermodynamic model.

As for the energy barriers that often prevent reactions from occurring, thermodynamics has nothing whatever to say about them. It pretends they do not exist. More exactly, thermodynamics simply deals with energy levels, energy differences. It does not concern itself with whether a system actually lowers its energy level or not. This is an important omission, of course, and is the subject of much study in other branches of science, particularly *kinetics*.

The most important question now is what kind of energy is released during these reactions? If it is not heat energy, then what is it? We have called it "chemical energy," but this is just because we haven't said yet what it really is. This is the topic of the next two chapters.

3

The first law of thermodynamics

3.1 Temperature and pressure scales

3.1.1 Temperature

One of the early triumphs of the study of thermodynamics was the demonstration that there is an absolute zero of temperature. However, there are several different temperature scales, for historical reasons. All you need to know about this is that the kelvin scale (named after William Thompson, Lord Kelvin) has an absolute zero of 0 K[1] and a temperature of 273.16 K at the triple point where water, ice, and water vapor are at equilibrium together. The melting point of ice at one atmosphere pressure is 0.01 degrees less than this, at 273.15 K (Figure 3.1). The Celsius scale (named after Anders Celsius, a Swedish astronomer) has a temperature of 0 °C at the ice point (273.15 K) and absolute zero at −273.15 °C. This gives almost exactly 100 °C between the freezing and boiling points of water at one atmosphere, so water boils at 100 °C (373.15 K). Thus the numerical conversion between the two scales is

$$T\,\mathrm{K} = T\,{}^\circ\mathrm{C} + 273.15$$

Remember that *all* equations in thermodynamics use the absolute or kelvin temperature scale, so that if you are given temperatures in °C, you must convert them to the kelvin scale before using them. The “standard” temperature of 25 °C for example is 298.15 K.

3.1.2 Pressure

Force is measured in newtons (N), where 1 newton will give a mass of 1 kg an acceleration of 1 $\mathrm{m\,sec^{-2}}$. Pressure is defined as force per unit area, and a pressure of 1 newton per square meter (1 $\mathrm{N\,m^{-2}}$) is called 1 pascal (1 Pa). This is a very small pressure, and older, larger pressure units are still in use.

---

[1] It is not a point of major importance, but by international agreement, temperatures on the kelvin scale are so many “kelvins,” not “degrees kelvin,” while on the Celsius scale (no longer called the centigrade scale) temperatures are “degrees Celsius.”

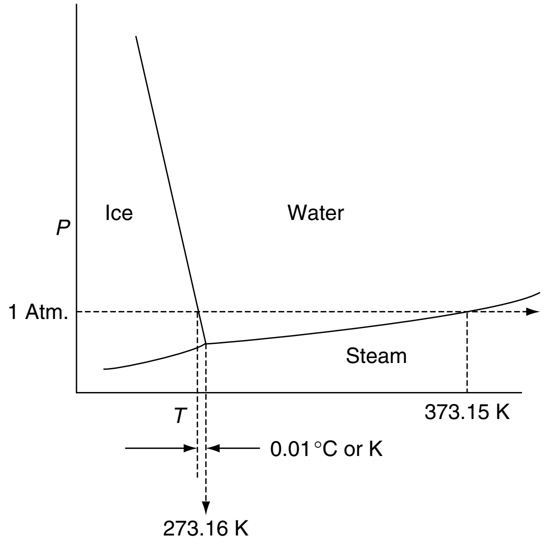


Figure 3.1 Schematic *P*–*T* phase diagram for the system H_2O. The temperature of the triple point is defined as 273.16 K.

The bar, for example, is 10^5 Pa and is almost equal to the standard atmosphere (1 atm = 1.01325 bar). Weather reports in many countries give the atmospheric pressure in kilopascals (kPa), and it is usually close to 101 kPa, or 1 atm, or 1 bar. These units are summarized in Appendix A.

The standard temperature and pressure chosen for reporting values of thermodynamic variables is now 25 °C and 0.1 MPa. A pressure of 0.1 MPa is 100 kPa and 10^5 Pa, or 1 bar. It is convenient to use bars instead of pascals, because the bar is essentially the same as atmospheric pressure, and the notation is slightly simpler.

3.2 Internal energy

In everyday conversation we use words like heat, work, and energy quite frequently, and everyone has a sufficiently good idea of their meaning for our ideas to be communicated. Unfortunately, this type of understanding is not sufficient for the construction of a quantitative model of energy relationships like thermodynamics. To get quantitative about anything, or, in other words, to devise equations relating measurements of real quantities, you must first be quite sure what it is you are measuring. This is not too difficult if you are measuring the weight of potatoes and carrots; it is a more subtle problem when you are measuring heat, work, and energy. Historically, it took several decades of effort by many investigators in the nineteenth century to sort out the difficulties that you are expected to understand by reading this chapter!

3.2.1 Energy

Everyone knows what energy is, but it is an elusive topic if you are looking for a deep understanding. In fact, a Nobel Prize-winning physicist has affirmed that

> It is important to realize that in physics today, we have no knowledge of what energy *is*. (Feynman et al., 1963, pp. 4–2)

One of the most eminent of French scientists said

> As we cannot give a general definition of energy, the principle of the conservation of energy simply signifies that there is *something* which remains constant. (Poincaré, 1952, p. 166)

If you consult a dictionary as to the meaning of energy, you find that the scientific meaning is '*the ability to do work, i.e., move a body.*' In physics, work is not what you do from 9 to 5 every day, but the action of a force moving through a distance. So if you lift a book from the floor and put it on the table, you are performing work (the mass of the book (multiplied by the acceleration due to gravity) times the distance from the floor to the table), and we say that we expended energy to lift the book. It has proved tremendously useful to take the view that the energy we expended has not disappeared, but has been transferred to the book. In other words, the book on the table has more energy (potential energy) than it had on the floor, and the increase is exactly equal to the work we did in lifting it. Thus we can use energy to do work, and we can do work on a system to increase the energy of that system. Work and energy are thus very closely related concepts (note that they have the same dimensions in Appendix A).

If only things were that simple. However, we know that they are not, because the energy in a stick of dynamite on the table is not equal to the work expended in lifting it from the floor. Similarly, the energy in water is not the same as in ice, whether on the floor or the table. These complications are actually of two types.

1. There are many ways of doing work, because there are many kinds of forces. We are particularly concerned with the work involved in chemical reactions.
2. The second is that although work and energy are indeed closely related, doing work is not the only way of changing the energy of something, and changing the energy of something does not always produce work. For example, we could change the energy in our book by warming or cooling it.

We have to consider both work (in all its forms) and heat to get a consistent picture of energy changes. You will notice that although we have been illustrating energy and work by using the ball-in-valley idea (Chapter 2) and the book-and-table idea (this chapter), which emphasizes *potential* energy, this particular kind of energy/work is actually irrelevant in thermodynamics, except as an analogy. We will define the energy content of systems of importance to us to be the same whether they are on the floor or the table.

3.2.2 Absolute energy

In discussing energy, we always seem to be talking about *changes* in energy. The book has more energy on the table than on the floor, and presumably more energy on the roof than on the table. And we add energy by warming the book, too. But how much energy has the book *got* in any particular state – say, on the table at 25 °C? What is the absolute energy content of the book? This was a difficult question until 1905, when Einstein postulated the essential equivalence of mass and energy in his famous equation

$$E_r = mc^2$$

where E_r is the rest energy of a system, m is the mass, and c is the speed of light. Therefore, the energy contained in any macroscopic system is extremely large, and adding energy to a system (for instance by heating it) will in fact increase its mass. However, ordinary (i.e., nonnuclear) energy changes result in extremely small and unmeasurable changes in mass, so that relativity theory is not very useful to us, except in the sense that it gives energy an absolute kind of meaning, which is sometimes helpful in trying to visualize what energy *is*.

Thus in considering ordinary everyday kinds of changes and chemical reactions, we will continue to deal with energy *changes* only, never with how much energy is in any particular equilibrium state. This is entirely sufficient for our needs, but it does introduce some complications that would be avoided if we had a useful absolute energy scale.

3.2.3 The internal energy

All that is required to develop our model of energy relationships is that every equilibrium state of a system (such as our book on the table or the stick of dynamite on the table) have a fixed energy content, called the internal energy, $\mathbf{U}$ (or U, the molar internal energy) of the system. The numerical value of this energy content is not known, and not needed. It could be thought of as identical to the rest energy E_r, if that helps, or as some small subset of E_r; it doesn't really matter. All that matters is that when the system is at equilibrium, its energy content or energy level is constant. Formally, the relation between the total or rest energy and the internal energy used in thermodynamics is

$$E_r = \mathbf{U} + \text{constant}$$

where the value of the constant is unknown (and unimportant). Since we do not use absolute values of $\mathbf{U}$ or U, we cannot use absolute values of any quantities having U in their equations of definition.

Somewhat paradoxically, in spite of being possibly the most fundamental of thermodynamic quantities, changes in U are little used in geochemical applications. It is never listed in tables of thermodynamic values, for example, and one rarely needs to calculate ΔU. The reason for this will become apparent as we proceed. It has to do with the fact that we, the users of thermodynamics, have a great predilection for using temperature, pressure, and volume as our principal constraints or measured system parameters. It turns out that this requires that we use ΔU in slightly modified forms, that is, ΔU modified by what are often relatively small correction factors (such as $P\Delta V$), and these modified forms are given different names and symbols. It is then quite possible to rarely think about ΔU, since it seems only to arise in the development of the first law. For a better understanding of the subject, however, it is best to realize that in most energy transfers in the real problems that we will be considering, ΔU is by far the largest term involved. Just because we do not usually calculate its value does not mean it is not important.

3.3 Energy transfers

In the discussions in the previous chapters, we proposed the idea that changes or reactions occur because systems can lower their energy by such changes. However, we mentioned that the most obvious kind of energy, heat energy, was not the right kind of energy. There is another very common kind – energy expended as *work*, as when dynamite is used to break rock. However, work energy is not the answer to our questions either, nor is the combination of heat and work. Nevertheless, they are extremely important, and together form the basis of the first law.

- *Heat* ($\boldsymbol{q}$) is the energy that flows across a system boundary in response to a temperature gradient.
- *Work* ($\boldsymbol{w}$) is the energy that flows across a system boundary in response to a force moving through a distance (such as happens when a system changes volume).

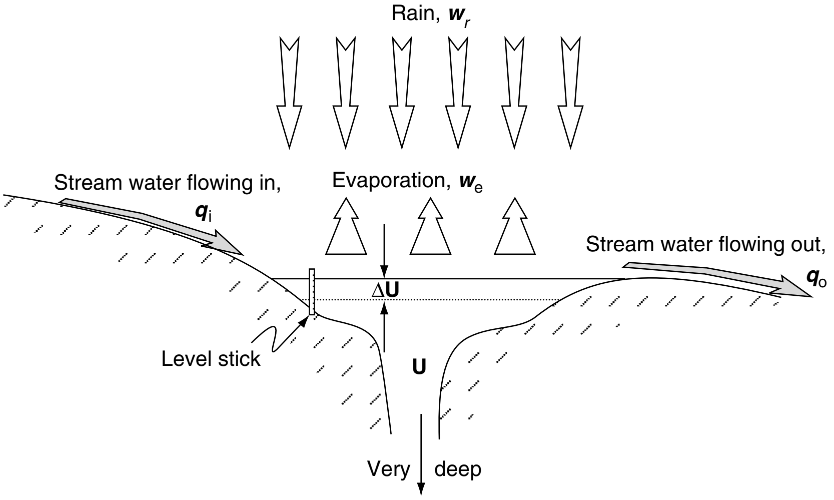
Heat and work[2] are therefore not separate entities as such but are forms of energy that are transferred in different ways. An enlightening analogy has been offered by Callen (1960). In Figure 3.2 we consider the water in a very deep pond (the amount of water is thus very great but finite and in principle could be exactly measured) to correspond to the internal energy **U** of a system.

Water may be added and subtracted from the pond either in the form of stream water (heat) or precipitation/evaporation (work). Both the inlet and outlet stream water can be monitored by flow gauges, and the precipitation

---

[2] We use $\boldsymbol{q}$ and $\boldsymbol{w}$ for increments or amounts of heat and work, and q and w where the molar form is appropriate. Many texts use δq and δw for the same things, but not dq or dw, which would imply they are exact differentials (see §C.2.1).

Figure 3.2 The pond analogy for the first law.



measured by a rain gauge. Evaporation would be trickier to measure, but we may assume that we have a suitable measure for it. Now if the volume of stream inlet water over some period of time is $\boldsymbol{q}_{\mathrm{i}}$, the stream outlet water $\boldsymbol{q}_{\mathrm{o}}$, the rain $\boldsymbol{w}_r$, and the evaporation $\boldsymbol{w}_{\mathrm{e}}$, then if there are no other ways of adding or subtracting water, clearly

$$\Delta\mathbf{U} = (\boldsymbol{q}_{\mathrm{i}} - \boldsymbol{q}_{\mathrm{o}}) + (\boldsymbol{w}_r - \boldsymbol{w}_{\mathrm{e}})$$

where $\Delta\mathbf{U}$ is the change in the amount of water in the pond, which could be monitored by a level indicator as shown. Thus

$$\Delta\mathbf{U} = \boldsymbol{q} + \boldsymbol{w}$$

where

$$\boldsymbol{q} = \boldsymbol{q}_{\mathrm{i}} - \boldsymbol{q}_{\mathrm{o}}$$

and

$$\boldsymbol{w} = \boldsymbol{w}_r - \boldsymbol{w}_{\mathrm{e}}$$

Once water has entered the pond, it loses its identity as stream or rain water. The pond does not contain any identifiable stream water or rain water, simply water. Similarly systems do not contain so much heat or work, just energy. Just as the water level in the pond can be raised *either* by stream water alone *or* by rain water alone, Joule showed in the nineteenth century that a temperature rise in a water bath of so many degrees can be caused *either* by heating (transferring energy due to a temperature difference) *or* by thrashing a paddle wheel about in it (transferring energy by force through distance, i.e., by deformation of the system boundary).

Another implication or assumption in our pond analogy is that water is conserved, that is, it cannot simply disappear as if by magic. The same proposition regarding energy is known as the first law of thermodynamics. We invoked this principle when we said that the energy expended in lifting the book from the floor to the table was not lost, but transferred to the book.

3.4 The first law of thermodynamics

The first law of thermodynamics is the law of conservation of energy. If **U** is the energy content of a system, and it may gain or lose energy only by the flow of heat (q) or work (w), then clearly, as in the pond analogy, $\Delta\mathbf{U}$ must be the algebraic sum of q and w. In order to express this algebraically, we must have some convention as to what direction of energy flow $+q$, $-q$, $+w$, and $-w$ refer to. In the pond analogy we assumed implicitly that addition of water to the pond was positive, whether as stream water or rain water. Thus heat added *to* a system is positive, and work done *on* a system is positive. This convention may be represented as in Figure 3.3a and is what we call the "scientific" convention – scientists like it because it is internally consistent. It results in the equation previously found,

$$\boxed{\Delta\mathbf{U} = q + w} \tag{3.1}$$

Another convention (Figure 3.3b) is to say that heat added to a system is positive, but that work done *by* a system is also positive, or that work done *on* a system is negative. This we call the "engineering" convention, because engineers prefer to think in terms of heat engines, and an engine doing work is something positive. This results in the relation

$$\Delta\mathbf{U} = q - w$$

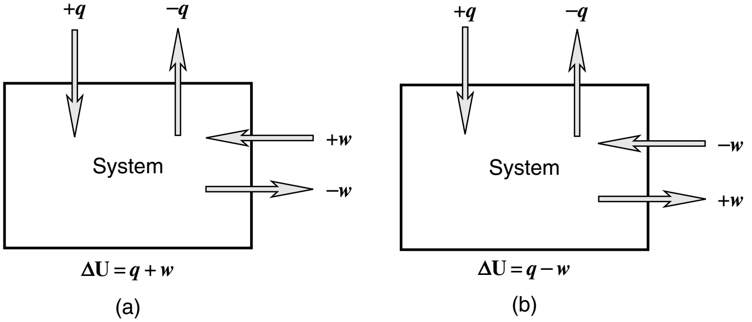


Figure 3.3 The two commonly used conventions for the sign of q and w, leading to two formulations of the first law.

and also results in slightly simpler equations expressing pressure–volume work (the minus signs in Equations (3.3) to (3.7) would be missing). In this text we will use the scientific sign convention. Any additions of matter and energy to the system are positive in sign and all losses are negative.

Note that we have not "proved" the first law. It is a principle that has been deduced from the way things work in our experience, but the fact that it has never been known to fail does not constitute a proof. Neither does the fact that the sun has never failed to rise in the east constitute a proof that it will rise in the east tomorrow, but I wouldn't bet against it.

3.4.1 Work

Types of work

There are many different ways of doing work on a system, and many different ways of having a system do work, depending on what kinds of forces are available. For example,

- The force of gravity means that we have to do work to lift objects, as mentioned above. If the mass is m, the acceleration due to gravity is g, and the distance is dh, the work $\boldsymbol{w}$ is $\boldsymbol{w} = mg \cdot dh$.[3]
- Tensile force can be used to stretch a wire. If the tensile force is f and the increase in length is dl, $\boldsymbol{w} = f \cdot dl$.[4]
- Tensile force can be used to increase the area of a soap film. If the surface tension is γ and the increase in surface area is dA, then $\boldsymbol{w} = \gamma \cdot dA$.
- Expansion due to the heating of a gas, or indeed of anything at all, produces a force. This case is of special interest to us, because the work done by expansion or contraction of systems due to a change in temperature at constant pressure cannot be avoided. We can choose to eliminate other forms of work, but not this one (unless we consider only constant volume systems, which is useful at times, but not very practical). It is treated in more detail below.
- Chemical work. For example, a battery can be used to do work, because a chemical reaction occurs in it which produces a voltage. The work done by chemical reactions is of course a principal focus of chemical thermodynamics, and the equations for it will be developed in later chapters.

There are others, such as work done by centripetal and frictional forces, that you can review in a physics text. Thermodynamics can accommodate all kinds of forces and types of work, but because they are in principle all the same, and are treated in the same way, it is simpler to develop the subject by considering

---

[3] Don't be confused by the differential term dh here. This is *not* (or at least not necessarily) an infinitesimal distance. It is *any* distance in the h direction. We could call it Δh, but we will later include it in a differential equation (§4.7), so we might as well use the differential notation here too. It is only when we integrate that dh must become very, very small. This is treated in more detail in Appendix C.

[4] This assumes the wire deforms elastically.

only those forms of work that we cannot avoid. Therefore the basic structure of thermodynamics is always developed using heat and pressure–volume work, and other forms of work considered afterward. In our case, the only other form of work of any importance is chemical reaction work.

Pressure–volume work

Work in natural environments is for the most part only of one kind – the work of expansion, or pressure–volume work. Pressure–volume work is always discussed using a piston-and-cylinder arrangement as shown in Figure 3.4. This seems natural to engineers, but may seem rather artificial or even useless to someone interested in processes that happen in nature or in the environment. You have to realize that virtually *all* processes in *all* natural systems involve some change in volume, and therefore work is done against the pressure on the system, whatever that is (it is very often atmospheric pressure). We use a piston–cylinder arrangement for convenience – any system that changes volume could be used. Once we have found the appropriate equations for pressure–volume work, we can use them in our models of any system, whether or not they have pistons and cylinders.

The piston–cylinder arrangement as shown in Figure 3.4 is not a real piston in a real cylinder, of course, but a conceptual one, so we can give it whatever properties we like. We must be careful about this, however, otherwise the results will be useless. The cylinder is fitted with some devices that can hold the piston in position at various levels. When the piston is held stationary, the forces tending to move the piston are balanced (force pushing up equals force pushing down). If this were not the case, the piston would move. The two forces are acting on opposite sides of the same piston, having the same area (and force/area = pressure) so the pressure of the gas, P_{int}, is exactly balanced by the external pressure, P_{ext}. The external pressure is provided partly by the stops that are holding it in place and partly by the weight of the piston itself, plus any weights on the piston. If the stops are removed, then all of a sudden P_{ext} is reduced to that produced by the piston and weights only, $P_{\text{int}} \gg P_{\text{ext}}$,

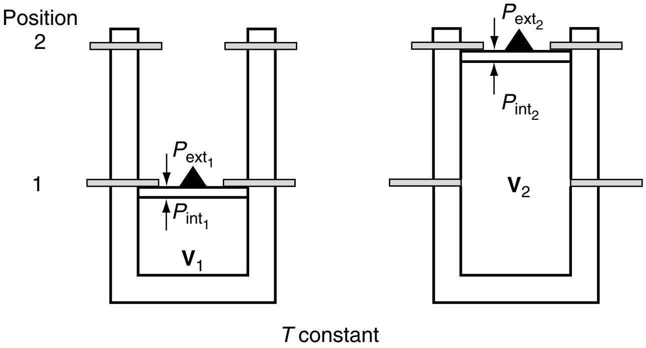


Figure 3.4 Irreversible expansion of a gas from external pressure P_{ext_1} to P_{ext_2}. During expansion, external pressure is fixed by the weight of the piston plus the weights on the piston.

Other forms of work

A block weighing 10.0 kg is lifted 4 meters at a place where $g = 9.80\,\mathrm{m\,sec^{-2}}$. The work done is

$$\begin{aligned} \boldsymbol{w} &= mg \cdot dh \\ &= 10.0 \times 9.80 \times 4 \\ &= 98.0\,\text{newtons} \times 4\,\text{meters} \\ &= 392\,\text{joules} \end{aligned}$$

A film of water has a surface tension of $\gamma = 72 \times 10^{-3}\,\mathrm{N\,m^{-1}}$. The work done in expanding its area by $1\,\mathrm{cm}^2$ $(= 10^{-4}\,\mathrm{m}^2)$ is

$$\begin{aligned} \boldsymbol{w} &= \gamma \cdot dA \\ &= 72\,\mathrm{N\,m^{-1}} \times 10^{-4}\mathrm{m}^2 \\ &= 72 \times 10^{-4}\,\text{joules} \end{aligned}$$

and the piston moves up until it encounters more stops – WHAP! – and all of a sudden $P_{\text{int}} = P_{\text{ext}}$ once more, though at a different (lower) pressure (the experiment has been arranged such that the gas pressure is 10 pressure units at the upper stops, which is position 2, and 20 pressure units at the lower stops, position 1). Real gases tend to cool during expansion, so if we want the initial and final states to be at the same temperature, some heat must flow into the cylinder from the surroundings.

At this stage, one normally says something like "If the piston is well-lubricated and well-constructed, we can ignore friction effects" but we know we are conducting a model experiment, so we just say there is no friction in our model. The pressure–volume history of the change can be illustrated as in Figure 3.5. The external pressure during expansion is constant, since it is fixed by the mass of the piston. The work done during the expansion is[5]

$$\begin{aligned} \boldsymbol{w} &= \text{force} \times \text{distance} \\ &= -(\text{total mass} \cdot g) \cdot \Delta h \\ &= -(P_{\text{ext}} \cdot A) \cdot \Delta h \\ &= -P_{\text{ext}}(A \cdot \Delta h) \\ &= -P_{\text{ext}} \cdot \Delta \mathbf{V} \end{aligned} \tag{3.2}$$

---

[5] The work done will also include a term $(\frac{1}{2}mv^2)$ for the work done in accelerating the piston. If we let the stops be part of the system, this kinetic energy is returned to the system at the upper stops, and can be neglected (Kivelson and Oppenheim 1966).

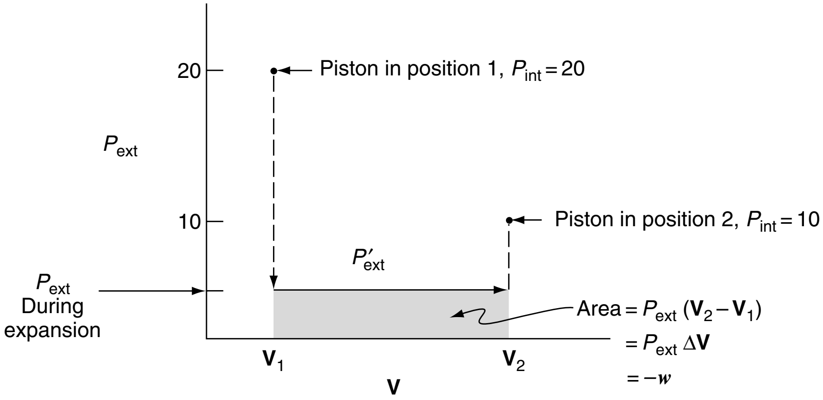


Figure 3.5 External pressure (P_{ext}) versus volume (**V**) plot for the irreversible expansion of the gas in Figure 3.4. The vertical dashed lines indicate an instantaneous change in pressure. The solid horizontal lines indicate change in volume at constant pressure.

where A is the area of the piston and Δh the distance it travels, so w is the area under the path of expansion or expansion curve in Figure 3.5. The minus sign is because the system is doing work. If we repeat the process, but this time we place a larger weight on the piston, exactly the same thing will happen, but more work is done because a greater mass was lifted through the same volume.

If another weight is added for the next expansion, we may have a total weight that is too great to allow the piston to reach the upper stops (position 2) and it will come to rest (equilibrium) somewhere in between. Then if the second weight is removed, the piston will proceed upward again as before, giving an expansion path as shown in Figure 3.6. If we use a lot of weights and remove them one at a time, letting the piston come to rest after each step, we will get a path such as shown in Figure 3.7.

Clearly we are approaching a limit of maximum work obtainable from the expansion of our gas, and clearly too, the maximum will be when we take an infinite number of infinitesimally small incremental steps from $\mathbf{V}_1$ to $\mathbf{V}_2$.

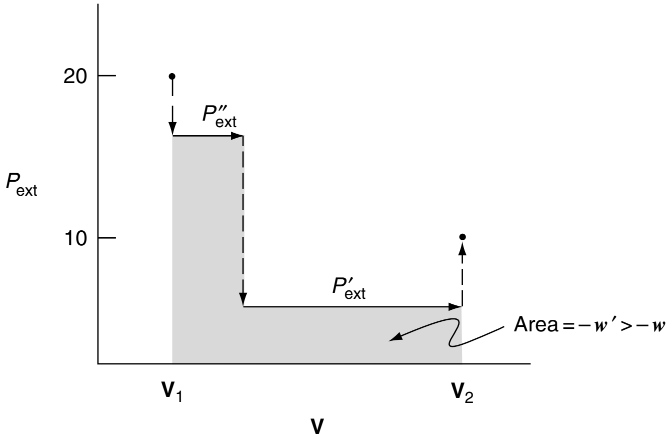


Figure 3.6 External pressure (P_{ext}) versus volume (**V**) for a two-stage expansion of gas. After an initial expansion at P''_{ext}, some weight was removed from the piston and the expansion continued at P'_{ext}.

$P\Delta V$ work

In Figure 3.5, suppose the pressure units are bars, $\mathbf{V}_1 = 1000\,\text{cm}^3$ of ideal gas, and during expansion P_{ext} is 5 bars. How much work is done?
The pressure on the gas is decreased by half (from 20 to 10 bars), so the (ideal) gas will expand to twice its volume ($P\mathbf{V} =$ constant), so $\mathbf{V}_2 = 2000\,\text{cm}^3$. Then

$$\begin{aligned} w &= -P_{\text{ext}}(\mathbf{V}_2 - \mathbf{V}_1) \\ &= -5 \times (2000 - 1000) \\ &= -5000\,\text{bar cm}^3 \end{aligned}$$

To convert this to joules, Appendix A gives the conversion $1\,\text{bar} = 0.10\,\text{J cm}^{-3}$, so

$$w = -500\,\text{J}$$

Note the minus sign, which indicates the system is doing work. If $\mathbf{V}_2$ were less than $\mathbf{V}_1$, $\Delta\mathbf{V}$ would be negative and w would be positive, meaning work is done on the system.

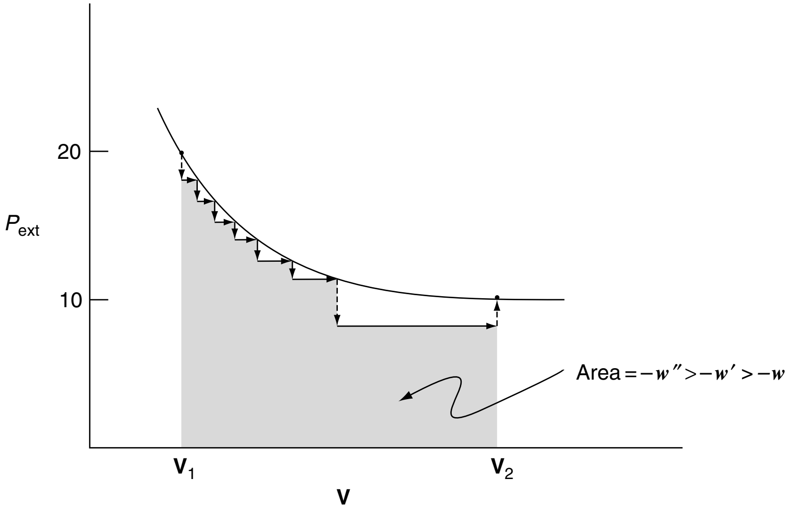


Figure 3.7 External pressure (P_{ext}) versus volume ($\mathbf{V}$) for a multistage expansion of gas. After each constant P_{ext} expansion, some weight was removed, allowing a further expansion.

Since we have been letting the piston come to rest or equilibrium after every weight removal, in the limit we will have an infinite number or continuous succession of equilibrium states, giving us an example of a reversible process. In this particular case the name "reversible" is particularly appropriate since at any stage in the expansion the direction of movement can be reversed by changing the external pressure infinitesimally.

In the limit when infinitesimal increments of **V** are taken, the work of expansion is (see Figure 3.8)

$$\boldsymbol{w}_{\text{rev}} = \boldsymbol{w}_{\max} = -\int_{\mathbf{V}_1}^{\mathbf{V}_2} P\,d\mathbf{V} \tag{3.3}$$

Here we need make no distinction between P_{ext} and P_{int} because they are never more than infinitesimally different in our continuous succession of equilibrium states. Again, note the negative sign required to comply with the scientific sign convention.

Since the end positions 1 and 2 of our expansion in every case consisted of our gas at stable equilibrium at a fixed P and T, then according to the first law there is a fixed energy difference $\Delta\mathbf{U}$ between the two states. We have gone to some length to show that there is no fixed "difference in work," or work available from the change from one state to the other. Thus we are led to believe that the amount of heat flowing into our thermostatted cylinder must at all times, once equilibrium was established, have compensated for the variations in work performed, giving the same total $\boldsymbol{q}+\boldsymbol{w}$ in every case. We could verify this, of course, by making calorimetric measurements, but this is basically what Joule and many other workers have already done.

Our intent here is not so much to illustrate the constant energy change between states, but that this energy change, while accomplished by heat and work, can be made up of an infinite variety of combinations of heat and work. When the process is made reversible, we get the maximum work of expansion, and this will be given by Equation (3.3), but even so, we are unable to calculate this amount of work (evaluate the integral) without more information (we need to know P as a function of **V** so that we can integrate Equation 3.3).

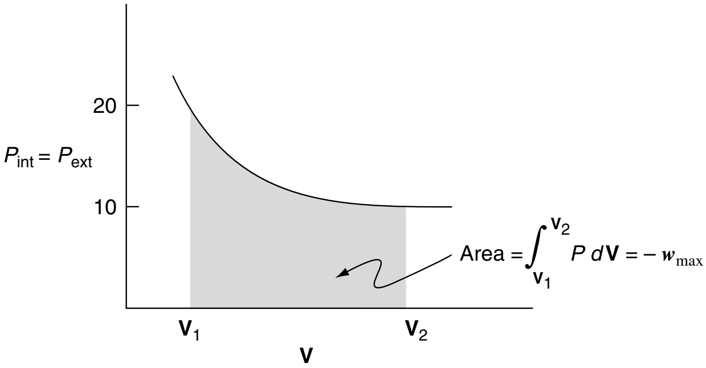


Figure 3.8 Pressure versus volume for the reversible expansion of a gas. The limiting case where an infinite number of constant P_{ext} steps are taken gives the maximum area under the curve. During the expansion, internal pressure and external pressure are never more than infinitesimally different, or $P_{\text{int}} = P_{\text{ext}}$ at all times.

The integration of (3.3) at constant external pressure results in

$$\boldsymbol{w} = -P_{\mathrm{ext}}(\mathbf{V}_2 - \mathbf{V}_1) \tag{3.4}$$

$$= -P_{\mathrm{ext}}\,\Delta\mathbf{V} \tag{3.5}$$

as in Equation (3.2). The *internal* pressure necessarily varies during this expansion, as discussed above. It is also possible to integrate at constant internal pressure P, but this is necessarily a reversible process, because the only way to do this is to have the external and internal pressures equal at all times. This results in

$$\begin{aligned} \boldsymbol{w}_{\mathrm{rev}} &= -P(\mathbf{V}_2 - \mathbf{V}_1) \\ &= -P\Delta\mathbf{V} \end{aligned} \tag{3.6}$$

In this case, for both P and $\mathbf{V}$ to vary, the temperature would have to vary as well. See Figure 4.10 for an example, although for a reversible compression rather than expansion.

A point worth emphasizing is that in any real or nonreversible expansion, as shown in our example, the work obtained is less than the maximum obtainable (from a reversible expansion). Thus in general, rewriting (3.3),

$$-\boldsymbol{w} \le \int_{\mathbf{V}_1}^{\mathbf{V}_2} P\,d\mathbf{V} \tag{3.7}$$

where the $<$ part of the $\le$ sign refers to any irreversible change in $\mathbf{V}$. This can also be expressed as[6]

$$-\boldsymbol{w} \le -\boldsymbol{w}_{\mathrm{max}}$$

or

$$-\boldsymbol{w} \le -\boldsymbol{w}_{\mathrm{rev}}$$

For the opposite case of compressing the gas from position 2 back to position 1, the inverse series of steps can be employed. Thus, if at position 2 a heavy weight is placed on the piston, it will WHAP down to the stops at position 1, describing a path such as in Figure 3.9. Obviously, much more work has had to be done in compressing the gas than we obtained, even in the reversible case, from expansion. However, by adding a larger number of smaller weights one at a time we can reduce the amount of work required

---

[6] These minus signs can certainly be confusing. In equations like (3.4), which have an = sign, it doesn't matter whether the minus sign is on the left or the right. In inequality expressions like (3.7), it does. The "engineering convention" (page 37) is perhaps less confusing in this respect.

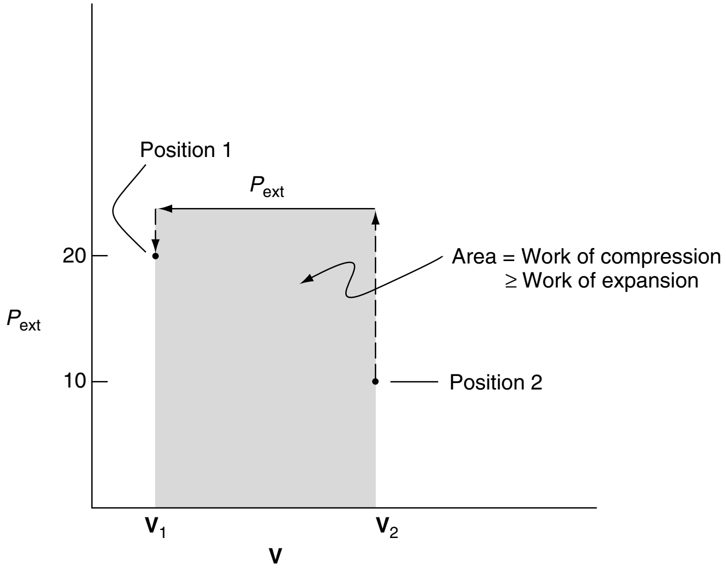


Figure 3.9 External pressure (P_{ext}) versus volume for the irreversible compression of gas at constant P_{ext}.

for the compression, gradually approaching the stable equilibrium curve from above, rather than from below as before. In the limit, of course, we find that for a reversible compression the work required is exactly the same as the work available from a reversible expansion.

Considering this work stuff in such detail may make it look complex, but it really is not. Just remember that if you need to calculate work (which happens surprisingly little in geochemistry), you need either a constant pressure process, or a reversible one (so you can integrate). For "real" work processes, the work done is invariably less than the reversible work (Equation 3.7), usually much less, and usually of more interest to engineers than to geochemists.

3.4.2 Heat

It might be expected that since

$$\Delta \mathbf{U} = q + w$$

and

$$-w \leq \int_{\mathbf{V}_1}^{\mathbf{V}_2} P \, d\mathbf{V} \tag{3.8}$$

perhaps there is a very similar story for the heat transfers in the gas expansion cases we have been considering. That is, perhaps

$$-q \leq \int_{\mathbf{Z}_1}^{\mathbf{Z}_2} T \, d\mathbf{Z} \tag{3.9}$$

where **Z** is some property of the gas. This is indeed the case (except for a sign change), but we must await the development of the second law, which will introduce us to entropy ($-\mathbf{Z}$ in Equation 3.9).

3.4.3 The molar forms

Up to here we have considered total work, total heat and total energy involved in processes. In future chapters where we deal primarily with chemical reactions, it will often be more convenient to use the molar forms (§2.4.1), for example

$$\Delta U = q + w \tag{3.10}$$

$$w = -P_{\text{ext}}\Delta V \tag{3.11}$$

$$w_{\text{rev}} = -P\Delta V \tag{3.12}$$

$$-w \leq \int_{V_1}^{V_2} P\,dV \tag{3.13}$$

and

$$-q \leq \int_{Z_1}^{Z_2} T\,dZ \tag{3.14}$$

Forms of work other than the work done by chemical reactions are not usually expressed in a molar form, so we could be more explicit by writing (3.12), for example, as

$$w_{\text{rev}} = -P\Delta_r V$$

but the less explicit form is not incorrect.

3.5 Enthalpy, the heat of reaction

In processes at constant external pressure, the work done, as we have seen (Equation 3.11), is $-P_{\text{ext}}\Delta V$.[7] In the present derivation it doesn't matter whether we use P_{ext} or P, as long as it is constant. Therefore the first law can be written

$$\Delta U = q_P - P\,\Delta V$$

---

[7] You might see Equation (3.11) written as $w = -P\,\Delta V$. The difference is, to repeat, that P_{ext} is the (constant) external pressure (often it is atmospheric pressure) on the system which may be undergoing any sort of process, whereas unsubscripted P refers to the *system* pressure, which can have meaning only when the system is at equilibrium. Therefore if the system is undergoing a process, and P is constant, that process must be a reversible one, in which the heat flow in or out of the system constantly compensates for any tendency of the pressure to change. See also Figure 4.10.

Heat of reaction

The standard heat of reaction for reaction (2.2) is

$$\begin{aligned}\Delta_r H^\circ &= \Delta_f H^\circ_{\mathrm{Al_2O_3 \cdot 3H_2O}(s)} - \Delta_f H^\circ_{\mathrm{Al_2O_3}(s)} - 3\Delta_f H^\circ_{\mathrm{H_2O}(l)} \\ &= -2586.67 - (-1675.7) - 3(-285.83) \\ &= -53.48\,\mathrm{kJ\,mol^{-1}} \\ &= -53\,480\,\mathrm{J\,mol^{-1}}\end{aligned}$$

Again, the minus sign means heat is evolved (exothermic reaction). This amount of heat would raise the temperature of a liter of water about 12 °C.

where q_P is the heat transferred in the constant pressure process. Thus

$$q_P = \Delta U + P\,\Delta V$$

from which we can see that the heat transferred in constant pressure processes is equal to a function involving only state variables, and so it is itself a state variable. Don't forget that we have gone to some trouble to show that in general neither q nor w is a state variable; it is only in the special case of constant pressure processes that they both become state variables. Because of this, it is useful to define a new term, enthalpy,

$$H = U + PV \tag{3.15}$$

which has the differential form

$$dH = dU + P\,dV + V\,dP$$

At constant pressure, this becomes

$$dH_P = dU + P\,dV \tag{3.16}$$

or

$$\Delta H_P = \Delta U + P\,\Delta V \tag{3.17}$$

and since

$$\Delta U = q_P - P\,\Delta V$$

therefore

$$\Delta H_P = q_P \tag{3.18}$$

All we have done here is notice that, because work becomes a fixed quantity in constant pressure processes, then heat does too, by the first law. And because constant pressure processes are so common (including all reactions carried out at atmospheric pressure, such as most biochemical reactions), it is convenient to have a state variable defined to equal this heat term. Defining enthalpy as in (3.15) accomplishes this, and we now have a “heat of reaction” term, which will be useful in all constant pressure processes.

Note that because H is a state variable, ΔH is perfectly well defined between any two equilibrium states. But when the two states are at the same pressure, ΔH becomes equal to the total heat flow during the process from one to the other, and in practice enthalpy is little used except in this context.[8] Processes having a negative $\Delta_r H$ ($\Delta_r H < 0$) are termed *exothermic*, and those having a positive $\Delta_r H$ are termed *endothermic*.

3.5.1 Additivity of state variables

At several points in our discussions so far, we have mentioned or assumed that we can add and subtract state variables such as ΔH, for example as shown in Figure 5.3. This is perhaps obvious, but it is so fundamental that we will emphasize it here.

We said in Chapter 2 (§2.4) that a state variable is a property of a system that has a fixed value when the system is at equilibrium, whether we know the value of that property or not. For example, a mole of water at 25 °C, 1 atm has a fixed but unknown enthalpy H, and fixed values of all other state variables. We also said that this means that the changes in these properties between equilibrium states depends only on the equilibrium states, and not on what happens between the time the system leaves one equilibrium state and the time it settles down in its new equilibrium state. Therefore, if two different reactions produce the same compound, we can subtract the $\Delta_r H^\circ$, for example, of these reactions to get the Δ of the combined reaction, and the properties of that compound will cancel out. For example, carbon dioxide, CO_2, might be produced from the oxidation of either graphite or carbon monoxide, CO:

$$\mathrm{C} + \mathrm{O_2} = \mathrm{CO_2} \qquad \Delta_r H^\circ = -393.509\,\mathrm{kJ\,mol^{-1}}$$

$$\mathrm{CO} + \tfrac{1}{2}\mathrm{O_2} = \mathrm{CO_2} \qquad \Delta_r H^\circ = -282.984\,\mathrm{kJ\,mol^{-1}}$$

Subtracting the reactions and the $\Delta_r H^\circ$ values (reverse the second reaction, change the sign of $\Delta_r H^\circ$, and add), we get

$$\mathrm{C} + \tfrac{1}{2}\mathrm{O_2} = \mathrm{CO} \qquad \Delta_r H^\circ = -110.525\,\mathrm{kJ\,mol^{-1}}$$

Thus we get the properties of a reaction that is impossible to carry out experimentally from two reactions that are relatively easy do experimentally.

---

[8] Another use of the enthalpy is discussed in §6.2.3.

3.5.2 Enthalpy of formation from the elements

A major problem arises from the definition of enthalpy, Equation (3.15). The problem is that we cannot measure it. This arises from the nature of energy itself, because we can only measure energy changes, not absolute energies. Therefore we can only measure enthalpy changes, and changes in any other property which includes the energy **U** (or U).

The problem this creates is that we do not want to have to tabulate an enthalpy change for every process or chemical reaction which might become of interest to us – there are too many. We would like to be able to associate an enthalpy with every substance – solids, liquids, gases, and solutes – for some standard conditions, so that having tabulated these, we could then easily calculate an enthalpy change between any such substances under those standard conditions. After that, we could deal with the changes introduced by impurities and other nonstandard conditions. The method developed to allow this is to determine, for every pure compound, the difference between the enthalpy of the compound and the sum of the enthalpies of the elements, each in its most stable state, which make up the compound. This quantity is called $\Delta_f H^\circ$, the standard molar enthalpy of formation from the elements. For aqueous ions, the quantity determined is a little more complicated (Chapter 15), but the principle is the same. It is this enthalpy quantity which is invariably tabulated in compilations of data.

For example, the standard enthalpy of formation of anhydrite is

$$\Delta_f H^\circ_{\mathrm{CaSO_4}(s)} = H^\circ_{\mathrm{CaSO_4}(s)} - H^\circ_{\mathrm{Ca}(s)} - H^\circ_{\mathrm{S}(s)} - 2\,H^\circ_{\mathrm{O_2}(g)} \tag{3.19}$$

where the superscript ° refers to the standard conditions (see below). None of the individual H° quantities is determinable, but the difference is determinable by calorimetry. Now if we want to know the heat liberated or absorbed in a chemical reaction, we need only look up these $\Delta_f H^\circ$ values for each reactant and product. For example, for the formation of gypsum from anhydrite, we write

$$\mathrm{CaSO_4}(s) + 2\,\mathrm{H_2O}(l) = \mathrm{CaSO_4}\cdot 2\,\mathrm{H_2O}(s) \tag{3.20}$$

for which the "standard molar heat of reaction," $\Delta_r H^\circ$, is

$$\Delta_r H^\circ = \Delta_f H^\circ_{\mathrm{CaSO_4\cdot 2H_2O}} - \Delta_f H^\circ_{\mathrm{CaSO_4}} - 2\Delta_f H^\circ_{\mathrm{H_2O}(l)} \tag{3.21}$$

$$= H^\circ_{\mathrm{CaSO_4\cdot 2H_2O}} - H^\circ_{\mathrm{CaSO_4}} - 2\,H^\circ_{\mathrm{H_2O}(l)} \tag{3.22}$$

Note that in balanced reactions the H° terms for all the elements cancel out, and we are left with the "real" enthalpy difference (Equation 3.22) between products and reactants, with no contribution from arbitrary conventions or assumptions. It is, however, a heat of reaction for standard conditions only.

Standard states

It is often assumed that the "standard conditions" are 25 °C, 1 bar. Actually it is a bit more complicated in two respects.

1. Knowing the T and P of the state is not sufficient – we must also specify the physical state of the substance. For solids and liquids, it is simply the pure substance (as in our anhydrite-gypsum-water example) but for gases it is the gas acting ideally at one bar (or 10^5 Pa), and for solutes it is the solute acting ideally at a concentration of one molal. The reasons for these choices will be discussed later (Chapter 8).
2. While the temperature and pressure of the standard conditions are indeed 25 °C and one bar for purposes of tabulating data, we can and often do have standard conditions at any T and P.

These more complete definitions of our "standard conditions" define our *standard states*, which will be seen to become particularly useful when we later define the concept of *activity* (§8.2).

3.5.3 The heat capacity

An older name for the enthalpy is the "heat content." This name is somewhat discredited for good reasons, but nevertheless it helps a little in conveying the essential idea behind the next concept, the heat capacity. The molar heat capacity can be defined as the amount of heat required to raise the temperature of one mole of a substance by one degree. Of course, some substances require much more heat to do this than do others.

The formal definition is

$$\left(\frac{dH}{dT}\right)_P = C_P \tag{3.23}$$

or

$$\left(\frac{d\Delta H}{dT}\right)_P = \Delta C_P \tag{3.24}$$

or

$$\left(\frac{d\Delta H^\circ}{dT}\right)_P = \Delta C_P^\circ \tag{3.25}$$

Thus heat capacity is the rate of change in H or ΔH with T. A large C_P means that H or ΔH changes a lot for a given change in T, i.e., it takes a lot of heat to raise the temperature.

It takes a different amount of heat to raise the temperature of a system depending on whether the volume or the pressure is kept constant, giving two different quantities, C_P and C_V. C_V is rarely used in geochemistry, but the heat capacity at constant pressure, C_P, is a surprisingly important quantity. It can be used to calculate not only $\Delta_r H$ at high temperatures, as in the next section, but the high temperature values of several other important quantities that we will be considering.

Temperature dependence of the heat capacity

Many different equations have been suggested to represent the variation of C_P with temperature, and several are in current use. No differences in principle are involved, so we will consider only three of these equations.

Maier–Kelley

This equation was suggested by Maier and Kelley (1932), and is used in the program SUPCRT92 (Johnson et al., 1992) (except for a different sign for the c term) for minerals and gases, to be described later. It is

$$C_P = a + bT - cT^{-2} \tag{3.26}$$

Thus to know how C_P for a substance varies with T we need only look up the values of a, b, and c for that substance. However, it is important to note that these coefficients are only available for pure solids, liquids and gases, because C_P for pure substances increases in a fairly simple way with T. Aqueous solutes, however, have a much more complex behavior we will describe later.

The Maier–Kelley coefficients are available from various literature sources, but the most convenient source is SUPCRT92. For example, for quartz, SUPCRT92 shows the coefficients in Table 3.1. This means that $a = 11.22$, $b \times 10^3 = 8.2$, and $c \times 10^{-5} = -2.7$, or

$$a = 11.22$$

$$b = 0.0082$$

$$c = -270\,000$$

These values are only valid up to 574.85 °C, where a phase transition (from α-quartz to β-quartz) takes place. At temperatures above the transition, $a = 14.410$, $b = 0.001\,94$, and $c = 0$.

Inserting these values into equation (3.26) (and converting °C to kelvins), you find that C_P varies as in Table 3.2, showing that the amount of heat required

Table 3.1 *Part of some* SUPCRT92 *output.*

```
                          MAIER-KELLY COEFFICIENTS
NAME                      a(10**0)    b(10**3)    c(10**-5)    T limit (C)
----------------          --------    --------    --------     ---------
QUARTZ                     11.220       8.200       -2.700        574.85
  post-transition 1        14.410       1.940         .000       1726.85
```

Table 3.2 *Heat capacity of quartz as a function of temperature.*

	SUPCRT92		Berman (1988)	Shomate
T °C	cal mol^{-1} K^{-1}	J mol^{-1} K^{-1}	J mol^{-1} K^{-1}	J mol^{-1} K^{-1}
25	10.63	41.19	44.74	44.58
100	12.34	51.63	51.56	51.39
200	13.89	58.13	57.76	58.16
300	15.10	63.17	61.79	63.22
400	16.14	67.55	64.53	67.58
500	17.11	71.58	66.50	72.26

to raise the temperature of quartz (and most pure substances) increases as the temperature increases.

In most geochemical work, these numbers are not of much interest by themselves. However, they are the key to using thermodynamics at temperatures above 25 °C.

There are two things to remember when using Maier–Kelley coefficients from SUPCRT92.

1. SUPCRT92 uses calories throughout. Normally you will want to convert to joules.
2. SUPCRT92 uses the equation $C_P = a + bT + cT^{-2}$, rather than Equation (3.26) so if you are using (3.26) you must change the sign of the SUPCRT92 c-term (or, if you prefer, SUPCRT92 does use Equation (3.26), but includes the minus sign with the tabulated c-term).

For chemical reactions in which solutes are not involved, the change in each coefficient between products and reactants is evaluated in the usual way. For example, for reaction of anhydrite plus water to form gypsum, Equation (3.20),

$$\begin{aligned}\Delta_r a &= a_{\mathrm{CaSO_4 \cdot 2H_2O}} - a_{\mathrm{CaSO_4}} - 2\,a_{\mathrm{H_2O}(l)}\\ \Delta_r b &= b_{\mathrm{CaSO_4 \cdot 2H_2O}} - b_{\mathrm{CaSO_4}} - 2\,b_{\mathrm{H_2O}(l)}\\ \Delta_r c &= c_{\mathrm{CaSO_4 \cdot 2H_2O}} - c_{\mathrm{CaSO_4}} - 2\,c_{\mathrm{H_2O}(l)}\end{aligned}$$

so the change in C_P between products and reactants is

$$\Delta_r C_P^\circ = \Delta_r a + \Delta_r b\,T - \Delta_r c\,T^{-2} \tag{3.27}$$

Berman–Brown

Another widely used heat capacity equation was suggested by Berman and Brown (1985). They claim it reproduces calorimetric data better than does Maier–Kelley, and also ensures that C_P approaches the high temperature limit predicted by lattice vibration theory. However, the main reason for knowing

Table 3.3 *The entry for α-quartz from the Ge0-Calc database.*

```
A-QUARTZ              SI(1)O(2)                    aQz         AQTZ
ST         -856288. -910700.000        41.460        2.269
C1          80.01199   -240.276 -3546684.000      491568384.    0.0
V1        2.38945698  0.0          -0.24339298    0.00101375 0.0
T1            848.00      373.00  -0.09186959     0.00024607       0.0
T2          0.023743    0.0       0.0         -0.0              0.0
```

about this equation is probably that it is used in the useful Ge0-Calc software (Brown et al., 1988).[9] The equation analogous to Equation (3.27) is

$$\Delta C_P^\circ = \Delta_r k_0 + \Delta_r k_1\,T^{-0.5} + \Delta_r k_2\,T^{-2} + \Delta_r k_3 T^{-3} \tag{3.28}$$

The values for quartz are shown in Table 3.3, which is part of the database for the Ge0-Calc software. The C_P coefficients are listed on the line labeled `C1`. They are also available from Table 3a in Berman (1988), where they are shown as $k_0 = 80.01$, $k_1 \times 10^{-2} = -2.403$, $k_2 \times 10^{-5} = -35.467$, and $k_3 \times 10^{-7} = 49.157$.

The calculated C_P using these numbers is shown in Table 3.2. The agreement is very good at low temperatures, but the difference increases towards the transition point. This undoubtedly reflects a different choice of experimental C_P values to fit, rather than the ability of the equation to fit them. Differences of this type abound throughout databases. Choices are made for you that you may not agree with. You hope that this disagreement will not change the sense of your results.

The measurement of C_P will be considered in Chapter 5.

Shomate

In the 1940s C. Howard Shomate began using an equation for C_P that has subsequently been adopted by the National Institute for Science and Technology (NIST) for their archives of thermochemical data, and available for many substances at the "chemistry webbook" website, http://webbook.nist.gov. The equation is

$$C_P^\circ = A + Bt + Ct^2 + Dt^3 + E/t^2 \tag{3.29}$$

where A, B, C, D and E are constants, and $t = T(\mathrm{K})/1000$, the temperature in kelvins divided by 1000. The webbook site also lists equations for $H_T^\circ - H_{298}^\circ$ and S_T°, which require additional constants. The listing for quartz is shown in Table 3.4.

Equations using these constants are

$$\begin{aligned} H_T^\circ - H_{298}^\circ &= At + Bt^2/2 + Ct^3/3 + Dt^4/4 - E/t + F - H \\ S^\circ &= A\ln t + Bt + Ct^2/2 + Dt^3/3 - E/(2t^2) + G \end{aligned}$$

---

[9] The "Ge0" in this unusual name is intended to signify the thermodynamic quantity G°, not the prefix Geo-, as in Geo-engineering.

Table 3.4 *The NIST data for quartz.*

Temperature (K)	298.0–847.0	847.0–1996
A	−6.076 591	58.753 40
B	251.6755	10.279 25
C	−324.7964	−0.131 384
D	168.5604	0.025 210
E	0.002 548	0.025 601
F	−917.6893	−929.3292
G	−27.969 62	105.8092
H	−910.8568	−910.8568
Reference	Chase, 1998	Chase, 1998
Comment	quartz phase data last reviewed in June, 1967	quartz phase data last reviewed in June, 1967

where $H^\circ_T - H^\circ_{298}$ is in kJ mol^{-1} and S° is in J mol^{-1} K^{-1}. Values for C°_P using equation (3.29) are listed in Table 3.2.

The Maier–Kelley and the Berman–Brown equations are intended for temperatures above 298.15 K, while the Shomate equation is valid down to 0 K. The upper temperature limit for all three equations varies depending on the experimental data available.

3.5.4 Temperature dependence of the enthalpy

Equation (3.25) for our more explicit situation (a chemical reaction) now becomes

$$\left(\frac{\partial \Delta_r H^\circ}{\partial T}\right)_P = \Delta_r C^\circ_P \tag{3.30}$$

where subscript "*r*" has been added to signify that ΔH° refers to a chemical reaction. Using (3.27), (3.30) can now be integrated to give the standard enthalpy of reaction at any elevated temperature T at one bar pressure:

$$\int_{T_\mathrm{r}}^{T} d\Delta_r H^\circ = \int_{T_\mathrm{r}}^{T} \Delta_r C^\circ_P \, dT \tag{3.31}$$

$$\begin{aligned}\Delta_r H^\circ_T - \Delta_r H^\circ_{T_\mathrm{r}} &= \int_{T_\mathrm{r}}^{T} (\Delta_r a + \Delta_r bT - \Delta_r cT^{-2}) dT \\ &= \Delta_r a(T - T_\mathrm{r}) + \frac{\Delta_r b}{2}(T^2 - T_\mathrm{r}^2) + \Delta_r c\left(\frac{1}{T} - \frac{1}{T_\mathrm{r}}\right)\end{aligned} \tag{3.32}$$

where $\Delta_r H^\circ_T$ is the standard enthalpy of reaction at temperature T, and $\Delta_r H^\circ_{T_\mathrm{r}}$ is the standard enthalpy of reaction at the reference temperature, T_r, normally 298.15 K. This is determined by calorimetry, usually by determining $\Delta_f H^\circ$

for each compound in the reaction (§3.5.2). It is readily available for most compounds. The effect of pressure on enthalpy will be considered later.

Using the Berman–Brown formulation, the analogous enthalpy equation is

$$\Delta_r H^\circ_T - \Delta_r H^\circ_{T_\mathrm{r}} = \Delta_r k_0 (T - T_\mathrm{r}) + 2\Delta_r k_1 (T^{\frac{1}{2}} - T_\mathrm{r}^{\frac{1}{2}}) - \Delta_r k_3 \left(\frac{1}{T} - \frac{1}{T_\mathrm{r}}\right) - \frac{\Delta_r k_4}{2}\left(\frac{1}{T^2} - \frac{1}{T_\mathrm{r}^2}\right) \tag{3.33}$$

These k coefficients are of course quite different from the Maier–Kelley coefficients.

These equations are a bit awkward if you are using a calculator, but are simple to program, and values for many reactions can be obtained directly from the software, e.g., SUPCRT92 in the case of the Maier–Kelley coefficients. As an example, consider the reaction

$$\mathrm{CaCO_3}(s) + \mathrm{SiO_2}(s) = \mathrm{CaSiO_3}(s) + \mathrm{CO_2}(g) \tag{3.34}$$

which is a common metamorphic reaction in which wollastonite is formed when carbonate rocks are intruded by granite magma. It is really only of interest at high temperatures. From SUPCRT92, we find the Maier–Kelley coefficients in Table 3.5, so

$$\begin{aligned}\Delta_r a &= a_{\mathrm{CaSiO_3}} + a_{\mathrm{CO_2}} - a_{\mathrm{CaCO_3}} - a_{\mathrm{SiO_2}}\\ &= 26.64 + 10.57 - 24.98 - 11.22\\ &= 1.01\end{aligned}$$

Similarly, $\Delta_r b = -0.00774$, and $\Delta_r c = 32000$.

Inserting these values in Equation (3.32), not forgetting to change the sign of $\Delta_r c$, we find the values of $\Delta_r H^\circ_T - \Delta_r H^\circ_{T_\mathrm{r}}$ shown in Table 3.6. Exactly the same values are produced by SUPCRT92 for this reaction. To obtain the standard heat of reaction at T, $\Delta_r H^\circ_T$, of course you need to know $\Delta_r H^\circ_{T_\mathrm{r}}$, which in turn is generally available from tabulated values of $\Delta_f H^\circ$, as discussed in §3.5.2. In this case, SUPCRT92 reports that $\Delta_r H^\circ_{298} = 22\,561$ cal mol^{-1}, so that $\Delta_r H^\circ$ at 500 °C, for example, is $-1424 + 22\,561 = 21\,137$ cal mol^{-1}.

Table 3.5 *Maier–Kelley coefficients for reaction (3.34).*

Compound	a	b	c
calcite	24.98	0.005 24	−620 000
quartz	11.22	0.0082	−270 000
wollastonite	26.64	0.0036	−652 000
$\mathrm{CO_2}$	10.57	0.0021	−206 000

Table 3.6 $\Delta_r H^\circ_T - \Delta_r H^\circ_{T_r}$ *values for reaction (3.34) from Equation (3.32). The results from* SUPCRT92 *are identical.*

	$\Delta_r H^\circ_T - \Delta_r H^\circ_{T_r}$	
T °C	cal mol^{-1}	J mol^{-1}
25	0	0
100	−98	−408
200	−306	−1280
300	−598	−2502
400	−971	−4063
500	−1424	−5956

Now we know how to determine the standard enthalpy change for reactions at high temperatures, which, as a matter of fact, you may not wish to do very often, unless you get involved in heat flow problems. However, we will see later on that the method is very similar for other important properties, so it is the first of a set of procedures which are very useful.

But there are complications, as usual. If your field area contained wollastonite at a granite contact, and you were really interested in this reaction, you would almost certainly need values above 500 °C. This raises several points:

- You need to perform a different integration above 575 °C, using different coefficients for quartz.
- All these data are for 1 bar pressure. Your field situation undoubtedly requires a higher pressure, which has a big effect on the properties of CO_2.
- You can get $\Delta_r H^\circ$ values at practically any T and P from SUPCRT92 without using Equation (3.32) yourself (SUPCRT92 uses it for you), but you have to realize that you get values of $\Delta_r H^\circ$, the enthalpy change under standard conditions, not the conditions of your intrusive contact. The standard conditions for the minerals are the pure minerals at T and P, which is fine, but the standard conditions for CO_2 are T and 1 bar. In other words, you usually cannot assume that an unmodified $\Delta_r H^\circ$ applies to some field situation.

We are getting a bit ahead of ourselves here. We consider these standard conditions and how to change them in Chapter 8.

3.5.5 Standard and apparent enthalpy of reaction

Hopefully the material up to here is not too confusing. The enthalpy of reaction is the heat that is liberated or absorbed when a reaction takes place at constant pressure, and normally at constant temperature as well. We don't have to measure the enthalpy change for every reaction of interest, because the "formation from the elements" method allows us to tabulate enthalpies of formation for every compound, and combining these gives us (standard) enthalpies of reaction. The heat capacity tells us how enthalpy changes with temperature, and the Maier–Kelley (or Berman–Brown) coefficients allow us to handle this easily. Integrating the relevant expression gives us an equation for the difference in $\Delta_r H°$ at T and at T_r, a reference temperature, so as long as you know $\Delta_r H°$ at T_r you can get $\Delta_r H°$ at T.

But now let's discuss something that *is* confusing. The numbers we calculated in Table 3.6 for the wollastonite reaction are *exactly* the same numbers you get by running SUPCRT92 for this reaction. Now let's try the same comparison for reaction (3.19); that is, we want to know the standard heat of formation of anhydrite from the elements at temperatures above 25 °C, both from Equation (3.32) and from SUPCRT92. The Maier–Kelley coefficients are shown in Table 3.7.

The first problem we run into is that SUPCRT92 does not contain data for the element calcium.[10] In fact it has data for only a few naturally occurring elements, such as silver, gold, sulfur, oxygen, etc. Geochemists are not usually interested in the properties of native calcium, for good reason – it never occurs in nature. However, SUPCRT92 does contain anhydrite, and if you enter the reaction as

```
1 ANHYDRITE
0
```

SUPCRT92 will report something called `STANDARD STATE PROPERTIES OF THE REACTION AT ELEVATED TEMPERATURES`, which will include a `Delta H`. The value for this `Delta H` at 25 °C is definitely the standard enthalpy of formation from the elements, $\Delta_f H°$, which refers to Equation (3.19).

Table 3.7 *Maier–Kelley coefficients for reaction (3.19).*

Compound or element	$a \times 10^0$	$b \times 10^3$	$c \times 10^{-5}$	T limit, K
$CaSO_4$	16.78	23.6	0	1449.0
Ca(α)	5.25	3.44	0	713.0
S(rhombohedral)	3.58	6.24	0	368.6
S(monoclinic)	6.2	0	0	392.0
S(liquid)	8.73	0	0	717.8
$O_2(g)$	7.16	1.0	−0.4	3000.0

---

[10] The coefficients for calcium in Table 3.7 were obtained from Kelley (1960), which is one of the same sources used by SUPCRT92.

But if there are no data for calcium, what reaction do the `Delta H` values at higher temperatures refer to?

The next problem we run into is that sulfur has two transitions in our temperature range, which means we have to do three separate calculations to get to 500 °C. This is a nuisance, but using a spreadsheet we can manage. The fact that we would like to avoid this complication illustrates a point we will make later on.

The results for $\Delta_r H^\circ_T - \Delta_r H^\circ_{T_r}$ from both the equation and SUPCRT92 are shown in Table 3.8 and in Figure 3.10. Clearly, SUPCRT92 is not using Equa-

Table 3.8 *Values of* $\Delta_f H^\circ$ *for anhydrite from Equation (3.32),* SUPCRT92*, and Robie* et al. *(1978).*

	$\Delta_r H^\circ_T - \Delta_r H^\circ_{T_r}$, cal mol^{-1}		
T °C	SUPCRT92	Equation (3.32)	Robie et al. (1978)
25	0	0	0
30	119	−10	
40	360	−28	
50	603	−46	
60	848	−63	
70	1096	−79	
80	1346	−95	
90	1598	−110	
100	1853	−163	
125	2500	−452	
126.85			−625
150	3161	−542	
175	3838	−622	
200	4529	−691	
225	5235	−750	
226.85			−917
250	5956	−799	
275	6692	−837	
300	7442	−864	
325	8207	−880	
326.85			−979
350	8987	−885	
375	9781	−880	
400	10591	−863	
425	11415	−836	
426.85			−834

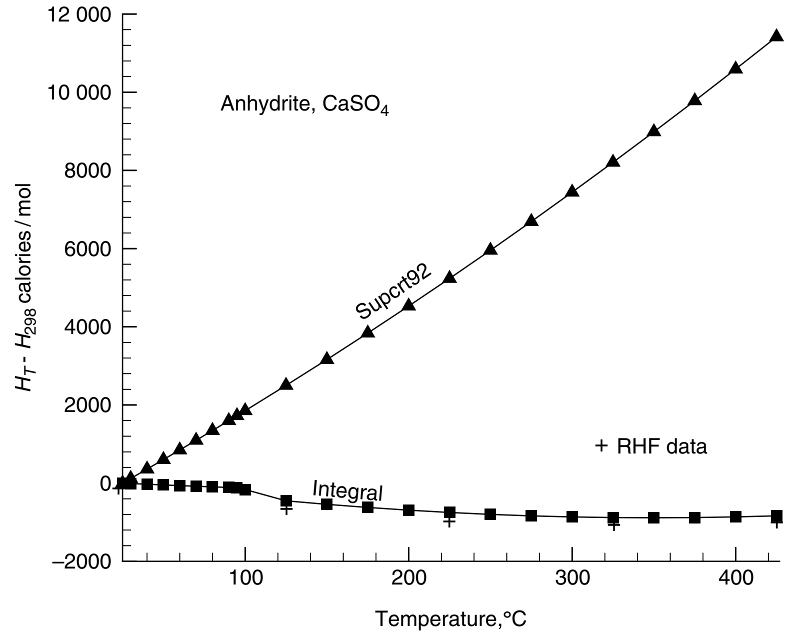


Figure 3.10 Values of $\Delta_r H^\circ_T - \Delta_r H^\circ_{T_r}$ for the mineral anhydrite from SUPCRT92 and from Equation (3.32) (labeled "Integral"). Plus signs are data from Robie, Hemingway and Fisher (1978).

tion (3.32). Not only are the values different, but they tend in opposite directions as T increases.

A clue is provided by plotting the same quantity ($\Delta_r H^\circ_T - \Delta_r H^\circ_{T_r}$) from Robie, Hemingway and Fisher (1978), a well-known compilation of thermodynamic data at elevated temperatures (many chemical sources have data only at 25 °C). We see that they show the same trend as the results from Equation (3.32), with very similar values. They are not exactly the same because the Robie et al. data are not derived from Equation (3.32) or any other equation. They are smoothed experimental data. Nevertheless, we may deduce that they refer to reaction (3.19), as do the results from Equation (3.32) in Table 3.8. So the question is, what do the data from SUPCRT92 mean? What reaction does its `Delta H` refer to?

Another clue is provided by calculating $\Delta_r H^\circ_T - \Delta_r H^\circ_{T_r}$ for anhydrite itself, by substituting anhydrite a, b, and c values for $\Delta_r a$, $\Delta_r b$ and $\Delta_r c$ in Equation (3.32), i.e., by integrating Equation (3.23), resulting in Equation (3.35):

$$\begin{aligned}\int_{T_r}^{T} dH^\circ &= \int_{T_r}^{T} C^\circ_P \, dT \\ H^\circ_T - H^\circ_{T_r} &= \int_{T_r}^{T} (a + bT - cT^{-2}) dT \\ &= a(T - T_r) + \frac{b}{2}(T^2 - T_r^2) + c\left(\frac{1}{T} - \frac{1}{T_r}\right) \qquad (3.35)\end{aligned}$$

The results from this calculation are identical with the SUPCRT92 results, Figure 3.10.

What this means is that in the expression $\Delta H^\circ_T - \Delta H^\circ_{T_r}$, the value of ΔH°_T that SUPCRT92 calculates is *not* the difference in H° between anhydrite and $\sum H^\circ$ of its constituent elements, all at T, but the difference between the H° of anhydrite at T and $\sum H^\circ$ of its constituent elements at T_r, 298.15 K. This called the *apparent* enthalpy of formation, $\Delta_a H^\circ$.[11] The value of $\Delta_r H^\circ_{T_r}$, on the other hand, is what it claims to be, the difference in H° of anhydrite and $\sum H^\circ$ of its constituent elements, all at T_r, which is in fact the heat of formation from the elements, $\Delta_f H^\circ$.

The Benson–Helgeson convention

This way of doing things was started by Benson (1968) for enthalpy, and was adopted by Helgeson in subsequent years for both enthalpy and other energy terms. Thus the Robie et al. (1978) enthalpy and free energy values include provision for the change in the properties of the elements as temperature increases, while SUPCRT92, and any data in publications by Helgeson and the Berkeley group do not. There are distinct advantages in “neglecting” the elements at elevated temperatures, as in the Benson–Helgeson method. You can dispense with finding coefficients for the elements for whatever heat capacity equation you are using, and you don’t have to deal with all the phase transitions in the elements, such as those in sulfur that we complained about above. As long as you deal only with balanced chemical reactions, the elements all cancel out anyway, so it does not matter what properties you assign to them. So if it doesn’t matter, why not just leave them out?

Probably the vast majority of geochemists deal only with balanced reactions in considering geological problems, so they need never worry about whether the “traditional,” Robie et al. method or the Benson–Helgeson method is being used in their databases. But if you search the literature for data on specific compounds, you must make sure you know how data above 25 °C are derived.

The Berman–Brown convention

Berman (1988), in presenting a widely used set of thermodynamic data for metamorphic minerals, not only used a different heat capacity equation (Equation 3.28), but also a somewhat different method for “neglecting” the properties of the elements. The Berman–Brown method is in fact the same as Benson–Helgeson for enthalpy, but different for Gibbs energy (§5.7.1, page 144).

---

[11] The use of *apparent* quantities is now so general that we will not usually bother with the separate notation $\Delta_a H$ or $\Delta_a G$. That is, we will use $\Delta_f H$ and $\Delta_f G$, whatever convention is used.

3.6 How far have we got?

We have defined internal energy as some unspecified subset of the total energy in a system and considered the two common ways of changing this energy content. Along the way, we have noted that energy never disappears, and this is called the first law of thermodynamics. How far have we got toward finding the "chemical energy," that always decreases in spontaneous changes?

Well, we've made the first vital step, but if you think about the previous chapters you'll realize that we cannot have the answer yet. Why? Because we noted that some processes occur spontaneously with *no* energy change (ink spreading in water). Obviously, then, clarifying our thoughts about energy changes will not help in explaining processes that happen with no change in energy of any kind.[12] We still have some way to go toward defining a useful "chemical energy."

3.7 The model again

In this chapter we have discussed some very practical operations. There is nothing particularly theoretical about gases expanding in cylinders and performing work. It happens countless times every day all over the world. Equations such as (3.2) belong to the real world. However, the result of the limit-taking, when the number of expansions or compressions in a single cycle is increased without limit, is a reversible process that belongs not to the real world but to the thermodynamic model. This is another illustration of the point made in §2.6.2, that energy differences between states can be calculated only for *reversible* processes.

The equation

$$\boldsymbol{w}_{\text{rev}} = -\int_{\mathbf{V}_1}^{\mathbf{V}_2} P\, d\mathbf{V} \qquad [3.4]$$

is an extremely simple one, considered mathematically. If P can be expressed as an integrable function of $\mathbf{V}$, then the integration is carried out and $\boldsymbol{w}_{\text{rev}}$ is determined for a given change from $\mathbf{V}_1$ to $\mathbf{V}_2$. This presents absolutely no conceptual difficulties (beyond those in understanding calculus) if P and $\mathbf{V}$ are mathematical variables. However, if P and $\mathbf{V}$ represent measured pressures and volumes from a real system in the real world, then even if P has been determined as an integrable function of $\mathbf{V}$ for a number of individual measurements of P and $\mathbf{V}$, the integration represents a variation of P with $\mathbf{V}$ that is impossible to carry out in the system. It is, however, simple to carry it out in the thermodynamic

---

[12] Actually, we will note in the next chapter that the internal energy U is in fact the energy we need to predict which way reactions will go under certain unusual conditions, but it is rarely used in this sense.

model, that is essentially mathematical and in that P as a function of $\mathbf{V}$ is simply a line in P–$\mathbf{V}$ space. This line represents a reversible process, a perfectly simple and understandable facet of the thermodynamic model.

3.7.1 Applicability of the equations

Don't forget – this conclusion about the work done due to a change in volume is not only applicable to piston–cylinder arrangements. Virtually all chemical reactions involve some change in volume between reactants and products, and the equations are applicable no matter what the physical form of the reactants and products. In other words, when corundum and water react to form gibbsite (Figure 2.2), the gibbsite occupies a different volume than does the sum of the volumes of the water and the corundum; therefore, some work is done during the reaction, and this work can be calculated using Equations (3.3) and (3.5). Even in reactions in living cells there will generally be a difference in volume between products and reactants, and a constant pressure environment, and so some work is done during each and every biochemical reaction. This work energy may be relatively small compared to the heat evolved or absorbed during the same reactions, but it must always be considered. In reactions at higher pressures, it of course becomes even more important.

3.7.2 Clarifying notation

We have introduced quite a few subscripts and superscripts all at once here, which can be confusing. The logical relationships among these terms is shown in Figure 3.11, using H as an example. The same relationships will hold for other parameters we will introduce later.

The most general term for a change in H is simply ΔH. This refers to any change in the enthalpy of any system between two equilibrium states (stable or metastable), not necessarily associated with a chemical reaction. A special case is the ΔH between the products and reactants of a chemical reaction, called $\Delta_r H$, so this represents a subset of the more general term ΔH. A special kind of chemical reaction involves only pure compounds, whose thermodynamic parameters can be found in tables, and so a subset of all $\Delta_r H$ values can be called $\Delta_r H^\circ$, to indicate that all products and reactants are in their pure reference states.[13] A special case of $\Delta_r H^\circ$ is the reaction in which a compound is formed from its elements, all in their pure reference states, and this is called $\Delta_f H^\circ$. Finally, we found that there are two conventions for defining the enthalpy of formation from the elements, one being the traditional or "common sense" method, where the compound and all its elements are at

---

[13] Later on (Chapter 8) we will find that strictly speaking superscript ° refers to a more general "standard state," and that "pure reference states" are just one kind of standard state.

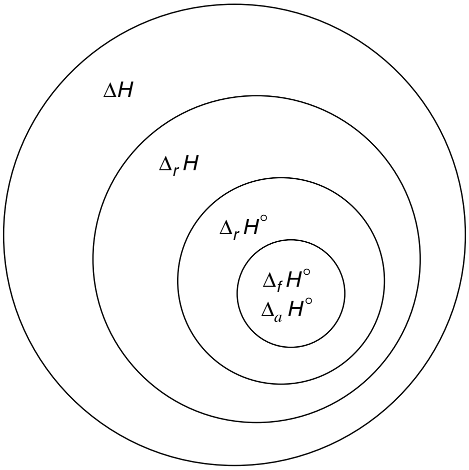


Figure 3.11 The hierarchy of ΔH terms.

the same temperature, and one, called $\Delta_a H^\circ$, in which they are at different temperatures. The difference is only important if you are dealing with the properties of single phases, rather than with balanced chemical reactions, and in most discussions we will use the $\Delta_f H^\circ$ notation to include the $\Delta_a H^\circ$ variety.

3.8 Summary

This chapter attempts to make precise our use of the terms *energy*, *heat*, and *work*. The line of thought we are pursuing has to do with systems that spontaneously decrease their energy content, and so we have started to get quite clear about what kinds of energy we mean. Relativity theory tells us that the *total* energy of all kinds contained in any system is given by multiplying the mass of the system by the square of the speed of light, but this approach is not very useful except in the study of nuclear processes. None of the chemical reactions we are interested in are of this type. However, apart from relativity theory there is no way of knowing the energy content of a system, so we have to be content with knowing *changes* in the energy content.

When we consider by what means the energy content of systems can change, we find that there are only two – we can heat/cool the system, or we can do work on the system/have the system do work. There are several ways of doing work on systems, depending on the forces we choose to consider (magnetic, electrostatic, surface tension, etc.), and so we start out by choosing the most common, pressure–volume work. The others are all handled in the same way and can be brought in when the situation calls for them.

Then by appealing to long experience with energy transfers, we propose the first law of thermodynamics, the law of conservation of energy. Systems (that is, *any* system) can change their energy content by having energy subtracted or added in the two forms – heat and work. Any combination of the two can result in the same total energy change; there is no specific "difference in heat" or "difference in work" between two different states of the same system.

Finally, we went into some detail on one special kind of energy transfer – the heat absorbed or released during a chemical reaction, where the initial and final states have the same pressure, generally referred to as a constant pressure reaction. This quantity of heat is the enthalpy, and it is one of the fundamental building blocks of our model. The fact that there are no absolute values for H is a decided nuisance, but a very simple way around this is by using the "formation from the elements" convention. This means that for every compound, we measure ΔH for the reaction in which a compound is formed from its elements, each in its most stable form, and these quantities are given the symbol $\Delta_f H^\circ$, where the subscript "f" stands for "formation from the elements," and the superscript ° means all substances are in their pure (except for dissolved substances) reference states. The ΔH for any other reaction can then be found by combining these $\Delta_f H^\circ$ terms for compounds.

Defining what we mean by energy and energy transfers is, of course, important, but it does not by itself answer our questions about why reactions go one way and not the other.

4

The second law of thermodynamics

4.1 Introduction

The first sentence of Gibbs' (1875) classic memoir "On the equilibrium of heterogeneous substances" is

> The comprehension of the laws which govern any material system is greatly facilitated by considering the energy and entropy of the system in the various states of which it is capable.

Given the fact that virtually all of equilibrium thermodynamics can be derived by doing exactly that, as Gibbs did, this must rank as one of the world's great understatements. In this chapter, we begin to explore what Gibbs was referring to. By considering "the laws which govern any material system," we should be able to find the answers to the questions we posed in Chapter 1.

4.2 The problem restated

Having taken a couple of chapters to get our terminology settled and to get used to discussing energy changes in systems, we must now get back to our main problem – what determines whether chemical processes will go or not go? Our method of determining this might be considered a bit simple-minded – we will simply determine the "chemical energy" differences between equilibrium states. Processes can take place spontaneously if they are in the direction of lowering the chemical energy. They cannot take place spontaneously in the opposite direction.

We have seen that the first great principle of energy transfers is that energy never disappears; it simply takes on different forms. It is the second principle or law that more directly addresses our main problem. It is observed that once the conditions of the beginning and ending states are decided upon, processes can proceed spontaneously in only one direction between these states and are never observed to proceed in the other direction unless they are "pushed" with an external energy source. Thus for beginning and ending conditions of $P = 1$ bar and $T = 5\,°\mathrm{C}$, ice will melt, but water will never spontaneously change to ice.

We are looking for a "chemical energy" term that will always decrease in such spontaneous reactions and will enable us to systematize and predict what way reactions will proceed under given conditions. This may seem like a simple problem, but it is not.

The greatest single step forward in the development of thermodynamics was the recognition and definition of a parameter, entropy, that enables such predictions and systematizations to be made. And yet, entropy still is not the energy term we have been looking for; the energy that always decreases in spontaneous reactions. In fact, it is not even an energy term. Nevertheless, it is the secret to an understanding of spontaneous reactions.

4.2.1 What's ahead

In this chapter, we will try to explain why this is so. As with the first law, there is no way of proving the second law. It is a principle that is distilled from our experience of how things happen. It can be stated in a lot of different ways, usually having something to do with the impossibility of perpetual motion or with the availability of energy, topics that seem to have little to do with the problem we have set for ourselves – that of finding an energy term that always decreases in spontaneous reactions. We will choose to state it in a way that emphasizes its role as a directionality parameter. This leads to the shortest possible path to the practical applications we wish to consider.

So here's the plan. We will define entropy as a parameter in our model systems, having certain properties. This definition is not, of course, pulled out of the blue, but is based on many years of work by many scientists. It should be accepted at first on faith, as simply a useful parameter, because it will not have any intuitive meaning as do our other terms such as energy, work, and so on. Then we will show how the "chemical energy" term we have been looking for is related to entropy. Finally, we will discuss what entropy is (and what it is not), and in Chapter 5, we will discuss how to measure it.

4.3 Thermodynamic potentials

At the start we should note that "directionality parameters" have a technical name. They are called *thermodynamic potentials*. A potential in this sense is a quantity (a state variable) which is minimized (or maximized) at equilibrium, *subject to certain constraints*. This means that if you want to compare two states of the same system to see which is the more stable, i.e., in which direction the spontaneous change will "go," two state variables must be the same in both states, and we call these two variables the constraints on the system. But not just any two. Which two depends on which state variable you have chosen to be maximized or minimized, i.e., on your choice of thermodynamic potential. Why two, and not one or three? Fundamentally it is because we chose to limit the ways our model systems can exchange energy to two: heat and only one kind of work, and we need one constraint for each of these.

For example, an appropriate thermodynamic potential would have a lower value for calcite than for aragonite at 25 °C, 1 bar (note that in saying 25 °C, 1 bar we have chosen two constraints; one for heat and one for work). Generally, however, we have problems similar in principle but more complex. We may have several phases, including a solution with many compositional variables, and we will want to know the equilibrium compositions of all the phases at T and P. In these cases there is a range of values for the thermodynamic potential, one for every possible composition of the phases involved, and we need to find the *minimum* value of the potential. For any other value, greater than this minimum, some change in compositions (phases will dissolve, precipitate, etc.) will take place until the minimum value is achieved. We then speak of *minimizing* the potential at a given T and P, i.e., subject to the given constraints. This is not a hypothetical problem, but a real problem in applied mathematics. We will see one way of doing this (*speciation*) in Chapter 16.

In this chapter we will identify four thermodynamic potentials. The number could in fact be extended – several other state variables with certain unusual constraints can also be considered thermodynamic potentials. However, they are never used in this sense, so we won't bother with them. As a matter of fact, only one thermodynamic potential is ever used in geochemistry, but understanding is increased by learning about the others.

4.4 Entropy

4.4.1 Analogy

The first and most important thermodynamic potential we need is entropy.[1] One way to define entropy would be to simply say that the Z-term in Equation (3.14) does indeed exist, where entropy is called S, and $Z = -S$. This provides a useful analogy between pressure–volume and temperature–entropy, and we will see these terms linked together in many equations. They represent work and heat energy in many processes we will be considering.

This way of defining entropy is also useful in explaining a somewhat puzzling feature of thermodynamics. In the next section, we will see that although entropy is a state variable of the kind we are looking for (one that can be used to tell which way reactions will go), it is unfortunately one that *increases* in spontaneous reactions, rather than decreasing, as we had supposed. This turns out

---

[1] Strictly speaking, when identifying a thermodynamic potential, it is necessary to specify the constraints. In other words, entropy (**S**) by itself is not a potential; entropy at constant energy and volume ($\mathbf{S}_{\mathrm{U,V}}$) is a potential. In still other words, entropy is not maximized in systems at constant T and P, only in systems at constant **U** and **V**.

Potentials in mechanics

Our term "thermodynamic potential" is so named because it is similar to the "potential" quantities in mechanics. The simplest and most relevant is potential energy. A body has a potential or potential energy because a force is acting on it that is capable of causing the body to move. For a given mass and force, the potential energy is a function only of the position of the body.

Force is a vector quantity, but if we restrict ourselves to one dimension, the functional relationship between the potential energy of a body E_p, its position r, and the force $\mathcal{F}$ is

$$dE_\mathrm{p}/dr = -\mathcal{F} \tag{4.1}$$

Since the potential E_p is a result of the force $\mathcal{F}$ and will decrease if the body is allowed to move, the terms are given opposite signs. The functional form in Equation (4.1) is common to all potential quantities. Any change in the potential will appear as work, either done on the body to increase the potential, or by the body in lowering its potential. Thus

$$\begin{aligned} w &= \int_{r_1}^{r_2} dE_\mathrm{p} \\ &= E_{\mathrm{p},r_2} - E_{\mathrm{p},r_1} \end{aligned}$$

where r_1 and r_2 are two positions of the body, and work is considered negative when done by the body.

We will see that in thermodynamics there are quite a few equations in the form of (4.1), such as

$$(\partial \mathbf{U}/\partial \mathbf{S}))_\mathbf{V} = T, \qquad [4.16]$$

and

$$(\partial \mathbf{U}/\partial \mathbf{V}))_\mathbf{S} = -P \qquad [4.17]$$

and in all these cases, the numerator (in this case **U**) is a thermodynamic potential, the denominator (in this case **S** or **V**) is a configuration term analogous to distance, and the right hand side (in this case T or $-P$) is analogous to a force.

to be simply because entropy was historically defined as $-\mathbf{Z}$ in Equation (3.9), rather than as **Z**. In other words, Equations (3.9) and (3.14)

$$-\boldsymbol{q} \le \int_{\mathbf{Z}_1}^{\mathbf{Z}_2} T\, d\mathbf{Z} \qquad [3.9]$$

$$-q \le \int_{Z_1}^{Z_2} T\, dZ \qquad [3.14]$$

are actually written

$$\boldsymbol{q} \leq \int_{\mathbf{S}_1}^{\mathbf{S}_2} T\,d\mathbf{S} \tag{4.2}$$

and

$$q \leq \int_{S_1}^{S_2} T\,dS \tag{4.3}$$

that is, without the minus sign. This is a sort of "historical accident." Because of the complete analogy between work–pressure–volume and heat–temperature–entropy (except for the minus sign), we can also write

$$\boldsymbol{q}_{\text{rev}} = T\,\Delta\mathbf{S} \tag{4.4}$$

$$q_{\text{rev}} = T\,\Delta S \tag{4.5}$$

analogous to (3.6) and (3.12),

$$\boldsymbol{w}_{\text{rev}} = -P\Delta\mathbf{V} \qquad [3.6]$$

$$w_{\text{rev}} = -P\Delta V \qquad [3.12]$$

A somewhat more mathematical way to express this analogy between work–pressure–volume and heat–temperature–entropy, and which is discussed in more detail in §C.2.1, is to note that although w_{rev} (or δw_{rev} if you prefer that notation) is not the differential of any function (not an exact differential) w_{rev}/P is, being equal to $-\Delta V$, or $-dV$ in differential notation. So one might suppose that similarly although q_{rev} is not exact, perhaps q_{rev}/T is. And indeed it is, being equal to dS.

4.4.2 Definition

However, a better way to define entropy is as follows. If there is indeed "something missing," that is, only one thing missing from the energy-decreasing analogy, it is something that causes reactions to "go," *even when no energy change whatsoever occurs*. Now, we have defined a type of system (the *isolated* system) that does not permit energy changes to occur in the system. Therefore, all we have to do is define a parameter with which we can predict reaction directions in this kind of system, combine it with the energy-decreasing idea, and we should have our answer. This is what we do with the following definition, paraphrased after Callen (1960). It can also serve as a statement of the second law of thermodynamics. The statement reflects Callen's "postulational" approach to thermodynamics, that is, let's postulate that there is a parameter that includes everything we need, and see how it works out:

> There exists an extensive property of systems, entropy (**S**), which for isolated systems achieves a maximum when the system is at stable equilibrium. Entropy is a smoothly varying function of the other state variables and is an increasing function of the internal energy **U**.

Note that using an isolated system automatically ensures that we will compare states having the same values of two state variables, **U** and **V**, our two constraints. We insert the postulate that entropy increases with **U** to ensure that the other directionality parameters to be derived decrease (have minima) rather than increase. This can be shown by considering the isolated (model) system in Figure 4.1.

4.4.3 The U–S–V/(V_1+V_2) surface

The exterior wall in Figure 4.1 is impermeable to energy and rigid, so the system is of constant **U** and **V**.[2] The piston is movable and can be locked in any position. It is impermeable but it conducts heat so that the two sides are at the same temperature. If there are equal amounts of the same gas in the two compartments, the equilibrium position of the piston when it is free to move is where $\mathbf{V}_1 = \mathbf{V}_2$. Also, according to our definition of **S**, the equilibrium position of the piston is one of maximum entropy for the system, and any other position has lower entropy.

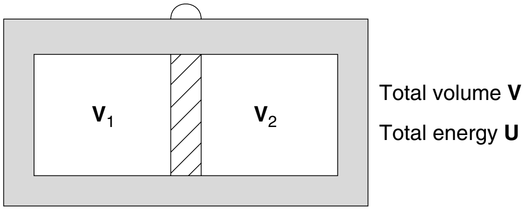


Figure 4.1 An isolated system having a movable partition. The partition is impermeable to matter but conducts heat. The volume **V** of the system is the sum of the volumes of the two subsystems, that is, $\mathbf{V} = \mathbf{V}_1 + \mathbf{V}_2$.

---

[2] The discussion and diagrams in this section use the total quantities **U**, **S**, and **V**. We could equally well use the molar properties U, S, and V. As mentioned in §2.4.1, most equations remain true whether using total or molar properties. However, some don't, and those that don't are very important, e.g., (2.1); (4.66). In the case of total and molar work and heat ($\boldsymbol{w}$ and w; $\boldsymbol{q}$ and q), the molar form usually only makes sense when considering chemical reactions, but this is the only case considered in much of geochemistry.

In the equilibrium position, the piston is not locked in place, i.e., there is no constraint other than **U** and **V**, no third constraint. In any other position the piston must be locked in place, because the pressure on one side is greater than the pressure on the other side.[3] Nevertheless, such locked positions are unchanging, and are equilibrium states. We want to distinguish between these states having an extra constraint, and those equilibrium states that have only two constraints, so we call the three-constraint states *metastable equilibrium states*.

Then if we consider the same situation but with successively greater energy contents $\mathbf{U}'$, $\mathbf{U}''$, and $\mathbf{U}'''$ (which might be achieved by increasing the gas temperature), we will have entropy–volume curves as in Figure 4.2, where **S** is plotted against $\mathbf{V}_1/(\mathbf{V}_1+\mathbf{V}_2)$, which varies between 0 and 1. The maximum value of **S** is at $\mathbf{V}_1/(\mathbf{V}_1+\mathbf{V}_2)=0.5$, where $\mathbf{V}_1=\mathbf{V}_2$, and the curves for $\mathbf{U}'$, $\mathbf{U}''$, and $\mathbf{U}'''$ are arranged with increasing entropies because we defined **S** to be an increasing function of **U**. In Figure 4.3, the curves of Figure 4.2 are drawn in three dimensions, and in Figure 4.4 the complete surface is shaded with a number of contours – the horizontal ones being contours of constant **S** and **V**, the vertical ones contours of constant **U** and **V** (recall that the whole diagram is for conditions of constant **V**).[4] In Figure 4.5 two

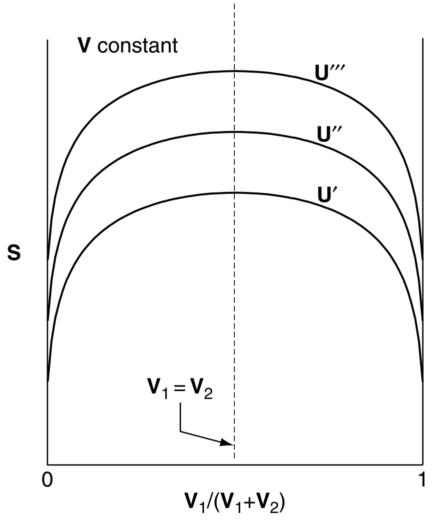


Figure 4.2 Entropy (**S**) versus volume fraction $\mathbf{V}_1/(\mathbf{V}_1+\mathbf{V}_2)$ for the system in Figure 4.1 at three different energy levels, where $\mathbf{U}'''>\mathbf{U}''>\mathbf{U}'$. Volume **V** is constant.

---

[3] If you find yourself wondering how one could lock and unlock the piston if the system is truly isolated, you have not yet fully grasped the fact that thermodynamics deals with mathematical models, not real things. In the model, the position of the piston is, or could be, a mathematical variable.

[4] Don't confuse this surface, $\mathbf{U}$–$\mathbf{S}$–$\mathbf{V}/(\mathbf{V}_1+\mathbf{V}_2)$, which includes metastable equilibrium states, with the *USV* (or **USV**) surface (§4.6), which has only stable equilibrium states. This is important.

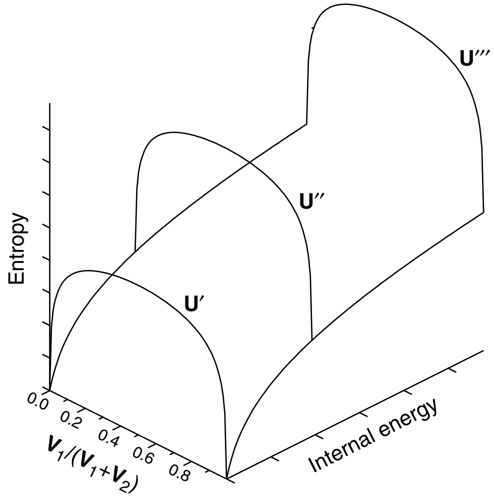


Figure 4.3 The **U**–**S**–**V**/(**V**$_1$ + **V**$_2$) curves of Figure 4.2 in three dimensions.

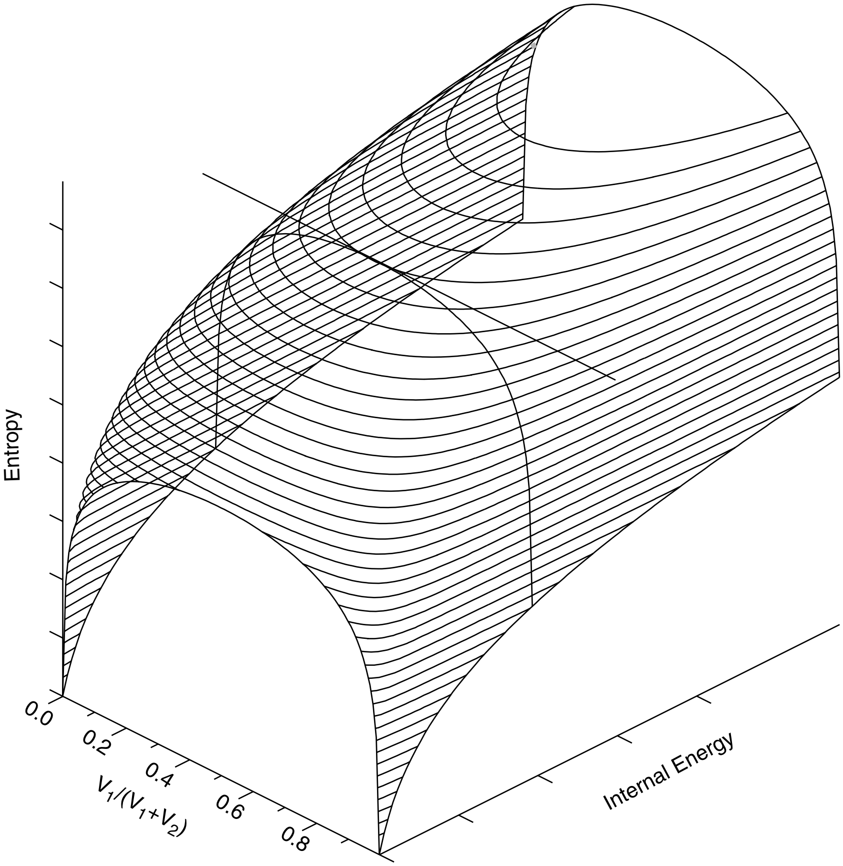


Figure 4.4 The **U**–**S**–**V**/(**V**$_1$ + **V**$_2$) surface, calculated for an ideal gas. All points on the surface other than those at maximum **S**, minimum **U** represent metastable states of the system.

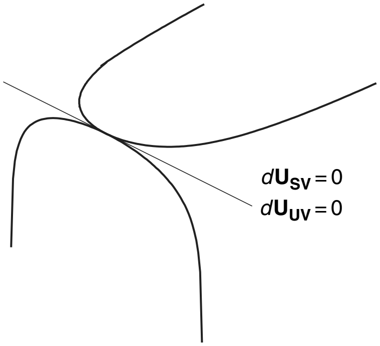


Figure 4.5 Constant **S**, **V**, and constant **U**, **V** sections from the **U**–**S**–**V**/($\mathbf{V}_1 + \mathbf{V}_2$) surface shown in Figure 4.4, with their common tangent, which is simultaneously $d\mathbf{S}_{\mathbf{U},\mathbf{V}}$ and $d\mathbf{U}_{\mathbf{S},\mathbf{V}}$. The tangent point represents a position of stable equilibrium for the system.

of these contours are abstracted to show more clearly that the two contours, which meet at a point, have a common tangent.

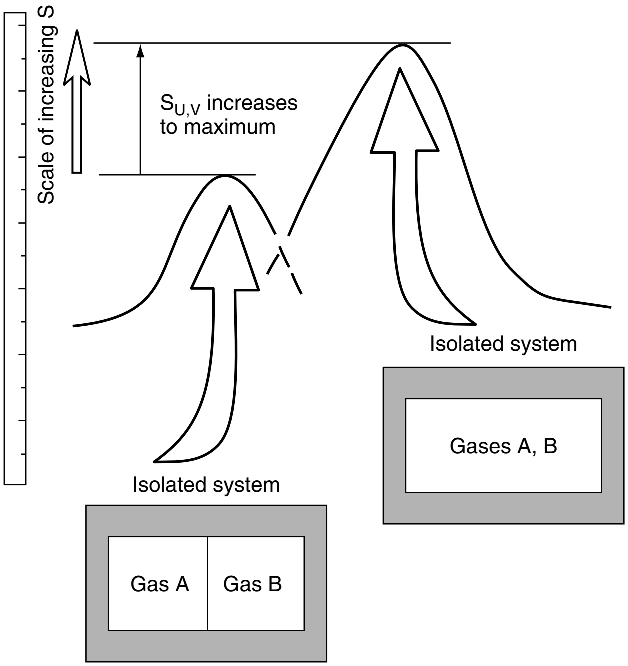
This tangent is located at the extremum in both curves, and so is in mathematical terms both $d\mathbf{U}_{\mathbf{S},\mathbf{V}} = 0$ and $d\mathbf{S}_{\mathbf{U},\mathbf{V}} = 0$. In other words, the condition that at equilibrium **S** is a maximum for given **U**, **V** *implies* the condition that **U** is a minimum for given **S**, **V** at a given equilibrium point.

Not only is **S** a parameter that always *increases* when a metastable state changes to a stable state at given values of **U** and **V**, but **U** is a parameter (a *state variable*) that always *decreases* when a metastable state changes to a stable state at given values of **S** and **V**. Thus as long as we consider only systems at constant **U** and **V** or constant **S** and **V**, both $\mathbf{S}_{\mathbf{U},\mathbf{V}}$ and $\mathbf{U}_{\mathbf{S},\mathbf{V}}$ are parameters of the type we have been looking for – they are thermodynamic potentials. These kinds of systems are very rare; still, we're getting closer.

4.4.4 Mixing example

Figure 4.6 shows another system having three constraints. As before, the system is isolated, so **U** and **V** are two constraints. The third constraint is a partition separating two gases. A spontaneous process (mixing) occurs when the partition is removed, and according to our definitions, entropy therefore increases. The importance of the isolated system is that it *prevents* energy changes from taking place in systems undergoing reactions, and it is therefore the clue to the missing factor we have mentioned several times (e.g., §1.3.1). We knew there was something missing because some spontaneous processes (like melting ice) take place while absorbing heat energy, and some (like mixing gases) take place with little or no energy change at all. If reactions can take place with *no* energy change, and if entropy is the directionality parameter that predicts which way reactions will go when there is no energy change, then perhaps *combining* entropy with our other parameters such as heat and work will lead to more useful directionality parameters. This is exactly the case.

Figure 4.6 Entropy increases in spontaneous processes in isolated systems.



In other words, the reason we were a little off-base with suggesting that there is an analogy between a ball rolling in a valley and spontaneous chemical reactions is that some reactions can happen with no drop in energy at all. Chemical systems are more complex than simple mechanical systems, and analogies are dangerous. Entropy is a state variable that always increases in spontaneous reactions in which there is no energy change (those which take place in isolated systems), and is the missing factor. Combined with other state variables, we will have directionality parameters (our "chemical energy" term) for all kinds of systems.

Despite the apparent usefulness of entropy, we have not yet discussed what entropy actually *is*. We have no intuitive feeling for it such as we have for energy, pressure, and the other terms we have used so far. We must for the moment just use it as a parameter that we have defined in such a way as to be useful, and figure out what it *means* later.

4.5 The fundamental equation

The first and second laws can be combined into a single equation, which lies fairly close to the very heart of thermodynamics, called the Fundamental Equation.

Combining Equation (3.1) ($\Delta\mathbf{U} = \boldsymbol{q} + \boldsymbol{w}$) with Equations (3.6) ($\boldsymbol{w}_{\text{rev}} = -P\,\Delta\mathbf{V}$) and (4.4) ($\boldsymbol{q}_{\text{rev}} = T\,\Delta\mathbf{S}$) (or Equations 3.12, 4.5 and 3.10) we obtain

$$\Delta\mathbf{U} = T\,\Delta\mathbf{S} - P\,\Delta\mathbf{V} \tag{4.6}$$

and

$$\Delta U = T\,\Delta S - P\,\Delta V \tag{4.7}$$

These can also be written in differential notation as

$$\boxed{d\mathbf{U} = T\,d\mathbf{S} - P\,d\mathbf{V}} \tag{4.8}$$

and

$$\boxed{dU = T\,dS - P\,dV} \tag{4.9}$$

This is probably the single most important equation in thermodynamics, and for this reason it is called the *Fundamental Equation*.

It is worth mentioning that the “fundamental” nature of (4.9) does not mean that we often use it directly. We usually use it after a little manipulation, so that we can use integration limits of T and P, rather than S and V. In other words, we usually use Equation (4.37). Equation (4.9) is fundamental because it is directly linked to the first and second laws, and because Equation (4.37) and many others are derived from it.

4.6 The USV surface

4.6.1 Geometrical meaning of the fundamental equation

The easiest way to get a clear geometric picture of the fundamental equation is to realize that by virtue of the first law, every system at equilibrium, whether as simple as an ideal gas or as complex as a bacterium, has a single fixed energy content for given values of two independent variables, and by virtue of the second law we are able to use as variables S and V. Every system can therefore be represented by a surface in USV space, such as shown in Figure 4.7. Note that, in order to make the surface easier to draw, U increases downward in Figure 4.7. At every point on the surface, such as point A, there will be a tangent surface, the equation for which is the fundamental equation, $dU = T\,dS - P\,dV$. Figure 4.8 shows how, starting at point A, increments of dS and dV are combined with the slopes $\partial U/\partial S$ and $\partial U/\partial V$ to produce the total change in U at any other point on the tangent plane. In this case, each of dU, dS, and dV has any magnitude, however large. Because of this, you may visualize Equation (4.9) as having the form $Z = aX + bY$, which is a

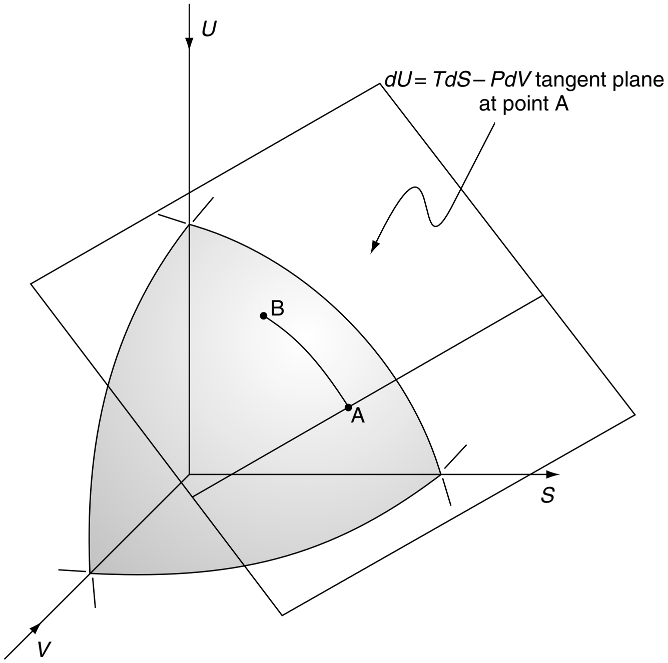


Figure 4.7 Every system has a unique *USV* surface. The fundamental equation (4.9) represents a tangent to this surface when *dS* and *dV* are of arbitrary magnitude, and it can be integrated to give the change in *U*, ΔU, between any two points on the surface, such as A and B. The tangent plane is illustrated further in Figure 4.8.

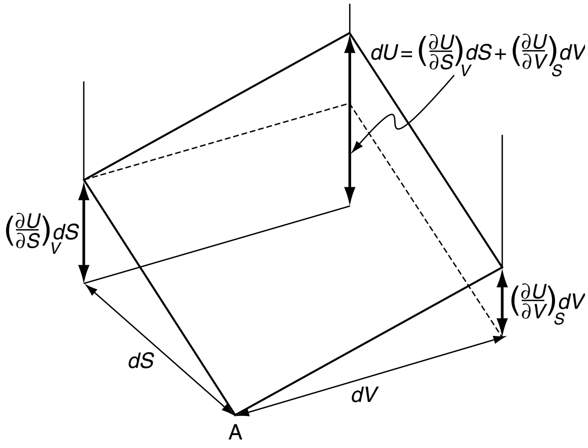


Figure 4.8 The tangent surface at point A in Figure 4.7, showing how *dU* is geometrically related to *dS* and *dV*.

combination of the straight-line equations $Z = aX$ and $Z = bY$, and is the equation of a plane in *XYZ* space.[5]

On the other hand, the fundamental equation can also be used to calculate values of ΔU between any two points on the *USV* surface itself, such as between points A and B. Because *U* follows some complex function of *S* and *V* between A and B, the fundamental equation must be *integrated* between

---

[5] *Z* here has no connection with the *Z* in Equation (3.14).

A and B, and this means allowing dS and dV to take on infinitesimal values[6] and performing a summation, symbolized by the $\int$ symbol. This is written

$$\begin{aligned}\Delta U &= U_\mathrm{B} - U_\mathrm{A}\\ &= \int_\mathrm{A}^\mathrm{B} T\,dS - \int_\mathrm{A}^\mathrm{B} P\,dV\end{aligned}$$

The *calculation* of this difference follows the reversible path shown on the *USV* surface – a continuous succession of equilibrium states, because that's what integration implies. However, the calculated ΔU is the same no matter how the change from A to B is actually carried out.

The difference in **U** or in fact any thermodynamic property between a quartz crystal at 25 °C and the same crystal at 50 °C has nothing to do with how that difference is achieved, but the calculation of that difference follows a reversible path. That just means we calculate the difference by integrating an equation. If you have difficulty understanding differential equations as used here, study Figures 4.7 and 4.8.

4.7 Those other forms of work

Even though we hardly ever need to include other kinds of work in our fundamental equation, it will prove useful to have a brief look at what this equation looks like when they are included. Because many work terms don't make much sense in the molar form, we use the total energy form of the fundamental Equation (4.8). A very general formulation, including all forms of work, is

$$d\mathbf{U} = \underbrace{T\,d\mathbf{S}}_{\text{heat}} - \underbrace{\textstyle\sum_i X_i\,dx_i}_{\text{all forms of work}} \tag{4.10}$$

where X is a generalized force and dx_i a generalized displacement. Normally, we consider X_i to be $-P$ and x_i to be $\mathbf{V}$ or V, and we get the fundamental equation. We considered a few other possibilities for the $\sum_i X_i\,dx_i$ term in §3.4.1. So, for example, if a system could exchange energy because of gravity acting on a mass m in addition to P–$\mathbf{V}$ work, Equation (4.9) would become

$$d\mathbf{U} = \underbrace{T\,d\mathbf{S}}_{\text{heat}} - \underbrace{P\,d\mathbf{V} + mg\,dh}_{\text{work}} \tag{4.11}$$

For the tensile force and surface tension forms of work we would have

$$d\mathbf{U} = \underbrace{T\,d\mathbf{S}}_{\text{heat}} - \underbrace{P\,d\mathbf{V} + f\,dl}_{\text{work}} \tag{4.12}$$

---

[6] See Appendix C, page 589. for a note on infinitesimals and integration.

and

$$d\mathbf{U} = \underbrace{T\,d\mathbf{S}}_{\text{heat}} - \underbrace{P\,d\mathbf{V} + \gamma\,dA}_{\text{work}} \tag{4.13}$$

The chemical work term that we mentioned in §3.4.1 is $-\mathcal{A}\,d\xi$, where $\mathcal{A}$ is the affinity and ξ is the progress variable, resulting in

$$d\mathbf{U} = \underbrace{T\,d\mathbf{S}}_{\text{heat}} - \underbrace{\underbrace{P\,d\mathbf{V}}_{PV\text{ work}} - \underbrace{\mathcal{A}\,d\xi}_{\text{chemical work}}}_{\text{work}} \tag{4.14}$$

and, of course we could have any or all combinations, at least in principle. These equations should be fairly intuitive – $d\mathbf{U}$ represents the change in total energy of a system, and the three terms on the right are the three sources of energy change – heat, P–$\mathbf{V}$ work, and another kind of work. Because we are so interested in chemical reactions, we have a great deal more to say about Equation (4.14).

4.8 Applicability of the fundamental equation

For such a simple relationship, Equation (4.9) traditionally generates quite a bit of confusion. This is of two types, or perhaps two aspects of the same problem. That problem is reversibility versus irreversibility.

Equations (4.7) and (4.9)

$$\Delta U = T\,\Delta S - P\,\Delta V \qquad [4.7]$$

$$dU = T\,dS - P\,dV \qquad [4.9]$$

appear to be limited to reversible processes, as both (3.12) and (4.5)

$$w_{\text{rev}} = -P\Delta V \qquad [3.12]$$

$$q_{\text{rev}} = T\,\Delta S \qquad [4.5]$$

refer to reversible processes. But this is not the case, which can be explained in various ways.

1. For one thing, Equations (4.7) and (4.9) contain only state variables (U, S, and V, in addition to T and P). Therefore, because the changes in state variables do not depend on the nature of the change, Equation (4.7) (or the integration of Equation 4.9) is true for any change between two equilibrium states which have the same composition, with one important exception (see §4.8.1).
2. Although Equations (3.13) and (3.14) will both be untrue in an irreversible process, the amount by which they become untrue (resulting in $-w < P\Delta V$ and $q < T\Delta S$) will cancel when they are added together. We could give examples of this, but it must be true because of item 1.

3. In Figure 4.7, points A and B represent two stable equilibrium states of a system. Integration of the fundamental equation from A→B takes place on the *USV* surface, as shown, which is necessarily a reversible process. An irreversible process A→B would cause the system to leave the surface at A, be not representable in *USV* space between A and B, then the system would reappear at B. In either case, ΔU, ΔS, and ΔV are related by Equation (4.7).

4.8.1 Two kinds of irreversibility

So the Equation (4.9) applies to reversible and irreversible processes. There remains the "important exception" just mentioned. The problem here is that there are two very different kinds of irreversible processes.[7]

1. One kind of irreversibility involves a system changing from one equilibrium state to another. The equilibrium state can be stable or metastable (in the thermodynamic sense, discussed below), as long as there is no change from one to the other (i.e., metastable→stable). For example Equation (4.9) can be applied to increments of changes such as the irreversible gas expansion in Figure 3.4, or perhaps the irreversible heating of a crystal from 25 to 50 °C by simply putting it in an oven at 50 °C. The crystal could be (stable) calcite or (metastable) aragonite. Integrating Equation (4.9) applies to all these processes. They could be thought of as processes in response to a change in the first and/or second constraints.
2. But Equation (4.9) does *not* apply to irreversible processes in which a metastable state changes to a stable state, such as a crystal of aragonite at some T and P within the calcite stability field recrystallizing to calcite. Equations (4.14) and (4.46) apply to these cases. In these cases a second form of work (non-PV work) is being done.

In one sense, it is perfectly obvious that if there is more than one form of work involved, you need one of the equations in §4.7, not Equation (4.9). In geochemistry, and most of chemistry, the only other form of work of interest is chemical work, and so the equation we use to represent a chemical reaction progressing towards stable equilibrium is (4.14), or more usually (looking ahead) (4.46).

The most general way of expressing the applicability of the fundamental equation is that it applies to any process which does not involve release of a third constraint.

4.9 Constraints and metastable states

We introduced the idea of constraints and metastable states in §4.4.3. Because of the importance of these concepts, we take a closer look at them here.

---

[7] We mentioned these briefly in §2.6, page 21.

4.9.1 Metastable states

It is here that distinguishing between real systems and our model systems becomes most valuable.

Real metastable states

The distinction between stable and metastable equilibrium according to most sources is that the stable equilibrium state is "truly unchanging," or unchanging given indefinite time, whereas the metastable state may be changing, but too slowly to be observed. This distinction clearly refers to real systems, and is often very difficult to make. We know that at 25 °C, 1 bar, calcite is the most stable form of $CaCO_3$. Aragonite is another form, and although it does not change to calcite on museum shelves, it does change in nature under some conditions, given very long time periods, so is it metastable or unstable? Is volcanic glass an unstable or a metastable phase? These questions can start arguments among geochemists. There are many reactions for which kinetic rate constants are known at high temperatures (the reacting assemblage is therefore unstable), but not at 25 °C, where the assemblage is considered metastable. At what temperature does metastable change to unstable? If anyone was interested, the answer would of course be completely arbitrary. The kinetics of very slowly changing systems is a problem for real systems, but it is not a problem in thermodynamics.

Thermodynamic metastable states

In thermodynamics, a metastable equilibrium state has at least three constraints. Two of these constraints apply to a stable equilibrium state, and the third prevents the system from achieving that state.[8] On releasing the third constraint the system experiences a spontaneous process and achieves the stable equilibrium state. We have seen two examples so far, in Figures 4.1 and 4.6. These examples were chosen to follow from our definition of entropy, and show spontaneous processes having no overall energy change in the system. They show entropy acting as a thermodynamic potential.

However, processes happening in isolated systems are pretty rare, in fact strictly speaking they are nonexistent. We need to develop other, more useful thermodynamic potentials (§4.12), to deal with the usual type of metastable→stable processes such as aragonite→calcite which invariably involve a decrease in system energy. But even without doing that, we can see several other examples of thermodynamic metastable states by looking

---

[8] Actually, to be quite accurate, we should put the previous sentence in mathematical terms. It would then read *Two of these constraints represent the independent variables of a function simulating the energy content of a physical system, and the third represents a possible additional independent variable.* However, that has the disadvantage of obscuring the physical meaning of the constraints.

at the equations in §4.7. In each equation we see the fundamental equation, $d\mathbf{U} = T\,d\mathbf{S} - P\,d\mathbf{V}$. Taking only one example, Equation (4.12), applying force f to stretch a wire obviously does work on the wire (adds energy to the system), and releasing the force allows the wire to do work and resume its stable equilibrium state. While stretched, the system is in a metastable state. Similar statements apply to the other equations. In each case, doing work on the system puts it in a higher energy state (**U** is increased) and is held there by the third constraint. Releasing the constraint allows the system to do work in a spontaneous process. We give a more precise definition of constraints in §4.9.2. These relationships are illustrated in Figure 4.9a for the system Al_2O_3–H_2O, where the arrow $A \rightarrow A'$ refers to a "chemical work" term, as in Equation (4.14).

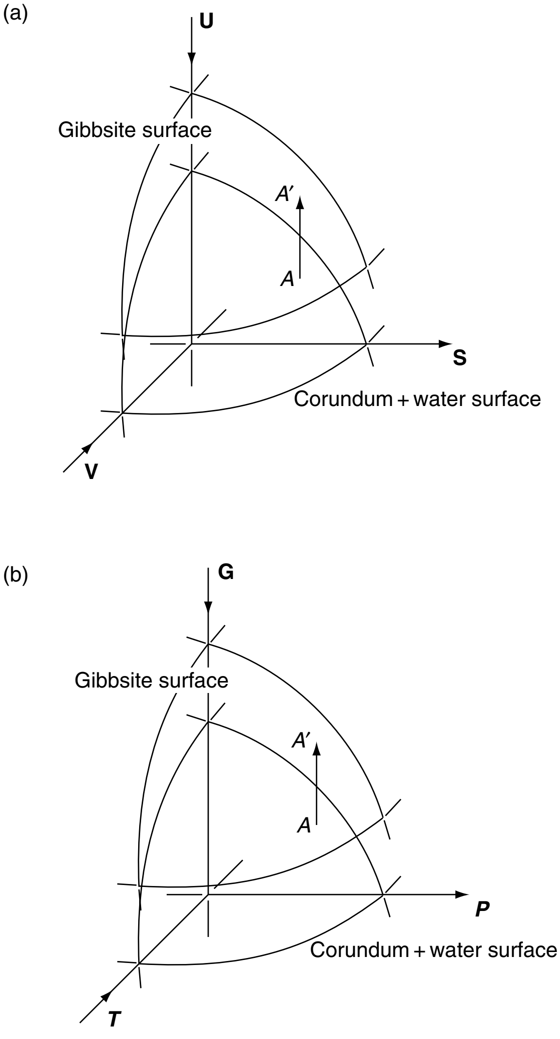


Figure 4.9
(a) Two equilibrium states of the system Al_2O_3–H_2O in **USV** space: metastable corundum plus water, and stable gibbsite. The process $A \rightarrow A'$ is reaction (2.2). Changes in **U** along either surface can be described by Equation (4.9). Changes in **U** between the surfaces ($A \rightarrow A'$) are described by Equation (4.14). (b) The same system in **G***TP* space. Note that **U** and **G** increase downwards.

Real systems may be truly unchanging and metastable, or they may be unstable and changing very, very slowly. Often we don't know which. But our models of systems have no such uncertainty. Metastable systems are modeled as in complete thermodynamic equilibrium, with (at least) three constraints. If the real system is really unstable, our model is in this respect incorrect, but generally useful nonetheless.

4.9.2 Constraints

Constraint, like several other terms, has a slightly different meaning in real and model systems. Real systems, such as aragonite, are said to be constrained from reacting to a more stable state (calcite) by an activation energy barrier. The usage is rather imprecise, because we don't actually know if the aragonite is changing on some very long time scale or not. In model systems, the meaning is much more exact. A constraint in mathematics is a condition that must be observed. For example you might want to find the minimum value of a function of many variables, while assigning constant values to some of the variables. Each variable held constant is a constraint on the solution. In thermodynamics it is essentially the same thing. We minimize some function of many variables while constraining two state variables to constant (known) values. We constrain two, because in the first law we define only two ways of changing the energy of a system, heat and P–V work. If there is a third way of changing energy, a third state variable must be constrained – a third constraint. The system must be in some equilibrium state for that variable, or any variable, to be defined. The only thermodynamic equilibrium state that is not a stable equilibrium state is a metastable equilibrium state. A thermodynamic constraint, then, is defined as *a state variable, associated with some method of changing system energy, that is held constant while minimizing a thermodynamic potential*. Stable equilibrium

Constraints in mechanics

Constraints, along with potentials, work, and energy, is another topic common to thermodynamics and mechanics. The motion of bodies in mechanics is subject not only to the force applied, but also to whatever constraints are present. A marble rolling on the surface of a bowl is constrained to remain on the surface. Beads on a string are constrained to move in a straight line, and so on.

Constraints in mechanics can be classified into various types, for example as to whether the equation of constraint contains time as a variable or not. In thermodynamics, which has only scalar variables, and which has no time variable, constraints are simpler, and are identified with ways in which systems can change their energy content.

requires two constraints. Any extra constraints result in equilibrium states with greater energy contents we call metastable.[9]

4.10 The energy inequality expression

Stable equilibrium states are the target destination for all metastable states, once their constraints are released. What equations refer to this process? We try here to make these important relationships fairly intuitive.

Equation (4.8) can also be written in the form of a total differential as

$$d\mathbf{U} = \left(\frac{\partial \mathbf{U}}{\partial \mathbf{S}}\right)_{\mathbf{V}} d\mathbf{S} + \left(\frac{\partial \mathbf{U}}{\partial \mathbf{V}}\right)_{\mathbf{S}} d\mathbf{V} \tag{4.15}$$

showing that, comparing with (4.8),

$$\left(\frac{\partial \mathbf{U}}{\partial \mathbf{S}}\right)_{\mathbf{V}} = T \tag{4.16}$$

and

$$\left(\frac{\partial \mathbf{U}}{\partial \mathbf{V}}\right)_{\mathbf{S}} = -P \tag{4.17}$$

which are the slopes of the **USV** surface in the **V** and **S** directions.

The **USV** surface defined by this equation has an energy **U** defined for given values of the two constraint variables **S** and **V**. It refers to changes in energy between stable equilibrium states of a system (which may take place reversibly or irreversibly). At given values of **S** and **V**, a metastable state will have greater values of **U** than those represented by this surface, and it will be held in that state by a third constraint of some kind. Examples of equations having a third constraint are shown in §4.7.

Consider the wire-stretching example (Equation 4.12) again. If we impose a third constraint, i.e., we do non-$P\mathbf{V}$ work on the system like stretching the wire, we increase the value of **U** (dl and therefore $f\,dl$ is positive).[10] When the constraint is released, the wire shrinks back to its former length, and the system lowers its **U** irreversibly to the stable value.[11] During this process, the shrinking

---

[9] Clearly, isolated systems are an exception to this last statement (i.e., the statement that applying a third constraint results in a state with greater energy). We applied a constraint in an isolated system in two cases, illustrated in Figures 4.1 and 4.6, resulting in metastable states, but with no increase in system energy. In all simulations of real processes, however, the statement is true.

[10] At this point you may well wonder why the PV work term has a negative sign, while all the other work terms are positive. In PV work adding energy decreases the volume, so dV is negative, but in all the other cases, increasing energy increases the displacement or differential term. For example in stretching a wire, dl is positive. Therefore only for PV work do we use the negative sign to make $P\,dV$ positive for a positive energy change.

[11] You might also be wondering, what if I release the constraint reversibly? Can't be done. That's like supposing a ball can roll downhill reversibly.

wire can do useful work. Therefore for any irreversible process (meaning any process in which the third constraint is released[12]) **U** decreases, the third term on the right side becomes negative (e.g., dl is negative), and if we don't include this negative term, we can only write

$$d\mathbf{U} < T\,d\mathbf{S} - P\,d\mathbf{V} \tag{4.18}$$

Combining (4.18) and (4.9), a general fundamental equation for all states of the system is

$$d\mathbf{U} \le T\,d\mathbf{S} - P\,d\mathbf{V} \tag{4.19}$$

If **S** and **V** are constant, $d\mathbf{S} = 0$ and $d\mathbf{V} = 0$, and it follows from (4.19) that

$$d\mathbf{U}_{\mathbf{S},\mathbf{V}} \le 0 \tag{4.20}$$

which is analogous to (4.28), and shows that the second law provides another criterion for spontaneous processes, this time that **U** is minimized at equilibrium in systems having constant values of **S** and **V**. It is another thermodynamic potential. We saw this previously in graphical form in Figure 4.5.

Let's just look at Equation (4.20) a little more closely. Breaking it into separate statements, it says

$$d\mathbf{U}_{\mathbf{S},\mathbf{V}} < 0 \quad \text{for irreversible processes,} \tag{4.21}$$

$$d\mathbf{U}_{\mathbf{S},\mathbf{V}} = 0 \quad \text{at equilibrium} \tag{4.22}$$

These equations actually *imply* the existence of a third constraint. They imply the existence of a function **U** with independent variables **S** and **V** and a third independent variable, because if **S** and **V** are constant, the system cannot change its energy by heat or PV work. **U** can only change using a second work term involving increments of a third constraint, and that change is always negative for irreversible processes (Equation 4.21). Equation (4.22) says that **U** is at a minimum (or maximum) of some continuous function. That function can only be **U** as a function of the third constraint, because **S** and **V** cannot change, so **U** cannot show a minimum with respect to either variable. It can only show a minimum with respect to changes in some third variable,[13] which can only be a second work term.[14]

---

[12] It is very hard to avoid intermixing real-life and mathematical terminology. "Releasing" a constraint is real-life terminology, and gives an intuitive feeling for the situation. Actually, releasing or imposing a constraint simply means changing the sign of the differential term in the second work term in Equations (4.11)–(4.14). If it is positive we impose the constraint, and **U** increases. If it is negative we release the constraint, and **U** decreases.

[13] Note that Figures 4.4 and 4.5 show this, i.e., the locked piston is the third constraint.

[14] This situation has not of course completely escaped the attention of other authors. It is put in a different way by Tisza, page 622.

We considered the wire-stretching example because it is so simple. The same discussion could be held using chemical work, such as the charging and discharging of a battery. In such cases the third constraint is the progress variable ξ. Integrating $\mathcal{A}\,d\xi$ gives the energy difference between the stable equilibrium state and any other (metastable) equilibrium state, and when $\mathcal{A}=0$, $\mathcal{A}\,d\xi=0$, and **U** has its minimum value. However, such examples are more complex because both $\mathcal{A}$ and ξ require quite a bit of explanation, which we come to in Chapter 18.

4.11 Entropy and heat capacity

So far, all we know about entropy is that it increases in spontaneous reactions in isolated systems, and that it appears in equations such as (4.55) and (4.56). Hidden in the equations we have derived so far is an important relationship between entropy and heat capacity, which we will see in Chapter 5 serves as a basis for the measurement of entropy.

Combining Equations (3.17) and (4.9),

$$dH = dU + P\,dV + V\,dP \qquad [3.17]$$

$$dU = T\,dS - P\,dV \qquad [4.9]$$

we find[15]

$$dH = T\,dS + V\,dP \qquad (4.23)$$

If we choose constant pressure conditions, dP becomes zero, so

$$dS = \frac{dH}{T} \qquad (4.24)$$

and substituting $C_P\,dT$ for dH (Equation 3.23), we have

$$C_P\,dT = T\,dS$$

or

$$dS = \frac{C_P}{T}dT \qquad (4.25)$$

Here we have the entropy defined in terms of something measurable, the heat capacity. Integrating (4.25), we have

$$S_{T_2} - S_{T_1} = \int_{T_1}^{T_2} \frac{C_P}{T}dT \qquad (4.26)$$

and so you see that assuming that you can get numbers for C_P at a series of temperatures, you could divide each C_P by its value of T and evaluate the integral, giving you the difference in entropy between two temperatures.

---

[15] Equation (4.23) happens to be another fundamental equation, one that is never used in this sense. It shows that enthalpy is a thermodynamic potential which is minimized for processes at constant (S, P).

4.11.1 Illustrating entropy changes

Our statement of the second law (page 70) implicitly includes the following relationships (4.27)–(4.30):

$$\Delta \mathbf{S}_{\mathbf{U,V}} > 0 \quad \text{for spontaneous processes} \tag{4.27}$$

or, switching to molar units and differential notation,

$$dS_{U,V} > 0 \quad \text{for spontaneous processes} \tag{4.28}$$

and, at the maximum value of S,

$$dS_{U,V} = 0 \quad \text{equilibrium} \tag{4.29}$$

or, combining (4.28) and (4.29),

$$dS_{U,V} \geq 0 \quad \text{for any constant } U,V \text{ process} \tag{4.30}$$

In addition, we have the relation between C_P and entropy, Equation (4.25). It refers to reversible processes. If in the derivation we used the more general Equation (4.19), we would find the more general relationships

$$dS \geq \frac{dH}{T} \tag{4.31}$$

$$dS \geq \frac{C_P}{T} dT \tag{4.32}$$

and

$$S_{T_2} - S_{T_1} \geq \int_{T_1}^{T_2} \frac{C_P}{T} dT \tag{4.33}$$

where the inequality (>) of course refers to irreversible processes.

The normal reaction for students reaching this point is to have not much idea what all these equations really mean, if anything. Before going on to discuss entropy in other terms, we can illustrate what they mean in terms of some simple measurements.

Example 1 Take Equation (4.24). It says that, for example, if you melt ice reversibly at 0 °C,[16] the ΔS will equal the heat of fusion, ΔH, divided by the

---

[16] How do you melt ice reversibly? I thought we said reversible processes were impossible. Well, they are, but phase transitions are a special case. Because the two phases can coexist at equilibrium, the change in thermodynamic properties such as ΔS and ΔH at that temperature between the two phases will be the same as the change that would be observed in a reversible change from one phase to the other, so phase transitions such as this are often said to be "reversible." Nevertheless, you cannot melt ice or freeze water at exactly 0 °C.

temperature, 273.15 K. The heat of fusion of ice is +6008 J mol^{-1} (positive, because we must *add* heat to the system), so

$$\begin{aligned}\Delta S_{\text{ice}\rightarrow\text{water}} &= S_{\text{water}} - S_{\text{ice}}\\ &= \frac{6008}{273.15}\\ &= 21.995\,\text{J mol}^{-1}\end{aligned}$$

Now the enthalpy change for the melting of ice changes very little from 273.15 to 274 K, in fact hardly at all, so we can assume it is still +6008 J mol^{-1}. So if you melt ice irreversibly, say at 274 K, then ΔS will be *greater* than the enthalpy change divided by the temperature, so

$$\begin{aligned}\Delta S_{\text{ice}\rightarrow\text{water}} &= S_{\text{water}} - S_{\text{ice}}\\ &> \frac{6008}{274}\\ &> 21.927\,\text{J mol}^{-1}\end{aligned}$$

Similarly, if you freeze water at 272 K, the enthalpy change is now −6008 kJ mol^{-1}, and

$$\begin{aligned}\Delta S_{\text{water}\rightarrow\text{ice}} &= S_{\text{ice}} - S_{\text{water}}\\ &> \frac{-6008}{272}\\ &> -22.088\,\text{J mol}^{-1}\end{aligned}$$

which means that

$$\Delta S_{\text{ice}\rightarrow\text{water}} < +22.088\,\text{J mol}^{-1}$$

In other words,

$$21.927 < \Delta S_{\text{ice}\rightarrow\text{water}} < 22.088\,\text{J mol}^{-1}$$

so we have determined the entropy change between ice and water at equilibrium at 273.15 K to within 1% with two irreversible measurements.

Example 2 As another example, consider the problem of determining the change in entropy of a substance X between 300 K and 350 K at one bar. We will suppose that the heat capacity of X is constant at exactly 10 J mol^{-1} so that 500 J are required to heat one mole of X from 300 to 350 K, 250 J to heat it from 300 to 325 K, and so on. It follows that on heating one mole of X from 300 to 350 K in a thermostat at 350 K,

$$S_{350} - S_{300} > \frac{500}{350}$$

and on cooling back to 300 K in a similar thermostat at 300 K

$$S_{300} - S_{350} > \frac{-500}{300}$$

from which we conclude that

$$\frac{500}{300} > S_{350} - S_{300} > \frac{500}{350}$$

or

$$1.67 > S_{350} - S_{300} > 1.43\,\mathrm{J\,mol^{-1}}$$

which means we have determined the ΔS to about one part in six with two measurements. To improve our precision, we could double the number of measurements, and heat and cool in two stages each. That is, because

$$S_{350} - S_{300} = (S_{350} - S_{325}) + (S_{325} - S_{300})$$

then

$$\frac{250}{325} + \frac{250}{300} > S_{350} - S_{300} > \frac{250}{350} + \frac{250}{325}$$

or

$$1.60 > S_{350} - S_{300} > 1.48\,\mathrm{J\,mol^{-1}}$$

Given enough patience, we could make 50 measurements at one degree intervals, in which case

$$\frac{10}{349} + \frac{10}{348} + \cdots + \frac{10}{301} + \frac{10}{300} > S_{350} - S_{300} > \frac{10}{301} + \frac{10}{302} + \cdots + \frac{10}{349} + \frac{10}{350}$$

or

$$1.543\,890 > S_{350} - S_{300} > 1.539\,128\,\mathrm{J\,mol^{-1}}$$

Clearly we are approximating an integral, which is of course Equation (4.26), which in this case becomes

$$\begin{aligned}\Delta S &= S_{350} - S_{300}\\ &= \int_{300}^{350} \frac{C_P}{T}\,dT\\ &= \int_{300}^{350} C_P\,d\ln T\\ &= 10 \times \ln(350/300)\\ &= 1.541\,507\,\mathrm{J\,mol^{-1}}\end{aligned}$$

This type of calculation is similar to the one we did for work using the piston–cylinder arrangement, in the sense that we approach the reversible process by taking more and more steps. We reiterate that the physical operation implied by the integral $\int (C_P/T)\,dT$ is a reversible process, which is impossible. This doesn't bother us, however, because the integration involves surfaces in the thermodynamic model, not physical reality. It is helpful though to see, as above, what sequence of physically real measurements could lead to the same result.

4.11.2 *T–S* diagrams

We mentioned a couple of times the complete analogy between *w–P–V* and *q–T–S* relationships. Thus isotherms on a *P–V* diagram (Figure 4.10) are quite analogous to isobars on a *T–S* diagram (Figure 4.11).

The isobars in Figure 4.11[17] have a positive slope because $\partial S/\partial T$, or C_P, is always positive, and the slope increases with T because C_P does.

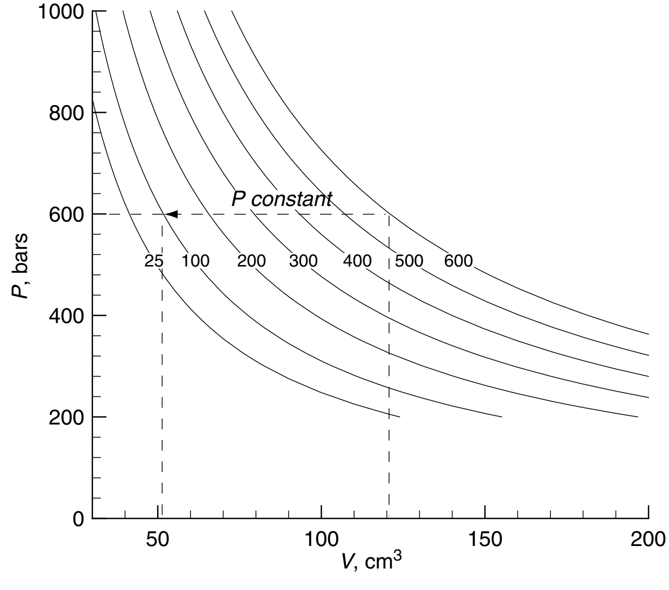


Figure 4.10 Ideal gas isotherms. The work done by a constant pressure reversible compression from 100 °C to 600 °C at 600 bars is the area under the *P constant* arrow.

---

[17] The entropy data in this diagram are from program STEAM, v.2.2, developed by Harvey et al. (2000). This program, in common with many other sources, reports values of entropy which are not third law entropies (Chapter 5), but the difference in entropy between the state of interest and the T and P of the triple point of water. Program SUPCRT92 on the other hand, reports third law entropies, that is, the entropy of water using zero as the entropy of perfectly crystalline ice at 0 K. The third law entropy of water at the triple point is 63.304 J mol^{-1} K^{-1} or 15.130 cal mol^{-1} K^{-1}, so this must be added to many tabulated values to get third law entropies. More detail on steam tables is presented in §13.6.1.

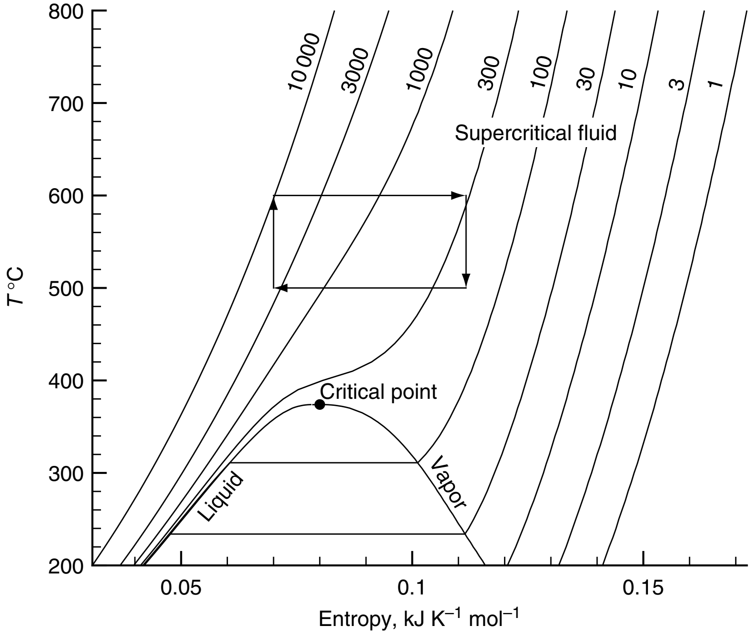


Figure 4.11 The entropy of water. The contours are labeled in bars. The arrows outline a Carnot cycle. Data from program STEAM (Harvey et al. 2000).

We won't go through all the details as we did with work and P–V diagrams, but you can see from equation (4.3) that the area under an isobar in Figure 4.11 between two points on the isobar will equal the reversible heat between those states, just as the area under an isotherm in Figure 3.8 (or Figure 4.10) is the reversible work. The only purpose in showing this *TS* diagram is to complete the analogy with *PV* diagrams. However, it is worth noting that this particular type of diagram is very useful in illustrating Carnot cycles, which we will not discuss in any detail. A Carnot cycle consists of an isothermal expansion (shown at 600 °C), then an adiabatic (S constant, $q = 0$) expansion, then an isothermal compression (shown at 500 °C), then an adiabatic compression. Each expansion and compression is reversible (continuous succession of equilibrium states), and the fact that it is a cycle, ending up at the starting point, means that the total energy change is zero. And by the first law, if $\Delta U = 0$, then q and w must be equal and opposite, or $q = -w$. Because the heat transfer in each step of the cycle equals the area under the curve (under the arrow in this case), it follows that the area enclosed by the rectangle is the total heat input during the cycle.

Because steam engines and internal combustion engines operate in cycles of compression and expansion of some working substance, you can readily imagine that this idealized cycle is of great interest to engineers. It represents the ideal, or maximum work that can be attained by any engine.

4.12 A more useful thermodynamic potential

Now we have two parameters, $S_{U,V}$ and $U_{S,V}$, that will tell us which way processes will go, but they refer to processes which virtually never occur, except perhaps in classroom exercises – that is, processes which occur at constant values of U and V, or of S and V, our two constraints. We need a parameter which will refer to processes at constant T and P, our most common case.[18]

4.12.1 Gibbs energy

From Equation (4.19) (in the molar form),

$$dU - T\,dS + P\,dV \leq 0 \tag{4.34}$$

we can see that if we define a function

$$G = U - TS + PV \tag{4.35}$$

called the Gibbs energy or the Gibbs free energy, the differential of which is

$$dG = dU - T\,dS - S\,dT + P\,dV + V\,dP \tag{4.36}$$

or

$$dG_{T,P} = dU - T\,dS + P\,dV \tag{4.37}$$

Reversible versus irreversible compression

The reversible compression shown in Figure 4.10 looks very much like the irreversible compression shown in Figure 3.9, but it is quite different. In Figure 3.9 the initial and final states are at the same temperature. The compression is irreversible, and any amount of work and heat exchange can occur, depending on the weight on the piston. No isotherms can be shown, because this would imply that we knew the temperature at all points on the compression arrow, but we don't. In Figure 4.10 the initial and final temperatures are different, and during the compression the temperature is indicated by the isotherms. It is an example of Equation (3.6).

---

[18] It is usual to speak of processes occurring at constant U and V, or constant T and P. It would be more accurate to speak of processes having the same values of U and V, or of T and P, before and after the process. It doesn't really matter what the system does between the two states; that is, the system need not be at constant T and P during the process.

which, combined with (3.16), is

$$dG_{T,P} = dH - T\,dS \tag{4.38}$$

then we find by comparing (4.37) and (4.34) that

$$dG_{T,P} \leq 0 \tag{4.39}$$

Thus our definition of the second law has led to a function, G, which will always decrease to a minimum in spontaneous processes in systems having specified values of T and P. It is an extremely useful thermodynamic potential. All we have to do is to find a way to get measurable values of this function for all pure compounds and solutes, and to find how they change with T, P, and concentration, and we will then be able to predict the equilibrium configuration of any system by minimizing G.

To see how G changes with T and P is fairly simple. To see how it changes with composition is a little more difficult (see §§7.6; 8.2). Combining (4.9) and (4.36), we find another fundamental equation,[19]

$$dG = -S\,dT + V\,dP \tag{4.40}$$

Because (4.40) could also be written as a total differential,

$$dG = \left(\frac{\partial G}{\partial T}\right)_P dT + \left(\frac{\partial G}{\partial P}\right)_T dP \tag{4.41}$$

we see from (4.40) that

$$(\partial G/\partial T)_P = -S \tag{4.42}$$

and

$$(\partial G/\partial P)_T = V \tag{4.43}$$

which are the slopes of the G–T–P surface in the T and P directions (Figure 4.9b). We will see how to integrate these expressions shortly.

In addition,

$$dG_{T,P} = 0 \tag{4.44}$$

Equation (4.44) is simply the condition for a minimum in G, that is, the tangent is horizontal (see Figure 4.14). It follows too, that $\Delta G_{T,P} < 0$ for spontaneous processes in systems having the same T and P before and after the process,

---

[19] Equation (4.40) is derived in a more mathematically elegant way in §C.4. It is more elegant because it shows in a simple way how all thermodynamic potentials are intimately related.

because any spontaneous process must head toward this minimum from some higher point (some point of greater G value).[20]

The total form of Equation (4.36) is

$$d\mathbf{G} = d\mathbf{U} - T\,d\mathbf{S} - \mathbf{S}\,dT + P\,d\mathbf{V} + \mathbf{V}\,dP \tag{4.45}$$

and combining this with Equation (4.14), we get

$$d\mathbf{G} = -\mathbf{S}\,dT + \mathbf{V}\,dP - \mathcal{A}\,d\xi \tag{4.46}$$

which is analogous to (4.14), and so can be written

$$d\mathbf{G} = \underbrace{-\mathbf{S}\,dT}_{\text{heat}} + \underbrace{\underbrace{\mathbf{V}\,dP}_{PV\text{ work}} \quad - \quad \underbrace{\mathcal{A}\,d\xi}_{\text{chemical work}}}_{\text{work}} \tag{4.47}$$

Gibbs energy in chemical reactions

Let's review what we know about Equation (4.46). The $d\mathbf{G} = -\mathbf{S}\,dT + \mathbf{V}\,dP$ part describes an equilibrium surface in $\mathbf{G}TP$ space for a particular system. Recalling the discussion in §4.8.1, $d\mathbf{G} = -\mathbf{S}\,dT + \mathbf{V}\,dP$ describes *either* the stable equilibrium surface *or* a metastable equilibrium surface. Its just that the functional relationship between **S** and T and between **V** and P will be different in the two cases. In other words, you can apply $d\mathbf{G} = -\mathbf{S}\,dT + \mathbf{V}\,dP$ to either calcite or aragonite.

When the third constraint is added, as in Equation (4.46), the $d\mathbf{G} = -\mathbf{S}\,dT + \mathbf{V}\,dP$ part of the equation normally refers to the stable equilibrium surface. The $\mathcal{A}\,d\xi$ part then represents chemical work, such as the work you would obtain by discharging the battery (or changing aragonite to calcite), and so represents the difference in **G** between the two surfaces.[21]

The equation thus opens the way to considering chemical reactions trying to achieve stable equilibrium, not just the equilibrium states themselves, and would therefore appear to be extremely important. Well, it is, but perhaps not as much as you might think. In Chapter 18 we will show how to use the quantity $\mathcal{A}\,d\xi$ in considering chemical reactions, but in fact calculating the value of $\mathcal{A}$ is not usually required. Using the progress variable ξ itself is quite common, but although implicitly linked to $\mathcal{A}$, this link is not necessarily made explicit in calculations.

---

[20] Note that the mathematical conditions for a function minimum, such as $d\mathbf{U}_{\mathbf{S},\mathbf{V}} = 0$ and $d\mathbf{G}_{T,P} = 0$, are all true simultaneously at the minimum, as illustrated in Figure 4.5, and they imply nothing whatsoever about how the function (i.e., the system) reached or achieved that minimum.

[21] It is conceivable that the "$d\mathbf{G} = -\mathbf{S}\,dT + \mathbf{V}\,dP$ part of the equation" could refer to a metastable state or surface, in which case the "$\mathcal{A}\,d\xi$ part" would represent the work of charging the battery or changing calcite to aragonite, but in fact it is never used in this sense.

4.12.2 Helmholtz energy

Similarly, if we define a function

$$A = U - TS \tag{4.48}$$

called the Helmholtz energy, or the Helmholtz work function, the differential of which is

$$dA = dU - T\,dS - S\,dT \tag{4.49}$$

or

$$dA_T = dU - T\,dS \tag{4.50}$$

we find, comparing (4.34) and (4.50),

$$dA_T + P\,dV \leq 0 \tag{4.51}$$

or

$$dA_{T,V} \leq 0 \tag{4.52}$$

Thus we have another function, A, which will always decrease to a minimum in spontaneous processes in systems having specified values of T and V, and is another thermodynamic potential. The usefulness of this function will be discussed below.

4.13 Gibbs and Helmholtz functions as work

4.13.1 Gibbs energy as useful work

A ball in a metastable equilibrium valley (e.g., Figure 1.2) is capable of doing work as it rolls down to lower elevations (once it is pushed over the barrier). The maximum work it can do is exactly equal to the (minimum) work required to push the ball back up to its metastable elevation. One way of understanding the Gibbs energy is that it is equal to the maximum amount of useful work that chemical systems can do as they change from metastable states to stable states, underlining the usefulness of the ball-in-valley analogy.

However, we must first distinguish between *total* work and *useful* work. Chemical systems undergoing change (i.e., in which reactions occur) can do various kinds of work. For instance, batteries can do electrical work. While undergoing these reactions, the chemical system invariably has some change in volume, because it is most unlikely that the reaction products would have exactly the same volume as the reactants. This change in volume ΔV takes place under some ambient pressure P, so that $P\,\Delta V$ work is done during the reaction regardless of whether any other kind of work is done or not – if the reaction is to take place, it cannot be avoided. This “work against the atmosphere” (or against the confining pressure, whatever it is) usually is not *useful*; it simply takes place whether we like it or not, and at atmospheric pressure it is often a rather small part of the total energy change. Although we can decide to eliminate electrical work or other kinds of mechanical work from our systems, we cannot eliminate this $P\,\Delta V$ work (unless we consider only constant volume systems, which is not usually very practical). We can talk about the total work $\boldsymbol{w}$, or the total work per mole, w. Net work (per mole) other than $P\,\Delta V$ work can be written

$$\begin{aligned} w_{\text{net}} &= (w_{\text{total}} - w_{P\,\Delta V}) \\ &= (w + P\,\Delta V) \end{aligned}$$

and because $q \leq T\,\Delta S$ (integration of Equation 4.3), it follows from the first law ($\Delta U = q + w$) that

$$w \geq \Delta U - T\,\Delta S \tag{4.53}$$

Adding $P\,\Delta V$ to both sides,

$$w + P\,\Delta V \geq \Delta U - T\,\Delta S + P\,\Delta V$$

and so

$$w_{\text{net}} \geq \Delta U - T\,\Delta S + P\,\Delta V \tag{4.54}$$

Now,

$$\begin{aligned} G &= U - TS + PV \\ dG &= dU - T\,dS - S\,dT + P\,dV - V\,dP \\ dG_{T,P} &= dU - T\,dS + P\,dV \end{aligned} \tag{4.55}$$

and

$$\Delta G_{T,P} = \Delta U - T\,\Delta S + P\,\Delta V \tag{4.56}$$

So, combining (4.54) and (4.56),

$$w_{\text{net}} \geq \Delta G_{T,P} \tag{4.57}$$

or, the other way around,

$$\Delta G_{T,P} \leq w_{\text{net}}$$

Gibbs energy as maximum work

The maximum amount of work, other than the work done by the confining pressure, available from reaction (2.2) is

$$\begin{aligned}\Delta_r G^\circ &= \Delta_f G^\circ_{\mathrm{Al_2O_3\cdot 3H_2O}(s)} - \Delta_f G^\circ_{\mathrm{Al_2O_3}(s)} - 3\Delta_f G^\circ_{\mathrm{H_2O}(l)} \\ &= -2310.21 - (-1582.3) - 3(-237.129) \\ &= -16.523\,\mathrm{kJ\,mol^{-1}} \\ &= -16\,523\,\mathrm{J\,mol^{-1}}\end{aligned}$$

By comparison, $P\Delta V$ work done by atmospheric pressure during this reaction is $1.61\,\mathrm{J\,mol^{-1}}$.

However, don't forget that according to our sign convention, $-w$ is the work done by a system, or available from a system, so we should perhaps write

$$-\Delta G_{T,P} \geq -w_{\text{net}}$$

In other words, the net useful work available from a system cannot be greater than the decrease in G that the system undergoes (that's why it is often called Gibbs *free* energy). For example, if a battery is doing work by lighting the bulb in a flashlight, the maximum amount of useful work it can do is given by its decrease in G toward stable equilibrium, when the battery is dead (we continue this thought in §12.5.1). If the system does *no* work other than expanding or contracting against its confining pressure (no work other than $P\Delta V$ work), then $w_{\text{net}} = 0$, and

$$\Delta G_{T,P} \leq 0 \tag{4.58}$$

This result is not surprising, as it agrees with our conclusion in §4.10, but it does serve to link the Gibbs energy with an intuitive concept, the available work.

4.13.2 Helmholtz energy as total work

Comparing Equations (4.53) and (4.50), we see that

$$w \geq \Delta A_T \tag{4.59}$$

which is analogous to (4.57), only in this case we say that the *total* work per mole (rather than the *available* work) cannot be greater than the decrease in A, or that ΔA_T is an upper limit to the total work done in isothermal processes.

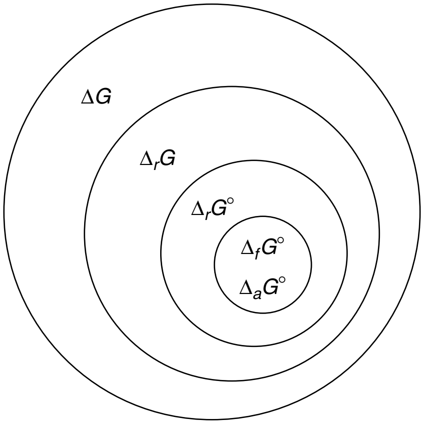


Figure 4.12 The hierarchy of ΔG terms.

If no work at all is done, then $w = 0$, which implies that $\Delta V = 0$, and

$$\Delta A_{T,V} \leq 0 \tag{4.60}$$

so that A always decreases in spontaneous processes under constant T, V conditions. It is another thermodynamic potential.

It might seem useful to have a function related to the total work available from a system, but in fact A is little used in this sense. It (that is, $A_{T,V}$) is also not much used as a thermodynamic potential, despite the fact that replacement processes in weathering, metamorphism and metasomatism are commonly interpreted to occur at constant volume (Nahon and Merino 1997, Carmichael 1986). However, replacement processes do not, by definition, take place in a closed system, so that the Helmholtz energy is not the appropriate potential. So what is the appropriate potential quantity in open systems? We consider this in §4.14.

Where the Helmholtz energy frequently *is* used is in constructing *equations of state* (Chapter 13), and that is the reason for including it here.

4.13.3 Notation again

As a reminder, Figure 4.12 shows the logical relationship between all our various ΔG terms, just as Figure 3.11 did for ΔH terms. Refer to §3.7.2 for a discussion.

4.14 Open systems

4.14.1 The open system equation

If you look at Figure 3.3 and Figure 4.1, or if you think about the fundamental equation (4.9), you realize that we have expended all this time and effort in defining thermodynamic potentials that are limited in one very important respect, and that is that they apply only to systems which do not change composition. Everything we have said is limited to constant composition systems.

Such systems are indeed important, and they can include changes in composition in a limited sense. For example we can use **G** to predict that in a system having 1 gram of halite plus a kilogram of water, the water will change composition as the halite dissolves to an equilibrium composition. But the solution must have the same composition as the halite + water that we started with. The system has not changed composition; components have just been redistributed between phases within the system, as in Figure 2.2b.

The total Gibbs energy **G** depends on the mass of system we consider. The **G** of 2 kg of halite is twice the **G** of 1 kg of halite (§2.4.1). So if the total energy of a system depends on the mass or number of moles there are in the system, surely it also depends on what that mass consists of. Because we deal only with energy changes, we have no way of knowing whether a mole of halite has more or less energy then a mole of sylvite (KCl) or a mole of anything else, and we don't need to know. We just need to know how the energy of a system *changes* if we add halite or sylvite to it. We know that if we add halite to a system consisting of halite, the change is linear, but how does it change if we add sylvite to halite (in solid solution)?

Because we have a mathematical model of energy changes, we have a simple mathematical answer to this question. **G** is a state variable, so $d\mathbf{G}$ is an exact differential (§C.2.1). This means that, among other things, we can write the total differential as in Equation (4.41), or in the total energy form as

$$d\mathbf{G} = \left(\frac{\partial \mathbf{G}}{\partial T}\right)_P dT + \left(\frac{\partial \mathbf{G}}{\partial P}\right)_T dP \tag{4.61}$$

This shows, as long as we can integrate these derivatives, how **G** changes with changes in T and P. To see how it changes when we add n_1 moles of component 1 and n_2 moles of component 2, we just add more derivatives, so

$$d\mathbf{G} = \left(\frac{\partial \mathbf{G}}{\partial T}\right)_{P,n} dT + \left(\frac{\partial \mathbf{G}}{\partial P}\right)_{T,n} dP + \left(\frac{\partial \mathbf{G}}{\partial n_1}\right)_{T,P,n_2} dn_1 + \left(\frac{\partial \mathbf{G}}{\partial n_2}\right)_{T,P,n_1} dn_2 \tag{4.62}$$

where n means (n_1, n_2), or in general, with c components,

$$d\mathbf{G} = \left(\frac{\partial \mathbf{G}}{\partial T}\right)_{P,n} dT + \left(\frac{\partial \mathbf{G}}{\partial P}\right)_{T,n} dP + \sum_{i=1}^{c} \left(\frac{\partial \mathbf{G}}{\partial n_i}\right)_{T,P,\hat{n}_i} dn_i \tag{4.63}$$

$$= -\mathbf{S}dT + \mathbf{V}dP + \sum_{i=1}^{c} \left(\frac{\partial \mathbf{G}}{\partial n_i}\right)_{T,P,\hat{n}_i} dn_i \tag{4.64}$$

where n now means $n_1, n_2, \ldots, n_c$ (all components), n_i refers to any individual component i, and $\hat{n}_i$ refers to all components *except* i.

Our new derivative terms $(\partial\mathbf{G}/\partial n_i)_{T,P,\hat{n}_i}$ look suspiciously like our definition of a partial molar property in Equation (2.1), and indeed they are partial molar Gibbs energies. They allow us to deal with compositional changes, and as such they are one of the most important quantities in chemical thermodynamics. They are given a name, chemical potentials, and a symbol, μ, i.e.,

$$\mu_i = \left(\frac{\partial \mathbf{G}}{\partial n_i}\right)_{T,P,\hat{n}_i}$$

We have a lot more to say about them, but for the moment we will just get the equations we need for future reference. We can now write (4.64) as

$$d\mathbf{G} = -\mathbf{S}dT + \mathbf{V}dP + \sum_i^c \mu_i dn_i \tag{4.65}$$

for c independent components.

4.14.2 The Gibbs–Duhem equation

An interesting result is obtained by integrating (4.63) at constant T and P, that is, by supposing that the quantity of system varies from zero up to some finite value. The result is[22]

$$\mathbf{G}_{T,P} = \sum_i \mu_i n_i \tag{4.66}$$

which shows that μ is not just some abstract partial derivative, but is the Gibbs energy per mole of a dissolved substance. That is, Equation (4.66) shows that the total Gibbs energy of a system is simply the sum of the number of moles of each component in the system (n_i) times the free energy per mole of that component (μ_i).

Our new "more complete" fundamental equation (4.63) is a bit inconvenient for some purposes, in that some of the differential terms (dT, dP) are intensive, but the others (dn_i) are extensive. We would like to have an equation which contains compositional terms, but which has differentials of intensive variables only. We get this by first differentiating (4.66),

$$d\mathbf{G} = n_1\, d\mu_1 + \mu_1\, dn_1 + n_2\, d\mu_2 + \mu_2\, dn_2 + \cdots + n_c\, d\mu_c + \mu_c\, dn_c \tag{4.67}$$

and subtracting from this Equation (4.63). The result is

$$0 = \mathbf{S}\, dT - \mathbf{V}\, dP + \sum_{i=1}^c n_i d\mu_i \tag{4.68}$$

which is called the Gibbs–Duhem equation, or "Gibbs 97" by readers of Gibbs original papers,[23] as that is its number in the original. One important application is in the derivation of the phase rule (Chapter 11).

---

[22] Equation (4.66) is also the result of Euler's theorem (§C.2.3) applied to **G** as a function homogeneous in the first degree in the masses of the components.

[23] Available in Dover reprint edition (Gibbs, 1875), and highly recommended for mind-stretching.

The integrated form

Another important application of this equation is to show that even though the components in the system are compositionally independent of one another, their chemical potentials are not always independent. In a binary solution of components 1 and 2, you can change n_1 without changing n_2, but doing this will change *both* μ_1 and μ_2. This turns out to be quite important in Chapter 10. Let's see how it works.

At a constant T and P, Equation (4.68) becomes

$$\sum_{i=1}^{c} n_i d\mu_i = 0 \tag{4.69}$$

so for a binary system

$$n_1\, d\mu_1 + n_2\, d\mu_2 = 0 \tag{4.70}$$

or, dividing by $n_1 + n_2$,

$$x_1\, d\mu_1 + x_2\, d\mu_2 = 0 \tag{4.71}$$

so

$$d\mu_1 = -\frac{x_2}{x_1} d\mu_2 \tag{4.72}$$

$$\mu_1'' - \mu_1' = -\int_{x_1'}^{x_1''} \frac{x_2}{x_1} d\mu_2 \tag{4.73}$$

showing that if you know how one potential changes as a result of a compositional change, you can calculate the change in the other. They are not independent.

This process can be extended to ternary or even higher order systems, and it can be shown that all chemical potentials in a multicomponent system can be evaluated if the potential of one component is known over the whole compositional space. However, the process becomes complex, and has been little used even in ternary systems. Pitzer and Brewer (1961, Chapter 34) have a discussion of this, with several useful references. The vast majority of uses of the Gibbs–Duhem relation have been in binary systems, using variations of Equation (4.73). We should mention that the term Gibbs–Duhem is commonly applied to any of Equations (4.68)–(4.73).

Equation (4.73) proves inconvenient in practice, because $\mu \to -\infty$ as $x \to 0$. That is, μ changes very rapidly in dilute solutions and does not approach a limit, so it's not convenient in integration. Equation (4.73) is therefore modified to a more convenient form, which we will see in Chapter 10.

4.14.3 Other kinds of open systems

The open system as a subsystem in an overall closed system (Figure 2.2b) is by far the most commonly used kind of open system in geochemistry. In any multi-phase system, the compositions of the phases must be adjusted to achieve a Gibbs energy minimum, and this type of calculation is very common. But there are other kinds of open systems.

Externally controlled potentials

You might say that the open systems discussed above are not "really" open; the phases can change composition, but that can only be handled thermodynamically in an overall closed system. The significant feature of this situation is that the overall system composition is constant. A truly open system has matter passing in and out, to and from the environment of the system, and we may not know exactly how much of this stuff is in the system at any one time – it may be quite variable. How do we handle that?

In real life, we may know no more than that – we know there are some components in our real system, but we don't know how much, and we don't know anything else about them, except their chemical composition and perhaps that they have the same *T* and *P* as all the other components. If that is the case, thermodynamics is of no help. To make a thermodynamic model we must know either the mass (or number of moles) or the chemical potential of each component. A system in which we know the masses of some components and the potentials of others can be visualized as in Figure 4.13, where for simplicity we have only one component of each type. In this system we have a hypothetical semipermeable membrane, through which component B can freely pass, but component A cannot. The external system containing B can have its pressure or composition controlled independently of the system containing composition A, so we have an open system with externally controlled potentials. Note that the amount or mass of component B in the external part of the system

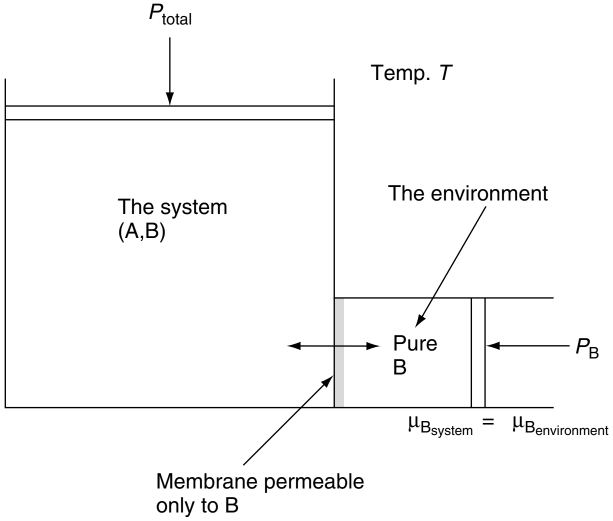


Figure 4.13 The system contains two components A and B, and the environment contains pure component B. The temperature is the same in both but the pressures are independent, so that the chemical potential of B in the system (A,B) is controlled by that in the environment.

is irrelevant. The use of such a "magical membrane" to model our system may seem bothersome to some. We can address this in two ways.

- There actually are such membranes in common use in experimental work. Platinum at high temperatures is permeable to hydrogen but not to other substances, and this fact is extensively used to control the oxidation state of experiments. This is discussed further in §11.5.2.
- Real geological systems have no such membranes, but they do have components (e.g. H_2O) which have externally controlled potentials, such as practically any rock undergoing hydrothermal alteration, and the membrane allows us to model this situation. The fact that the model system has a membrane but the real system does not is no more important than many other differences between the two kinds of systems.

The question is, what thermodynamic potential is appropriate to this situation? It cannot be the Gibbs energy, because if we write (4.65) for components A and B, and consider T, P, and n_A to be fixed, we get

$$d\mathbf{G}_{T,P,n_\mathrm{A}} = \mu_\mathrm{B}\, dn_\mathrm{B}$$

The closed system criterion for equilibrium is $d\mathbf{G}_{T,P} = 0$, but this is obviously inappropriate in this case because μ_B is certainly not zero and dn_B is not necessarily zero because we have said nothing about P_B. There are in fact no conditions pertaining only to the system containing both A and B that will specify the equilibrium state when the environment is attached to it. This is essentially self-evident. A thermodynamic potential appropriate for the system plus environment must evidently contain terms referring to both parts of the system.

Thompson's L function

The thermodynamics of this situation was worked out independently by Korzhinskii (1959) and Thompson (1970). The easiest way to do this is by a Legendre transform (discussed in §C.4) of a potential already established for other parameters. Starting with

$$\mathbf{G} = \mathbf{G}(T, P, n_\mathrm{A}, n_\mathrm{B})$$

we want a new function, which Thompson (1970) has called **L**, which preserves the potential qualities of **G** but has different independent variables; i.e., we want

$$\mathbf{L} = \mathbf{L}(T, P, n_\mathrm{A}, \mu_\mathrm{B})$$

We find the Legendre transform by subtracting from **G** the product of the old variable and the derivative of the old function with respect to that variable. Thus

$$\mathbf{L} = \mathbf{G} - n_\mathrm{B}\left(\frac{\partial \mathbf{G}}{\partial n_\mathrm{B}}\right)$$

or

$$\mathbf{L} = \mathbf{G} - n_{\mathrm{B}}\mu_{\mathrm{B}} \tag{4.74}$$

Differentiating (4.74) and adding (4.65) written for components A and B, we get

$$d\mathbf{L} = -\mathbf{S}\,dT + \mathbf{V}\,dP + \mu_{\mathrm{A}}dn_{\mathrm{A}} - n_{\mathrm{B}}d\mu_{\mathrm{B}} \tag{4.75}$$

showing that **L** has $(T, P, n_{\mathrm{A}}, \mu_{\mathrm{B}})$ as independent variables. **L** is a function that is minimized at equilibrium for given values of T, P, n_{A} and μ_{B}, in exactly the same way that **G** is minimized for given values of T, P, n_{A} and n_{B}. Since n_{B} is not zero, $d\mathbf{L}$ is clearly only zero when $d\mu_{\mathrm{B}}$ is zero, which will of course be when μ_{B} is the same in the system and in the environment.

The **L** function has not found much use in geochemistry to date, and we will not use it in this text. It is included here to demonstrate that thermodynamics can in fact handle truly open systems, and because it is an interesting exercise in the use of the Legendre transform. Also it seems quite posssible that inventive geochemists will find much greater use for this (and other thermodynamic potentials?) in the future. Its derivation and use in metamorphic studies has had a somewhat troubled history, nicely reviewed and discussed by Rumble (1982).

Engineering applications

Open systems are much more common in mechanical engineering. Basically this is because the applications do not involve chemical reactions to nearly the same extent. All kinds of heat engines, turbines, refrigeration machines, boilers, and so on can be modeled in terms of mass and energy balances. Conservation of mass and energy in the inputs and outputs, which is basically the first law, plus heat to work conversion, the second law, account for much of thermodynamic modeling in these applications. Mechanical engineering texts may not even mention the concept of activity (Chapter 8), so central to geochemistry.

4.15 The meaning of entropy

As long as we are dealing with pure compounds, we have answered just about all our questions. We have an energy parameter, the Gibbs energy, which always decreases in spontaneous reactions at a given T and P, and we know how to measure this energy term – calorimetry. We know, however, that this energy term, ΔG, is made up partly of a fairly comprehensible term ΔH, which is just a heat flow term, and another term ΔS, which is more mysterious. All we know about this one is that we defined the second law such that the entropy always increases in spontaneous reactions in isolated systems. The entropy is not itself an energy term, but the product of T and S, or T and ΔS, is an energy term.

If we had to rely on classical thermodynamics, we would know little more than we have already said about entropy. It is a parameter, with a method of measurement, which increases in spontaneous processes, *even when no energy changes are possible*, that is, in isolated systems. We would also notice, after measuring the entropy of many substances, that the entropies of gases are relatively large, those of solids relatively small, and those of liquids somewhere in between, but we would probably not have any mental picture of what entropy represents physically.

If you look at some processes that are quite irreversible but that involve little or no energy change, such as the mixing of two gases or the spreading of a colored dye in water, you observe a driving force for processes that is quite different from the energy-drop paradigm we have been pursuing (the ball rolling downhill). There is no energy drop when gases mix, but they do so invariably and irreversibly, and this shows that another driving force for reactions is *mixing*, or an increase in "mixed-upness," which will take place if it is possible. If you think about gases as collections of countless tiny molecules zipping around with the speed of rifle bullets, but with quite a lot of space between them, you realize that if two different gases are brought together, it is no more difficult to understand why they will always mix together than it is to understand why a ball will roll downhill. But this mixing process involves no energy change (at least for ideal gases), so the first law of thermodynamics is powerless – it cannot be the basis for a thermodynamic explanation. The second law and entropy do provide it. Entropy can be thought of as a *degree of mixed-upness*, and increasing the randomness or mixed-upness of systems is *one* of the driving forces for spontaneous reactions.

The confusion arises because it is not the only one – the ball rolling down the hill (energy decrease) is also one. It is the two together that provide the complete answer. In some processes energy decrease is the dominant factor, and in others, mixing or entropy increase is the dominant factor. The two are brought together in the Gibbs energy equation

$$dG_{T,P} = dH - T\,dS \tag{4.38}$$

which can also be written

$$\Delta G_{T,P} = \Delta H - T\,\Delta S$$

Here there are two factors that together determine whether $\Delta G_{T,P}$ will be positive or negative. One is $\Delta H (= \Delta U + P\,\Delta V)$, which is the energy change due to heat and work in the process represented by Δ, and the other is $T\,\Delta S$, the energy change due to the mixing factor. In many spontaneous reactions, ΔH is negative (the process is exothermic) and ΔS is positive (mixed-upness increases), and so both factors are negative ($T\,\Delta S$ positive, $-T\,\Delta S$ negative) and ΔG is negative. In other reactions, ΔH is positive (process is endothermic, e.g., melting ice), but the ΔS term is sufficiently large and positive so that $T\,\Delta S > \Delta H$, and ΔG is negative in spite of the positive ΔH. This will be especially true at high temperatures, when T is large. There are all varieties of combinations; the point is that whether or not any particular process is spontaneous is the result of the two competing factors. Systems want to lower their energy content, but they also want to maximize their mixed-upness. The balance between these factors decides the issue.

4.15.1 But what is entropy, *really?*

This question has been around since Clausius invented the term in 1865, and the answer takes on many forms. Some follow the historical route, from steam engines, to Carnot, Clausius, Thompson, Joule, Rankine, and so on. A particularly lucid, concise account of this history is Purrington (1997). A central feature of this approach is Carnot cycles, as used by Clausius to deduce the existence of the entropy parameter. This approach is rather abstract, and needs some manipulation to be seen to be connected to thermodynamic potentials and chemical reactions. Others emphasize the impossibility of some processes, or the "availability" of energy, and some have a rather unique viewpoint, such as Reiss (1965), who considers entropy as the "degree of constraint."

Virtually since the beginning, however, a popular viewpoint has been to see entropy as a measure of disorder. Helmholtz used the word "Unordnung" (disorder) in 1882. This results from familiar relationships such as $S_{\text{gases}} > S_{\text{liquids}} > S_{\text{solids}}$, and the universally positive entropy of mixing. We used this relationship in the previous section when we spoke of "degree of mixed-upness." However the "disorder" analogy can involve a serious fallacy, as made clear by Lambert (1999; see also http://www.entropysite.com).

The rather subjective concepts of disorder and mixed-upness are useful analogies in certain situations, such as melting solids and mixing gases, but they fail completely in most other situations. You cannot tell whether α- or β-quartz is "more disordered" or has the higher entropy by looking at their structures. Shuffling a deck of cards perhaps increases its disorder, but it does not increase the entropy of the cards. To see this, just imagine cooling the cards down to near zero K, measuring their heat capacity up to room temperature, and determining the entropy of the card deck from these measurements. Obviously, the result will be the same no matter in what state of order the cards are. The same is true for the configuration of any macroscopic system. Checkers on a board are ordered at the start of a game, and become progressively disordered during a game, but the entropy of the checkers remains the same no matter what their arrangement.

A complete understanding of this inevitably involves statistical and even some quantum mechanics. In statistical mechanics, the atoms in a system are postulated to have a number of "microstates," which are energetic states of the system, the definition of which involves the distribution of energy quanta over available energy levels in the particles in a macroscopic system. If a change occurs so as to increase the number of available microstates, energy will spread out among the newly available microstates. The basic postulate of statistical mechanics is that *over long periods of time, every possible microstate of an isolated assembly occurs with equal probability* (Nash, 1972). The quantity W in the famous Boltzmann–Planck equation

$$\mathbf{S} = k \ln W$$

is in fact the number of these microstates, where $W = 1$ for a perfect crystal at absolute zero K and k is Boltzmann's constant (which, significantly, includes temperature). The stable state simply has an immensely greater number of equivalent microstates.

Much of the discussion in statistical mechanics concerns probability distributions and their application to microstates. Information theory (Shannon, 1949) defines a quantity $-\sum_i p_i \log p_i$, called the entropy of a probability distribution, which has some remarkable similarities to the properties of classical or thermodynamic entropy. What seems to be overlooked by nonexperts is that in chemistry and physics, the probability distributions being discussed are those of *energy* distributions among microstates. When we speak of mixing or "mixed-upness," we actually refer to particles (atoms, molecules) distributing themselves in different energetic states, not to macroscopic things like cards or checkers. Explaining entropy does involve probability theory, but the probabilities refer to energetic states, not to the outcome of dealing cards or tossing dice.

The central fact about entropy as used in science is that it involves the distribution of energy in a system. Energy tends to become "spread out," or delocalized, if not prevented from doing so. The "configurational entropy" much used by mineralogists in discussing the various arrangements of atoms on a crystal lattice (Chapter 14) is fundamentally different fom the arrangement of checkers on a board because energy is transferred when atomic arrangements are changed – the heat capacity of each arrangement is different.

If you "really" want to understand entropy, you need to learn more than just equilibrium thermodynamics. In this book, we take the simple view that entropy is a parameter, having a clearly defined method of measurement, which enables us to define thermodynamic potentials in chemical systems. It is simply related to "disorder" in many simple situations, which is an intuitive aid, but this aid doesn't extend very far. Because of this resemblance to probability and disorder, entropy has been related to everything from shuffled cards to the fall of empires, but these connections for the most part have nothing to do with the second law of thermodynamics.

4.16 A word about Carnot

Some readers may have noticed that, despite the numerous equations and derivations, the introduction of entropy and the second law is less than completely rigorous in this chapter. This is because at a crucial point, the introduction of the relationship between entropy, temperature, and heat (Equation 4.3), we rely on analogy rather than demonstration or argument. Most treatments rely on a discussion of Carnot cycles, but as one chemist has observed,

> Arguments based on heat engines have little appeal to chemists
>
> (McGlashan, 1979, p. 111).

In my experience, such arguments have even less appeal to geochemists. This is perhaps unfortunate, because approaching the subject from various points of view certainly aids understanding, and in fact the heat engine approach is very useful in the derivation of Equation (4.3), and also the kelvin temperature scale.

However, such discussions tend to be rather long and abstract, and exactly how conclusions about irreversible heat transfers in Carnot cycles get transferred to irreversibility in chemical reactions requires even more discussion. This detracts from the task attempted here – the shortest possible, intuitively clear, development of the concepts necessary to use thermodynamics in solving Earth science problems. Readers who want to know more about the historical development of the entropy concept and its deeper meaning must consult the many excellent treatises on this subject.

4.17 The end of the road

We pause here to note that, in case you hadn't noticed, we have arrived at the answer to the question posed in Chapters 1 and 2. The question was, what controls whether a reaction or a process will happen or not happen? Why does water freeze below 0 °C and ice melt above 0 °C? What is the "chemical energy" term that always decreases to a minimum, like the ball rolling down the hill? In answering this question, we first had to define fairly carefully some terminology such as system, equilibrium, and process. We then noted (Chapter 3) that systems have fixed energy contents (**U**, or U) at equilibrium, but this didn't help, because although this energy is conserved, it doesn't distinguish at all between directions processes take (bricks could cool themselves, and use this energy to fly, as far as U is concerned).

The missing ingredient to understanding why reactions go one way and not the other is entropy. Entropy is defined as a state variable that always *increases* in spontaneous processes in isolated systems. But a parameter that is useful only in isolated systems is not of much practical use, so we defined another state variable, the Gibbs energy, that always decreases in spontaneous process in systems at a given T and P (see Figure 4.14). This is the parameter we have been looking for. Figure 4.14 shows two different states of a system at

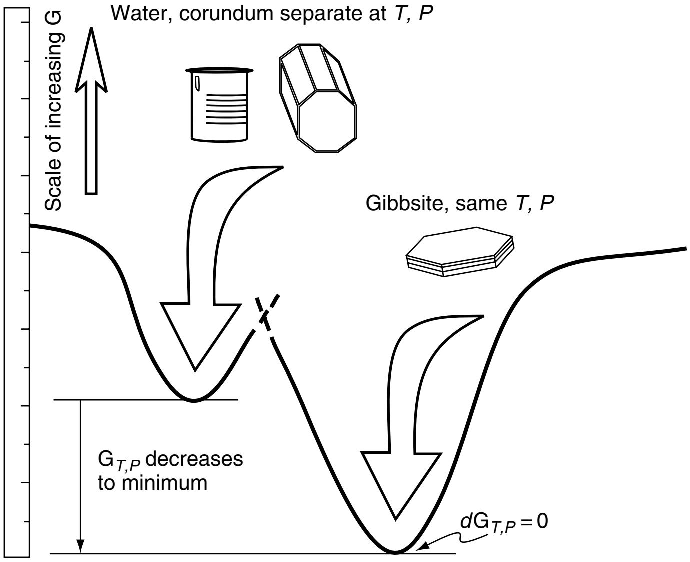


Figure 4.14 Gibbs energy decreases in spontaneous processes at a given temperature and pressure.

the same temperature and pressure. A spontaneous process (the formation of gibbsite) occurs when the constraint keeping the reactants separated is removed (corundum and water are mixed together).

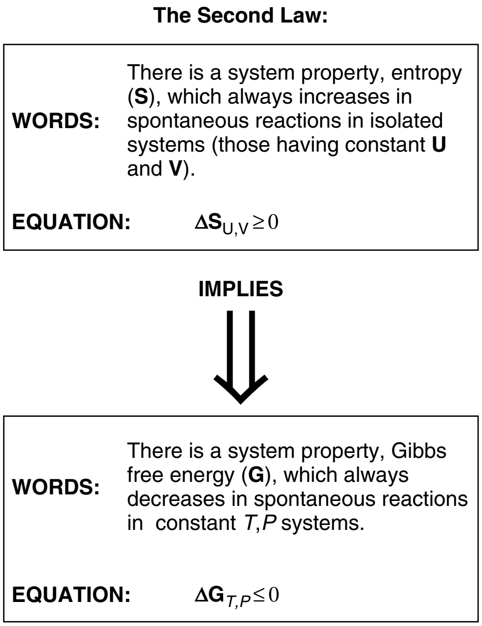
Similarly, in Figure 2.4, the "chemical energy" term is in fact the Gibbs energy. At +5 °C, 1 bar, $G_{\text{ice}} > G_{\text{water}}$, and at −5 °C, 1 bar, $G_{\text{water}} > G_{\text{ice}}$. In Figure 2.5 $G_{\text{diamond}} > G_{\text{graphite}}$, assuming both have the same P and T. Mathematically, the entropy and Gibbs energy potentials are two sides of the same coin – one implies the other, as shown in Figure 4.15. Looking at it in still another way, you see from Figure 4.5 that a single system can have $dU_{S,V} = 0$ and $dS_{U,V} = 0$ simultaneously. If we were not limited to three dimensions, we could show that the same system also has $dG_{T,P} = 0$. Each condition implies all the others.

In biochemistry, processes having a negative $\Delta_r G$ ($\Delta_r G < 0$) are termed *exergonic*, and those having a positive $\Delta_r G$ are termed *endergonic*. For some reason, these terms are not common in geochemistry.

The problem at the moment is that these new state variables S and G will have no "feeling of reality" for a reader new to the subject. That is, what *is* entropy or Gibbs energy, and how does one measure these things? Only by actually using these concepts will one become familiar with them. The next chapter is a first attempt at describing these variables in more familiar terms.

4.18 Summary

What you should know at this point is that we have defined a parameter, entropy, which can tell us which way reactions will go, but only in isolated



The Second Law:

WORDS: There is a system property, entropy (**S**), which always increases in spontaneous reactions in isolated systems (those having constant **U** and **V**).

EQUATION: $\Delta\mathbf{S}_{U,V} \geq 0$

IMPLIES

$\Downarrow$

WORDS: There is a system property, Gibbs free energy (**G**), which always decreases in spontaneous reactions in constant *T*,*P* systems.

EQUATION: $\Delta\mathbf{G}_{T,P} \leq 0$

Figure 4.15 All you need to know about the second law and the derivation of the Gibbs energy function. These statements and equations can be written equally well in their molar forms (using *U*, *S*, *V*, and *G*).

systems.[24] The statement defining entropy is one way of stating the second law of thermodynamics. Combining entropy with the first law, we then defined another parameter, the Gibbs energy, which can tell us which way reactions go in systems at a given temperature and pressure. We also showed that the Gibbs energy is equal to the maximum amount of useful energy or work available from such reactions, but we have not yet seen how to measure any of these apparently useful quantities.

It is normal at this point for newcomers to this subject to be rather confused, or perhaps impatient. If we think about natural processes that we would like to understand, such as occur in living plants and animals, or even simpler inorganic processes such as occur in creating our weather patterns or in erupting volcanoes, we could be forgiven for wondering what earthly use the kind of material we have considered up to now can be. We seem to have restricted ourselves to ridiculously simple cases such as balls rolling in valleys, and even though we have claimed that certain simple inorganic processes such as melting ice and polymorphic mineral changes are analogous, we haven't shown how to do anything remotely useful.

Not only is it not yet useful, but even after restricting ourselves to simple cases and claiming not to be dealing with reality but with models of reality, we

---

[24] That is, only in isolated systems if used by itself as a thermodynamic potential. Indirectly, i.e., combined with other state variables to define other thermodynamic potentials, it gives directionality parameters for any kind of equilibrium system. All thermodynamic potentials include entropy in some way.

have introduced at least one concept (entropy) that is rather difficult to fully comprehend and have used a level of mathematics that, although not exceeding that taught in introductory calculus courses, has physical implications that are hard to grasp. The best remedy for this is to review the material to some extent, and to plunge ahead even if it is not entirely clear. After some familiarity with practical applications is attained, some of the earlier material will become clearer, so the best approach is continuous review, in addition to assimilation of new material.

It may well seem that we have made no progress toward understanding complex processes, but this is not true. Most natural processes are so complex that we simply must start with the very simplest ones we can think of and define our terms very carefully. Our goal of finding the secret to why reactions go in one direction and not the other may seem overly simple, but it is in fact the basic concept necessary to build up an understanding of all the natural phenomena mentioned above. Of course, even when we have mastered thermodynamics, we will find that we don't have all the answers to all our questions; in fact, we will find that the things thermodynamics can tell us are fairly limited. They have, however, a level of certainty which surpasses that of most other ways of looking at the same problems, and this makes the subject an absolutely essential element of all research into problems that involve energy transfers. You may wish to know much more than thermodynamics can tell you, but you need to know what it can tell you.

5
Getting data

5.1 Introduction

We have had quite enough theoretical discussion for now. Let's see how to get some numbers into our equations so as to be able to calculate something useful. Welcome to the world of experimental thermochemistry.

In this chapter we will have a look at a few of the ways in which the thermodynamic parameters we have derived are measured; i.e., where the numbers in the tables and databases come from. A deep knowledge of this subject is not necessary in order to use thermodynamics to model chemical, geological, or environmental systems, in the same sense that a knowledge of a composer's life and times is not necessary to enjoy his or her music. But it does enrich the experience, and in the case of using thermodynamics, such knowledge does serve to make the user conscious of the many, many reasons why his or her data might be incorrect. It enables the user to truly believe in the "modeler's motto" – never trust your data absolutely.

Thermochemical data are produced for the most part by dedicated scientists, who devote a good part of their lives to tracking down elusive sources of error, and devising ever-improved methods for determining nature's fundamental parameters as defined by thermodynamic theory. When determined by independent methods and/or independent laboratories, the results are often satisfyingly in agreement, but almost as often they are not, meaning that there is some source of error, and identifying it can take a lot of discussion (perhaps arguments would be a better term) and a long time.

Beginning with the establishment of the Geophysical Laboratory in Washington in 1905, scientists primarily interested in geological processes have contributed to our knowledge of these thermodynamic properties, especially, as might be expected, those of the rock and soil-forming minerals, and aqueous solutions such as sea water and hydrothermal solutions. In addition to the measurement of mineral and solution properties, collecting these data into databases for use by computer programs, and the testing of these databases for internal consistency has become increasingly important.

The experimental part and the self-consistent database part of using thermodynamics in the Earth sciences today have both become large and complex subjects, each worthy of separate study. Geologists and geochemists primarily interested in Earth processes, but who need to do some thermodynamic modeling in order to better understand these processes, can be forgiven for not wanting to be involved with either experiment or database development. Nevertheless, rigorous thermodynamic methods can produce nonsensical results, given incorrect data, so a blind faith in the database attached to some computer program can be a recipe for disaster.

In this chapter (nor in this book) we cannot address these areas of study to any great extent, but understanding thermodynamics for most people is helped by having some knowledge of how fundamental data are determined. The aim of this chapter is thus to trace the connection between laboratory experiments and thermodynamic modeling. The idea is not only to impart some idea of how it is done, but some idea of the great difficulties involved at every stage, and therefore some respect for data. Good data are a precious commodity; you should have some interest in where yours come from.

The discussion of methods in this chapter is very sketchy. Many details and variations are omitted, because all we want to do is illustrate what is involved. Do not conclude from the simplicity of the presentation that there is nothing much to this experimental business. It is not too difficult to get data that are not much good; it is extremely difficult, time consuming, frustrating and often expensive, to get excellent data, ones that stand the test of time.[1]

Sections 5.2 to 5.7 deal primarily with obtaining data for solid phases. Section 5.8 deals with liquids, gases and solutes. An overview of many methods used by Earth science experimenters can be found in Ulmer and Barnes (1987).

5.2 What to measure?

Imagine that you have an interesting field problem, and you want to do some thermodynamic modeling to better understand it, but for some reason there are no data for the mineral you are most interested in, gibbsite. You go to a fully equipped laboratory, but what do you do there?

5.2.1 What not to do

Perhaps in your field study you have concluded that gibbsite is being formed from some other mineral by some alteration process you have figured out, and

---

[1] Much the same can of course be said about field work, or indeed about any serious branch of science. If you don't know much about it, it looks easy. I am reminded of the story of the geochemist, well versed in theory and experiment, who said that field mapping could be carried out by a bunch of monkeys trained to collect samples. Let's hope that such appalling ignorance is not widespread.

you want to use thermodynamics to see if it makes sense. Maybe thermodynamics will tell you that this process is not permissible at your favorite T and P, or that gibbsite is not the stable phase.

But your alteration process involves other minerals, as well as an aqueous solution. The most obvious thing to do then is to set up an experiment in which you try to duplicate this reaction. Perhaps you put some groundwater and some minerals together in some reaction vessel for a while, then look at the results. In other words, you try to duplicate nature. This approach has been tried many times, and generally the results are of very limited usefulness. The results are usually quite complex and difficult to interpret, and in any case they apply only to one specific set of starting materials – it is difficult to draw general conclusions from the results. Besides, the measurements you can make in such experiments, which are usually the compositions of some complex phases, have nothing to do with measuring heat and work, which as we have seen is the basis of all our thermodynamic parameters.

5.2.2 What to do

If you wish to use thermodynamics, perhaps you should look at the equations to get a clue as to what to measure. Looking at the first law, $\Delta U = q + w$, you see that you will be involved in measuring heat and work. Looking at the second law, either $dS = \int (C_P/T)\, dT$ or $\Delta S = \Delta H/T$, you see that measuring heat at various temperatures is also involved. Measuring quantities associated with q and with w are therefore fundamental to experimental thermodynamics. There are quite a number of ways of doing this, but for the moment we can divide them into direct and indirect methods.

Direct methods

Based on our discussion so far, it seems that we would like most to know $\Delta_r G$ for reactions of interest to us, because this will tell us which way the reactions will go, assuming that pressure and temperature are fixed. For example, if we were not sure which of the two forms of carbon was the stable form at 25 °C, 1 bar, we could measure $G_{\text{graphite}} - G_{\text{diamond}}$, which is $\Delta_r G$ for the reaction

$$\text{C(diamond)} = \text{C(graphite)}$$

and if this quantity was negative, then graphite would be stable, and if it was positive, diamond would be stable. The same reasoning would hold for any complex reaction involving gibbsite, as long as we know the Gibbs energy of every species in the reaction.

There are quite a number of ways of determining changes in Gibbs energy, but we will discuss only the most common one here. Others are associated with determining the equilibrium constant or cell voltages, as we will see in

Chapters 9 and 12. To see how changes in G are measured, consider first Equations (4.56) and (3.17)

$$\Delta G_{T,P} = \Delta U - T\,\Delta S + P\,\Delta V \tag{4.56}$$

$$\Delta H_P = \Delta U + P\,\Delta V \tag{3.17}$$

Combining these, we have

$$\Delta G_{T,P} = \Delta H_P - T\,\Delta S \qquad (5.1)$$

From this, we see that we can calculate ΔG for a process if we know ΔH_P and ΔS for that process. We know (from §3.5) that ΔH_P is simply the heat transferred to or from the closed system during a constant pressure process, so all we have to do is carry out some process (reaction) at some constant P, probably atmospheric P, and measure how much heat is evolved or absorbed.

This ΔH is also connected to ΔS, but only for a reversible process, as we have seen. This might be a little difficult to do, experimentally. But we also have an equation involving ΔS, C_P and T (Equation 4.25), and C_P is related to ΔH (Equation 3.25), so it begins to look like measuring heat is pretty important, even if you are not concerned with heat flow in your field situation. Therefore, *calorimetry*, the art and science of measuring heat flows, is the secret to determining values of ΔG.

Indirect methods

So far we have talked about using thermodynamics to determine phase relationships. But the opposite approach can also be used; phase relationships determined under strictly controlled conditions (meaning controlled by the phase rule, Chapter 11) can be used to deduce fundamental thermodynamic properties. For example, for gibbsite, you might conduct experiments to determine the temperature at which gibbsite changes to corundum,

$$2\,\mathrm{Al(OH)_3}(s) = \mathrm{Al_2O_3}(s) + 3\,\mathrm{H_2O}(l) \qquad (5.2)$$

or, you might determine the solubility of gibbsite in water at various temperatures and pH values. Both these relationships are interesting to Earth scientists even without thermodynamic manipulation, but they can also be used to determine the thermodynamic properties of gibbsite and related species. This is a subject for later chapters, but intuitively it would seem that knowing the properties of corundum and water, you might be able to deduce those for gibbsite from the requirement that reaction (5.2) be at equilibrium.

Even more importantly, families of such relationships, that is, a number of relationships involving the same minerals, can be used to test the consistency of thermodynamic data. For example, suppose you have determined $\Delta_f H°$ for each of gibbsite, corundum, and water from calorimetry. If you get a different result from studying reaction (5.2), then there is some error. Or you might have several different reactions involving gibbsite, so your data for gibbsite must fit all those reactions. In fact all the data for all the minerals in your database must fit all the reactions that you know about. Finding errors in these cases can be difficult, but the point is, phase relationships are a good way to examine how consistent your data are.

For now, let's concentrate on calorimetry, the classical method for determining thermodynamic properties.

5.3 Solution calorimetry

Heat flows can be measured in various ways. One way is to observe some process in which heat is liberated under controlled conditions, resulting in a rise in temperature, and then duplicate that temperature rise using an electrical heater. The energy used by the heater can be measured exactly, and will equal the energy released by the process considered. This is the principle used in the calorimeter in Figure 5.1.

This apparatus is used to measure how much heat is liberated when a known amount of solid material, such as a mineral, dissolves. Most minerals are notoriously insoluble in water and so an acid, such as hydrofluoric acid (HF), is used. The method is called *solution calorimetry.*

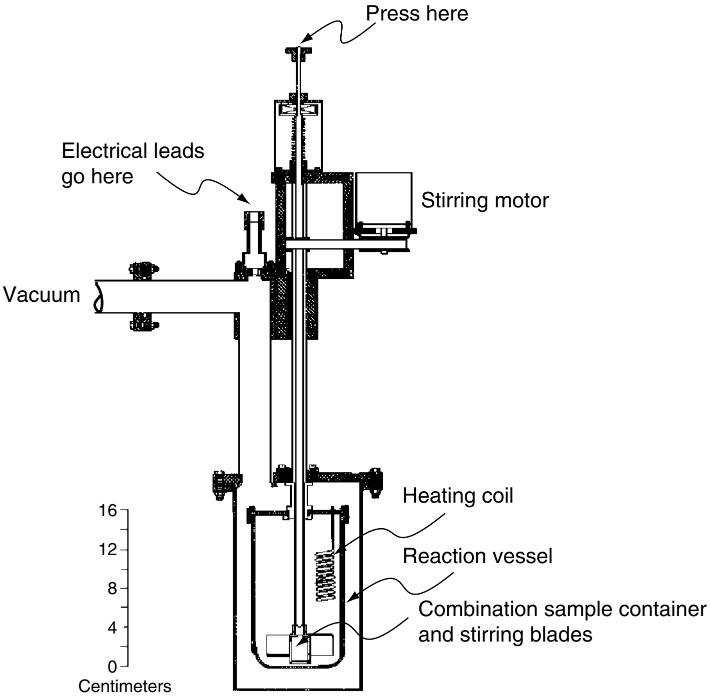


Figure 5.1 An adiabatic heat-of-solution calorimeter. The reaction vessel contains acid. Pushing down on the handle at the top punctures the upper seal and pushes out the bottom of the sample container, allowing the sample to dissolve. A thermometer is wound around the reaction vessel and records the change in temperature. No electrical leads are shown. (Simplified from Robie and Hemingway, 1972.)

5.3.1 The method

A few grams of crushed mineral are put into the *sample holder* and sealed with gold foil. The sample holder is then place in the *reaction chamber*, which is then filled with acid. A long rod reaches from the top of the sample holder through various seals to the top of the apparatus. This assembly is then sealed and placed in a vacuum chamber, which goes into a water bath. The purpose of the vacuum and water bath is to minimize the loss of heat from the reaction chamber.

When everything has settled down, the mineral sample and the acid are at the same temperature, but are separated. The long rod is then pushed down. This punctures the seal on the sample container, and the bottom also falls out, allowing the sample to mix with the acid and dissolve. The sample holder also has fins, and rotation stirs the solution and speeds up the dissolution process. The dissolution of the mineral releases heat, which raises the temperature of the acid, and the amount of temperature change is measured by a resistance thermometer, which is wrapped around the reaction vessel. The apparatus is calibrated by using an *electrical heating coil* to raise the temperature in a different experiment, but using exactly the same setup. The voltage drop across the electrical heater and the current flowing through it are known, and so the amount of heat required to raise the temperature of the calorimeter by any given amount is known exactly by turning on the heater for a short time and observing the temperature increase. By comparing the temperature change caused by the heating coil to that caused by the mineral dissolution, the heat liberated by the mineral dissolution can be determined quite precisely. A number of small corrections must be made for various heat losses in the apparatus, plus a correction to the heat measured over the temperature interval to what would have been observed if the process had occurred at a constant temperature of 25 °C. These calculations require a knowledge of the *heat capacity* of the calorimeter (see below). Because we know the mass of mineral grains used, the heat of solution per mole of mineral at 25 °C then can be calculated. The whole process is exacting and painstaking.

5.3.2 The interpretation

What are the meaning and use of this heat of solution? In terms of the processes we have been discussing, we have observed an irreversible reaction between a metastable state (pure acid and mineral grains, separated, at T_1) and a stable state (mineral dissolved in acid at T_2), made some measurements, and calculated from this the heat that would be released in the reaction

$$\text{mineral, HF separated} \rightarrow \text{mineral dissolved in HF} + q_{\text{dissolution}}$$

at 25 °C. If the calorimeter is open to the atmosphere, then the mineral dissolution process happens at a constant pressure, and by Equation (3.18), the heat measured, $q_{\text{dissolution}}$, is equal to the change in enthalpy of the system, ΔH. This change is illustrated in Figure 5.2.

Of course, a heat of solution is not exactly what we wanted, although it is a ΔH. We want a ΔH that is the difference between products and reactants of reactions of all kinds, such as our corundum–water–gibbsite or diamond–graphite reactions, and innumerable others. But the heat of solution technique allows us to do this. Note that in any balanced chemical reaction, the total or bulk composition of the reactants must be exactly the same as that of the products. That's what "balanced" means – all the atoms on the left side of the reaction must appear also on the right. Therefore, if in separate experiments we dissolve the reactants and the products in the same kind of acid, we will get identical solutions. We will, however, measure different heats of solution, because the products and reactants have different structures and different energy contents. Therefore, the difference in the heats of solution must be equal to the difference in enthalpies of the products and reactants themselves.

To put this argument in formal terms, suppose our reaction is

$$\mathrm{A} + \mathrm{B} = \mathrm{AB}$$

for example, $SiO_2 + Al_2O_3 = Al_2SiO_5$. First we dissolve the reactants, and then, in a separate experiment, we dissolve the products:

$$\mathrm{A} + \mathrm{B} + \text{solution}' \rightarrow \text{solution}'' + \text{heat} \quad (\Delta H_1)$$

$$\mathrm{AB} + \text{solution}' \rightarrow \text{solution}'' + \text{heat} \quad (\Delta H_2)$$

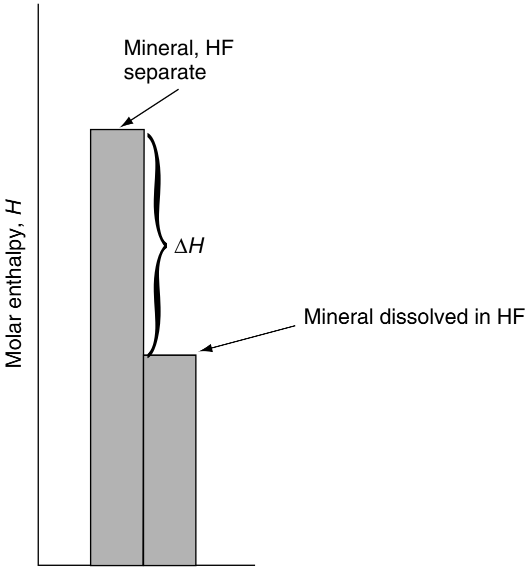


Figure 5.2 Enthalpy is a state variable of fixed but unknown value in the beginning and final equilibrium states. ΔH is obtained by measuring the heat liberated in the reaction at constant pressure.

As long as both solution′ and solution″ have the same compositions in both reactions, the reactions may be subtracted giving

$$\mathrm{A} + \mathrm{B} \rightarrow \mathrm{AB} + \Delta H_3$$

where $\Delta H_3 = \Delta H_1 - \Delta H_2$ and is the heat of reaction ($\Delta_r H$) of the reaction $\mathrm{A} + \mathrm{B} = \mathrm{AB}$, as shown in Figure 5.3. A, B, and AB can also refer to complex organic compounds of any kind. We need only be able to separate them into their pure forms, so as to be able to work with them.

Because of practical difficulties, the determination of $\Delta_f H^\circ$ of a compound is rarely the sum of only two heats of solution, as in Figure 5.3. Quite often 10 or 15 solution reactions may have to be carried out to determine one $\Delta_f H^\circ$, and the whole process may take several weeks.

5.3.3 A real example

The enthalpy of formation from the elements has been determined for gibbsite most recently by Hemingway and Robie (1977), and we look at their results here. The reactions they used, at a calorimeter temperature of 303.5 K, were:

$$\mathrm{Al}(s) + 3\,\mathrm{HF}(aq) = \mathrm{AlF_3}(aq) + \tfrac{3}{2}\mathrm{H_2}(g) \qquad \Delta_r H_1^\circ = -595195 \pm 1192\,\mathrm{J\,mol^{-1}} \tag{5.3}$$

$$\mathrm{Al(OH)_3}(s) + 3\,\mathrm{HF}(aq) = \mathrm{AlF_3}(aq) + 3\,\mathrm{H_2O}(l) \qquad \Delta_r H_2^\circ = -2046 \pm 3\,\mathrm{J\,mol^{-1}} \tag{5.4}$$

$$3\,\mathrm{H_2}(g) + \tfrac{3}{2}\mathrm{O_2}(g) = 3\,\mathrm{H_2O}(l) \qquad \Delta_r H_3^\circ = -857490 \pm 75\,\mathrm{J\,mol^{-1}} \tag{5.5}$$

The enthalpy of formation from the elements for gibbsite is the reaction

$$\mathrm{Al}(s) + \tfrac{3}{2}\mathrm{H_2}(g) + \tfrac{3}{2}\mathrm{O_2}(g) = \mathrm{Al(OH)_3}(s) \qquad \Delta_f H^\circ_{\text{gibbsite}} \tag{5.6}$$

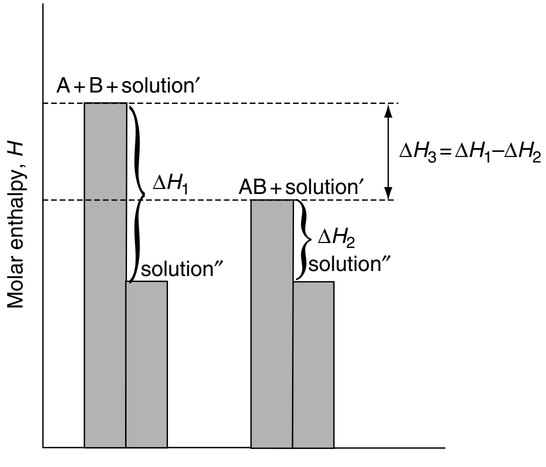


Figure 5.3 Compounds $\mathrm{A} + \mathrm{B}$ together have heat of solution ΔH_1. Compound AB has heat of solution ΔH_2. Both processes result in solution″, so the heat of reaction $\mathrm{A} + \mathrm{B} \rightarrow \mathrm{AB}$, which cannot be carried out in a calorimeter, is $\Delta H_1 - \Delta H_2$.

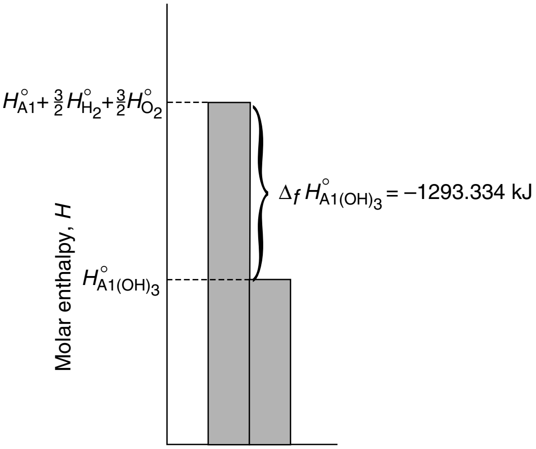


Figure 5.4 The meaning of $\Delta_f H^\circ_{\text{gibbsite}}$.

so *apparently* we get this from the sum $\Delta_r H^\circ_1 - \Delta_r H^\circ_2 + \Delta_r H^\circ_3$.[2] Actually, we do not. We said that the aqueous solution would cancel out if it was identical in all cases. But in reaction (5.3) one mole of Al dissolves in HF solution and H_2 gas is evolved, which leaves the calorimeter. But in reaction (5.4), a mole of Al dissolves in the same kind of HF solution, but it brings with it some H_2O from the gibbsite, which dilutes the acid solution. So the solutions after reactions (5.3) and (5.4) do not have the same composition. This dilution of the acid solution is not a neutral process, but evolves heat, and this must be measured. But this raises another question – how do you write a dilution reaction? And what is the composition of the solutions in those reactions, anyway? The way we usually write reactions gives no clue as to the composition of the solutions after the dissolution reactions. The calorimetry people actually write their reactions in a much more explicit way, as shown in the box on page 121. A graphical idea of the meaning of $\Delta_f H^\circ_{\text{gibbsite}}$ is shown in Figure 5.4.

5.4 The third law

Now we must consider how to measure entropy. So far, all we know about entropy is that it increases in spontaneous reactions in isolated systems, and that it appears in equations such as (4.55) and (4.56). To get it into a form that would suggest a method of measurement, we combine Equations (3.17) and (4.9),

$$dH = dU + P\,dV + V\,dP \qquad [3.17]$$

$$dU = T\,dS - P\,dV \qquad [4.9]$$

---

[2] Remember that to subtract a reaction the best way is to *reverse* it, in this case reaction (5.4), change the sign of $\Delta_f H^\circ_2$ and then add the reactions and the $\Delta_f H^\circ$ values.

giving

$$dH = T\,dS + V\,dP \tag{5.7}$$

If we choose constant pressure conditions, dP becomes zero, and substituting $C_P\,dT$ for dH (Equation 3.23), we have

$$C_P\,dT = T\,dS$$

or

$$dS = \frac{C_P}{T}dT \tag{5.8}$$

Here at last we have the entropy defined in terms of something measurable, the heat capacity. Integrating (5.8), we have

$$S_{T_2} - S_{T_1} = \int_{T_1}^{T_2} \frac{C_P}{T}dT \tag{5.9}$$

and so you see that assuming that you can get numbers for C_P at a series of temperatures, you could divide each C_P by its value of T and evaluate the integral, giving you the difference in entropy between two temperatures.

By now you are probably accustomed to being told that we cannot know the absolute values of thermodynamic parameters, only differences. But this applies only to the internal energy, U, and any parameters that contain U, such as H and G. Entropy is different in that we *can* get absolute values, by virtue of the third law of thermodynamics.

5.4.1 The third law – historical aspects

Lacking an absolute value in some state, entropy is in the same boat as enthalpy and Gibbs energy, having only differences rather than absolute values. Differences in S can be determined from Equation (5.9), but this means that the only way to determine the difference in entropy for any chemical reaction is for there to be some *equilibrium* path between the products and reactants. For example, to determine ΔS between rhombic and monoclinic sulfur at 298.15 K and 1 bar, you would need to measure heat capacities from 25 to 95 °C (the equilibrium phase transition temperature) for both phases, then integrate (5.9) up to 95 °C for rhombic S and back to 25 °C for monoclinic S. To determine the ΔS between calcite and aragonite, you would need to integrate up to the transition pressure for calcite and back down for aragonite, perhaps using Equation (5.39). That is, you would do this in principle, but in practice it would not work well, because V changes so little with pressure.

This is but one of a host of difficulties you would have in finding an equilibrium path between states you were interested in. In some cases an equilibrium path exists but kinetic or other factors make the experimental determinations

The enthalpy of gibbsite

The gibbsite reactions in §5.3.3 were carried out by dissolving 0.005 moles of Al (0.1349 g Al; 0.3900 g $Al(OH)_3$) in 920.0 g of 20.1% HF solution. This HF solution has 184.92 g of HF (9.243 moles), and 735.08 g H_2O (40.8029 moles). To report measurements per mole of Al, these amounts are multiplied by $1/0.005 = 200$. Therefore the starting HF solution for the experiments, per mole of Al, has $9.243 \times 200 = 1849$ moles HF and $40.8029 \times 200 = 8163$ moles H_2O, so the dissolution reactions are written

$$\begin{aligned} \mathrm{Al}(s) + [1849\,\mathrm{HF} + 8263\,\mathrm{H_2O}](aq) &= [\mathrm{AlF_3} + 1846\,\mathrm{HF} + 8163\,\mathrm{H_2O}](aq) \\ &\quad + \tfrac{3}{2}\mathrm{H_2}(g) \end{aligned}$$

and

$$\begin{aligned} \mathrm{Al(OH)_3}(s) + [1849\,\mathrm{HF} + 8263\,\mathrm{H_2O}](aq) &= [\mathrm{AlF_3} + 1846\,\mathrm{HF} \\ &\quad + 8166\,\mathrm{H_2O}](aq) \end{aligned}$$

The dilution reaction was performed by dissolving water into the solution produced by reaction (5.3), thus

$$\begin{aligned} 3\,\mathrm{H_2O}(l) + [\mathrm{AlF_3} + 1846\,\mathrm{HF} + 8163\,\mathrm{H_2O}](aq) &= [\mathrm{AlF_3} + 1846\,\mathrm{HF} \\ &\quad + 8166\,\mathrm{H_2O}](aq) \end{aligned}$$

This results in a correction of $-622\,\mathrm{J\,mol^{-1}}$. In addition, the heat capacities of $Al(OH)_3$, Al, H_2O, and H_2 were required to correct the measurement from 303.5 K to 298.15 K, ($+615\,\mathrm{J\,mol^{-1}}$).

Finally, note that the actual calorimetric measurement of reaction (5.3) was not $-595\,195\,\mathrm{J\,mol^{-1}}$, but $-592\,952\,\mathrm{J\,mol^{-1}}$. The problem is that the H_2 gas which is evolved during the experiment, and which escapes from the calorimeter, carries with it some HF vapor and some H_2O vapor, and this causes some cooling due to evaporation. Correcting for this requires knowledge of the enthalpy of vaporization of HF and H_2O from the experimental solution. The final result is $\Delta_f H^\circ_{\mathrm{gibbsite}} = -1\,293\,130\,\mathrm{J\,mol^{-1}}$.

The $\Delta_f H^\circ$ of gibbsite in Appendix B, from the NBS tables of Wagman et al. (1982), is very slightly different, $-2586.67\,\mathrm{kJ\,mol^{-1}}$ for $Al_2O_3 \cdot 3H_2O$, or $-1293\,334\,\mathrm{J\,mol^{-1}}$ for $Al(OH)_3$, and is from Hemingway et al. (1982). The difference is due to the use of different data fitting techniques and is not statistically significant.

Thermodynamic data are not written in stone, they are written in blood, so to speak. Many obscure errors are possible, but you should not change data to suit your latest theory.

difficult. In other cases, such as with virtually all organic compounds, no equilibrium path is possible, so using thermodynamics would be greatly inhibited. Until development and general acceptance of the third law, this was in fact the case.

This meant that the driving force for reactions, ΔG, could only be determined for substances for which such equilibrium paths could be used. Otherwise a "constant of integration" was always involved. Some method of determining ΔG using only thermal data was greatly desired. We should perhaps note that the emphasis here is on chemical reactions. You will commonly find statements to the effect that a reversible path is required to calculate entropy differences, but these statements are generalities, applicable to many processes such as the gas expansions in Chapter 3. An example of such a statement is found in the quotation from Schottky (1929) in Appendix G. As mentioned, such reversible paths are difficult or impossible to find for chemical reactions, the focus of a large part of chemistry and geochemistry, and it is in these cases that the third law is useful.

It was widely believed that determining the "constant of integration" for entropy was possible, beginning with LeChatelier in 1888. Contributions by Lewis (1899), Richards (1902), van't Hoff (1904), Haber (1905), Nernst (1906), and Planck (1912), not to mention Einstein's (1907) fundamental work on heat capacities and Boltzmann's development of an atomistic approach,[3] led to the general belief that the heat capacity, and perhaps entropy, became zero at absolute zero temperature. However, convincing calorimetric data supporting this idea, as well as confirming the limitations, only developed during the 1920s and 1930s. W.F. Giauque received the Nobel prize in Chemistry in 1949 for his lifelong contributions to our understanding of the third law.

The physics of materials at low temperatures is now a large and important topic, and a complete understanding of the third law requires some knowledge of statistical mechanics and even some quantum mechanics. A fairly brief overview is Wilks (1961). However, for those whose interests lie at the other end of Earth's temperature spectrum and are mainly interested in having accurate thermochemical data, the only important aspect of the third law is that it provides an absolute reference point for entropy data.

The structure of thermodynamics is based on the first and second laws, but it is the third law which allows the structure to be useful for chemical reactions. By far the most data on Gibbs energy differences and equilibrium constants has been through use of third law entropies. An illustration of the difference between the old "equilibrium method" and the use of the third law in determining the Gibbs energy difference between the two solid forms of sulfur is given in the box on page 126.

---

[3] References in Lewis and Randall (1923, Chapter 31).

Lewis and Randall (1914) point out that the change in Gibbs energy at 25 °C for the sulfur transition was also determined by Brönsted in 1906 using a completely different method. In separate experiments he dissolved both forms of sulfur in four different organic solvents. The solubilities of each form in each solvent were different, but the *ratio* [solubility of S(mono)/solubility of S(rhombic)] was constant, the average value being 1.28. He also determined that the sulfur in solution in each case was S_8. In §9.4.1 we will see that this is a special case of the equilibrium constant being equal to a solubility, and that $\Delta_r G^\circ = (RT/8) \ln [\text{solubility ratio}]$, resulting in $\Delta_r G^\circ = 18.3\,\text{cal mol}^{-1}$, roughly confirming the Lewis and Randall determination, probably within the combined uncertainty of the experimental results. The interest is not the value of $\Delta_r G^\circ$, but that it was necessary to use an equilibrium in the determination. Before the third law came into general use, there was no other way.

5.4.2 Statement of the third law

The statement of the third law by Lewis and Randall (1923) is still useful:

> If the entropy of each element in some crystalline state be taken as zero at the absolute zero of temperature: *every substance has a finite positive entropy, but at the absolute zero of temperature the entropy may become zero, and does so become in the case of perfect crystalline substances.*

The reference to perfect crystalline substances means that a nonzero entropy may be "frozen-in" at low temperatures, and is so in the case of glasses, gels, and various other cases of substances having some configurational disorder. The most important case for geologists is that of solid solutions, in which two or more atoms occupying a crystal lattice site may be disordered. This disorder is undoubtedly still present when the crystal is cooled down to cryogenic temperatures for heat capacity measurements, so the entropy does not approach zero at 0 k. This "residual entropy" must be calculated and added to the entropy evaluated from Equation (5.10). Also note the condition that the entropy of the (perfectly crystalline) elements is assumed to be zero. The heat capacity is certainly zero at 0 K, but all we can really say about entropy is that the entropy of all perfectly crystalline substances becomes the same at 0 K, and is called zero by convention (Melrose, 1970). Of course, giving it any other number would make no difference to equilibrium calculations.

If we let T_1 be absolute zero in Equation (5.9), the entropy of minerals at any temperature, say our standard temperature of 298.15 K, is (assuming no residual entropy)

$$S_{298} = \int_{T=0}^{T=298} \frac{C_P}{T}\, dT \tag{5.10}$$

and all that is required to determine "absolute" values for the entropy of minerals is to measure their heat capacity at a series of temperatures between zero and 298.15 K and to evaluate the integral. This gives rise to another kind of calorimetry, *cryogenic*, or low-temperature calorimetry.

5.4.3 Cryogenic calorimetry

A cryogenic calorimeter (Figure 5.5) is an apparatus designed for the determination of heat capacities at very low temperatures. The procedure is to cool the sample down to a temperature within a few degrees of absolute zero (a temperature of absolute zero itself is actually impossible to achieve, a fact actually implicit in the third law), introduce a known quantity of heat using an electrical heating coil, and observe the resulting increase in temperature (usually a few degrees). The quantity of heat is equal to ΔH, and this divided by difference gives an approximate value of C_P at the midpoint of the temperature range. Corrections are then made to compensate for heat leaks, for the heat absorbed by the calorimeter, and to get exact C_P values from the approximate ones. The integration of C_P/T values to obtain the entropy at 298.15 K is illustrated in Figure 5.6. A much more detailed description of the calorimeter and its operation is in Robie (1987).

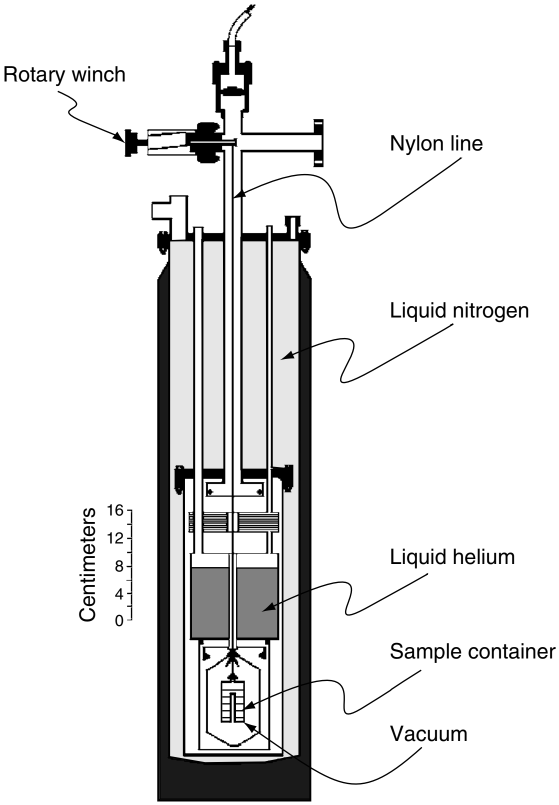


Figure 5.5 A cryogenic or low-temperature calorimeter. The sample container can be raised by the rotary winch so as to be in contact with the liquid helium reservoir for cooling to 4.2 K, or lowered into the vacuum for heating. The re-entrant well in the sample container contains a heating coil. (Simplified from Robie and Hemingway, 1972.)

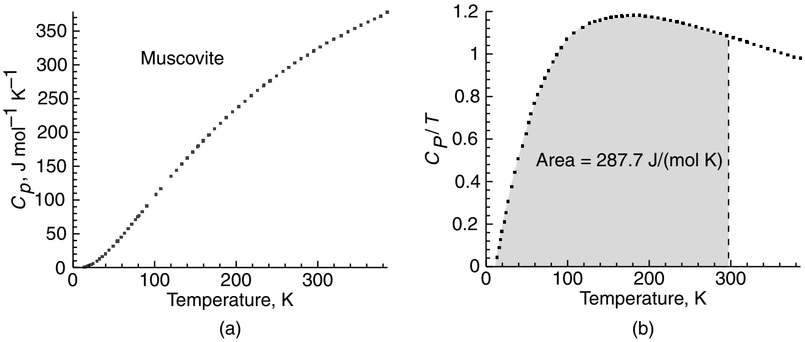


Figure 5.6 (a) Measured heat capacity of muscovite as a function of temperature (Robie et al., 1976). (b) C_P/T versus T for the same data. Integration gives the shaded area under the curve, which is equal to the entropy at the upper limit of integration, in this case, $S^\circ_{298.15} = 287.7\,\text{J}\,\text{mol}^{-1}$.

5.5 The problem resolved

We now know how to tell which way reactions will go, not just in a theoretical way (they will decrease their Gibbs energy at T, P), but in a practical way (how do we get this ΔG?). For example, SiO_2 comes in several crystalline varieties (polymorphs), such as the minerals quartz and cristobalite. They have the same composition SiO_2, but different crystallographic structures and energy contents, and one is stable and one is metastable at 25 °C, 1 bar. The problem is analogous to the diamond – graphite problem, and the reaction is

$$\text{SiO}_2^{\text{cristobalite}} = \text{SiO}_2^{\text{quartz}}$$

Which way does this reaction go at 25 °C, 1 bar? You could answer this question thermodynamically as follows:

1. Dissolve quartz in a solvent (HF acid) and measure the heat released.
2. Dissolve the same amount of cristobalite in the same amount of the same solvent and measure the heat released.
3. Because the solution after dissolution in the two cases has exactly the same composition and is identical in all respects, the difference in the two measured heat terms must be the difference in enthalpy between quartz and cristobalite, $\Delta_r H = H_{298}^{\text{quartz}} - H_{298}^{\text{cristobalite}}$ (remember, products minus reactants).
4. Measure the heat capacities of both quartz and cristobalite from near absolute zero to 298 K, and calculate S_{298}^{quartz} and $S_{298}^{\text{cristobalite}}$.
5. Subtract these two entropies to give $\Delta_r S = S_{298}^{\text{quartz}} - S_{298}^{\text{cristobalite}}$.
6. Calculate $\Delta_r G = \Delta_r H - 298.15 \cdot \Delta_r S$. If this is negative, quartz is stable; if it is positive, cristobalite is stable.

Sulfur and the third law of thermodynamics

The problem is which of the two crystalline forms of sulfur is more stable? This is equivalent to asking what is ΔG° for the reaction

$$\mathrm{S(rhomb)} = \mathrm{S(mono)}$$

This is the example used by Lewis and Randall (1923) to illustrate calculation of Gibbs energies.

Equilibrium method The only data available are the heat capacities of rhombic and monoclinic sulfur from 0 to 100 °C, and the enthalpy of transition between the two forms at 0 °C.

$$\Delta C_P^\circ = -0.50\,T + 0.0025\,T^2\,\mathrm{cal\,mol^{-1}\,K^{-1}}$$

$$\Delta_{\mathrm{trans}}H^\circ = 77.0\,\mathrm{cal\,mol^{-1}}$$

General integration of Equation (3.31) gives an integration constant on each side, which can be combined into one, and at 273.15 K ΔH° is 77.0 cal mol^{-1}, allowing calculation of the constant. General integration of

$$\frac{d\Delta S^\circ}{dT} = \frac{\Delta C_P^\circ}{T}$$

results in another integration constant, and combining these expressions with $\Delta G = \Delta H - T\Delta S$ gives

$$\Delta G^\circ = 120.3 + 0.50T\ln T - 0.001\,25\,T^2 + IT$$

The crucial point is that the equilibrium temperature at which $\Delta G^\circ = 0$ is known to be 95 °C, allowing I to be calculated. Thus ΔG° at 25 °C is calculated to be 17.5 cal mol^{-1}.

The third law method Modern data from the JANAF tables (Chase, 1998) give

$$S^{\mathrm{rhomb}} = 32.056\,\mathrm{J\,mol^{-1}}$$

$$S^{\mathrm{mono}} = 33.028\,\mathrm{J\,mol^{-1}}$$

$$\Delta_{\mathrm{trans}}H^\circ = 0.360\,\mathrm{kJ\,mol^{-1}}$$

(continued)

Then

$$\begin{aligned}\Delta G^\circ &= \Delta H^\circ - T\Delta S^\circ \\ &= 360 - 298.15(33.028 - 32.056) \\ &= 70.2\,\mathrm{J\,mol^{-1}} \\ &= 16.8\,\mathrm{cal\,mol^{-1}}\end{aligned}$$

Putting $\Delta G^\circ = 0$ the equilibrium temperature is *predicted* to be 97.2 °C.

You are no doubt quite sure that you don't have to do this incredible amount of work to answer such a simple question; there must be an easier way. Well, there is, but only because other people have already done this incredible amount of work, and lots more like it. In other words, you can look up the data in tables. However, you do not look up heats of solution.

The trick we have just used to get the difference in enthalpy between two minerals (that is, to dissolve them both and subtract the heats of solution) is a very useful way of determining heats of reaction, because many reactions proceed very slowly, or not at all, and so you cannot measure the heat of reaction directly in a calorimeter. You cannot measure the heat of reaction as cristobalite changes directly into quartz at 25 °C, because it never does – it is a truly metastable form of SiO_2. However, most minerals will dissolve fairly rapidly in some kind of solvent, providing an indirect means of getting their enthalpy differences.

To have tables of data that enable you to calculate $\Delta_r H$ for any reaction, it would seem that all you need to do is tabulate heats of solution. But if you think about this for a minute or so, you find that although fine in theory, this will not work well in practice. For one thing, you would need to tabulate heats of solution for all combinations of substances that might be of interest. That is, the heats of solution of gibbsite, corundum, and water separately are not enough – you need the heat of solution of corundum + water in a 1:1 ratio. But for other reactions, you would need corundum + water in other proportions. Then you would need to be sure that the solution compositions were identical, and given the variety and concentrations of solvents used, your database would soon become very large and unmanageable. This problem is resolved, of course, by using "formation from the elements" properties, as discussed in Chapter 3. This enables us to tabulate a single number for each property for each compound, and makes calculation of reaction deltas easy, at least for standard conditions.

5.5.1 Gypsum – anhydrite example

The change in enthalpy for any reaction between compounds for which there are formation-from-the-element data is given by a simple algebraic addition of these $\Delta_f H^\circ$ terms, because in balanced reactions, the elements always cancel out. To take another example, consider reaction (3.20) again

$$\mathrm{CaSO_4}(s)\ \text{(anhydrite)} + 2\,\mathrm{H_2O}(l) = \mathrm{CaSO_4\cdot 2H_2O}(s)\ \text{(gypsum)} \tag{5.11}$$

Both gypsum and anhydrite occur at the Earth's surface, and it is not always clear which is the stable phase. To determine the enthalpy change in this reaction, we consider the reactions in which each phase in the reaction is formed from its elements:

$$\begin{aligned}
\mathrm{Ca}(s) + \mathrm{S}(s) + 2\,\mathrm{O_2}(g) &= \mathrm{CaSO_4}(s); & \Delta_f H^\circ_{\text{anhydrite}} &= -1434.11\,\mathrm{kJ\,mol^{-1}}\\
\mathrm{H_2}(g) + \tfrac{1}{2}\mathrm{O_2}(g) &= \mathrm{H_2O}(l); & \Delta_f H^\circ_{\text{water}} &= -285.830\,\mathrm{kJ\,mol^{-1}}\\
\mathrm{Ca}(s) + \mathrm{S}(s) + 3\,\mathrm{O_2}(g) + 2\,\mathrm{H_2}(g) &= \mathrm{CaSO_4\cdot 2H_2O}(s); & \Delta_f H^\circ_{\text{gypsum}} &= -2022.63\,\mathrm{kJ\,mol^{-1}}
\end{aligned}$$

So for reaction (5.11) we have

$$\begin{aligned}
\Delta_r H^\circ &= \Delta_f H^\circ_{\text{gypsum}} - \Delta_f H^\circ_{\text{anhydrite}} - 2\,\Delta_f H^\circ_{\text{water}}\\
&= -2022.63 - (-1434.11) - 2\,(-285.830)\\
&= -16.86\,\mathrm{kJ\,mol^{-1}}
\end{aligned}$$

from which we see that the reaction between anhydrite and water to form gypsum is exothermic; that is, 16.86 kJ of heat would be released for every mole of anhydrite reacted.

As mentioned in Chapter 3, it is important to realize that this heat of reaction, $\Delta_r H^\circ$, is equal to the difference in the *absolute* enthalpies of the reactants and products – the enthalpies of the elements have nothing to do with it, because they all cancel out. Thus

$$\begin{aligned}
\Delta_r H^\circ &= \Delta_f H^\circ_{\text{gypsum}} - \Delta_f H^\circ_{\text{anhydrite}} - 2\,\Delta_f H^\circ_{\text{water}}\\
&= H^\circ_{\mathrm{CaSO_4\cdot 2H_2O}} - H^\circ_{\mathrm{Ca}} - H^\circ_{\mathrm{S}} - 2H^\circ_{\mathrm{H_2}} - 3H^\circ_{\mathrm{O_2}}\\
&\qquad - (H^\circ_{\mathrm{CaSO_4}} - H^\circ_{\mathrm{Ca}} - H^\circ_{\mathrm{S}} - 2H^\circ_{\mathrm{O_2}})\\
&\qquad - 2(H^\circ_{\mathrm{H_2O}} - H^\circ_{\mathrm{H_2}} - \tfrac{1}{2}H^\circ_{\mathrm{O_2}})\\
&= H^\circ_{\mathrm{CaSO_4\cdot 2H_2O}} - H^\circ_{\mathrm{CaSO_4}} - 2\,H^\circ_{\mathrm{H_2O}}\\
&= -16.86\,\mathrm{kJ\,mol^{-1}}
\end{aligned} \tag{5.12}$$

If you look carefully, you'll see that all the H° terms for the elements cancel out. But if you think before you look, you'll realize that they *must* cancel out if the reaction is balanced. Otherwise there is a mistake somewhere. Despite many statements to the contrary, the absolute enthalpies of the elements (H°_{Ca}, $H^\circ_{\mathrm{O_2}}$, etc.) are not assumed to be zero. There is no need to do so, because they all cancel out in balanced reactions.[4]

As you might expect, all the hard work of cryogenic calorimetry has already been done too, for most common substances, and the results are obtainable in tables and compilations of data. From Appendix B, we find that

$$\begin{aligned} S^\circ_{\mathrm{CaSO_4}}\ (\text{anhydrite}) &= 106.7\,\mathrm{J\,mol^{-1}\,K^{-1}} \\ S^\circ_{\mathrm{CaSO_4\cdot 2H_2O}}\ (\text{gypsum}) &= 194.1\,\mathrm{J\,mol^{-1}\,K^{-1}} \\ S^\circ_{\mathrm{H_2O}}\ (\text{water}) &= 69.91\,\mathrm{J\,mol^{-1}\,K^{-1}} \end{aligned}$$

and the entropy of reaction for (5.11) is

$$\begin{aligned} \Delta_r S^\circ &= S^\circ_{\text{gypsum}} - S^\circ_{\text{anhydrite}} - 2\,S^\circ_{\text{water}} \\ &= 194.1 - 106.7 - 2(69.91) \\ &= -52.42\,\mathrm{J\,mol^{-1}\,K^{-1}} \end{aligned}$$

So is gypsum or anhydrite stable?

Now that we have numerical values for both $\Delta_r H^\circ$ and $\Delta_r S^\circ$ for reaction (5.11), it is a simple matter to calculate $\Delta_r G^\circ$ to see which way the reaction goes. Our number for enthalpy is in kJ and that for entropy is in J, so we must convert one of them to be consistent. Converting kJ to J, we have

$$\begin{aligned} \Delta_r G^\circ &= \Delta_r H^\circ - T\,\Delta_r S^\circ \\ &= -16\,860 - 298.15(-52.42) \\ &= -1231\,\mathrm{J\,mol^{-1}} \end{aligned}$$

which is negative; therefore, gypsum is more stable than anhydrite in the presence of water at Earth surface conditions. We repeat that what we have found is that the *assemblage* of anhydrite plus water is metastable with respect to gypsum at 25 °C, 1 bar.[5] Anhydrite by itself is not metastable, as there is no other form of $\mathrm{CaSO_4}$ that has a lower energy.

---

[4] In the derivation of the enthalpies and Gibbs energies of aqueous ions, the assumption that the properties of the elements are zero is convenient, but even there it is not necessary (Chapter 15).

[5] Actually, we are not really sure whether it is metastable or unstable. The reaction between water and anhydrite to form gypsum is probably very, very slow at 25 °C, but much faster at slightly higher temperatures. This is an example of the discussion in §4.9.1. Metastable is a rather fuzzy concept in geology, but crystal clear in thermodynamics.

Gibbs energy From tables

Although we must do the calorimetry experiments in order to calculate free energy differences, there is usually no need to use $\Delta_r H^\circ$ and $\Delta_r S^\circ$ values from tables to calculate $\Delta_r G^\circ$. Values of $\Delta_f G^\circ$ for most compounds have been calculated and are also to be found in the same tables of data, and so we can use these values directly, instead of going through the $\Delta_r H^\circ - T\,\Delta_r S^\circ$ calculation.

For example, $\Delta_f G^\circ$ for anhydrite can be calculated from

$$\Delta_f G^\circ_{\mathrm{CaSO_4}} = \Delta_f H^\circ_{\mathrm{CaSO_4}} - T\,\Delta_f S^\circ_{\mathrm{CaSO_4}}$$

where $\Delta_f S^\circ_{\mathrm{CaSO_4}}$ is

$$\Delta_f S^\circ_{\mathrm{CaSO_4}} = S^\circ_{\mathrm{CaSO_4}} - S^\circ_{\mathrm{Ca}} - S^\circ_{\mathrm{S}} - 2\,S^\circ_{\mathrm{O_2}}$$

Don't forget that absolute entropies are obtainable for the elements just as well as for compounds, and these numbers are available in tables of data, such as Appendix B. These numbers are

Substance	S°, J mol^{-1} K^{-1}
$\mathrm{CaSO_4}(s)$	106.7
$\mathrm{Ca}(s)$	41.42
$\mathrm{S}(s)$	31.80
$\mathrm{O_2}(g)$	205.138

So

$$\begin{aligned}\Delta_f S^\circ_{\mathrm{CaSO_4}} &= 106.7 - 41.42 - 31.80 - 2 \times 205.138 \\ &= -376.796\,\mathrm{J\,mol^{-1}\,K^{-1}}\end{aligned}$$

Therefore, the Gibbs energy of formation of anhydrite is

$$\begin{aligned}\Delta_f G^\circ_{\mathrm{CaSO_4}} &= \Delta_f H^\circ_{\mathrm{CaSO_4}} - T\,\Delta_f S^\circ_{\mathrm{CaSO_4}} \\ &= -1434\,110 - 298.15(-376.796) \\ &= -1321\,768\,\mathrm{J\,mol^{-1}} \\ &= -1321.77\,\mathrm{kJ\,mol^{-1}}\end{aligned}$$

which is the number for $\Delta_f G^\circ$ in Appendix B ($-1321.79\,\mathrm{kJ\,mol^{-1}}$), within the limits of accuracy of the data.

The calculation for determining whether gypsum or anhydrite is stable is therefore a little easier – we just look up the $\Delta_f G^\circ$ numbers instead of both the $\Delta_f H^\circ$ and S° numbers. Thus

$$\begin{aligned}\Delta_r G^\circ &= \Delta_f G^\circ_{\mathrm{CaSO_4\cdot 2H_2O}} - \Delta_f G^\circ_{\mathrm{CaSO_4}} - 2\,\Delta_f G^\circ_{\mathrm{H_2O}} \\ &= -1797\,280 - (-1321\,790) - 2(-237\,129) \\ &= -1232\,\mathrm{J\,mol^{-1}} \qquad (5.13)\end{aligned}$$

which is what we got before ($-1231\,\mathrm{J\,mol^{-1}}$), within the limits of accuracy of the tabulated data.

Again, although we use the Gibbs energies of formation, the G values for the elements all cancel out, and what we calculate is the difference between the absolute Gibbs energies of the compounds in the reaction. Free energy and enthalpy are similar in this respect.

5.5.2 An aqueous organic example

To emphasize that our model is just as useful for organic or biochemical processes as for mineralogical ones, let's take another look at the reaction involving amino acids we considered in §2.6.1. Equations (2.3) and (2.5) are, to repeat,

$$\mathrm{C_8H_{16}N_2O_3}(aq) + \mathrm{H_2O}(l) = \mathrm{C_6H_{13}NO_2}(aq) + \mathrm{C_2H_5NO_2}(aq) \qquad (5.14)$$

$$\mathrm{C_6H_{13}NO_2}(aq) + \mathrm{C_2H_5NO_2}(aq) = 2\,\mathrm{H_2}(g) + 2\,\mathrm{NH_3}(g) + 4\,\mathrm{H_2O}(l) + 8\,\mathrm{C_{graphite}} \qquad (5.15)$$

From the tables in Appendix C, we find the following properties:

Substance	Formulas	$\Delta_f G^\circ$, $\mathrm{J\,mol^{-1}}$
leucine	$\mathrm{C_6H_{13}NO_2}(aq)$	−343 088
glycine	$\mathrm{C_2H_5NO_2}(aq)$	−370 778
leucylglycine	$\mathrm{C_8H_{16}N_2O_3}(aq)$	−462 834
hydrogen	$\mathrm{H_2}(g)$	0
ammonia	$\mathrm{NH_3}(g)$	−16 450
water	$\mathrm{H_2O}(l)$	−237 129
graphite	$\mathrm{C}(s)$	0

Therefore for reaction (5.14),

$$\begin{aligned}\Delta_r G^\circ &= \Delta_f G^\circ_{\mathrm{leucine}} + \Delta_f G^\circ_{\mathrm{glycine}} - \Delta_f G^\circ_{\mathrm{leucylglycine}} - \Delta_f G^\circ_{\mathrm{water}} \\ &= -343\,088 - 370\,778 - (-462\,834) - (-237\,129) \\ &= -13\,903\,\mathrm{J\,mol^{-1}}\end{aligned}$$

and for reaction (5.15)

$$\begin{aligned}\Delta_r G^\circ &= 2\,\Delta_f G^\circ_{\mathrm{H_2}(g)} + 2\,\Delta_f G^\circ_{\mathrm{NH_3}(g)} + 4\,\Delta_f G^\circ_{\mathrm{H_2O}(l)} + 8\,\Delta_f G^\circ_{\mathrm{C}(s)} \\ &\quad - \Delta_f G^\circ_{\mathrm{leucine}} - \Delta_f G^\circ_{\mathrm{glycine}} \\ &= 2\,(0) + 2\,(-16\,450) + 4\,(-237\,129) + 8\,(0) \\ &\quad - (-343\,088) - (-370\,778) \\ &= -267\,550\,\mathrm{J\,mol^{-1}}\end{aligned}$$

Thus we see that, as illustrated in Figure 2.8, both reactions have a negative "chemical energy," or $\Delta_r G^\circ$. However, to say any more about these reactions, we must emphasize a factor we have not yet mentioned, and which we cannot develop fully until Chapter 9 (§9.9).

The problem with solutions

It matters not a bit whether the substances we consider are organic or inorganic, stable or metastable, as long as we have data for them. But it matters a great deal whether they are pure substances (such as gypsum, quartz, diamond, liquid water, etc.) or are dissolved in some solvent, as with all substances designated (aq) in the tables. The problem is that the Gibbs energy (and all other properties) of a pure substance is a fixed and known quantity, but the Gibbs energy of a substance in solution depends on its concentration. The tabulated values of $\Delta_f G^\circ$ for (aq) substances are for one particular standard concentration. Therefore, although we have calculated a negative $\Delta_r G^\circ$ for our two reactions above, they both involve at least some dissolved substances and, therefore, the conclusion that the reactions should proceed spontaneously applies only when all the (aq) substances have the standard concentrations. We look more carefully at these problems in Chapters 7, 8, and 9.

Enzymes as catalysts

One more thing to note about chemical reactions is that living organisms have evolved mechanisms involving enzymes that overcome the energy barriers between reactants and products for reactions required by the organism. Such reactions, therefore, proceed easily and quickly, whereas in the inorganic world, diamond persists forever in its metastable state. No organism needs to change diamond to graphite, so no enzymes exist for this reaction. Living organisms also have mechanisms that drive some reactions "uphill," or against the Gibbs energy gradient. Thus peptide bonds are formed in organisms, as well as broken. The energy required to do this is obtained ultimately from the sun, but the exact mechanisms are complex. The study of such reactions forms a large part of the science of biochemistry.

5.6 Data at higher temperatures

Everything we have discussed so far is about determining data for "standard conditions," which usually means pure phases at 25 °C, 1 bar, although we will see later that it *can* mean something else. But as geologists we often deal with reactions at metamorphic and igneous temperatures of many hundreds of degrees. Specialized calorimeters can be used up to a few hundred degrees, but the experimental difficulties become great.

Obviously other methods are needed. As usual, there are several, but we will mention just two.

5.6.1 Drop calorimetry

The amount of heat required to raise the temperature of a mole of substance from T_{r} to T at constant pressure is simply $H_T - H_{T_{\mathrm{r}}}$ (or $H_T^\circ - H_{T_{\mathrm{r}}}^\circ$ for a standard reference substance); again, a difference between two unknown quantities. This quantity is conveniently determined by *cooling* the substance from T to T_{r} and measuring the amount of heat given up by the substance during this process.[6] To do this, a calorimeter is placed directly under a furnace and the sample is dropped from the furnace where it has temperature T_1, into the calorimeter, where it gives up its heat and achieves temperature T_2 (Figure 5.7). The amount of heat given up by the sample is determined by using this heat to melt a working substance in the calorimeter (either H_2O or diphenyl ether $(C_6H_5)_2O$), and measuring the volume change of this substance by the displacement of mercury. The relationship between the volume change and the ΔH of the solid→liquid phase transition (T_2 in the calorimeter is 273.15 K for H_2O; 303.03 K for diphenyl ether) is accurately known, so this amount of heat equals $H_{T_1} - H_{T_2}$. Small corrections are then applied using heat capacities to adjust this ΔH to $H_T - H_{T_{\mathrm{r}}}$, where T_{r} is invariably 298.15 K. More details of the method are given by Robie (1987). Experimental results for muscovite are shown in Figure 5.8.

Values of $H_T - H_{T_{\mathrm{r}}}$ can be combined to give $\Delta_f H^\circ$ for substances at high temperatures. Thus for any substance

$$\Delta_f H_T^\circ = \Delta_f H_{T_{\mathrm{r}}}^\circ + \Delta_f (H_T^\circ - H_{T_{\mathrm{r}}}^\circ) \tag{5.16}$$

---

[6] As a matter of fact, drop calorimetry has been largely superseded by differential scanning calorimetry (DSC) (§5.6.2), but I include a description here because it illustrates the acquisition of high temperature enthalpies and heat capacities more intuitively than does DSC, and because much of presently used data were obtained by this method.

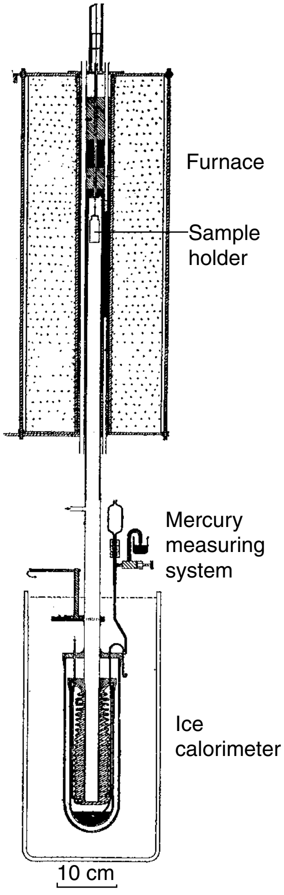


Figure 5.7 A drop calorimeter. (Simplified from Douglas and King (1968).)

where Δ_f refers to the reaction in which the substance is formed from its elements. For example,

$$\begin{aligned}\Delta_f(H^\circ_T - H^\circ_{T_r})_{\mathrm{SiO_2}} &= (H^\circ_T - H^\circ_{T_r})_{\mathrm{SiO_2}} - (H^\circ_T - H^\circ_{T_r})_{\mathrm{Si}} - (H^\circ_T - H^\circ_{T_r})_{\mathrm{O_2}}\\ &= (H^\circ_{T,\mathrm{SiO_2}} - H^\circ_{T,\mathrm{Si}} - H^\circ_{T,\mathrm{O_2}}) - (H^\circ_{T_r,\mathrm{SiO_2}} - H^\circ_{T_r,\mathrm{Si}} - H^\circ_{T_r,\mathrm{O_2}})\\ &= \Delta_f H^\circ_{T,\mathrm{SiO_2}} - \Delta_f H^\circ_{T_r,\mathrm{SiO_2}} \qquad (5.17)\end{aligned}$$

and therefore

$$\Delta_f H^\circ_{T,\mathrm{SiO_2}} = \Delta_f H^\circ_{T_r,\mathrm{SiO_2}} + \Delta_f(H^\circ_T - H^\circ_{T_r})_{\mathrm{SiO_2}} \qquad (5.18)$$

To get heat capacities from these measurements, the experimental values of $(H^\circ_T - H^\circ_{T_r})$ for the substance and its elements are first fitted to a function, which is commonly

$$(H^\circ_T - H^\circ_{T_r}) = A + BT + CT^2 + DT^{-1} \qquad (5.19)$$

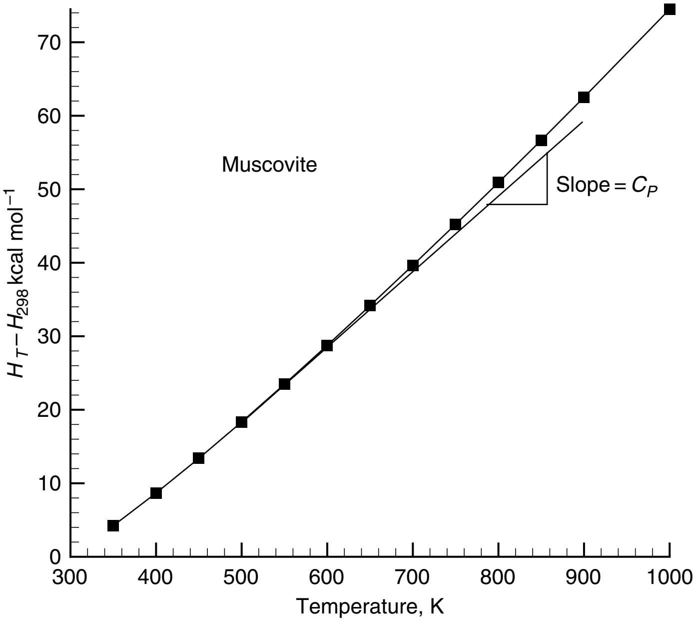


Figure 5.8 Values of $H_T - H_{298}$ for muscovite as measured in a drop calorimeter. The slope of the curve at any point equals the heat capacity at that temperature. Data from Pankratz (1964).

Once the "best fit" values of A, B, C, and D are calculated, $(H^\circ_T - H^\circ_{T_r})$ may be computed for any desired temperature. For example, the equation for the muscovite data in Figure 5.8 is

$$H^\circ_T - H^\circ_{298} = -38\,793 + 97.65\,T + 13.19 \times 10^{-3}\,T^2 + 25.44 \times 10^5\,T^{-1}$$

The heat capacity

A knowledge of how the quantity $H^\circ_T - H^\circ_{298}$ varies with T is useful because the first derivative, or the slope of the curve, is the heat capacity, C_P. As we have said, $H^\circ_{T_r}$ is an unknown quantity, but it is certainly a constant, so that

$$\frac{d}{dT}(H^\circ_{T_r}) = 0$$

Therefore

$$\begin{aligned}\frac{d}{dT}(H^\circ_T - H^\circ_{T_r}) &= \frac{d}{dT}(H^\circ_T)\\ &= C^\circ_P\\ &= B + 2CT - DT^{-2}\end{aligned}$$

or

$$C^\circ_P = a + bT - cT^{-2} \tag{5.20}$$

where

$$a = B,\ b = 2C, \text{ and } \ c = D.$$

Thus the Maier–Kelley coefficients for the muscovite data mentioned above are $a = 97.65$, $b = 26.38 \times 10^{-3}$, and $c = 25.44 \times 10^{5}$, and these are the values for muscovite in Helgeson et al. (1978).

We have carried the superscript ° throughout this derivation, so clearly we intend "standard conditions" to include high temperatures at times. In this case, it means simply that we are measuring some pure compound, rather than any arbitrary mixture, for which H and C_P would be more appropriate than H° and C_P°.

5.6.2 Differential scanning calorimetry

Determination of C_P by differentiating an experimental curve introduces an uncertainty greater than the uncertainty of the measurements themselves. A differential scanning calorimeter (DSC) measures C_P directly. In this method, a sample and a reference material are slowly heated simultaneously with separate heating elements (Figure 5.9). Care is taken to keep the temperature of each sample exactly the same, but because the samples are of different materials, the power delivered to each heater is different, and the difference is a direct function of the difference in the heat capacity of the two materials. Knowing the C_P of the reference material, the C_P of the sample may be determined. See Höhne et al. (1996) for an overview of the many different variations of differential scanning calorimetry.

Results for the C_P of muscovite from DSC measurements (Krupka et al., 1979) are compared to C_P calculated from the Maier–Kelley coefficients of Pankratz (1964) in Figure 5.10. The slight difference may be due to the fact

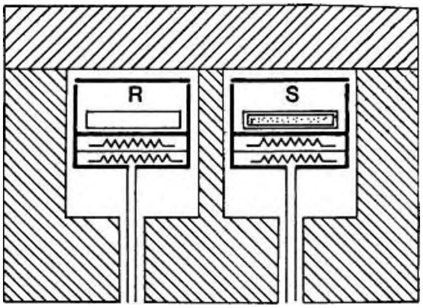


Figure 5.9 Schematic cross-section of a power compensated differential scanning calorimeter (modified from Robie (1987)). R – reference; S – sample. Under each sample pan is a platinum resistance thermometer and a platinum heater. The large metal block helps to keep the temperatures in the two chambers equal.

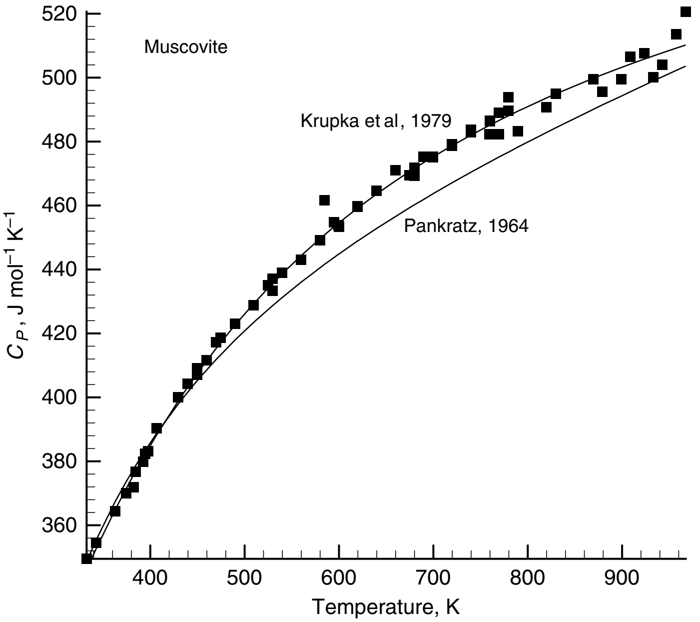


Figure 5.10 The heat capacity of muscovite, determined from drop calorimetry and from differential scanning calorimetry.

that the DSC measurements were corrected for the deviation of the sample composition from the stoichiometric formulas $KAl_2(AlSi_3O_{10})(OH)_2$, whereas the Pankratz measurements were not. The line through the data points of Krupka et al. was calculated using equation (3.28), with the muscovite coefficients from Berman and Brown (1985).

5.6.3 Entropies above 298 K

For temperatures above 298 K, entropies can be calculated by combining $S^\circ_{T_r}$ and the $(H^\circ_T - H^\circ_{T_r})$ measurements previously described. Since we know $S^\circ_{T_r}$, all we need are values of $S^\circ_T - S^\circ_{T_r}$, which equals $(H^\circ_T - H^\circ_{T_r})/T$.

Thus

$$\begin{aligned} d(S^\circ_T - S^\circ_{T_r}) &= dS^\circ_T \text{ (because } S^\circ_{T_r} \text{ is constant)} \\ &= d\left(\frac{H^\circ_T - H^\circ_{T_r}}{T}\right) \end{aligned}$$

so

$$\int_{T_r}^{T} dS^\circ = \int_{T_r}^{T} d\left(\frac{H^\circ_T - H^\circ_{T_r}}{T}\right)$$

The right-hand side is integrated by parts, giving

$$S^\circ_T - S^\circ_{T_r} = \left(\frac{H^\circ_T - H^\circ_{T_r}}{T}\right) + \int_{T_r}^{T} \left(\frac{H^\circ_T - H^\circ_{T_r}}{T^2}\right) dT$$

Since $(H^\circ_T - H^\circ_{T_\mathrm{r}})$ and therefore $(H^\circ_T - H^\circ_{T_\mathrm{r}})/T^2$ is known as a function of T, the integral can be evaluated, and S°_T values calculated for elevated temperatures. As in the case of ΔH° values, an alternative and usually preferable method is to calculate S°_T values, or more likely $\Delta_r S^\circ$ values, at elevated temperatures by means of the Maier–Kelley heat capacity coefficients. In other words, since

$$\frac{d}{dT}(S^\circ) = \frac{C^\circ_P}{T}$$

then

$$\frac{d}{dT}(\Delta_r S^\circ) = \frac{\Delta_r C^\circ_P}{T}$$

where $\Delta_r S^\circ$ refers to the entropy change of a balanced chemical reaction. Integrating,

$$\int_{T_\mathrm{r}}^{T} d\Delta_r S^\circ = \int_{T_\mathrm{r}}^{T} \frac{\Delta_r C^\circ_P}{T} dT \tag{5.21}$$

Combining this with the Maier–Kelley equation

$$\Delta C^\circ_P = \Delta a + \Delta b T - \Delta c T^{-2} \tag{5.22}$$

we have

$$\Delta_r S^\circ_T - \Delta_r S^\circ_{T_\mathrm{r}} = \int_{T_\mathrm{r}}^{T} \left(\frac{\Delta_r a}{T} + \Delta_r b - \frac{\Delta_r c}{T^3} \right) dT$$

or

$$\Delta_r S^\circ_T - \Delta_r S^\circ_{T_\mathrm{r}} = \Delta_r a \ln\left(\frac{T}{T_\mathrm{r}}\right) + \Delta_r b(T - T_\mathrm{r}) + \frac{\Delta_r c}{2}\left(\frac{1}{T^2} - \frac{1}{{T_\mathrm{r}}^2}\right) \tag{5.23}$$

In this equation $\Delta_r S^\circ_T$ refers to the entropy change of any balanced chemical reaction at temperature T. If the reaction is the formation of a compound from its elements, $\Delta_r S^\circ_T$ becomes $\Delta_f S^\circ_T$.

The apparent entropy of formation can be calculated from

$$\Delta_a S^\circ_T = \Delta_f S^\circ_{T_\mathrm{r}} + \int_{T_\mathrm{r}}^{T} \frac{C^\circ_P}{T} dT$$

where C°_P/T refers to the compound only.

5.6.4 Gibbs energies above 298 K

Standard Gibbs energies of formation from the elements at 298 K are computed from

$$\Delta_f G^\circ_{T_\mathrm{r}} = \Delta_f H^\circ_{T_\mathrm{r}} - T_\mathrm{r} \Delta_f S^\circ_{T_\mathrm{r}} \tag{5.24}$$

and at higher temperatures from

$$\Delta_f G^\circ_T = \Delta_f H^\circ_T - T\Delta_f S^\circ_T$$

Using the heat capacity approach, for apparent Gibbs energies at higher temperatures for a compound i

$$\begin{aligned}\Delta_a G^\circ_{T,i} &= \Delta_f G^\circ_{T_\mathrm{r},i} + \int_{T_\mathrm{r}}^{T} (\partial G^\circ_i/\partial T)dT \qquad (5.25)\\ &= \Delta_f G^\circ_{T_\mathrm{r},i} + \int_{T_\mathrm{r}}^{T} -S^\circ_i\, dT\\ &= \Delta_f G^\circ_{T_\mathrm{r},i} + \int_{T_\mathrm{r}}^{T} \left[-S^\circ_{T_\mathrm{r},i} - \int_{T_\mathrm{r}}^{T} \frac{C^\circ_{Pi}}{T} dT\right] dT\\ &= \Delta_f G^\circ_{T_\mathrm{r},i} - S^\circ_{T_\mathrm{r},i}(T - T_\mathrm{r}) - \int_{T_\mathrm{r}}^{T} \left[\int_{T_\mathrm{r}}^{T} \frac{C^\circ_{Pi}}{T} dT\right] dT\end{aligned}$$

The integration in the last term is performed by parts. That is

$$\int u\, dv = uv - \int v\, du$$

where

$$u = \int_{T_\mathrm{r}}^{T} \frac{C^\circ_{Pi}}{T} dT \quad \text{and} \quad v = T$$

This results in

$$\Delta_a G^\circ_{T,i} = \Delta_f G^\circ_{T_\mathrm{r},i} - S^\circ_{T_\mathrm{r},i}(T - T_\mathrm{r}) + \int_{T_\mathrm{r}}^{T} C^\circ_{Pi}\, dT - T\int_{T_\mathrm{r}}^{T} \frac{C^\circ_{Pi}}{T} dT \qquad (5.26)$$

which, after substitution of

$$\int_{T_\mathrm{r}}^{T} C^\circ_P dT = a(T - T_\mathrm{r}) + \frac{b}{2}(T^2 - T_\mathrm{r}^2) + c\left(\frac{1}{T} - \frac{1}{T_\mathrm{r}}\right) \qquad (5.27)$$

and

$$\int_{T_\mathrm{r}}^{T} \frac{C^\circ_P}{T} dT = a \ln\left(\frac{T}{T_\mathrm{r}}\right) + b(T - T_\mathrm{r}) + \frac{c}{2}\left(\frac{1}{T^2} - \frac{1}{T_\mathrm{r}^2}\right) \qquad (5.28)$$

and collection of terms, results in

$$\begin{aligned}\Delta_a G^\circ_{T,i} &= \Delta_f G^\circ_{T_\mathrm{r},i} - S^\circ_{T_\mathrm{r},i}(T - T_\mathrm{r})\\ &\quad + a_i\left[T - T_\mathrm{r} - T \ln\left(\frac{T}{T_\mathrm{r}}\right)\right]\\ &\quad + \frac{b_i}{2}\left(2TT_\mathrm{r} - T^2 - {T_\mathrm{r}}^2\right) \qquad (5.29)\\ &\quad + \frac{c_i\left(T^2 + T_\mathrm{r}^2 - 2TT_\mathrm{r}\right)}{2TT_\mathrm{r}^2}\end{aligned}$$

For a mineral reaction, $\Delta_r G^\circ$ is obtained by substituting $\Delta_r S^\circ_{T_\mathrm{r}}$ for $S^\circ_{T_\mathrm{r},i}$ and $\Delta_r a$, $\Delta_r b$, and $\Delta_r c$ for a_i, b_i, and c_i, where $\Delta_r a$, etc. are the usual product−reactant terms.
Thus

$$\begin{aligned}\Delta_r G^\circ_T = {} & \Delta_r G^\circ_{T_\mathrm{r}} - \Delta_r S^\circ_{T_\mathrm{r}} (T - T_\mathrm{r}) \\ & + \Delta_r a \left[T - T_\mathrm{r} - T \ln \left(\frac{T}{T_\mathrm{r}} \right) \right] \\ & + \frac{\Delta_r b}{2} \left(2TT_\mathrm{r} - T^2 - T_\mathrm{r}{}^2 \right) \\ & + \frac{\Delta_r c \left(T^2 + T_\mathrm{r}^2 - 2TT_\mathrm{r} \right)}{2TT_\mathrm{r}^2}\end{aligned} \tag{5.30}$$

A special case would be the reaction in which i is formed from its elements, all at T, in which case $\Delta_r S^\circ_{T_\mathrm{r}}$ becomes $\Delta_f S^\circ_{T_\mathrm{r},i}$, $\Delta_r a$ becomes $\Delta_f a_i$, and so on, and Equation (5.30) then gives $\Delta_f G^\circ$, the traditional Gibbs energy of formation as a function of T for compound i, but this would perhaps only be of interest to compare with older data compilations which use this quantity.

High temperature muscovite free energies

Equation (5.29) is used to calculate Gibbs energy values in many programs, including SUPCRT92. To illustrate this consider muscovite again. The required information is shown in Table 5.1. We have by now seen how all of these quantities were (or might have been) derived.

Table 5.2 shows the result of plugging the values from Table 5.1 into Equation (5.29), compared to the values from SUPCRT92. The values are identical up to 100 °C, but then begin to deviate. This is because the values from SUPCRT92 are not all at 1 bar. Above 100 °C they are at the pressure of the water – steam equilibrium. Clearly this pressure difference has an effect on the Gibbs energy, to be discussed next.

Table 5.1 *Data for muscovite for Equation (5.29), plus the molar volume.*

$\Delta_f G^\circ_{T_\mathrm{r}}$	−1336 301	$\mathrm{cal\,mol^{-1}}$
$S^\circ_{T_\mathrm{r}}$	68.8	$\mathrm{cal\,mol^{-1}\,K^{-1}}$
a	97.56	$\mathrm{cal\,mol^{-1}}$
b	0.026 38	$\mathrm{cal\,mol^{-1}\,K^{-1}}$
c	2 544 000	$\mathrm{cal\,mol^{-1}\,K^{2}}$
V	140.7	$\mathrm{cm^3\,mol^{-1}}$

Table 5.2 *The apparent Gibbs energy of formation of muscovite, from* SUPCRT92 *and from Equation (5.29). Data in cal mol^{-1}.*

T°C	P bars	SUPCRT92	Equation (5.29)	Difference	P correction $= V(P-1)$
0.01	1.000	−1334 662	−1334 662	0	0
25	1.000	−1336 301	−1336 301	0	0
50	1.000	−1338 101	−1338 101	0	0
75	1.000	−1340 059	−1340 059	0	0
100	1.013	−1342 171	−1342 171	0	0
125	2.320	−1344 428	−1344 432	4	4
150	4.757	−1346 825	−1346 838	13	13
175	8.918	−1349 357	−1349 383	26	27
200	15.536	−1352 015	−1352 063	48	49
225	25.479	−1354 792	−1354 874	82	82
250	39.736	−1357 680	−1357 811	131	130
275	59.431	−1360 673	−1360 869	196	197
300	85.838	−1363 760	−1364 046	286	285
325	120.458	−1366 935	−1367 337	402	402
350	165.211	−1370 187	−1370 739	552	552

5.7 Data at higher pressures

Fundamentally, to know ΔG° at any T and P, we need to be able to evaluate

$$\Delta_a G^\circ_{T,P,i} = \Delta_f G^\circ_{T_\mathrm{r},P_\mathrm{r},i} + \int_{T_\mathrm{r}}^{T} (\partial G^\circ_i/\partial T)_{P=P_\mathrm{r}}\, dT + \int_{P_\mathrm{r}}^{P} (\partial G^\circ_i/\partial P)_{T=T}\, dP \qquad (5.31)$$

and there are similar equations for enthalpy and entropy. The integrals in Equation (5.31) take care of the change in G with T and P, respectively, and we have just seen how the temperature integral is handled in Equation (5.25). The second integral shows that we need to know how G varies with pressure at high temperatures.

In this section we will discuss only the effect of pressure on solid phases, i.e., minerals. The evaluation of the pressure integral is done in quite a different way for gases, water, and aqueous solutes, and will be treated in later chapters.

5.7.1 Effect of P on Gibbs energy

As shown previously (Equation 4.43), the derivative of G with respect to P is V, i.e.,

$$(\partial G_i/\partial P)_T = V_i \qquad [4.43]$$

so that to calculate the effect of P on G_i we must know how V_i varies as a function of P.

Constant V

When substance i is a solid phase and thus has relatively small variation of V with both P and T (relative that is to liquids and gases), the errors introduced by the assumption that V is not affected by P or T tend to cancel one another, and very little error is introduced by assuming that V_i is a constant at all P, T values. As a result, the assumption of constant V for solids is often adopted for minerals, and results in

$$\begin{aligned}\int_{P_\mathrm{r}}^{P} (\partial G_i/\partial P)_T dP &= G_{i,P} - G_{i,P_\mathrm{r}} \\ &= V_i(P-1) \end{aligned} \tag{5.32}$$

where V_i is the molar volume of the solid phase in J bar^{-1}, and $P_\mathrm{r} = 1$ bar. Determination of the molar volume can be done experimentally, but is more commonly calculated from X-ray determination of the mineral structure. This term $V_i(P-1)$ is then added to the right side of Equation (5.29) to evaluate $\Delta_a G^\circ$ at high pressures as well as high temperatures, giving

$$\begin{aligned}\Delta_a G^\circ_{T,i} = \Delta_f G^\circ_{T_\mathrm{r},i} - S^\circ_{T_\mathrm{r},i}(T-T_\mathrm{r})a_i\left[T - T_\mathrm{r} - T\ln\left(\frac{T}{T_\mathrm{r}}\right)\right] \\ + \frac{b_i}{2}\left(2TT_\mathrm{r} - T^2 - T_\mathrm{r}{}^2\right) + \frac{c_i\left(T^2 + T_\mathrm{r}^2 - 2TT_\mathrm{r}\right)}{2TT_\mathrm{r}^2} + V_i(P-1)\end{aligned} \tag{5.33}$$

An example of this for muscovite is shown in Table 5.2.

If we are dealing with a mineral reaction (that is, a reaction involving only solid phases for which (4.43) is valid) instead of a pure compound, we substitute $\Delta_r S^\circ_{T_\mathrm{r}}$ for $S^\circ_{T_\mathrm{r},i}$ and $\Delta_r a$, $\Delta_r b$, and $\Delta_r c$ for a_i, b_i, and c_i, where $\Delta_r a$, etc. are the usual product–reactant terms, just as we did in Equation (5.30). Equation (5.33) then becomes

$$\begin{aligned}\Delta_r G^\circ_{T,i} = \Delta_r G^\circ_{T_\mathrm{r},i} - \Delta_r S^\circ_{T_\mathrm{r},i}(T-T_\mathrm{r}) + \Delta_r a_i\left[T - T_\mathrm{r} - T\ln\left(\frac{T}{T_\mathrm{r}}\right)\right] \\ + \frac{\Delta_r b_i}{2}\left(2TT_\mathrm{r} - T^2 - T_\mathrm{r}{}^2\right) + \frac{\Delta_r c_i\left(T^2 + T_\mathrm{r}^2 - 2TT_\mathrm{r}\right)}{2TT_\mathrm{r}^2} + \Delta_r V_i(P-1)\end{aligned} \tag{5.34}$$

Variable V

Alternatively, some attempt at modeling the temperature and pressure effects on mineral volumes can be attempted (see Helgeson et al., 1978, and Berman, 1988 for lengthy discussions of this topic). The only attempt at this to find its way into a widely used database is that of Berman (1988). Berman fit the available data for rock-forming minerals to the expression

$$\frac{V_{P,T}}{V_{P_\mathrm{r},T_\mathrm{r}}} = 1 + v_1(P-P_\mathrm{r}) + v_2(P-P_\mathrm{r})^2 + v_3(T-T_\mathrm{r}) + v_4(T-T_\mathrm{r})^2 \tag{5.35}$$

where $v_1 - v_4$ are fit parameters, $P_\mathrm{r} = 1$ bar, and $T_\mathrm{r} = 298.15$ K. The equation has no theoretical basis, and Berman cautions against its use at conditions beyond those used in the derivation of the parameters. Nevertheless, it fits available data quite well, and should lead to improved accuracy of phase equilibrium calculations.

To give some idea of the volume changes involved using Equation (5.35), Figure 5.11 shows the % change in the molar volume of muscovite over the range of P and T of validity of the fit parameters. The maximum change is less than 2%, and over a wide range of metamorphic conditions it is much less than that, so it may seem that it is not worth bothering with. However, in reactions involving only solids, both $\Delta_r V$ and $\Delta_r S$ will be small. These quantities represent the slopes of $\Delta_r G$ versus P and T, so small errors in either $\Delta_r V$ or $\Delta_r S$ will mean relatively large errors in $\Delta_r G$ and hence in the T and P of computed phase transitions.

To get the change in G from P_r, T to P,T, we need $\int V\,dP$ from P_r to P at temperature T, where V is given by (5.35). Thus

$$\int_{T,P_\mathrm{r}}^{T,P} V\,dP = \int_{T,P_\mathrm{r}}^{T,P} V_{T_\mathrm{r},P_\mathrm{r}}\left[1 + v_1(P-P_\mathrm{r}) + v_2(P-P_\mathrm{r})^2 + v_3(T-T_\mathrm{r}) + v_4(T-T_\mathrm{r})^2\right]dP$$

Integrating and collecting terms then results in

$$\int_{T,P_\mathrm{r}}^{T,P} V\,dP = V_{T_\mathrm{r},P_\mathrm{r}}\Big[\left\{1 - v_1 + v_2 + v_3(T-T_\mathrm{r}) + v_4(T-T_\mathrm{r})^2\right\}(P-P_\mathrm{r}) + \left(\frac{v_1}{2} - v_2\right)(P^2 - P_\mathrm{r}) + \frac{v_2}{3}(P^3 - P_\mathrm{r})\Big] \quad (5.36)$$

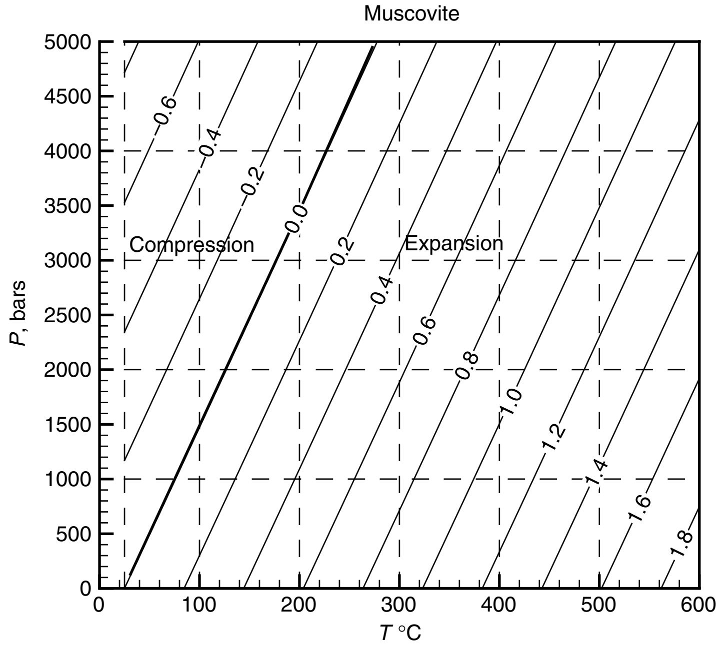


Figure 5.11 The % change in the molar volume of muscovite as a function of P and T, from Equation (5.35).

The right side of Equation (5.36) could be substituted for $V_i(P-1)$ in Equation (5.33). Realistically, however, it should probably be used only with the Berman (1988) database of mineral properties, as it was derived using this formula.

The Berman–Brown convention

We mentioned in Chapter 3 the two conventions used to define "apparent" properties, which are used as a form of "formation from the element" properties at temperatures above 298.15 K. The Benson–Helgeson convention was described there. The Berman–Brown convention, although the same for ΔH°, is different for ΔG°.

In the Berman–Brown convention, the apparent Gibbs energy, $\Delta_a G^\circ$, is defined as

$$\Delta_a G^\circ_{P,T} = \Delta_a H^\circ_{P,T} - T \cdot S^\circ_{P,T} \tag{5.37}$$

Notice the difference between this equation and Equation (5.1) or Equation (5.24), that is, it uses S°, not ΔS°. In other words, the entropies of the elements at 298.15 K are omitted because they are not needed to define S°. However, it means that while Benson–Helgeson values of $\Delta_a G^\circ$ are the same as traditional $\Delta_f G^\circ$ values at 298.15 K, 1 bar, Berman–Brown values of $\Delta_a G^\circ$ are not.

If you integrate Equation (5.31) using both the Berman and Brown (1985) heat capacity Equation (3.28) and the Berman–Brown convention for $\Delta_a G^\circ$, you find

$$\begin{aligned}
\Delta_a G^\circ_{P,T} = {} & \Delta_f G^\circ_{P_\mathrm{r},T_\mathrm{r}} - TS^\circ_{P_\mathrm{r},T_\mathrm{r}} \\
& + k_0\left[(T-T_\mathrm{r}) - T(\ln T - \ln T_\mathrm{r})\right] \\
& + 2k_1\left[(T^{0.5} - T_\mathrm{r}{}^{0.5}) + T(T^{-0.5} - T_\mathrm{r}{}^{-0.5})\right] \\
& - k_2\left[(T^{-1} - T_\mathrm{r}{}^{-1}) - \frac{T}{2}(T^{-2} - T_\mathrm{r}{}^{-2})\right] \\
& - k_3\left[\frac{T^{-2} - T_\mathrm{r}{}^{-2}}{2} - \frac{T}{3}(T^{-3} - T_\mathrm{r}{}^{-3})\right] \\
& + V_{P_\mathrm{r},T_\mathrm{r}}\left[\left(\frac{v-1}{2} - v_2\right)(P^2 - P_\mathrm{r}{}^2) + \frac{v_2}{3}(P^3 - P_\mathrm{r}{}^3)\right. \\
& \left. + \left\{1 - v - 1 + v_2 + v_3(T-T_\mathrm{r}) + v_4(T-T_\mathrm{r})^2\right\}(P-P_\mathrm{r})\right]
\end{aligned} \tag{5.38}$$

which is the equivalent of SUPCRT92's Equation (5.33). The numbers for individual minerals these two equations produce are quite different, but for balanced reactions they are very nearly the same.

5.7.2 Effect of *P* on enthalpy and entropy

For most geochemical modeling purposes, only the pressure effect on G is required. The effect of pressure on S is[7]

$$\left(\frac{\partial S}{\partial P}\right)_T = -\left(\frac{\partial V}{\partial T}\right)_P \tag{5.39}$$

showing that the effect of pressure on entropy can be obtained by measuring the effect of temperature on volume (or density), which is usually a much simpler task. Fairly simple manipulations then show the effect of pressure on enthalpy to be

$$\left(\frac{\partial H}{\partial P}\right)_T = V - T\left(\frac{\partial V}{\partial T}\right)_P \tag{5.40}$$

Therefore the effect of pressure on H and S using the constant molar volume assumption ($(\partial V/\partial T)_P = 0$) is particularly simple. Equation (5.39) shows that there is no effect on S, and integration of Equation (5.40) with $(\partial V/\partial T)_P = 0$ shows that the pressure effect on H is the same as that on G, that is, $V(P-1)$.

Equations (5.39) and (5.40) result in particularly simple expressions for an ideal gas, which are often useful. Substituting RT/P for V in Equation (5.39) leads to

$$\begin{aligned}\Delta S &= S_{P_2} - S_{P_1}\\ &= R\ln\frac{P_2}{P_1}\end{aligned} \tag{5.41}$$

and making the same substitution in (5.40) results in

$$\begin{aligned}\Delta H &= H_{P_2} - H_{P_2}\\ &= 0\end{aligned} \tag{5.42}$$

These results should make intuitive sense. Changing the pressure on an ideal gas does not change the fact that there are no interparticle forces, so there should be no effect of pressure on energy terms. However, it does change the ordering or arrangement of the particles, and hence the entropy.

5.8 Other methods

Up to here in this chapter we have discussed getting data only for solid phases. We have seen that the goal is to determine $\Delta_a G^\circ$ over a wide range of T and P (and that $\Delta_a G^\circ = \Delta_f G^\circ$ at 25 °C, 1 bar for one widely used convention, but

---

[7] This equation is obtained by applying the reciprocity relation (Equation C.10) to Equation (4.40).

not for another). Doing this requires integration of Equation (5.31), which in turn requires that we know how V and C_P vary with T and P. For solids, V is either considered a constant, or its slight variation can be estimated, and C_P is measured directly (§5.6.2) or indirectly (§5.6.1) by calorimetry.

For gases and liquids the fundamentals are the same (we still must integrate (5.31)), but methods are quite different because V is far from constant. For dissolved substances, more difficulties arise because solution properties vary with the concentration of the solute, and for electrolytes, there is also a variable degree of association of the charged particles. We will mention a few experimental methods here, but we cannot discuss the data obtained until we learn more about how we deal with the properties of dissolved substances.

5.8.1 Density measurement

The vibrating tube densitometer is an instrument designed to precisely measure liquid or gas density. A tube or spool is vibrated mechanically at its natural frequency and sensors measure the frequency of that vibration. The measured frequency will decrease as density increases, and can be calculated after calibrating the instrument with fluids of known density. Other density related variables such as specific gravity, molecular weight, and concentration may also be calculated. It is now very widely used in industry as well as scientific laboratories. See, e.g., Majer et al. (1991).

A more direct method is provided by a dilatometer, basically a pressure vessel connected to a mercury reservoir and a glass capillary column. Changes in the height of the mercury in the column are directly related to the change in density of the fluid in the pressure vessel. The apparatus is calibrated using pure water, with its known volumetric properties. Of course, no brief description can give any idea of the multitude of details of operation and correction factors that go into making precise measurements.

5.8.2 Calorimetry

Calorimetry of liquids and solutes has been revolutionized in recent years by the combination of the differential scanning technique, in which some difference between a sample and a standard is observed, with the continuous flow of fluids through the calorimeter. Instead of having two mineral samples (§5.6.2), two columns or tubes are used, through which a reference solution and a sample solution flow at a controlled rate (Figure 5.12). As before, the difference in the power required to keep the columns at the same temperature is directly related to the difference in the heat capacities of the two fluids. See Wood (1989) for a history of the development of these methods and their advantages.

Heat capacities

Figure 5.12 shows a schematic cross-section through a twin-tube flow calorimeter. The shaded area represents a large copper block with two cavities, through which pass two thin-walled tubes, each with a heating coil. On the left is the reference tube or cell, and on the right is the working cell. The pump delivers fluid at a known rate, and the back-pressure regulator allows the escape of fluid at a fixed pressure. A six-port, two-position valve controls the fluid flow. In one position, fluid (pure water) from the pump passes through both cells, establishing the base-line operation; in the other position the water is diverted into the sample loop, pushing 10 cm^3 of sample solution from the sample loop into the working cell. The different heat capacity of the sample solution causes a change in the power delivered to the working cell heater to maintain the temperatures of the two cells at the same value. The ratio of the power required to keep the working cell temperature at the baseline value to the power required during the establishment of the baseline is directly proportional to the ratio of the heat capacities of water and sample, multiplied by the ratio of the mass flow rates.

Heat of solution/dilution

Without knowing the exact reasons, to be discussed in Chapter 10, it seems natural that we would need to know the heat change when a solute dissolves

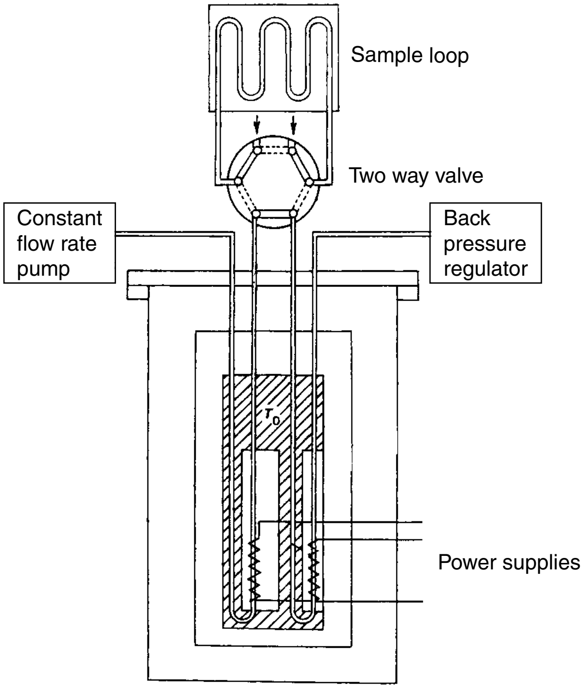


Figure 5.12 Schematic cross-section of the twin-flow calorimeter of Smith-Magowan and Wood (1981). Modified from Robie (1987).

into a fluid, such as NaCl dissolving in water, i.e., calorimetric heats of solution, either solid into liquid, or liquid into liquid, i.e., mixing of liquid solutions.

A practical consideration in conducting this kind of heat of solution measurements is that it is frequently much simpler and more accurate to measure heat flows associated with diluting solutions, rather than with making them more concentrated. Dissolving NaCl in a concentrated solution of NaCl in water is a rather slow process, so that the heat is absorbed over a long period of time, and it is difficult to know when the process is complete. Adding water to a concentrated NaCl solution is on the other hand accomplished fairly quickly and the same information can be deduced from these dilution measurements as from "solution" measurements. Of course, what is "dilution" and what is "solution" is to some degree quite arbitrary – a "heat of dilution" on adding water to NaCl solutions could be considered a "heat of solution" of water in the system NaCl–H_2O. Normally, however, water is considered the solvent, and adding water is called dilution. Heat of dilution measurements have also benefited from the twin-tube differential method. One such instrument is described by Busey et al. (1984).

5.8.3 Conductance measurements

Electrolyte solutions are those in which the solute (e.g., NaCl) dissociates into charged particles called ions (e.g., Na^+ and Cl^-). Naturally, such solutions are much better conductors of electricity than is the solvent alone (normally water), and the conductivity of the solution will depend on the extent to which the dissociation takes place. That is, if only some of the NaCl dissociates into Na^+ and Cl^-, the solution will be less conducting than if all of it does. Measurement of conductivity is therefore a means of determining the degree of dissociation of solutes. Significant improvements in the speed and accuracy of conductance measurements has been achieved by using a flow-through cell (Zimmerman et al. 1995).

5.8.4 Isopiestic measurements

Isopiestic means, roughly, equal pressures. In Figure 5.13 a number of containers contain solutions of various concentrations of a salt A, enclosed in a sealed, evacuated chamber. One of the containers contains a standard solution of a salt B, for which the properties are accurately known. As the solutions all have different vapor pressures, water will slowly evaporate from some and condense in others until all containers have exactly the same vapor pressure. The properties of the salt and the water in each solution are related in an interesting way by thermodynamics. Because you know the properties of the standard solution, and now all solutions have the same water vapor pressure, it is possible to calculate the properties (well, one property anyway) of salt A as

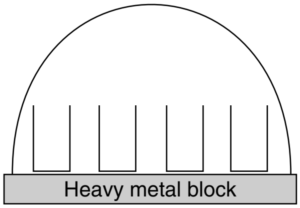


Figure 5.13 Isopiestic measurement in which sample solutions containing various known weights of salt A are equilibrated with a standard solution of salt B.

a function of concentration by analyzing the (changed) concentrations of all the containers. How this is done is discussed in §4.14.2 and §10.9. An overview of the isopiestic method is given in Rard and Platford (1991).

5.9 Summary

This chapter contains the transition from somewhat abstract theory to usable numbers. The Gibbs energy and enthalpy are forms of energy, closely related to the "energy in the deep pond," **U** (Figure 3.2). Energy can be transferred by heat and/or work, and assuming only mechanical ($P\,\Delta V$) work is involved, the fundamental properties we need to know in order to know the energy change are the thermal and volumetric properties ΔH (or more fundamentally C_P) and V.

Thermal properties are measured by some form of calorimetry, an exacting experimental procedure in which some kind of reaction is carried out, such as dissolution of a solid phase, and the heat (q) released or absorbed is measured. If the reaction occurred at constant pressure, the measured q is a ΔH, and if not, it is fairly easily converted into a ΔH. Entropy can also be measured by calorimetry, though of a different type, and combining the enthalpy and entropy measurements gives ΔG numbers. Values of ΔG° can also be obtained by other methods, to be discussed in later chapters. All these quantities are related to the heat capacity, which turns out to be a very fundamental and important parameter. If pressure changes are important, then the volume or density is also required.

The use of these concepts in modern computer programs adds some complications which, although not required to understand thermodynamics itself, are required to understand how the programs use data. These complications include the choice of algorithm to represent heat capacity as a function of temperature, how to represent the effect of pressure, and the various conventions for "formation from the elements" quantities.

At this point, much of the theory and practice of chemical thermodynamics has been presented. It is worth pausing to reflect on just how it is that delicate measurements near absolute zero temperature, combined with a bunch of differential equations which refer to unattainable conditions, are essential in deciphering the origins of ore deposits, metamorphic rocks, and other geological phenomena.

6

Some simple applications

6.1 Introduction

We now know how to determine in which direction any chemical reaction will proceed at a given temperature and pressure, at least when all the products and reactants are pure phases. When even one of the products or reactants is a solute, that is, part of a solution, we would be stuck because although we have had a brief look at how calorimetry can be used with liquids and liquid solutions, we haven't yet seen how to use the data obtained. We will start considering this problem in the next chapter. Before going on, however, we should explore some relationships using the concepts we have defined so far, so as to make sure we fully understand them. Naturally, we will only be able to consider some simple properties of pure phases, and reactions between pure phases.

6.2 Some properties of water

Water is an extraordinarily important substance in many ways in Earth processes, both organic and inorganic. The fact that it is a liquid in the relatively narrow range of T and P found at the Earth's surface was of course essential to the evolution of life. The reason for this, that is, the polar structure of the H_2O molecule and its many unusual properties such as hydrogen bonding, is a fascinating subject in physical chemistry. In geochemical thermodynamics, we are concerned only with its macroscopic properties and its role in mineral reactions, both at and near the surface and at depth in the crust, that is, the weathering, metamorphic, and igneous environments.

So what macroscopic properties do we mean? Well, those that are connected with heat ($\boldsymbol{q}$) and work ($\boldsymbol{w}$), of course. Water is an important agent in the *transport* of heat, such as in convecting systems, but this is not really a thermodynamic subject. By "connected with heat" I mean things like the heat capacity, entropy, and enthalpy of water itself, and how changes in these properties are of interest. The property connected with work is of course the molar or specific volume.

6.2.1 Volume

So how is the molar volume connected to work that a geologist might be interested in? You might immediately think of exploding volcanos, but our interest in volume is a more subtle one.

We have seen previously the connection between Gibbs energy and volume, Equation (4.43), which can also be written as

$$\int_{P_1}^{P_2} dG = \int_{P_1}^{P_2} V\,dP \tag{6.1}$$

and which, if the volume is a constant, results in Equation (5.32). But what if V is not constant, as with fluids? In this case you need an equation expressing V as a function of P, and if you are interested in a range of T, the equation must express V as a function of both P and T. Such an equation is an *equation of state*.

The equation of state for water is discussed in Chapter 13. Here we can look at diagrams showing the properties of water, generated by the equation. The simplest of these diagrams shows molar or specific volume of water as a function of P and T (Figure 6.1). We see a fairly simple-looking fanning of specific volume lines (isochores) away from the liquid–vapor curve. At the critical point (374 °C, 221 bar), the specific volume is 3.22 cm^3 g^{-1}, but this isochore has no particular significance away from the critical point.

Over most of this diagram, water is above the critical pressure and/or the critical temperature, and is thus "supercritical". It will expand to fill any

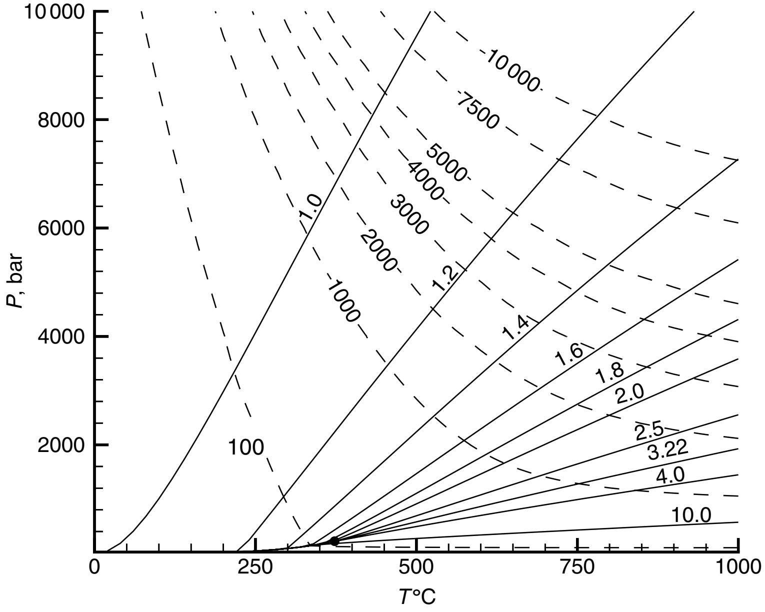


Figure 6.1 The specific volume (cm^3 g^{-1}) of water as a function of P and T. Dashed contours are water fugacity in bars. Small dot indicates the critical point. Data from the program STEAM (§13.6.1)

containing volume, thus is gas-like, but its density is far greater than normal gases, and is therefore liquid-like. Before this was well understood, hydrothermal ore deposits thought to have formed above 374 °C had a special name – "pneumatolytic," meaning formed from a gas. However we know now that above the critical pressure this temperature has no special significance – water properties vary smoothly and continuously from low to high temperatures.

6.2.2 Entropy

Figure 6.2 shows the entropy of water as a function of T and P. Clearly, isentropic expansion (that is, lowering P) will result in cooling, because the contours have a positive slope. This will be true for any pure substance, because the slope of the isentropes is

$$\left(\frac{\partial P}{\partial T}\right)_S = \frac{T}{C_P}\left(\frac{\partial V}{\partial T}\right)_P \tag{6.2}$$

and T, C_P, and $(\partial V/\partial T)_P$ are all inherently positive. However, isentropic processes can only occur reversibly, which is not possible in real life, and although they might be approximated under some conditions, this is not likely to be very common, so we will not pursue this kind of cooling process. Essentially the same information is available from Figure 4.11.

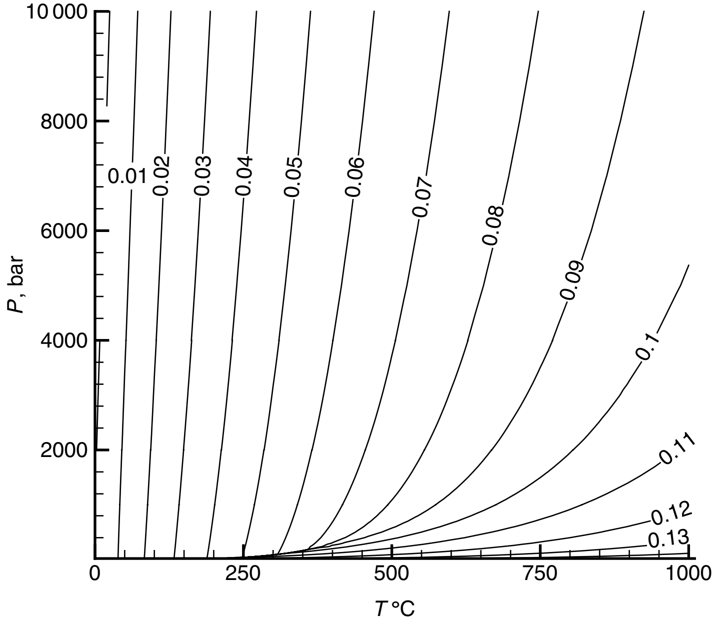


Figure 6.2 The entropy of water in kJ mol^{-1} as a function of P and T. Data from the program STEAM of Harvey et al. (2000).

6.2.3 Enthalpy

A more useful cooling process is Joule–Thompson or isenthalpic expansion. This is a "throttling" process, whereby a fluid is pushed through an orifice (originally a "porous plug") into a chamber at a lower pressure. Geologically, it would correspond to a fluid encountering a constriction while rising through a crack or fissure, and expanding as it passes the constriction. If the process is adiabatic, that is, if it happens sufficiently fast that conduction of heat from the wall rocks into the fluid is negligible, and if the pressures on the two sides of the orifice remain constant, the process will be isenthalpic.

To see this, consider a given mass of gas at pressure P_1 on the high pressure side of the orifice, plug, or constriction. As it is pushed through the orifice, the piston pushing it sweeps out a volume (ΔV) which we can call V_1, so the work done in pushing the gas is P_1V_1 (positive because work is done *on* the gas). On the other side, this same mass of gas pushes the other piston back at constant pressure through a volume (or a ΔV) V_2, doing work $-P_2V_2$ (negative because it is work done *by* the gas). The total work for the process is therefore $P_1V_1 - P_2V_2$. Because no heat is transferred, $q = 0$, so that by the first law, the internal energy change is equal to the work done, $\Delta U = w$.

$$\begin{aligned} U_2 - U_1 &= w \\ &= P_1V_1 - P_2V_2 \end{aligned}$$

so

$$U_2 + P_2V_2 = U_1 + P_1V_1$$

or

$$\begin{aligned} H_2 &= H_1 \\ \Delta H &= 0 \end{aligned}$$

To investigate the variation of temperature during an isenthalpic change of pressure, one is naturally interested in the derivative $(\partial T/\partial P)_H$, called the Joule–Thompson coefficient, μ_{JT}. Omitting the derivation, this is

$$\begin{aligned} \left(\frac{\partial T}{\partial P}\right)_H &= \frac{-V + T(\partial V/\partial T)_P}{C_P} \\ &= \frac{V(T\alpha - 1)}{C_P} \end{aligned}$$

or,

$$\mu_{\mathrm{JT}} = \frac{V(T\alpha - 1)}{C_P} \tag{6.3}$$

where α is the coefficient of thermal expansion, $(dV/dT)_P/V$, and C_P is the isobaric heat capacity. Since both V and C_P are intrinsically positive quantities, the sign of the derivative $(dT/dP)_H$ clearly depends on $(T\alpha - 1)$, which in turn depends on $T\alpha$.

As shown by Waldbaum (1971), all minerals have values of α so small that $T\alpha \ll 1$ for all Earth conditions, so that $(dT/dP)_H$ is always negative, and rocks will always be warmed by isenthalpic expansions (dP negative, dT positive). For most common minerals, the effect is about 20 to 30 degrees per kilobar of pressure change, although if this pressure change is due to a change in depth in the Earth, the effect is less (Ramberg, 1971). Some mantle convection models incorporate provision for adiabatic expansion/compression, but most do not, and under crustal conditions it is negligible.

For fluids, which have much larger values of α, $(T\alpha - 1)$ can be positive or negative. Because fluids become less dense and hence show greater thermal expansion at higher temperatures and lower pressures, this is the range of conditions where they exhibit positive values of μ_{JT}. For gases, with the largest α values, this transition takes place at very low temperatures and is of limited interest to Earth scientists. It means, though, that gases invariably have a positive μ_{JT} under ordinary conditions.

The most important Earth fluid, water, has a very interesting behavior in this as in many other respects. In Figure 6.3 we show contours of constant enthalpy

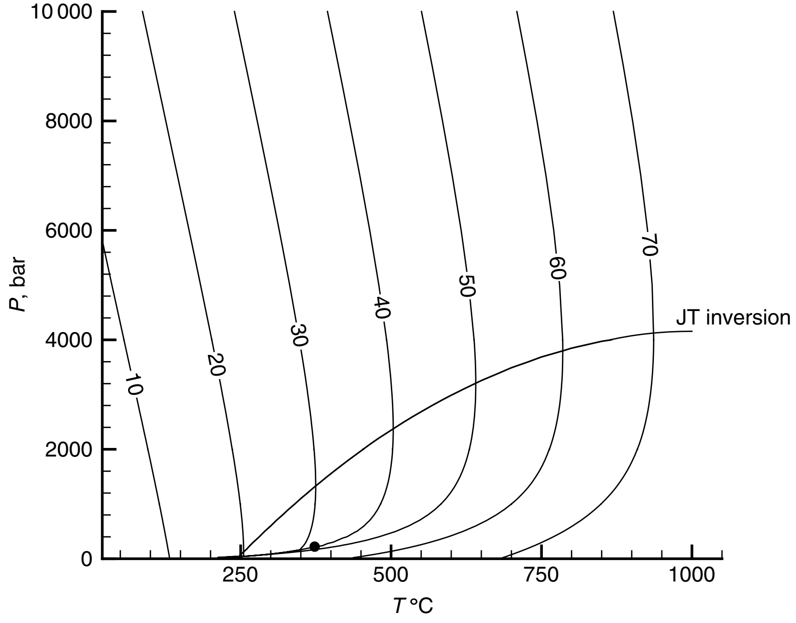


Figure 6.3 The enthalpy of water (J mol^{-1}) as a function of P and T. The small dot indicates the critical point. Data from the program STEAM of Harvey et al. (2000).

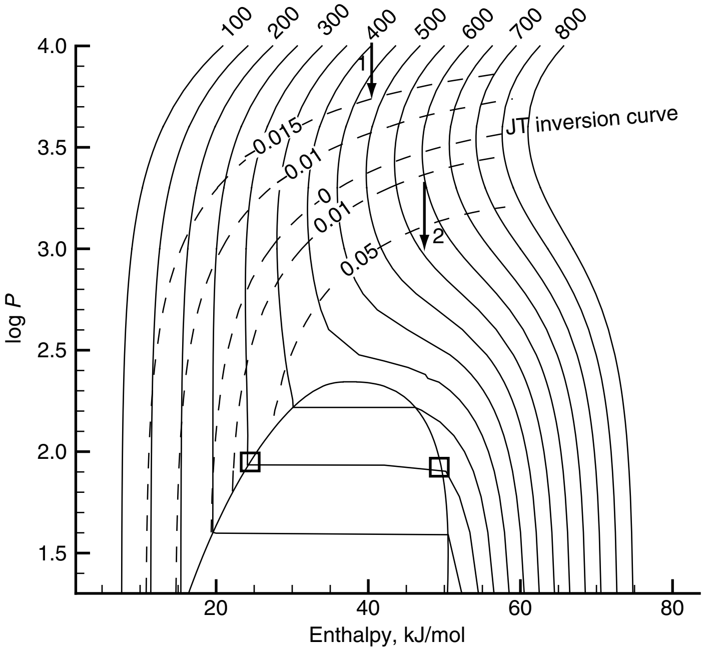


Figure 6.4 Isotherms on log P versus enthalpy in kJ mol^{-1} for water. Dashed contours are the Joule–Thompson coefficient, μ_{JT}. The contour for $\mu_{\mathrm{JT}} = 0$ is the Joule–Thompson inversion curve. Data from the program STEAM of Harvey et al. (2000).

of water as a function of P and T.[1] Where the contours have a negative slope, water warms on adiabatic expansion and cools if adiabatically compressed. Where they have a positive slope, water cools on adiabatic expansion and warms if adiabatically compressed, and the two regions are separated by the Joule–Thompson inversion curve. Much the same information is contained in the enthalpy–pressure diagram (Figure 6.4), where it can be seen that constant enthalpy changes in pressure lead to increases in temperature in one region (arrow 1) and decreases in another (arrow 2).

The effect of dissolved NaCl on the Joule–Thompson coefficient has been calculated by Wood and Spera (1984), and the effect will be similar for other electrolytes. Because the addition of most electrolytes to water results in a decrease in V and in α, μ_{JT} is smaller, and the net effect is to move the inversion curve to higher temperatures and lower pressures.

Isenthalpic boiling

A special case of adiabatic isenthalpic volume change of particular interest to geologists concerned with fluids in the upper levels of the Earth's crust is the case where a phase change, especially boiling, occurs along the cooling path of the fluid (Figure 6.5). The onset of boiling in a hot aqueous fluid in the crust

---

[1] As in Figures 4.11 and 6.4, these enthalpy numbers use the normal Steam Table convention, in which $U_{\mathrm{H_2O}} = S_{\mathrm{H_2O}} = 0$. Converting Steam Table values to apparent formation from the elements values (e.g., $\Delta_a H°$) is discussed in Chapter 13, page 388.

Joule–Thompson expansion

As an example of how the Joule–Thompson coefficient might be used, consider some hot spring fluids (approximated by pure water) rising vertically in the crust. When boiling begins, the pressure is 165 bars and the temperature is 350 °C. At this point, the properties are

V	31.35	$\mathrm{cm^3\,mol^{-1}}$
	$= 3.135$	$\mathrm{J\,bar^{-1}\,mol^{-1}}$
α	0.010 39	$\mathrm{K^{-1}}$
C_P°	182.2	$\mathrm{J\,mol^{-1}\,K^{-1}}$

so

$$\begin{aligned}\mu_{\mathrm{JT}} &= 3.135(623.15 \times 0.01039 - 1)/182.2 \\ &= 0.0942\ \mathrm{K\,bar^{-1}}\end{aligned}$$

meaning that the fluid is cooling at the rate of 0.094 °C per bar change in pressure as it rises, simply due to adiabatic decompression. If the water has risen under a hydrostatic head for two kilometers, the pressure change is about 200 bars, giving $200 \times 0.094 \approx 19$ degrees of cooling. Actually, the change is less (about 12 °C), because μ_{JT} is not constant but decreases with increasing pressure in this range.

is the cause of a number of important changes in the properties of the liquid phase because of the partitioning of the aqueous constituents between the two phases. In addition there is a new factor in the heat budget, the heat required to vaporize the liquid to steam, i.e., the heat of vaporization.

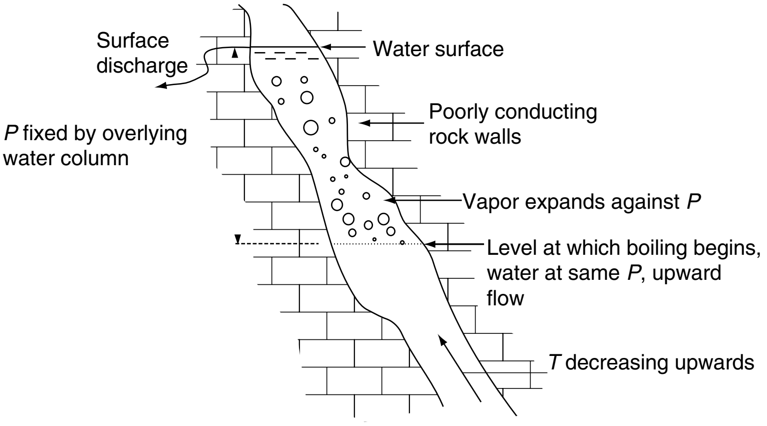


Figure 6.5 A hydrothermal fluid rising through a fissure begins to boil.

Adiabatic cooling

As shown in Figure 6.4 there is quite a large difference between the enthalpy of water and steam at equilibrium. At 300 °C, 85.88 bars for example, from Figure 6.4, $H_{\text{water}} = 24.23\,\text{kJ mol}^{-1}$ and $H_{\text{steam}} = 49.54\,\text{kJ mol}^{-1}$ (these two points marked with squares). This increase in enthalpy of the steam must come from the system itself, which is cooled as a result, but the enthalpy of the water + steam remains the same. For example, if mole fraction x_{w} of the water vaporizes at 300 °C, and the resulting water + steam is thereby cooled to T °C, then the energy balance is

$$H_{\text{w}}(300\,°\text{C}) = x_{\text{w}} \cdot H_{\text{w}}(T\,°\text{C}) + (1 - x_{\text{w}}) \cdot H_s\,(T\,°\text{C})$$

where subscripts "w" and "s" refer to water and steam respectively. To solve this we must know either T or x_{w}. If we know that the system cooled to 275 °C, where $H_{\text{w}} = 21.81\,\text{kJ mol}^{-1}$ and $H_s = 50.18\,\text{kJ mol}^{-1}$ then x_{w} is 0.91. In other words, when only 9 percent of the water boils isenthalpically the temperature drops 25 °C.

The calculation works equally well using values of $\Delta_a H°$ and/or mass fractions rather than mole fractions, but commonly numbers are obtained from Steam Tables, which use the convention $U_{\text{triple}} = S_{\text{triple}} = 0$, and joules or calories per gram. Remember, we do not know values of H_{w} and H_s. We follow the usual engineering custom above in using values for these quantities which are actually differences. These differences in enthalpy are the same as differences using the "apparent formation from the elements" convention, but Steam Table differences in Gibbs energy are not (see §13.6.1 and page 390).

Whatever the means of cooling the fluid to the point of phase separation (boiling), the boiling process is generally thought to be fast relative to the conduction of heat through the walls of the fluid reservoir, so that the fluid must itself supply the heat of vaporization, with no help from the wall rocks. This would result in essentially adiabatic conditions and the fluid would therefore be cooled as a result of boiling. Quite possibly the steam–water system would be confined at a constant pressure, exerted by the overlying column of water, so that the steam would expand at a constant pressure and the water would also be at the same constant pressure. These are the conditions required for isenthalpic expansion, and the boiling process is quite commonly assumed to be isenthalpic. This does not of course mean that the water and steam have the same enthalpy (see example). This subject is treated in more detail by Henley et al. (1984).

6.2.4 Dielectric constant and the Born function

The dielectric constant is not exactly a thermodynamic property of the types we have been discussing, but it is included here because of its importance in understanding the properties of electrolytes in water, and especially the HKF model (Chapter 15), which is incorporated in the program SUPCRT92, a widely used geochemical tool. So a brief foray into physical chemistry is perhaps warranted.

The dielectric constant is defined as the ratio of the capacitance of a substance to the capacitance of a vacuum in the same space. Bockris and Reddy (1970) offer a more enlightening definition. They say

$$\begin{array}{ccccc} \begin{array}{l}\text{Electric force in}\\ \text{a medium}\\ = \frac{q_+q_-}{\epsilon r^2}\end{array} & \Longrightarrow & \begin{array}{l}\text{is } \epsilon \text{ times}\\ \text{less than}\end{array} & \Longrightarrow & \begin{array}{l}\text{Electric force in}\\ \text{a vacuum}\\ = \frac{q_+q_-}{r^2}\end{array} \end{array}$$

where q_+ and q_- are electrostatic charges separated by distance r. If the material between them (the medium) has a dielectric constant ϵ, the attractive force between the charges is reduced. The relevance to electrolytes in water becomes obvious if you think of q_+ as a cation and q_- as an anion.

The dielectric constant of water at various temperatures and pressures is shown in Figure 6.6. The remarkable decrease from ≈ 80 at 25 °C to ≈ 20 at 300 °C at saturation vapor pressures is due to the behavior of the water molecules which, because they are miniature dipoles, are quite strongly lined up between the charges at low T, but become more thermally agitated and less strongly lined up at higher T. In consequence, oppositely charged ions

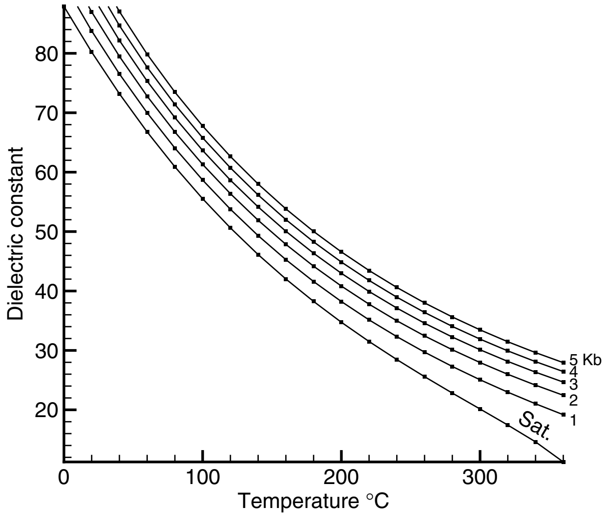


Figure 6.6 The dielectric constant of water. Data from NIST program STEAM.

are more strongly attracted to one another at higher temperatures, resulting in larger concentrations of neutral ion-pairs.

The Born function

This dramatic change in the dielectric constant has important implications for the thermodynamic properties of electrolytes in water. Born (1920) gives a very simple calculation of the work required to insert a charged sphere from a vacuum into a medium having a dielectric constant ϵ. This takes place in three steps:

1. Uncharging the sphere in the vacuum produces work energy

$$w = -\frac{(Ze)^2}{2r}$$

2. Moving the uncharged sphere from the vacuum into the medium requires no work, and
3. Charging the sphere in the medium requires work

$$w = \frac{(Ze)^2}{2\epsilon r}$$

where Z is the number of electronic charges (electrons) on the sphere (corresponding to the valence of an ion); e is the charge on an electron; r is the radius of the sphere; and ϵ is the dielectric constant of the medium. As there are no thermal effects, the combined work done is in fact the Gibbs energy of the transfer of the charged sphere from the vacuum to the medium,

$$\begin{aligned}\Delta \mathbf{G} &= \frac{(Ze)^2}{2\epsilon r} - \frac{(Ze)^2}{2r} \\ &= \frac{(Ze)^2}{2r}\left[\frac{1}{\epsilon} - 1\right] \qquad (6.4)\end{aligned}$$

Changing from a sphere to a mole (Avogadro's number, N_A) of ions, r becomes the ionic radius of, say, the jth ion, and the total Gibbs energy becomes the molar Gibbs energy of ion–solvent interaction, $\Delta_{I-S}G$, so

$$\begin{aligned}\Delta_{I-S}G &= \frac{N_A(Ze)^2}{2r_j}\left[\frac{1}{\epsilon} - 1\right] \\ &= \omega^{\text{absolute}}\left[\frac{1}{\epsilon} - 1\right] \qquad (6.5)\end{aligned}$$

We call the fraction $[N_A(Ze)^2/2r_j]$ ω^{absolute} because later on (Chapter 15) we will define another, relative, ω used in the HKF model. $\Delta_{I-S}G$ is also called $\Delta_{\text{solvation}}G$ or $\Delta_{\text{hydration}}G$.

Despite the simplifying assumptions in the derivation, such as assuming that the medium, water, is a continuum with no structure, and that the only work is electrostatic, and even more assumptions in calculating the properties of individual ions from the measured properties of electrolytes, $\Delta_{\mathrm{I}-S}G$ as estimated by the Born function comes reasonably close to the measured Gibbs energy of ion solvation, as shown in Figure 6.7. Other thermodynamic properties such as the volume, entropy and enthalpy of solvation can also be obtained by appropriate differentiation of Equation (6.5). As a result, ever since its inception the Born equation has been used as a primitive model for the electrostatic contribution to the properties of an ion in a dielectric solvent.

When data at high temperatures and pressures began to be available, it was realized that the Born model was also capable of accounting for the large negative values of various partial molar properties of electrolytes at high temperatures (such as the partial molar volume, §10.2.4), and Helgeson and Kirkham (1976) used it in combination with other terms in their equation of state for aqueous species (Chapter 15).

6.2.5 Geological applications

A familiarity with the volumetric and thermal properties of water, as presented here, is a part of standard thermodynamic knowledge for geochemists. Theoretically, both might be highly useful in interpreting geological processes.

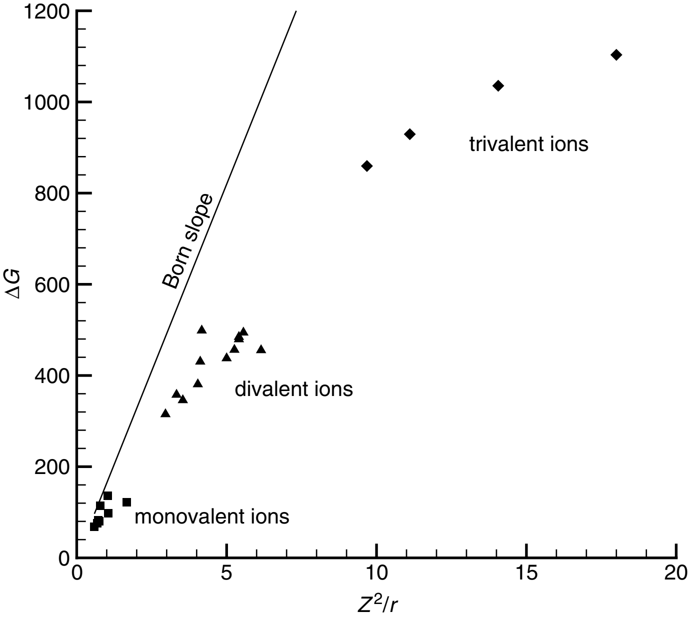


Figure 6.7 Theoretical Gibbs energies of ion–water interaction (solvation). Data from Millero (1996), Table 4.3.

The volumetric properties are in fact useful in interpreting fluid inclusions, and in knowing how much water to put in a pressure vessel, but the thermal properties are more difficult to use. The problem, as usual, is not in the thermodynamics, but in the difficulty of interpreting what we see in the field in terms of Earth processes that happened perhaps millions of years ago. Fluids rising through fissures in the upper few kilometers of the crust might well pass by an obstruction, which would have the effect of "throttling" the fluid, hence cooling it whether or not boiling was involved. The problem usually is establishing that that sort of thing actually happened.

Another aspect of this problem is that there are two different fluid regimes in the crust, an upper regime where fluid pressure is fixed by the mass of the overlying water column, and a lower regime where the fluids are trapped in pores and grain boundaries where they are not free to circulate and where the pressure is fixed by the prevailing stress pattern in the rocks. This stress is thought to be homogeneous at depths where the rocks behave plastically, but may be far from this in the region where the rocks are brittle. The boundaries between these various regimes is a continuous topic of debate, and does not concern our treatment of the thermodynamics involved. However, it is unlikely that fluids can pass reversibly between the two regimes; at some point there will be an irreversible release of pressure, with its attendant thermal effects. This change need not be adiabatic, of course, and the question of the heat flow to and from geological processes is a continual problem for modelers.

6.3 Simple phase diagrams

The reason we are interested in knowing $\Delta_r G$ for reactions is that we can then tell which way the reaction will go, or which side is more stable at one particular T and P. If we know how $\Delta_r G$ varies with T and P, we might find that under some conditions $\Delta_r G$ changes sign, so that the other side is more stable. This implies that there is a boundary between regions of T and P, with one side of the reaction stable on one side of the boundary, and the other side of the reaction stable on the other side of the boundary. A phase diagram shows which phases are stable as a function of T, P, composition, or other variables.

For example, calcium carbonate ($CaCO_3$) has two polymorphs, calcite and aragonite. Their properties (from Appendix B) are shown in Table 6.1. Because $\Delta_f G^\circ_{\text{calcite}} < \Delta_f G^\circ_{\text{aragonite}}$, we conclude immediately that calcite is the stable form of $CaCO_3$ at 25 °C, 1 bar, and that aragonite is a metastable form. But what about other temperatures and pressures? Is aragonite stable at high temperature? At high pressure? How can we tell?

Table 6.1 *Thermodynamic properties of calcite and aragonite, from Appendix B.*

Formulas	Form	$\Delta_f H^\circ$ (kJ mol^{-1})	$\Delta_f G^\circ$ (kJ mol^{-1})	S° J mol^{-1} K^{-1}	V° cm^3 mol^{-1}
$CaCO_3$	calcite	−1206.92	−1128.79	92.9	36.934
$CaCO_3$	aragonite	−1207.13	−1127.75	88.7	34.150

6.3.1 LeChatelier's principle

When looking at thermodynamic data, or the results of some thermodynamic calculation, it is always a good idea to ask yourself if it makes sense, if it is reasonable. To some extent this is a matter of experience, but in another way, "making sense" means obeying LeChatelier's principle. This simply says that *if a change is made to a system, the system will respond such as to absorb the force causing the change*. For example, if the pressure on a system is raised, the system will respond by lowering its volume, that is, by being compressed. Systems never expand as a result of increased pressure. The result of a change in temperature is less obvious, though equally certain. If the temperature of a system is raised, the enthalpy and the entropy of the system will both increase. This is because of Equations (3.24) and (5.8), which show that the temperature derivative of each is a simple function of C_P, the heat capacity, which is always positive for pure compounds.

Therefore by looking at V° and $\Delta_f H^\circ$ or S° for calcite and aragonite, and assuming that the relative magnitudes of these properties do not change much with T and P, we can tell something about their relative positions on the phase diagram. We note that $V_{\text{aragonite}} < V_{\text{calcite}}$; therefore, increasing the pressure on calcite should favor the formation of aragonite. Also, $\Delta_f H^\circ$ and S° for calcite are greater than the values for aragonite, and so raising the temperature of calcite will *not* favor the formation of aragonite. In other words, *lowering* the temperature of calcite should favor the formation of aragonite. If the stability field of aragonite lies somewhere at higher pressure and lower temperature than 25 °C, 1 bar, the boundary between the two phases must have a positive slope, as shown in Figure 6.8. This is the common case for phase boundaries; it is normal for the high-pressure, lower volume side to be the lower enthalpy, lower entropy side. The most common exception to this is the ice–water transition, as shown in Figure 3.1.

In Figure 6.8 we see that a phase diagram is a kind of free energy map – it shows a T–P region where calcite is stable ($G_{\text{calcite}} < G_{\text{aragonite}}$), and another where aragonite is stable ($G_{\text{aragonite}} < G_{\text{calcite}}$). These two regions are necessarily separated by a line where $G_{\text{aragonite}} = G_{\text{calcite}}$, the phase boundary. We have a lot more to say about phase diagrams in Chapter 17.

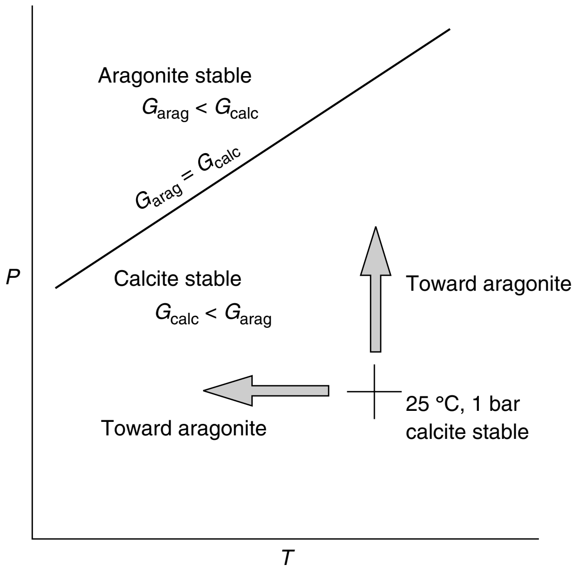


Figure 6.8 The form of the calcite–aragonite phase diagram deduced from LeChatelier's principle.

6.3.2 The effect of pressure on $\Delta_r G°$

Having figured out the relationship between calcite and aragonite qualitatively, the next step is to define the stability field of aragonite, that is, to calculate the position of the phase boundary. This should be possible, because we know that

$$\partial G/\partial P = V \qquad [4.43]$$

and thus

$$\partial \Delta G/\partial P = \Delta V$$

ΔG and ΔV refer to the difference in G and V between any two equilibrium states. In this case we are dealing with a chemical reaction between two compounds in their pure states, so we can also write

$$\partial \Delta_r G°/\partial P = \Delta_r V°$$

Integrating this equation between 1 bar and some higher pressure P, we have

$$\Delta_r G_P^\circ - \Delta_r G_{1\text{ bar}}^\circ = \int_{1\text{ bar}}^{P} \Delta_r V°\, dP \qquad (6.6)$$

and if we assume that $\Delta_r V°$ is a constant, this becomes

$$\begin{aligned}\Delta_r G_P^\circ - \Delta_r G_{1\text{ bar}}^\circ &= \Delta_r V° \int_{1\text{ bar}}^{P} dP\\ &= \Delta_r V°(P-1)\end{aligned}$$

We could use this to evaluate $\Delta_r G^\circ_P$ at any chosen value of P. However, we are particularly interested in a value of $\Delta_r G^\circ_P = \Delta_r G^\circ_{P_{\text{eqbm}}} = 0$, that is, on the phase boundary. We know the values of

$$\begin{aligned}\Delta_r G^\circ_{1\text{ bar}} &= \Delta_f G^\circ_{\text{aragonite}} - \Delta_f G^\circ_{\text{calcite}}\\ &= -1127.75 - (-1128.79)\\ &= 1.04\ \text{kJ mol}^{-1}\\ &= 1040\,\text{J mol}^{-1}\end{aligned}$$

and

$$\begin{aligned}\Delta_r V^\circ &= V^\circ_{\text{aragonite}} - V^\circ_{\text{calcite}}\\ &= 34.150 - 36.934\\ &= -2.784\ \text{cm}^3\,\text{mol}^{-1}\end{aligned}$$

So we can solve the equation for P_{eqbm}, the pressure of the calcite–aragonite equilibrium at 25 °C.

However, there is one little problem.

The units of volume

Volumes are generally measured in cubic centimeters, milliliters, liters, and so on. But if you look at an equation such as

$$w = -P\,\Delta V$$

you see that we have a problem with our units. Work (w) and $P\Delta V$ are obviously energy terms (J mol^{-1}), but the product of P in bars and ΔV in $\text{cm}^3\,\text{mol}^{-1}$ is not joules. We must always convert our volumes to joules bar^{-1}, so that the product of P and V or ΔV is J mol^{-1}. The conversion factor (Appendix A) is

$$1\text{cm}^3 = 0.10\ \text{J bar}^{-1}$$

so now our $\Delta_r V^\circ$ is $-2.784 \times 0.1 = -0.2784\ \text{J bar}^{-1}$.

Now we can solve for pressure P_{eqbm}:

$$\begin{aligned}\Delta_r G^\circ_{P_{\text{eqbm}}} - \Delta_r G^\circ_{1\text{bar}} &= \Delta_r V^\circ (P_{\text{eqbm}} - 1)\\ 0 - 1040 &= -0.2784(P_{\text{eqbm}} - 1)\\ P_{\text{eqbm}} &= 3737\ \text{bar}\end{aligned}$$

The relationship between G and P in this calculation is shown in Figure 6.9. This gives us one point on the calcite–aragonite phase boundary. We also know that the boundary has a positive slope, and so we could sketch a diagram that would be approximately right, but we really need one more piece of information – either another point on the boundary or its slope.

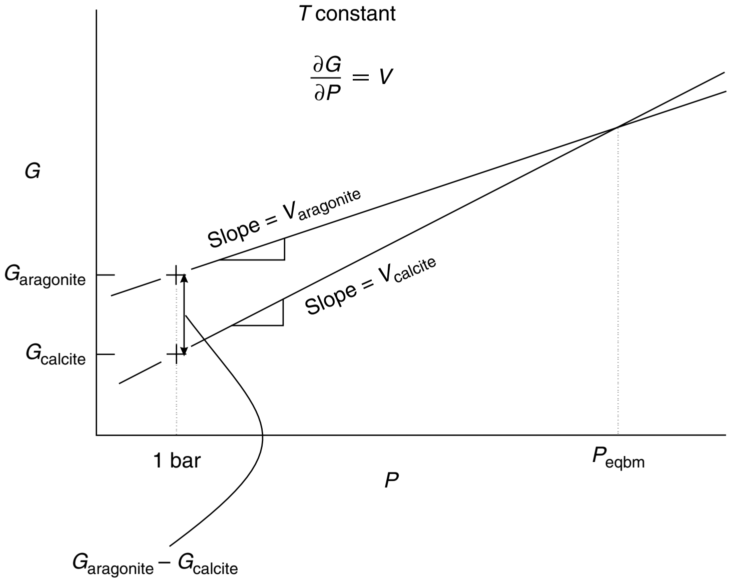


Figure 6.9 The relationship between G and P. Note that we don't know individual G values, so there are no numbers on the y-axis. We do know $G_{\text{calcite}} - G_{\text{aragonite}}$ and the slopes of the lines (the molar volumes), and this is sufficient to solve for P_{eqbm}. At P_{eqbm}, $G_{\text{calcite}} = G_{\text{aragonite}}$, the two phases can coexist, and we have a phase boundary.

6.4 The slope of phase boundaries

The phase boundary is the locus of T and P conditions where $\Delta_r G = 0$, i.e., where

$$G_{\text{calcite}} = G_{\text{aragonite}} \tag{6.7}$$

It follows that on the boundary,

$$dG_{\text{calcite}} = dG_{\text{aragonite}} \tag{6.8}$$

This simply says that as you move along the boundary, the change in G_{calcite} has to be the same as the change in $G_{\text{aragonite}}$; otherwise you won't stay on the boundary. From Equation (4.40) we have

$$dG = -S\,dT + V\,dP \qquad [4.40]$$

This applies to each mineral, and combining with (6.8) gives

$$-S_{\text{calcite}}dT + V_{\text{calcite}}dP = -S_{\text{aragonite}}dT + V_{\text{aragonite}}dP$$

Rearranging this gives

$$\frac{dP}{dT} = \frac{(S_{\text{calcite}} - S_{\text{aragonite}})}{(V_{\text{calcite}} - V_{\text{aragonite}})}$$

or, for any reaction

$$\frac{dP}{dT} = \frac{\Delta_r S}{\Delta_r V} \tag{6.9}$$

which gives the slope of an equilibrium phase boundary in terms of the entropy and volume changes between the phases involved in the reaction. This is called the Clapeyron equation.

Equation (5.1) says

$$\Delta G_{T,P} = \Delta H - T\,\Delta S \tag*{[5.1]}$$

This applies to any change between two equilibrium states at the same T and P. If those two equilibrium states have the same value of G, such as calcite and aragonite do on their phase boundary (6.7), then $\Delta G_{T,P} = 0$, and

$$\Delta H = T\,\Delta S \tag{6.10}$$

or

$$\frac{\Delta H}{T} = \Delta S \tag{6.11}$$

This is a useful relationship for any phase boundary,[2] which is the usual place to find $\Delta G_{T,P} = 0$. This gives an alternative form of the Clapeyron equation,

$$\frac{dP}{dT} = \frac{\Delta H}{T\,\Delta V} \tag{6.12}$$

6.4.1 The slope of the calcite–aragonite boundary

We have one point on the calcite–aragonite boundary at 3737 bar, 25 °C. If we assume that the $\Delta_r S$ and the $\Delta_r V$ at this P and T are the same as those at 1 bar, 25 °C, we can calculate the slope from the data in our tables. Thus

$$\begin{aligned}\Delta_r S &= S_{\text{aragonite}} - S_{\text{calcite}} \\ &= 88.7 - 92.9 \\ &= -4.2\,\text{J mol}^{-1}\,\text{K}^{-1}\end{aligned} \tag{6.13}$$

and

$$\begin{aligned}\Delta_r V &= V_{\text{aragonite}} - V_{\text{calcite}} \\ &= 34.150 - 36.934 \\ &= -2.784\,\text{cm}^3\,\text{mol}^{-1} \\ &= -0.2784\,\text{J bar}^{-1}\end{aligned}$$

---

[2] That is, any phase boundary in a one-component system. With two or more components, the relationship is in principle the same but becomes more complicated, and less useful.

Therefore

$$\begin{aligned}\frac{dP}{dT} &= \frac{\Delta_r S}{\Delta_r V}\\ &= \frac{-4.2}{-0.2784}\\ &= 15.09\ \text{bar/°C}\end{aligned}$$

Therefore, to get another point on the calcite–aragonite phase boundary, we simply choose an arbitrary temperature increment, say 100 °C, calculate the corresponding pressure increment, $100 \times 15.09 = 1509$ bar, and add these increments to our first point. We now have a second point at 125 °C, $3737 + 1509 = 5246$ bar, and we can plot the boundary as in Figure 6.10.

Keep in mind that we have assumed that the $\Delta_r S$ and $\Delta_r V$ from the tables are unchanged at all temperatures and pressures, that is, that they are constants. This is quite a good approximation for a reaction involving only solid phases such as this one, but you would not use it for reactions involving liquids, gases, or solutes. In general, all thermodynamic parameters do vary with T and P, so phase boundaries are in principle curved and not straight as we have assumed. However, the amount of curvature is quite small in some cases, such as this one.

6.4.2 Comparison with experimental results

Table 6.2 shows the results of some experiments on the stability of $CaCO_3$ at elevated temperatures and pressures. A mixture of calcite and aragonite was

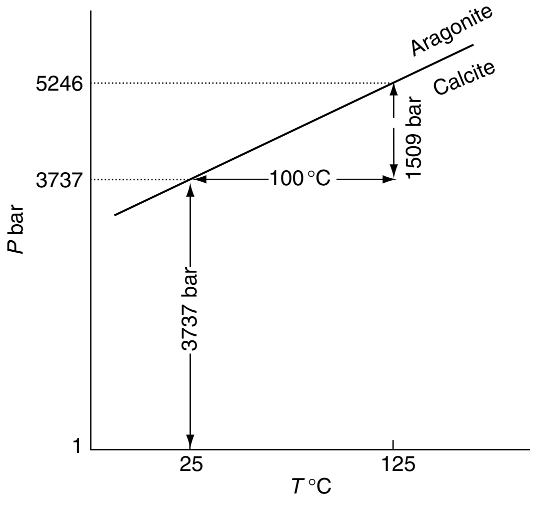


Figure 6.10 Calculation of the calcite–aragonite phase diagram.

Table 6.2 *Experimental results for the system $CaCO_3$ from Crawford and Hoersh (1972).*

Temperature °C	Pressure bars	Experimental result	Duration days
128	5180	A	21
132	5180	A	21
153	4830	C	35
76	4480	A	3
90	4140	C	28
93	4140	C	17
56	4140	A	28
70	4140	A	17
70	3690	C	8
81	3520	C	36

held at the indicated *T* and *P* for the length of time shown, then quenched and examined. The stable phase is shown as C (calcite) or A (aragonite). These points are plotted in Figure 6.11.

Also shown in this figure are the two points we have just calculated at 25 and 125 °C, plus results using data from Helgeson et al. (1978), and a line showing the experimenter's best estimate of the phase boundary. As you see,

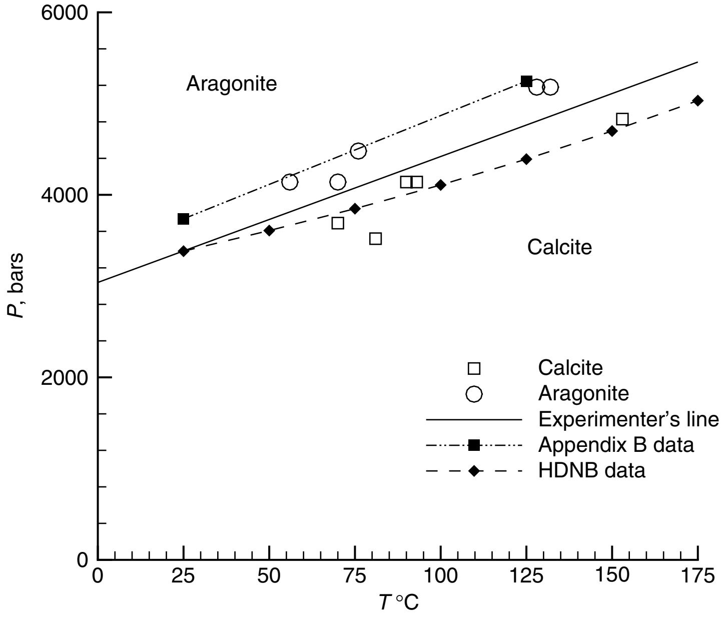


Figure 6.11 Comparison of experiment and calculation of the calcite–aragonite phase diagram.

the calculated results using data from Appendix B are a little high, and the Helgeson et al. results are a bit low.[3]

Using a reaction involving only solids, such as our calcite–aragonite example, is actually quite a stringent test of accuracy of ΔG° values. The reason is that minerals have relatively small entropies, and entropies of reaction, which are differences, are thus even smaller. The standard entropy of reaction for calcite–aragonite for example is only $-4.2\,\mathrm{J\,mol^{-1}\,K^{-1}}$ (Equation 6.13), or $-2.5\,\mathrm{J\,mol^{-1}\,K^{-1}}$ from SUPCRT92. Why should the entropy have such an effect?

Figure 6.12 shows $\Delta_r G^\circ$ for the calcite $\rightleftharpoons$ aragonite reaction from HDNB data, for 1 bar and 4140 bars.[4] Also shown are the four experimental points at that pressure. As you can see, shifting the $\Delta_r G^\circ$ curve down by only $15\,\mathrm{cal\,mol^{-1}}$ would shift the $\Delta_r G^\circ = 0$ intersection by about 20 °C, which would be within the experimental bracket at 4140 bars. Clearly, the amount of temperature change for a given change in $\Delta_r G^\circ$ depends on the slope of the line as it passes through $\Delta_r G^\circ = 0$, and the slope of the line is of course $-\Delta_r S^\circ$. The relationship is just a variation of Equation (4.42), or

$$\left(\frac{\partial \Delta_r G^\circ}{\partial T}\right)_P = -\Delta_r S^\circ \tag{6.14}$$

where in this case, $\partial \Delta_r G^\circ$ is the assumed error in $\Delta_r G^\circ$, ∂T is the resulting ΔT, and $-\Delta_r S^\circ$ is the slope of the $\Delta_r G^\circ$ line as it passes through zero. In this case, the slope (at 102.6 °C, from SUPCRT92, or by fitting a polynomial to the ΔG° line in Figure 6.12 and differentiating) is $0.73\,\mathrm{cal\,mol^{-1}\,K^{-1}}$, so an assumed error of $-15\,\mathrm{cal\,mol^{-1}}$ gives $\Delta T = -15/0.73 = -20\,^\circ\mathrm{C}$.

6.4.3 Errors

An error in $\Delta_r G^\circ$ of $15\,\mathrm{cal\,mol^{-1}}$ ($\approx 63\,\mathrm{J\,mol^{-1}}$) is extremely small. Standard errors in Gibbs energies determined by statistical methods are normally

---

[3] Data from Appendix B are mostly from calorimetric sources, and have not been "refined" using high pressure experimental data, therefore it is not surprising that they do not fit the experimental points, although the agreement is actually fairly good. Data from Helgeson et al. (1978), called HDNB data, were so refined, and normally fit experimental data very well. That they miss somewhat in this case may be due to the fact that other data at much higher pressures, as well as solubility data, were also used in refining the ΔG° data.
This phase boundary may also be calculated directly by SUPCRT92, using the "univariant curve option." Surprisingly, the results are worse than using the HDNB data. For many minerals, SUPCRT92 uses HDNB data but for calcite and aragonite the values of ΔG° are slightly different.

[4] In this discussion of the calcite $\rightleftharpoons$ aragonite reaction we use the terms $\Delta_r G^\circ$ and $\Delta_r S^\circ$. Figure 6.12 uses the term ΔG°. We might also have used ΔG and ΔS. This can cause confusion, unless you refer to Figure 4.12 to see that there is no logical problem, just a difference in how explicit we wish to be. That is, $\Delta_r G^\circ$ is just a special case of ΔG. In this case, we are dealing with a chemical reaction (subscript "*r*"), and with pure phases in their (high *T*) standard states (superscript °), so $\Delta_r G^\circ$ is correct, but ΔG° or ΔG is not incorrect.

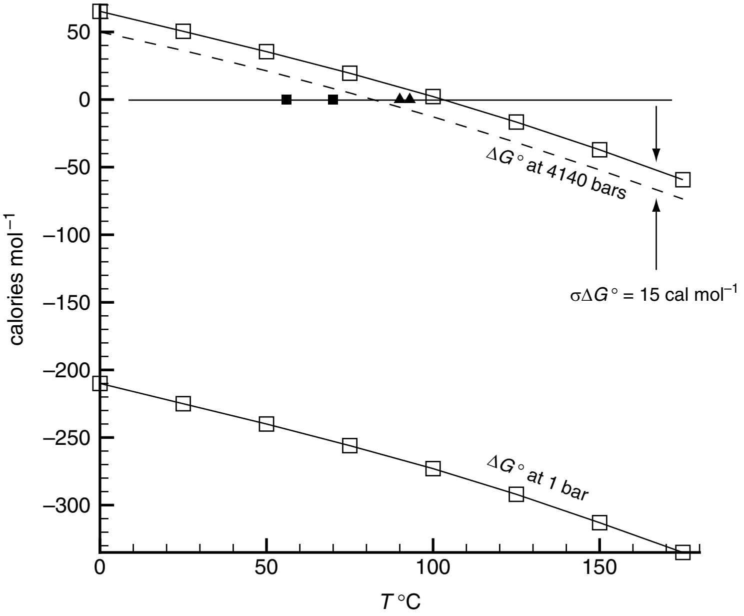


Figure 6.12 $\Delta_r G°$ for the reaction calcite $\rightleftharpoons$ aragonite. Lower curve at 1 bar, upper curve corrected to 4140 bars. Dashed curve shows the effect of an error in $\Delta G°$ of $-15\,\text{cal}\,\text{mol}^{-1}$. Squares (aragonite) and triangles (calcite) from Table 6.2.

measured in kilocalories, or kilojoules. With uncertainties of that magnitude, possible errors in calculated phase boundaries are obviously huge. On the other hand, if you know the position of a phase boundary accurately, you can deduce very accurate free energy values from it. For example, if the determinations of Crawford and Hoersh (1972) in Table 6.2 are accurate, we could determine $\Delta_r G°$ for the calcite–aragonite reaction with extremely little error just by reversing the calculation we have just done. In other words, phase stability determinations are a very powerful tool in determining the best values of thermodynamic terms.

This fact was first used in a systematic way by Helgeson et al. (1978) to determine values of most of the common rock-forming minerals, using certain data as “bedrock,” or well-known, and calculating free energies from phase equilibria, as well as other sources such as solubilities and calorimetry, to develop a network of self-consistent numbers. Self-consistent in this sense means that the calculated free energies would reproduce all the phase equilibria and other data that were used in determining them. Since then several other investigators have also done this, using a variety of methods, including in some cases observations on natural assemblages as well as experimental results. These data have also been assembled into databases for use by computer programs.

Because of these efforts, the situation with regard to data of geochemical interest has improved greatly. Nevertheless, many problems remain – the properties of the minerals in experiments may be different from those of natural minerals; there are many experimental difficulties that may result in incorrect data, and so on. We might say that there will *always* be problems with data – there will never be complete agreement as to the "best" numbers to use, and there will always be new data to consider.

Accuracy and precision

The problem will generally be one of accuracy, and not precision. Accuracy refers to how close to "the truth" your determination is, and precision refers to the dispersion of your determinations – i.e., the reproducibility. Statistical methods resulting in "standard errors" refer to precision, but this is no longer a problem in this field. To illustrate this, consider the case where ten laboratories cooperate in determining the composition of a sample they wish to establish as a standard. When the results are in, it is found that nine laboratories agree very closely, but the tenth gives a very different result. Using standard statistical methods, the tenth result can be eliminated as an "outlier," and the nine results combined to give a very impressive "standard error." However, it may be that the tenth result is actually more accurate; that the nine laboratories used some method or technique that introduced a systematic error.

The problem of systematic error in analytical and experimental work will always be with us, which is why there are always arguments about the validity of various methods, and which results are "the best." Reproducibility is generally very good, but who cares, until confidence is established that the accuracy is good. Of course once there is confidence in the methods, measurements of precision become important in assembling data from various sources, for example in developing equations of state. Estimates of precision are commonly used to assign weighting factors to data from different sources.

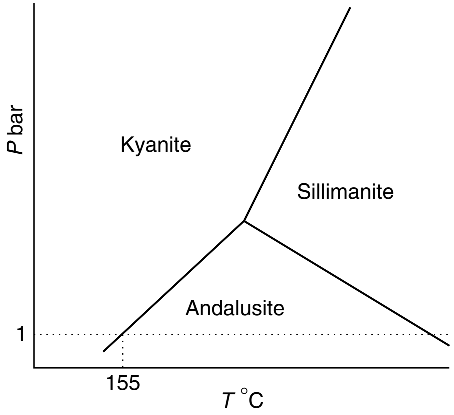
Thus having self-consistent data is no guarantee, because they may be self-consistent with flawed experiments or analyses. This subject is now a large and very important part of geochemical research, but it is not particularly a thermodynamic subject, so we will not pursue it further here.

6.5 Another example

6.5.1 The effect of temperature on $\Delta_r G^\circ$

To illustrate the effect of temperature on $\Delta_r G^\circ$, we could continue with the calcite–aragonite case and try to calculate the temperature where the phase boundary crosses the 1 bar pressure line (Figure 6.10). Unfortunately, this turns out to be close to absolute zero, so it is not a very useful example. As another case let's consider the polymorphs of Al_2SiO_5. There are three of these, kyanite, andalusite, and sillimanite. Therefore there are three two-phase boundaries, and these three boundaries meet at a single point, where $G^{\text{kyanite}} = G^{\text{andalusite}} = G^{\text{sillimanite}}$ as shown in Figure 6.13. These minerals, which form quite commonly in rocks subjected to high temperatures and pressures in the Earth's crust, are of special interest to geologists who study these rocks because the "triple point," the point where the three phase boundaries meet, is

Figure 6.13 The phase diagram for the aluminum silicate polymorphs.



in the middle of a rather common range of T–P conditions. If a rock contains one of these minerals, the geologist immediately has a general idea of the T and P conditions at the time the rock formed. It is only a "general idea" because it is not safe to assume that rocks reach chemical equilibrium at some P and T, and then remain unaltered as they are exhumed and are exposed at the Earth's surface. Many complications can occur, which are not within the subject of thermodynamics.

According to Figure 6.13, the kyanite–andalusite boundary crosses the 1 bar line at some elevated temperature. We should be able to calculate what this is by methods perfectly analogous to those we used for calcite–aragonite. The data we need (from Appendix B) are shown in Table 6.3.

First, we note that all seems to be well to start with, in that there is no conflict between the data and Figure 6.13. Kyanite has the lower value of $\Delta_f G^\circ$, and so it should be the stable phase at 25 °C, 1 bar, as shown in the diagram. The values of $\Delta_r S^\circ$ and $\Delta_r V^\circ$ would indicate that kyanite is the high-pressure phase, and that the kyanite – andalusite boundary has a positive slope, also as shown by the diagram.

To calculate the temperature of the kyanite – andalusite boundary at 1 bar, we start with Equation (4.42),

$$(\partial G/\partial T)_P = -S \qquad [4.42]$$

Table 6.3 *Thermodynamic data for the Al_2SiO_5 minerals, from Appendix B.*

Formulas	Form	$\Delta_f H^\circ$ kJ mol^{-1}	$\Delta_f G^\circ$ kJ mol^{-1}	S° J mol^{-1} K^{-1}	V° cm^3 mol^{-1}
Al_2SiO_5	kyanite	−2594.29	−2443.88	83.81	44.09
Al_2SiO_5	andalusite	−2590.27	−2442.66	93.22	51.53

from which we can write immediately[5]

$$(\partial \Delta_r G^\circ/\partial T)_P = -\Delta_r S^\circ$$

Integrating this from 298.15 K to some higher temperature T, we get

$$\Delta_r G^\circ_T - \Delta_r G^\circ_{298} = \int_{298}^{T} -\Delta_r S^\circ dT \tag{6.15}$$

and if we assume that $\Delta_r S^\circ$ is a constant, this becomes

$$\begin{aligned}\Delta_r G^\circ_T - \Delta_r G^\circ_{298} &= -\Delta_r S^\circ_{298} \int_{298}^{T} dT \\ &= -\Delta_r S^\circ_{298}(T - 298.15) \qquad (6.16)\end{aligned}$$

Now if we let $\Delta_r G^\circ_T = 0$, T becomes T_{eqbm}, and we can solve for this. From the tables, $\Delta_r G^\circ_{298} = 1220\,\mathrm{J\,mol^{-1}}$ and $\Delta_r S^\circ_{298} = 9.41\,\mathrm{J\,mol^{-1}\,K^{-1}}$, so

$$\begin{aligned}\Delta_r G^\circ_T - \Delta_r G^\circ_{298} &= -\Delta_r S^\circ_{298} \int_{298}^{T_{\mathrm{eqbm}}} dT \\ 0 - 1220 &= -9.41(T_{\mathrm{eqbm}} - 298.15) \\ T_{\mathrm{eqbm}} &= 427.8\ \mathrm{K} \\ &= 154.6\,^\circ\mathrm{C}\end{aligned}$$

The relationship between G and T is shown in Figure 6.14. Note that in Figure 6.9, the slope of G versus P is positive, whereas in Figure 6.14, the slope of G versus T is negative. This is because for pure substances V is always

---

[5] It is not immediately clear to many students why, if $(\partial G/\partial T)_P = -S$, we can "write immediately" $(\partial \Delta_r G^\circ/\partial T)_P = -\Delta_r S^\circ$, that is, why we can just stick in a Δ whenever we wish. It is because the derivative relationship can be applied to all terms of any balanced reaction. For example, if the reaction is $\mathrm{A} + 2\mathrm{B} = \mathrm{C}$ (e.g., reaction (5.11)),

$$\begin{aligned}\Delta_r G^\circ &= \Delta_f G^\circ_{\mathrm{C}} - \Delta_f G^\circ_{\mathrm{A}} - 2\,\Delta_f G^\circ_{\mathrm{B}} \\ &= G^\circ_{\mathrm{C}} - G^\circ_{\mathrm{A}} - 2\,G^\circ_{\mathrm{B}}\end{aligned}$$

so the derivative with respect to T is

$$\begin{aligned}(\partial \Delta_r G^\circ/\partial T) &= (\partial G^\circ_{\mathrm{C}}/\partial T) - (\partial G^\circ_{\mathrm{A}}/\partial T) - 2\,(\partial G^\circ_{\mathrm{B}}/\partial T) \\ &= -S^\circ_{\mathrm{C}} + S^\circ_{\mathrm{A}} + 2\,S^\circ_{\mathrm{B}} \\ &= -(S^\circ_{\mathrm{C}} - S^\circ_{\mathrm{A}} - 2\,S^\circ_{\mathrm{B}}) \\ &= -\Delta_r S^\circ\end{aligned}$$

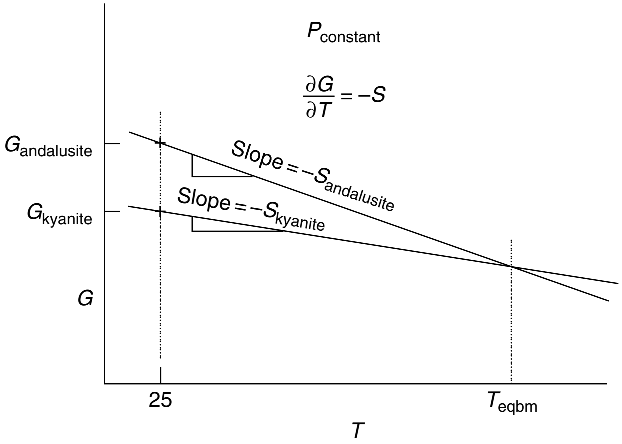


Figure 6.14
The relationship between G and T. Note the general similarity to Figure 6.9, with the exception that the slopes are negative.

positive, and S is always positive by virtue of the third law. This is only true in general for pure substances; for differences (i.e., $\Delta_r V$, $\Delta_r S$) or for solutes, these quantities may be negative, as we will see.

6.5.2 A different formulas for $\Delta_r G_T^\circ$

Another useful way of expressing the effect of temperature on G is given by expanding (6.16). Thus

$$\Delta_r G_T^\circ - \Delta_r G_{298}^\circ = -\Delta_r S^\circ (T - 298.15)$$

$$\Delta_r G_T^\circ - (\Delta_r H_{298}^\circ - 298.15\,\Delta_r S_{298}^\circ) = -T\,\Delta_r S_{298}^\circ + 298.15\,\Delta_r S_{298}^\circ$$

Collecting and rearranging terms gives

$$\Delta_r G_T^\circ = \Delta_r H_{298}^\circ - T\,\Delta_r S_{298}^\circ \tag{6.17}$$

In other words, you can calculate $\Delta_r G^\circ$ at some temperature T using the values of $\Delta_r H^\circ$ and $\Delta_r S^\circ$ at 298.15 K. However, this is subject to the same restriction as before, that both $\Delta_r H^\circ$ and $\Delta_r S^\circ$ are not functions of temperature. Of course, both these terms always *are* functions of temperature, but often this can be neglected without introducing much error, especially if T is not very different from 298 K.

Both (6.16) and (6.17) are therefore approximations, to be used only over a small temperature interval, or in cases where the result need only be approximate. More accurate formulae involve the heat capacity, but as there are a variety of equations expressing the heat capacity as a function of T, there are a variety of more accurate expressions for $\Delta_r G^\circ$. We looked at two of these in Chapter 5, Equations (5.29) and (5.38).

6.6 Summary

The main idea in this chapter is to illustrate the uses of our thermodynamic variables, using only pure phases. We did this by considering some properties of water at elevated T and P, and by calculating simple phase diagrams. In phase diagrams, the condition $\Delta_r G = 0$ becomes a central concern, and for pure phases, this is the same as $\Delta_r G^\circ = 0$.

Many reactions involving only pure phases also involve water, but calculating the Gibbs energy of water at high T and P is more difficult than for minerals, because we cannot assume that it is incompressible. How to handle this, as well as how to deal with solutions, is an important topic in later chapters.

7

Ideal solutions

7.1 Introduction

If the world were made of pure substances, our development of the thermodynamic model would now be complete. We have developed a method, based on measurements of heat flow, that enables predictions to be made about which way reactions will go in given circumstances. But one of the reasons that the world is so complex is that pure substances are relatively rare, and strictly speaking they are nonexistent (even "pure" substances contain impurities in trace quantities). Most natural substances are composed of several components, and the result is called a *solution*. Therefore, we need to develop a way to deal with components in solution in the same way that we can now deal with pure substances – we have to be able to get numerical values for the Gibbs energies, enthalpies, and entropies of components in solutions. We will then be able to predict the outcome of reactions that take place entirely in solution, such as the ionization of acids and bases, and reactions that involve solids and gases as well as dissolved components, such as whether minerals will dissolve or precipitate. Our thermodynamic model will then be complete.

In this chapter we have a look at how to deal with dissolved substances – solutes. When we mix two substances together, sometimes they dissolve into one another, like sugar into coffee or alcohol into water, and sometimes they do not, like oil and water. In the former case, if we thought about it at all, we would probably expect that the properties of the mixture or solution would be some kind of average of the properties of the two separate substances. This is more or less true for some properties, but decidedly not true for the most important one, Gibbs energy.

After making sure we understand how to express the composition of solutions, we begin by considering properties of *ideal solutions*, which are, as you might expect, the simplest possible properties that solutions might have. As you might also expect, no real solutions are in fact ideal, although some come fairly close. But the properties of ideal solutions are of interest not only because some solutions are almost ideal, but because what we often do for more complex solutions is to *subtract* the properties of the ideal solution from the corresponding property of the real solution, and then deal only with the *difference* between the two, which simplifies things quite a bit.

7.2 Measures of concentration

A number of concentration terms are used in describing solutions, and it is naturally important to be able to change from one to another.

Mole fraction

Consider a solution containing a number of components, n_1 moles of component 1, n_2 moles of component 2, and so on. If it is an aqueous solution, then water is one of the components, normally the major component. The mole fraction of any one component i is defined as

$$x_i = \frac{n_i}{\sum_i n_i} \tag{7.1}$$

where $\sum_i n_i$ is the total number of moles of components, $n_1 + n_2 + n_3 + \cdots$.

The mole fraction is very commonly used, especially in theoretical discussions, because it is perfectly general, and it can cover the entire range of compositions from dilute solutions to pure components. It is inconvenient for aqueous solutions because these are usually quite dilute on the mole fraction scale; that is, water is by far the dominant component, and the mole fractions of the solutes are numerically very small.

The mole fraction is a simple concept, but there is one important thing to note. In any mole fraction the question is, n_1, n_2, etc., are moles of *what*? This is not as simple as it might seem. Let's say you have a solution containing 1 mole (≈58 g) of NaCl, 1 mole (≈75 g) of KCl, and 50 moles (≈900 g) of water. What is the mole fraction of NaCl? Is it

$$\begin{aligned} x_{\mathrm{NaCl}} &= \frac{n_{\mathrm{NaCl}}}{n_{\mathrm{NaCl}} + n_{\mathrm{KCl}} + n_{\mathrm{H_2O}}} \\ &= \frac{1}{1+1+50} \\ &= 0.0192 \end{aligned}$$

or is it

$$\begin{aligned} x_{\mathrm{NaCl}} &= \frac{2n_{\mathrm{NaCl}}}{n_{\mathrm{Na^+}} + n_{\mathrm{K^+}} + n_{\mathrm{Cl^-}} + n_{\mathrm{H_2O}}} \\ &= \frac{2}{1+1+2+50} \\ &= 0.0370 \end{aligned}$$

Completely dissociated electrolytes under ambient conditions has long been a major topic in solution chemistry, so the second option is traditionally used. That is, for the mole fraction of an aqueous solution of a strong electrolyte such as NaCl or KCl, Equation (7.1) is modified to

$$x_i = \frac{\nu n_i}{\nu n_i + n_{\mathrm{H_2O}}} \tag{7.2}$$

where ν is the number of moles of ions resulting from the dissolution of one mole of solute (assuming complete dissociation; e.g., ν is 2 for NaCl, 3 for Na_2SO_4, 4 for $AlCl_3$, etc.). The corresponding mole fraction of the solvent, water, is then

$$x_{\mathrm{H_2O}} = \frac{n_{\mathrm{H_2O}}}{\nu n_i + n_{\mathrm{H_2O}}}$$

However, the situation is not as clear under conditions of high T and P, where even "strong" electrolytes dissociate to a variable extent, and may hardly dissociate at all.

So the message is, if you use mole fractions, make sure you know how they are defined.

Molality

The molality (m_i) of component i is the number of moles of i (n_i) per kilogram of pure solvent, usually water. Even if the aqueous solution contains several solutes, the molality is the number of moles of one of them in 1000 g of *pure* water. It is less general than mole fraction in the sense that you cannot express the composition of the pure solute in molal units, because m_i becomes infinite for pure i.

The use of molality is virtually universal for aqueous solutions because it is independent of the temperature and pressure of the solution, and equations in molality are usually simpler than equations in mole fractions.

Molarity

The molarity (M_i) of component i is the number of moles of i in 1 liter of solution (not a liter of pure solvent). This is a convenient unit in the laboratory, where solutions are prepared in volumetric flasks. It has the disadvantage that as temperature or pressure changes, the volume of the solution changes but the definition of the liter does not, and the molarity is therefore a function of temperature and pressure. The conversion to molality requires a knowledge of the density of the solution, which is readily available in handbooks for binary solutions at 25 °C, but usually not available for natural solutions. For dilute

Table 7.1 *Relationship between NaCl molality, weight %, and molality at 25 °C. Density data from Pitzer* et al. *(1984).*

Molality mol kg^{-1}	Weight % g/100 g	Density g cm^3	Molarity mol L^{-1}
0.01	0.058	0.997 46	0.0100
0.10	0.58	1.001 17	0.0995
0.25	1.44	1.007 22	0.2482
0.50	2.84	1.017 10	0.4941
0.75	4.20	1.026 76	0.7377
1.0	5.52	1.036 23	0.9790
2.0	10.46	1.072 28	1.9201
3.0	14.92	1.105 77	2.8225
4.0	18.95	1.137 05	3.6864
5.0	22.61	1.166 44	4.5133
6.0	25.96	1.194 23	5.3051

solutions at ambient conditions, m and M are about the same. The relationship for NaCl is shown in Table 7.1

Weight percent

Measurement in weight %, the grams of solute per 100 grams of solution, is used in the metallurgical literature, and in some areas of geochemistry. Natural solutions found in fluid inclusions, basinal brines and evaporitic environments can reach concentrations of several molal. To convert to or from molality you need to know the molecular weight of the component. The relationship between NaCl weight % and molality is shown in Table 7.1.

Parts per million

For trace components a million grams of solution rather than 100 grams may be used, giving parts per million (ppm). For example, an aqueous solution that is 10^{-4} molal in Zn contains $0.0001 \times 65.37 = 0.006\,537$ grams of NaCl in $(1000 + 0.006\,537)$ grams of solution, or about 10^3 grams of solution. Therefore there would be 6.537 grams of Zn in 10^6 grams of solution, or 6.5 ppm. If the solution contains a number of other solutes, they should all be included in the denominator, but it is common practice to ignore all components except the solute of interest and water.

Roughly equivalent and perhaps more common units are milligrams per liter and milligrams per kilogram of solvent. Being a volumetric unit, conversion of mg/L should involve the density, but for dilute solutions, mg/L, mg/kg and ppm are about the same. Table 7.2 gives a comparison for Zn.

Table 7.2 *Zinc concentration units.*

Molality	ppm	mg/kg
0.0001	6.5	6.5
0.001	65.0	65.0
0.005	327.0	327.0
0.010	653.0	654.0
0.050	3 258.0	3269.0
0.100	6 495.0	6537.0
0.150	9 710.0	9806.0
0.200	12 905.0	13 074.0

7.3 Properties of ideal solutions

What are the properties of true ideal solutions and why do real solutions not behave this way? The picture differs for gases, liquids, and solids. Before developing the equations, it will help to have a mental picture of what an ideal solution is.

7.3.1 Ideal gaseous solutions

Taking the simplest case first, an ideal gas consists of hypothetical, vanishingly small particles that do not interact in any way with each other. They are unaware of the existence of the other particles and there are no forces or energies of attraction or repulsion. An ideal gas obeys the ideal gas law, $P\mathbf{V} = nRT$, where n is the number of moles, T is related to the movement and individual energies of the particles, $\mathbf{V}$ is the volume occupied by the particles, and P comes from the only interaction allowed in the system – particles bouncing off the walls or boundaries. A *solution* of two ideal gases will also obey the ideal gas law since the particles of the different constituents remain unaware of all other particles, just as with an ideal single-component gas. You might say that molecules in an ideal gas, whether pure or a solution, think they are in a perfect vacuum. Of course, real gases do interact at the molecular scale and can only be expected to approach ideal behavior at very low densities and pressures, or in the limit as $P \to 0$.

7.3.2 Ideal liquid solutions

Liquids are necessarily more complicated than gases. To start with, they have much greater cohesiveness than gases; for example, a liquid equilibrated with its gaseous vapor develops a meniscus. This boundary has a measurable surface tension caused by the fact that particle interactions in the liquid are stronger than those in the vapor. A liquid must have significant interaction among its particles – if it did not, it would disperse and become a gas.

In an ideal liquid solution, the forces of interaction between all molecules, whether of one type or another, are exactly the same. For example in a liquid solution of constituents A and B, interactions A–A, A–B, and B–B must be identical. This means that all constituents A,B,…, must have the same molecular properties (size, charge, polarity, bonding characteristics). This is never the case, of course, but mixtures of some organic compounds come fairly close.

Given this uniformity of intermolecular forces in the ideal liquid solution (as opposed to the *absence* of such forces in the ideal gas), it follows that many properties of the solution are very simply related to the properties of the pure compounds. Thus the volume of the solution is the sum of the volumes of the pure components before mixing, and no heat is absorbed or given off when the solution is prepared (because such effects are caused by changes in the particle interactions, which we have just ruled out).

7.3.3 Ideal solid solutions

A solid has a rigid structure, and its component molecules, ions, or atoms are confined to specific structural sites. The regularity of the structure varies, of course, from glassy to fully crystalline materials, but whatever the degree of ordering, the positions of the particles are fixed. Whereas ideal gases and gaseous solutions have a complete absence of interparticle forces and ideal liquid solutions have a complete uniformity, solids must have highly specific interactions between different constituents. We speak of specific sites in crystals, such as tetrahedral silicon–oxygen bonds and octahedral aluminum sites, and the same is true (although to a lesser extent) of glassy solids. The interactions between Si–Si, Si–O, Si–Al, O–O, and Al–O in an aluminosilicate are all quite different. However, within the framework of a perfectly crystalline compound it is frequently possible to substitute one element for another. This substitution and the corresponding solid solution would be ideal *if the two substituting elements or species were completely indistinguishable.* The closest approximation to an ideal solid solution would be the substitution of two isotopes of the same element on the same crystal site. Like ideal gaseous and liquid solutions, there would be no heat evolved on mixing the components and the total volume of the solution must simply be the sum of the volumes of the pure constituents before mixing. We go into more detail in Chapter 14.

7.3.4 Two kinds of ideal solution

There is only one kind of ideal gas solution, as discussed above, but there are two kinds of ideal liquid and solid solutions.

Our discussion of liquid and solid solution ideality above makes no provision for the possibility that there might be more than one definition of ideality; that a solution might act ideally in one way but not in another. If in a liquid solution of A and B there are A–A and A–B interactions but no B–B interactions we have another kind of ideal solution, in which changing the concentration of B results in a perfectly linear change in the properties of B, and has no effect on the properties of A. In order for there to be no B–B interaction, particles of B must be quite widely separated – the concentration of B in A must be very small. This leads to the concept of the "infinitely dilute solution," in which there is only one particle of B in a sea of A, and therefore there is interaction between the particle of B and the surrounding A, and of course A–A interaction, but here being no other B particles, there is no B–B interaction. We can't deal with a single molecule of B, so we need to think of a mole of B particles and have so much A that no B particle is influenced by another B particle – perhaps a roomful of A, as in the room analogy on page 278.

These two kinds of ideality permeate discussions of liquid and solid solution properties, and are formalized by two ideal solution laws – Raoult's law and Henry's law.

7.4 Ideal solution laws

These relationships or laws were discovered in the nineteenth century by investigations of gas or vapor pressures associated with solutions of known composition. Because the gas or vapor was at a fairly low pressure, it acted as an ideal gas, and because it was in equilibrium with the solution, it provided information on the nature of the liquid solution. Today, the original connection with an associated vapor or gas phase is a secondary concern. The relationship between the ideal solution components themselves proves to be more useful, a subject to be discussed in terms of *activities*, an important topic introduced here and treated more fully in Chapter 8. Before discussing these relationships, we look first at solutions of ideal gases.

7.4.1 Dalton's law

The simplest imaginable system other than a vacuum is undoubtedly an ideal gas. One mole of ideal gas occupies 22.41 liters at 0 °C, 1 atm, so that (from the ideal gas law) one mole of ideal gas occupying one liter at 0 °C would have a pressure of 22.41 bars.

It was an early discovery (Dalton, 1811) that mixtures of gases would exert a pressure equal to the sum of the pressures that each of the species gases would have if each alone occupied the same volume. This was established using gases at relatively low pressures where they behave close to ideally, and in fact it is only strictly true for mixtures of ideal gases, which are also then ideal gases. Thus for each species gas 1, 2, 3, etc.

$$\begin{aligned}&P_1\mathbf{V} = n_1RT\\&P_2\mathbf{V} = n_2RT\\&\text{etc}\ldots\end{aligned}$$

and for the gas mixture

$$P_{\text{total}}\mathbf{V} = \sum_i n_iRT$$

Thus

$$\begin{aligned}&\frac{P_1}{P_{\text{total}}} = \frac{n_1}{\sum_i n_i} = x_1\\&\frac{P_2}{P_{\text{total}}} = \frac{n_2}{\sum_i n_i} = x_2\\&\text{etc}\ldots\end{aligned}$$

or,

$$\left.\begin{aligned}&P_1 = x_1 \cdot P_{\text{total}}\\&P_2 = x_2 \cdot P_{\text{total}}\\&\ \vdots \qquad\quad \vdots\\&\text{etc}\ldots\end{aligned}\right\} \tag{7.3}$$

P_1, P_2, etc., are called the *partial pressures* of the solution gases and equations (7.3) are now normally used as the definition of partial pressure even though in real, nonideal solutions they give a quantity that is not equivalent to the original meaning, i.e., the pressure a gas would exert if it alone occupied the total volume.[1]

7.4.2 Henry's law

Henry's law in its original form stated that the solubility of a gas in a liquid is proportional to the pressure on the gas. In Figure 7.1 is shown an apparatus for controlling the pressure on a gas i in contact with a liquid. As the pressure on the gas P_i increases, more of it dissolves in the liquid, and so x_i increases. When x_i is sufficiently small, it is directly proportional to P_i, and the constant

---

[1] There are several definitions of partial pressure. de Heer (1986, §23.4) says there are five, and explains the three most common. In my experience, only the definition in Equations (7.3) is ever used in geochemistry.

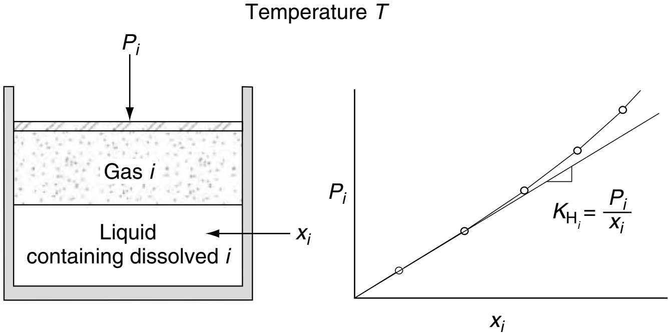


Figure 7.1 Illustration of Henry's law. As the pressure P_i on the gas i increases, more of it goes into solution in the liquid, increasing x_i.

of proportionality is called the Henry's law constant, K_{H_i}. As x_i gets larger, there is inevitably some deviation from strict proportionality, as shown.

Mathematically, this is expressed as

$$P_i = K_{\mathrm{H}_i} x_i \tag{7.4}$$

where P_i is the pressure or partial pressure of some component i, x_i is its mole fraction in solution, and K_{H_i} is a constant, specific for component i, the Henry's law constant. Actually, it is more often used in terms of molality,

$$P_i = K_{\mathrm{H}_i} m_i \tag{7.5}$$

where m_i is the molality of i in solution. K_{H_i} will have a different numerical value in the two cases. It follows that in the concentration range where Henry's law is obeyed,

$$\begin{aligned}\frac{\partial P_i}{\partial m_i} &= K_{\mathrm{H}_i} \\ &= \frac{P_i}{m_i}\end{aligned} \tag{7.6}$$

a result we will use later (§7.5.3).

In this experimental situation, it will be noted that the total pressure is not strictly speaking P_i, because some of the liquid solvent will evaporate into gas i, so that the piston is supported partly by gas i and partly by vaporized liquid. In other words there are always at least two partial pressures in a gas in contact with a liquid. However, if the vapor pressure of the liquid is small compared to the gas pressure, it can be neglected, and the pressure on the piston equated with P_i. This was the case in the early experiments of Henry and others.

Henry's law results from the lack of interaction between the solute particles, and represents the limiting behavior as solute concentrations approach zero. It has been generalized to refer not only to gas concentrations and pressures, but to any linear proportionality between the activity and concentration.

Finally, it is important to be aware of the physical meaning of the tangent in Figure 7.1. As $x_i \to 0$, it represents the values of P_i in equilibrium with the solution in which a particle of i interacts with the solvent but is unaware of any other particles of i; i is "infinitely dilute." As we go out along the tangent to larger values of x_i, the value of P_i given by the tangent at any x_i of course deviates from the actual measured P_i, but still represents the P_i which *would be* in equilibrium with i in solution, *if i continued to fail to be aware of other i particles*. It represents the P_i for a hypothetical solution of i which displays dilute solution behavior at all concentrations. This rather esoteric sounding situation proves to be surprisingly useful when we generalize P_i to the activity a_i (§8.3.4).

7.4.3 Raoult's law

Raoult's law originally concerned the composition of a vapor phase in equilibrium with a solution of two or more components. This sounds quite different from the Henry's law situation, but the two are intimately related. In fact, Raoult's law can be considered to be just a special case of Henry's law. Many combinations of components A and B (e.g., water and alcohol, or two organic liquids) were dissolved into one another in various proportions, and the composition and pressure of the coexisting vapor phase was measured (Figure 7.2). The results of these measurements varied widely, but a very few systems showed a particularly simple relationship. When the two liquids A and B were very similar, the vapor pressure of their mixture was a simple function of the vapor pressures of the pure liquids,

$$P_{\text{mixture}} = x_{\text{A}} P_{\text{A}}^{\circ} + x_{\text{B}} P_{\text{B}}^{\circ} \tag{7.7}$$

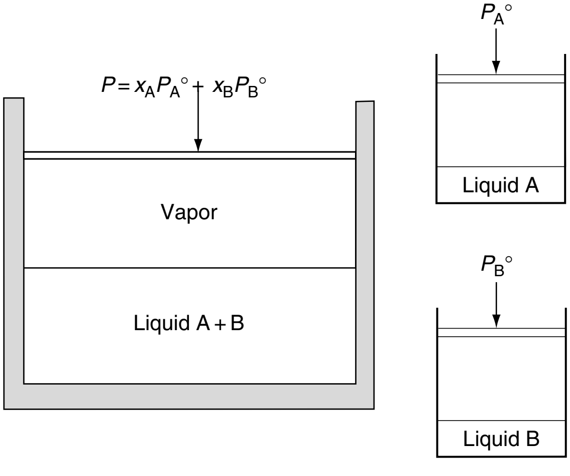


Figure 7.2 The vapor pressure of a solution of A and B that obeys Raoult's law.

and the partial pressures of A and B in the vapor were found to be directly proportional to their concentration in the liquid (Figure 7.2).

$$P_A = x_A^{\text{liquid}} P_A^\circ$$

$$P_B = x_B^{\text{liquid}} P_B^\circ$$

The only way that these simple relationships can hold is for the liquid solution to be ideal, in the sense discussed in §7.3.2. That is, intermolecular forces between A–A, B–B, and A–B must be identical, so that a molecule of A behaves exactly the same way whether it is surrounded mostly by A or mostly by B. Thus the relationship

$$P_i = x_i^{\text{liquid}} P_i^\circ \tag{7.8}$$

can be taken as a statement of Raoult's law, which means the solution is ideal. In Figure 7.3b we see that in an ideal solution of this type, P/P° for both components (i.e., P_A/P_A° and P_B/P_B°) are represented by diagonal lines, so that if we define *activity* as $a = P/P^\circ$, then activity equals mole fraction ($a_A = x_A$; $a_B = x_B$) in Raoultian systems. The concept of activity is actually a bit more complicated. We get to that in Chapter 8.

Raoult's law has therefore been generalized to refer, not only to the partial pressures of gases, but to any solution, including solid solutions, in which component activities equal their mole fractions.

Although some authors use different terminology to distinguish between Raoultian and Henryan ideality, many do not. We must always be clear whether we are referring to Raoultian or Henryan ideality.

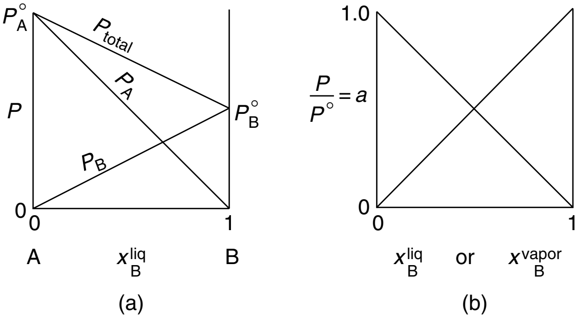


Figure 7.3 (a) The vapor pressure (P_{total}) of a binary solution that obeys Raoult's law is $P = P_A + P_B = x_A P_A^\circ + x_B P_B^\circ$. The partial pressure of each component is given by the diagonal lines, e.g., between 0 at $x_B = 0$ and P_B° at $x_B = 1$. (b) The partial pressure of each component divided by the vapor pressure of the pure component.

7.5 Ideal solution equations

7.5.1 Volume, enthalpy, heat capacity

These terms are simply additive in ideal solutions. Volume is the most intuitive of these terms, and we illustrate the additivity of volumes in ideal solutions in §10.2.1. This leads to Equation (10.4), the generalized form of which is

$$V_{\text{ideal sol'n}} = \sum_i x_i V_i^\circ \tag{7.9}$$

where V is the molar volume of an ideal solution, and V_i° is the molar volume of the pure component i.

The form of Equation (7.9) applies also to any thermodynamic parameter which does not contain entropy in its definition. The important ones are enthalpy and heat capacity, so that

$$H_{\text{ideal sol'n}} = \sum_i x_i H_i^\circ \tag{7.10}$$

and

$$C_{P\text{ideal sol'n}} = \sum_i x_i C_{Pi}^\circ \tag{7.11}$$

The difference between the property of a solution and the combined properties of the pure components is called a *mixing property*. For an ideal solution, from Equations (7.9)–(7.11)

$$\begin{aligned}\Delta_{\text{mix}} V_{\text{ideal sol'n}} &= V_{\text{ideal sol'n}} - \sum_i x_i V_i^\circ \\ &= \left(\frac{\partial \Delta_{\text{mix}} G}{\partial P}\right)_T \\ &= 0\end{aligned} \tag{7.12}$$

$$\begin{aligned}\Delta_{\text{mix}} H_{\text{ideal sol'n}} &= H_{\text{ideal sol'n}} - \sum_i x_i H_i^\circ \\ &= \left(\frac{\partial \Delta_{\text{mix}} G/T}{\partial (1/T)}\right)_P \\ &= 0\end{aligned} \tag{7.13}$$

$$\begin{aligned}\Delta_{\text{mix}} C_{P\text{ideal sol'n}} &= C_{P\text{ideal sol'n}} - \sum_i x_i C_{Pi}^\circ \\ &= \left(\frac{\partial \Delta_{\text{mix}} H}{\partial T}\right)_P \\ &= 0\end{aligned} \tag{7.14}$$

7.5.2 Entropy, Gibbs energy

Entropy and other potential quantities which contain entropy (such as G) are specifically defined so as to change in spontaneous processes, and two or more substances dissolving into one another is a perfect example of a spontaneous process.

Consider two ideal gases which are allowed to mix at constant P and T. In the final mixture, which is also an ideal gas, the partial pressure of gas 1 is $P_1 = x_1 P$, and of gas 2 is $P_2 = x_2 P$, where x_1 and x_2 are the mole fractions. The change in entropy on mixing, $\Delta_{\text{mix}} S$, is equal to the ΔS involved in expanding each gas from its initial pressure P to its partial pressure in the gas mixture. From Equation (5.41), this process is, for each gas

$$\Delta S_1 = R \ln(P/P_1)$$

and

$$\Delta S_2 = R \ln(P/P_2)$$

and the total change in entropy is

$$\begin{aligned}\Delta_{\text{mix}} S_{\text{ideal sol'n}} &= x_1 R \ln(P/P_1) + x_2 R \ln(P/P_2)\\ &= x_1 R \ln(1/x_1) + x_2 R \ln(1/x_2)\\ &= -R(x_1 \ln x_1 + x_2 \ln x_2) \qquad (7.15)\end{aligned}$$

or, in general terms,

$$\begin{aligned}\Delta_{\text{mix}} S_{\text{ideal sol'n}} &= S_{\text{ideal sol'n}} - \sum_i x_i S_i^\circ \qquad (7.16)\\ &= -R \sum_i x_i \ln x_i \qquad (7.17)\\ &= -\left(\frac{\partial \Delta_{\text{mix}} G}{\partial T}\right)_P \qquad (7.18)\end{aligned}$$

Because the x terms are fractional, $\Delta_{\text{mix}} S_{\text{ideal sol'n}}$ is inherently positive. Equation (7.17) can also be derived from reasonably simple statistical considerations which have nothing to do with the physical state of the particles. In other words it applies equally to ideal gas, liquid, and solid solutions.

It is important to note the fundamental difference between the ideal mixing of volumes and other terms not containing entropy (Equations 7.9, 7.10, 7.11), which are just linear combinations of the pure end-member terms, and the ideal entropy of mixing, Equations (7.16), (7.17), which are nonlinear, and result in all mixtures having a higher entropy than points on the $\sum_i x_i S_i^\circ$ line or plane. It is this property which gives the entropy, Gibbs energy and other thermodynamic potentials (all of which contain an entropy term, either as part of the definition or as a constraint) their ability to predict energy differences, and hence reaction directions.

The Gibbs paradox

Equation (7.17) has resulted in a quite remarkable literature on what is called “the Gibbs paradox.” Consider the case of two gases, A and B. The mixing expression for $-\Delta S/R$ is then $(x_A \ln x_A + x_B \ln x_B)$. Gibbs pointed out (Gibbs, Dover 1961 edition, p. 166), using slightly different expressions, that this is independent of the nature of the gases,

> …except that the gases which are mixed must be of different kinds. If we should bring into contact two masses of the same kind of gas, they would also mix, but there would be no increase in entropy.

This situation, which Gibbs explains on the same page, has been discussed by innumerable authors, often on a molecular or statistical basis. The best treatment, I think, is by E.T. Jaynes (1992). He says that

> Usually, Gibbs’ prose style conveys his meaning in a sufficiently clear way, using no more than twice as many words as Poincaré or Einstein would have used to say the same thing.

but that on this point he is more than usually obscure, although perfectly right.

Jaynes then examines the case where, unknown to us, there are actually two different kinds of argon, Ar1 and Ar2, identical in all respects, except that Ar2 is soluble in Whifnium, while Ar1 is not. Whifnium is one of the rare superkalic elements; in fact, it is so rare that it has not yet been discovered. Until the discovery of Whifnium in the next century, we would not know there were such things as Ar1 and Ar2, and even if by chance we happened to mix Ar1 and Ar2 we would have no way of knowing that they were two different entities, and we would correctly describe the mixing process as having zero change in entropy. After the discovery of Whifnium, we can prepare pure Ar1 and Ar2, and carry out the same mixing process, and we can even suppose that every molecule follows exactly the same path that might have happened previously by pure chance (when we knew nothing about Ar1 and Ar2). But in this case the entropy change is given by Equation (7.17), and is nonzero, perfectly illustrating why there is a paradox. Jaynes then goes on to discuss Whafnium, in which Ar1 is soluble but Ar2 is not.

In his usual elegant style, Jaynes uses this situation to clarify not only the macroscopic nature of thermodynamics, but the role of information, reversible and irreversible mixing and the work available. Particularly interesting is his use of this example to emphasize the importance of defining the constraints (which he calls macrovariables) defining a system. When we know about Ar1 and Ar2 and are able to separate them by doing work on the system, we have an extra constraint in the sense of §4.9.2.

It follows from Equations (5.1), (7.16), and the fact that $\Delta_{\text{mix}}H_{\text{ideal sol'n}} = 0$ (Equation 7.13), that the Gibbs energy of (ideal) mixing is

$$\begin{aligned}\Delta_{\text{mix}}G_{\text{ideal sol'n}} &= \Delta_{\text{mix}}H_{\text{ideal sol'n}} - T\Delta_{\text{mix}}S_{\text{ideal sol'n}} \\ &= G_{\text{ideal sol'n}} - \sum_i x_i G_i^\circ \qquad (7.19) \\ &= RT\sum_i x_i \ln x_i \qquad (7.20)\end{aligned}$$

The $\sum_i x_i G_i^\circ$ term defines a straight line (or plane surface) between points representing end-member components, and the $\sum_i x_i \ln x_i$ term (which is inherently negative) describes how far *below* this line or plane is the surface representing the G of the (ideal) solution.

Figure 7.4 shows $\Delta_{\text{mix}}H$, $\Delta_{\text{mix}}S$, and $\Delta_{\text{mix}}G$ for an ideal solution. There are several things to note in this diagram.

1. Because the mixing of A and B takes place at a constant T and P and is a spontaneous process, $\Delta_{\text{mix}}G$ must be negative. The curve shown is an expression of Equation (7.20). No experimental data are required; just mole fraction numbers from 0 to 1.
2. The $\Delta_{\text{mix}}S$ shown is an expression of Equation (7.16), but because it is a much smaller quantity (reaching a maximum of 5.76 J mol^{-1} K^{-1} at $x_B = 0.5$) it is exaggerated in the diagram.
3. $\Delta_{\text{mix}}H$ is from Equation (7.13) and is of course zero at all x_B. $\Delta_{\text{mix}}G$ is therefore a mirror reflection of $T\Delta_{\text{mix}}S$, because $\Delta_{\text{mix}}G = \Delta_{\text{mix}}H - T\Delta_{\text{mix}}S$. The $T\Delta_{\text{mix}}S$ curve is not shown.

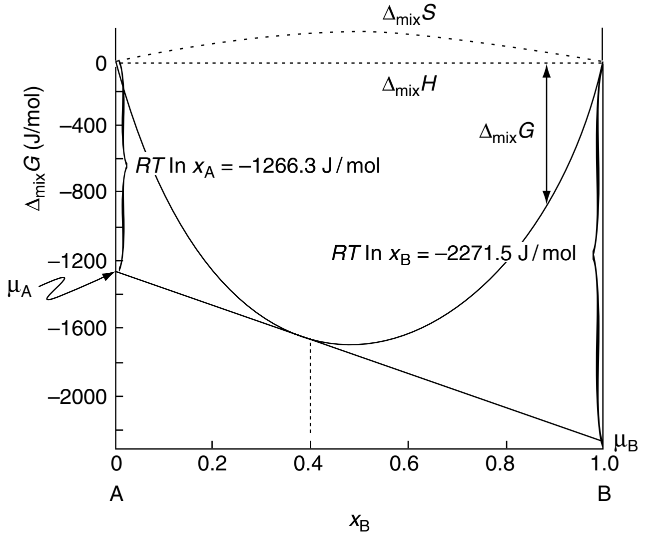


Figure 7.4 Gibbs energy of ideal mixing, from Equation (7.20).

4. We don't have absolute values for G, so we must always measure the difference in G from some other state at the same T and P. In Figure 7.4 this other state is pure A ($x_A = 1$) and pure B ($x_B = 1$).
5. Despite appearances, the molar Gibbs energy curve in Figure 7.4 is actually asymptotic to each vertical axis, whereas the volume curve in Figure 10.2 is not. This is not of great importance by itself, but it is connected to the fact that we cannot use the infinite dilution standard state for Gibbs energy. We leave this important topic to Chapter 8.

Chemical potentials

Figure 7.4 also introduces us to a new problem. Because the mixing curve is concave downward, the Gibbs energies of components A and B in the solution are necessarily less than the corresponding values of G_A° and G_B°, the molar Gibbs energies of the pure compounds. The fact that this is so provides the (thermodynamic) reason why A and B form a solution. If we mix n_A moles of A and n_B moles of B the reaction is

$$n_A\,\mathrm{A} + n_B\,\mathrm{B} = (n_A\,\mathrm{A},\, n_B\,\mathrm{B})_{\text{solution}} \tag{7.21}$$

and the Gibbs energy change for this reaction is $\Delta_{\text{mix}}G$, which is negative. During this reaction, the Gibbs energy of both A and B become lower. If the mixing line lies *above* the straight line joining G_A° and G_B°, then ΔG_{mix} would be positive, the dissolution reaction (7.21) would not be spontaneous, and no solution would form – A and B would be immiscible, like oil and water.

So $G_A^{\text{solution}} < G_A^\circ$, and $G_B^{\text{solution}} < G_B^\circ$. But this raises a few questions, like what are these quantities G_A^{solution} and G_B^{solution}? Where are they on the diagram? How can you separate ΔG_{mix} into these two separate quantities? If we know that, then perhaps we could evaluate $G_A^\circ - G_A^{\text{solution}}$ and $G_A^\circ - G_A^{\text{solution}}$.

You can see by simple inspection of Figure 7.4 that at any point on the $\Delta_{\text{mix}}G$ curve, the tangent at that point allows calculation of the numerical value of that point on the curve by combining two points on the tangent in a linear combination. In Figure 7.4 we show the tangent to the curve at $x_B = 0.4$, and identify the intersections of the tangent with the two ordinate axes as μ_A and μ_B. In Equation (7.19) we see that if G_A° and G_B° are zero, then $\Delta_{\text{mix}}G = G_{\text{solution}}$, and in drawing Figure 7.4, this is exactly what we have assumed. We have made $\Delta_{\text{mix}}G$ the difference between the G of the solution and zero. Evidently then,

$$G^{\text{solution}} = x_A\mu_A + x_B\mu_B \tag{7.22}$$

But what is the physical meaning of μ? There are two ways of answering this.

Euler's answer

Mathematically minded people simply invoke Euler's theorem for homogeneous functions (§C.2.3). In plain language, this says that for any extensive (total) function such as **V** or in this case **G**, having n_A moles of component A and n_B moles of component B

$$\mathbf{G} = n_\mathrm{A}\left(\frac{\partial \mathbf{G}}{\partial n_\mathrm{A}}\right)_{n_\mathrm{B}} + n_\mathrm{B}\left(\frac{\partial \mathbf{G}}{\partial n_\mathrm{B}}\right)_{n_\mathrm{A}} \tag{7.23}$$

or, dividing by $(n_\mathrm{A} + n_\mathrm{B})$,

$$G = x_\mathrm{A}\left(\frac{\partial \mathbf{G}}{\partial n_\mathrm{A}}\right)_{n_\mathrm{B}} + x_\mathrm{B}\left(\frac{\partial \mathbf{G}}{\partial n_\mathrm{B}}\right)_{n_\mathrm{A}} \tag{7.24}$$

Comparing (7.22) and (7.24), we have

$$\left.\begin{aligned} \mu_\mathrm{A} &= \left(\frac{\partial \mathbf{G}}{\partial n_\mathrm{A}}\right)_{T,P,n_\mathrm{B}} \\ \mu_\mathrm{B} &= \left(\frac{\partial \mathbf{G}}{\partial n_\mathrm{B}}\right)_{T,P,n_\mathrm{A}} \end{aligned}\right\} \tag{7.25}$$

so the tangent intersections are in fact the partial derivatives of the total Gibbs energy of the solution **G** with respect to n_A and n_B. These are in fact our first examples of *partial molar* terms, introduced in §2.4.1 and §4.14.1. μ is a partial molar Gibbs energy, and it is our answer to the question, what is the Gibbs energy per mole of a dissolved substance? It is important to develop an intuitive understanding of partial molar terms, so we devote quite a bit more discussion to this in Chapter 10.

Our answer

But we don't have to use Euler's theorem. We can simply expand our definition of **G**, which so far is restricted to closed (constant composition) systems. If we exclude chemical work, which means we deal only with systems at complete stable equilibrium, we know from Equation (4.65)

$$d\mathbf{G} = -\mathbf{S}dT + \mathbf{V}dP + \sum_i \mu_i dn_i$$

At constant T and P [by implication a condition of Equations (7.23) and (7.24)] the first two terms on the right drop out, and integrating the other terms from an increment of solution up to the whole system, we get, for two components 1 and 2,

$$\mathbf{G} = n_1\left(\frac{\partial \mathbf{G}}{\partial n_1}\right)_{T,P,\hat{n}_1} + n_2\left(\frac{\partial \mathbf{G}}{\partial n_2}\right)_{T,P,\hat{n}_2}$$

which is Equation (7.23), and from there we get Equations (7.25) again.

So that answers one question–how do we split G^solution into $G_\mathrm{A}^\mathrm{solution}$ and $G_\mathrm{B}^\mathrm{solution}$, and where are these things on the diagram? The next question is, how do we evaluate $G_\mathrm{A}^\circ - G_\mathrm{A}^\mathrm{solution}$ and $G_\mathrm{B}^\circ - G_\mathrm{B}^\mathrm{solution}$?

7.5.3 The relation between composition and free energy

Figure 7.4 also shows a very important relationship between Gibbs energy and composition, namely, that

$$\left.\begin{aligned}\mu_{\mathrm{A}} - G_{\mathrm{A}}^{\circ} &= RT\ \ln x_{\mathrm{A}}\\ \mu_{\mathrm{B}} - G_{\mathrm{B}}^{\circ} &= RT\ \ln x_{\mathrm{B}}\end{aligned}\right\} \tag{7.26}$$

Equations (7.26) provide a relationship between the concentration of an ideal solution component and its Gibbs energy. This is an important milestone. Equations (4.42) and (4.43)

$$(\partial G/\partial T)_P = -S \qquad [4.42]$$

$$(\partial G/\partial P)_T = V \qquad [4.43]$$

showed how Gibbs energy varies with T and P, respectively (expanded upon in Figures 6.9, 6.14); now we can see how Gibbs energy varies with concentration of something in solution. If we can calculate the Gibbs energy of solids, liquids, gases, and solutes over a range of T, P, and composition (x), we have just about solved all our problems, in principle. Basically, from here on we will be amplifying and coming to grips with practical matters, such as the fact that Equation (7.26) only applies to Raoultian solutions.

But where do Equations (7.26) come from?

Gibbs energy and mole fraction I

The most direct way to derive the relationship between Gibbs energy and mole fraction is to simply differentiate the total form of Equation (7.20) with respect to n_i. For two components, $G = \mathbf{G}/(n_1 + n_2)$, so multiplying both sides of (7.20) by $(n_1 + n_2)$, then differentiating, gives

$$\frac{\partial}{\partial n_1}(\mathbf{G}_{\text{sol'n}} - n_1 G_1^{\circ} - n_2 G_2^{\circ}) = \frac{\partial}{\partial n_1}[RT\,(n_1 \ln x_1 + n_2 \ln x_2)]$$

which, with n_2, G_1°, and G_2° constant, gives

$$\begin{aligned}\mu_1 - G_1^{\circ} &= RT\left[\frac{\partial}{\partial n_1}(n_1 \ln x_1) + \frac{\partial}{\partial n_1}[n_2 \ln(1 - x_1)]\right]\\ &= RT\,(\ln x_1 + x_2 - x_2)\\ &= RT \ln x_1 \end{aligned} \tag{7.27}$$

which is (7.26) for component 1.

Applying the same method to Equation (7.15) we get

$$\overline{S}_i - \overline{S}_i^{\circ} = -R \ln x_i \tag{7.28}$$

and recall that $\Delta_{\text{mix}} H = 0$, so

$$\overline{H}_i - \overline{H}_i^{\circ} = 0 \tag{7.29}$$

as well, for ideal solutions.

There is a complication in (7.27) and (7.28) which is actually very general but will only become important in the study of solid solutions (Chapter 14). We discuss it in more detail in §9.10.2, but briefly, the problem is that in Equation (7.27) the left side is a difference in energy *per mole*, which obviously depends on how the mole is defined, but the right side has a mole fraction, which is independent of the definition of the mole as long as all components of the solution are treated in the same way.

For example, normally we define a mole of nitrogen as Avogadro's number of N_2 in nitrogen gas, and oxygen as O_2. But we could define these as the same number of N and O, or of N_4 and O_4, without affecting the mole fraction. Whether these forms exist or not is not relevant. The mole fraction of N_2 in a solution of nitrogen and oxygen does not depend on which of these ways we define the mole, but the value of μ does. Avogadro's number of N_4 particles has twice the mass and twice the energy of the same number of N_2 particles, so

$$\mu_{N_4} - \mu^{\circ}_{N_4} = 2[\mu_{N_2} - \mu^{\circ}_{N_2}] \tag{7.30}$$

Therefore, if there is any question as to the size of the mole, which usually only arises in defining components in solid solutions, Equation (7.27) is generalized to

$$\begin{aligned} \mu_1 - G^{\circ}_1 &= n\,RT\ln x_1 \\ &= RT\ln x^n_{x_1} \end{aligned} \tag{7.31}$$

where n is the factor relating the two definitions of the mole.

Gibbs energy and mole fraction II

It will also prove useful to derive (7.27) another way, especially when we use molal units of concentration instead of mole fractions. So first we will re-derive (7.27), and then use the same method for molal units in §8.2.3.

Equation (7.27) expresses the relationship between μ and x, or concentration. To derive this, it would seem natural to find an expression for the derivative of μ with a concentration term, and then integrate. In other words, what is the value of $(\partial\mu_i/\partial n_i)_{\hat{n}_i}$? How does μ_i (the G of i in solution) vary with the amount of i in solution in the ideal case? This is a partial derivative, so if n_i is the number of moles of i, we need to keep the concentrations of all other components constant. We denote all other components by $\hat{n}_i$.

If we expand $(\partial\mu_i/\partial n_i)_{T,P,\hat{n}_i}$ by introducing P_i, the pressure on gaseous i which is, or might be, in equilibrium with solute i (whether or not there is such a gas phase is irrelevant), we get

$$\left(\frac{\partial\mu_i}{\partial n_i}\right)_{\hat{n}_i} = \frac{\partial\mu_i}{\partial P_i}\frac{\partial P_i}{\partial n_i} \tag{7.32}$$

where μ_i is the same in the solution and in the (perhaps hypothetical) gas phase, where it can be called G_i (the gas being assumed ideal), so that $(\partial\mu_i/\partial P_i) = (\partial G_i/\partial P_i) = V_i = RT/P_i$, and where $(\partial P_i/\partial n_i) = P_i/n_i$ is an expression of Henry's law (Equation 7.6).[2] Combining all this we get

$$\left(\frac{\partial\mu_i}{\partial n_i}\right)_{\hat{n}_i} = \frac{RT}{n_i} \tag{7.33}$$

for ideal solutions. Integrating this equation between two values of n_i, n_i' and n_i'', we get

$$\begin{aligned}\mu_i'' - \mu_i' &= RT\ln\frac{n_i''}{n_i'}\\ &= RT\ln\frac{P_i''}{P_i'} \quad \text{by Henry's law}\\ \mu_i'' - \mu_i^\circ &= RT\ln\frac{P_i''}{P_i^\circ} \quad \text{if state } ' \text{ is pure } i\end{aligned}$$

and, from Equation (7.8)

$$\mu_i - \mu_i^\circ = RT\ln x_i \quad \text{(state } '' \text{ no longer needs superscript)} \tag{7.34}$$

which is Equation (7.26) for component *i*.

When $x_B = 0.4$ and $x_A = 0.6$, $R = 8.31451\,\mathrm{J\,mol^{-1}\,K^{-1}}$, and $T = 298.15\,\mathrm{K}$, Equation (7.26) gives

$$\mu_A - G_A^\circ = -1266\,\mathrm{J\,mol^{-1}}$$

and

$$\mu_B - G_B^\circ = -2271\,\mathrm{J\,mol^{-1}}$$

This says that a mole of A has 1266 $\mathrm{J\,mol^{-1}}$ less in the dissolved state than in the pure state, and this is the "thermodynamic explanation" for why A dissolves in B.

Constructing the tangents

Although not essential to understanding the thermodynamics involved, it is of some interest to know how to calculate the position of the tangents and the chemical potential intercepts in Figure 7.4.

---

[2] OK, Equation (7.6) shows the ratio of P_i to m_i, not n_i. But as m_i is just n_i per kilogram of solvent, the ratio is the same.

The tangent is the slope of the curve calculated from Equation (7.20). Therefore the equation for the slope of the tangent is (not forgetting that $x_A = 1 - x_B$)

$$\left(\frac{d\Delta_{\text{mix}}G}{dx_B}\right) = \frac{d}{dx_B}\,[RT(x_A \ln x_A + x_B \ln x_B)]$$

$$= RT \ln\left(\frac{x_B}{x_A}\right) \tag{7.35}$$

and the equation for the tangent line itself is

$$\mu_A = \Delta_{\text{mix}}G - x_B\left(\frac{d\Delta_{\text{mix}}G}{dx_B}\right) \tag{7.36}$$

If you substitute (7.35) into (7.36) you get, not surprisingly, $\mu_A = RT \ln x_A$, which is Equation (7.26), except that we take $\mu^\circ = 0$ in the diagram. This may make the whole thing look like an exercise in algebra. Well, it is, but it is algebra that is exceptionally useful as a basis for understanding solutions.

7.6 Next step – the activity

We have now developed the relationship between the Gibbs energy of a component of a solution and the concentration of that component (Equations 7.26, 7.27, 7.34). However, it only applies to ideal solutions, and only for concentrations in mole fractions. Obviously we need to expand the range of applicability of this relationship.

Doing this gets complicated, because we have gaseous, liquid and solid solutions, a variety of concentration scales, nonideal solutions, and several different standard states that μ° refers to. That is, the quantity $\mu_i - \mu_i^\circ$ need not always refer to the difference between i in solution and i in its pure state. At the same time, the form of Equation (7.34) is very convenient, and we want to retain it for all these conditions. We do this by defining the *activity*, already mentioned in §7.4.3, as

$$\mu_i - \mu_i^\circ = RT \ln a_i \tag{7.37}$$

All the complications are accommodated by this parameter, and we try to sort it all out in Chapter 8.

7.7 Summary

In Chapter 4 (§4.12.1) we saw that

$$(\partial G/\partial T)_P = -S \qquad [4.42]$$

and

$$(\partial G/\partial P)_T = V \qquad [4.43]$$

We didn't bother to write, though it is equally true, that

$$(\partial \mathbf{G}/\partial T)_P = -\mathbf{S} \tag{7.38}$$

and

$$(\partial \mathbf{G}/\partial P)_T = \mathbf{V} \tag{7.39}$$

and, differentiating by n_i, that

$$(\partial \mathbf{S}/\partial n_i)_{T,P} = \overline{S_i} \tag{7.40}$$

and

$$(\partial \mathbf{V}/\partial n_i)_{T,P} = \overline{V_i} \tag{7.41}$$

meaning that any solution property, not just Gibbs energy, can be split up into the contributions of the individual components. Finally, because

$$\frac{\partial}{\partial T}\left(\frac{\partial \mathbf{G}}{\partial n_i}\right) = \frac{\partial}{\partial n_i}\left(\frac{\partial \mathbf{G}}{\partial T}\right)$$

and similarly for P, then

$$\frac{\partial \mu_i}{\partial T} = -\overline{S_i}$$

and

$$\frac{\partial \mu_i}{\partial P} = \overline{V_i}$$

So it turns out that

$$\left.\begin{aligned}(\partial \mu_i/\partial T)_{P,n} &= -\overline{S_i}\\(\partial \mu_i/\partial P)_{T,n} &= \overline{V_i}\\(\partial \mu/\partial n_i)_{T,P,\hat{n}_i} &= RT/n_i\end{aligned}\right\} \tag{7.42}$$

are the relationships we need to be able to evaluate (by integration) to know the Gibbs energy of any substance as a function of T, P, and composition. But we're not finished yet. The first two equations apply to any solution, but the $(\partial \mu/\partial n_i)$ equation deals only with ideal solutions. We need to find ways to evaluate the partial molar properties in the first two equations, and how to modify the third one so that it works for any solution. We do this in Chapter 10. A central concept in all of this is the activity, which we take up in Chapter 8.

8

Fugacity and activity

8.1 Fugacity

8.1.1 Introduction

The fugacity was introduced by G.N. Lewis in 1901, and became widely used after the appearance of *Thermodynamics*, a very influential textbook by Lewis and Randall in 1923. Lewis describes the need for such a function in terms of an analogy with temperature in the attainment of equilibrium between phases. Just as equilibrium requires that heat must flow such that temperature is the same in all parts of the system, so matter must flow such that chemical potentials are also equalized. He referred to the flow of matter from one phase to another as an "escaping tendency," such as a liquid escaping to the gas form to achieve an equilibrium vapor pressure. He pointed out that in fact vapor pressure is equilibrated between phases under many conditions (and in fact is the basis for the isopiestic method of activity determinations, §5.8.4), and could serve as a good measure of escaping tendency if it behaved always as an ideal gas.

The chemical potential is of course another measure of "escaping tendency," but Lewis pointed out that there are "certain respects in which this function is awkward." This refers to the fact that $\mu \to -\infty$ as $a \to 0$, because activity is defined as $\mu - \mu^\circ = RT \ln a$. Lewis defined a function which would be much like a vapor pressure, which would be equilibrated between phases at equilibrium, even in nonideal cases, and even if no vapor phase actually existed. It is an absolute property, in the sense that it does not depend on a standard state, and it has units of pressure. Lewis and Randall (1923) called it a kind of "ideal or corrected vapor pressure."

Fugacity has proved useful in a number of ways. One way is to provide a relatively simple way to evaluate the integral $\int V dP$. In §5.7.1 we saw one way to do this. That is, for solids, we often assume that the molar volume is constant, making the integration very simple. Another way, for gases, is to assume the ideal gas law (see below). This is actually a special case of the most general method, which is to develop an equation of state for the system (Chapter 13), from which you can generate all its thermodynamic properties.

Lacking an equation of state, how do we evaluate the pressure integral for a fluid such as H_2O or CO_2, either in the pure form, mixed with other fluid components, or reacting with solid phases? A possible way to proceed would be to express V as a function of P in some sort of power series, just as we did for C_P as a function of T (Equation 3.26). $\int V\,dP$ could then be integrated, and we could determine the values of the power series coefficients for each gas or fluid and tabulate them as we do for the Maier–Kelley coefficients.

8.1.2 Definition of fugacity

Fortunately, thanks to the insight of Lewis, we can proceed in a simpler and completely different fashion. To see how the inspiration for such a function might have arisen, consider the form of the volume integral $\int V\,dP$ for an ideal gas. Starting with (4.43),

$$\begin{aligned} dG &= V\,dP \\ &= \frac{RT}{P}dP \\ &= RTd\ln P \qquad (8.1) \end{aligned}$$

$$\int_{P_1}^{P_2} dG = \int_{P_1}^{P_2} \frac{RT}{P}dP$$

$$G_{P_2} - G_{P_1} = RT\ln\frac{P_2}{P_1} \qquad (8.2)$$

If P_1 is 1 bar and this is designated a standard or reference state denoted by a superscript °, then P_2 becomes simply P, and

$$\begin{aligned} G - G^\circ &= RT\ln\frac{P}{P^\circ} \\ &= RT\ln P \qquad \text{since } P^\circ = 1 \qquad (8.3) \end{aligned}$$

Thus for ideal gases $RT\ln P$ all by itself gives the value of $\int_{P=1}^{P} dG$. Unfortunately, this doesn't work for real gases although it's not a bad approximation at low pressures and high temperatures where real gases approach ideal behavior. However, the *form* of the relationship

$$dG = V\,dP = RT\,d\ln P$$

(Equation 8.1) is sufficient to suggest that we could define a function such that the relationship *would* hold true for real gases. This function is the fugacity, f, where

$$dG = V\,dP = RTd\ln f \qquad (8.4)$$

and

$$\int_{P_1}^{P_2} dG = \int_{P_1}^{P_2} V\,dP$$

$$G_{P_2} - G_{P_1} = \int_{P_1}^{P_2} RT\,d\ln f$$

$$= RT \ln \frac{f_{P_2}}{f_{P_1}} \tag{8.5}$$

Because f appears as a ratio in (8.5), this equation cannot serve as a full definition of f. We must specify how to determine the fugacity at some P and T, then the integral can be used to calculate it at other pressures at that T. To complete the definition, it would be convenient for f to approximate P for gases that behave more or less ideally, that is, at low values of P. We accomplish this by stipulating that

$$\lim_{P\to 0}\left(\frac{f}{P}\right) = 1 \tag{8.6}$$

This means that for an ideal gas, $f = P$, and for gases at low pressures, $f \approx P$. Equations (8.6) plus (8.4) or (8.5) make up the definition of fugacity.

Fugacity coefficient

The ratio f/P is called the fugacity coefficient, γ_f. Thus

$$f_i = \gamma_{f_i} P \tag{8.7}$$

where P is the pressure of a pure fluid compound, or the *partial pressure* of a compound in a solution. The partial pressure is a measure of concentration, so the fugacity coefficient is another kind of *activity coefficient*, introduced in §10.6.1.

8.1.3 Calculation of fugacity

From the preceding equations we see that to measure fugacity, we need to know the molar volume as a function of P. In other words we have to measure gas densities. Also we should measure these densities down to very low pressures – pressures sufficiently low that we can say $f = P$ to whatever degree of accuracy is required. But at these low pressures, the volume of a given mass of gas becomes very large and very difficult to measure accurately. Thus, while in principle evaluating $\int V\,dP$ is straightforward, in practice it is difficult when the lower limit of integration is a very low pressure. To avoid evaluating this integral at very low pressures where $V \to \infty$, it is convenient to define a *residual function* α, where

$$\alpha = V^{\text{real}} - V^{\text{ideal gas}} \tag{8.8}$$

$$= V - RT/P$$

Thus α is the difference between the molar volume of the gas and the volume it would have if it behaved ideally. Then

$$V = \alpha + RT/P$$

and

$$\begin{aligned} dG &= V\,dP \\ &= RT\,d\ln f \\ RT\,d\ln f &= (\alpha + RT/P)dP \\ d\ln f &= d\ln P + (\alpha/RT)dP \end{aligned}$$

Integrating from P_0 to P

$$\int_{P_0}^{P} d\ln f = \int_{P_0}^{P} d\ln P + \frac{1}{RT}\int_{P_0}^{P} \alpha\,dP$$

$$\ln f_P - \ln f_{P_0} = \ln P - \ln P_0 + \frac{1}{RT}\int_{P_0}^{P} \alpha\,dP \tag{8.9}$$

where P_0 means some unspecified value of P, sufficiently low such that $P = f$ to a good approximation. At this low value of P, the two terms $\ln f_{P_0}$ and $\ln P_0$ are equal, so that Equation (8.9) becomes

$$\ln f = \ln P + \frac{1}{RT}\int_{P_0}^{P} \alpha\,dP$$

and

$$\begin{aligned} \ln\frac{f}{P} &= \ln\gamma_f \\ &= \frac{1}{RT}\int_{P_0}^{P}\left(V - \frac{RT}{P}\right)dP \end{aligned} \tag{8.10}$$

This is for a pure gas or fluid. If the gas is in a mixture of gases, total pressure P becomes the partial pressure $x_i P$, and molar volume V becomes the partial molar volume $\overline{V}_i$, so

$$\ln\gamma_{f_i} = \ln\frac{f_i}{x_i P} = \frac{1}{RT}\int_{P_0}^{P}\left(\overline{V}_i - \frac{RT}{P}\right)dP \tag{8.11}$$

This is Equation (3.14) in Prausnitz et al. (1999), a standard reference on fugacities in fluids. Equation (8.10) shows that for an ideal gas, in which $V = RT/P$, $f = P$, that is, fugacity is the same as pressure, which we also saw from Equation (8.6). Similarly, Equation (8.11) shows that in an ideal mixture of gases in which $\overline{V}_i = RT/P$, fugacity is equal to partial pressure.

Equation (8.11) has P and T as independent variables. More commonly, however, P is known as some function of V and T. In this case Equation (8.11) becomes

$$\ln\gamma_{f_i} = \ln\frac{f_i}{x_iP} = \frac{1}{RT}\int_V^\infty\left[\left(\frac{\partial P}{\partial n_i}\right) - \frac{RT}{P}\right]dV - RT\ln Z \tag{8.12}$$

where Z is the compressibility factor (§13.2.4). This is Equation (3.53) in Prausnitz et al. (1999).

Because

$$\frac{Z-1}{P} = \frac{V-RT/P}{RT}$$

we can rewrite Equation (8.10) as

$$\ln\gamma_f = \int_{P_0}^{P}(Z-1)d\ln P \tag{8.13}$$

so the fugacity coefficient can be calculated as a function of Z, the compressibility factor, and since Z is known and is the same for most gases in terms of their P_r and T_r, then f/P can be calculated for any gas given its P_c and T_c. Several authors have prepared charts showing f/P as a function of P_r at various T_r. This is a useful device for obtaining fugacities of gases in the lower range of geologically interesting P–T conditions. The charts do not extend far into the metamorphic range of P–T conditions, but nevertheless the compressibility factor has found very wide use in equations of state and other treatments of fugacity.

The Lewis fugacity rule

In Chapter 7 we saw that if substances mix ideally ($\Delta_m V = 0$),

$$V = x_\mathrm{A}\,V_\mathrm{A}^\circ + x_\mathrm{B}\,V_\mathrm{B}^\circ \qquad [10.4]$$

This volumetric ideal mixing is known as Amagat's law. If substances mix nonideally ($\Delta_m V \neq 0$),

$$V = x_\mathrm{A}\,\overline{V}_\mathrm{A} + x_\mathrm{B}\,\overline{V}_\mathrm{B} \qquad [10.7]$$

It follows, because (10.7) is perfectly general, that for ideal solutions, the partial molar volume is the same as the molar volume, $V_i = \overline{V}_i$. Therefore, comparing (8.10) and (8.11), we see that

$$\frac{f^\mathrm{pure}}{P} = \frac{f^\mathrm{mix}}{x_iP}$$

or

$$f_i^\mathrm{mix} = x_i\cdot f^\mathrm{pure} \tag{8.14}$$

This is the Lewis fugacity rule, and is seen to be true if Amagat's law is true. So fluids can mix ideally volumetrically, but might still be nonideal mixtures. Note that it assumes the additivity of the molar volumes of all components of

Calculation of fugacity

What is the fugacity of water at 800 °C between 1 and 1000 bars?

Using Equation (8.10) and volume data from the NIST program STEAM, v. 2.2 the calculated values of $V - RT/P$ are shown in Figure 8.1a. Integrating this curve from $P_0 = 1$ bar (where $\gamma_f = 0.9998$, close enough to ideal gas) to various pressures gives values of $RT \ln (f/P)$ as shown in Figure 8.1b.

To do the integration, volume data were obtained from the NIST program at 0.1 bar intervals from 1 to 10 bars, and at 10 bar intervals from 10 to 1000 bars. Integration was done numerically in MATLAB®, after fitting the curve with a spline function. If you try this, don't forget to get V, RT/P, and P in compatible units. The easiest way is to change volumes in $\text{cm}^3\,\text{mol}^{-1}$ to $\text{J}\,\text{bar}^{-1}\,\text{mol}^{-1}$ by multiplying by 0.1.

The results at various pressures are shown below, compared with values calculated by the NIST program.

P bars	$RT \ln (f/P)$ $\text{J}\,\text{mol}^{-1}$	γ_f —	f bars	f from NIST bars
1	0.000	1.0	0.9998	1.0
100	−160.772	0.9821	98.2	98.2
200	−322.532	0.9645	192.9	192.9
300	−482.704	0.9473	284.2	284.2
400	−639.849	0.9308	372.3	372.3
500	−792.594	0.9150	457.5	457.5
600	−939.696	0.9000	540.0	540.0
700	−1080.076	0.8860	620.2	620.1
800	−1212.853	0.8729	698.3	698.3
900	−1337.339	0.8608	774.7	774.7
1000	−1453.034	0.8497	849.7	849.7

the mixture, not only at pressure P but at all pressures from P_0 to P. In effect, it proposes that the fugacity coefficient of i is independent of composition and of the nature of the other components in the solution. The Lewis fugacity rule is often used in the absence of better alternatives, but can be very inaccurate for small concentrations (partial pressures) of i. It becomes more accurate under certain limiting conditions (see Prausnitz et al., 1999, Chapter 5).

8.1.4 Fugacities from equations of state

Determining fugacities in fluid mixtures is essentially equivalent to determining an equation of state for the mixture. Deriving thermodynamic properties from

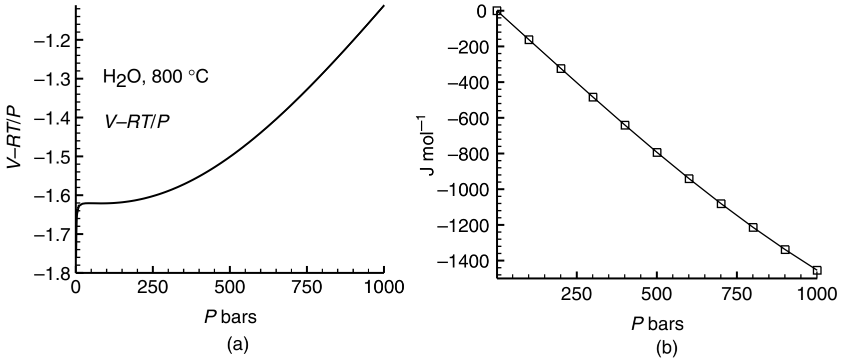


Figure 8.1 (a) The value of $\alpha = V - RT/P$ for water at 800 °C, from 1 bar to 1000 bars, from the NIST program STEAM. (b) The value of the integral $\int_1^P \alpha\, dP$ for water at 800 °C, i.e., the area enclosed by the curve in (a).

that equation is then straightforward, if often computationally tedious. We will look at only a couple of examples.

Fugacities from the van der Waals equation

Thermodynamic properties can be derived from any equation of state, but because of the differentiation and integration involved, the resulting expressions rapidly get surprisingly complex. For example, Prausnitz et al. (1999, Equation 3.65) show that the expression for the fugacity of i in a mixture of gases obeying the relatively simple van der Waals equation is

$$\begin{aligned} RT\ln\frac{f_i}{x_iP} &= RT\ln\frac{\mathbf{V}}{\mathbf{V}-n_Tb} + n_TRT\left[\frac{\partial(n_Tb)}{\partial n_i}\right]\frac{1}{(\mathbf{V}-n_Tb)} \\ &\quad - \left[\frac{\partial(n_T^2a)}{\partial n_i}\right]\frac{1}{\mathbf{V}} - RT\ln Z \end{aligned} \tag{8.15}$$

where n_T is the total number of moles of gas components, and a and b are the van der Waals constants for the mixture. But here another problem arises. The a and b parameters are known for each of the individual gases in the mixture, but how are they combined to give an overall a and b for the mixture of gases? This is the problem of the *mixing rules*, encountered for all equations of state of this type, which have measured or fit parameters for individual gases, which must be combined in some fashion when the gases are mixed. There is no "correct" way of combining these parameters, and a mixing rule which works in one case may not work in another.

Properties derived from extensions of the van der Waals equation get even more complex. For example, the equation for fugacity coefficients in H_2O–CO_2 mixtures using the MRK Equation (13.17) occupies a whole page (Kerrick and Jacobs, 1981).

Fugacities from virial equations

Another commonly used equation of state is the virial equation. We discuss the calculation of fugacity from virial coefficients in Chapter 13, §13.5.1.

Fugacity standard states

The fact that fugacities often appear in the ratio f/f° has led to a common confusion about "fugacity standard states." For example in one textbook the authors say

> The fugacity is a relative function because its numerical value is always relative to that of an ideal gas at unit fugacity; in other words, the standard state fugacity f_i° in [...the equation $\mu_i - \mu_i^\circ = RT\ln(f_i/f_i^\circ)$...] is arbitrarily set equal to some fixed value, usually 1 bar.

This is a bit misleading. The numerical value of f_i is not relative to anything, but $\mu - \mu^\circ$ is. Fugacity is calculated from measured densities or molar volumes, as on page 203, and the fact that the lower limit of integration is some very low pressure does not change the fact that fugacity is a system property, while activity is not. Fugacity is independent of whatever one chooses as f°, but the choice of f° of course governs the corresponding value of $\mu - \mu^\circ$.

The expression $RT\ln f_i$ gives the difference between μ_i at T, P, and μ_i of ideal gas i at T, 1 bar, just as $RT\ln(m_i\gamma_i)$ gives the difference between μ_i at T, P, and μ_i in an ideal one molal solution at T, P. f_i is no more a relative value than is $m_i\gamma_i$. So fugacities do not have standard states any more than corrected concentrations have standard states.

In other words, $a_i = 0.01$ is meaningless unless the standard state is known, but $f_i = 0.01$ bars is unambiguous.

8.1.5 Summary

The fugacity is a property of systems and of system constituents that was invented in order to facilitate the evaluation of $\int dG$ for gaseous compounds. However, the fact that its only practical use is for gases or supercritical fluids does not change the fact that in principle it is a property of *all* system components or species. Each species fugacity is therefore a system property or state variable, whether measurable or not.

G.N. Lewis, who invented the fugacity, suggested that it be called an "escaping tendency," but in most solid and liquid systems it is more intuitively thought of as a (corrected) vapor pressure or partial pressure, but remember that it is at the same time a thermodynamic model parameter. Thus it still has significance at values of say 10^{-60} bars, whereas a pressure of this value would have no physical significance. Its dimensions are the dimensions of pressure, and it is most often measured in bars or pascals. In spite of its appearing as a ratio in one of its equations of definition (8.5), it is nevertheless an absolute quantity for a given species in a given system, as is **G**, thanks to the other equation of definition, (8.6). In other words, fugacities do not have standard states in the way that activities do. They are absolute system properties.

Fugacities in gas mixtures are important in understanding metamorphic mineral equilibria, and the composition of fluids evolving from crystallizing magmas and boiling hydrothermal fluids. The difficulties are formidable, both experimental and theoretical. That is, determining densities at high temperatures and pressures in fluid mixtures is difficult, and the treatment in terms of equations of state is always a problem. Far more detail is available in Prausnitz et al. (1999).

Another thing to note is that in Equation (8.5) we see the beginnings of a way to attack our primary goal, a simple way of relating system composition to the Gibbs energy. A ratio of fugacities is related in a simple way to a difference in Gibbs energies. These fugacities could, at least in principle, refer to species in a solution. Perhaps we can generalize this ratio to a quantity which will refer to a difference in Gibbs energies in all kinds of situations. This is the role of the *activity*.

8.2 Activity

8.2.1 Introduction

In Chapter 7 (§7.6) we introduced the *activity*, and we said that in the form $RT \ln a_i$, it gives the quantity $\mu_i - \mu_i^\circ$. If you think about this, you will realize that there are not many more important concepts related to using thermodynamics in chemical systems. The goal of finding the minimum value of our thermodynamic potential G (or μ) in chemical systems is made complicated by the variety and complexity of our systems, and the fact that we use a variety of standard states in calculating our difference in Gibbs energy. In a sense, all these complexities are transferred to a single quantity, the activity, a dimensionless number which is directly related to $\mu_i - \mu_i^\circ$ for any component i in any system under any conditions. How to calculate a_i for various kinds of components (pure phases, associated and dissociated solutes, etc.) in various kinds of systems (multiphase solid, liquid and gaseous solutions) system is therefore an important topic.

Let us first summarize our development of the concept of the fugacity, f. Starting with the definition

$$dG = RTd \ln f \tag{8.4}$$

and

$$\lim_{P \to 0} (f/P) = 1 \tag{8.6}$$

we found

$$G_{P_2} - G_{P_1} = RT \ln \frac{f_{P_2}}{f_{P_1}} \tag{8.5}$$

which expresses the relationship between the Gibbs energy and fugacity of a gas at two different pressures at the same T. However, changing the pressure on a pure phase is not the only way of changing the fugacity. Because fugacity approximates partial pressure, we might, for instance, simply introduce other components at the same P and T, which will also change the fugacity. Dealing with a solution rather than a pure phase, though, means we should use μ rather than G. So generalizing from a single gas to a gas i in a mixture of gases, and from two states at different pressures to any two states $'$ and $''$ at the same temperature, this becomes

$$\mu_i'' - \mu_i' = RT \ln \frac{f_i''}{f_i'} \tag{8.16}$$

One implication of this is that the fugacity of i is the same in any two states or phases that are in mutual equilibrium, because if $\mu' = \mu''$ then $f' = f''$. This of course was Lewis's intention in defining the fugacity in the first place. Because in principle any substance or species has a fugacity, Equation (8.16) seems to offer a general method for determining Gibbs energy differences. The problem with that is that the fugacities ($\approx$ vapor pressures) of substances other than gases are far too small to measure, and are mostly unknown. However, even in systems where species fugacities are unknown, the *ratio* of a species fugacity to its fugacity in some other state is quite often a measurable and useful quantity, and comparing (8.16) with Equation (7.37), we see that this ratio is in fact a way of expressing the *activity*.

8.2.2 Definition I: gases

Rewriting (8.16) so that state $''$ is any (unsuperscripted) state and state $'$ is a standard state designated by superscript $^\circ$, we have

$$\mu_i - \mu_i^\circ = RT \ln \frac{f_i}{f_i^\circ} \tag{8.17}$$

This is a simple generalization of (8.16), and hence a direct result of the definition of fugacity. We now define the activity of species i as

$$a_i = \frac{f_i}{f_i^\circ} \tag{8.18}$$

where f_i and f_i° are the fugacities of i in the particular solution or state of interest to us and in some reference state at the same temperature. Thus

$$\mu_i - \mu_i^\circ = RT \ln a_i \tag{8.19}$$

which is of course Equation (7.37), arrived at in a different way. We begin now to see why using the activity can be confusing. In Chapter 7 (Equation 7.34) the state that μ_i° refers to is i as a pure liquid or solid, and in this case (8.16) μ_i° refers to i existing as a gas or perhaps fluid in some as yet undefined state, which might be, and is in fact, completely arbitrary. It will be interesting to see how it is that we can use μ_i in a multicomponent, multiphase system, where at equilibrium μ_i must be the same in every phase, while limited by the fact that we can only know μ_i as the difference between it in whatever state it is and some other, arbitrary state which will be different for each kind of phase. We will try to do this in the remainder of this chapter and the next chapter, where activities become part of the equilibrium constant.

8.2.3 Definition II: solutes

We use the same method we used in §7.5.3. We need an expression for the derivative of μ with a concentration term, which we can integrate. The derivative of μ_i with respect to the molality of i, m_i, is $(\partial\mu_i/\partial m_i)_{T,P,\hat{m}_i}$, where $\hat{m}_i$ means the molality of all solution components *except* i.

If we expand $(\partial\mu_i/\partial m_i)_{T,P,\hat{m}_i}$ by introducing P_i, the pressure on gaseous i which is, or might be, in equilibrium with solute i (whether or not there is such a gas phase is irrelevant), we get

$$\left(\frac{\partial\mu_i}{\partial m_i}\right)_{\hat{m}_i} = \frac{\partial\mu_i}{\partial P_i}\frac{\partial P_i}{\partial m_i} \tag{8.20}$$

where μ_i is the same in the solution and in the vapor phase, where it can be called G_i (the vapor being assumed an ideal gas), so that $(\partial\mu_i/\partial P_i) = (\partial G_i/\partial P_i) = V_i = RT/P_i$, and where $(\partial P_i/\partial m_i) = P_i/m_i$ is an expression of Henry's law (§7.4.2), as mentioned earlier. Combining all this we get

$$\left(\frac{\partial\mu_i}{\partial m_i}\right)_{\hat{m}_i} = \frac{RT}{m_i} \tag{8.21}$$

for ideal (Henryan) solutions. Integrating this equation between two values of molality, m_i' and m_i'', we get

$$\mu_i'' - \mu_i' = RT\ln\frac{m_i''}{m_i'} \tag{8.22}$$

showing the effect of changing solute concentration on the chemical potential, as we wanted. However, it is limited to ideal (Henryan) solutions. The relationship is generalized to any kind of solution by introducing a correction factor at each concentration. Thus

$$\mu_i'' - \mu_i' = RT\ln\frac{\gamma_{\mathrm{H}}'' m_i''}{\gamma_{\mathrm{H}}' m_i'} \tag{8.23}$$

where $\gamma_{\rm H}$ is the Henryan activity coefficient, and Equation (8.23) now refers to any real solution at a given temperature in which species i changes concentration, all other species remaining unchanged.[1]

Equation (8.23) can be generalized and so made more useful by choosing a single concentration m'_i for all solutes. In choosing this concentration, we should realize that

1. $\gamma_{\rm H}$ in the denominator will be different for all different solutes unless we choose some idealized state, and
2. it would be convenient to have the denominator $(\gamma'_{\rm H} m')$ disappear, i.e., be unity.

The only state which satisfies these conditions and is equal to 1 molal for all solutes is the ideal (Henryan) one molal solution, and this is universally used as the standard state for solutes. Introducing superscript ° for the standard state, and dropping the now unnecessary superscript ″, we get

$$\mu_i - \mu_i^\circ = RT \ln \frac{\gamma_{\rm H_i}\, m_i}{\gamma_{\rm H_i}^\circ m_i^\circ} \tag{8.24}$$

and because $\gamma_{\rm H_i}^\circ = 1$ and $m_i^\circ = 1$, this is usually written

$$\mu_i - \mu_i^\circ = RT \ln (\gamma_{\rm H_i}\, m_i) \tag{8.25}$$

The quantity $(\gamma_{\rm H_i}\, m_i)/(\gamma_{\rm H_i}^\circ m_i^\circ)$ is another definition of the *activity*, a_i, so

$$\mu_i - \mu_i^\circ = RT \ln a_i \tag{8.26}$$

The activity thus allows calculation of the difference between the μ_i in a solution and μ_i in the ideal one molal standard state at the same T and P as the solution. This sounds like a fairly esoteric thing to do, but because standard Gibbs energies of formation are determined for this ideal standard state (albeit at 25 °C, 1 bar), it is immensely useful, as we will see.

At this point you should note that we have not used the infinite dilution standard state for aqueous solutes, as we will for other properties in Chapter 10. Having $m_i^\circ \to 0$ in Equation (8.24) would obviously be inconvenient.

8.2.4 Definition III: solids and liquids

Now that we know about the fugacity, we can derive Equation (7.27) in still another way, because for an ideal gas f_i/f_i° is equal to P_i/P_i°, which is equal to the mole fraction, x_i (Equation 7.8). So, for ideal gaseous and liquid solutions, and by extension, for any ideal (Raoultian) solution,

$$\mu_i - \mu_i^\circ = RT \ln x_i \tag{8.27}$$

---

[1] Note that you can do this, i.e., change the concentration of i without changing any other concentrations, because molality is moles per kg of *solvent*. You cannot do it using mole fraction or molarity.

For solutions covering a wide range of compositions, such as many solid and liquid solutions, this equation can be used by introducing another correction factor, the Raoultian activity coefficient, γ_{R}. Thus

$$\mu_i - \mu_i^\circ = RT \ln (x_i \gamma_{\mathrm{R}_i}) \tag{8.28}$$

As before, we now define another activity term

$$a_i = x_i \gamma_{\mathrm{R}_i} \tag{8.29}$$

which is useful for solutions covering a wide range of concentrations, and for which γ_{R} is known or can be estimated. In geochemistry, this tends to be for solid and gaseous solutions only, but it is widely used in metallurgy for liquids as well. The standard state, as before, is that state for which $a = 1$, in this case the pure liquid or solid ($x = 1$; $\gamma_{\mathrm{R}} = 1$ in this state by definition).

In §8.3.4 we will see that we can also use γ_{H} instead of γ_{R} with the mole fraction in Equation (8.28). This rarely happens in geochemistry, but is instructive nonetheless.

8.2.5 Summary

Here are our various definitions of activity:

$$\left.\begin{aligned} \mu_i - \mu_i^\circ &= RT \ln a_i \\ \mu_i - \mu_i^\circ &= RT \ln \frac{f_i}{f_i^\circ} \\ \mu_i - \mu_i^\circ &= RT \ln \frac{\gamma_{\mathrm{H}_i} m_i}{\gamma_{\mathrm{H}_i}^\circ m_i^\circ} \\ \mu_i - \mu_i^\circ &= RT \ln x_i \gamma_{\mathrm{R}_i} \end{aligned}\right\} \tag{8.30}$$

In any equilibrium state, both μ_i and μ_i° are absolute, finite quantities with a fixed difference between them. If the same standard state is chosen for each of these equations, then $\mu_i - \mu_i^\circ$ is the same in each equation, and the activity would be the same in all phases at equilibrium. This would be nice, but it would mean using a vapor pressure as the standard state for activity in solids, or an ideal one molal solution standard state for activities in a gas, or perhaps an ideal gas at one bar for an aqueous solute. This would be not only inconvenient, but impossible in many cases. So we accept the small inconvenience of having different activities for the same species in different phases.

In a multicomponent, multiphase system at equilibrium, μ_i is the same in every phase, but in most cases μ_i° and therefore $\mu_i - \mu_i^\circ$ is different for solids, liquids, gases, and solutes (we know this without knowing the numerical value of either term). Thermodynamic properties are determined and tabulated for substances in these various standard states, and how they relate to one another in chemical reactions can be seen when we consider the equilibrium constant (Chapter 9).

Finally, note that fugacities have units of pressure (e.g., bars), but that activities and activity coefficients are always dimensionless.

8.3 Standard states and activity coefficients

8.3.1 Introduction

Standard states are simply a special sort of reference state for physical properties, made necessary, as we have mentioned several times, by our lack of knowledge of absolute values for the properties U, μ, H, and A. We have to express values of these quantities in some real state as differences from some other state, the standard state.

If you mention the subject of standard states to most geochemists, they will normally assume that you are referring to the activity–chemical potential relationship, because this is by far the most important usage of standard states. But there is another important usage of standard states, and that is for enthalpy. The Helmholtz energy uses the same standard states as the Gibbs energy, and using U quantitatively is rare, so standard states for μ and H are the only ones we really need to worry about. But of course once these states are defined they have other properties as well, so we have standard state volumes, heat capacities and entropies, in addition to Gibbs energies and enthalpies. The difference with these properties is that we know their values in all states, and probably would not define standard states for them if it were not for our Gibbs energy–enthalpy problem.

In this chapter, we discuss the various standard states used for the Gibbs energy and the activity. The standard state used for enthalpy, volume and heat capacity is quite different, and is discussed, along with a more detailed look at partial molar properties, in Chapter 10.

Standard states are states of matter in specified conditions. The definition must be sufficiently complete as to determine the thermodynamic parameters of the substance, and therefore must have at least four attributes:

1. temperature
2. pressure
3. composition
4. state of aggregation (solid polymorph, liquid, gas, ideal gas, ideal solution, etc.).

Thus “25 °C, 1 bar” is *not* a standard state. The question is, what *system* at 25 °C, 1 bar?

Because the goal of the definition is to specify the thermodynamic parameters of the substance, it frequently happens that the standard state chosen is a hypothetical, perhaps physically unrealizable state. The importance of these states lies in our knowing their properties, not in being able to actually achieve them. Certain standard states are so commonly used that one need not always elaborate on the definition, i.e., it may be obvious from the context. If there is any deviation from usual practice, it would be necessary to be quite specific about the nature of the standard state.

In the following discussion of standard states we must distinguish between the properties of the standard state (T°, P°, m°, etc.) and those of the state of interest (i.e., the equilibrium state in which the activity of some component or species is of interest to us; T, P, m, etc.) because the two states are often completely different. We will also refer to the state of interest as “the system.”

8.3.2 Definitions

The general relationship between activity and Gibbs energy is Equation (8.19),

$$\mu_i - \mu_i^\circ = RT \ln a_i \qquad [8.19]$$

Thus a_i is a measure of the difference between the chemical potential of i in some equilibrium state, and the chemical potential of i in its standard state. When i is in its standard state, $a_i = 1$ and this difference is zero.[2] As we have seen, the activity can take on several different forms, depending on whether we are using fugacities, molalities, or mole fractions. Therefore the definition of the standard state will be different in each case, because $a_i = 1$ will imply a different state for i in each case.

We will now have a look at the standard states for each of these cases. These states are very different from each other, and some are very hypothetical and seemingly very unrealistic. We will try to show how this arises, and that the standard states we use are actually quite reasonable. To do this with real numbers rather than just symbols, we need either experimental data or some equations that simulate or fit experimental data in a realistic way. Real data unfortunately have uncertainties, so we will borrow the concept of regular solutions from Chapter 10. For our purposes here, all we need to know is that in such solutions, activity coefficients follow the relationships

$$\left.\begin{aligned} RT \ln \gamma_{\mathrm{R_A}} &= w_G x_{\mathrm{B}}^2 \\ RT \ln \gamma_{\mathrm{R_B}} &= w_G x_{\mathrm{A}}^2 \end{aligned}\right\} \qquad (8.31)$$

where w_G is a constant. These equations are discussed in §10.6.1, page 306. We choose an arbitrary but convenient value for w_G of 2000 J mol^{-1}. This gives us some activity coefficient numbers to play with and to illustrate several things about these coefficients and about standard states.

---

[2] In the standard state, a_i is 1.0, but the reverse is not always true. Activity coefficients and concentrations can sometimes combine to give a value of a_i equal to 1.0 in some real state which is not the standard state.

8.3.3 $RT \ln(f_i/f_i^\circ)$

The first thing to note is that in all three cases, the state of interest and the standard state must be at the same temperature. You can see this by looking at the derivation, in which RT was treated as a constant during integration. Thus the standard state temperature must be defined as the temperature of whatever state we are comparing with it. Because we are often interested in a series of equilibrium states at different temperatures, we therefore have a corresponding series of standard states, one for each temperature. This can be regarded rather as a single standard state having a variable temperature.

Fixed pressure

We can also see from Equation (8.17) why it often proves convenient to choose a standard state for i, which is not only not its most stable state, but one that is extremely hypothetical. If f_i° is set to 1.0, Equation (8.17) becomes

$$\mu_i - \mu_i^\circ = RT \ln f_i \tag{8.32}$$

which is a very convenient form of the equation, provided that fugacities of i are available or measurable. The physical significance of setting the denominator to 1.0 is that constituent i is said to be in a state in which the fugacity is 1.0 at all temperatures. The only substance for which this is true is an ideal gas at $P = 1$, so Equation (8.32) implies the choice of "ideal gaseous i at one bar and temperature T" as the standard state. This perhaps seems reasonable enough for a gas, but it can be used for any substance including solids and liquids. The only reason it is not universally used is that fugacities (f_i) are not known for many constituents of interest, especially solids[3] and dilute solutes.

To put this in other terms, if you have the fugacity of some substance i in some system, then $RT \ln f_i$ is the difference in Gibbs energy per mole of i in the system at T and i as an ideal gas at T and 1 bar. Whether i could ever come close to existing as an ideal gas is irrelevant. Other examples of hypothetical standard states are discussed below.

Note that if you do set $f^\circ = 1$, then μ_i° and f_i° are independent of the system pressure. That is to say, they depend on P° but not on P. Once the standard state is chosen, it is a function only of the value of T, held constant during the integration. Another way of putting this is that we have a *fixed pressure* standard state.

Finally, comparing Equations (8.32) and (8.19) we see that it is quite possible for the activity of i to be numerically identical to the fugacity of i. It just requires that the standard state chosen for i is ideal gas i at T and 1 bar. We will find (Chapter 9) that this is a very common situation when i is a gas or a gaseous component, and in fact it is the standard state used in program SUPCRT92, as shown in Table 8.1. In this table, note that at each of the three temperatures

---

[3] Interest in the condensation of solids from the solar nebula in the early history of the solar system has made the fugacity of solids a more relevant topic.

Table 8.1 *Part of the output from* SUPCRT92 *for pure* CO_2 *gas. Some rows and columns have been removed.*

```
                                    DELTA G      DELTA H      DELTA S
PRES(bars)  TEMP(degC)    LOG K     (cal)        (cal)        (cal/K)
---------   ---------   -------   ---------    ---------    ---------
  500.000        .000    74.398     -92986.      -94267.         50.3
  500.000     100.000    57.494     -98166.      -93344.         53.2
  500.000     200.000    47.857    -103611.      -92315.         55.6
---------------------------------------------------------------------
 1000.000        .000    74.398     -92986.      -94267.         50.3
 1000.000     100.000    57.494     -98166.      -93344.         53.2
 1000.000     200.000    47.857    -103611.      -92315.         55.6
---------------------------------------------------------------------
 5000.000        .000    74.398     -92986.      -94267.         50.3
 5000.000     100.000    57.494     -98166.      -93344.         53.2
 5000.000     200.000    47.857    -103611.      -92315.         55.6
```

0, 100 and 200 °C, the thermodynamic parameters $\Delta_f G^\circ$, $\Delta_f H^\circ$, and S° are identical at all pressures, and the same holds true at other temperatures.[4] This is because SUPCRT92 supplies standard state properties, and the standard state for gases is the ideal gas at T and 1 bar. The properties shown refer to this state, not to real CO_2 at T and P.

Variable pressure

But it is not necessary to set $f^\circ = 1$ bar, just convenient in many cases. Another option, fairly common in geochemistry though not in chemistry, is to let $f^\circ = f^{\text{pure}}$, and to define the standard state pressure as the system pressure. In this case, we compare the fugacity of i with the fugacity of pure i at the same T and P. If the Lewis fugacity rule (§8.1.3) holds, this is the same as using the mole fraction of i, but normally this will be only approximately true.

Again, because we often consider various system pressures, the standard state pressure will vary, so we have a *variable pressure* standard state.

8.3.4 $RT \ln x_i$

This formulation is used for liquid and solid solutions, where relatively large ranges of compositional change occur. In this case, we have less choice as to the nature of the standard state – if $a_i = 1$, then $x_i = 1$.[5] In other words, the

---

[4] Careful readers will note that the table heading for entropy is `DELTA S`, whereas what is reported is S°, the standard third law entropy for CO_2 gas. This is because the tables in SUPCRT92 are set up to report data for complete balanced reactions, although you can, as in this case, get data for single compounds.

[5] As before, there may be some conditions where x_i and γ_{R_i} just happen to combine to give $a_i = 1.0$. This is not another standard state.

standard state must be pure *i*, and it must be at the same *T* and *P* as the system, because only *n* was allowed to vary during integration.

But we do have a choice as to what kind of activity coefficient we want to use, Raoultian or Henryan. You might think that if the standard state consists of pure *i*, normally a pure solid or liquid, it might be difficult to use a dilute solution (Henryan) standard state. However, using hypothetical states makes it quite simple, and quite instructive. But first we consider the Raoultian standard state.

Raoultian γ

If we choose $x_i = 1$, $\gamma_{R_i} = 1$, then the standard state is pure *i* at the system *T* and *P*. This is possibly the simplest of all standard states. The activity of a solid or liquid solution is simply equal to its mole fraction, if activity coefficients are ignored (a common occurrence), or its mole fraction multiplied by some correction factor γ_{R_i}. And yes, this standard state is the actual, existing, pure phase, unless you want to get very picky and hold that there are no "really" pure phases, so that the pure phase standard state is just as hypothetical as all the others. This is a useless distinction we can leave to purists.

To illustrate the Raoultian activity–activity coefficient relationship we use activity coefficients defined by Equations (8.31), shown in Figure 8.2. In real systems these are measured quantities with associated uncertainties, and the shape of the activity curve may not fit any simple function. Figure 8.2 shows

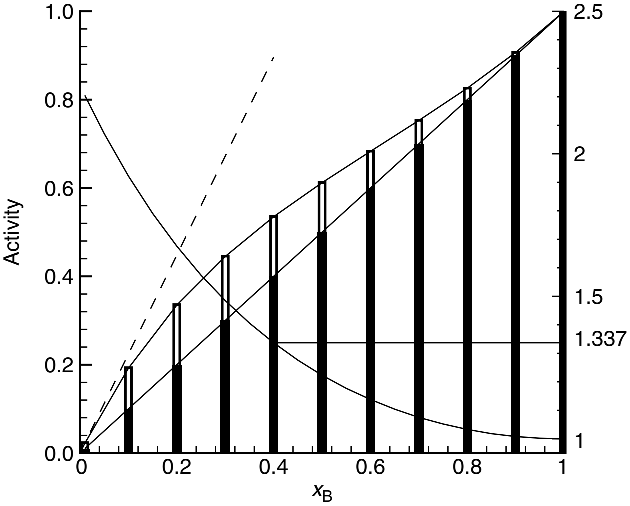


Figure 8.2 Ideal or Raoult's law activity (= x_B) at intervals is shown by the solid bars, and the activity of B by the hollow bars, The ratio of these two quantities (a_B/x_B) is the activity coefficient. The value of $\gamma_{R_B} = 1.337$ at $x_B = 0.4$ is shown by the arrow. The dashed line is tangent to the activity of B at $x_B = 0$, and indicates Henry's law.

the activity of component B, and the ratio of the hollow and black bars shows that the activity coefficient is a measure of how far the activity is from the ideal at any given mole fraction.[6] This value of this ratio, the activity coefficient, is shown on the right-hand axis, and the value of 1.337 from §10.6.2 is indicated. Note that the activity coefficient $\gamma_{\mathrm{R_B}} \to 1$ as $x_{\mathrm{B}} \to 1$, consistent with $a_{\mathrm{B}} = 1$ for pure B. The dashed line is the tangent to the activity curve at $x_{\mathrm{B}} = 0$ (the Henry's law slope), and is discussed further below.

Henryan γ

However, if we choose $x_i = 1$, $\gamma_{\mathrm{H}_i} = 1$, the standard state becomes a pure phase ($x_i = 1$) which forms ideal Henryan solutions with another component. But the only pure phase which does this is a hypothetical one, defined by the intersection of the Henry's law slope at the other component composition with the $x_i = 1$ axis. All compositions along this tangent line obey Henry's law. Because our example is using regular solution theory for simplicity, we know the equation for the activity curve, so we can differentiate it and find the value of this at $x_{\mathrm{B}} = 0$. With $w_G = 2000\,\mathrm{J\,mol^{-1}}$ this slope (K_{H}, discussed in §7.4.2, but expressed in terms of fugacity rather than pressure) turns out to be 2.2407, giving an intercept at $x_{\mathrm{B}} = 1$ of 2.2407 as shown in Figure 8.3.

Because we know the equations for both the Raoult's law line and the Henry's law line, we can express any measured fugacity in terms of its deviation

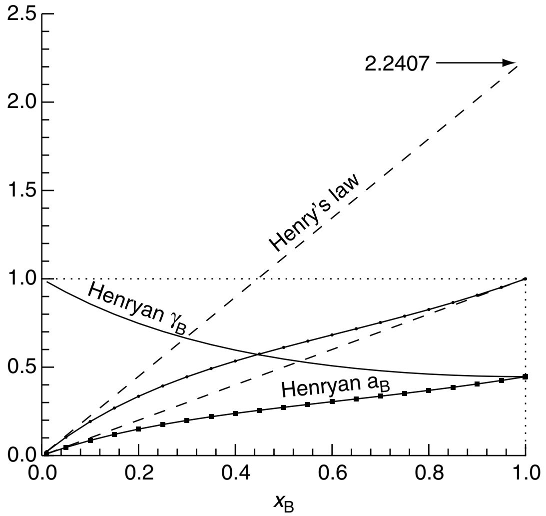


Figure 8.3 Henry's law tangent (Figure 8.2) extended to intersect the $x_{\mathrm{B}} = 1$ axis. Raoultian activities from Figure 8.2 are also shown. Henryan activities shown are $\gamma_{\mathrm{H}} \cdot x_{\mathrm{B}}$.

---

[6] The same data are shown in Figure 10.15. There the emphasis is on regular solution theory. Here it is on activity coefficients and standard states.

from both lines, and hence find the relationship between the two kinds of activity coefficients, γ_{R} and γ_{H}. The two equations are

$$f_{\mathrm{B}} = x_{\mathrm{B}}\gamma_{\mathrm{R_B}}, \quad \text{and} \quad f_{\mathrm{B}} = x_{\mathrm{B}}\,K_{\mathrm{H}}\,\gamma_{\mathrm{H_B}}$$

from which

$$\gamma_{\mathrm{H_B}} = \gamma_{\mathrm{R_B}}/K_{\mathrm{H}} \tag{8.33}$$

In Figure 8.4, where the activities and activity coefficients of B are the previously defined Raoultian values, all divided by 2.2407. The data are shown in Table 8.2.

A final point. The standard state could equally well be chosen anywhere on the Henry's law slope, all points on which have the properties of the infinitely dilute solution. However if, say, $x_{\mathrm{B}} = 0.5$ were chosen, the activity coefficient in the standard state would be 2 instead of 1. This is not a reasonable thing to do, but the idea of sliding the standard state along the Henry's law slope while at the same time changing the concentration axis from mole fraction to something else certainly is reasonable. Metallurgists sometimes use a weight percent axis, and define the standard state at 1% on the Henryan slope (see e.g., Lupis, 1983, Chapter 7). In chemistry and geochemistry where aqueous solutions are used, a molality axis is best, and 1 molal on the Henryan slope becomes the standard state, as discussed in the next section.

Fixed versus variable pressure

An ideal binary solid solution A–B at temperature T and pressure of 1 bar having $x_{\mathrm{B}} = 0.5$ has an activity of B $a_{\mathrm{B}} = 0.5$ on a Raoultian basis. What is

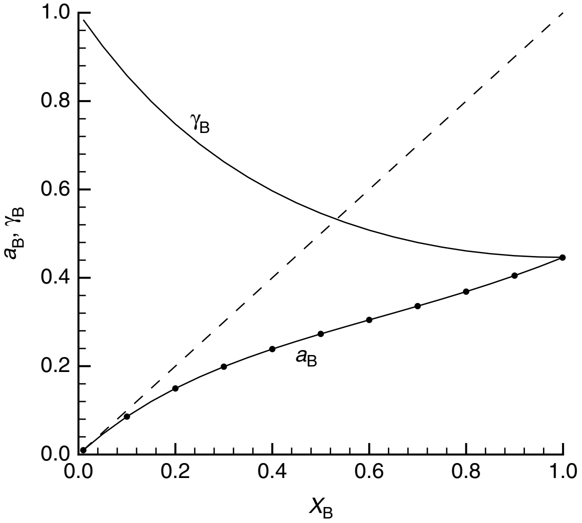


Figure 8.4 Henryan activity and activity coefficient data from Figure 8.3, with the Henryan slope normalized to 1.0.

Table 8.2 *Raoultian and Henryan activities and activity coefficients for a regular solution having* $w_G = 2000\,\mathrm{J\,mol^{-1}}$. *Raoultian data are from Table 10.2.*

	Raoultian		Henryan	
x_B	a_B	γ_{R_B}	a_B	γ_{H_B}
0.1	0.192	1.922	0.086	0.858
0.2	0.335	1.676	0.150	0.748
0.3	0.446	1.485	0.199	0.663
0.4	0.535	1.337	0.239	0.597
0.5	0.612	1.224	0.273	0.546
0.6	0.683	1.138	0.305	0.508
0.7	0.753	1.075	0.336	0.480
0.8	0.826	1.033	0.369	0.461
0.9	0.907	1.008	0.405	0.450
1.0	1.000	1.000	0.446	0.446

the activity of B if the solid solution is at a pressure of 1000 bars? This is of course dependent on how we define the standard state. It is done in two ways:

1. Most chemistry texts define all standard states as having a fixed pressure of 1 bar. Therefore a component of an ideal solid solution will have a different activity at 1000 bars than it has at 1 bar.
2. In geochemistry it is more common to define the standard state as having a standard state pressure equal to the pressure of interest, $P^\circ = P$, so that a component of an ideal solid solution will have the same activity at 1000 bars that it has at 1 bar.

The details of these calculations will be discussed below (§8.4).

8.3.5 $RT \ln[(\gamma_{H_i} m_i)/(\gamma^\circ_{H_i} m^\circ_i)]$

There are some very good reasons why the thermodynamics of aqueous solutions looks quite different from the cases already considered.

- We can and do use fugacities for the solvent, water, because its vapor pressure is a measurable quantity, but in most cases we cannot use fugacities for the solute because most solutes are nonvolatile – they have no measurable vapor pressure.
- For water vapor, a gas, the $f^\circ = 1$ bar standard state is used, but for liquid water, f° is the fugacity of pure water (often approximately equal to the vapor pressure of pure water). This results in $f = f^\circ$ and $a = 1$ for pure water, the Raoultian standard state.
- Solute concentrations measured in mole fractions tend to be very small, which is inconvenient. So we use molalities which, being independent of density and hence of T and P are by far the most convenient measure of concentration. But this means

that we cannot use the pure solute as a standard state, because m becomes infinite for the pure solute. We must use different standard states for the solvent and solute, a situation often called "unsymmetrical," but which in itself causes no problems.

- We can use a Raoultian standard state (pure water) for the solvent, but its deviation from ideal behavior, whether based on a mole fraction or a molality scale, is often converted to the osmotic coefficient ϕ, which does not actually have a standard state. It is an absolute system property.

We look first at the solute and then at the solvent.

8.3.6 Activity coefficients of solutes

Activities of the solute, whether electrolyte or nonelectrolyte, use molalities and the $(\gamma^{\circ}_{\mathrm{H}_i} m^{\circ}_i)$ standard state. We must now examine what this means.

We want to use the Henryan standard state, which is a state in which the solute exhibits dilute solution behavior. That is, no matter what the actual concentration, the solute behaves as if there is absolutely no solute–solute interactions – each solute molecule thinks it is alone in the solvent. It is a state which obey's Henry's law, which at real concentrations is obviously a hypothetical state, and it lies anywhere on the Henry's law tangent. In §8.3.4 we saw that we could choose a point on this tangent having $x_i = 1$, or we could choose any other point. What the other points on this slope mean depends on what concentration scale we are using – if we use a weight percent scale we can choose a weight percent standard state, and if we use a molality scale we can choose a molal standard state.

Although theoretically we could choose any value for m°, any choice except $m^{\circ} = 1$ would introduce complications, and of course we want $\gamma^{\circ}_{\mathrm{H}} = 1$ so that the standard state lies on the tangent and refers to properties at infinite dilution. This leads to the adoption of the "hypothetical ideal one molal" standard state for aqueous solutes. If in Figure 8.4 we change the concentration scale to molality, and focus on the lower left corner of the diagram, we have Figure 8.5. We have assumed that B is a nonelectrolyte with $\nu = 1$ such as sucrose or oxygen, and the conversion is

$$m_{\mathrm{B}} = \frac{55.51 \cdot x_{\mathrm{B}}}{1 - x_{\mathrm{B}}}$$

In this case, in which we have continued our example of a regular solution having $w_{\mathrm{G}} = 2000\,\mathrm{J\,mol^{-1}}$, the solution does not deviate from the ideal very greatly until molalities well above $1\,m$. Activity coefficients based on mole fractions and molalities are shown in Table 8.3. The reason for two slightly different values of γ_{H}, $\gamma_{\mathrm{H}x}$ and $\gamma_{\mathrm{H}m}$, is that molality is not exactly proportional to mole fraction except in the limit of infinite dilution, so that a Henryan tangent

Table 8.3 *Activity coefficients of component B on various scales for a regular solution having* $w_G = 2000\,J\,mol^{-1}$.

Mole fraction x_B	Molality m_B	γ_R Figure 8.2	γ_{Hx} Figure 8.4	γ_{Hm} Figure 8.5
0	0.000	2.241	1.000	1.000
0.001	0.056	2.237	0.998	0.997
0.002	0.111	2.233	0.997	0.995
0.003	0.167	2.230	0.995	0.992
0.004	0.223	2.226	0.994	0.990
0.005	0.279	2.223	0.992	0.987
0.006	0.335	2.219	0.990	0.984
0.007	0.391	2.216	0.989	0.982
0.008	0.448	2.212	0.987	0.979
0.009	0.504	2.209	0.986	0.977
0.010	0.561	2.205	0.984	0.974
0.050	2.921	2.071	0.924	0.878
0.100	6.168	1.922	0.858	0.772
0.150	9.796	1.791	0.799	0.679
0.200	13.877	1.676	0.748	0.598

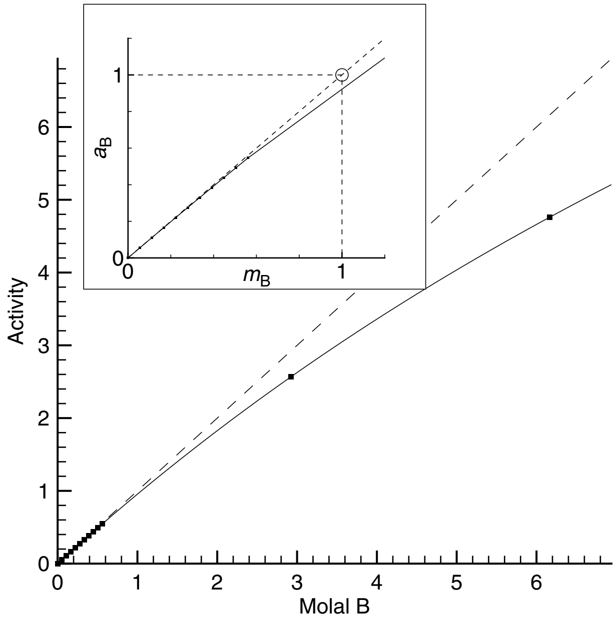


Figure 8.5 Activity of B from Figure 8.4 expressed on a molality scale. The inset is an enlargement of the region up to 1 molal, and the circle shows the position of the ideal 1 molal standard state.

on a mole fraction diagram will appear to be slightly curved on a molality diagram. The correction factor for activity coefficients is

$$\ln(\gamma_{\mathrm{H}x}) = \ln(\gamma_{\mathrm{H}m}) + \ln(1 + 0.018\,0154\,m)$$

and, as you can see, it is insignificant until quite large molalities are reached.

In this section the magnitude of our activity coefficients has been fixed by our choice of a regular solution model. We did this so as to produce realistic numbers, and to concentrate on the meanings of activity coefficients and standard states. In most situations, however, activity coefficients are either measured in a laboratory situation, or they are calculated on a theoretical basis to serve as correction factors to some known concentrations in modeling a real solution or equilibrium. They will not always fit a regular solution model, although the results of Aranovich and Newton are interesting in this regard, as mentioned above. The calculation of activity coefficients from electrostatic considerations will be discussed in Chapter 15.

8.3.7 Osmotic coefficients

In aqueous solutions the activity of the pure solvent, measured as f/f°, is 1.0. In dilute solutions, the focus of a great deal of solution chemistry, the activity remains very close to 1.0, so its accurate expression requires several significant figures. The example given by Robinson and Stokes (1959), a classic reference, is a 2 molal KCl solution, which has a water mole fraction of 0.9328 and a water activity of 0.9364. Its activity coefficient is therefore 1.004, a figure "which fails to emphasize the departure from ideality indicated by the activity coefficient of the solute," which is 0.614. To overcome this problem use of the osmotic coefficient has become standard practice, even though its slope changes sign at concentrations above about 1 molal, as shown in Figure 8.6a. The same pattern, this time for NaCl solutions, becomes more subdued at higher temperatures and pressures (Figure 8.6b). A less complicated pattern is shown by many nonelectrolytes such as sucrose, shown in Figure 8.7.

Osmotic pressure

The osmotic coefficient has its origin, as you might suspect, through its connection with osmotic pressures. The chemical potential of water in an aqueous solution is inherently less than that of the pure solvent, as you can see from $\mu_1 - \mu_1^\circ = RT \ln x_1$. If this solution is separated from the pure solvent by a membrane permeable only to the solvent, pure solvent will pass through the membrane into the solution in an attempt to remove this difference. If the

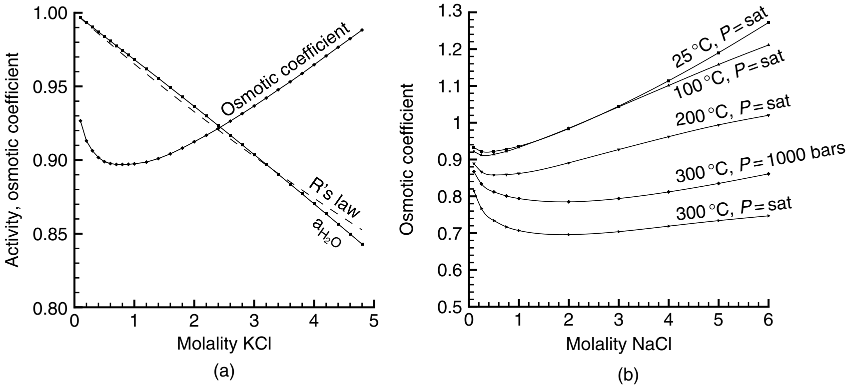


Figure 8.6 (a) The osmotic coefficient and activity of KCl in aqueous KCl solutions at 25 °C, 1 bar. Data from Robinson and Stokes (1959).
(b) Osmotic coefficient for aqueous NaCl solutions at various temperatures and pressures. Data from Pitzer and Peiper (1984).

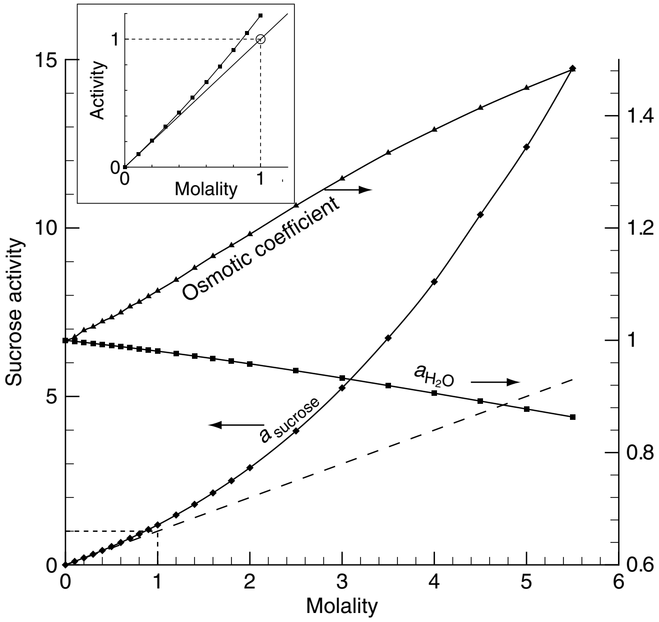


Figure 8.7 Henryan activity of sucrose, Raoultian activity of water, and osmotic coefficients in sucrose solutions at 25 °C. The inset is an enlargement of the region up to 1 molal, and the circle shows the ideal one molal standard state.

solution is arranged such that this raises the level of the solution above that of the pure solvent, the added pressure on the solution will increase the potential μ_1 in the solution, and the process will continue until μ_1 is the same on both sides of the membrane. This difference in pressure on the two sides of the membrane is called the osmotic pressure. An approximate expression for the osmotic pressure is

$$P - P^\circ = -\frac{RT \ln a_1}{V_1^\circ} \tag{8.34}$$

where $P - P^\circ$ is the osmotic pressure and V_1° is the molar volume of the pure solvent. The expression is approximate because the volume term should strictly be the partial molar volume of the solvent rather than its molar volume. The molar osmotic coefficient is defined as

$$\begin{aligned}\ln a_{\mathrm{H_2O}} &= \ln \frac{f_{\mathrm{H_2O}}}{f^\circ_{\mathrm{H_2O}}} \\ &= -\frac{\nu m}{55.51}\phi \end{aligned} \tag{8.35}$$

and, combining Equations (8.34) and (8.35) we get

$$P - P^\circ = \phi\left(\frac{RT\,\nu m}{55.51 V^\circ_1}\right) \tag{8.36}$$

where m is the molality of the solute and ν is the number of moles of ions produced by dissociation of one mole of solute (e.g., $\nu = 2$ for NaCl). This shows that ϕ is a factor correcting the approximate expression to the true osmotic pressure, hence its name. There is another version of the osmotic coefficient used with the mole fraction scale of concentrations.

Equation (8.35) is written for one solute. If there are several, we write

$$\begin{aligned}\phi &= -\frac{55.51}{\nu m}\ln a_{\mathrm{H_2O}}, \quad \text{or for many solutes,} \\ &= -\frac{55.51}{\sum_i m_i}\ln a_{\mathrm{H_2O}} \end{aligned} \tag{8.37}$$

where m_i is the molality of any solute species, and $\sum_i m_i$ is the sum of all species molalities (e.g., $2\,m_{\mathrm{NaCl}}$ for completely dissociated NaCl).

H_2O Activity at High T and P

The fact that the activity of water in even quite concentrated salt solutions is not greatly different from 1.0, as mentioned above, is the reason that $a_{\mathrm{H_2O}} = 1$ is routinely used in geochemical calculations, no matter what the solution composition. Until recently there have been no measurements of water activity in concentrated solutions at high T and P to modify this practice. Aranovitch and Newton (1996, 1997) measured the activity of water in concentrated NaCl, KCl, and NaCl+KCl brines to temperatures of 550–900 °C and pressures up to 15 000 bars, and found that in contrast to its behavior at ambient conditions, $a_{\mathrm{H_2O}}$ can drop to very low values at high T and P. These solutions, which are close to saturation with halite and/or sylvite, can have $a_{\mathrm{H_2O}}$ as low as 0.15 on the Raoultian scale. This is important in petrology, as it means that hydration of anhydrous minerals such as garnets and pyroxenes to biotites and amphiboles as well as partial melting may begin at much higher temperatures than have been found experimentally using pure water. In the present context, the work of Aranovitch and Newton is interesting for other reasons as well. They find that regular solution theory works quite well for these solutions, and they find that the dissociation of NaCl and KCl, which decreases remarkably from 25 to 350 °C at solution vapor pressures, increases again to complete dissociation at 10 to 15 kbar, 550–900 °C. Therefore they do not use $\nu = 2$ in calculating mole fraction (see §7.2), but must use a variable ν.

As an Earth scientist using thermodynamics, you may never actually use the osmotic coefficient. However in reading the chemical literature you will find it mentioned quite a lot, so you have to know what it is. It is another measure of the deviation of the solvent, usually water, from ideal behavior.

8.3.8 Activity coefficients as deviation functions

We noted in §13.2.3 that the fugacity coefficient, in the form $RT \ln \gamma_f$, is a residual function, defined (§13.2.3) as the difference between a "real system" thermodynamic function and the same function for an ideal gas under the same conditions.

The same can be said for all the expressions for $\mu - \mu^\circ$ in Equations (8.30). They all express the difference between the chemical potential of a solute species in a real system, and the same potential in an ideal system under the same conditions. The term "residual function" is strictly speaking applied only when the ideal system is an ideal gas, so differences from other states such as infinitely dilute solutions or pure phases are called "deviation functions" (Ewing and Peters, 2000).

We begin to see just how pervasive the concept of ideality is in thermodynamics.

8.4 Effect of temperature and pressure on activities

The change in activities as a response to changes in temperature and pressure of the system naturally depends to a large extent on the standard states involved. Note too, that any variation of activity with change in T or P is actually due to variation of the activity coefficient, because these effects are normally calculated for constant composition conditions.

8.4.1 Temperature

We have by now defined activity in four different but equivalent ways, Equations (8.30). Differentiating with respect to temperature, we have

$$\frac{\partial}{\partial T}(R \ln a_i) = \frac{\partial}{\partial T}(R \ln \gamma_i) \tag{8.38}$$

$$= \frac{\partial}{\partial T}\left(\frac{\mu_i}{T} - \frac{\mu_i^\circ}{T}\right)$$

$$= \frac{-(\overline{H}_i - \overline{H}_i^\circ)}{T^2} \tag{8.39}$$

$$= \frac{-\overline{L}_i}{T^2} \tag{8.40}$$

where γ_i may be γ_f, γ_R, or γ_H. Thus the temperature derivative of the activity is a simple function of the *relative partial molar enthalpy*, $\overline{L}_i$, which was discussed in §9.3.3. The numerical value of $\overline{L}_i$ of course depends on the form of activity being used, because this determines the value of $\overline{H}_i^\circ$. For standard states based on Raoult's law, that is, where the standard state is the pure substance i at T and P, $\overline{H}_i^\circ$ is the same as H_i°, the molar enthalpy of the substance (an unknown quantity, but as usual it appears in a difference term, so we don't need to know its absolute value).

For the ideal gaseous standard state, $\overline{H}_i^\circ$ is evidently the molar enthalpy of an ideal gas. For standard states based on Henry's law, where $\gamma_H \to 1$ as x or $m \to 0$, $\overline{H}_i^\circ$ is the partial molar enthalpy of the solute in the hypothetical pure substance having $\gamma_H = 1$ or the hypothetical ideal one molal solution respectively. Substances in these strange states have partial molar enthalpies (and volumes) equal to that at infinite dilution, hence providing a method of measurement. This can be seen by considering Equations (8.38) and (8.39), which show that $\overline{H}_i^\circ$ becomes equal to $\overline{H}_i$ when γ_i is 1.0. Therefore for Henryan standard states where $\gamma_i \to 1$ as x_i or $m \to 0$, $\overline{H}_i^\circ$ must be the partial molar enthalpy of i at infinite dilution, and for Raoultian standard states where $\gamma_i \to 1$ as $x_i \to 1$, $\overline{H}_i^\circ$ must be the partial molar enthalpy (the molar enthalpy) of pure i (confirming what we stated by simple inspection, above).

Note that in the case of multicomponent solutions infinite dilution means infinite dilution of all components, not just of component i. Thus the Henryan standard states, which seem so unattainable, are actually convenient because some of their properties are the same as those of the infinitely dilute solution, and these are obtainable by extrapolation from measurements at finite concentration.

8.4.2 Pressure

In considering the effect of pressure on activity, we must recall that the standard state pressure (P°) is not always the same as the system pressure (P), so that the differentiation with respect to pressure is not always completely analogous to differentiation with respect to temperature. The argument here is very similar to that in §6.3.2.

First of all, for variable pressure standard states, those that do have $P^\circ = P$, we have

$$\begin{aligned}\frac{\partial}{\partial P}(RT\ln a_i) &= \frac{\partial}{\partial P}(RT\ln \gamma_i)\\ &= \frac{\partial}{\partial P}(\mu_i - \mu_i^\circ)\\ &= \overline{V}_i - \overline{V}_i^\circ\end{aligned}$$

where γ_i may be γ_f, γ_R, or γ_H, and $\overline{V}_i^\circ$ is either the molar volume of pure i or the partial molar volume of i at infinite dilution, depending on the standard state used (see above). Integration of this expression requires a knowledge of the variation of the *relative partial molar volume* $\overline{V}_i - \overline{V}_i^\circ$ with pressure for which there is no general expression. There is also no special symbol for $\overline{V}_i - \overline{V}_i^\circ$ as there is for $\overline{H}_i - \overline{H}_i^\circ$.

However for the fixed pressure standard states,

$$\frac{\partial}{\partial P}(\mu_i^\circ) = 0 \tag{8.41}$$

because changing P does not change the pressure on the standard state (P°). Therefore

$$\frac{\partial}{\partial P}(R \ln a_i) = \overline{V}_i$$

A case of particular interest to us is the activity of solids, for which in many cases the assumption that $\overline{V}_i$ is unaffected by pressure (solid i is incompressible) is reasonable. If $\overline{V}_i$ is a constant, then

$$\int_{P_1}^{P_2} d \ln a_i = \frac{1}{RT} \int_{P_1}^{P_2} \overline{V}_i dP$$

$$\ln(a_i)_{P_2} - \ln(a_i)_{P_1} = \overline{V}_i(P_2 - P_1)/RT$$

Furthermore, if $P_1 = 1$ bar and the standard state of i is pure i at T and one bar, then pure i at P_1 has an activity of 1, and

$$\ln(a_i)_{P_2} = \ln(a_i)_P = \overline{V}_i(P - 1)/RT$$

or for *pure* solids

$$\ln(a_i)_P = V_i(P - 1)/RT \tag{8.42}$$

This permits calculation of the activity of a pure mineral at any pressure, relative to the same mineral at one bar.

For example, consider calcite and aragonite at equilibrium at 3737 bars and 25 °C, from our example in Chapter 6. The molar volumes of calcite and aragonite are taken from Appendix B (or Table 6.1), so for calcite Equation (8.42) gives

$$\begin{aligned}\ln a_{\text{calcite}, P=3737} &= 3.6934 \cdot (3737 - 1)/(8.31451 \times 298.15)\\ &= 5.566\end{aligned}$$

so

$$a_{\text{calcite}, P=3737} = 261$$

and the activity of aragonite is

$$\begin{aligned}\ln a_{\text{aragonite}, P=3737} &= 3.4150 \cdot (3737 - 1)/(8.31451 \times 298.15)\\ &= 5.147\end{aligned}$$

so

$$a_{\text{aragonite},P=3737} = 172$$

Thus $CaCO_3$ in the form of calcite and aragonite at 3737 bars, although having the same chemical potential, has two different activities because there are two different standard states. Recalling that activity is the ratio f_i/f_i°, we see too that the physical meaning of an activity of calcite of 261 is that the fugacity (which ≈ vapor pressure) of calcite is increased by a factor of 261 when it is squeezed from 1 bar to 3737 bars. We know this without knowing either f at 3737 bars or f° at one bar for calcite.

8.5 Activities and standard states: an overall view

We have now said everything necessary about activities and standard states, but the overall effect for the newcomer is often one of confusion at this stage. To try to draw the various threads together we consider in Figure 8.8 a hypothetical three-phase equilibrium at temperature T and pressure P. A solid crystalline solution of B in A is in contact with an aqueous solution of A(*aq*) and B(*aq*), which is in turn in contact with a vapor phase containing A(*v*) and B(*v*) in addition to water vapor. We can suppose the dissolution of (A,B)(*s*) to be stoichiometric so that the ratio of A to B is the same in all three phases, but this is irrelevant to our development as we consider only component A. Let's say that for a solid solution composition of $x_A = 0.5$, $x_B = 0.5$, the concentration of A(*aq*) at equilibrium (m_A) is 10^{-2} molal, and the fugacity of A in the vapor (f_A) is 10^{-5} bars. Assuming activity coefficients in the solid and liquid phases to be 1.0, the activity of A in the solid solution (using a standard state of pure crystalline A at T and P) is 0.5, the activity of A in the aqueous solution (using a standard state of the hypothetical ideal one molal solution of A at T and P) is 10^{-2}, and the activity of A in the vapor (using a standard state of pure ideal gaseous A at T and one bar) is 10^{-5}. Because the system is at equilibrium, the chemical potential of A (μ_A) is the same in each of the three phases, but because the three standard states are different, the standard chemical potential of A (μ_A°) is different for the three phases. The difference ($\mu_A - \mu_A^\circ$) is calculable from the equations we have just derived. Thus, letting $T = 25\,°C$,

$$\begin{aligned}
(\mu_A - \mu_A^\circ)_{\text{solid}} &= RT\ln x_A \\
&= 8.314\,51 \times 298.15 \times \ln(0.5) \\
&= -1.72\,\text{kJ mol}^{-1} \\
(\mu_A - \mu_A^\circ)_{\text{aq}} &= RT\ln m_A \\
&= 8.314\,51 \times 298.15 \times \ln(10^{-2}) \\
&= -11.4\,\text{kJ mol}^{-1}
\end{aligned}$$

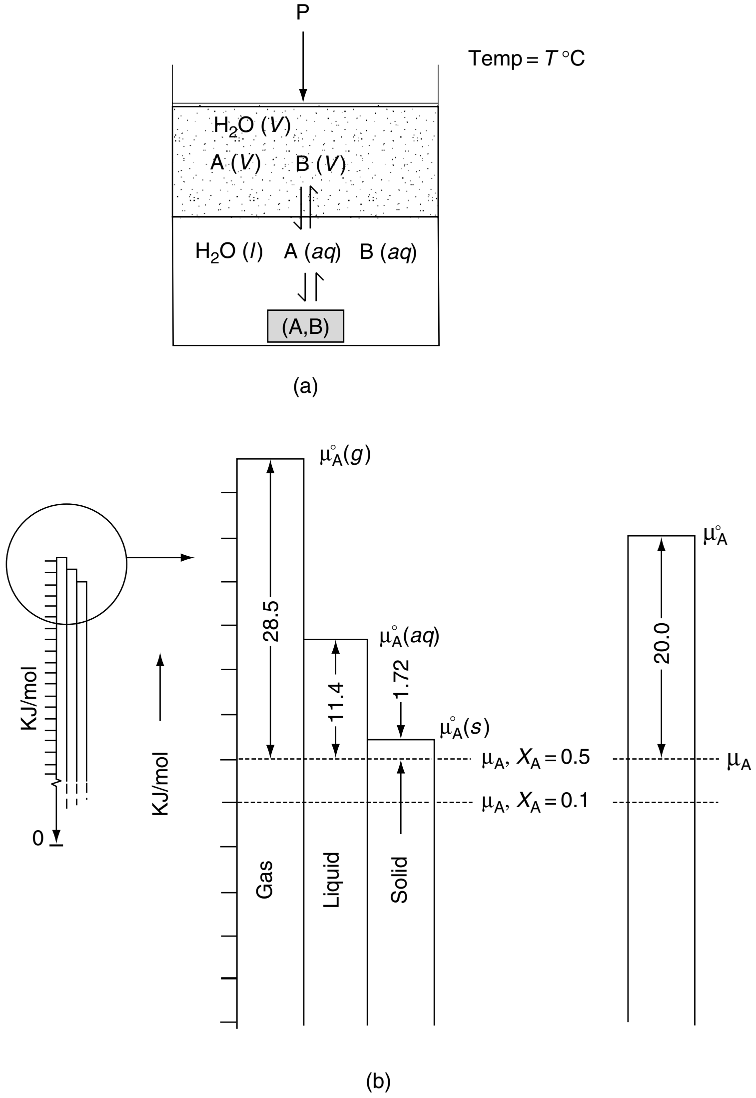


Figure 8.8 (a) A hypothetical three-phase system at equilibrium at pressure *P* and temperature *T*. (b) The top part of the histogram of chemical potentials in kJ mol^{-1}. The length of the bar for each phase is fixed when the standard state is chosen, and the chemical potential of A in the equilibrium system is represented by a line across the histogram at a level depending on the amount of B in the system. The lengths of the bars on the left represent traditional standard states, but any position for the top of the bars could be chosen, such as the one on the right, thus defining a new standard state.

$$
\begin{aligned}
(\mu_{\mathrm{A}} - \mu_{\mathrm{A}}^{\circ})_{\mathrm{gas}} &= RT \ln f_{\mathrm{A}} \\
&= 8.31451 \times 298.15 \times \ln(10^{-5}) \\
&= -28.5\,\mathrm{kJ\,mol^{-1}}
\end{aligned}
$$

It is instructive to consider these differences on a histogram (Figure 8.8) in which the ordinate is a scale of $\mathrm{kJ\,mol^{-1}}$, on which we plot the *absolute* chemical potentials of A. These absolute potentials may be very large, so we look at only the tops of the bars in the histogram, and we unfortunately don't know the values of the absolute potentials individually, so we can't put an absolute scale on the ordinate. But we *can* plot the relative positions of the tops of the bars, and the position of the equilibrium chemical potential of A in the system.

If we now consider systems having more and more B in the solid solution (and hence in the other two phases), but always at equilibrium, the histogram bars stay where they are (because we are not changing standard states) but the level of the (absolute) chemical potential of A is lowered, increasing the distance between the top of the histogram bar for each phase and the level of μ_{A}, that is, increasing the (negative) value of $(\mu_{\mathrm{A}} - \mu_{\mathrm{A}}^{\circ})$ as the activity of A is lowered.

This diagram is worth careful thought. It illustrates several things that are useful in understanding activities, chemical potentials, and standard states, such as the absolute nature of chemical potentials and the necessity of using differences, the equality of chemical potentials in each phase, and the arbitrary nature of the standard state.

This is all very nice, but wouldn't it be a lot simpler to have the same activity in every phase, just as we have the same chemical potential in every phase? This is worth some thought too. Having the same activity in each phase means having the same value of $\mu - \mu^{\circ}$ in each phase, and presumably we could choose whatever value we like for this. If we chose this to be $-20\,\mathrm{kJ\,mol^{-1}}$, the histogram for every phase would look like that shown in Figure 8.8. Then because the three standard states have the same value of μ°, they could all coexist at equilibrium (if they could exist at all). However, arbitrarily choosing a_i means arbitrarily choosing f_i° and m_i°, and this results in some standard states that are even more weird than the traditional ones. For example, if $\mu - \mu^{\circ} = -20\,\mathrm{kJ\,mol^{-1}}$, then $a_i = 10^{-3.50}$, and if $m_i = 10^{-5}$ and $m_i/m_i^{\circ} = 10^{-3.5}$, then $m_i^{\circ} = 10^{1.5}$ or 31.9. So the standard state becomes a hypothetical ideal 31.9 molal solution.

So probably it is better to stick to having different standard states and different activities in each phase.

8.5.1 Changing standard states

Another way to think about standard states is to consider how to change from one to another. For example, let's say we have a real system consisting of a gas in contact with a liquid. Methane in the gas has a measured partial pressure of 0.01 bar, which we take to be its fugacity, and we want to calculate how much methane is dissolved in the water. In the gas phase, $f_{\mathrm{CH_4}} = 0.01\,\mathrm{bar}$, and using a standard state of ideal methane gas at 1 bar ($f^\circ_{\mathrm{CH_4}} = 1.0\,\mathrm{bar}$), we have an activity of methane of 0.01 ($a_{\mathrm{CH_4}} = 0.01$). The relevant reaction is

$$\mathrm{CH_4}(g) = \mathrm{CH_4}(aq)$$

and getting numbers from the tables we have

$$\begin{aligned}\Delta_r G^\circ &= \Delta_f G^\circ_{\mathrm{CH_4}(aq)} - \Delta_f G^\circ_{\mathrm{CH_4}(g)}\\ &= -34\,451 - (-50\,720)\\ &= 16\,269\,\mathrm{J\,mol^{-1}}\end{aligned}$$

This means that, *because the elements all cancel out* (§3.5),

$$G^\circ_{\mathrm{CH_4}(aq)} - G^\circ_{\mathrm{CH_4}(g)} = 16\,269\,\mathrm{J\,mol^{-1}}$$

where $G^\circ_{\mathrm{CH_4}(aq)}$ is the absolute G of methane in an ideal 1 molal solution, and $G^\circ_{\mathrm{CH_4}(g)}$ is the absolute G of methane as an ideal gas at 1 bar. Then

$$\begin{aligned}\log K &= -\Delta_r G^\circ/(2.30\,259\,RT)\\ &= -16\,269/(2.302\,59 \times 8.314\,51 \times 298.15)\\ &= -2.850\end{aligned}$$

where

$$K = \frac{a_{\mathrm{CH_4}(aq)}}{a_{\mathrm{CH_4}(g)}}$$

Therefore if $a_{\mathrm{CH_4}(g)} = 0.01$, then

$$10^{-2.850} = \frac{a_{\mathrm{CH_4}(aq)}}{0.01}$$

and

$$\begin{aligned}a_{\mathrm{CH_4}(aq)} &= 0.01 \times 10^{-2.850}\\ &= 10^{-4.850}\end{aligned}$$

So the calculated activity of dissolved methane is $10^{-4.85}$. With a standard state of ideal 1 molal methane, this means $m_{\mathrm{CH_4}}\gamma_{\mathrm{CH_4}} = 10^{-4.85}$, and on the reasonable assumption that $\gamma_{\mathrm{CH_4}} = 1.0$, then $m_{\mathrm{CH_4}} = 10^{-4.85}$. So in spite of the fact that two different standard states are used for the same component in the same reaction, we arrive at a useful answer. This is because the standard states used do not influence the final result in any way. They are simply useful fictions to hang tabulated data on.

To illustrate this, we can choose a completely different standard state for the gaseous methane and see what happens. Our new standard state is ideal gaseous methane at some pressure chosen at random, such as 978.4852 bar. First, we need $\Delta_f G^\circ$ of methane in this new state. Because the gas is ideal, this is easy. From Equation (8.5), and because for an ideal gas, $f = P$,

$$\begin{aligned} G_{978.4852\ \text{bars}} - G_{1\ \text{bar}} &= RT\ln(978.4852/1) \\ &= 17\,070.2\,\text{J}\,\text{mol}^{-1} \end{aligned}$$

This is the change in G as ideal methane is squeezed from 1 bar to 978.4852 bar. The difference in G between ideal methane at 1 bar and the elements C and $2\,H_2$ at 1 bar is $-50\,720\,\text{J}\,\text{mol}^{-1}$ (Figure 8.9). Therefore, the difference in G between methane ideal gas at 978.4852 bar and the elements at 1 bar is $-50\,720 + 17\,070 = -33\,650\,\text{J}\,\text{mol}^{-1}$. We can call this $\Delta_f G^\star$, our new standard state Gibbs energy. Our calculation is now

$$\begin{aligned} \Delta_r G^{\circ\star} &= \Delta_f G^\circ_{\text{CH}_4(aq)} - \Delta_f G^\star_{\text{CH}_4(g)} \\ &= -34\,451 - (-33\,650) \\ &= -801\,\text{J}\,\text{mol}^{-1} \\ G^\circ_{\text{CH}_4(aq)} - G^\star_{\text{CH}_4(g)} &= -801\,\text{J}\,\text{mol}^{-1} \end{aligned}$$

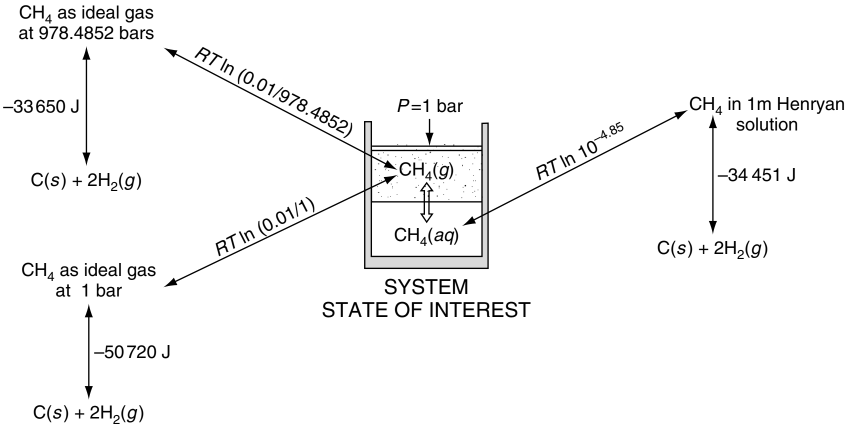


Figure 8.9 The pressure on the gas phase is 1 bar, but the partial pressure of CH_4 in the gas is 0.01 bar. The Gibbs energy (μ) of CH_4 is the same in each phase, and a variety of standard state data may be used.

Then

$$\begin{aligned}\log K &= -\Delta_r G^{\circ\star}/(2.302\,59\,RT)\\ &= 801/(2.302\,59 \times 8.314\,51 \times 298.15)\\ &= 0.140\end{aligned}$$

or

$$\begin{aligned}K &= \frac{a_{\mathrm{CH_4}(aq)}}{a_{\mathrm{CH_4}(g)}}\\ &= 10^{0.140}\\ &= 1.38\end{aligned}$$

Now $a_{\mathrm{CH_4}(g)}$ is no longer 0.01. Because of the change of standard state,

$$\begin{aligned}a_{\mathrm{CH_4}(g)} &= \frac{f_{\mathrm{CH_4}(g)}}{f^{\star}_{\mathrm{CH_4}(g)}}\\ &= \frac{0.01}{978.4852}\\ &= 10^{-4.99}\end{aligned}$$

and so, solving for $a_{\mathrm{CH_4}(aq)}$ with our new K,

$$\begin{aligned}\frac{a_{\mathrm{CH_4}(aq)}}{a_{\mathrm{CH_4}(g)}} &= 1.38\\ a_{\mathrm{CH_4}(aq)} &= 1.38 \times a_{\mathrm{CH_4}(g)}\\ &= 1.38 \times 10^{-4.99}\\ &= 10^{-4.85}\end{aligned}$$

as before. So using any arbitrary standard state makes no difference at all. If you follow the calculations closely, you will see that the properties of the standard state cannot affect the results. They are a convenient repository for tabulated data derived from experimental work. As shown in Figure 8.9, all paths from the elements to the equilibrium state must give the same total change in G, because G of all products and reactants is fixed in both states. The standard state is merely a repository somewhere along the elements→equilibrium state path.

Units again

A reminder:

- Activities and activity coefficients (a, γ) have no units, but fugacity (f) does.
- Activities have standard states, but fugacities do not.

8.6 Summary

It would be hard to overemphasize the importance of the concepts of fugacity and activity and their relationship to the chemical potential, at least in chemical thermodynamics. In thermodynamics for engineers or physicists, chemical reactions play a smaller role, but for anyone interested in processes involving chemical changes, this is a central topic. That is because the fugacity and activity are the parameters which relate the composition of a system to its Gibbs energy; how changes in phase compositions change the Gibbs energy of the system, and so determine the equilibrium phase compositions at a given T and P.

Fugacity and activity are basically compositional terms. In ideal solutions they are not necessary; pressure and various composition terms can be directly linked to the Gibbs energy. Real solutions have a variety of intermolecular forces, so that ideal solution models need correction factors. These corrections can be made either to the composition terms (fugacity and activity coefficients) or to the thermodynamic potentials (excess functions), and efforts to model these correction factors in mathematical terms have always been, and likely always will be, an important research field.

Because there are three main kinds of solutions and there are several commonly used methods of expressing concentrations, activity can take on several different forms requiring different kinds of correction factors, and because Gibbs energy is always a difference in energy between two states, there are several different standard or reference states in common use. This all adds up to possible confusion, and although it is possible to learn how things are done and to follow the rules, it is of course much better to understand the reasons for why things are done this way.

9
The equilibrium constant

9.1 Reactions in solution

Reactions involving dissolved compounds are different in an important way from reactions involving only pure compounds, such as pure solids. To see why, consider two reactions, one between pure compounds and one between dissolved substances.

The first is a reaction in which all products and reactants are pure substances, the kind of reaction we have been considering up to now. It is

$$\mathrm{NaAlSiO_4}(s) + 2\,\mathrm{SiO_2}(s) = \mathrm{NaAlSi_3O_8}(s) \tag{9.1}$$

The second is a reaction in which all products and reactants are dissolved in water and are capable of changing their concentration:

$$\mathrm{H_2CO_3}(aq) = \mathrm{HCO_3^-}(aq) + \mathrm{H^+}(aq) \tag{9.2}$$

The temperature and pressure are normal, 298.15 K and 1 bar. As usual, we want to know which way each reaction will go. Reaction (9.1) presents no problem. We look up the values of $\Delta_f G^\circ$ for each compound, and calculate$\Delta_r G^\circ$:

$$\begin{aligned}\Delta_r G^\circ &= \Delta_f G^\circ_{\mathrm{NaAlSi_3O_8}} - \Delta_f G^\circ_{\mathrm{NaAlSiO_4}} - 2\,\Delta_f G^\circ_{\mathrm{SiO_2}}\\ &= -3711.5 - (-1978.1) - 2(-856.64)\\ &= -20.12\,\mathrm{kJ\,mol^{-1}}\end{aligned}$$

We see that the reaction as written is spontaneous; $\mathrm{NaAlSiO_4}$ (nepheline) and $\mathrm{SiO_2}$ (quartz) at 1 bar pressure should react together to form $\mathrm{NaAlSi_3O_8}$ (albite). If the reaction does proceed (thermodynamics doesn't tell us whether it will or not, only that the energy gradient favors it), then nepheline and quartz get used up during the reaction. However, while being used up, they do not change their Gibbs energies. The reaction should actually proceed as long as any reactants are left. When either the nepheline or the quartz is used up completely, the reaction must stop. This reaction can be represented graphically as in Figure 9.1. Here we use bars to represent the magnitude of the combined Gibbs energy of the products and of the reactants. The difference in the height of the bars

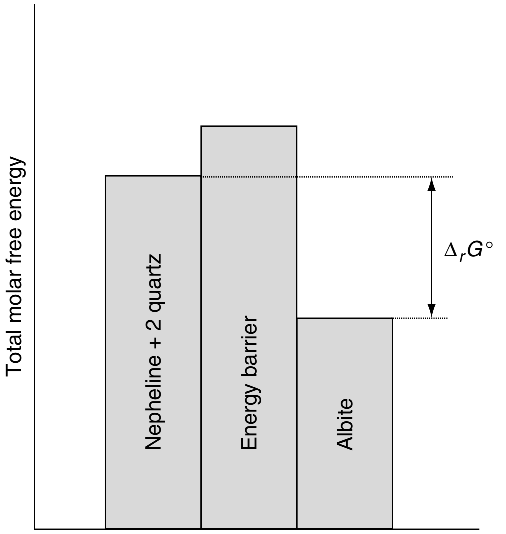


Figure 9.1 Molar Gibbs energies when all products and reactants are pure compounds. The Gibbs energy of reaction is given by $\Delta_r G°$ because all products and reactants are in their reference states, and this does not change during the reaction until one of the reactants disappears.

represents $\Delta_r G°$, the driving force for the reaction. The middle bar represents an activation energy barrier that prevents the reaction from occurring. It is put there to form a link with the discussion in Chapter 2, but thermodynamics is unable to calculate the size of this barrier, or anything whatever about it. Nevertheless it is often there, and is the reason that one of the states is metastable. The point here is that the size of the bars does not change during the reaction, if it proceeds, because none of the products or reactants changes in any way–only the amounts present change. The value of $\Delta_r G°$ never goes to zero.

Reaction (9.2) is different. We can start off the same way, by looking up the values of $\Delta_f G°$ for each compound:

$$\begin{aligned}\Delta_r G° &= \Delta_f G°_{\mathrm{HCO}_3^-(aq)} + \Delta_f G°_{\mathrm{H}^+(aq)} - \Delta_f G°_{\mathrm{H_2CO_3}(aq)}\\ &= -586.77 + 0 - (-623.109)\\ &= 36.339\,\mathrm{kJ\,mol^{-1}}\end{aligned}$$

This is positive, and so the reaction goes spontaneously to the left. So far, so good. But as soon as the reaction starts, the concentrations of H^+ and HCO_3^- start to decrease, the concentration of H_2CO_3 starts to increase, and the Gibbs energies of all three change, as shown in Figure 9.2. All we can say from the tabulated data is that if all three aqueous species were present in their reference state concentrations, the reaction would start to go to the left. But suppose we are interested in some other concentrations? And what happens to the reaction after it starts? Because the solutes can change their concentrations and their Gibbs energies, the situation is quite different from the "all pure substances" situation. These problems are all handled easily by the *equilibrium constant*.

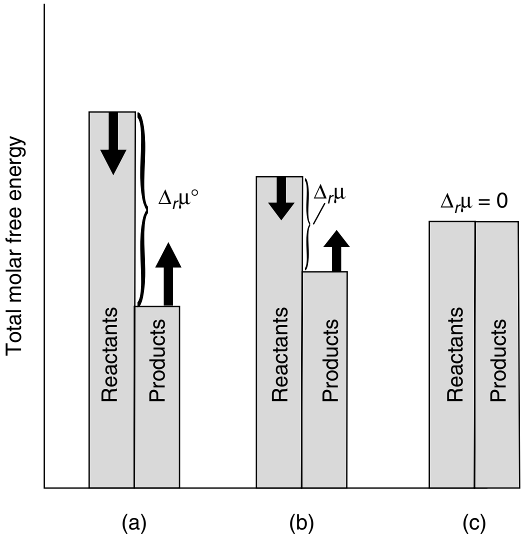


Figure 9.2 Molar Gibbs energies of reactants and products of a reaction between aqueous solutes. (a) A hypothetical starting condition, represented by the numbers in the tables of data (note superscript ° on $\Delta_r\mu°$, indicating standard conditions for all products and reactants). (b) Either the reaction in (a) after it has proceeded for some time, or a beginning state where the reactants and products are not all at 1 molal (no superscript °). The point is that, as shown by the arrows, the Gibbs energy of the reactants decreases and that of the products increases during the reaction. (c) Sooner or later, a state of equilibrium is reached, when the Gibbs energies of reactants and products are equal.

9.2 Reactions at equilibrium

Chemical reactions can not only go one way or the other (our main problem), but they can stop going for two reasons. Either one of the reactants is used up, or the reaction can reach an equilibrium state, with all products and reactants present in a balanced condition. The second possibility is the subject of this chapter–how much can we predict about this balanced state of equilibrium?

In Chapter 4 we defined the molar Gibbs energy, G, which always decreases in spontaneous reactions ($\Delta G < 0$). In Chapter 6, we used the fact that a reaction at equilibrium (e.g., calcite $\rightleftharpoons$ aragonite) does not go either way ($\Delta G = 0$) to calculate the T and P of equilibrium between phases. The expression $\Delta G = 0$ expresses a balance between the Gibbs energies of calcite and aragonite, that is, $G_{\text{calcite}} = G_{\text{aragonite}}$ (§6.3.1). If there is more than one reactant or product, the same relationship must hold (the G of reactants and products are equal), but each side is now a sum of G terms, and the G terms for solutes are properly written as μ rather than G. Of course, not all products and reactants need be solutes. For example, the reaction

$$\mathrm{SiO_2}(s) + 2\mathrm{H_2O} = \mathrm{H_4SiO_4}(aq) \qquad (9.3)$$

shows what happens when quartz dissolves in water. Molecules of SiO_2 dissolve and combine with water molecules to form the solute species H_4SiO_4. This dissolution process continues until the solution is saturated with silica, and then stops. The system is then at equilibrium, because

$$\mu_{H_4SiO_4} = \mu_{SiO_2} + 2\mu_{H_2O} \tag{9.4}$$

If we added some H_4SiO_4 to this solution it would then be supersaturated, $\mu_{H_4SiO_4}$ would be greater than its equilibrium value, and the reaction would tend to go to the left, precipitating quartz.[1]

9.3 The most useful equation in thermodynamics

To find out what we can say about this balanced equilibrium state when several solutes and other phases are involved, let's consider a general chemical reaction

$$a\mathrm{A} + b\mathrm{B} = c\mathrm{C} + d\mathrm{D} \tag{9.5}$$

where A, B, C, and D are chemical formulas, and *a*, *b*, *c*, *d* (called stoichiometric coefficients) are any numbers (usually small integers) that allow the reaction to be balanced in both composition and electrical charges, if any. When this reaction reaches equilibrium,

$$c\mu_C + d\mu_D = a\mu_A + b\mu_B$$

and

$$\begin{aligned}\Delta_r\mu &= c\mu_C + d\mu_D - a\mu_A - b\mu_B \\ &= 0\end{aligned} \tag{9.6}$$

By our definition of activity, Equation (7.37),

$$\mu_A = \mu_A^\circ + RT \ln a_A$$
$$\mu_B = \mu_B^\circ + RT \ln a_B$$
$$\mu_C = \mu_C^\circ + RT \ln a_C$$

and

$$\mu_D = \mu_D^\circ + RT \ln a_D$$

---

[1] At the risk of becoming repetitious, we note that it is in our model reaction that quartz precipitates. In real life, something else might happen – nothing might precipitate, or some other SiO_2 phase such as silica gel might precipitate.

Substituting these expressions into (9.6), we get

$$\begin{aligned}
\Delta_r\mu &= c\mu_C + d\mu_D - a\mu_A - b\mu_B \\
&= c(\mu_C^\circ + RT\ln a_C) + d(\mu_D^\circ + RT\ln a_D) \\
&\quad - a(\mu_A^\circ + RT\ln a_A) - b(\mu_B^\circ + RT\ln a_B) \\
&= (c\mu_C^\circ + d\mu_D^\circ - a\mu_A^\circ - b\mu_B^\circ) + RT\ln a_C^c + RT\ln a_D^d \\
&\quad - RT\ln a_A^a - RT\ln a_B^b \\
&= \Delta_r\mu^\circ + RT\ln\left(\frac{a_C^c a_D^d}{a_A^a\, a_B^b}\right)
\end{aligned}$$

There may be any number of reactants and products, and so to be completely general we can write

$$\Delta_r\mu = \Delta_r\mu^\circ + RT\ln\prod_i a_i^{\nu_i} \tag{9.7}$$

where i is an index that can refer to any product or reactant, ν_i refers to the stoichiometric coefficients of the products and reactants, with ν_i positive if i is a product, and negative if i is a reactant. $\prod\limits_i$ (or $\prod_i$) is a symbol meaning "product of all i terms," which means that all the $a_i^{\nu_i}$ terms are to be multiplied together (much as $\sum_i a_i$ would mean that all a_i terms were to be added together). So in our case, the ν terms are c, d, $-a$, and $-b$, and

$$\begin{aligned}
\prod_i a_i^{\nu_i} &= a_C^c\, a_D^d\, a_A^{-a}\, a_B^{-b} \\
&= \frac{a_C^c\, a_D^d}{a_A^a\, a_B^b}
\end{aligned}$$

In the general case, $\prod_i a_i^{\nu_i}$ is given the symbol Q, so (9.7) becomes

$$\Delta_r\mu = \Delta_r\mu^\circ + RT\ln Q \tag{9.8}$$

We must be perfectly clear as to what (9.8) means. In Figure 9.3 (a variation of Figure 9.2) are pictured the possible relationships between the Gibbs energies of the products and reactants in reaction (9.5).

First, the term $\Delta_r\mu^\circ$ refers to the difference in Gibbs energies of products and reactants when each product and each reactant, whether solid, liquid, gas, or solute, is in its pure reference state. This means the pure phase for solids and liquids [e.g., most minerals, $H_2O(s)$, $H_2O(l)$, alcohol, etc.], pure ideal gases at 1 bar [e.g., $O_2(g)$, $H_2O(g)$, etc.], and dissolved substances [solutes, e.g., $NaCl(aq)$, Na^+, etc.] in ideal solution at a concentration of 1 molal. Although we do have at times fairly pure solid phases in our real systems (minerals such as quartz and calcite are often quite pure), we rarely have pure liquids or gases, and we never have ideal solutions as concentrated as 1 molal.

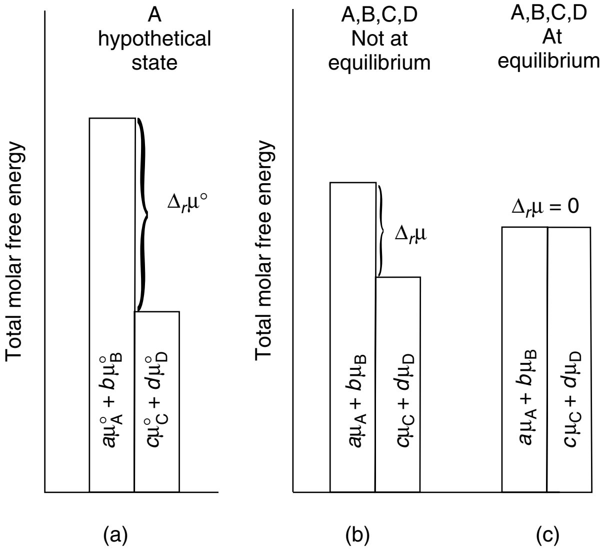


Figure 9.3 Possible Gibbs energy relationships in the reaction $a\text{A} + b\text{B} \rightarrow c\text{C} + d\text{D}$.

Therefore, $\Delta_r\mu^\circ$ usually refers to quite a hypothetical situation. It is best not to try to picture what physical situation it might represent, but to think of it as just the difference in numbers that are obtained from tables.

$\Delta_r\mu$, on the other hand, is the difference in Gibbs energy of reactants and products as they actually occur in the system you are considering, which may or may not have reached stable equilibrium.[2] The activities in the Q term (the concentrations, fugacities, mole fractions, etc. of the products and reactants) change during the reaction as it strives to reach equilibrium and at any particular moment result in a particular value of $\Delta_r\mu$. Thus $\Delta_r\mu^\circ$ is a number obtained from tables that is independent of what is happening in the real system you are considering, but $\Delta_r\mu$ and Q are linked together – whatever activities (think *concentrations*) are in Q will result in a certain value of $\Delta_r\mu$.

If it makes more sense, you can write Equation (9.8) as

$$\Delta_r\mu - \Delta_r\mu^\circ = RT \ln Q \tag{9.9}$$

which means that whatever terms are in Q control how different the chemical potentials ($\Delta_r\mu$) are from their standard tabulated values ($\Delta_r\mu^\circ$). When all activities in Q are 1.0, then there is no difference, $\Delta_r\mu = \Delta_r\mu^\circ$.

We are especially interested in the value of Q when our systems reach equilibrium, that is, when the product and reactant activities have adjusted

---

[2] Strictly speaking, μ has meaning only in equilibrium states, so we cannot really consider the reaction which is not at equilibrium. What we really do is to consider the reaction as taking place in a series of metastable equilibrium states, as discussed more fully in connection with the progress variable in Chapter 18. At this stage, however, you may consider this a mere quibble, and think of reacting substances as having μ values if you wish.

themselves spontaneously such that $\Delta_r\mu = 0$. In this state, the $\prod_i a_i^{\nu_i}$ term is called K, instead of Q, and (9.8) becomes

$$0 = \Delta_r\mu^\circ + RT\ln K$$

or

$$\Delta_r\mu^\circ = -RT\ln K \tag{9.10}$$

Standard states usually refer to pure substances (except for the aqueous standard states) in which $\mu = G$, so this equation is often written

$$\boxed{\Delta_r G^\circ = -RT\ln K} \tag{9.11}$$

This equation has been called, with some reason, the most useful in chemical thermodynamics, and it certainly merits the most careful attention. Most important is the fact that the activity product ratio (K) on the right-hand side is independent of variations in the system composition. Its value is controlled completely by a difference in standard (tabulated) state Gibbs energies ($\Delta_r G^\circ$) and so is a function only of the temperature and pressure. It is a constant for a given system at a given temperature or temperature and pressure and is called the *equilibrium constant*. Its numerical value for a given system is not dependent on the system actually achieving equilibrium, or in fact even existing. Its value is fixed when the reacting substances are chosen. The left-hand side refers to a difference in Gibbs energies of a number of different physical and ideal states, which do not represent any real system or reaction. The right-hand side, on the other hand, refers to a single reaction that has reached equilibrium, or more exactly, to the activity product ratio that would be observed if the system had reached equilibrium.

The great usefulness of Equation (9.11) lies in the fact that knowledge of a few standard state Gibbs energies allows calculation of an indefinite number of equilibrium constants. Furthermore, these equilibrium constants are very useful pieces of information about any reaction. If K is very large, it tends to shows that a reaction will tend to go "to completion," that is, mostly products will be present at equilibrium, and if K is small, it tends to show that the reaction hardly goes at all before enough products are formed to stop it.[3] If you are a chemical engineer designing a process to produce some new chemical, it is obviously of great importance to know to what extent reactions should theoretically proceed. The equilibrium constant, of course, will never tell you whether reactants will actually react, or at what rate; there may be some reason for reaction kinetics being very slow. It indicates the activity product ratio at equilibrium, not whether equilibrium is easily achievable.

---

[3] These are just generalizations which are not always true, because the activities of products and reactants at equilibrium can be very large or very small. We emphasize this a bit more in the boxed statement on page 266.

Finally, NEVER write Equation (9.11) as

$$\Delta_r G = -RT \ln K$$

that is, omitting the superscript °, because doing so indicates a complete lack of understanding of the difference between $\Delta_r G$ and $\Delta_r G^\circ$ and is just about grounds for failing any course in this subject.

Let's go over it once more. $\Delta_r G^\circ$ (or $\Delta_r \mu^\circ$) is the difference in Gibbs energy between products and reactants when they are all in their reference states (pure solids and liquids, solutes at ideal 1 molal, gases at 1 bar), determined directly from the tables. Products and reactants are virtually never at equilibrium with each other under these conditions ($\Delta_r G^\circ$ or $\Delta_r \mu^\circ$ never becomes equal to zero). $\Delta_r G$ (or $\Delta_r \mu$) is the difference in free energy between products and reactants in the general case (when at least one of the products or reactants is *not* in its reference state) and becomes equal to zero when the reaction reaches equilibrium. $\Delta_r G$ cannot be used in place of $\Delta_r G^\circ$ in (9.11) because this would mean, among other things, that every reaction at equilibrium ($\Delta_r G = 0$) would have an equilibrium constant of 1.0.

9.3.1 A first example

Let's calculate the equilibrium constant for reaction (9.2),

$$H_2CO_3(aq) = HCO_3^-(aq) + H^+(aq)$$

First we write, as before,

$$\Delta_r G^\circ = \Delta_f G^\circ_{HCO_3^-} + \Delta_f G^\circ_{H+} - \Delta_f G^\circ_{H_2CO_3}$$

Getting numbers from the tables, we find

$$\begin{aligned}\Delta_r G^\circ &= -586.77 + 0 - (-623.109)\\ &= 36.339\,\text{kJ mol}^{-1}\\ &= 36\,399\,\text{J mol}^{-1}\end{aligned}$$

The fact that this number is positive is not as significant as in our previous examples. In this case it means that *if* H_2CO_3, HCO_3^-, and CO_3^{2-} were all present in an ideal solution, and each had a concentration of 1 molal, the reaction would go to the left. This hypothetical situation is not of much interest. We want the value of K.

Inserting this result in Equation (9.11), we get

$$\begin{aligned}\Delta_r G^\circ &= -RT \ln K\\ 36\,339 &= -(8.3145 \times 298.15) \ln K\end{aligned}$$

so

$$\begin{aligned}\ln K &= -36\,339/(8.3145 \times 298.15)\\ &= -14.659\end{aligned}$$

or

$$\begin{aligned}K &= 4.30 \cdot 10^{-7}\\ &= 10^{-6.37}\end{aligned}$$

If you don’t like dealing with natural logarithms, you can use the conversion factor $\log x = \ln x/2.302\,59$ (Appendix A). This gives

$$\begin{aligned}\log K &= -36\,339/(2.302\,59 \times 8.3145 \times 298.15)\\ &= -6.37\end{aligned}$$

directly.

This means that when these three aqueous species are at equilibrium,

$$\frac{a_{\mathrm{HCO_3^-}} \cdot a_{\mathrm{H^+}}}{a_{\mathrm{H_2CO_3}}} = 10^{-6.37}$$

This is the answer to our question in §9.1 (“what happens to the reaction after it starts?”). The reaction continues until the ratios of the activities of the products and reactants equals the equilibrium constant, in this case $10^{-6.37}$. It doesn’t matter what the starting activities were, and individual activities at equilibrium can be quite variable. In other words the values of $a_{\mathrm{H_2CO_3}}$ and of $(a_{\mathrm{HCO_3^-}} \cdot a_{\mathrm{H^+}})$ are not determined, nor are the values of $a_{\mathrm{HCO_3^-}}$ or $a_{\mathrm{H^+}}$ individually; only the ratio expressed by K is fixed. In specific cases, the values of these individual activities are determined by the bulk composition of the solution, and can be determined by *speciation* (Chapter 16). For now, we are content to determine K. In this case K is the ionization constant for carbonic acid, H_2CO_3. It is a very small number, meaning that carbonic acid is a weak acid.

9.4 Special meanings for *K*

Equilibrium constants are also sometimes equal to system properties of interest, such as vapor pressures, solubilities, phase compositions, and so on. This is because quite often it can be arranged that all activity terms drop out (are equal to 1.0) except the one of interest, which can then be converted to a pressure or composition.

9.4.1 *K* equal to a solubility

Quartz–water example

In our quartz–water example (Equation 9.3), the equilibrium constant expression is

$$K = \frac{a_{\mathrm{H_4SiO_4}}}{a_{\mathrm{SiO_2}}\, a^2_{\mathrm{H_2O}}} \tag{9.12}$$

At this point the expression is perfectly general, valid for any conditions, and K is calculable from Equation (9.11) if we know the Gibbs energies of the three species in their reference states. In this case we are dealing with pure quartz and water saturated with quartz. The quartz is in its reference state, and the water contains so little silica that it is almost pure.[4] By our definitions then (Equations 8.30) $a_{\mathrm{SiO_2}} = 1$ and $a_{\mathrm{H_2O}} = 1$. Therefore

$$\begin{aligned} K &= a_{\mathrm{H_4SiO_4}} \\ &= (m_{\mathrm{H_4SiO_4}}\,\gamma_{\mathrm{H_4SiO_4}}) \\ &= m_{\mathrm{H_4SiO_4}} \qquad \text{assuming } \gamma_{\mathrm{H_4SiO_4}} = 1.0 \end{aligned}$$

This shows that assuming $\gamma_{\mathrm{H_4SiO_4}}$ is 1.0, which happens to be an excellent approximation in this case, we can calculate the concentration of silica ($m_{\mathrm{H_4SiO_4}}$) in equilibrium with quartz, that is, the solubility of quartz.

Following our routine, we write for the reaction as written

$$\Delta_r G^\circ = \Delta_f G^\circ_{\mathrm{H_4SiO_4}} - \Delta_f G^\circ_{\mathrm{SiO_2}} - 2\,\Delta_f G^\circ_{\mathrm{H_2O}} \tag{9.13}$$

Then, getting numbers from the tables,

$$\begin{aligned} \Delta_r G^\circ &= -1307.7 - (-856.64) - 2(-237.129) \\ &= 23.198\,\mathrm{kJ\,mol^{-1}} \\ &= 23\,198\,\mathrm{J\,mol^{-1}} \end{aligned}$$

Then

$$\begin{aligned} \Delta_r G^\circ &= -RT\ln K \\ 23\,198 &= -(8.3145 \times 298.15)\ln K \end{aligned}$$

so

$$\begin{aligned} \log K &= -23\,198/(2.302\,59 \times 8.3145 \times 298.15) \\ &= -4.064 \end{aligned}$$

Thus the molality of $\mathrm{SiO_2}$ in a solution in equilibrium with quartz is about $10^{-4.064}$, or about 5.2 ppm.[5]

---

[4] Dissolving minerals in water changes $a_{\mathrm{H_2O}}$ very little. So while strictly speaking $a_{\mathrm{H_2O}}$ is not 1.0 when saturated with some mineral, this assumption is usually quite good.

[5] If the molality of $\mathrm{H_4SiO_4}(aq)$ is x, then the molality of $\mathrm{SiO_2}(aq)$ is also x, as there is 1 mole of $\mathrm{SiO_2}$ in each.

Doing it backwards

So we see that Gibbs energies can sometimes be used to calculate a solubility. The same calculation also works in the other direction, that is, measuring a solubility can be used to calculate a value of $\Delta_r G^\circ$. In the quartz–water case, the reaction is particularly simple, in that because the quartz and water are essentially pure phases, not only are a_{H_2O} and a_{SiO_2} equal to 1.0, but their values of $\Delta_f G^\circ$ are known, as shown above. Therefore, a value of $\Delta_r G^\circ$ calculated from a solubility measurement can be used to calculate $\Delta_f G^\circ$ for aqueous silica. Thus if you measured the solubility of quartz to be 5.2 ppm at 25 °C, you could use (9.13) in the form

$$\Delta_f G^\circ_{H_4SiO_4} = \Delta_f G^\circ_{SiO_2} + 2\,\Delta_f G^\circ_{H_2O} + \Delta_r G^\circ \tag{9.14}$$

to calculate $\Delta_f G^\circ_{H_4SiO_4} = -1307.7\,\mathrm{kJ\,mol^{-1}}$, and as a matter of fact that is usually how this quantity is determined.

A strange procedure

Note the strangeness of what we are doing here. On the left-hand side of $\Delta_r G^\circ = -RT\ln K$ (Equation 9.11) we enter the standard Gibbs energies of the reactants and products, which in this case includes $\Delta_f G^\circ$ of H_4SiO_4 at a concentration of one molal (its concentration in its standard state) in a hypothetical ideal solution, and on the right-hand side calculated its equilibrium concentration, only a few ppm. Remember what we said in deriving the equilibrium constant–the left-hand side consists of tabulated reference state data; it has nothing to do with real systems or with equilibrium. But from these data, equilibrium activity ratios and sometimes compositions can be calculated. Think about it.

9.4.2 *K* equal to fugacity of a volatile species

Hematite–magnetite example

The next example is the same in principle. Consider the reaction[6]

$$6\,Fe_2O_3(s) = 4\,Fe_3O_4(s) + O_2(g) \tag{9.15}$$

for which the equilibrium constant is

$$K = \frac{a^4_{Fe_3O_4}\,a_{O_2}}{a^6_{Fe_2O_3}}$$

---

[6] It is important in this reaction to note that we write oxygen as $O_2(g)$, that is, oxygen gas. There are also data for dissolved oxygen, written $O_2(aq)$ which are of course completely different. The same ambiguity does not exist for hematite and magnetite, but it is always a good idea to append the (g), (aq), (s), or (l) symbols for clarity.

If the reaction involves pure hematite Fe_2O_3 and pure magnetite Fe_3O_4, then $x_{Fe_2O_3} = 1$ and $x_{Fe_3O_4} = 1$, so $a_{Fe_2O_3} = 1$ and $a_{Fe_3O_4} = 1$. Therefore

$$\begin{aligned} K &= a_{O_2} \\ &= f_{O_2} \\ &= P_{O_2}\gamma_f \end{aligned}$$

Assuming that the activity coefficient γ_f is 1.0, again in this case an excellent approximation, we can calculate the partial pressure of oxygen in a gas phase in equilibrium with the two minerals hematite and magnetite.

Following the routine, we write

$$\begin{aligned} \Delta_r G^\circ &= 4\,\Delta_f G^\circ_{Fe_3O_4} + \Delta_f G^\circ_{O_2} - 6\,\Delta_f G^\circ_{Fe_2O_3} \\ &= 4(-1015.4) + 0 - 6(-742.2) \\ &= 391.6\,\mathrm{kJ\,mol^{-1}} \\ &= 391600\,\mathrm{J\,mol^{-1}} \end{aligned}$$

$$\begin{aligned} \Delta_r G^\circ &= -RT\ln K \\ 391\,600 &= -(8.3145 \times 298.15)\ln K \end{aligned}$$

$$\begin{aligned} \log K &= -391\,600/(2.302\,59 \times 8.3145 \times 298.15) \\ &= -68.40 \end{aligned}$$

So the oxygen fugacity in equilibrium with hematite and magnetite at 25 °C and 1 bar is $10^{-68.40}$ bar. This is an incredibly small quantity, which would have absolutely no significance if it were simply a partial pressure, unconnected to thermodynamics. A partial pressure of this magnitude would be produced by one molecule of oxygen in a volume larger than that of a sphere with a diameter of the solar system (§12.12). However, it is in fact a parameter in the thermodynamic model, just as valid as any other part of the model. It can be used, for example, to calculate other parameters that might be more easily measurable. For example, the reaction

$$CH_4(g) + O_2(g) = CO_2(g) + 2\,H_2(g) \tag{9.16}$$

is one that you might be interested in if you were studying the bottom muds in Figure 2.1c. The equilibrium constant for this is

$$\begin{aligned} \Delta_r G^\circ &= -394.359 + 2(0) - (-50.72) - 0 \\ &= -343.639\,\mathrm{kJ\,mol^{-1}} \\ &= -343\,639\,\mathrm{J\,mol^{-1}} \end{aligned}$$

$$\begin{aligned} \log K &= 343\,639/(2.302\,59 \times 8.3145 \times 298.15) \\ &= 60.203 \end{aligned}$$

which means that at equilibrium,

$$\frac{f_{CO_2} \cdot f_{H_2}^2}{f_{CH_4} \cdot f_{O_2}} = 10^{60.203}$$

Now $10^{60.203}$ is just as ridiculous as $10^{-68.40}$ in a sense. But if we insert the value $f_{O_2} = 10^{-68.40}$ into this expression, we get

$$\begin{aligned} \frac{f_{CO_2}}{f_{CH_4}} f_{H_2}^2 &= 10^{60.203} \cdot 10^{-68.40} \\ &= 10^{-8.20} \end{aligned}$$

which begins to look a little more reasonable. This tells you something about how the CO_2/CH_4 ratio varies with f_{H_2}. For example, you could say that according to the thermodynamic model, if f_{O_2} is controlled by hematite–magnetite, the CO_2 and CH_4 fugacities (partial pressures) are equal when f_{H_2} is $10^{-4.1}$ bar, and this might in fact be a measurable quantity in the muds.

The point is that by writing a few reactions and using thermodynamics, your thoughts about what might be happening in the bottom muds or any other environment take shape in a controlled fashion – controlled, that is, by the implied hypothesis of chemical equilibrium. Your system may not be at complete equilibrium, but your model is, because that is a good place to start. And the fact that one of your thermodynamic parameters, such as f_{O_2}, turns out to be impossibly small or large does not make it ridiculous; it just means you won't be able to measure it directly, and you might want to concentrate on other parameters to which your impossible one is connected by the model.

Muscovite example

The hematite–magnetite example is just one of a great variety of geologically important reactions in which the fugacity of one species is numerically equal to an equilibrium constant. That one species can be O_2 as in the example, but it can also be another species, typically H_2O, CO_2, or H_2. For example, the assemblage muscovite plus quartz reacts at high temperatures to andalusite plus K-feldspar in the reaction

$$KAl_3Si_3O_{10}(OH)_2(s) + SiO_2(s) = Al_2SiO_5(s) + KAlSi_3O_8(s) + H_2O(g) \tag{9.17}$$

for which the equilibrium constant is

$$\begin{aligned} K &= \frac{a_{Al_2SiO_5} \cdot a_{KAlSi_3O_8} \cdot a_{H_2O}}{a_{KAl_3Si_3O_{10}(OH)_2} \cdot a_{SiO_2}} \\ &= a_{H_2O} \quad \text{(minerals are pure so } a = 1\text{)} \\ &= \frac{f_{H_2O}}{f^\circ_{H_2O}} \\ &= f_{H_2O} \quad \text{(ideal gas std. state at } P = 1 \text{ so } f^\circ = 1\,\text{bar)} \end{aligned}$$

so that as long as the minerals are all quite pure and are only very slightly soluble so that the water is also quite pure, K is numerically equal to the fugacity of pure water at the specified T and P. These conditions are met in experimental work (though perhaps not in nature) so that determination of the equilibrium T and P for the reaction allows determination of K, because the fugacity of pure water is well known (Figure 6.1).

The equilibrium diagram for this reaction (calculated by SUPCRT92) is shown in Figure 9.4. The water fugacity at 2000 bars, 599.75 °C (indicated by the square) is 1052 bars, so $K = 1052$ at this point. However, be sure to note the assumptions we have made in saying $K = 1052$. They are:

1. The standard state for all minerals is the pure mineral at $T = 599.75\,°\mathrm{C}$ and 2000 bars.
2. The standard state for water is ideal gaseous H_2O at $T = 599.75\,°\mathrm{C}$ and a pressure of one bar.

These are the common choices in geochemical calculations, but of course others are possible. Every value of an equilibrium constant implies that standard state choices have been made.

Humidity buffer example

Another example which is the same in principle though different in experimental practice and in application, is the case of the metal sulfate minerals. Sulfate minerals such as melanterite ($FeSO_4 \cdot 7H_2O$), rozenite ($FeSO_4 \cdot 4H_2O$), calcanthite ($CuSO_4 \cdot 5H_2O$), bonattite ($CuSO_4 \cdot 3H_2O$) and many others, are

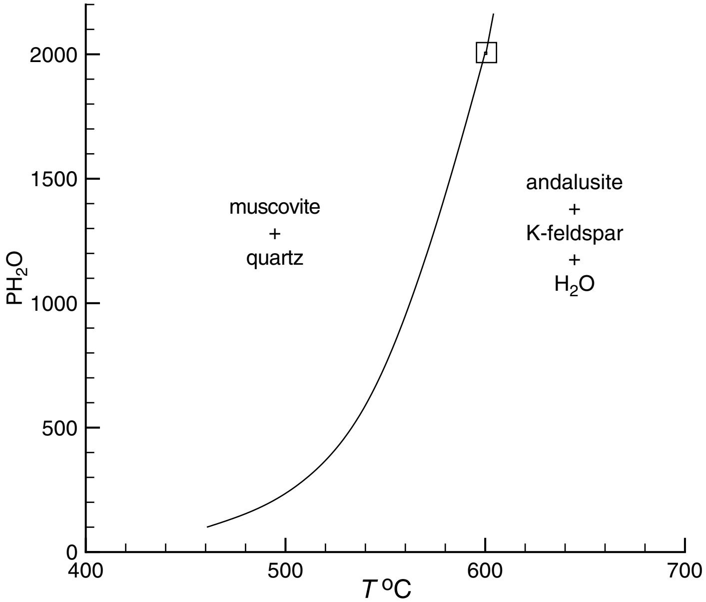


Figure 9.4 Equilibrium pressures and temperatures for reaction (9.17) as calculated by SUPCRT92. The square indicates the equilibrium state at 599.75 °C, 2000 bars.

Relative humidity and water activity

The relative humidity (RH) is defined as

$$\mathrm{RH} = \frac{\text{water vapor pressure, or } P_{\mathrm{H_2O}}}{\text{vapor pressure of pure liquid water, or } P^\circ_{\mathrm{H_2O}}}$$

This is usually reported as a percentage by multiplying by 100, i.e., %RH = RH × 100. Because the water vapor pressure and the water fugacity are almost identical (Table 9.1), we can write

$$\begin{aligned} a_{\mathrm{H_2O}} &= \frac{f_{\mathrm{H_2O}}}{f^\circ_{\mathrm{H_2O}}} \quad \text{(std. state pure liquid water)} \\ &= \frac{P_{\mathrm{H_2O}}}{P^\circ_{\mathrm{H_2O}}} \\ &= \mathrm{RH} \end{aligned}$$

where $f^\circ_{\mathrm{H_2O}}$ is the fugacity of pure water. So measurements of RH are one form of water activity. Alternatively, we can use ideal gaseous H_2O at $P = 1$ bar standard state, giving

$$\begin{aligned} a_{\mathrm{H_2O}} &= \frac{f_{\mathrm{H_2O}}}{1} \quad \text{(std. state ideal gas H}_2\text{O at 1 bar)} \\ &= \mathrm{RH} \times f^\circ_{\mathrm{H_2O}} \end{aligned}$$

Both standard states are used in this type of work (Jerz and Rimstidt, 2003). They give totally different values of $\ln K$ for reactions (Table 9.1).

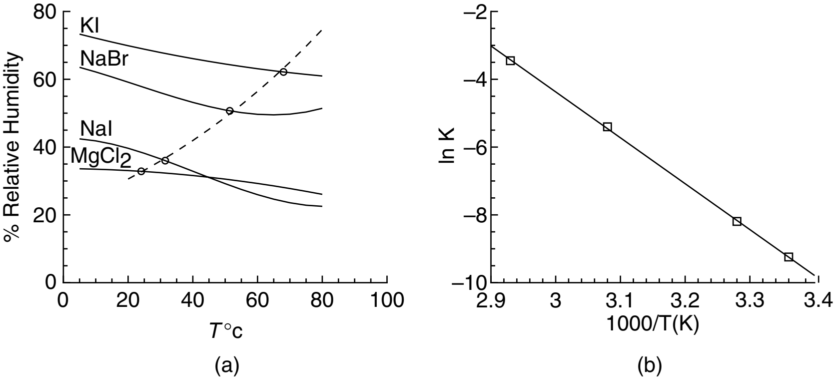
common in mines, on mine tailings, and on sulfide minerals exposed to air. When these phases dissolve they create very acid solutions. They store metals and sulfate during dry periods and dissolve readily during flushing events, causing sudden increases in the acidity of drainage from abandoned mine sites with attendant environmental damage (Nordstrom and Alpers, 1999). It is therefore important to have thermodynamic data for them in order to understand acid mine drainage situations.

The pressure of interest for reactions between these minerals is naturally atmospheric pressure (1 bar) and the temperature of interest ranges from ambient to the somewhat elevated temperatures that would occur in mine waste dumps (oxidation of sulfides is exothermic). Many of these hydrated sulfates are stable only at water fugacities less than that of liquid water (they are not stable in water), so we need a way of controlling f_{H_2O} in the range of, say, 25–100 °C. A convenient way to do this is with the humidity buffer technique.

Consider, for example, the calcanthite–bonattite reaction, which is

$$CuSO_4 \cdot 5H_2O(s) = CuSO_4 \cdot 3H_2O(s) + 2\,H_2O(g) \qquad (9.18)$$

To determine the equilibrium constant for this reaction, all we need to do is determine the f_{H_2O} (or P_{H_2O}) at which both minerals are stable, at some fixed *T* and *P*. This could be done by actually measuring the vapor pressure in a closed space containing the two minerals by some suitable method, and this can be done. A more convenient and very accurate method is to connect the space containing the two minerals with another space containing a saturated salt solution. A saturated salt solution has a fixed vapor pressure at any temperature (a consequence of the phase rule, Chapter 11), and these vapor pressures (fugacities) have been compiled for many salt solutions by Greenspan (1977). Those for four salts are shown in Figure 9.5a. There is then a fixed f_{H_2O} above the two minerals. One of them will be unstable at that f_{H_2O} so the reaction between them will begin to proceed one way or the other, and the container with the minerals will then either gain or lose weight, depending on which way the reaction proceeds. By varying the temperature, a temperature at which there is no weight change can be found, and the fugacity of the saturated salt solution at that temperature will be the equilibrium fugacity for the mineral reaction. In the sense that two or more containers reach equilibrium by transfer of water through the vapor phase, the method is similar in principle to the isopiestic method mentioned in §5.13.



(a) (b)

Figure 9.5 (a) The % relative humidity established by saturated solutions of four different salts as a function of temperature. The circles indicate the equilibrium temperature for reaction (9.18) at the indicated % RH. (b) The % RH values from (a) converted to ln*K* and plotted versus the inverse absolute equilibrium temperature. From this plot, values of $\Delta_r G°$ and $\Delta_r H°$ for reaction (9.18) may be obtained. The line is a least-squares fit to the data, and is shown as the dashed line in (a).

Table 9.1 *Data for the calcanthite–bonattite reaction from Chou* et al. *(2002). Fugacity ($f^\circ_{H_2O}$) and vapor pressure ($P^\circ_{H_2O}$) of pure water are from IAPWS-95 in the form of program* STEAM *(see §13.6.1).*

	Buffer			
	$MgCl_2 \cdot 6H_2O$	$NaI \cdot 2H_2O$	$NaBr \cdot 2H_2O$	KI
Equilibrium T °C	24.06	31.47	51.41	68.02
$f^\circ_{H_2O}$ (bars)	0.02994	0.04612	0.1320	0.2845
$P^\circ_{H_2O}$ (bars)	0.02997	0.04619	0.1324	0.2862
% RH	32.85	36.00	50.69	62.16
$\ln K$ (a_{H_2O} std. state liquid water)	−2.226	−2.043	−1.359	−0.951
$\ln K$ (a_{H_2O} std. state ideal gas)	−9.242	−8.193	−5.402	−3.453

Data for the calcanthite–bonattite reaction from Chou et al. (2002) are shown in Table 9.1 and Figure 9.5.[7]

9.5 *K* in solid–solid reactions

It should be evident by now that the equilibrium constant is most useful in reactions between dissolved substances, those that change their activities during the reaction. Reactions of the other kind, between pure substances that do not change their activities during the reaction [e.g., reaction (9.1)] have no need of an equilibrium constant because in general they do not reach an equilibrium; they proceed until one of the reactants disappears. But what happens if you *do* calculate K for such a reaction–what does it mean? Let's do this for reaction (9.1) and see what happens.

$$\begin{aligned}\Delta_r G^\circ &= \Delta_f G^\circ_{NaAlSi_3O_8} - 2\,\Delta_f G^\circ_{SiO_2(s)} - \Delta_f G^\circ_{NaAlSiO_4}\\ &= -3711.5 - 2(-856.64) - (-1978.1)\\ &= -20.12\,\mathrm{kJ\,mol^{-1}}\\ &= -20\,120\,\mathrm{J\,mol^{-1}}\end{aligned}$$

---

[7] The values of $f^\circ_{H_2O}$ and $P^\circ_{H_2O}$ in Table 9.1 indicate that they are very nearly equal. However for the purists, we point out that these two quantities do not refer to the exact same conditions. $f^\circ_{H_2O}$ is the fugacity of water at T (which is $< 100\,°\mathrm{C}$) and a pressure of 1 bar. At these conditions, water has no vapor phase, and so cannot have a vapor pressure. The values of P_{H_2O} refer to the vapor pressure of water at T under its own vapor pressure. The difference between the vapor pressure and 1 bar is small but not totally insignificant. For example, at $T = 68.02\,°\mathrm{C}$, the fugacity of water at the vapor pressure of water (0.2862 bars) is 0.2843 bars, while at a pressure of 1 bar, it is 0.2845 bars.

$$\log K = 20\,120/(2.302\,59 \times 8.3145 \times 298.15)$$

$$= 3.52$$

$$K = 3349$$

As usual, this means that *at equilibrium*,

$$\frac{a_{\mathrm{NaAlSi_3O_8}}}{a_{\mathrm{NaAlSiO_4}}\,a^2_{\mathrm{SiO_2}}} = 10^{3.52}$$

If we in fact have pure nepheline, pure albite and pure quartz involved in the reaction, then we come up with the same answer as before. The activities of $NaAlSi_3O_8$, $NaAlSiO_4$, and SiO_2 are 1.0 by our definitions, so the ratio $a_{\mathrm{NaAlSi_3O_8}}/(a_{\mathrm{NaAlSiO_4}}\,a^2_{\mathrm{SiO_2}})$ is fixed at 1.0 and can never be equal to 3349 – the three pure minerals can never reach equilibrium at 25 °C, 1 bar. But suppose the minerals are not pure – suppose they are solid solutions. Albite ($NaAlSi_3O_8$) forms a solid solution with anorthite ($CaAl_2Si_2O_8$) in the mineral plagioclase, and nepheline also usually occurs in a solid solution with kalsilite ($KAlSiO_4$), so the mole fractions and hence the activities of both $NaAlSi_3O_8$ and $NaAlSiO_4$ will generally be less than 1.0, even though there are pure minerals with these compositions. The activity is less than 1.0 when the minerals are *not* pure, but occur as a components of solid solutions.

In this particular case, having $a_{\mathrm{NaAlSi_3O_8}}$ less than 1.0 would not help to achieve equilibrium. Equilibrium could only be achieved by lowering $a_{\mathrm{NaAlSiO_4}}$ or $a_{\mathrm{SiO_2}}$. For example, in the presence of pure nepheline and pure albite, $a_{\mathrm{SiO_2}(s)}$ would have to be 0.0173 to achieve equilibrium. This of course could not happen if quartz was present, but $a_{\mathrm{SiO_2}}$ might be controlled in some other way, such as by the amount of dissolved SiO_2 in a solution that is undersaturated with quartz. To calculate this SiO_2 concentration, the reaction would be written using $SiO_2(aq)$ rather than $SiO_2(s)$ (see page 253).

There is an important lesson here. When we write a chemical reaction, we look up a value of $\Delta_f G^\circ$ for each chemical formulas. The values of $\Delta_f G^\circ$ are determined for those chemical species in very particular states – pure solids, ideal 1 molal solution, and so on. If our reaction is concerned with those species in those particular states, then the result is directly applicable to our problem – the value of $\Delta_r G$ is the same as the value of $\Delta_r G^\circ$, and the reaction accordingly will go or not go. This case basically arises only when dealing with pure solids. When dealing with solutions (solid, liquid, or gaseous), $\Delta_r G^\circ$ is only a starting point. The reacting species are never in their reference states and have values of Gibbs energy that add up to $\Delta_r G$, not $\Delta_r G^\circ$. The chemical formulas in our reactions represent species in some kind of solution, and we deal with these solutions with our activity terms, which are basically concentrations.

Reactions between solid phases such as (9.1) are in principle no different from any other kind of reactions, such as (9.2). The only difference is that there is in fact such a thing as relatively pure albite and quartz, and the like, to which the numbers in the tables apply directly, and we are sometimes interested in reactions between these pure compounds. In principle, however, each chemical formulas in a chemical reaction, whether $Mg_2SiO_4(s)$ or $HCO_3^-(aq)$, can and usually does occur in a solution of some kind, with an activity controlled by its concentration.

9.6 Change of *K* with temperature I

To get the effect of temperature on K, assuming as before (§6.5) that $\Delta_r H^\circ$ and $\Delta_r S^\circ$ are constants (not affected by temperature), we need only combine Equations (9.11) and (6.17),

$$\begin{aligned}\Delta_r G^\circ &= -RT\ln K\\ &= \Delta_r H^\circ_{298} - T\,\Delta_r S^\circ_{298}\end{aligned}$$

so

$$\ln K = \frac{-\Delta_r H^\circ_{298}}{RT} + \frac{\Delta_r S^\circ_{298}}{R} \tag{9.19}$$

or

$$\log K = \frac{-\Delta_r H^\circ_{298}}{2.302\,59\,RT} + \frac{\Delta_r S^\circ_{298}}{2.302\,59\,R} \tag{9.20}$$

As both $\Delta_r H^\circ$ and $\Delta_r S^\circ$ are assumed constant, this can be rewritten

$$\log K = a(1/T) + b$$

where a and b are constants, which is an equation in the form $y = ax + b$, meaning that $\log K$ is a linear function of $1/T$. An example of this is shown in Figure 9.6.

In Figure 9.6a are shown some solubility data for quartz, measured at a constant pressure of 1000 atm. As discussed in §9.4.1, these numbers can be

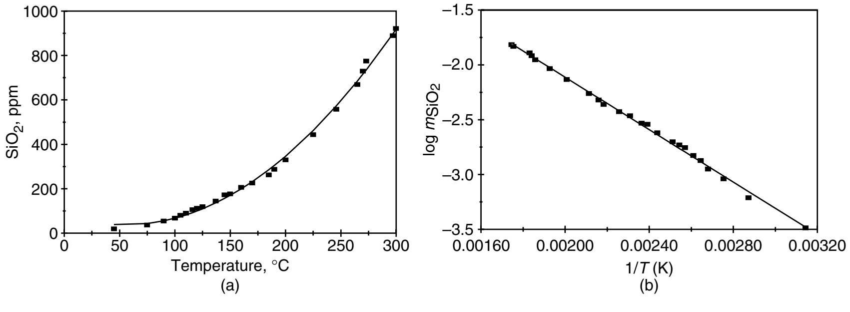


Figure 9.6 (a) The solubility of quartz in water as a function of temperature at a pressure of 1000 bar. (From Morey et al., 1962.) (b) The same data converted to $\log m_{SiO_2(aq)}$ and plotted versus the reciprocal of absolute temperature.

Calculation of aqueous silica

What is the silica content of a solution in equilibrium with plagioclase and nepheline solid solutions at 25 °C and 1 bar? The activities of the solid solution components are

$$a_{\mathrm{NaAlSiO_4}} = 0.75$$
$$a_{\mathrm{NaAlSi_3O_8}} = 0.5$$

The reaction is Equation (9.1) as before, *except* that we must use $\mathrm{SiO_2}(aq)$ rather than $\mathrm{SiO_2}(s)$. Thus we write

$$\mathrm{NaAlSiO_4}(s) + 2\,\mathrm{SiO_2}(aq) = \mathrm{NaAlSi_3O_8}(s)$$

Getting data from Appendix B,

$$\begin{aligned}\Delta_r G^\circ &= \Delta_f G^\circ_{\mathrm{NaAlSi_3O_8}} - 2\,\Delta_f G^\circ_{\mathrm{SiO_2}(aq)} - \Delta_f G^\circ_{\mathrm{NaAlSiO_4}}\\ &= -3711.5 - 2(-833.411) - (-1978.1)\\ &= -66.578\,\mathrm{kJ\,mol^{-1}}\end{aligned}$$

Then

$$\begin{aligned}\log K &= -(-66\,578)/(2.302\,59 \times 8.314\,5 \times 298.15)\\ &= 11.664\\ K &= 10^{11.664}\\ &= \frac{a_{\mathrm{NaAlSi_3O_8}}}{a_{\mathrm{NaAlSiO_4}}\,a^2_{\mathrm{SiO_2}(aq)}}\\ &= 0.5/(0.75 \times a^2_{\mathrm{SiO_2}(aq)})\\ a_{\mathrm{SiO_2}(aq)} &= [0.5/(10^{11.664} \times 0.75)]^{1/2}\\ &= 10^{-5.92}\end{aligned}$$

So assuming $\gamma_{\mathrm{SiO2}(aq)} = 1$, $m_{\mathrm{SiO2}(aq)}$ in equilibrium with these two solid solutions is $10^{-5.92}$ molal, slightly less than would be in equilibrium with the pure minerals.

- This is quite a hypothetical situation. Solid solutions with these activities are common, but based on experience, they would never equilibrate with aqueous silica at 25 °C. However at higher temperatures, such equilibria are common.
- We used $\mathrm{SiO_2}(aq)$ here, but $\mathrm{H_4SiO_4}$ in Equation (9.12). We discuss this in §9.10.

interpreted as values of the equilibrium constant for the quartz dissolution reaction. The same data plotted as $\log m_{\mathrm{SiO_2}(aq)}$ versus $1/T$, where T is in kelvins, are shown in Figure 9.6b. Obviously this shows a good linear correlation, indicating that $\Delta_r H^\circ$ does not change greatly over the temperature range of 25–300 °C.

9.6.1 Another example

As an example of the effect of T on K, as well as some of the other points we have made, consider the reaction

$$\mathrm{CaCO_3}(s) + \mathrm{SiO_2}(s) = \mathrm{CaSiO_3}(s) + \mathrm{CO_2}(g) \tag{9.21}$$

This is an important reaction at high temperatures, when granites intrude limestones at depth in the Earth, but we will consider it at low temperatures and 1 bar pressure.

log K versus $1/T$

A plot of $\log K$ versus $1/T$ can be used to obtain an estimate of $\Delta_r H^\circ$ for the reaction for which K is the equilibrium constant. According to the authors (Morey et al., 1962), the slope of the line in Figure 9.6b (fitted by the method of least squares) is -1180 K, and so from (8.14),

$$\frac{\Delta_r H^\circ}{2.30259\,R} = -(-1180)$$

and

$$\begin{aligned}\Delta_r H^\circ &= 1180 \times 2.302\,59 \times 8.3145\\ &= 22\,590\,\mathrm{J\,mol^{-1}}\\ &\approx\ 22.6\,\mathrm{kJ\,mol^{-1}}\end{aligned}$$

However, although the data may appear to be quite linear, confirming a constant $\Delta_r H^\circ$ and $\Delta_r S^\circ$, you must realize that a gentle curvature can easily be obscured by small random experimental errors, and even a gentle curvature implies a significant change in slope and of $\Delta_r H^\circ$. In this case, a theoretical treatment of these and other data (contained in program SUPCRT92, Johnson et al., 1992) shows that $\Delta_r H^\circ$ can vary (at 1 kbar) from 35.2 kJ mol^{-1} at 25 °C to 23.2 kJ mol^{-1} at 300 °C, while retaining an excellent fit to the data. It follows that exceptionally accurate values of K are needed to give accurate values of ΔH° and ΔS° by this method (Prue, 1969). The assumption of constant $\Delta_r H^\circ$ is not suitable for accurate work, but is often useful nonetheless.

First, we get the equilibrium constant, as usual,

$$\begin{aligned}
\Delta_r G^\circ &= \Delta_f G^\circ_{\mathrm{CaSiO_3}} + \Delta_f G^\circ_{\mathrm{CO_2}} - \Delta_f G^\circ_{\mathrm{CaCO_3}} - \Delta_f G^\circ_{\mathrm{SiO_2}} \\
&= -1549.66 + (-394.359) - (-1128.79) - (-856.64) \\
&= 41.411\,\mathrm{kJ\,mol^{-1}} \\
&= 41\,411\,\mathrm{J\,mol^{-1}}
\end{aligned}$$

$$\begin{aligned}
\Delta_r G^\circ &= -RT\ln K \\
41\,411 &= -(8.3145 \times 298.15)\ln K
\end{aligned}$$

$$\begin{aligned}
\log K_{298} &= -41\,411/(2.302\,59 \times 8.3145 \times 298.15) \\
&= -7.25
\end{aligned}$$

This means, as usual, that

$$\frac{a_{\mathrm{CaSiO_3}}\,a_{\mathrm{CO_2}}}{a_{\mathrm{CaCO_3}}\,a_{\mathrm{SiO_2}}} = 10^{-7.25}$$

and because all the solid phases are pure, their activities are all 1.0, and we write

$$\begin{aligned}
a_{\mathrm{CO_2}} &= 10^{-7.25} \\
&= f_{\mathrm{CO_2}}
\end{aligned}$$

Of course, both $CaSiO_3$ (wollastonite) and $CaCO_3$ (calcite) often form solid solutions and in natural situations might have activities less than 1.0, as discussed above. However, we are interested here in the pure phases.

The calculated $f_{\mathrm{CO_2}}$ of $10^{-7.25}$ can be thought of as meaning that if calcite, wollastonite, and quartz were at equilibrium with a gas phase having a pressure of 1 bar at 25 °C, the partial pressure of CO_2 in that gas would be about $10^{-7.25}$ or 5.6×10^{-8} bar. As long as the three minerals remain pure and at equilibrium, the equilibrium constant will continue to be equal to $f_{\mathrm{CO_2}}$, and so we can calculate the temperature at which the CO_2 pressure (fugacity) will reach 1 bar by calculating the change in K with T.

To do this, we will first get another expression for the effect of T on K that will be more convenient. From (8.13) you can see that the slope of the graph of $\ln K$ as a function of $1/T$ is $-\Delta_r H^\circ/R$, which is to say that at a temperature of 298 K,

$$\frac{d\ln K}{d(1/T)} = -\frac{\Delta_r H^\circ_{298}}{R}$$

Integrating this between 298 K and T, we get

$$\int_{298}^{T} d\ln K = -\frac{\Delta_r H^\circ_{298}}{R}\int_{298}^{T} d(1/T)$$

and so

$$\ln K_T - \ln K_{298} = -\frac{\Delta_r H^\circ_{298}}{R}\left(\frac{1}{T} - \frac{1}{298.15}\right)$$

or

$$\log K_T = \log K_{298} - \frac{\Delta_r H^\circ_{298}}{2.302\,59\,R}\left(\frac{1}{T} - \frac{1}{298.15}\right) \tag{9.22}$$

By substituting terms, you can easily show that these are equivalent to our previous equations, (9.19) and (9.20). Remember, they are valid only for constant $\Delta_r H^\circ$ and $\Delta_r S^\circ$.

Now we need $\Delta_r H^\circ$ for reaction (9.21). This is

$$\begin{aligned}\Delta_r H^\circ &= \Delta_f H^\circ_{\mathrm{CaSiO_3}} + \Delta_f H^\circ_{\mathrm{CO_2}} - \Delta_f H^\circ_{\mathrm{CaCO_3}} - \Delta_f H^\circ_{\mathrm{SiO_2}} \\ &= -1634.94 + (-393.509) - (-1206.92) - (-910.94) \\ &= 89.411\,\mathrm{kJ\,mol^{-1}}\end{aligned}$$

If we want to calculate the temperature T at which $f_{\mathrm{CO_2}}$ reaches 1 bar while in equilibrium with calcite, quartz, and wollastonite, then $K_T = 1$, $\log K_T = 0$, and using our value of $\Delta_r H^\circ$, we get

$$0 = -7.25 - \frac{89\,411}{2.302\,59 \times 8.3145}\left(\frac{1}{T} - \frac{1}{298.15}\right)$$

from which $T = 555$ K or about 282 °C. The meaning of these calculations is illustrated in Figure 9.7.

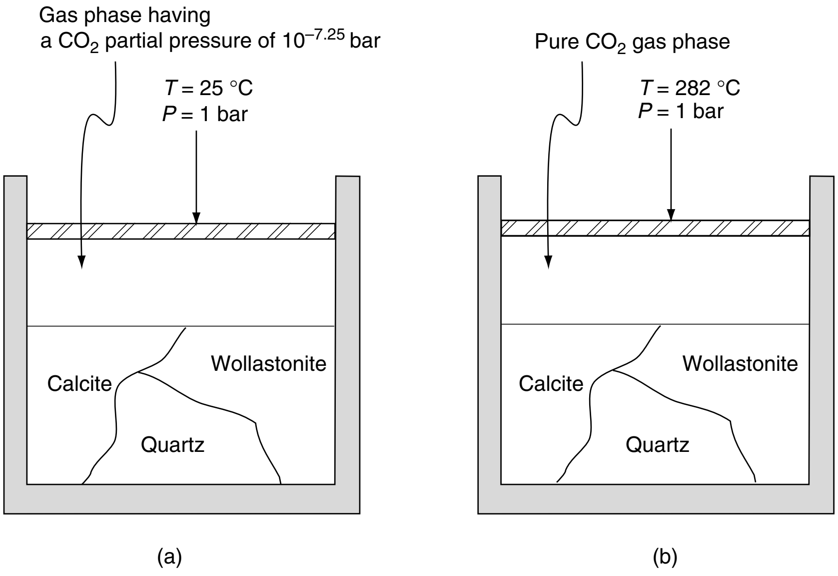


Figure 9.7 The fugacity of CO_2 in equilibrium with calcite, wollastonite, and quartz.

A common error

You must remember that you CANNOT calculate $\ln K$ at T from

$$\Delta_r G^\circ_{298} = -RT\ln K \qquad [9.11]$$

where $\Delta_r G^\circ$ comes from the normal tables, simply by changing T from 298.15 to some other value. In this Equation (9.11) , $\Delta_r G^\circ$ and K must refer to the same temperature. If you want K at some temperature other than 25 °C, *first* get $\Delta_r G^\circ$ at that new temperature from (6.17) or some other method, *then* get K from (9.11), using your new value of $\Delta_r G^\circ_T$ in place of $\Delta_r G^\circ_{298}$. Of course, this procedure has essentially been done for you in equations such as (9.20) and (9.22). An interesting exception to this rule is the case of some isocoulombic reactions (§9.7.2).

9.7 Change of *K* with temperature II

In Equation (9.11) we saw that $\ln K = -\Delta_r G^\circ/RT$, so that the effect of temperature on K evidently depends on the effect of temperature on $\Delta_r G^\circ$. In §9.6 we assumed a constant ΔH° and ΔS°, resulting in a simple linear relationship between $\log K$ and $1/T$. However, we also pointed out in §3.5.3 that

$$\frac{d\Delta H}{dT} = \Delta C_P \qquad [3.23]$$

or

$$\frac{d\Delta_r H^\circ}{dT} = \Delta_r C^\circ_P \qquad (9.23)$$

and

$$\frac{d\Delta_r S^\circ}{dT} = \frac{\Delta_r C^\circ_P}{T} \qquad (9.24)$$

In other words, what we have really assumed is that $\Delta_r C^\circ_P$ is zero, or that the heat capacities of reactants and products are equal. However, if you look at measured heat capacities, they look something like those shown in Figure 9.8. Pure solids, liquids, and gases generally have C°_P values that increase monotonically (constantly increasing) with T as illustrated by the mineral corundum. Over a temperature range of only 200 °C, the variation of C°_P may be fairly linear; over a larger range of T, it will show a distinct curvature. If the particular values of C°_P of products and reactants in the reaction of interest happen to be about the same and so cancel out, the $\Delta_r C^\circ_P = 0$ approximation works well, but realistically this only happens in fairly simple reactions such as calcite$\rightleftharpoons$aragonite, or by chance.

Furthermore, C°_P for aqueous species behaves quite differently, always showing a change in curvature as illustrated by HCl(*aq*) in Figure 9.8, so that reactions having both minerals and aqueous species have virtually no chance of having $\Delta_r C^\circ_P$ constant over a range of temperatures.

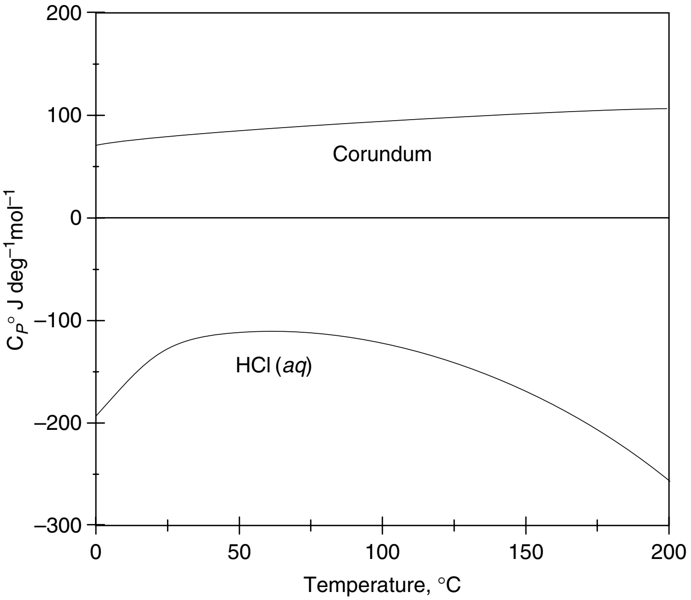


Figure 9.8 The heat capacities of corundum (Al_2O_3) and HCl(*aq*) as a function of temperature.

9.7.1 Nonaqueous reactions

In §3.5.3 we introduced both the Maier–Kelley and the Berman–Brown equations to describe the variation of heat capacity with temperature, and in Equation (5.30) showed an equation for the variation of $\Delta_r G^\circ$ with T, using the Maier–Kelley formulation. This equation is

$$\begin{aligned}\Delta_r G^\circ_T = {}& \Delta_r G^\circ_{T_\mathrm{r}} - \Delta_r S^\circ_{T_\mathrm{r}}(T - T_\mathrm{r}) \\ & + \Delta_r a\left[T - T_\mathrm{r} - T\ln\left(\frac{T}{T_\mathrm{r}}\right)\right] \\ & + \Delta_r \frac{b}{2}\left(2TT_\mathrm{r} - T^2 - T_r^2\right) \\ & + \frac{\Delta_r c\left(T^2 + T_\mathrm{r}^2 - 2TT_\mathrm{r}\right)}{2TT_\mathrm{r}^2}\end{aligned}$$

and substituting this into $\Delta_r G^\circ = -RT\ln K$ results in

$$\begin{aligned}\ln K_T = {}& \ln K_{T_\mathrm{r}} - \frac{\Delta_r H^\circ}{R}\left(\frac{1}{T} - \frac{1}{T_\mathrm{r}}\right) + \frac{\Delta_r a}{R}\left(\ln\frac{T}{T_\mathrm{r}} + \frac{T_\mathrm{r}}{T} - 1\right) \\ & + \frac{\Delta_r b}{2R}\left(T + \frac{T_\mathrm{r}^2}{T} - 2T_\mathrm{r}\right) \\ & + \frac{\Delta_r c}{R}\,\frac{\left(-T^2 - T_\mathrm{r}^2 + 2TT_\mathrm{r}\right)}{2T^2T_\mathrm{r}^2}\end{aligned} \tag{9.25}$$

This rather long equation is simple to use if the Maier–Kelley a, b, and c values are available for all products and reactants, and gives accurate results.

9.7.2 Aqueous reactions

However, if your reactions contain aqueous species, the Maier–Kelley formulation does *not* apply, and a, b, and c values are not available because they don't work. The C_P° of HCl(aq) shown in Figure 9.8 is fairly typical of aqueous species, which are convex upward and have a maximum somewhere around 50–100 °C (see also Figures 10.12; 15.10). If the aqueous and mineral species in your reaction are all in the SUPCRT92 database, you should use that program, which incorporates the HKF model, introduced in §13.6.2, and discussed more fully in Chapter 15. Failing that you must estimate your equilibrium constant as a function of T and P, and there are only a couple of ways of doing that.

Isocoulombic reactions

Generally speaking, to model the change in $\log K$ with T you must account for the change in C_P° for each species in the reaction individually, especially for aqueous species. However, it is sometimes possible to write the reaction such that equal numbers of ions appear on each side of the reaction or, failing that, the same total charge on each side, in which case a great deal of the variation will cancel out. Such reactions are called "isocoulombic" (Mesmer and Baes, 1974; Lindsay, 1980).

It is understood that the isocoulombic reaction has no solids or gases, because their variation of C_P° with T is quite different, as discussed above. The transformation from the reaction you are interested in to an isocoulombic reaction is made by combining your reaction with another one, which is often the ionization of water reaction. For example,

$$\mathrm{H_3PO_4}(aq) = \mathrm{H^+} + \mathrm{H_2PO_4^-}$$

is a typical ionization reaction with all charges appearing on the right side. The equilibrium constant for this reaction shows a considerable curvature, plotted as $\log K$ versus $1/T$. Add to this the reaction

$$\mathrm{OH^-} + \mathrm{H^+} = \mathrm{H_2O}(l)$$

and we have

$$\mathrm{H_3PO_4}(aq) + \mathrm{OH^-} = \mathrm{H_2PO_4^-} + \mathrm{H_2O}(l)$$

which has the same charge on both sides, and $\log K$ versus $1/T$ is rather close to a straight line. Of course, to recover the desired ionization constant at some higher temperature, you need to know the ionization constant for water as a function of T in order to be able to "uncombine" it at the higher T. $\log K$ as a function of $1/T$ for isocoulombic reactions can be surprisingly accurate in many cases.

Gu et al. (1994) have extended this method by noticing that the ΔC_P° and ΔS° terms for many isocoulombic reactions are not only small, but are often of opposite sign, and tend to cancel one another. Omitting the ΔS° term leads to what they call the one-term extrapolation method. What this amounts to is that for well-balanced aqueous reactions, $\Delta_r G^\circ$ is independent of temperature, so that Equation (9.11) becomes

$$\Delta_r G^\circ_T = \Delta_r G^\circ_{T_\mathrm{r}} = -RT \ln K \tag{9.26}$$

so that only the Gibbs energy change of a reaction at 25 °C is sufficient to calculate the value at any other temperature. A discussion of this method with examples is given by Wood and Samson (1998).

The density model

Franck (1956, 1961) observed that the ionization constants of water and of many aqueous solutes at elevated temperatures and pressures showed a remarkably linear behavior when plotted as $\log K$ versus $\log \rho$ over wide ranges of T and P, where K is the equilibrium constant and ρ is the density of pure water. Since then this relationship and variations of it have been used in many studies (several references to these are in Anderson et al., 1991). Marshall and Franck (1981) used the expression

$$\log K = a + b/T + c/T^2 + d/T^3 + (e + f/T + g/T^2)\log \rho \tag{9.27}$$

to represent the ionization constant of water to 1000 °C with much success. Mesmer (1985, 1986) then showed that when this is simplified to include only the constants a, b and f, these parameters take on values fixed by the properties of water at some reference condition, and in Mesmer et al. (1987, 1988, 1989) that the resulting equation is quite successful in predicting solute properties to about 300 °C. Anderson et al. (1991) give a more complete description, and include the data required to use the model equation.

Using only the a, b, and f terms, Equation (9.27) can be rewritten

$$\ln K = p_1 + \frac{p_2}{T} + \frac{p_3 \ln \rho}{T} \tag{9.28}$$

where p_1, p_2, and p_3 are constants and ρ is the density of water at T and P.

Because $\ln K$ is proportional to $\Delta_r G^\circ$ (Equation 9.11) and $\ln \rho$ is specified at specific values of T and P, it follows that an expression relating $\ln K$ and $\ln \rho$ is logically equivalent to one giving $\Delta_r G^\circ$ as a function of T and P. In Chapter 4 we saw that an equation giving $\Delta_r G^\circ$ as a function of T and P is called a *fundamental equation*, and that it implicitly contains information on the variation of *all* thermodynamic parameters with T and P. Therefore there are implicit relationships between the parameters in both Equations (9.27) and (9.28) and all other thermodynamic parameters. These are given by Gates et al. (1982) for Equation (9.27).

Differentiation of Equation (9.28) with respect to temperature yields

$$\Delta H^\circ = -R[p_2 + p_3(T\alpha + \ln \rho)] \tag{9.29}$$

where α is the coefficient of thermal expansion of H_2O (and V is the molar volume),

$$\alpha = (1/V)(\partial V/\partial T)_P = (\partial \ln V/\partial T)_P = -(\partial \ln \rho/\partial T)_P$$

and ΔH° is the standard enthalpy of reaction in a chosen reference state. Differentiating Equation (9.29) with respect to T at constant P gives

$$\Delta C_P^\circ = -RTp_3(\partial\alpha/\partial T)_P, \quad \text{or}$$

$$p_3 = \frac{-\Delta C_P^\circ}{RT(\partial\alpha/\partial T)_P} \tag{9.30}$$

showing that Equation (9.28) implies that the quantity $\Delta C_P^\circ/RT(\partial\alpha/\partial T)_P$ is a constant. This means that ΔC_P° must mimic $(\partial\alpha/\partial T)_P$. The value of p_3 can be determined by choosing values for T and P, that is, a reference state. Denoting the reference state by subscript r, then

$$\Delta C_P^\circ/RT(\partial\alpha/\partial T)_P = \Delta C_{P_\mathrm{r}}^\circ/RT_\mathrm{r}(\partial\alpha/\partial T)_{P_\mathrm{r}}$$

so that at any T and P, Equation (9.28) implies that

$$\Delta C_P^\circ = \Delta C_{P_\mathrm{r}}^\circ \cdot \frac{T(\partial\alpha/\partial T)_P}{T_\mathrm{r}(\partial\alpha/\partial T)_{P_\mathrm{r}}} \tag{9.31}$$

Similarly, differentiation of Equation (9.28) with respect to pressure gives

$$\begin{aligned}\Delta V^\circ &= -p_3R\beta \\ &= \frac{\Delta C_{P_\mathrm{r}}^\circ \beta}{T_\mathrm{r}(\partial\alpha/\partial T)_{P_\mathrm{r}}}\end{aligned} \tag{9.32}$$

where β is the compressibility coefficient of water,

$$\beta = (1/V)(\partial V/\partial P)_T = (\partial \ln V/\partial P)_T = -(\partial \ln \rho/\partial P)_T$$

Inserting this expression for the heat capacity into the standard expressions for enthalpy end entropy, Equations (3.31) and (5.21) gives expressions for ΔH°ãnd ΔS°, and combining these as $\Delta H^\circ - T\Delta S^\circ$ gives ΔG° and from this $\log K$. Comparing the expression for $\log K$ with Equation (9.28) gives expressions for p_1 and p_2.

These and other relationships that follow from the model are summarized in the box on page 263. This empirical formulas apparently works well because $(\partial\alpha/\partial T)$ as a function of temperature for water is U-shaped, which gives to the expression for $\Delta_r C_P^\circ$ (if ΔC_P° is negative) an inverted-U shape with a maximum around 100 °C, which is the same shape that the C_P° of many aqueous ions have (see Chapter 15). Furthermore the variation of ΔC_P° at higher pressures is also fairly faithfully modeled by the expression for ΔC_P°, and the equations fit the

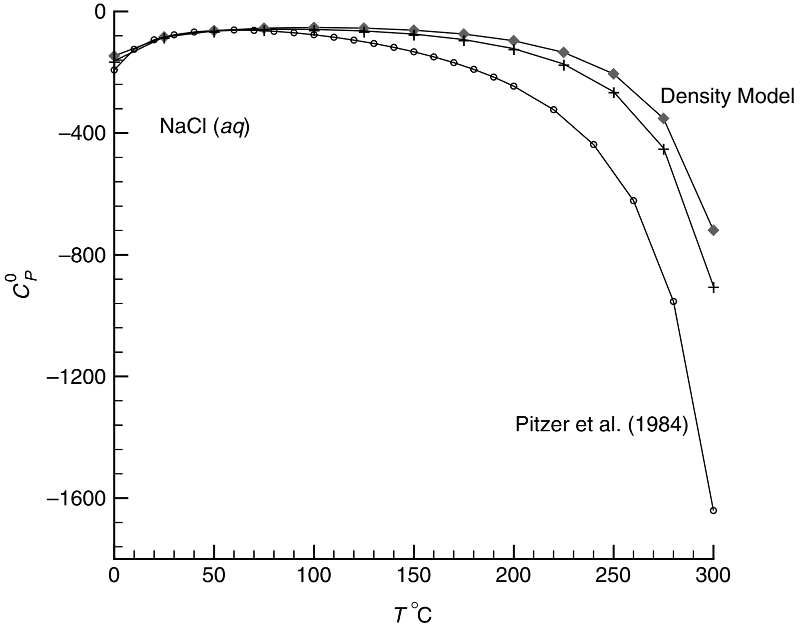


Figure 9.9 The standard (infinite dilution) heat capacity of aqueous NaCl. Data from Pitzer et al. (1984) (lower curve). Upper two curves are from the density model: crosses – data from Anderson et al. (1991); diamonds – data from program STEAM, Harvey et al. (2000). The Pitzer et al. data are also shown in Figure 10.12.

C_P° of individual electrolytes just as well as the ΔC_P° of reactions. A comparison of heat capacities for NaCl(*aq*) is shown in Figure 9.9.

To use the equation, one needs only the values of $\ln K$, ΔH°, and ΔC_P° for the reaction at the reference conditions (which will frequently be 25 °C, 1 bar, but could easily be some other conditions in cases where experimental data at high temperatures or pressures are involved), as well as the density of the solvent, water, at the desired P, T conditions. To obtain estimates of $\Delta_r V^\circ$ and $\Delta_r C_P^\circ$ for the reaction at T, P, values of α and β for the solvent are also required. These data and several other examples and additional details are given by Anderson et al. (1991). However, note that data for $(\partial\alpha/\partial T)_P$ are now more easily available from the NIST program STEAM (Harvey et al., 2000). Program STEAM does not provide values of $(\partial\alpha/\partial T)_P$ directly, but does give values of $(\partial^2\rho/\partial T^2)_P$, which can be modified to $(\partial\alpha/\partial T)_P$ by multiplying by $-V$. That is,

$$\begin{aligned}\left(\frac{\partial\alpha}{\partial T}\right)_P &= -\left(\frac{\partial^2 \ln\rho}{\partial T^2}\right)_P\\ &= -\frac{1}{\rho}\left(\frac{\partial^2\rho}{\partial T^2}\right)_P\\ &= -V\left(\frac{\partial^2\rho}{\partial T^2}\right)_P\end{aligned}$$

The density model equations

$$p_1 = \ln K_\mathrm{r} + \frac{\Delta H^\circ_\mathrm{r}}{RT_\mathrm{r}} - \frac{\Delta C^\circ_{P_\mathrm{r}} \cdot \alpha_\mathrm{r}}{RT_\mathrm{r}(\partial\alpha/\partial T)_{P_\mathrm{r}}}$$

$$p_2 = -\frac{\Delta H^\circ_\mathrm{r}}{R} + \frac{(T_\mathrm{r}\alpha_\mathrm{r} + \ln\rho_\mathrm{r})\Delta C^\circ_{P_\mathrm{r}}}{RT_\mathrm{r}(\partial\alpha/\partial T)_{P_\mathrm{r}}}$$

$$p_3 = \frac{-\Delta C^\circ_{P_\mathrm{r}}}{RT_\mathrm{r}(\partial\alpha/\partial T)_{P_\mathrm{r}}}$$

$$\begin{aligned}
\Delta H^\circ &= -R\left[p_2 + p_3(T\alpha + \ln\rho)\right] \\
&= \Delta H^\circ_\mathrm{r} + \frac{\Delta C^\circ_{P_\mathrm{r}}}{T_\mathrm{r}(\partial\alpha/\partial T)_{P_\mathrm{r}}}\left(T\alpha - T_\mathrm{r}\alpha_\mathrm{r} + \ln\frac{\rho}{\rho_\mathrm{r}}\right)
\end{aligned}$$

$$\begin{aligned}
\Delta S^\circ &= R(p_1 - p_3\alpha) \\
&= \Delta S^\circ_r + \frac{\Delta C^\circ_{P_\mathrm{r}}}{T_\mathrm{r}(\partial\alpha/\partial T)_{P_\mathrm{r}}}(\alpha - \alpha_\mathrm{r})
\end{aligned}$$

$$\begin{aligned}
\Delta V^\circ &= -p_3 R\beta \\
&= \frac{\beta\Delta C^\circ_{P_\mathrm{r}}}{T_\mathrm{r}(\partial\alpha/\partial T)_{P_\mathrm{r}}}
\end{aligned}$$

$$\begin{aligned}
\Delta G^\circ &= -R(p_1 T + p_2 + p_3\ln\rho) \\
&= \Delta G^\circ_r + \Delta H^\circ_\mathrm{r}\left(1 - \frac{T}{T_\mathrm{r}}\right) + \frac{\Delta C^\circ_{P_\mathrm{r}}}{T_\mathrm{r}(\partial\alpha/\partial T)_{P_\mathrm{r}}}\left(\alpha_\mathrm{r}(T - T_\mathrm{r}) + \ln\frac{\rho}{\rho_\mathrm{r}}\right) \qquad (9.33)
\end{aligned}$$

$$\ln K = \ln K_\mathrm{r} - \frac{\Delta H^\circ_\mathrm{r}}{R}\left(\frac{1}{T} - \frac{1}{T_\mathrm{r}}\right) + \frac{\Delta C^\circ_{P_\mathrm{r}}}{RT_\mathrm{r}(\partial\alpha/\partial T)_{P_\mathrm{r}}}\left(\frac{1}{T}\ln\frac{\rho_\mathrm{r}}{\rho} - \frac{\alpha_\mathrm{r}}{T}(T - T_\mathrm{r})\right) \qquad (9.34)$$

The data in Anderson et al. (1991) are based on IAPS-84, while the NIST data are based on IAPWS-95, discussed in §13.6.1. A comparison of the two sources of data for saturation pressures is shown in Figure 9.9.

For reactions for which $\log K$, ΔH°, and ΔC°_P at 25 °C are available but little else, the density model is one of the best ways of obtaining estimates of $\log K$ and other parameters at higher temperatures and pressures.

9.7.3 Combined reactions

Obviously both the isocoulombic and the density model methods are suitable only for reactions having only aqueous species, because solids, liquids and gases have heat capacities with a quite different temperature variation. For reactions having both pure phases and aqueous species, the heat capacity expressions can be simply combined.

For reactions in which the aqueous species are "isocoulombic" but minerals or other phases are also present, one simply uses the Maier–Kelley expression for $\log K$, Equation (9.25), whether or not the pure phases in the reaction are compositionally balanced. For the density model, two expressions for the variation of $\log K$ are combined, giving

$$\begin{aligned}\ln K_T = {} & \ln K_{T_\mathrm{r}} - \frac{\Delta_r H^\circ}{R}\left(\frac{1}{T}-\frac{1}{T_\mathrm{r}}\right) + \frac{\Delta_r a}{R}\left(\ln\frac{T}{T_\mathrm{r}}+\frac{T_\mathrm{r}}{T}-1\right)\\ & +\frac{\Delta_r b}{2R}\left(T+\frac{T_\mathrm{r}^2}{T}-2T_\mathrm{r}\right)\\ & +\frac{\Delta_r c}{R}\frac{\left(-T^2-T_\mathrm{r}^2+2TT_\mathrm{r}\right)}{2T^2T_\mathrm{r}^2}\\ & +\frac{\Delta C^\circ_{P_\mathrm{r}}}{RT_\mathrm{r}(\partial\alpha/\partial T)_{P_\mathrm{r}}}\left(\frac{1}{T}\ln\frac{\rho_\mathrm{r}}{\rho}-\frac{\alpha_\mathrm{r}}{T}(T-T_\mathrm{r})\right)\end{aligned} \tag{9.35}$$

where a, b, c are Maier–Kelley heat capacity coefficients and $\Delta C^\circ_{P_\mathrm{r}}$ refers to aqueous species only.

However, Anderson (1995) notes that it is often advantageous to use the model, not on the overall reaction, but on individual species. This makes it easier to add parameters calculated by Maier–Kelley or some other method. In using the density model with individual species and minerals, it is advisable to calculate high T, P Gibbs energies and later combine these to the desired $\log K$ values. This should be done with a slightly different version of Equation (9.33) [substitute $-S^\circ_\mathrm{r}(T-T_\mathrm{r})$ for $\Delta H^\circ_\mathrm{r}(1-T/T_\mathrm{r})$] and Equation (5.33) for the pure phases. This ensures that all species use the same convention for Gibbs energies.

9.8 Change of *K* with pressure

The variation of the equilibrium constant with pressure depends of course on how standard states are defined, because this controls how $\Delta_r G^\circ$ and the activity terms in K vary with pressure. This was discussed in some detail in connection with activities (§8.3). To briefly summarize, if the standard states of all the reaction species are defined as fixed at 1 bar (or conceivably some other pressure), then K does not vary with P; it is a constant at a given T. If the standard states are defined as variable in P, then K does vary with P. This variation takes different forms, depending on whether the reaction species are solid, liquid, gaseous, or aqueous, and is best handled by calculating the variation of each individual species activity with P, then combining them into K.

Surprisingly, it is not often necessary to do this, because applications commonly provide data along isobars, so that only the temperature variation is of interest. Each isobar has its own standard state pressure, so it is not necessary to calculate the change in activities or equilibrium constants between isobars.

The only commonly used relation between pressure and the equilibrium constant is the case where all reactants and products are solids. Assuming the solids are incompressible, applying relation (5.32) to each of the terms in $\Delta_r G^\circ$ results in

$$\begin{aligned}\frac{\partial \ln K}{\partial P} &= -\frac{\partial \Delta_r G^\circ}{\partial P}/RT \\ &= -\Delta_r V^\circ/RT\end{aligned}$$

so that

$$\ln K_{P_2} = \ln K_{P_1} - \frac{\Delta_r V^\circ}{RT}(P_2 - P_1) \tag{9.36}$$

9.9 The amino acid example again

Let's write Equation (5.14) one more time.

$$\mathrm{C_8H_{16}N_2O_3}(aq) + \mathrm{H_2O}(l) = \mathrm{C_6H_{13}NO_2}(aq) + \mathrm{C_2H_5NO_2}(aq) \tag{9.37}$$

or

$$\text{leucylglycine} + \text{water} = \text{leucine} + \text{glycine} \tag{9.38}$$

$$\begin{aligned}\Delta_r G^\circ &= \Delta_f G^\circ_{\text{leucine}} + \Delta_f G^\circ_{\text{glycine}} - \Delta_f G^\circ_{\text{leucylglycine}} - \Delta_f G^\circ_{\text{water}} \\ &= -13\,903\,\mathrm{J\,mol^{-1}}\end{aligned}$$

We now know that our calculation of this $\Delta_r G^\circ$ (§5.5.2), the reaction in which a peptide bond between two amino acids is broken, was only a beginning. The value of $-13\,903\,\mathrm{J\,mol^{-1}}$ means that if all reactants and products had unit activity (leucine, glycine, and leucylglycine had concentrations of 1 molal, and water was pure), the reaction would start to go to the right; leucylglycine would start to break down to leucine and glycine. But we note again the fundamental difference between this reaction between dissolved compounds, and reaction (9.1) between solid compounds. Repeating (9.1) here,

$$\mathrm{NaAlSiO_4}(s) + 2\,\mathrm{SiO_2}(s) = \mathrm{NaAlSi_3O_8}(s) \tag{9.39}$$

$$\begin{aligned}\Delta_r G^\circ &= \Delta_f G^\circ_{\mathrm{NaAlSi_3O_8}} - \Delta_f G^\circ_{\mathrm{NaAlSiO_4}} - 2\,\Delta_f G^\circ_{\mathrm{SiO_2}} \\ &= -20.12\,\mathrm{kJ\,mol^{-1}}\end{aligned}$$

The value of $\Delta_r G^\circ$ of $-20\,120\,\mathrm{J\,mol^{-1}}$ means that reaction (9.39) will also go to the right. But this reaction will continue to go (strictly, it *should* continue to go, according to our model) until either $NaAlSiO_4(s)$ or $SiO_2(s)$ is used up. Thus $NaAlSiO_4(s)$ and $SiO_2(s)$ *are not stable together* – one of them must disappear.

This is not the case with leucylglycine. We cannot say that leucylglycine is not stable in water–what happens to it depends entirely on its concentration and on the concentrations of other things in solution such as leucine and glycine. The unit activities are only a starting point, and a very unrealistic one at that. The next step is to calculate the equilibrium constant for (9.37)

$$\begin{aligned}\Delta_r G^\circ &= -RT\ln K\\ -13\,903 &= -(8.3145\times 298.15)\ln K\\ \log K_{298} &= 13\,903/(2.302\,59\times 8.3145\times 298.15)\\ &= 2.436\end{aligned}$$

Thus

$$\frac{a_{\text{leucine}}\,a_{\text{glycine}}}{a_{\text{leucylglycine}}\,a_{\text{water}}} = 10^{2.436}$$

The activity (mole fraction) of water in biochemical systems is usually close to 1.0, so we see that although leucylglycine is not “unstable” in water, its concentration at equilibrium must be quite a bit less than that of its constituent amino acids. For example, if leucine and glycine had concentrations of say $10^{-3}\,m$ (activities of 10^{-3}), the equilibrium activity of leucylglycine would be $10^{-8.436}$ (concentration $10^{-8.436}\,m$). So with concentrations of 10^{-3}, 10^{-3}, and $10^{-8.436}$, leucine, glycine, and leucylglycine would not react at all, but would be at equilibrium. In fact, with a concentration of leucylglycine of less than $10^{-8.436}$, the reaction as written would go to the left – leucylglycine would form from the two amino acids. So remember this – unless the reaction consists only of pure phases,

You cannot reliably tell which way the reaction will go by looking at $\Delta_r G^\circ$.

You can always tell which way the reaction will go by looking at $\Delta_r G$.

Look at Equation (9.8) one more time. When leucine, glycine, leucylglycine, and water all have unit activities, (9.8) becomes

$$\begin{aligned}\Delta_r\mu &= \Delta_r\mu^\circ + RT\ln Q\\ -13\,903 &= -13\,903 + RT\ln\left(\frac{1\times 1}{1\times 1}\right)\end{aligned}$$

In other words, $\Delta_r\mu$ is the same as $\Delta_r\mu^\circ$; the driving force for the reaction can be obtained directly from the tables, as for solid–solid reactions. When products and reactants have reached equilibrium,

$$\begin{aligned}\Delta_r\mu &= \Delta_r\mu^\circ + RT\ln K\\ 0 &= -13\,903 + RT\ln\left(\frac{10^{-3}\times 10^{-3}}{10^{-8.436}\times 1}\right)\end{aligned}$$

Now the $\ln K$ term exactly balances the $\Delta_r\mu^\circ$ term, and the driving force for the reaction is zero. If $a_{\text{leucylglycine}} < 10^{-8.436}$, the driving force ($\Delta_r\mu$) becomes positive.

9.9.1 Peptides favored at higher temperatures

To round out our discussion of this reaction, let's calculate the effect of temperature on the equilibrium constant in reaction (8.18). From Appendix B we find the following data:

Substance	Formulas	$\Delta_f H^\circ$, J mol^{-1}	S°, J mol^{-1} K^{-1}
leucine	$C_6H_{13}NO_2(aq)$	−632 077	215.48
glycine	$C_2H_5NO_2(aq)$	−513 988	158.32
leucylglycine	$C_8H_{16}N_2O_3(aq)$	−847 929	299.16
water	$H_2O(l)$	−285 830	69.91

$$\begin{aligned}\Delta_r H^\circ &= \Delta_f H^\circ_{\text{leucine}} + \Delta_f H^\circ_{\text{glycine}} - \Delta_f H^\circ_{\text{leucylglyine}} - \Delta_f H^\circ_{\text{water}}\\ &= -632\,077 - 513\,988 - (-847\,929) - (-285\,830)\\ &= -12\,306\ \text{J mol}^{-1}\end{aligned}$$

These aqueous species are not ionized, so perhaps our constant $\Delta_r C_P^\circ$ assumption will not be too bad over small temperature intervals. Suppose we wanted the value of K at 100 °C. Equation (9.22) then becomes

$$\begin{aligned}\log K_T &= \log K_{298} - \frac{\Delta_r H^\circ_{298}}{2.302\,59\,R}\left(\frac{1}{T} - \frac{1}{298.15}\right)\\ &= 2.436 - \frac{-12\,306}{2.302\,59\times 8.314\,5}\left(\frac{1}{373.15} - \frac{1}{298.15}\right)\\ &= 2.00\end{aligned}$$

Alternatively, by calculating $\Delta_r S^\circ$, you could use Equation (6.17) first, then Equation (9.11). Thus

$$\begin{aligned}\Delta_r S^\circ &= S^\circ_{\text{leucine}} + S^\circ_{\text{glycine}} - S^\circ_{\text{leucylglyine}} - S^\circ_{\text{water}}\\ &= 215.48 + 158.32 - 299.16 - 69.91\\ &= 4.73\ \text{J mol}^{-1}\end{aligned}$$

Then

$$\begin{aligned}\Delta_r G^\circ_{373} &= \Delta_r H^\circ_{298} - T\,\Delta_r S^\circ_{298}\\ &= -12306 - 373.15 \times 4.73\\ &= -14071\ \text{J mol}^{-1}\end{aligned}$$

from which

$$\begin{aligned}\log K_{373} &= \frac{-\Delta_r G^\circ_{373}}{2.30259\,RT}\\ &= -\frac{-14071}{2.30259 \times 8.3145 \times 373.15}\\ &= 1.97\end{aligned}$$

There will often be a small discrepancy in $\log K$ calculated in different ways, as here (2.00 versus 1.97), because of slight inconsistencies in the data. In other words, to get answers that are exactly the same no matter which way the calculation is done, the data in the tables for each compound must satisfy the relation

$$\Delta_f G^\circ = \Delta_f H^\circ - 298.15 \times \Delta_f S^\circ$$

Because enthalpy, entropy, and Gibbs energy data come from different experiments, using a variety of methods, this relation is often not satisfied exactly in the tabulated data.

The interesting aspect of this calculation of K is that according to the data, leucylglycine (and perhaps all peptide bonds in proteins) becomes *more stable* as temperature increases. Thus for the same concentrations of leucine and glycine ($10^{-3}m$) as before, we find the leucylglycine concentration is $10^{-8.0}\,m$ at 100 °C, compared to $10^{-8.436}\,m$ at 25 °C. That is, its concentration is more than doubled. This result is quite interesting to those scientists trying to figure out how life could have begun in the early days of the Earth, 3.5 billion years ago. The fact that increasing temperatures do not impair but in fact aid the bonding of simple amino acids, the building blocks of life, has led to thoughts that perhaps life began when the oceans were at higher temperatures, or in particular locations (volcanic environments) where heat was available.

This result is typical of the value of thermodynamics. It does not and cannot tell you how life began, but it can tell you which processes are possible and which impossible, and what the effects of changing the constraints on your system will be. This guides the development of scientific ideas in an essential way and provides a universally agreed-upon bedrock from which to start. However, it is up to you to think of the processes to ask thermodynamics about, and this is the creative part of science.

9.10 Some conventions regarding components

There are two ways in which the way chemical formulas are used which may prove confusing:

1. Aqueous species are used in both hydrated and nonhydrated forms. For example, dissolved silica is written as $SiO_2(aq)$ or as $SiO_2 \cdot 2H_2O$ (or H_4SiO_4, which is the same thing).
2. Formulas can be presented in various multiples. For example forsterite may be listed as Mg_2SiO_4 or as $MgSi_{0.5}O_2$.

These are quite simple relationships, but they can cause quite a bit of confusion.

9.10.1 Hydrated versus nonhydrated species

H_4SiO_4 example

One way to write the dissolution reaction for quartz is

$$SiO_2(s) + 2H_2O = H_4SiO_4(aq) \tag{9.40}$$

Another way to write the same reaction is

$$SiO_2(s) = SiO_2(aq) \tag{9.41}$$

The only difference between these two ways of writing a reaction for the dissolution of quartz is that in (9.40) we have assumed that the dissolved silica is in the form of a molecule containing one SiO_2 attached to two H_2O molecules, whereas in (9.41) we have made no assumption as to the form of the dissolved silica. So what do we actually know about dissolved silica? What we know, besides the concentrations under various conditions, is that

1. Under most conditions, the aqueous silica molecule has only one Si (i.e., it is *monomeric*, not *polymeric*), and
2. Except in very basic solutions it is uncharged, electrically neutral.

We might as well then write a formulas for this species that is as simple as possible, while observing these two facts, and $SiO_2(aq)$ does this. Therefore, $SiO_2(aq)$ does *not* refer to a species of dissolved silica which is not attached to any H_2O or other molecules; it refers to the silica that exists as a monomeric uncharged species of whatever nature in solution. It might be attached to two H_2O, or six H_2O, or be a mixture of several such species; it doesn't matter.

The other common formulas, $H_4SiO_4(aq)$, originates historically in the belief that Si in water must be tetrahedrally coordinated by oxygens, as it is in crystals. That may well be true, but there may be other oxygens in the form of H_2Os also attracted to the Si. The exact nature of the complexes of Si and many other elements of interest is a continuing research topic. The important point from the thermodynamic point of view is that what we *call* the dissolved silica, whether $H_4SiO_4(aq)$ or $SiO_2(aq)$, doesn't matter, because as long as we derive the properties of each in a consistent manner, each will give the right answer in calculations.

In what way are the properties of these species different? Because these two formulas refer to the same physical substance, dissolved silica, their concentrations and activities are identical. But because they are related by the equation

$$\mathrm{SiO_2}(aq) + 2\,\mathrm{H_2O}(l) = \mathrm{H_4SiO_4}(aq)$$

their standard state properties such as $\Delta_f G^\circ$ and $\Delta_f H^\circ$ must be different by exactly twice the corresponding property of $H_2O(l)$. Thus

$$\Delta_f G^\circ_{\mathrm{H_4SiO_4}(aq)} = \Delta_f G^\circ_{\mathrm{SiO_2}(aq)} + 2\,\Delta_f G^\circ_{\mathrm{H_2O}}(l)$$

The Gibbs energy of formation of H_4SiO_4 is *defined* as the sum of the Gibbs energies of $SiO_2(aq)$ and (twice that of) $H_2O(l)$. In other words, the relationship between $H_4SiO_4(aq)$ and $SiO_2(aq)$ is strictly a formal one. They are derived from the same experimental data and will yield the same results in calculations.

H_2CO_3 example

The same relationship also holds for other species. For example, when CO_2 gas dissolves in water, it hydrolyzes (reacts with water) to a very small extent, forming some H_2CO_3 molecules in solution. It is rather difficult to determine the exact amount of H_2CO_3, and this problem is avoided by simply calling the total amount of carbon dioxide in solution either $CO_2(aq)$ or $H_2CO_3(aq)$, exactly as the dissolved silica is called $SiO_2(aq)$ or $H_4SiO_4(aq)$. Then for the same reason as before, we find that

$$\Delta_f G^\circ_{\mathrm{H_2CO_3}(aq)} = \Delta_f G^\circ_{\mathrm{CO_2}(aq)} + \Delta_f G^\circ_{\mathrm{H_2O}(l)}$$

Again, this relationship is strictly formal, although in this case it can be more confusing, because there is in fact some literature on the subject of how much dissolved CO_2 actually hydrolyzes to the species H_2CO_3 and how much remains as CO_2 molecules. In other words, H_2CO_3 is sometimes used as a species, and sometimes in the conventional sense we are discussing. The thermodynamic properties of H_2CO_3 in these two senses will of course be completely different. In this book we use the conventional sense for $H_2CO_3(aq)$. Another way of looking at this is to see H_2CO_3 as an alternative component, rather than as an aqueous species.

$Al(OH)_3$ example

Another example is the aluminum species $\mathrm{Al(OH)_3}(aq)$. Again, this is a monomeric uncharged species of Al in solution. There is really no need to assume that it has three oxygens and three hydrogens attached to it. Whatever is attached to it, we can *call* it $\mathrm{AlO_{1.5}}(aq)$, whose properties will differ from those of $\mathrm{Al(OH)_3}(aq)$ by those of $1.5\,\mathrm{H_2O}(l)$, because

$$\mathrm{AlO_{1.5}}(aq) + 1.5\,\mathrm{H_2O}(l) = \mathrm{Al(OH)_3}(aq)$$

and similarly for other aqueous aluminum species. Thus in basic solutions the dominant Al species can be referred to as $\mathrm{Al(OH)_4^-}$ or as $\mathrm{AlO_2^-}$. The essential information is that the species is monomeric and has one negative charge.

9.10.2 What is a mole of olivine?

Another way that the choice of formulas can differ is that some choices can be multiples of other choices. This is most often seen in choosing solute species in solids, because there are no "real" species, just a crystal structure that is a solid solution. For example, the mineral olivine is a solid solution of two components, forsterite ($\mathrm{Mg_2SiO_4}$) and fayalite ($\mathrm{Fe_2SiO_4}$). The solution is represented by $\mathrm{(Mg, Fe)_2SiO_4}$, because the Mg and Fe atoms share the same positions in the crystal structure.

But what reason do we have to choose $\mathrm{Mg_2SiO_4}$ and $\mathrm{Fe_2SiO_4}$ as our components? The formulas simply shows us the stoichiometry of the components – the ratios or relative amounts of the elements. Why not $\mathrm{MgSi_{0.5}O_2}$, or $\mathrm{Mg_4Si_2O_8}$? The same question could also arise in discussing aqueous species, except in that case we often have experimental evidence about the nature of the species in solution. That kind of evidence does not exist for the three-dimensional crystal structures of solid solutions – we are free to choose any component that is stoichiometrically correct. Does it make any difference? Yes.

In §7.2 we noted that some confusion might arise in the definition of mole fractions, but here the choice makes no difference. That is,

$$\frac{n_{\mathrm{FeSi_{0.5}O_2}}}{n_{\mathrm{FeSi_{0.5}O_2}} + n_{\mathrm{MgSi_{0.5}O_2}}} = \frac{n_{\mathrm{Fe_2SiO_4}}}{n_{\mathrm{Fe_2SiO_4}} + n_{\mathrm{Mg_2SiO_4}}}$$

So that is not the problem. The problem is that the choice of formulas, the mole of substance, affects the energy content per mole and hence the activity.

Suppose our system consists of a certain mass (a crystal) of pure forsterite, $\mathrm{Mg_2SiO_4}$. The Gibbs energy of the system is a finite, unknown quantity, which depends on the mass of the crystal. A crystal with twice the mass has a $\mathbf{G}$ twice as large. But the *molar* G does not vary with the size of the crystal. The molar G is defined as $G = \mathbf{G}/n$ (§2.4.1), where n is the number of moles in the crystal. The point is, the number of moles of *what*? Obviously the number of moles of $\mathrm{Mg_2SiO_4}$ in the crystal will be exactly half the number of moles of

$MgSi_{0.5}O_2$ in the crystal, because Mg_2SiO_4 contains twice the number of atoms that $MgSi_{0.5}O_2$ does. Therefore, $G_{Mg_2SiO_4} = 2\,G_{MgSi_{0.5}O_2}$. Or, if you prefer, you can say that $G_{Mg_2SiO_4} = 2\,G_{MgSi_{0.5}O_2}$ simply because it contains twice the mass and, therefore, twice the energy of whatever kind.

This difference in the Gibbs energy of the mole is translated into a difference in activities. Because $G_{Mg_2SiO_4} = 2\,G_{MgSi_{0.5}O_2}$ and $G^\circ_{Mg_2SiO_4} = 2\,G^\circ_{MgSi_{0.5}O_2}$, then

$$G_{Mg_2SiO_4} - G^\circ_{Mg_2SiO_4} = 2\,(G_{MgSi_{0.5}O_2} - G^\circ_{MgSi_{0.5}O_2})$$

and

$$RT \ln a_{Mg_2SiO_4} = 2\,RT \ln a_{MgSi_{0.5}O_2}$$

and therefore

$$a_{Mg_2SiO_4} = a^2_{MgSi_{0.5}O_2}$$

The problem this poses can be seen in considering Raoult's law, which we said was $a_i = x_i$. But if $a_i = a^2_{0.5i}$ we have a problem. Because x_i is independent of how we write the formulas for i, we see that a_i and $a_{0.5i}$ cannot both be equal to x_i, even if Raoult's law is followed exactly. If $a_{0.5i}$ versus x_i is a straight line, then a_i versus x_i will describe a parabola.

This is a well-known problem, and generally the formulas for components is chosen such that the simple statement of Raoult's law is followed as closely as possible. Again, this relationship between activities is entirely formal and tells us nothing about forsterite or olivine. However, it is important to remember that choosing a formulas for your components has consequences for activities.

The problem is more difficult in other systems. How does one choose components in a complex silicate melt, for example? In a melt there are no stoichiometric restrictions to be observed, but the formal relationship between the activities of various component choices that we have discussed remains true. So if you measure the activity of some component in a melt, and determine the deviations of these activities from Raoult's law by calculating activity coefficients, the question is, what part of these activity coefficients represents nonideal behavior, and what part represents a poor choice of components? Generally speaking, extremely large or extremely small activity coefficients mean that the component involved has been badly chosen, which is to say that it does not come very close to representing the "real" situation in the system. In these situations, thermodynamics provides no help whatsoever. It points out the consequences of choices relative to each other, and from there on the investigator is on her own. In other words, the choice of components, as much as the choice of system to investigate, is a part of the "art of doing science," that part which relies on skill and intuition, and can never be taught.

9.11 Summary

This chapter contains a sudden increase in the amount of practical, usable material. If you ever have occasion to use thermodynamics in a practical situation, it will very likely involve the use of the equilibrium constant.

The molar Gibbs energy of a dissolved substance changes with the concentration of the substance. The activity is a dimensionless concentration-like term that is used to give the Gibbs energy in a particular state, in terms of its difference from its value in some reference state (Equation 7.37). When a reaction has reached equilibrium, the activities of the various products and reactants can have a variety of values individually, but their ratio, as expressed in the equilibrium constant K, has a fixed and calculable value.

The equilibrium constant is calculated from numbers (Gibbs energies) taken from tables of standard data (derived experimentally, as discussed in Chapter 5). These standard data give the term $\Delta_r\mu^\circ$ or Δ_rG°, which is a constant for a given T and P. It has nothing to do with whether your system or reaction has reached equilibrium ($\Delta_r\mu = 0$) or not. However, it can be used to calculate K, which gives the ratio of product and reactant activities your reaction will have if it ever reaches equilibrium.

The superscript ° therefore has considerable significance. It should not be omitted or inserted carelessly in your calculations.

10

Real solutions

10.1 Introduction

We have now considered both ideal solution behavior and deviations from this, but in a rather generalized way, using activity coefficients. We now have to start to consider how to measure these things, and doing this means we have to consider partial molar properties in much more detail.

We start with a fairly detailed look at the volumetric properties of solutions, because these are the most intuitive. Partial molar properties of the other state variables are the same in principle, but become more complicated in the case of enthalpy measurements because of its relative nature. The Gibbs energy is also a relative property, but is treated in quite a different way.

Most of the material in this chapter is quite general, and can be applied to any kind of solution, although most of our examples are for aqueous solutions. The properties of electrolyte solutions introduce complications, discussed in Chapter 15. The properties of real gaseous solutions are often handled by "equations of state," the subject of Chapter 13, and those of solid solutions have some unique aspects, discussed in Chapter 14.

10.2 Solution volumes

All real solutions are of course nonideal. Our discussion of their properties will be concerned for the most part with deviations from the properties of ideal solutions, whether Henryan or Raoultian.

10.2.1 Partial and apparent properties

The properties of a dissolved substance are described in terms of *partial*, *apparent*, and *excess* total or molar properties, so we begin by discussing these terms, using volume as an example.

The volume of mixing

If two substances are immiscible (they do not dissolve into one another to any appreciable extent, like oil and water), obviously the volume of the two together is simply the sum of the two volumes separately. But if they are completely miscible (they dissolve into one another completely, forming a solution), this may be more or less true, but probably not exactly true. Why?

If you mix white sand and black sand together, there is no interaction or chemical reaction at all between the two kinds of sand, and the volume of the mixture is the same as the two volumes separately. If the volume of the white sand is $\mathbf{V}_\mathrm{w}$ and the volume of the black sand is $\mathbf{V}_\mathrm{b}$, the total volume is

$$\mathbf{V} = \mathbf{V}_\mathrm{w} + \mathbf{V}_\mathrm{b}$$

It's sort of like stacking boxes as in Figure 10.1. There is no change in total volume just because they are together.

However, using total volumes usually turns out to be inconvenient. If the volume per mole of white sand is V_w and of black sand is V_b, then the total volume is

$$\mathbf{V} = n_\mathrm{w} V_\mathrm{w} + n_\mathrm{b} V_\mathrm{b} \tag{10.1}$$

where n_w and n_b are the number of moles of white and black sand in the mixture. The molar volume is defined as the total volume divided by the number of moles of all components in the system (i.e., the molar volume of pure white sand is therefore $\mathbf{V}_\mathrm{w}/n_\mathrm{w}$); so if the mixture contains n_w moles of white sand and n_b moles of black sand, the total number of moles in the mixture is $n_\mathrm{w} + n_\mathrm{b}$. Dividing both sides of Equation (10.1) by $n_\mathrm{w} + n_\mathrm{b}$, we get

$$V = x_\mathrm{w} V_\mathrm{w} + x_\mathrm{b} V_\mathrm{b} \tag{10.2}$$

Here, V is the *molar volume* of the mixture and x is the *mole fraction*, where

$$\begin{aligned} x_\mathrm{w} &= \frac{n_\mathrm{w}}{\sum n} \\ &= \frac{n_\mathrm{w}}{n_\mathrm{w} + n_\mathrm{b}} \end{aligned} \tag{10.3}$$

and similarly for x_b. This equation simply says that the volume of the mixture is the same as the volume of the two things separately. The introduction of

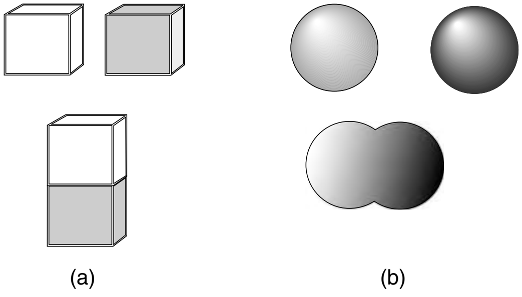

(a) (b)

Figure 10.1 (a) There is no volume change when boxes are stacked together – they do not interact. (b) When molecules are mixed together, they may occupy less volume than they did separately.

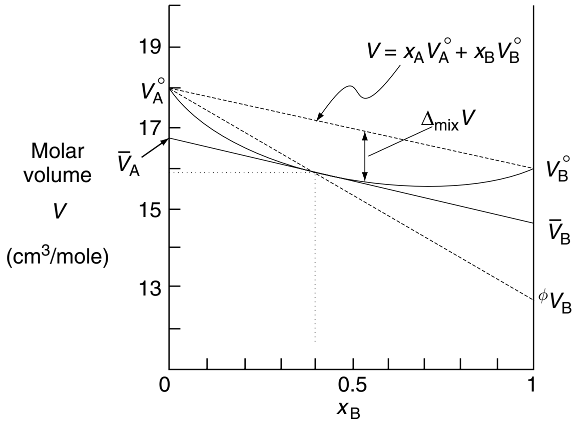


Figure 10.2 The molar volume of solutions of A and B. The molar volume of pure A (V_A°) is 18.0 cm^3 mol^{-1} and that of pure B (V_B°) is 16.0 cm^3 mol^{-1}. The molar volume of an ideal solution having $x_B = 0.4$ is $0.6 \times 18.0 + 0.4 \times 16.0 = 17.2$ cm^3 mol^{-1}. The molar volume of a real solution having $x_B = 0.4$ is actually 15.9 cm^3 mol^{-1}. It may be calculated in the same way, but using $\overline{V}_A$ and $\overline{V}_B$ instead of V_A° and V_B°. The difference between the real and the ideal molar volumes is the change in V on mixing $\Delta_{mix}V$. The apparent molar volume of B ($^\phi V$) in a solution is the intercept on the $x_B = 1$ axis of a line joining the molar volume of pure A and the molar volume of the solution. $^\phi V = 12.75$ cm^3 mol^{-1} from Equation (10.12).

n and *x* is just to determine how much of each is used. If we plot molar volume against mole fraction of either component sand, we get a straight line (Figure 10.2), called the *ideal mixing* line.

Clearly these relations do not depend on the grain size of the sands;[1] they depend on the fact that the sands do not react in any way with each other. Each grain of white sand is indifferent to what kind of sand is next to it. Now imagine that the grain size of the sands gets smaller and smaller. Soon they get so small that you can no longer distinguish the colors – the mixture becomes gray. Imagine the grain size continuing to get smaller and smaller – right down to atomic proportions, so that instead of having a mechanical mixture of black and white sand, we have a true solution of black and white atoms. If the black and white atoms continue to have no attraction, repulsion, or chemical reaction with one another, the volume of the two together will continue to be exactly the same as the sum of the two separately. Actually, we have oversimplified a bit – normally the white molecules interact with each other even in the pure state, and similarly with the black molecules. If these interactions are very similar in nature, then when they are mixed together the molecules will continue to interact with each other in the same way, and the volumes will be additive. In other words, it is not necessary for there to be no molecular interactions for

---

[1] Actually, only as long as the grain size of the black and white sands are the same.

ideal mixing, only that white molecules react with black molecules in exactly the same way that they do with other white molecules.

But suppose that at this molecular size, white (w) and black (b) particles are attracted to one another more than to others of the same kind, perhaps even forming a new kind of particle (wb). Because of this attraction, the particles will be closer together than they would otherwise be, and the total volume of the mixture will be smaller, as shown in Figure 10.1b, and instead of getting a straight mixing line as in Figure 10.2, the line is curved downward as in Figure 10.2. Alternatively, if the white and black particles repel each other, the total volume will be greater, and in Figure 10.2b the curved line for the molar volume of the mixture will lie above the straight line that represents no interaction. The volume change on mixing ($\Delta_{\text{mix}}V$, Figure 10.2) caused by the attraction between A and B is the difference between the straight line and the curved line. The straight line

$$V = x_A V_A^\circ + x_B V_B^\circ \tag{10.4}$$

is called *ideal mixing* and is rarely observed.[2] The curved line represents nonideal mixing, the general case. The difference between the ideal mixing line and the actual molar volume V is called the change in volume on mixing, $\Delta_{\text{mix}}V$. Thus

$$\Delta_{\text{mix}}V = V - (x_A V_A^\circ + x_B V_B^\circ) \tag{10.5}$$

10.2.2 Partial molar volumes

Now suppose in our mixture of white and black particles that attract each other, that we are not satisfied to have the total volume or the molar volume of the mixture as a whole. We would like to know the volume of each component in the mixture, not just the combined volume. But how can this be done, when each is dispersed at the molecular level and is interacting strongly with another component? Simple. Just draw the tangent to the molar volume curve at the composition you are interested in. The intercepts of this tangent give the volumes of each component in the solution, called *partial molar volumes*, which are combined to give the total molar volume in exactly the same way as the black and white sands in Equation (10.2) and Figure 10.2.

Looking at partial molar volumes in this way, they seem to be just a sort of geometrical construct. They are defined such that they can be substituted for V_A and V_B in Equation (10.2) in cases where mixing results in a curved line for the molar volumes; thus[3]

$$V = x_A \overline{V}_A + x_B \overline{V}_B \qquad \text{(Figure 10.2)} \tag{10.6}$$

---

[2] Volumetric ideal mixing (Equation 10.4) is also called Amagat's law, which we saw was connected to the Lewis fugacity rule in Chapter 8.

[3] These equations are also derived in §C.2.3.

or, multiplying both sides by $(n_A + n_B)$,

$$\mathbf{V} = n_A \overline{V}_A + n_B \overline{V}_B \tag{10.7}$$

In Figure 10.2 we have shown a case where A and B are attracted to each other, and their partial molar volumes are both *less* than the volumes of the pure components ($\overline{V}_A < V_A^\circ$). If A and B repelled one another, the mixing line would lie above the straight line and the partial molar volumes would be *larger* than the pure volumes. There is no general rule for the shapes and positions of these mixing curves; they must be measured experimentally. This would be done by density measurements in the case of volume, and calorimetry in the case of enthalpy and entropy. It is quite possible for the mixing curve to be shaped such that in a certain range of composition one of the tangent intercepts is at less than zero volume – a negative partial molar volume. This is why some of the tabulated thermodynamic parameters in Appendix B are negative for some solute components. It is, of course, not possible for pure components to have a negative volume.

The room analogy

But there is another way of looking at partial molar volumes which shows that they really are the volume of a mole of each component in solution. Just for a change we will switch from components A and B to a solution of salt (NaCl) in water. Consider an extremely large quantity of water – say enough to fill a large room (Figure 10.3). Now let's add enough salt to make the concentration exactly 1 molal, and adjust the volume of the solution so that the room is full and a little excess solution sticks up into a calibrated tube inserted into the ceiling. By observing changes in the level of solution in the tube, we can accurately record changes in the $\mathbf{V}$ of the solution in the room.

Now, when we add a mole of NaCl (58.5 g of NaCl occupying 27 cm^3) to the solution, the change in concentration is very small. In fact, if we can

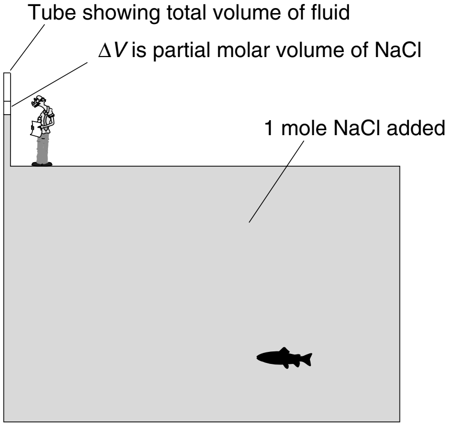


Figure 10.3 A roomful of 1 molal salt solution. The observer sees the change in volume caused by adding one mole of salt, which is the partial molar volume of salt in the 1 molal solution.

detect *any* change in concentration by the finest analytical techniques available, then our room is too small, and we must find and inundate a larger one. Eventually, we will fill a sufficiently large room with salt solution that on adding 58.5 g of NaCl we are unable to detect any change in concentration – it remains at 1.000 mole NaCl/kg H_2O. But although the concentration remains unchanged, the volume of course does not. The salt added cannot disappear without a trace. The level in the tube in the ceiling changes, and the $\Delta\mathbf{V}$ seen there is evidently the volume occupied by 1 mole of NaCl in a 1 molal NaCl solution, in this case about $19.47\,\mathrm{cm}^3\,\mathrm{mol}^{-1}$ of NaCl. This is, in quite a real sense, the volume occupied by a mole of salt in that salt solution and has a right to be thought of as a *molar volume* (just as much as $27\,\mathrm{cm}^3\,\mathrm{mol}^{-1}$ is the molar volume of crystalline salt) rather than as an arbitrary mathematical construct. It is referred to as the *partial* molar volume of NaCl in the salt solution, $\overline{V}_{\mathrm{NaCl}}$.

Some readers will have difficulty in seeing how, on adding our salt, the concentration does not change but the total volume does. If this is the case, think of the room as containing not a solution, but nine million white tennis balls and one million black tennis balls, all mixed together. The room is full, the balls are arranged so that no space is available for another ball, and a few balls overflow into the tube in the ceiling. The total volume is the volume of ten million tennis balls. Now we add one more black tennis ball, somewhere in the middle of the room. The fractional concentration of black balls changes from $10^6/10^7$ to $(10^6+1)/(10^7+1)$, or from 0.1 to 0.100 000 09, a change so small it is completely negligible.[4] But the total volume has changed by the volume of one tennis ball, and this change must be reflected by the level of the balls in the tube, which will rise by the volume of one ball. We can even extend the analogy by imagining that the balls in the room are compressed by the pressure, so that when we add another ball, it becomes compressed too, and the level in the tube rises by the volume of a compressed tennis ball, not a normal (standard state) tennis ball.

The formulas for partial molar properties

The partial molar concept is applied to most thermodynamic properties, not just volume. The mathematical expression, introduced in §2.4.1 on page 19, is

$$\left(\frac{\partial \mathbf{Z}}{\partial n_i}\right)_{\hat{n}_i} = \overline{Z}_i \tag{10.8}$$

where $\mathbf{Z}$ is a thermodynamic parameter such as $\mathbf{V}$, $\mathbf{S}$, $\mathbf{G}$, etc., n_i is the moles of component i, and $\hat{n}_i$ is the moles of all solution components *except* i. It is

---

[4] If you don't find it negligible, just imagine a bigger room and more tennis balls, until the change *is* negligible.

important to note that the derivative is taken of the *total* quantity, $\mathbf{Z}$, not the molar property, Z. It is the change in the *total* volume of the solution in the room that is measured, not the molar volume.

Put in this partial differential form, partial molar properties look somewhat obscure. However, it is important to have an intuitive grasp of their meaning, and you will be well advised to think of them in the sense of the room analogy, as *molar* properties of solutes in solutions of particular compositions, rather than in terms of Equation (10.8).

10.2.3 Apparent molar volume

The first thing we come across when looking at real data is that quite often the data are reported as "apparent" molar volumes, enthalpies, entropies or heat capacities. If we call component 1 the solvent (usually water in our cases), component 2 the solute (say, NaCl), $\mathbf{Z}$ and Z the total and molar forms of any of these properties, then apparent molar properties are defined as

$$^{\phi}Z = \frac{\mathbf{Z} - n_1 Z_1^\circ}{n_2} \tag{10.9}$$

or, in the case of volume,

$$^{\phi}V = \frac{\mathbf{V} - n_1 V_1^\circ}{n_2} \tag{10.10}$$

where V_1° is the molar volume of the pure phase. Thus the apparent molar volume is the volume that should be attributed to a mole of solute, if one assumes that the solvent contributes the same volume it has in its pure state (Figure 10.2). Alternatively,

$$\mathbf{V} = n_1 V_1^\circ + n_2 {}^{\phi}V \tag{10.11}$$

or, dividing by $(n_1 + n_2)$,

$$V = x_1 V_1^\circ + x_2 \cdot {}^{\phi}V \tag{10.12}$$

The apparent molar volume is known as accurately and as easily as the molar volume or the total volume of a solution whose composition is known, whereas finding the partial molar volume always involves some manipulation of the raw data (such as determining a tangent) and requires a knowledge of a range of compositions, not just a single one. Therefore measurement of the density of a solution (§5.8.1) enables you to calculate $\mathbf{V}$, the total volume of the solution. Then because you know the molar volume of the pure solvent V_1°, Equation (10.10) gives you the apparent molar volume of the solute in the solution you measured.

To obtain partial molar volumes from measurements of apparent molar volumes, differentiate Equation (10.11),

$$\left(\frac{\partial \mathbf{V}}{\partial n_2}\right)_{T,P,n_1} = \overline{V}_2 \tag{10.13}$$

$$= n_2\left(\frac{\partial {}^{\phi}V}{\partial n_2}\right)_{T,P,n_1} + {}^{\phi}V \tag{10.14}$$

$$= m\left(\frac{\partial {}^{\phi}V}{\partial m}\right) + {}^{\phi}V \quad \text{if } n_1 = 55.51 \tag{10.15}$$

so that evidently we need an equation to represent ${}^{\phi}V$ as a function of m, in order to calculate $(\partial {}^{\phi}V/\partial m)$. For electrolytes, it is found that using $\sqrt{m}$ rather than m gives a more linear plot, so that Equation (10.15) becomes

$$\overline{V}_2 = m\left(\frac{\partial {}^{\phi}V}{\partial m^{1/2}}\right)\left(\frac{\partial m^{1/2}}{\partial m}\right) + {}^{\phi}V$$

$$= \tfrac{1}{2}m^{1/2}\left(\frac{\partial {}^{\phi}V}{\partial m^{1/2}}\right) + {}^{\phi}V \tag{10.16}$$

Extrapolating values of ${}^{\phi}V$ back to $m = 0$ will therefore provide a value of $\overline{V}_2^{\circ}$, the partial molar volume at infinite dilution, which is the standard state value.

10.2.4 An example from NaCl–H_2O

As an example of the various terms we have defined, consider the system NaCl–H_2O. This system differs from the system A–B in Figure 10.2 only in the sense that A and B are completely miscible (they dissolve in each other in all proportions), while in NaCl–H_2O water becomes saturated with NaCl at a concentration which depends on P and T. This is 6.1 molal at 25 °C, 10.4 molal at 300 °C, so we can only look at concentrations below this value.

Partial molar volume

Apparent molar volumes at concentrations up to 5 molal are shown in Table 10.1 and Figure 10.4. If the mass of solvent, water, is 1 kg, then n_1 is $1000/18.0154 = 55.51$, and n_2 is the NaCl molality, m. The volume of the pure solvent V_1° is 18.068 $\text{cm}^3\,\text{mol}^{-1}$, so measurements of the total volume of the solution $\mathbf{V}$ give molar volumes [from $\mathbf{V}/(55.51 + m)$] and apparent molar volumes from Equation (10.11).

Apparent molar volumes ${}^{\phi}V$ can be converted into partial molar volumes in several ways. One way would be to actually do the operation illustrated in Figure 10.2, that is, construct the tangent to the molar volume curve, and determine the intercept on the $x_{\text{NaCl}} = 1$ axis.[5] The tangent at 3 molal is shown

---

[5] The mole fractions in Table 10.1 and Figure 10.5 are calculated using $\nu = 1$ rather than $\nu = 2$ as defined in Equation (7.2). Extrapolating a tangent to $x_{\text{NaCl}} = 1$ using $\nu = 2$ results in a value of $\frac{1}{2}\overline{V}_{\text{NaCl}}$, and the diagram is less intuitive.

Table 10.1 *Volumetric data for NaCl– H_2O at 25 °C, 1 bar. From the equation of state of Archer (1992).*

Molality m mol kg^{-1}	Mole fraction	Apparent molar volume $^{\phi}V$ cm^3 mol^{-1}	Molar volume V cm^3 mol^{-1}	Total volume **V** cm^3	Partial molar volume $\overline{V}$ cm^3 mol^{-1}
0.0	0.0000	16.62	18.068	1002.95	16.62
0.2	0.0036	17.43	18.066	1006.44	17.85
0.4	0.0072	17.78	18.066	1010.07	18.39
0.6	0.0107	18.05	18.068	1013.79	18.80
0.8	0.0142	18.28	18.071	1017.59	19.15
1.0	0.0177	18.49	18.076	1021.45	19.47
1.5	0.0263	18.94	18.091	1031.37	20.15
2.0	0.0348	19.33	18.112	1041.59	20.75
2.5	0.0431	19.67	18.136	1052.09	21.27
3.0	0.0513	19.98	18.165	1062.84	21.75
3.5	0.0593	20.27	18.197	1073.81	22.19
4.0	0.0672	20.52	18.232	1085.00	22.59
4.5	0.0750	20.76	18.270	1096.39	22.97
5.0	0.0826	20.97	18.311	1107.97	23.31

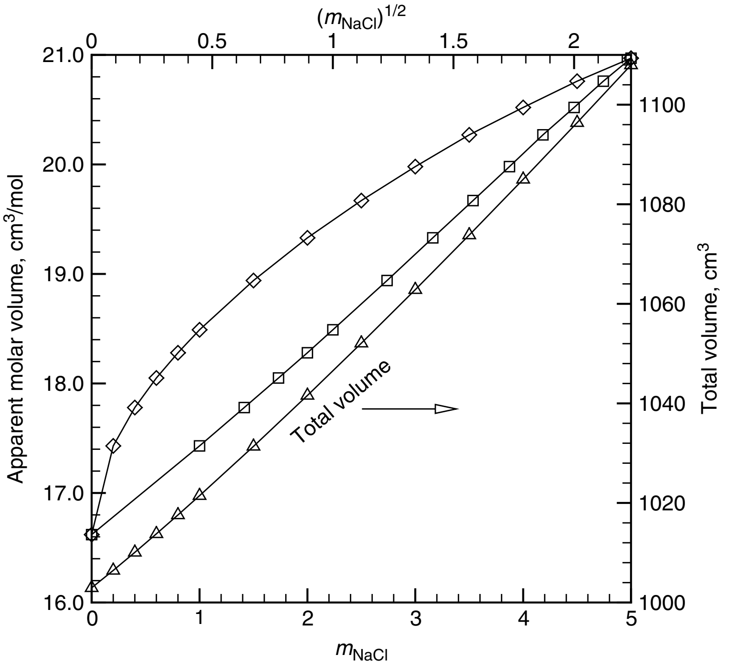


Figure 10.4 Apparent molar volume of NaCl ($^{\phi}V$) in water, and the total volume of solution, as a function of NaCl molality. Top curve: $^{\phi}V$ versus m_{NaCl}; middle curve: $^{\phi}V$ versus $m_{\mathrm{NaCl}}^{1/2}$; bottom curve: total volume of NaCl solution.

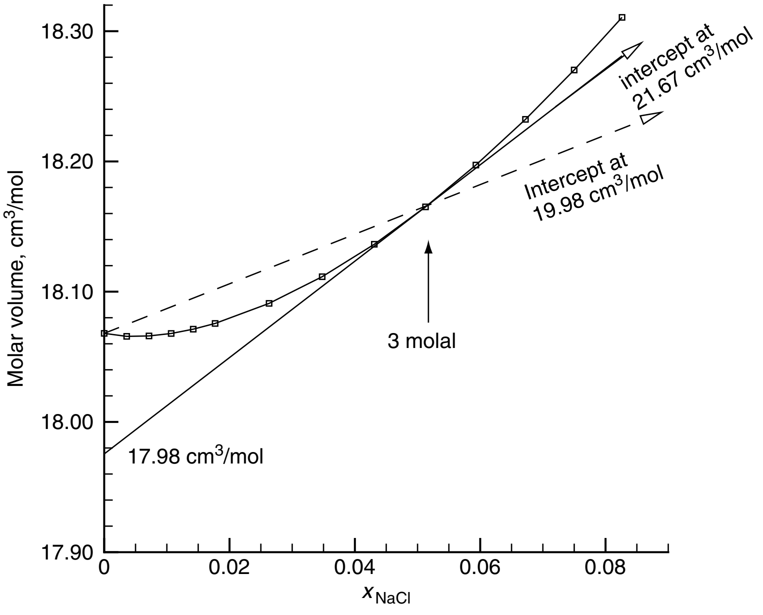


Figure 10.5 Molar volume of NaCl solutions versus mole fraction NaCl. The tangent to the curve at 3 molal intercepts the $x_{\mathrm{NaCl}} = 1$ axis at $\overline{V}^{\circ}_{\mathrm{NaCl}}$ and the $x_{\mathrm{NaCl}} = 0$ axis at $\overline{V}^{\circ}_{\mathrm{H_2O}}$. The dashed line intercepts the $x_{\mathrm{NaCl}} = 1$ axis at the apparent molar volume ${}^{\phi}V$. The tangent to the curve at $x_{\mathrm{NaCl}} = 0$ intercepts the $x_{\mathrm{NaCl}} = 1$ axis at $\overline{V}^{\infty}_{\mathrm{NaCl}}$.

in Figure 10.5. Note that the intercept at $x_{\mathrm{NaCl}} = 1$, 21.67 cm^3 mol^{-1}, is a bit different from the value in Table 10.1, 21.75 cm^3 mol^{-1}. This is because the tangent method involves differentiation plus a very long distance from $x_{\mathrm{NaCl}} = 0.0513$ (the value at 3 molal) to $x_{\mathrm{NaCl}} = 1$.

Another way would be to determine the slope of the total volume curve (Equation 10.13), which gives another slightly different value.

But the commonest method has been to fit an equation to ${}^{\phi}V$ data. If we use a linear equation and $m^{1/2}$ in place of m, we get

$$ {}^{\phi}V = \overline{V}^{\infty} + \mathrm{S}^{*}_{V}\, m^{1/2} \tag{10.17} $$

known as the Masson equation, where the slope is S^{*}_{V} and the intercept at $m = 0$ is the partial molar volume at infinite dilution, $\overline{V}^{\infty}$. Using the data in Table 10.1, S^{*}_{V} is 1.978, and $\overline{V}^{\infty}$ is 16.54 cm^3 mol^{-1}. However, Millero (1971) shows that despite the fact that this equation has the correct theoretical slope at infinite dilution, and was used extensively for many years, it often gives incorrect values of $\overline{V}^{\infty}$.

Other approaches can be used based on corrections to this equation (e.g., Helgeson and Kirkham, 1976), but in recent years the tendency has been to use the Pitzer equations (Chapter 15). Determining the intercept of this equation, or any nonlinear equation, at $m = 0$ places great emphasis on measurements of very dilute solutions, where they are most difficult. Clearly, some theoretical knowledge of what the slope at the intercept (the "limiting slope") should be is important, and all modern treatments of data of this type use the limiting slopes predicted from Debye–Hückel theory, which will be discussed in Chapter 15. The values in Table 10.1 from Archer (1992) are from an equation of state which uses the theoretical limiting slopes, and in addition involves not just the data at 25 °C, 1 bar, but a great many other data as well.

Electrostriction

Millero (1971) provides an interesting history of ideas on the volume occupied by dissolved electrolytes. Before 1770, a salt was thought to dissolve in water simply by filling up void spaces, thereby not changing the water volume. The experimental evidence used to support this idea was that a glass of water did not overflow when a spoonful of salt was added. In 1770 Watson (reference in Millero, 1971) showed that the volume definitely decreases when various salts are added to water, showing that the prevailing theory of the "porous" nature of water was incorrect. However, his work was soon forgotten, and the older ideas prevailed for another 70 years. Millero reports that Watson tired of chemistry and entered the ministry.

These days we look at the partial molar volume of salts. The molar volume of pure crystalline NaCl is 27.015 $\mathrm{cm^3\,mol^{-1}}$, so you see from the values of the partial molar volume in Table 10.1 that NaCl occupies less volume per mole in solution than it does in the solid form, at all concentrations. This fact is even more striking if you look at the partial molar volume at infinite dilution at high temperatures, as shown in Figure 10.6. At every pressure up to 1000 bars, $\overline{V}_{\mathrm{NaCl}}$ becomes negative, reaching almost $-100\,\mathrm{cm^3\,mol^{-1}}$ at 300 °C, saturation vapor pressure (which is 86 bars). This means that if you were the observer on top of the room full of pure water in Figure 10.3, and if the water was at 300 °C, you would observe the volume of water in the room to *decrease* by almost 100 $\mathrm{cm^3}$ when 58 grams of NaCl was added to the water. The volume of water is so large that after adding the salt, its presence is undetectable, i.e., it is at infinite dilution.

Figure 10.1 suggests that this effect is due to attraction between the NaCl and the H_2O. This is true, but in the case of electrolytes in water, it is somewhat oversimplified. Water is not gas-like, with a complete absence of structure, but largely due to its polar nature and hydrogen bonding, it has some kind of structure which has been the subject of much debate. Addition of charged particles in the form of a dissociated electrolyte disrupts and "collapses" this structure, and the resulting ion–water interactions of several types usually results in an overall decrease in volume, known as "electrostriction", despite the additional volume of the electrolyte. These interactions form the basis of the HKF model embedded in SUPCRT92, to be discussed in Chapter 15.

10.3 The infinite dilution standard state

So far we have just assumed that the standard state for our mixing components is the pure phase, just as it was in Chapter 3. This presents no problem

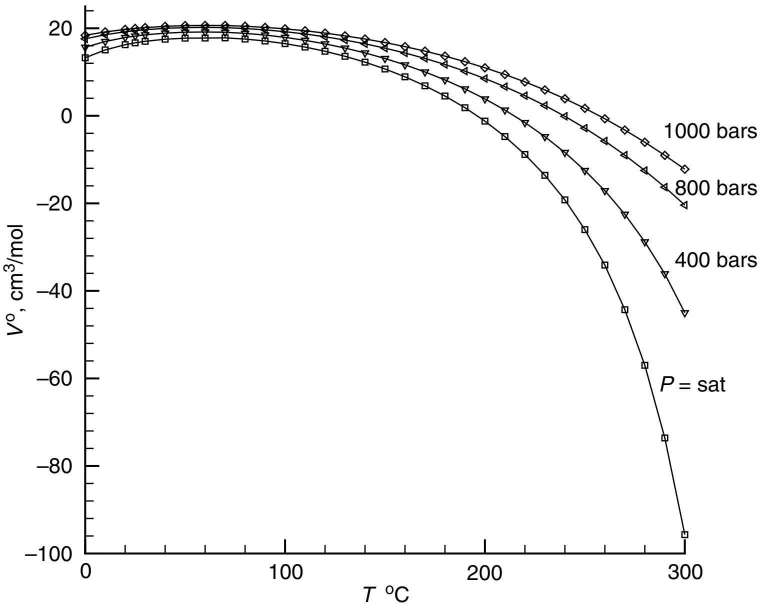


Figure 10.6 The partial molar volume of NaCl at infinite dilution as a function of temperature and pressure. P = sat is 1 bar up to 100 °C, and the saturation pressure of water above 100 °C. Data from Pitzer and Peiper (1984).

for water and NaCl – the standard states are, or could be, pure water and pure halite at the T and P of interest. But this doesn't work in the many cases where the solute does not exist as a pure liquid or solid phase, such as H_2SO_4, or any gaseous solute. An alternative is the "infinitely dilute solution," which is always available by extrapolation, and has the advantage that in the standard state the solute particles interact only with the solvent, not with other solute particles. Deviations from the standard state value are then some measure of solute–solute (or solute-affected solvent–solute) interactions. So infinite dilution is the standard state chosen for enthalpy, volume and heat capacity.

A final reminder about standard states

Standard states are necessary because G, A, H, and their partial derivatives, as well as the activity (functionally) related to a difference in Gibbs energies), can only express the energy difference between a state of interest and some other state. The standard state is used to answer the question, the difference from what other state? Once this state is defined, it of course also has values of V° and C°_P, which don't really require standard states, because their absolute values are (or can be) known.

10.3.1 The entropy standard state

In summary, then, for dissolved substances we use the ideal one molal standard state for Gibbs energy, and the infinite dilution standard state for enthalpy, volume and heat capacity. What about entropy?

By looking at one of our Equations (7.26)

$$\mu_{\mathrm{A}} - G^{\circ}_{\mathrm{A}} = RT \ln x_{\mathrm{A}}$$

it is easy to see why we cannot use infinite dilution for Gibbs energy. G°_{A} is a constant, so $\mu_{\mathrm{A}} \to -\infty$ as $x_{\mathrm{A}} \to 0$. So infinite dilution is hard to deal with. Equation (7.28) is

$$\overline{S}_i - \overline{S}^{\circ}_i = -R \ln x_i$$

so the same problem arises for entropy. Now we can write

$$G = H - TS$$

$$G^{\circ} = H^{\circ} - TS^{\circ}$$

and

$$G - G^{\circ} = (H - H^{\circ}) - T(S - S^{\circ})$$

where G° is for an ideal one molal solution and H° is for infinite dilution, and the question is, to what does S° refer?

It happens that for H (and V, C_P), the value at infinite dilution is equal to the value in an ideal one molal solution (and anywhere else on the Henryan tangent), so if G, H, and S refer to an ideal one molal solution, then $G - G^{\circ}$ and $H - H^{\circ}$ are both zero, and $S - S^{\circ}$ is zero only if S° also refers to an ideal one molal solution. Because entropy is calculated from other measurements (e.g., $\overline{S} = (\overline{H} - \overline{G})/T$) rather than being measured directly, this fact is perhaps not as useful as the others we have been discussing.

10.3.2 The reference state

Because partial molar volume, enthalpy, and heat capacity are the same anywhere on the Henry's law tangent, including both the state of infinite dilution and the ideal one molal solution, either of these states can serve as the standard state for these properties. We have chosen to say that the infinitely dilute solution is the standard state, but many treatments prefer to say that the standard state for these properties, as well as for the Gibbs energy and entropy, is the ideal one molal solution. For some reason, these treatments (e.g., Klotz, 1964, p. 361) then define the "reference state" for enthalpy, volume and heat capacity as the state of infinite dilution. This appears to have little usefulness, whatever standard state definition is used.

If the standard state is defined as having a fixed pressure of 1 bar, the reference state is also sometimes referred to as a state reached from the standard state by a change in pressure (Pitzer and Brewer, 1961, p. 249). Because in this text we use a variable pressure standard state, we have no need of the reference state in this sense either.

10.3.3 Symbols for the standard state

Superscript °, as in V°, indicates the standard state. Up to this point, this has generally been synonymous with the pure state (pure solid, liquid, or gas), so one might get the impression that ° indicates the pure phase. However, with solutions, we must be more careful, because the pure phase is not always the standard state.

In solutions, particularly electrolyte solutions, the standard state for the solvent is always the pure phase (pure water), so that, for example, V_1° refers to the molar volume of pure component 1, that is, pure water. For the solute, the standard state for most properties is, as just mentioned, the state of infinite dilution, so we could use $\overline{V}_2^\circ$ for the partial molar volume of the solute in the standard state. However, this proves a bit confusing, so for clarity we introduce superscript $^\infty$ to indicate the infinite dilution state ($\overline{V}_2^\infty$), and we understand that this is also the standard state for most properties. This raises the question of what symbol to use for the solute in its pure state. The IUPAC recommends the use of * for pure substances, but our examples involve only minerals so we will just use the mineral name. Thus we use V^{halite} for the molar volume of pure NaCl.

In the case of the Gibbs energy for aqueous solutes we saw (§8.2.3) that the standard state is *neither* the pure phase nor the infinitely dilute state, and the usual ° symbol is appropriate, as in G° and μ°.

10.4 Excess properties

In this section we extend our discussion of solution volumes (§10.2) to other properties, and introduce the excess properties. The difference between the property (V, H, etc.) of a real solution, and what that property would be if the solution was ideal, is called an *excess* property.

Thus, from Equations (7.9), (7.10), (7.11),

$$\begin{aligned} V^{\text{EX}} &= V_{\text{real sol'n}} - V_{\text{ideal sol'n}} \\ &= V_{\text{real sol'n}} - \sum_i x_i V_i^\circ \qquad (10.18) \\ &= \Delta_{\text{mix}} V \qquad (10.19) \end{aligned}$$

and similarly

$$H^{\mathrm{EX}} = \Delta_{\mathrm{mix}} H \tag{10.20}$$

$$C_P{}^{\mathrm{EX}} = \Delta_{\mathrm{mix}} C_P \tag{10.21}$$

For ideal solutions, these quantities are of course all zero (Equations 7.12, 7.13, 7.14). They are also true for the total properties, so that for ideal solutions

$$\mathbf{V}^{\mathrm{EX}} = 0 \tag{10.22}$$

$$\mathbf{H}^{\mathrm{EX}} = 0 \tag{10.23}$$

$$\mathbf{C}_{\boldsymbol{P}}{}^{\mathrm{EX}} = 0 \tag{10.24}$$

Relations between excess properties are the same as between their parent properties. For example,

$$\left(\frac{\partial G^{\mathrm{EX}}}{\partial P}\right)_T = \left(\frac{\partial G_{\mathrm{real}}}{\partial P}\right) - \left(\frac{\partial G_{ideal}}{\partial P}\right) \tag{10.25}$$

$$= V_{\mathrm{real}} - V_{\mathrm{ideal}} \tag{10.26}$$

$$= V^{\mathrm{EX}} \tag{10.27}$$

so that in addition,

$$\left(\frac{\partial G^{\mathrm{EX}}}{\partial T}\right)_P = -S^{\mathrm{EX}} \tag{10.28}$$

$$\left(\frac{\partial H^{\mathrm{EX}}}{\partial T}\right)_P = C_P{}^{\mathrm{EX}} \tag{10.29}$$

and so on.

The total excess enthalpy is also called the *relative enthalpy*, **L**, and is related to the excess total Gibbs energy by

$$\begin{aligned}\mathbf{L} &= \mathbf{H}^{\mathrm{EX}} \\ &= \mathbf{G}^{\mathrm{EX}} + T\,\mathbf{S}^{\mathrm{EX}} \\ &= \mathbf{G}^{\mathrm{EX}} - T\left(\frac{\partial \mathbf{G}^{\mathrm{EX}}}{\partial T}\right)_{P,m} && (10.30) \\ &= -T^2\left(\frac{\partial \mathbf{G}^{\mathrm{EX}}/T}{\partial T}\right)_{P,m} && (10.31)\end{aligned}$$

The total excess heat capacity, the *relative heat capacity*, also gets its own symbol

$$\mathbf{J} = \mathbf{C}_{\boldsymbol{P}}{}^{\mathrm{EX}} \tag{10.32}$$

$$= \left(\frac{\partial \mathbf{H}^{\mathrm{EX}}}{\partial T}\right)_{P,m} \tag{10.33}$$

For entropy,

$$S^{\mathrm{EX}} = S_{\text{real sol'n}} - S_{\text{ideal sol'n}} \tag{10.34}$$

$$= \Delta_{\mathrm{mix}} S_{\text{real sol'n}} - \Delta_{\mathrm{mix}} S_{\text{ideal sol'n}}$$

$$= \Delta_{\mathrm{mix}} S_{\text{real sol'n}} + R \sum_i x_i \ln x_i \tag{10.35}$$

Excess total entropy is also related to the relative enthalpy and excess total Gibbs energy as

$$\mathbf{S}^{\mathrm{EX}} = (\mathbf{L} - \mathbf{G}^{\mathrm{EX}})/T \tag{10.36}$$

For Gibbs energy,

$$G^{\mathrm{EX}} = G_{\text{real sol'n}} - G_{\text{ideal sol'n}}$$

$$= \Delta_{\mathrm{mix}} G_{\text{real sol'n}} - \Delta_{\mathrm{mix}} G_{\text{ideal sol'n}}$$

$$= \Delta_{\mathrm{mix}} G_{\text{real sol'n}} - RT \sum_i x_i \ln x_i \tag{10.37}$$

Partial molar excess properties

These have the same meaning as other partial molar properties. The general formulation then is

$$\left(\frac{\partial \mathbf{Z}^{\mathrm{EX}}}{\partial n_i}\right)_{\hat{n}_i} = \overline{Z}_i^{\mathrm{EX}}$$

$$= \left(\frac{\partial \mathbf{Z}^{\mathrm{real}}}{\partial n_i}\right) - \left(\frac{\partial \mathbf{Z}^{\mathrm{ideal}}}{\partial n_i}\right)$$

$$= \overline{Z}_i^{\mathrm{real}} - \overline{Z}_i^{\mathrm{ideal}}$$

For volume, using $i = 2$ to indicate the solute, this becomes

$$\overline{V}_2^{\mathrm{EX}} = \overline{V}_2^{\mathrm{real}} - \overline{V}_2^{\mathrm{ideal}}$$

$$= \overline{V}_2^{\mathrm{real}} - \overline{V}_2^{\infty}$$

superscript “real” is not generally needed, so we have

$$\overline{V}_2^{\mathrm{EX}} = \overline{V}_2 - \overline{V}_2^{\infty} \tag{10.38}$$

where the substitution $\overline{V}_2^{\mathrm{ideal}} = \overline{V}_2^{\infty}$ means that we are using the infinitely dilute solution of the solute as the ideal solution. It is therefore a Henryan sort of ideal solution. Similarly for enthalpy and heat capacity, the only difference being that they get special symbols,

$$\overline{H}_2^{\mathrm{EX}} = \overline{H}_2 - \overline{H}_2^{\infty} \tag{10.39}$$

$$= \overline{L}_2 \tag{10.40}$$

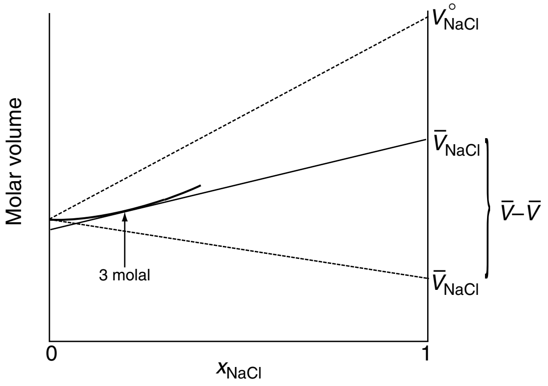


Figure 10.7 Schematic illustration of the meaning of $\overline{V}^{\text{EX}}$. The diagram is a modification of Figure 10.5 but with the dilute region of the x_{NaCl} scale greatly exaggerated and not to scale.

$$\overline{C_{P2}}^{\text{EX}} = \overline{C_{P2}} - \overline{C_{P2}}^{\infty} \tag{10.41}$$

$$= \overline{J}_2 \tag{10.42}$$

The meanings of these terms is illustrated in Figures 10.7 and 10.8. In Figure 10.5 we see that in a 3 molal solution of NaCl in water, $\overline{V}_{\text{NaCl}} = 21.67\,\text{cm}^3\,\text{mol}^{-1}$, and $\overline{V}^{\infty} = 16.62\,\text{cm}^3\,\text{mol}^{-1}$, so that in Figure 10.7, $\overline{V} - \overline{V}^{\infty} = 5.05\,\text{cm}^3\,\text{mol}^{-1}$, but note that we know this difference because we know both $\overline{V}$ and $\overline{V}^{\infty}$.

In Figure 10.8, on the other hand, we know that $\overline{L}_{\text{NaCl}} = \overline{H} - \overline{H}^{\infty} = 45\,\text{J}\,\text{mol}^{-1}$, but although we can measure $\overline{L}$ (see below), we don't know either $\overline{H}$ or $\overline{H}^{\infty}$. When we can only measure differences, the standard state becomes important. As shown, $\overline{H} - \overline{H}^{\infty}$ is negative, meaning that $\Delta_{\text{mix}}H$ is positive (H of the solution is greater than the combined H of the pure phases), and so dissolution of NaCl is endothermic (absorbs heat). Entropy and Gibbs energy

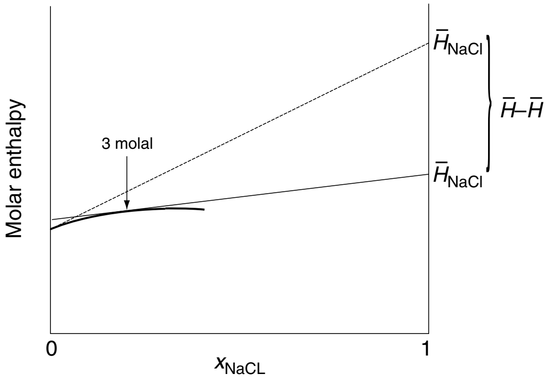


Figure 10.8 Schematic illustration of the meaning of $\overline{H}^{\text{EX}}$.

on the other hand do not get special symbols.

$$\overline{S}^{\text{EX}} = \overline{S} - \overline{S}^{\text{ideal}} \tag{10.43}$$

and is not measured but calculated from other quantities, and

$$\begin{aligned}\overline{G}^{\text{EX}} &= \mu^{\text{EX}} \\ &= \overline{G} - \overline{G}^{\text{ideal}} \\ &= \mu - \mu^{\text{ideal}}\end{aligned} \tag{10.44}$$

But there are some important differences besides the lack of special symbols.

1. We write superscript ° rather than ∞ because the standard state for these quantities is not the infinitely dilute solution, and it is not always the pure phase, and
2. Gibbs energy has a completely different functional relationship to concentration than do the other properties. They all use the *apparent* properties, which we defined in §10.2.3, whereas with Gibbs energy we use Equations (8.30).

For aqueous solutions, it is more useful to write

$$\begin{aligned}\overline{G}_i^{\text{EX}} &= \mu_i^{\text{EX}} \\ &= \mu_i - \mu_i^{\text{ideal}} \\ &= [\mu_i^\circ + RT\ln(m_i\gamma_i)] - [\mu_i^\circ + RT\ln m_i] \\ &= RT\ln\gamma_i\end{aligned} \tag{10.45}$$

or

$$\frac{\mu_i^{\text{EX}}}{RT} = \ln\gamma_i \tag{10.46}$$

or, as it often appears,

$$\frac{\partial \mathbf{G}^{\text{EX}}/RT}{\partial n_i} = \ln\gamma_i \tag{10.47}$$

Another expression for excess Gibbs energies

The molar Gibbs energy of ideal mixing was given as in Chapter 7 as

$$\Delta_{\text{mix}}G_{\text{ideal sol'n}} = RT\sum_i x_i\ln x_i \tag*{[7.20]}$$

It will be useful to have an expression for the Gibbs energy of mixing of real solutions. This is a bit more complicated for aqueous systems which are "unsymmetrical," that is, which have different standard states for the solvent and solute, and a completely different method of expressing deviations from ideality – osmotic coefficients for the solvent, and Henryan activity coefficients for the solutes. This development follows Pitzer (1991).

To start, think about Figure 7.4, but instead of ideal mixing, the mixing is for a real case, and the curve has any shape. Whatever the shape of the mixing curve, the chemical potentials of the two components are still given by the tangent intercepts, and the molar Gibbs energy of mixing of a real solution can be expressed as

$$\Delta_{\text{mix}} G = x_A(\mu_A - \mu_A^\circ) + x_B(\mu_B - \mu_B^\circ)$$

Because we will need excess chemical potentials as well, we change to the total form of the Gibbs energy (recall that chemical potentials are derivatives of the total Gibbs energy **G**, not of the molar Gibbs energy G). Multiplying both sides by $(n_A + n_B)$, we get

$$\Delta_{\text{mix}} \mathbf{G} = n_A(\mu_A - \mu_A^\circ) + n_B(\mu_B - \mu_B^\circ)$$

or, where component A is H_2O and there are several solutes B,

$$\begin{aligned}\Delta_{\text{mix}} \mathbf{G} &= n_{H_2O}(\mu_{H_2O} - \mu_{H_2O}^\circ) + \sum_i n_i(\mu_i - \mu_i^\circ) \\ &= RT\left(n_{H_2O} \ln a_{H_2O} + \sum_i n_i \ln a_i\right) \qquad (10.48)\end{aligned}$$

If the number of moles of water in a system is n_{H_2O}, the number of kilograms of water is

$$\text{kg}_w = \frac{n_{H_2O}}{55.51}$$

and for n_i moles of solute i, n_i/kg_w is the number of moles of i per kg of water, i.e., the molality, m_i. Substituting this relation, the relation $a_i = m_i\gamma_i$, and Equation (8.37)

$$\phi = -\frac{55.51}{\sum_i m_i} \ln a_{H_2O} \qquad [8.37]$$

into (10.48), we get

$$\Delta_{\text{mix}} \mathbf{G} = RT \sum_i n_i[-\phi + \ln(m_i\gamma_i)] \qquad (10.49)$$

Next we divide this into an ideal part containing only concentration terms and another part containing the correction terms ϕ and γ,

$$\Delta_{\text{mix}} \mathbf{G} = \underbrace{-RT \sum_i n_i(1 - \ln m_i)}_{\text{ideal part}} + \underbrace{RT \sum_i n_i(1 - \phi + \ln \gamma_i)}_{\text{corrective part}}$$

The "corrective part" is evidently the excess Gibbs energy, so

$$\begin{aligned}\mathbf{G}^{\text{EX}} &= \Delta_{\text{mix}} \mathbf{G} + RT \sum_i n_i(1 - \ln m_i) \qquad (10.50) \\ &= RT \sum_i n_i(1 - \phi + \ln \gamma_i), \qquad (10.51)\end{aligned}$$

and

$$\mathbf{G}^{\mathrm{EX}}/\mathrm{kg_w} = RT\sum_i m_i(1-\phi+\ln\gamma_i) \tag{10.52}$$

and, as in (10.47)

$$\left[\frac{\partial \mathbf{G}^{\mathrm{EX}}/(\mathrm{kg_w}RT)}{\partial m_i}\right]_{n_{\mathrm{H_2O}}} = \ln\gamma_i \tag{10.53}$$

[Pitzer 1973, Equation (23); 1991, Equation (34)] but now we also have an expression for ϕ in terms of $\mathbf{G}^{\mathrm{EX}}$,

$$\phi - 1 = -\frac{(\partial \mathbf{G}^{\mathrm{EX}}/\partial \mathrm{kg_w})_{n_i}}{RT\sum_i m_i} \tag{10.54}$$

[Pitzer 1973, Equation (22); 1991, Equation (35)].

Because Gibbs energy as a function of T and P is a fundamental equation (§4.12.1), excess Gibbs energies can be used to calculate not only activity coefficients but all other deviations from ideal behavior, such as osmotic coefficients, excess enthalpies, excess heat capacities, excess volumes, and so on (§10.4), so it is potentially quite informative. Also, properties calculated from $\mathbf{G}^{\mathrm{EX}}$ in this way will be entirely self-consistent, which might not be the case if each property was determined separately.

However, determining $\mathbf{G}^{\mathrm{EX}}$ for multicomponent electrolyte aqueous solutions has proved to be a difficult task. Virtually all applications these days use the formulation of K.S. Pitzer and his colleagues, developed during the 1970s and 1980s, which we discuss in Chapter 15. Equation (10.52) and its derivatives are the essential first step in the development of the Pitzer equations.

10.5 Enthalpy and heat capacity

Starting with Equation (10.7) as applied to enthalpy rather than volume, and adopting the usual convention where component 1 is the solvent and component 2 the solute,[6]

$$\mathbf{H} = n_1\overline{H}_1 + n_2\overline{H}_2 \tag{10.55}$$

where the partial molar enthalpies are of course

$$\overline{H}_1 = \left(\frac{\partial \mathbf{H}}{\partial n_1}\right)_{T,P,n_2}$$

$$\overline{H}_2 = \left(\frac{\partial \mathbf{H}}{\partial n_2}\right)_{T,P,n_1}$$

---

[6] Equation (10.55) is self-evident to someone familiar with Euler's theorem (§C.2.3).

And, as with the other partial molar properties (except μ), it is convenient to use apparent properties, so from Equation (10.9),

$$^{\phi}H_2 = \frac{\mathbf{H} - n_1 H_1^\circ}{n_2} \tag{10.56}$$

and

$$^{\phi}H_2^\infty = \frac{\mathbf{H}^\infty - n_1 H_1^\circ}{n_2} \tag{10.57}$$

where H_1° is the molar enthalpy of pure water. As before, the apparent and partial molar enthalpies are related by

$$\overline{H}_2 = n_2 \left(\frac{\partial^{\phi}H_2}{\partial n_2} \right)_{T,P,n_1} + {}^{\phi}H_2 \tag{10.58}$$

Equation (10.55) remains true as $n_2 \to 0$, so we can write

$$\begin{aligned} \mathbf{H}^\infty &= n_1 \overline{H}_1^\infty + n_2 \overline{H}_2^\infty \\ &= n_1 H_1^\circ + n_2 \overline{H}_2^\infty \end{aligned} \tag{10.59}$$

where we can write H_1° in place of $\overline{H}_1^\infty$ because infinitely dilute component 1 is just pure water, and partial molar properties are the same as molar properties for pure phases.

Subtracting (10.59) from (10.55),

$$\mathbf{H} - \mathbf{H}^\infty = n_1(\overline{H}_1 - H_1^\circ) + n_2(\overline{H}_2 - \overline{H}_2^\infty) \tag{10.60}$$

$$= n_2({}^{\phi}H_2 - {}^{\phi}H_2^\infty) \tag{10.61}$$

Now if we define

$$\mathbf{L} = \mathbf{H} - \mathbf{H}^\infty \tag{10.62}$$

$$\overline{L}_1 = \overline{H}_1 - H_1^\circ \tag{10.63}$$

$$\overline{L}_2 = \overline{H}_2 - \overline{H}_2^\infty \tag{10.64}$$

$$^{\phi}L_2 = {}^{\phi}H_2 - {}^{\phi}H_2^\infty \tag{10.65}$$

we have

$$\mathbf{L} = n_1 \overline{L}_1 + n_2 \overline{L}_2 \tag{10.66}$$

$$= n_2 {}^{\phi}L_2, \tag{10.67}$$

so

$$^{\phi}L_2 = \frac{\mathbf{L}}{n_2} \tag{10.68}$$

where **L** is called the *relative enthalpy* of the solution, $\overline{L}_2$ is the *relative partial molar enthalpy* and ${}^{\phi}L_2$ is the *apparent relative molar enthalpy* of component 2. Equation (10.66) is clearly analogous to (10.7) and (10.55). At infinite dilution, both $\overline{L}_1$ and $\overline{L}_2$ are zero. Equation (10.62) can also be written

$$\begin{aligned}\mathbf{L} &= \mathbf{H} - \mathbf{H}_1^{\circ}\\ &= \mathbf{H} - n_1 H_1^{\circ}\end{aligned} \tag{10.69}$$

In other words **L** is the difference between the total enthalpy of a solution and the total enthalpy of the same amount of pure water. Differentiating Equation (10.67) with respect to n_2 we get a familiar expression for $\overline{L}_2$,

$$\begin{aligned}\overline{L}_2 &= \frac{\partial \mathbf{L}}{\partial n_2}\\ &= n_2\left(\frac{\partial {}^{\phi}L_2}{\partial n_2}\right)_{T,P,n_1} + {}^{\phi}L_2\end{aligned} \tag{10.70}$$

and, as with volume, using $\sqrt{m}$ is often advantageous, so Equation (10.70) becomes

$$\overline{L}_2 = \tfrac{1}{2}m^{1/2}\left(\frac{\partial {}^{\phi}L_2}{\partial m^{1/2}}\right) + {}^{\phi}L_2 \tag{10.71}$$

An expression for $\overline{L}_1$ can be obtained by substituting Equations (10.70) and (10.67) into (10.66),

$$\begin{aligned}n_1\overline{L}_1 &= n_2{}^{\phi}L_2 - n_2\overline{L}_2\\ &= n_2{}^{\phi}L_2 - n_2\left[{}^{\phi}L_2 + n_2\left(\frac{\partial {}^{\phi}L_2}{\partial n_2}\right)_{n_1}\right]\\ \overline{L}_1 &= -\frac{n_2^2}{n_1}\left(\frac{\partial {}^{\phi}L_2}{\partial n_2}\right)\end{aligned} \tag{10.72}$$

Values of $\overline{L}_2$ can be obtained by measuring heats of dilution. The meaning of some of these terms is illustrated in Figure 10.9.

10.5.1 Heat of dilution

When a solution of initial molality m_{i} is diluted to a final molality m_f, the observed heat effect is

$$\begin{aligned}\Delta_{\mathrm{dilution}}H &= {}^{\phi}L_2(m_f) - {}^{\phi}L_2(m_{\mathrm{i}})\\ &= \Delta{}^{\phi}L_2\end{aligned} \tag{10.73}$$

which, if you insert the definitions of ${}^{\phi}L_2$, and ${}^{\phi}H_2$, means that

$$\Delta_{\mathrm{dilution}}H = \frac{\mathbf{H}_f - \mathbf{H}_{\mathrm{i}}}{n_2}$$

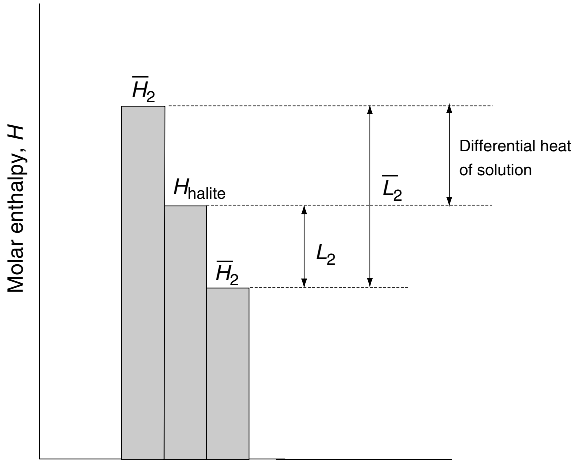


Figure 10.9 Schematic illustration of the meaning of some enthalpy terms.

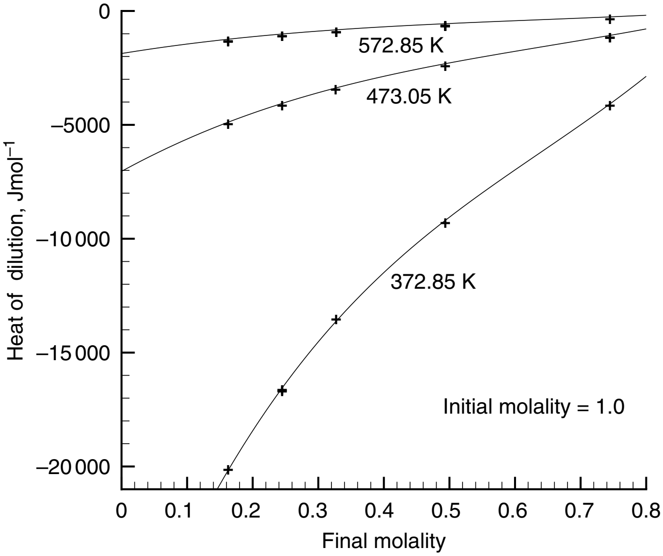


Figure 10.10 The heat of dilution of NaCl from 1 molal to the final molality shown, at the indicated temperatures. Data from Busey et al. (1984).

or, if you prefer,

$$\Delta_{\text{dilution}}\mathbf{H} = \mathbf{H}_f - \mathbf{H}_\text{i} \tag{10.74}$$

Some experimental heat of dilution data for NaCl are shown in Figure 10.10. The negative values show that the enthalpy of a dilute NaCl solution is less than that of a more concentrated one. The data in this diagram are an answer to the question "what is the (nonideal) heat of mixing of NaCl and water?" In this case the NaCl is at two different concentrations – one is 1.0 molal, and the other is as shown on the *x*-axis.

After adjusting to a common pressure, the data were fitted with a Pitzer equation, the slopes determined, and Equation (10.71) used to calculate values of $\overline{L}_2$, which are shown in Figure 10.11.

Another question would be "what is the heat of mixing of water and solid halite?"

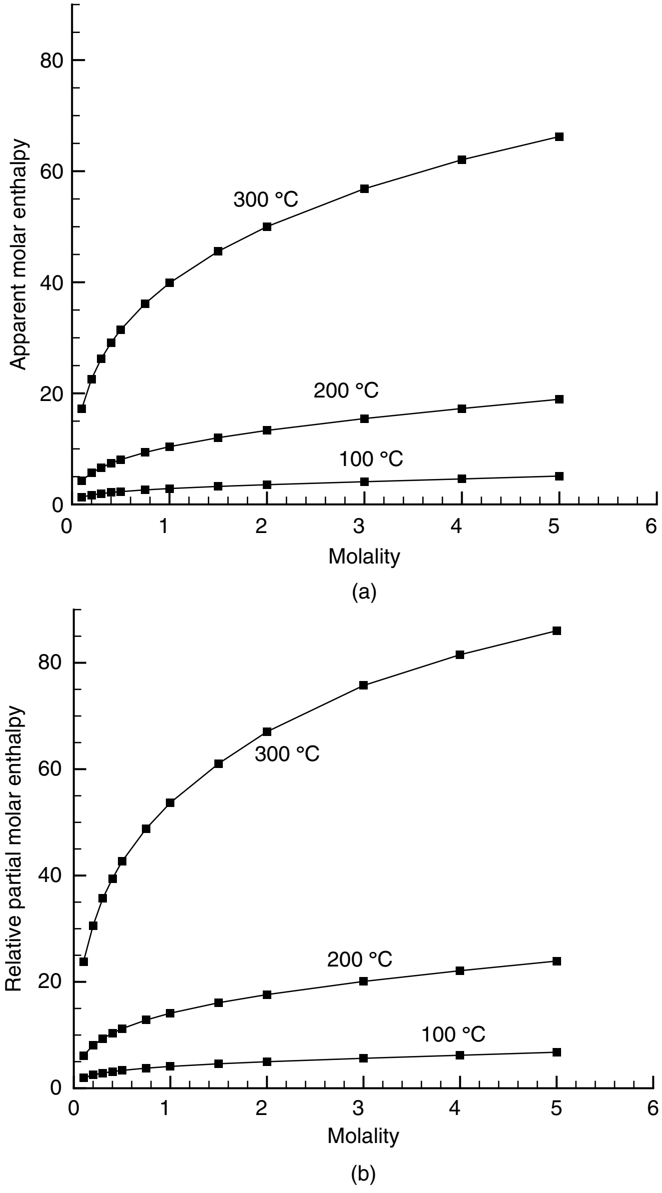


Figure 10.11 (a) The apparent relative molar enthalpy of NaCl (${}^{\phi}L_2$), from the data in Figure 10.10. (b) The relative partial molar enthalpy ($\overline{L}_2$) of NaCl from the data in Figure 10.10. Data from Busey et al.(1984).

10.5.2 Integral heat of solution

The dissolution of halite in water can be written as

$$n_2\,\mathrm{NaCl}(s) + n_1\,\mathrm{H_2O}(l) = [\text{solution of } n_2 \text{ moles of NaCl and } n_1 \text{ moles of H}_2\text{O}] \tag{10.75}$$

or alternatively

$$n_2\,\mathrm{NaCl}(s) + 55.51\,\mathrm{H_2O}(l) = n_2\,\mathrm{NaCl}(aq, m) \tag{10.76}$$

where (aq, m) signifies an aqueous solution of concentration m.

This process of forming a solution from pure solvent and pure solute is called the *integral heat of solution*, and the heat effect in this process is

$$\begin{aligned}
\Delta_{\text{integral}}\mathbf{H} &= \Delta_{\text{mix}}\mathbf{H} \\
&= \underbrace{(n_1\overline{H}_1 + n_2\overline{H}_2)}_{\mathbf{H}\text{ solution}} - \underbrace{(n_1H_1^\circ + n_2H^{\text{halite}})}_{\mathbf{H}\text{ pure phases}} \\
&= n_1(\overline{H}_1 - H_1^\circ) + n_2(\overline{H}_2 - H^{\text{halite}}) && (10.77) \\
&= n_1\overline{L}_1 + n_2(\overline{H}_2 - \overline{H}_2^\infty) - n_2(H^{\text{halite}} - \overline{H}_2^\infty) \\
&= \underbrace{n_1\overline{L}_1 + n_2\overline{L}_2}_{\mathbf{L}} - n_2(H^{\text{halite}} - \overline{H}_2^\infty) && (10.78)
\end{aligned}$$

where $(n_1\overline{H}_1 + n_2\overline{H}_2)$ is the absolute enthalpy of the solution itself (Equation 10.55), $\mathbf{H}^{\text{solution}}$, and $(n_1H_1^\circ + n_2H^{\text{halite}})$ is the absolute enthalpy of the pure starting materials, water and Halite. It is actually a form of Equation (7.13), except that we write $\mathbf{H}^{\text{solution}}$ and mole numbers rather than $H^{\text{ideal sol'n}}$ and mole fractions, and it is a form of Equation (10.60) but using a different standard state for component 2. Many sources introduce another symbol for $(H^{\text{halite}} - \overline{H}_2^\infty)$, such as $\overline{L}^{\text{halite}}$, or $\overline{L}(s)$, but we will retain $(H^{\text{halite}} - \overline{H}_2^\infty)$ in a small and perhaps futile attempt to lessen confusion.

Integral heats of solution cannot be obtained directly from heat of dilution measurements, because they do not include the energy required to break the solid into its constituent ions, and then hydrate them. This is shown by the fact that a term for H^{halite} appears in our equations.

Dividing through by n_2 in Equation (10.78) we get a more intuitive expression for the integral heat of solution:

$$\frac{\Delta_{\text{integral}}\mathbf{H}}{n_2} = \frac{\mathbf{L}}{n_2} - (H^{\text{halite}} - \overline{H}_2^\infty),$$

or

$$\Delta_{\text{integral}}H = {}^{\phi}L_2 + (\overline{H}_2^\infty - H^{\text{halite}}) \tag{10.79}$$

In this equation $(\overline{H}_2^\infty - H^{\text{halite}})$ is the heat of solution of halite into an infinite amount of water, and ${}^\phi L_2$is the heat difference between this infinitely dilute solution (which is pure water) and the solution at whatever concentration we wish. When the concentration of interest is 1 molal, ${}^\phi L_2$ is given by the intercept of the curves in Figure 10.10 at $m = 0$, and because ${}^\phi L_2$ at infinite dilution is zero, it is shown for all concentrations in Figure 10.11a.

10.5.3 Differential heat of solution

The differential heat of solution is another name for the partial molar enthalpy of solution, i.e., the heat effect of dissolving a mole of solute in an infinite amount of solution of some particular concentration. It's like measuring the heat effect instead of the volume effect in Figure 10.3, and it's just the slope of the integral heat effect. The integral and differential heats of solution become identical at infinite dilution. According to the definition of partial molar properties (Equation 10.8), differentiation of an equation for the integral heat should give us the differential heat, so

$$\left(\frac{\partial \Delta_{\text{integral}}\mathbf{H}}{\partial n_2}\right)_{T,P,n_1} = \left(\frac{\partial \mathbf{L}}{\partial n_2}\right)_{T,P,n_1} + \left(\frac{\partial\, n_2(H^{\text{halite}} - \overline{H}_2^\infty)}{\partial n_2}\right)_{T,P,n_1}$$

$$= \overline{L}_2 + (H^{\text{halite}} - \overline{H}_2^\infty) \tag{10.81}$$

$$= \overline{H}_2 - H^{\text{halite}} \tag{10.82}$$

Klotz (1964, p. 291) gives the following expression for the heat absorbed (in cal mol^{-1}) when m moles of halite are dissolved in 1000 g of H_2O (the integral heat of solution of halite in water) at 25 °C

$$\Delta_{\text{integral}}H = 923\,m + 476.1\,m^{3/2} - 726.1\,m^2 + 243.5\,m^{5/2} \tag{10.83}$$

so the differential heat is

$$\left(\frac{\partial \Delta_{\text{integral}}H}{\partial m}\right)_{T,P} = 923 + 714.15\,m^{1/2} - 1452.2\,m + 608.75\,m^{3/2} \tag{10.84}$$

10.5.4 Heat capacity

Heat capacities can be measured directly in twin tube flow calorimeters (§5.8.2) or by manipulation of other measurements. For example, Criss and Cobble (1961) and Gardner, Mitchell and Cobble (1969) measured the heat of solution of halite in water at various temperatures, extrapolated to standard state conditions (infinite dilution), then determined the temperature derivative of the

Integral heat of solution of NaCl

How much heat is absorbed (or evolved?) when 1 mole of halite is dissolved in 1 kg of water at 100, 150, and 200 °C?

To calculate this we need to know

1. the enthalpy change when a mole of halite is dissolved in an infinite amount of water, and
2. the enthalpy change between this solution and the 1 molal solution, which is the negative of the heat of dilution from 1 *m* to infinite dilution.

This is summed up in Equation (10.79).

Criss and Cobble (1961, 0–95 °C) and Gardner et al. (1961, 100–200 °C) report values of the heat of solution of halite in dilute NaCl solutions, extrapolated to $m_{\text{NaCl}} = 0$, thus giving values of $\overline{H}_2^\infty - H^{\text{halite}}$. If we combine these data we find they are fit very well by a cubic polynomial, which is

$$\overline{H}_2^\infty - H^{\text{halite}} = 1833.66 - 38.8870\,T + 0.016\,1221\,T^2 - 0.000\,803\,184\,T^3 \quad (10.80)$$

where T is in °C.

From the tabulated values in Busey et al. (1984) which were used to plot Figure 10.11a, we find values of ${}^{\phi}L_{\text{NaCl}}$. Combining these data we get

T °C	$\overline{H}_2^\infty - H^{\text{halite}}$ J mol^{-1}	${}^{\phi}L_{\text{NaCl}}$ J mol^{-1}	$\Delta_{\text{integral}}H$ J mol^{-1}
25	+914	⇐ from Criss and Cobble (1960)	
100	−5213	2861	−2352
150	−12898	5731	−7167
200	−24771	10400	−14371

The remarkable decrease in both $\overline{H}_2^\infty - H^{\text{halite}}$ and $\Delta_{\text{integral}}H$ with temperature, including a change from endothermic to exothermic in $\overline{H}_2^\infty - H^{\text{halite}}$ between 25 and 100 °C, is but one of many indications that there is a great increase in energy liberating processes (hydration, solvent collapse, ion-pair formation) as T is increased.

heat of solution measurements. At infinite dilution, because $\overline{H}_1 = H_1^\circ$, Equation (10.77) becomes

$$\Delta_{\text{integral}}\mathbf{H} = n_2(\overline{H}_2^\infty - H^{\text{halite}})$$

or

$$\Delta_{\text{integral}}H = (\overline{H}_2^\infty - H^{\text{halite}}) \quad (10.85)$$

Heat capacity from integral heat of solution

As shown in Equation (10.86), the temperature derivative of an equation for $\overline{H}_2^\infty - H^{\mathrm{halite}}$ will give $\overline{C}_P^\infty - C_P{}^{\mathrm{halite}}$, and adding $C_P{}^{\mathrm{halite}}$ to this will give $\overline{C}_P^\infty$. The temperature derivative of Equation (10.80) is

$$\overline{C}_P^\infty - C_P{}^{\mathrm{halite}} = -38.8870 + 0.032\,2442\,T - 0.002\,409\,552\,T^3$$

where T is in °C and $\overline{C}_P^\infty - C_P{}^{\mathrm{halite}}$ is in $\mathrm{cal\,mol^{-1}\,K^{-1}}$. The heat capacity of halite is given by (NIST web site, Shomate equation)

$$\begin{aligned} C_P{}^{\mathrm{halite}} = {} & 50.723\,89 + 6.672\,267(T/1000) - 2.517\,167(T/1000)^2 \\ & + 10.159\,34(T/1000)^3 - 0.200\,675/(T1000)^2 \end{aligned}$$

where T is in kelvins and $C_P{}^{\mathrm{halite}}$ is in $\mathrm{J\,mol^{-1}\,K^{-1}}$. Converting calories to joules and combining these data, we get

T °C	$\overline{C}_P^\infty - C_P{}^{\mathrm{halite}}$ $\mathrm{J\,mol^{-1}\,K^{-1}}$	$C_P{}^{\mathrm{halite}}$ $\mathrm{J\,mol^{-1}\,K^{-1}}$	$\overline{C}_P^\infty$ $\mathrm{J\,mol^{-1}\,K^{-1}}$
25	−135.28	50.50	−84.78
50	−120.45	51.04	−69.41
75	−118.23	51.51	−66.71
100	−128.61	51.95	−76.66
125	−151.59	52.36	−99.23
150	−187.17	52.75	−134.43
175	−235.36	53.12	−182.24
200	−296.15	53.50	−242.65

These values of $\overline{C}_P^\infty$ are almost identical to those plotted in Figure 10.12. The inverted-U shape of the C_P curve as a function of T is typical of all electrolytes (Helgeson et al. 1984).

and the temperature derivative is

$$\left(\frac{\partial(\overline{H}_2^\infty - H^{\mathrm{halite}})}{\partial T}\right)_P = \overline{C}_P^\infty - C_P{}^{\mathrm{halite}} \tag{10.86}$$

where C_P^∞ is the heat capacity of aqueous NaCl at infinite dilution, and is normally called C_P°. The heat capacity of halite ($C_P{}^{\mathrm{halite}}$) as a function of temperature is well known, so this is added to the result to obtain C_P^∞.

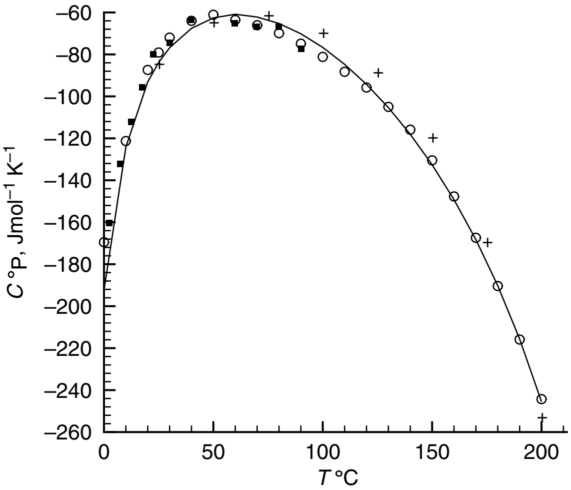


Figure 10.12 Standard state heat capacity of aqueous NaCl as a function of temperature. Squares – Criss and Cobble (1961). Open circles – Gardner, Mitchell and Cobble (1969). Crosses – Helgeson et al. (1981). Line – Pitzer et al. (1984).

Figure 10.12 shows their results, compared with a more recent compilation by Pitzer et al. (1984).

The heat capacity of solution at finite concentrations is analogous to volume, in that they can be measured directly in calorimeters, so that $(\overline{C_P} - \overline{C_P}^{\infty})$ can be evaluated by knowing both quantities, not just the difference. Like volume, $\overline{C_P}^{\infty}$ is evaluated by extrapolating measurements of $\overline{C_P}$ to infinite dilution, or, as above, by manipulating other infinite dilution quantities.

10.6 Gibbs energies

At this point we have a fundamental problem. Given the relationship between Gibbs energies and compositions for ideal solutions we have developed, how do we handle deviations from this behavior? What mathematical form should our equations for nonideality take? There is a variety of approaches for this. The most general is to develop an equation of state, and there is a variety of types of those (Chapter 13). Then there are different approaches for dilute and concentrated solutions, and for electrolytes and nonelectrolytes. In this section we look at some fairly general methods which have been applied to many solid and liquid solutions.

10.6.1 Regular solutions

The simplest form of excess Gibbs energy (Equation 10.44) is illustrated in Figure 10.13. This is a completely symmetrical G^{EX}. The ideal $\Delta_{\mathrm{mix}}G$ curve from Figure 7.4 is shown as a dashed line. The "real" $\Delta_{\mathrm{mix}}G$ curve is the sum of this and G^{EX}, and is therefore also symmetrical. Of course in real solutions both $\Delta_{\mathrm{mix}}G$ and G^{EX} will generally have more complicated shapes, but we can

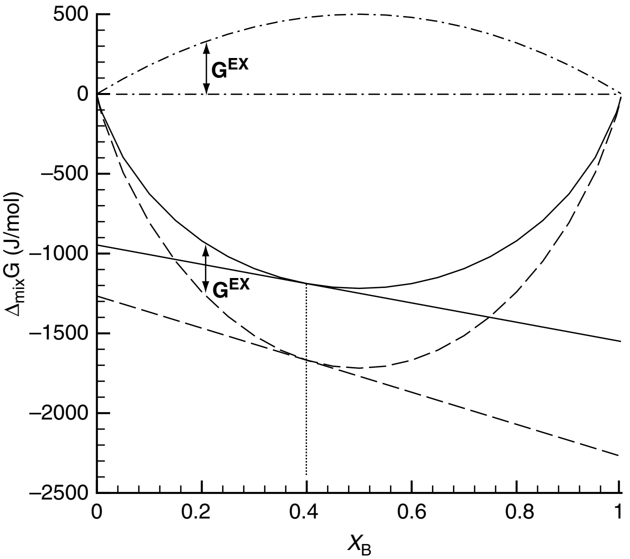


Figure 10.13 The excess Gibbs energy of mixing. The dashed curve is the ideal mixing line from Figure 7.4. The dash-dot curve shows values of G^{EX} for $w_G = 2000\,\text{J mol}^{-1}$ which are added to the ideal mixing values.

use this simple form of nonideal solution behavior to examine the relationship between G^{EX} and several other features of nonideal solutions.

How should we define the simplest possible deviation from ideal solution behavior? It should evidently have about the same parabolic shape as the ideal mixing curve, which means we need some kind of $y \propto x^2$ function. Then because excess properties are zero for pure components, this function should approach zero as the mole fraction, x_{A} or x_{B} of either of the two components, approaches 1.0. Finally, in a solution of two similar components such as benzene and toluene, we might expect the solution to be most nonideal when the components are mixed in equal proportions (because that's when nonuniform interactions between the two species are maximized). That means the equation should have a maximum or minimum at the 1:1 composition. The simplest equation which satisfies all these conditions is

$$G^{\text{EX}} = w_G x_{\text{A}} x_{\text{B}} \tag{10.87}$$

where w_G is a constant with units of energy which is independent of composition, but varies with T and P. If w_G is positive, then molecules A and B repel each other (or prefer energetically to be with molecules of the same type); this can lead to immiscibility and exsolution. If w_G is negative, there is an energy preference for A and B to associate in the solution, and this can produce intermediate compounds. Equation (10.87) is also called a *one-parameter Margules equation*. The equation for the mixing curve in Figure 10.13 is therefore

$$\Delta_{\text{mix}} G = RT\,(x_{\text{A}} \ln x_{\text{A}} + x_{\text{B}} \ln x_{\text{B}}) + w_G x_{\text{A}} x_{\text{B}} \tag{10.88}$$

Hildebrand and Scott (1964) discuss many (mostly binary) systems which can be conveniently treated with this formalism. It has been applied to ternary systems, but beyond that it gets quite cumbersome. Although not many real systems show such behavior (examples are systems like argon–oxygen, benzene–cyclohexane, Au–Cu), we can easily imagine adding more terms involving

Solution theories

There are literally thousands of articles on the theory of solutions, going back to the origins of physical chemistry itself, yet, as pointed out by Prausnitz et al. (1999, Chapter 7) after an extensive survey of the subject, "…we are still far from an adequate theory of liquid mixtures." Traditionally, and for good reason, theories on electrolyte and nonelectrolyte solutions are treated separately, but there is of course some common ground such as the laws of Raoult and Henry. Theories of the two types of solutions also have an interesting similarity in that they have generally been developed emphasizing either the "physical" or the "chemical" approach.

In nonelectrolyte theory, the physical approach goes back to van der Waals, and develops theories which deal with particle arrangements and generalized intermolecular forces, but nothing about the specific nature of any new particles formed due to these forces. The chemical approach concentrates on identifying these new particles, and in the extreme form, begun by Dolezalek in 1908, suggests that if all molecular species are correctly identified, with their respective equilibrium constants, Raoultian mixing is sufficient to account for the solution properties. The physical approach can be traced from van der Waals through van Laar, Scatchard and Hildebrand, resulting in regular solution theory (references in Prausnitz et al., 1999), which turns out to have some overlap with the equations proposed by Margules in 1895.

In electrolyte theory, so different in most respects because of the electrically charged nature of the particles, the same distinction between physical and chemical approaches can be seen. We will present more detail in Chapter 15, but basically the Pitzer equations represent the physical approach, having no reference to individual ionic species, and the HKF model of Helgeson et al. (1981) represents the chemical approach, in which the standard state properties of individual species are defined.

Modern theories of solutions combine aspects of both approaches, and the theories of Pitzer and Helgeson et al. are no different in this respect. The Pitzer approach must take into account strong complex formation, and the HKF approach uses activity coefficients, based on generalized electrostatic considerations, in addition to species information.

adjustable constants and mole fractions to account for more complex nonideality, and this approach has in fact been pursued extensively.

In Figure 10.13 we have assumed $w_G = 2000\,\mathrm{J\,mol^{-1}}$, so that at $x_B = 0.4$, $w_G x_A x_B = 480.0\,\mathrm{J\,mol^{-1}}$, and $\Delta_{\mathrm{mix}}G$ from Equation (10.88) is $-1668.4 + 480.0 = -1188.4\,\mathrm{J\,mol^{-1}}$.

Corrections to ΔH and ΔS

Because $\Delta G = \Delta H - T\Delta S$, applying a correction to ΔG raises the question as to how this correction is divided between ΔH and ΔS. The two extreme cases would be

1. that it all appears as a correction to ΔH with ΔS remaining ideal, or
2. it all appears as a correction to ΔS, with ΔH remaining zero

and, of course, there could be some adjustment in both terms. The original definition of *regular solutions* by Hildebrand in 1927 was case 1, with an ideal entropy of mixing and a nonideal enthalpy of mixing. The other case, retaining $\Delta H = 0$, produces *athermal solutions*. Many other descriptive terms have been used by other authors, and the term "regular solution" now covers both these cases, as well as some extensions.

Because $G^{\mathrm{EX}} = H^{\mathrm{EX}} - TS^{\mathrm{EX}}$, then

$$\begin{aligned} S^{\mathrm{EX}} &= -\frac{\partial G^{\mathrm{EX}}}{\partial T} \\ &= -\frac{\partial}{\partial T}(w_G\, x_A x_B) \\ &= x_A x_B\, w_S \end{aligned} \tag{10.89}$$

where $(\partial w_G/\partial T) = -w_S$, and H^{EX} is

$$\begin{aligned} H^{\mathrm{EX}} &= G^{\mathrm{EX}} + T\,S^{\mathrm{EX}} \\ &= x_A x_B (w_G + T\,w_S) \\ &= x_A x_B\, w_H \end{aligned} \tag{10.90}$$

Similarly, $(\partial w_G/\partial P) = w_V$, $V^{\mathrm{EX}} = x_A x_B w_V$, and

$$w_G = w_H - T w_S + P w_V \tag{10.91}$$

Measurements of any of these excess quantities could then contribute to a knowledge of a system's nonideality as a function of T and P. We give some examples of how this is done in Chapter 14.

$\mu - \mu^\circ$ from tangent

The slope of the tangent to the mixing curve is

$$\begin{aligned}\left(\frac{d\Delta_{\text{mix}}G}{dx_{\text{B}}}\right) &= \frac{d}{dx_{\text{B}}}\left[RT\left(x_{\text{A}}\ln x_{\text{A}} + x_{\text{B}}\ln x_{\text{B}}\right) + w_G x_{\text{A}} x_{\text{B}}\right] \\ &= RT\ln\left(\frac{x_{\text{B}}}{x_{\text{A}}}\right) + w_G\left(1 - 2x_{\text{B}}\right) \qquad (10.92)\end{aligned}$$

and the equation for the tangent line itself is, as before,

$$\mu_{\text{A}} = \Delta_{\text{mix}}G - x_{\text{B}}\left(\frac{d\Delta_{\text{mix}}G}{dx_{\text{B}}}\right) \qquad (10.93)$$

Calculating the intercept on the x_{A} axis of the tangent at $x_{\text{B}} = 0.4$ with these equations, we get $\mu - \mu^\circ = -946\,\text{J mol}^{-1}$.

$\mu - \mu^\circ$ from equation

Substitution of Equation (10.92) into Equation (10.93), remembering that $\Delta_{\text{mix}}G$ is now Equation (10.88), we find that

$$\mu_{\text{A}} = RT\ln x_{\text{A}} + w_G x_{\text{B}}^2$$

but we know that here μ_{A} is really $\mu_{\text{A}} - \mu^\circ$, because we have implicitly assumed that $\mu^\circ = 0$. In other words, the intercept we calculated is the numerical value on the x_{B} axis, but that numerical value is a ΔG of mixing, a difference, so our μ_{A} is also a difference. What we have calculated is in fact

$$\mu_{\text{A}} - \mu_{\text{A}}^\circ = RT\ln x_{\text{A}} + w_G x_{\text{B}}^2 \qquad (10.94)$$

and at $x_{\text{A}} = 0.6$, $x_{\text{B}} = 0.4$, $w_G = 2000\,\text{J mol}^{-1}$, Equation (10.94) gives $\mu_{\text{A}} - \mu_{\text{A}}^\circ = -946\,\text{J mol}^{-1}$, in agreement (of course) with the tangent method.

If in Equation (7.34) we change μ_i to $\mu_{\text{A}}^{\text{ideal}}$, and we compare this with Equation (10.94), we see that

$$\begin{aligned}\mu_{\text{A}} - \mu_{\text{A}}^{\text{ideal}} &= \mu^{\text{EX}} \\ &= w_G x_{\text{B}}^2 \qquad (10.95) \\ &= 2000 \times (0.4)^2 \\ &= 320\,\text{J mol}^{-1} \qquad (10.96) \\ &= -946 + 1266\,\text{J mol}^{-1} \quad (1266 \text{ is } RT\ln x_{\text{A}} \text{ from page 195})\end{aligned}$$

$\mu - \mu^\circ$ from activity coefficient

Also, it proves very convenient to define a correction factor for the mole fraction term in Equation (7.34), such that the nonideal effects are taken into account. In other words, instead of correcting the ΔG or $\mu - \mu^\circ$ side of the equation with G^{EX}, we correct the mole fraction on the other side of the equation to give the same effect (we looked at this approach in more detail in Chapter 8). So we define a factor γ_R such that

$$\begin{aligned}\mu_\text{A} - \mu_\text{A}^\circ &= RT \ln (x_\text{A} \gamma_{R_\text{A}}) \\ &= RT \ln x_\text{A} + RT \ln \gamma_{R_\text{A}}\end{aligned} \tag{10.97}$$

and comparing this with Equation (10.94) we see that

$$\left.\begin{aligned} RT \ln \gamma_{R_\text{A}} &= w_G x_\text{B}^2 \\ RT \ln \gamma_{R_\text{B}} &= w_G x_\text{A}^2 \end{aligned}\right\} \tag{10.98}$$

At $x_\text{B} = 0.4$ and $w_G = 2000\,\text{J mol}^{-1}$, $\gamma_{R_\text{A}} = 1.1378$ and $\mu_\text{A} - \mu_\text{A}^\circ = RT\, ln(0.6 \times 1.1378) = 946\,\text{J mol}^{-1}$ again.

So there is complete equivalence between the use of the activity coefficient and the excess Gibbs energy. In a sense there is hardly any difference at all. If the correction for nonideality is in a logarithmic form ($RT \ln \gamma$) it can be combined with the ideal term ($RT \ln x$) as a correction to x. If it does not have a logarithmic form (wx^2) it becomes a correction to the ΔG or $\Delta\mu$ term.

CO_2 solubility in NaCl solutions

An interesting application of regular solution theory is presented by Nesbitt (1984). He shows that activity coefficients for CO_2 in aqueous NaCl solutions to quite high temperatures ($\approx 500\,°\text{C}$) and NaCl concentrations ($\approx 6\,m$) can be fit very well by a slight modification of (10.98). As written, the activity coefficients in (10.98) are based on Raoultian activities; that is, $\gamma_\text{B} \to 1$ as $x_\text{B} \to 1$. Solubility studies on the other hand normally use Henryan coefficients, where $\gamma_\text{B} \to 1$ as $m_\text{B} \to 0$, where m_B is the molality of the solute.

Following Robinson and Stokes (1965), Nesbitt plots $\ln \gamma_{\text{CO}_2}$ versus $-(x_2^2 - 1)$ instead of x_2^2, where x_2 is the mole fraction of H_2O in the system. This has the desired effect of having $\gamma_{\text{CO}_2} \to 1$ as $x_2 \to 0$. Nesbitt calculates values of w_G and discusses applications to boiling hydrothermal solutions.

Unmixing

As w_G becomes larger, the mixing curve develops a flat portion, and eventually splits into two separate parabolic curves, one on each side of the diagram, as shown in Figure 10.14. In a real system this means that "unmixing" occurs. Instead of one homogeneous solution, two solutions form at equilibrium, the compositions of which are at the minima in the mixing curve. The chemical potentials in each of the two solutions are given by the tangent common to the two minima, showing that μ_A and μ_B are the same in each solution. Generally, the two compositions converge as temperature increases, generating a solvus (a subject continued in Chapter 17). The mixing curve and the solvus are not usually symmetrical as in this simple model. Somewhat more complex equations are required to generate asymmetry (§10.7).

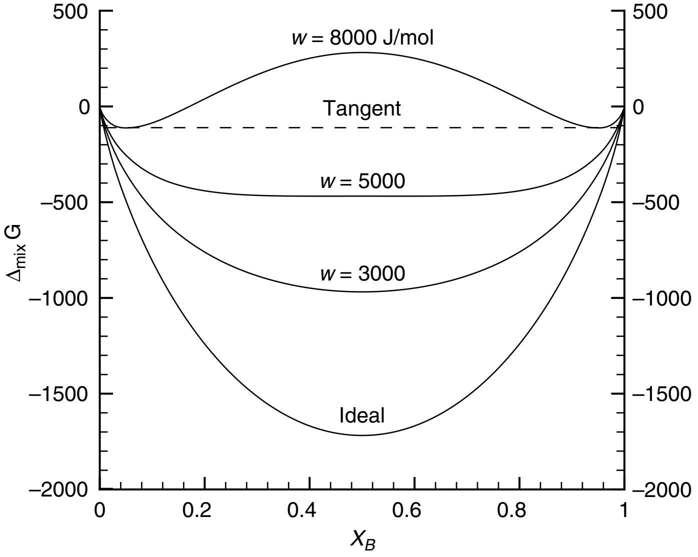


Figure 10.14 The excess Gibbs energy of mixing with various values of w_G, showing unmixing into two separate solutions when $w_G = 8000\,\mathrm{J\,mol^{-1}}$. The dashed line is a tangent common to both minima in this curve.

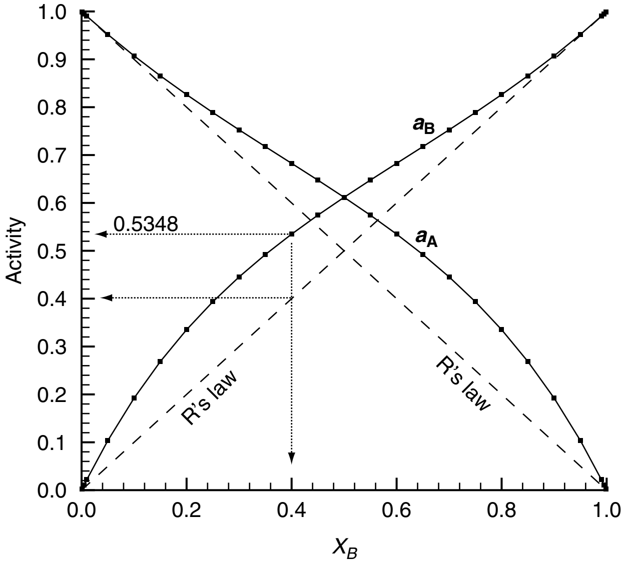


Figure 10.15 The activities of A and B as a function of x_B from Equations (10.98), with $w_G = 2000\,\mathrm{J\,mol^{-1}}$. The same data are shown in Figure 8.2.

10.6.2 The Raoultian activity

The activities of A ($a_A = x_A\gamma_{R_A}$) and B ($a_B = x_B\gamma_{R_B}$) as a function of mole fraction are shown in Figure 10.15 and Table 10.2. At $x_B = 0.4$ and $w_G = 2000\,\mathrm{J\,mol^{-1}}$, $\gamma_{R_B} = 1.337$ (Equation 10.98), so the activity of B is $0.4 \times 1.337 = 0.5348$, as shown.

The Margules equations

Margules (1895) suggested an equation for the vapor pressures of solution components (partial pressures) in the form of a series expansion, using the systems ethyl alchohol–water and methyl alchohol–water as examples. Using our notation, his equation was

$$P_i = P_i^\circ \, x_i^{\alpha_0} e^{[\alpha_1(1-x_i)+\frac{\alpha_2}{2}(1-x_i)^2+\frac{\alpha_3}{3}(1-x_i)^3+\cdots]} \tag{10.99}$$

where i is 1 or 2 (binary solutions, so $x_2 = 1 - x_1$), and where, to have the limiting condition that Henry's law is followed as $x_2 \to 0$, $\alpha_0 = 1$, and $\alpha_1 = 0$. This gives, as $(1 - x_1) \to 0$,

$$P_i = P_i^\circ \, x_i,$$

or

$$P_i = \text{constant} \cdot x_i$$

which is Henry's law (§7.4.2).

If we assume that the partial pressures are in fact fugacities ($f_i = P_i$), then using $a_i = f_i / f_i^\circ$

$$\begin{aligned} \ln \gamma_i &= \ln \frac{a_i}{x_i} \\ &= \sum_{j=2}^{n} \frac{1}{j}[\alpha_j (1 - x_i)^j] \end{aligned} \tag{10.100}$$

which, for a binary solution and $n = 3$, is

$$\ln \gamma_1 = \tfrac{1}{2}\alpha_2 \, x_2^2 + \tfrac{1}{3}\alpha_3 \, x_2^3,$$

and so

$$RT \ln \gamma_1 = \text{constant}_1 \cdot x_2^2 + \text{constant}_2 \cdot x_2^3$$

which is the same in form as Equations (10.105) or (10.107), and if truncated to the first term on the right, is the same as the regular solution expression, Equations (10.98).

There is an extensive literature in both chemistry and geochemistry on these and other similar equations and their interrelations. Useful guides are Grover (1977) and Thompson (1967). They apply to any solutions, solid, liquid, or gaseous. Thompson points out that the Margules equations bear the same relationship to Henry's law that the virial equation (§13.5) has to the ideal gas law, in that the first term in the virial series (in the form of 13.26) is the ideal gas law, and the Margules equation condenses to Henry's law, as shown above.

Table 10.2 *Activities and activity coefficients in a binary regular solution having* $w_G = 2000\,J\,mol^{-1}$.

x_A	x_B	γ_{R_A}	γ_{R_B}	a_A	a_B
0.9	0.1	1.008	1.922	0.907	0.192
0.8	0.2	1.033	1.676	0.826	0.335
0.7	0.3	1.075	1.485	0.753	0.445
0.6	0.4	1.138	1.337	0.683	0.535
0.5	0.5	1.223	1.223	0.612	0.612
0.4	0.6	1.337	1.138	0.535	0.683
0.3	0.7	1.485	1.075	0.445	0.753
0.2	0.8	1.676	1.033	0.335	0.826
0.1	0.9	1.922	1.008	0.192	0.907

The standard state we have just used is pure substance A, because our values of μ_A are measured on the $x_A = 1$ axis, and our values of $(\mu_A - \mu^\circ)$ refer to the difference between the chemical potential of A in the solution and A in its pure form, whatever that is (solid or liquid A). Standard states in other cases, where we will dealing with gases or electrolytes in solution, may not always be so obvious.

It is important to realize that all the features we have discussed – the tangent intercepts, activity coefficients, etc. – operate just the same no matter how complex the mixing curve becomes. They are just easier to discuss with a simple model.

10.7 Margules equations

Because most real solutions do not follow the parabolic, symmetrical form of the two-suffix Margules equations, it is necessary to add more flexibility to the equations by adding more coefficients and more terms. One way to do this is to simply define a second w_G term of the type in (10.87), then combine the two w_G terms in a linear equation in x. This can be imagined as the "mixing" of two solutions, each symmetrical but having different values of w_G:

$$G^{\text{EX}}_{\text{sol'n 1}} = w_{G_1} x_1 x_2 \tag{10.101}$$

$$G^{\text{EX}}_{\text{sol'n 2}} = w_{G_2} x_1 x_2 \tag{10.102}$$

Now mix x_2 moles of the first solution with $x_1 (= 1 - x_2)$ moles of the second, stir well, and make a new solution, which is described by the sum of (10.101) and (10.102):

$$G^{\text{EX}}_{\text{sol'n}} = x_1(w_{G_2} x_1 x_2) + x_2(w_{G_1} x_1 x_2) \tag{10.103}$$

or if you want to carry this process further, to m terms, the general expression is

$$G^{\mathrm{EX}}_{\mathrm{sol'n}} = x_1 x_2 \sum_{i=1}^{m} w_{G_i} x_2^{i-1} \tag{10.104}$$

What we have now is the (weighted) sum of two (or more) parabolic equations to describe a nonideal, asymmetric solution. Because of the second parameter, (10.103) is frequently called a two-parameter Margules equation, and (10.104) can be called the generalized Margules equation. It seems reasonable that this might have an appropriate form because (10.103) still goes to zero as either x_1 or x_2 approaches 1.0 (the excess Gibbs energy must be zero for pure substances); also with two parameters and two parabolas we should be able to fit a distorted parabolic shape, and that's what we might expect the excess Gibbs energy to look like for many real systems. In fact, two-parameter equations of this type do fit experimental data on real systems quite closely, and even work well with minerals that are sufficiently nonideal that they have miscibility gaps (a solvus region).

Activity coefficients can also of course be written in terms of the two-parameter equations. Without derivation, they are

$$\left.\begin{aligned} RT\ln\gamma_1 &= \overline{G}_1^{\mathrm{EX}} \\ &= (2w_{G_2} - w_{G_1})x_2^2 + 2(w_{G_1} - w_{G_2})x_2^3 \\ RT\ln\gamma_2 &= \overline{G}_2^{\mathrm{EX}} \\ &= (2w_{G_1} - w_{G_2})x_1^2 + 2(w_{G_2} - w_{G_1})x_1^3 \end{aligned}\right\} \tag{10.105}$$

A different extension of Equation (10.87) is the Redlich–Kister equation,

$$G^{\mathrm{EX}} = x_1 x_2 \left[A + B(x_1 - x_2) + C(x_2 - x_2)^2 + D(x_1 - x_2)^3 + \cdots\right] \tag{10.106}$$

which has the activity coefficients

$$\left.\begin{aligned} RT\ln\gamma_1 &= a^{(1)}x_2^2 + b^{(1)}x_2^3 + c^{(1)}x_2^4 + \cdots \\ RT\ln\gamma_2 &= a^{(2)}x_1^2 + b^{(2)}x_1^3 + c^{(2)}x_1^4 + \cdots \end{aligned}\right\} \tag{10.107}$$

in which the $a, b, c, \ldots$ parameters are simple combinations of the $A, B, C, \ldots$ parameters in (10.106) (see Prausnitz et al. 1999, p. 228).

These equations (10.103, 10.106) are just a small sample of the many that have been proposed and used in the chemistry and chemical engineering literature.

10.7.1 Margules equations and virial equations

It seems likely that various forms of the Margules equations work well in modeling solutions because they are actually forms of the virial equation (§13.5), which, as we said, has a secure theoretical foundation. To show that the Margules equations have the form of a virial equation, rearrange (10.103) for the total excess free energy of a real, asymmetric solution, remembering that $x_1 = 1 - x_2$, giving

$$G^{\mathrm{EX}} = 0 + w_{G_2}x_2 + (w_{G_1} - 2w_{G_2})x_2^2 + (w_{G_2} - w_{G_1})x_2^3 \tag{10.108}$$

Because excess Gibbs energy is zero for ideal solutions, this has exactly the required form – the first term on the right hand side (zero) is the value of G^{EX} for an ideal solution, and the remaining terms are an ascending polynomial in the variable x_2. Assuming that we can substitute mole fraction for volume (13.24) or pressure (13.26) by virtue of Henry's law, this has the form of a virial equation carried to the fourth coefficient. For symmetric solutions, $w_{G_2} = w_{G_1}$ and (10.108) reduces to

$$G^{\mathrm{EX}} = w_G x_2 - w_G x_2^2 \tag{10.109}$$

This has the form of a virial equation carried only to the third coefficient (which is why it doesn't fit more general, asymmetric solutions).

10.7.2 Margules equations for ternary and higher order systems

In the examples so far, we have used binary (two-component) solutions exclusively. Exactly the same reasoning may be used to derive equations for ternary, quaternary, and even more complex solutions. The equations become more unwieldy with each added component, and the properties they represent cannot be illustrated in two-dimensional graphs. However, the equations can be used in computer programs to generate properties of multicomponent systems for given conditions, and they should continue to fit real solutions adequately for the same reasons discussed above. The main problem with these higher order systems is not in using the Margules equations, but that a lot of experimental data are required if the fit is to be useful, and rather few solid solutions have been sufficiently investigated in this respect.

Rather than writing out all equations for all properties, we will save space by deriving only the Gibbs energy equations for ternary systems. The same approach can then be used to derive equations for other properties or for solutions containing more than three components. First, for a *ternary symmetric solution*, we can follow the analogous procedure used to derive (10.103) and imagine three regular binary solutions with a total of three components (1+2; 1+3; 2+3) being mixed to form a single ternary regular solution (1+2+3). The total excess Gibbs energy is the sum of the one-parameter equations (10.101) for each of these three solutions:

$$G^{\mathrm{EX}}_{\mathrm{sol'n}} = w_{G_{12}}x_1x_2 + w_{G_{23}}x_2x_3 + w_{G_{13}}x_1x_3 \tag{10.110}$$

Similarly, we can derive the analogous equation for a *ternary asymmetric solution* by summing expressions (10.103) for a mixture of three asymmetric binary solutions of the same three components:

$$\begin{aligned} G^{\mathrm{EX}}_{\mathrm{sol'n}} &= \left(w_{G_{23}} x_2^2 x_3 + w_{G_{32}} x_3^2 x_2\right) + \left(w_{G_{13}} x_1^2 x_3 + w_{G_{31}} x_3^2 x_1\right) \\ &\quad + \left(w_{G_{12}} x_1^2 x_2 + w_{G_{21}} x_2^2 x_1\right) \end{aligned} \tag{10.111}$$

The brackets here show the original three asymmetric binary solutions we have "mixed." Derivation of Equation (10.111) by series expansion shows that there is actually a seventh term on the right side, a constant, which is often arbitrarily set to zero.

Notice that (10.111) reduces to (10.110) by setting $w_{G_{ij}} = w_{G_{ji}}$ and that symmetric models therefore have half the number of w coefficients. A slightly more complex version of this equation, where the three terms are weighted in proportion to molar composition, is proposed by Grover (1977, p. 81); this is similar to the weighting technique used to derive the binary, asymmetric equation (10.103) above. Notice that the number of Margules w parameters increases three times relative to a binary system for both the symmetric and asymmetric equations. For a quaternary system they would increase four times, and so on. This means that we are going to need a great many data points in multicomponent systems so as not to over-fit the data. For example, you would not want to fit seven data points in a quaternary system with an eight-parameter Margules equation analogous to (10.111). In practice, this imposes an upper limit on the number of components one might wish to consider if data for a system is limited.

Fortunately, it is not often necessary to consider solid solutions of more than three or four components. Many minerals have fewer than four major components. Minor components do not have a very significant effect on excess properties of the overall solution, as you can see from Equations (10.110) and (10.111) for ternary systems above. For more detail, see Guggenheim (1952), Prigogine and Defay (1954), Saxena (1973), Thompson (1967), and Grover (1977). An algorithm for computer calculation of Margules equations based on higher degree virial equations and applicable to multicomponent systems is given by Berman and Brown (1984).

10.8 Beyond Margules

As useful and as influential as they have been, series expansions like the Margules, regular and virial equations can really only deal effectively with smooth, continuous curves. A high degree of molecular association, leading eventually to compound formation, as well as or combined with order–disorder effects, results in more complex types of curves which cannot be represented by series expansions without carrying the expansion to unmanageably high orders.

Kress (2003) uses the system CO–O_2 to illustrate the problem in a striking fashion. A naive consideration of the energetics of mixing of these two gases would result in a near-zero enthalpy of mixing and an entropy of mixing based on Equation (7.15)

$$\Delta_{\text{mix}} S_{\text{ideal sol'n}} = -R(x_1 \ln x_1 + x_2 \ln x_2) \qquad [7.15]$$

or some slight variation, resulting in a smooth curve of positive $\Delta_{\text{mix}} S$ values. In reality, the reaction

$$\mathrm{CO}(g) + \tfrac{1}{2}\mathrm{O}_2(g) = \mathrm{CO}_2(g)$$

occurs to a very great extent, and the entropy curve between CO and O_2 is curved downward with a cusp at the CO_2 composition, and the enthalpy and Gibbs energy variations are similarly totally different from ideal mixing.

This problem has long been recognized. Kress presents a useful short summary of the many approaches that have been suggested, beginning with Guggenheim's (1952) "quasi-chemical" approach, as well as algorithms allowing computation of species standard state and mixing properties in nonideal associated solutions. The problem extends beyond aqueous solutions to fused salts and silicate melts, and represents a research area that will be active for some time to come.

10.9 The Gibbs–Duhem equation

We must be aware of one very important relationship between solution components, which is that they are not all independent of one another. This seems reasonable enough qualitatively. You can well imagine that changing the concentration, say, of one component of a binary system would have some effect on the activities and activity coefficients of both components, not just one. These changes can be quantified, and this is a highly useful device, because it is very common to measure the activity of only one component in a binary system as a function of concentration, and then calculate the activity of the other component, instead of measuring it too. We mentioned one way of doing this in §5.8.4, the isopiestic method.

This relationship was introduced in §4.14.2. Equation (4.72) from that section is

$$d\mu_1 = -\frac{x_2}{x_1} d\mu_2 \qquad [4.72]$$

showing the interdependence of the chemical potentials in a binary system, but the equation is not used in this form. Now that we have the activity, in the form of Equation (8.19), which is

$$\mu_i = \mu_i^\circ + RT \ln a_i$$

we can substitute this into (4.72) (μ_i° is a constant so $d\mu_i^\circ = 0$, and RT cancels out), obtaining

$$d \ln a_1 = -\frac{x_2}{x_1} d \ln a_2 \tag{10.112}$$

This looks more useful, but it is difficult to integrate because a plot of x_2/x_1 versus $-\ln a_2$ is asymptotic to both axes. If we rewrite the equation with $x_i\gamma_i$ instead of a_i, and note that

$$\begin{aligned} x_1 d \ln x_1 + x_2 d \ln x_2 &= x_1 \frac{dx_1}{x_1} + x_2 \frac{dx_2}{x_2} \\ &= dx_1 + dx_2 \\ &= 0 \end{aligned}$$

it becomes

$$d \ln \gamma_1 = -\frac{x_2}{x_1} d \ln \gamma_2 \tag{10.113}$$

and a plot of x_2/x_1 versus $-\ln \gamma_2$ has only one asymptotic branch (for $x_1 \to 0$). Integrating,

$$\ln \gamma_1'' - \ln \gamma_1' = -\int_{x_1'}^{x_1''} \frac{x_2}{x_1} d \ln \gamma_2 \tag{10.114}$$

or if x'_1 is pure phase 1 (Raoultian standard state), $\ln \gamma_1' = 0$, and

$$\ln \gamma_1 = -\int_{x_1=1}^{x_1} \frac{x_2}{x_1} d \ln \gamma_2 \tag{10.115}$$

showing that if you know $\ln \gamma_2$ over a range of compositions, you can calculate $\ln \gamma_1$ over the same range.

There is a rather large literature on how to perform this integration graphically and analytically. We will show just one very effective method, introduced by Darken and Gurry (1953, Chapter 10); see also Lupis (1983, Chapter 5). If we define

$$\alpha_i = \frac{\ln \gamma_i}{(1 - x_i)^2} \quad (i = 1 \text{ or } 2) \tag{10.116}$$

then

$$\begin{aligned} d \ln \gamma_2 &= d(\alpha_2 x_1^2) \\ &= 2\,\alpha_2 x_1\, dx_1 + x_1^2\, d\alpha_2 \end{aligned}$$

and Equation (10.115) becomes

$$\ln \gamma_1 = -\int_{x_1=1}^{x_1} 2\,\alpha_2 x_2 dx_1 - \int_{x_1=1}^{x_1} x_1 x_2 d\alpha_2 \tag{10.117}$$

Integrating the second integral by parts[7] gives

$$\int_{x_1=1}^{x_1} x_1 x_2 \, d\alpha_2 = \alpha_2 x_1 x_2 - \int_{x_1=1}^{x_1} \alpha_2 \, d(x_1 x_2) \tag{10.118}$$

Combining this with (10.117) then gives, after some manipulation,

$$\ln \gamma_1 = -\alpha_2 x_1 x_2 - \int_{x_1=1}^{x_1} \alpha_2 \, dx_1 \tag{10.119}$$

The last term is quite easy to integrate, because the α function is always finite, even as $x_i \to 0$. That is,

$$\lim_{x_i \to 1} \left[\frac{\ln \gamma_i}{(1 - x_i)^2} \right]$$

is finite, despite the zero squared term in the denominator.

Function α_i in Equation (10.116) looks a lot like our use of w_G in Equation (10.98). In fact, if the solution is strictly regular, in which $\alpha_1 = \alpha_2 =$ a constant, then $\alpha = w_G/RT$, and Equation (10.119) becomes

$$\ln \gamma_1 = -\frac{w_G}{RT} x_1 x_2 - \frac{w_G}{RT}(x_1 - 1) \tag{10.120}$$

and similarly for $\ln \gamma_2$.

10.10 Summary

In previous chapters we have built an impressive theoretical structure, and seen how it could be applied to idealized systems, especially solutions, with a mention of real systems here and there. Eventually, of course, we have to take measurements on real solutions and somehow fit these into our theoretical structure. In this chapter, we start to do that.

We discussed solution volumes at length, because this seems to be the most intuitive way to understand partial molar quantities, and other solution properties. We then broadened the discussion to include enthalpies, heat capacities and Gibbs energies. A persistent theme is the attempt to find equations that will not only fit real data, but which have some relationship to our theory, so that we can extract thermodynamic data from them.

We found that equations for the Gibbs energy have a different form than equations for the other properties, basically because of its pressure derivative relationship to volume, and hence fugacity and activity. We introduced regular solution theory as the simplest way of dealing with deviations from ideal behavior, and then saw that the basic structure could be expanded to more complex solutions in the form of the Margules equations. Regular, Margules and virial equations have proven extremely useful, but they are inadequate in solutions with a high degree of association.

Finally we looked at the Gibbs–Duhem equation, and how it is used in binary systems. We will see some applications of this in Chapter 14.

---

[7] How to integrate by parts is shown in §5.6.4, except that here $u = \alpha_2$ and $v = x_1 x_2$.

11

The phase rule

11.1 Introduction

The phase rule, derived by Gibbs (1875), is a simple relationship between the composition of a system, the number of phases it has, and something called the variance of the system. It requires no thermodynamic data, just compositions, and although simple in principle, and easily applied to the simple systems usually used to explain it, it can be surprisingly difficult to use when considering geological systems. It is absolutely essential in discussing phase diagrams, as in Chapter 17.

In this chapter we will consider not only the "traditional" Gibbs phase rule, but how it becomes modified or extended when aqueous solutes are included in the phase compositions. We then have a look at buffered systems, which are essentially an application of the phase rule.

11.2 Derivation of the phase rule

11.2.1 Some definitions

Phase relations involve a small number of carefully defined terms.

Phases

A *phase* is defined as a homogeneous body of matter having distinct boundaries with adjacent phases, and so is in principle mechanically separable from them. Each mineral in a rock is therefore a single phase, as is a salt solution, or a mixture of gases.

Components

Each phase therefore has a definite chemical composition, and the various phases in a system may have the same (polymorphs) or different compositions. The compositions are described in terms of chemical formulas, such as SiO_2 or $CaMgSiO_3$. The smallest number of chemical formulas needed to describe the composition of all the phases in a system is called the number of *components* of the system. The *choice* of components is to some extent a matter of convenience, but the *number* of components is not. For example, consider the system MgO–SiO_2, which contains the intermediate compounds $MgSiO_3$ (or $MgO \cdot SiO_2$) and Mg_2SiO_4 (or $2\,MgO \cdot SiO_2$). Two formulas are needed to describe the composition of every phase, such as MgO and SiO_2. However, the choice could just as well be $MgSiO_3$ and SiO_2, or $MgSiO_3$ and Mg_2SiO_4, because MgO can be described as [$MgSiO_3$–SiO_2], or [Mg_2SiO_4–$MgSiO_3$], and so on. There are in fact an indefinite number of possible choices in any system, but the *number* that must be chosen is fixed.

To emphasize that it is the *minimum* number of formulas that make up components, consider the olivine solid solutions, i.e., the system Mg_2SiO_4–Fe_2SiO_4. The components could be, and normally are, those two formulas, but could also be any multiple of them, such as $MgSi_{0.5}O_2$ and $FeSi_{0.5}O_2$. Compositions on this join can also be described using the three formulas MgO, FeO and SiO_2, but these are not components of the olivine system, considered by itself. They are components of the olivine–pyroxene system, but not of just olivine or just pyroxene. A glance at Figure 14.5 will help to clarify this. The reason this is worth mentioning is that if the system is Mg_2SiO_4–Fe_2SiO_4, the chemical potentials of MgO, FeO and SiO_2 are undefined, and have no fixed meaning (Spear, 1995, Chapter 6).

System variance or degrees of freedom

To understand the phase rule and how to use it, you must first understand the concept of *variance* or *degrees of freedom*.

A single homogeneous phase such as an aqueous salt (say NaCl) solution has a large number of properties, such as temperature, density, NaCl molality, refractive index, heat capacity, absorption spectra, vapor pressure, conductivity, partial molar entropy of water, partial molar enthalpy of NaCl, ionization constant, osmotic coefficient, ionic strength, and so on. We know, however, that these properties are not all independent of one another. Most chemists know instinctively that a solution of NaCl in water will have all its properties fixed if temperature, pressure, and salt concentration are fixed. In other words, there are apparently three independent variables for this two-component system, or three variables that must be fixed before all variables are fixed. Furthermore, there seems to be no fundamental reason for singling out temperature, pressure, and salt concentration from the dozens of properties available – it's just more convenient; any three would do. The number of variables (system properties) that must be fixed in order to fix *all* system properties is known as the system variance or degrees of freedom.

Now consider two phases at equilibrium, say solid NaCl and a saturated salt solution. Again, intuition or experience tells us that we no longer have three independent variables, but two, because, for example, we cannot choose the composition of the salt solution once T and P are fixed – it is fixed for us by the solubility of NaCl in water. If we then consider the possibility of having a vapor phase in equilibrium with the salt and the solution, we see that we lose another independent variable because we can no longer choose the pressure on the system independently once the temperature is chosen – it is fixed by the vapor pressure of the system. So it would seem that, in general, we restrict the number of independent variables in a system by increasing the number of phases at equilibrium.

Mathematical analogy

The variance of a chemical system is exactly analogous to the variance of a system of linear equations. For example, for the function

$$x+y+z=0$$

if we choose $x=2$, $y=2$, then z is fixed at -4. The equation could be said to have a variance of two, because two variables must be fixed before all variables are fixed. Three variables minus one relationship between them (one equation) leaves two degrees of freedom. If in addition to this function we have another one involving the same variables, such as

$$2x-y+4z=-19$$

we now have three variables and two functional relationships, and we are free to choose only one of the three variables, the other two then being fixed. For example, if we choose $x=2$, then there is no further choice – $y=3$ and $z=-5$. If we choose $x=3$, then $y=2.6$ and $z=-5.6$. This situation can be said to be *univariant* or to have one degree of freedom.

And, of course, if we have a third functional relationship, for example,

$$-3x+2y-7z=35$$

then we have no choice: x, y and z are fixed at 2, 3, and -5, respectively, and the situation is *invariant*.

The reason that the linear equations analogy for phase relationships is so exact is that there is in fact a thermodynamic equation for each phase (see §11.2.2), and each of these equations has a number of independent variables equal to the number of components in the system plus two. And this, in turn, is because each component represents a degree of freedom (we can add or subtract each component), and there are two more because we defined our systems at the beginning as being able to exchange energy in only two ways – heat and one kind of work.[1] If the number of components is c, then the total number of independent system properties is $c+2$. If there are p phases in the

---

[1] If we included other kinds of work in our model, there would be an extra degree of freedom for each.

system, and each phase represents one equation, then there are p equations in $c+2$ variables, or $c+2-p$ degrees of freedom. This is the phase rule:

$$f = c - p + 2 \tag{11.1}$$

where f is the number of degrees of freedom. We have a more detailed look at this derivation below (§11.2.2).

Degrees of freedom can also be described as the number of intensive variables that can be changed (within limits) without changing the number of phases in a system. This point of view is perhaps more useful to someone looking at a phase diagram; thus divariant, univariant, and invariant systems correspond to areas, lines, and points in a *P–T* projection. I prefer, however, to emphasize the fact that coexisting phases reduce the number of independent variables and that some systems have all their properties determined. This fact is very useful in understanding why phase diagrams look the way they do.

11.2.2 Derivation

In Chapter 4 we derived the Gibbs–Duhem equation,

$$0 = \mathbf{S}\,dT - \mathbf{V}\,dP + \sum_{i=1}^{c} n_i d\mu_i \tag*{[4.68]}$$

From this equation we can see the number of independent intensive variables in any homogeneous phase. There are c terms containing μ, i.e., c independent compositional intensive variables, plus two other intensive terms, T and P, for a total of $c+2$ intensive variables. In a single homogeneous phase, these $c+2$ variables are linked by one equation (4.68), so only $c+2-1$ of them are independent. If there are p phases, there are still only $c+2$ intensive variables, because they all have the same value in every phase (at equilibrium), but now there is one equation (4.68) for each phase. Each additional equation reduces the number of independent variables by one, so there are now $c+2-p$ independent intensive variables. These independent intensive variables are called degrees of freedom, f, so

$$f = c - p + 2 \tag{11.2}$$

which is the phase rule. Because we usually consider systems at some fixed values of P and T, this "uses up" two degrees of freedom, so the phase rule becomes

$$f = c - p \tag{11.3}$$

which is sometimes called the "mineralogical phase rule."

To the extent that natural systems approach equilibrium, they obey the phase rule. You might reflect now and then on why natural systems should care about the results of this piece of mathematical reasoning.

11.3 Components and species

Some understanding of the concepts of species and components is essential to setting up geochemical models and interpreting their results, especially if aqueous solutions are involved.

11.3.1 Components and the basis

As discussed above, components are the minimum number of chemical formula units needed to describe the composition of all parts of the system. We say *formulas* rather than *substances* because the chemical formulas need not correspond to any actual compounds. For example, a solution of salt in water has two components, NaCl and H_2O, even if there is a vapor phase and/or a solid phase (ice or halite), because some combination of those two formulas can describe the composition of every phase. Similarly, a mixture of nitrogen and hydrogen needs only two components, such as N_2 and H_2, despite the fact that much of the gas may exist as *species* NH_3. Note that although there is always a wide choice of components for a given system (we could equally well choose N and H as our components, or N_{10} and H_{10}), the *number* of components for a given system is fixed. The components are simply "building blocks," or mathematical entities, with which we are able to describe the bulk composition of any phase in the system. The list of components chosen to represent a system is, in mathematical terms, a basis vector, or simply "the basis."

Components and species

If a gas contains 1 mole of component N_2 and 3 moles of component H_2 ($n_{N_2} = 1$; $n_{H_2} = 3$) at 25 °C, at equilibrium there will be 0.031 77 moles of species N_2, 0.0953 moles of species H_2, and 1.9365 moles of species NH_3 (we know this by solving some simultaneous equations in a process of *speciation*, discussed in Chapter 16). Thus

$$\begin{aligned} n_{\text{component } N_2} &= n_{\text{species } N_2} + \tfrac{1}{2} n_{\text{species } NH_3} \\ &= 0.031\,77 + \tfrac{1}{2} \times 1.9365 \\ &= 1.00 \end{aligned}$$

and

$$\begin{aligned} n_{\text{component } H_2} &= n_{\text{species } H_2} + \tfrac{3}{2} n_{\text{species } NH_3} \\ &= 0.095\,39 + \tfrac{3}{2} \times 1.9365 \\ &= 3.00 \end{aligned}$$

11.3.2 Species

We immediately see a possible confusion between *components* and *species*. Both N_2 and H_2 can exist as physical *species* in the gas, as well as being the chosen abstract *components*. This confusion would not exist if we had chosen N and H as our components, because nitrogen does not exist in the system as N, only as diatomic N_2, and as NH_3. Perhaps unfortunately, components are commonly chosen to be entities which also exist as species, but there is a big difference between, say, component N_2 and species N_2 in the mixture of N_2 and H_2 gases.

Clearly, *component* N_2 represents the total amount of nitrogen in the system, while *species* N_2 represents nitrogen present in that exact stoichiometry in the system. The distinction is important, but normally the meaning of "N_2" is evident from the context.

11.3.3 An alternative basis

Aqueous geochemists, however, are interested not only in the "composition of all parts of the system," but also in the concentrations of all aqueous species in the system, including charged ions. But we cannot describe the concentration of Na^+ or H^+ ions in a salt solution using the formulas NaCl and H_2O: no combination of the formulas NaCl and H_2O will result in the formula Na^+. Furthermore, modelers want not only to "describe" the composition of systems, but to control, or *constrain*, their evolution during some process, such as maintaining equilibrium with some solid or gas phase.

Basis species

To do these things, we need to use "building blocks" different from those we would choose to merely describe bulk compositions as illustrated above. We must use not only a different basis, but a different kind of basis. We could use the elements themselves (Al, B, N,..., etc.), plus the electronic charge, because this would certainly allow us to describe the composition of any species or phase. However, some economy of variables can be achieved by using as "building blocks," or descriptive composition terms, entities which do exist – ordinary charged ions such as HCO_3^- and Na^+. These are called *basis species*, *component species*, or *master species*, and they make up a new kind of basis, which is the minimum number of chemical formulas needed to describe the composition of all phases *and all species, charged and uncharged* in the system. If mineral or gas phases are present, their compositions must also be included in the basis, as described below.

Apart from trivial examples, calculation of species concentrations requires use of computers using a *speciation* program, discussed in Chapter 16. Most such programs have a selection of 30–80 or more basis species, plus a collection of minerals and gases, from which the modeler chooses those required to describe the composition of all aqueous species, gases, and minerals in a particular system. If an element, say rubidium (Rb), does not occur as a basis species (as say, Rb^+) in the database, the program is of course then unable to calculate the amounts of various rubidium species or minerals, even if we have an analysis for the rubidium content of our system.

Auxiliary or secondary species

Having chosen a few basis species to be used in "building" or describing the composition of all other species and phases in the system, all other remaining species are called auxiliary or secondary species. In modeling programs there must, of course, exist some relationship between the two sets of species.

Example Basis species

Suppose a system contains 1 mole of NaCl and 1 kg of water. The components as defined in §11.3.1 are NaCl and H_2O, but the *basis species* needed to describe all the ions present are (in most programs; other choices are always possible): Na^+, Cl^-, H^+, and H_2O. A speciation calculation gives the following results:

	Species	Molality
1	Cl^-	0.990056
2	Na^+	0.990056
3	$NaCl^\circ$	0.009944
4	H^+	1.556E – 07
5	OH^-	1.233E – 07
6	$NaOH^\circ$	3.231E – 08
7	HCl°	6.085E – 14

Note that the composition of all seven actual species can be described by some combination of the four basis species (e.g., $NaOH = Na^+ + H_2O - H^+$), and that each basis species (other than H^+ and H_2O) represents the total amount of some element. Thus

$$\begin{aligned} m_{\text{basis species Cl}^-} &= m_{\text{Cl}^-} + m_{\text{NaCl}^\circ} + m_{\text{HCl}^\circ} \\ &= 0.990056 + 0.009944 + 6.085 \times 10^{-14} \\ &= 1.00 \end{aligned}$$

$$\begin{aligned} m_{\text{basis species Na}^+} &= m_{\text{Na}^+} + m_{\text{NaCl}^\circ} + m_{\text{NaOH}^\circ} \\ &= 0.990056 + 0.009944 + 3.231 \times 10^{-8} \\ &= 1.00 \end{aligned}$$

Normally this consists of the stoichiometry of a secondary species in terms of basis species, plus the equilibrium constant of a reaction linking the two. Similarly, all minerals and gases included in the program must also be linked to the basis species by appropriate reactions and their equilibrium constants.

The distinction, then, between a species which actually exists in the real system, say, the sodium ion Na^+, and the *basis species*, Na^+, is very important. Just as in the nitrogen example above, *component* Na^+ represents the total amount of sodium in the system, and *species* Na^+ represents the sodium actually present as the univalent sodium ion in the solution. Similarly, in the output from the program, basis species and "real" species are commonly mixed together in some way, which is quite clear only if we are perfectly aware of the difference.

An example

For example, the analytical data for sample 912-18 from Merino (1975), referred to earlier, are shown in Table 11.1, and part of the output file from program SOLMIN88 is shown in Table 11.2.

There are 16 species which have numbers in the two "Analyzed" columns, one for each of the species analyzed (Table 11.1) except for Rb, Br, and I, which are ignored by SOLMIN88 because it has no basis species (no data) for these elements. The data for "alkalinity" turn up as data for HCO_3^- (this is under user control during input), and the pH is entered directly in the "$-\log_{10}$ activity" column (omitted from Table 11.2). Note too that, because the basis species for most elements is the same as the component reported in the analysis, the "mg/L" column contains the analytical numbers from Table 11.1. The exception is boron, which is reported in Table 11.1 as 55 $mg\,L^{-1}$ B, but as 314.6 $mg\,L^{-1}$ $B(OH)_3$ in the program output. In other words, the basis species chosen for boron is not B but $B(OH)_3$, and the program has made the conversion by multiplying 55 by the ratio of the molecular weights of $B(OH)_3$ and B; thus $55 \times 61.833/10.811 = 314.6$.

It seems reasonable, then, that the elements actually analyzed appear in the "Analyzed" columns in the program output. What may be confusing is

Table 11.1 *Analysis of sample 912-18 from the Kettleman North Dome oil field (Merino, 1975). Data are in mg L^{-1}.*

Li	1.58	Ca	904	Br	132
Na	9470	Sr	58	I	40.0
K	151	Ba	0.70	alkalinity	1410
Rb	0.376	Fe	1.78	SO_4^{2-}	335
NH_3	44.0	F	0.50	H_2S	2.05
Mg	59	Cl	16100	SiO_2	70
B	55	pH	6.8		

Table 11.2 *Some of the speciation results for sample 912-18 (Merino, 1975), as produced by program* SOLMIN88. *The original printout also includes columns "PPM" in the "ANALYZED" section and "log10 activity" in the "CALCULATED" sections. These have been removed to allow the data to fit the page.*

```
                ---ANALYZED------ -----------CALCULATED-----            ACTIVITY
       SPECIES    MG/L       MOLALITY      PPM         MG/L      MOLALITY    ACTIVITY     COEFF.
       -------  ---------   ---------   ---------   ---------   ---------   ---------    -------
  1 Ca ++        904.0000  0.2323E-01  0.6543E+03  0.6543E+03  0.1681E-01  0.4409E-02    0.2623
  2 Mg ++         59.0000  0.2499E-02  0.5392E+02  0.5392E+02  0.2284E-02  0.6385E-03    0.2796
  3 Na +        9470.0000  0.4242E+00  0.9170E+04  0.9170E+04  0.4108E+00  0.2801E+00    0.6819
  4 K +          151.0000  0.3977E-02  0.1474E+03  0.1474E+03  0.3881E-02  0.2496E-02    0.6432
  5 Cl -       16100.0000  0.4676E+00  0.1557E+05  0.1557E+05  0.4522E+00  0.2909E+00    0.6432
  6 SO4 --       335.0000  0.3591E-02  0.1932E+03  0.1932E+03  0.2071E-02  0.3897E-03    0.1881
  7 HCO3 -      1410.0000  0.2380E-01  0.1202E+04  0.1202E+04  0.2029E-01  0.1417E-01    0.6987
  8 H +                                0.1937E-03  0.1937E-03  0.1979E-06  0.1585E-06    0.8010
  9 OH -                               0.1581E-02  0.1581E-02  0.9575E-07  0.6353E-07    0.6635
 11 H4SiO4                             0.1118E+03  0.1118E+03  0.1198E-02  0.1604E-02    1.3392
 12 SiO2          70.0000  0.1200E-02
 15 Ba ++          0.7000  0.5249E-05  0.6109E+00  0.6109E+00  0.4581E-05  0.1036E-05    0.2261
 18 Fe ++          1.7800  0.3282E-04  0.1682E+01  0.1682E+01  0.3101E-04  0.8133E-05    0.2623
 22 Li +           1.5800  0.2345E-03  0.1577E+01  0.1577E+01  0.2341E-03  0.1734E-03    0.7410
 26 Sr ++         58.0000  0.6817E-03  0.5342E+02  0.5342E+02  0.6278E-03  0.1419E-03    0.2261
 30 F -            0.5000  0.2710E-04  0.4708E+00  0.4708E+00  0.2552E-04  0.1693E-04    0.6635
 31 B(OH)3       314.6000  0.5239E-02  0.3125E+03  0.3125E+03  0.5205E-02  0.5831E-02    1.1204
 32 NH3           44.0000  0.2660E-02
 33 H2S            2.0500  0.6194E-04  0.1023E+01  0.1023E+01  0.3091E-04  0.3185E-04    1.0305
 51 BaCO3                              0.2344E-03  0.2344E-03  0.1223E-08  0.1370E-08    1.1204
 52 BaHCO3                             0.1142E+00  0.1142E+00  0.5930E-06  0.4044E-06    0.6819
 53 BaOH +                             0.4680E-07  0.4680E-07  0.3123E-12  0.2230E-12    0.7140
 54 BaSO4                              0.1667E-01  0.1667E-01  0.7356E-07  0.8241E-07    1.1204
 55 CaCO3                              0.2655E+01  0.2655E+01  0.2732E-04  0.3061E-04    1.1204
 56 CaHCO3 +                           0.1067E+03  0.1067E+03  0.1086E-02  0.8051E-03    0.7410
 57 CaOH +                             0.2893E-03  0.2893E-03  0.5218E-08  0.3866E-08    0.7410
 61 CaSO4                              0.4045E+02  0.4045E+02  0.3060E-03  0.3428E-03    1.1204
 71 FeCl +                             0.9077E-01  0.9077E-01  0.1024E-05  0.6982E-06    0.6819
 72 FeCl2                              0.4662E-09  0.4662E-09  0.3787E-14  0.4243E-14    1.1204
 75 FeOH +                             0.1619E-02  0.1619E-02  0.2288E-07  0.1634E-07    0.7140
 76 Fe(OH)2                            0.6276E-07  0.6276E-07  0.7192E-12  0.8058E-12    1.1204
 77 FeOOH -                            0.2390E-08  0.2390E-08  0.2770E-13  0.1978E-13    0.7140
 78 FeSO4                              0.1123E+00  0.1123E+00  0.7614E-06  0.8531E-06    1.1204
 90 H2SiO4 --                          0.7150E-05  0.7150E-05  0.7825E-10  0.1884E-10    0.2408
 91 H3SiO4 -                           0.2027E+00  0.2027E+00  0.2195E-05  0.1497E-05    0.6819
 97 H2CO3                              0.2704E+03  0.2704E+03  0.4489E-02  0.5029E-02    1.1204
 98 CO3 --                             0.1012E+01  0.1012E+01  0.1737E-04  0.4183E-05    0.2408
281 CaCl +                             0.3628E+03  0.3628E+03  0.4946E-02  0.3373E-02    0.6819
282 CaCl2                              0.5308E+01  0.5308E+01  0.4925E-04  0.5517E-04    1.1204
```

that many of them appear as both "Analyzed" and "Calculated", and that the numbers in these two categories are completely different. For example, Ca^{2+} is analyzed at 904 $mg\,L^{-1}$, but is calculated to be 654.3 $mg\,L^{-1}$. It must be understood that in the "Analyzed" columns, Ca^{2+} represents the basis species chosen for calcium; it is the calcium *component*, which equals the total calcium content of the solution. In the "Calculated" columns, however, Ca^{2+} represents one of the calcium *species* actually present in the solution. Other calcium species are listed further down, and have numbers only in the "Calculated" columns. The sum of the molalities of all Ca-containing species is equal to the input value of 0.023 23 molal.

Another example of this important difference is provided by the results for silica. The analyzed silica content (Table 11.1) is 70 mg L^{-1}, and this appears as the basis species SiO_2. This has been converted by the program into a molality of 0.0012. In many applications, $SiO_2(aq)$ and H_4SiO_4 are used synonymously as the uncharged monomeric aqueous silica *species*. But here, SiO_2 is the total silica and the "Calculated" amount of H_4SiO_4 is 0.001 198, a bit less. This is because the basis species SiO_2 has been split up by the program into the three secondary species H_4SiO_4, $H_3SiO_4^-$, and $H_2SiO_4^{2-}$, seen further down in the results. The sum of the molalities of these three species is equal to the input value, 0.0012 molal.

11.4 Duhem's theorem

11.4.1 Traditional components versus basis species

The Gibbs phase rule deals only with *intensive* variables. It is the same whether you have 1 kg or 100 kg of your system. But if you are interested in how a solution changes composition when halite is precipitated from it, say in modeling evaporation in a closed basin, you need to know how much solution there is and how much halite precipitates. You need an *extensive* phase rule. This was provided by Pierre Duhem, a contemporary of Gibbs, in the form of Duhem's theorem. We will lead into this by looking more closely at our two kinds of components – "traditional" and basis species. Readers familiar with Morel and Hering (1993) will recognize that what we have termed "traditional" components are Morel and Hering's "recipes", and what we refer to as basis species they call simply components.

For the system NaCl–H_2O, we have either the two "traditional" components NaCl and H_2O, which allow us to describe the bulk composition of all phases in this system, or we have the four *basis species* Na^+, Cl^-, H^+, and H_2O, which allow us to describe not only the compositions of the phases but also the concentration of all dissolved species in the system. "Traditional" components and basis species are simply different choices of components, which have different purposes and different descriptive powers. We need more basis species because they are called upon to provide more information.

We discussed the system NaCl–H_2O in §11.2.1 in terms of the traditional components, NaCl and H_2O. If, however, we use basis species as components, we will have more degrees of freedom to deal with. For example, using components NaCl and H_2O, we have no control over the Na/Cl ratio, but using basis species Na^+, Cl^-, H^+, and H_2O, we do – we can specify Na^+ and Cl^- independently – an extra degree of freedom. The principles involved have not changed, so we can use a modified phase rule, but in dealing with aqueous solutions we invariably are at a fixed T and P, so we can start with the mineralogical phase rule (11.3). So instead of $f = c - p$, with c the number of traditional components, it becomes

$$f = b - p \tag{11.4}$$

where b is the number of basis species needed to define the system; and in this case $b = 4$.

Fortunately, this is fairly intuitive. It just says that to define an aqueous solution ($p = 1$) at a given T and P, we have to specify the concentration of each solute element. That is, since H_2O is always one of the basis species, then there are $(b - 1)$ degrees of freedom, which is evidently the number of *solute* basis species (Na^+, Cl^-, and H^+). Each additional phase present fixes the value of one basis species, and hence reduces by one the number of basis species that must be specified.[2]

Charge balance

These basis species are actually not completely independent of one another, as theoretically required, because the sum of the positive and negative charges must be equal – there must be a charge balance in real solutions. However, the charge balance requirement does not really reduce the degrees of freedom. It just means that one of the charged basis species, say Cl^-, should be adjusted to give the balance. The charge balance requirement is thus substituted for the requirement that basis species Cl^- be specified.

11.4.2 The extensive phase rule

Each of the phase rules above is used to "define the equilibrium state," which means that they each relate the number of *properties* (understood to be *intensive variables*) of the system to the number of degrees of freedom. This "defines" the equilibrium state, but it does not define *how much* of the equilibrium state we have. The "equilibrium state" of 1 kg of water saturated with halite is the same whether we have 1 g or 1 kg of halite. But modeling programs commonly want to do more than to define the equilibrium state. They want to dissolve or precipitate phases during processes controlled by the modeler, and to keep track of the masses involved, so as to know when phases should appear or disappear. To do this, the mass of each phase is required, not just its presence or absence. Therefore, an additional piece of information is required for each phase present, or p quantities. Almost invariably, the mass of H_2O is chosen

---

[2] For example, if halite was present at equilibrium, we would specify either the Na^+ or the Cl^- basis species, but not both. The other would be fixed through the solubility product. Thus an additional phase always reduces the number of independent basis species by one.

as 1 kg, so that the concentration of basis species defines the mass of each.[3] If solid or gas phases are specified, the mass is usually also specified. If we count these extra p pieces of data, the *extensive* phase rule becomes

$$f = b \tag{11.5}$$

This assumes T and P have been chosen, so they are not included as degrees of freedom. To include these degrees of freedom, we would write (11.5) as

$$f = b + 2 \tag{11.6}$$

This relationship is also fairly intuitive. Look at it this way. The number of phases is always at least one (a system with no phases is not very interesting). To define a system having only an aqueous solution phase ($p = 1$), we must specify each of the solutes in the water, or $b - 1$ quantities. If there is one mineral in equilibrium with the water ($p = 2$), it controls one basis species, and so reduces b by one, and similarly for all p mineral or gas phases. This is phase rule (11.4). But defining the equilibrium state is not usually enough. We want also to know the mass of each phase, so we need p extra data, giving phase rule (11.5), which says that for any system we need b pieces of information. These b pieces of information are

- the mass of water (almost always 1 kg);
- the mass of each mineral or gas phase;
- the concentration of basis species beyond those controlled by the mineral and gas phases.

An example will help to clarify this (see the box on page 329). However, if you find this confusing, you are not alone, and perhaps it is not terribly important. The traditional Gibbs phase rule is often used in a "paper and pencil" sense by petrologists interpreting some sequence of rocks, but when dealing with electrolyte solutions and basis species, well-established computer programs are generally used (Chapter 18). So although the extensive phase rule does tell you how many pieces of data are required as input to the program, most people do not figure this out before using the program.

If you were starting from scratch, with no program, this information would undoubtedly be useful, but when using well-established programs, it is often more efficient to bypass the phase rule, and rely on error messages from the modeling program to get things right. It is useful, though, to know that the phase rule used or implied in modeling aqueous solutions is somewhat different from the one derived by Gibbs.

---

[3] The unit of concentration used in modeling calculations is invariably molality, or the moles of solute species per kilogram of pure water. Therefore, if the mass of water is fixed at 1 kg, the molality of a species automatically equals the number of moles of the species, which is readily convertible to grams.

Example Phase rules

Consider a system consisting of water and dissolved Al, Si and K. If these solutes are sufficiently concentrated, aluminosilicate minerals will precipitate. Let's say that our T and P have been specified as 25 °C and 1 bar.

- "Traditional" components for this system would be K_2O, Al_2O_3, SiO_2, and H_2O, so $c = 4$. If the solutes are very dilute and we have only one phase (water), phase rule (11.3) says $f = c - 1 = 3$, so we have to specify the concentrations of K, Al and Si to define the system. But if we have three solid phases in equilibrium with the water, such as kaolinite, muscovite, and quartz, then $f = c - p = 0$, and the system is invariant (we don't have to specify anything). However, even though all properties of the system are fixed, including the species in the liquid phase, we are unable to describe or calculate the ionic species using these components.
- Alternatively, we choose as the basis the five basis species (or component species) H_2O, K^+, Al^{3+}, H^+, and $SiO_2(aq)$, thus $b = 5$. With only water present, phase rule (11.4) says that we need to specify $f = 5 - 1 = 4$ concentrations, which could be the total amounts of K, Al, Si, plus the pH. With kaolinite, muscovite and quartz present in any amount at equilibrium, phase rule (11.4) says that $f = 5 - 4 = 1$, so we need to specify only one concentration to be invariant. In either case, i.e., whether we have specified one or four concentration(s), we *are* able to describe and calculate all ionic species present.
- Extensive phase rule (11.5) then says that even though we are invariant, that is, even though we have adequately described the equilibrium state, to make our description more useful we need $f = b = 5$ pieces of information: either (with only water) the four concentrations and the mass of water, or (with water and three minerals) one concentration plus the masses of all four phases.

What Duhem said

The term "extensive phase rule" is more commonly referred to as Duhem's theorem. As expressed by Prigogine and Defay (1965), p. 188, Duhem's theorem says

> Whatever the number of phases, of components or of chemical reactions, the equilibrium state of a closed system, for which we know the initial masses $m_1^\circ \ldots m_c^\circ$, is completely determined by two independent variables.

The big difference from the Gibbs phase rule is of course the inclusion of "the initial masses $m_1^\circ \ldots m_c^\circ$," so that in addition to T and P we need to know the mass of every traditional component, if we are not concerned with ionic speciation, or every basis species, if we are (Equation 11.6).

The phase rule is a very beautiful example of the interface between mathematics and physical reality.

11.5 Buffered systems

A great deal of the power and usefulness of the phase rule in geochemistry comes from its demonstration of which systems are divariant, and which therefore have all their properties fixed at a given T and P. Changing the concentration of any component of a trivariant or multivariant solution will change all the properties of the solution, even if T and P do not change. However, consider a divariant system at fixed values of T and P.

11.5.1 Divariant systems

Divariant systems are of course those that have the same number of components and phases. Because the presence of a certain number of phases fixes all the properties of the system at given T and P, including the compositions of all phases, the system has a certain resistance to any process attempting to change these properties, and it will continue to resist change until the process in question succeeds in decreasing the number of phases. The system is therefore said to be "buffered." For example, a rock consisting of gypsum ($CaSO_4 \cdot 2H_2O$) and anhydrite ($CaSO_4$) has two phases and two components, so its properties are fixed at a given T and P. Thus the activity of water is fixed through the relation

$$\tfrac{1}{2}\mathrm{CaSO_4 \cdot 2H_2O} = \tfrac{1}{2}\mathrm{CaSO_4} + \mathrm{H_2O}$$

for which

$$\begin{aligned} K &= \frac{a^{0.5}_{\mathrm{CaSO_4}}\, a_{\mathrm{H_2O}}}{a^{0.5}_{\mathrm{CaSO_4 \cdot 2H_2O}}} \\ &= a_{\mathrm{H_2O}} \quad \text{when } a_{\mathrm{CaSO_4 \cdot 2H_2O}} = a_{\mathrm{CaSO_4}} = 1 \end{aligned}$$

If the standard states for the solid phases are the pure phases at T and P, their activities are both one, and the equilibrium constant equals the activity of water, i.e., the activity of water is a constant (fixed) as long as (pure) gypsum and anhydrite coexist. Water having some other greater activity introduced along cracks in this rock will tend to change the activity of water in the system, but no change will in fact take place (at least in the model) until all the anhydrite is converted to gypsum, at which point the system is no longer buffered and the water activity can assume a new value fixed by the introduced water.

Systems having fewer than two degrees of freedom (that is, one or zero) are also of course buffered in the same way. A pure component at its melting point for example ($c = 1$, $p = 2$, $f = 1$) is buffered against changes in temperature or pressure. If heat is added, the temperature does not rise until the solid phase has disappeared. Divariant systems at a given T and P are also in principle buffered against changes in T and P, but unless reactions with a large ΔH or ΔV are involved, which generally means a melting or vaporization reaction, the buffering effect will be small. Thus the principle of buffering should not be confused with the buffering capacity, which depends on the reactions involved and the amounts and proportions of the phases. Univariant systems are particularly important in metamorphic petrology where they are known as isograds. Invariant systems are quite rare in nature but at least one is fairly common in the laboratory; the ice–water–water vapor triple point is frequently used in the calibration of thermometers at 273.16 K.

11.5.2 Buffered experimental systems

This property (of resisting change) of systems having two (or fewer) degrees of freedom has been known for over 100 years, so one could be forgiven for being surprised at the great change (it's probably exaggerating to call it a revolution) in geochemical thought and practice brought about when Hans Eugster drew attention to its usefulness in experimental and theoretical work. The original application was a demonstration that the assemblages hematite/magnetite/water, quartz/fayalite/magnetite/water, nickel/nickel oxide/water and others could be used in experimental work to control the oxidation state of experimental systems by separating the "buffering system" (one of the above) from the experimental system by a membrane (platinum) permeable to hydrogen. The buffer system, having the same number of components as phases, as well as having an element (e.g., Fe, Ni) in two different valence states, fixed the properties of each of its phases, and in particular fixed the oxygen and hydrogen fugacities in the water. Typical buffering reactions are

$$\mathrm{H_2O} + 2\mathrm{Fe_3O_4}\ (\text{Magnetite}) = 3\,\mathrm{Fe_2O_3}\ (\text{Hematite}) + \mathrm{H_2}(g)$$

$$\mathrm{Ni}(s) + \mathrm{H_2O} = \mathrm{NiO}\ (\text{Bunsenite}) + \mathrm{H_2}(g)$$

The hydrogen diffused through the capsule walls and controlled the oxidation state of the experimental system, which also included an aqueous phase. This arrangement is illustrated in Figure 11.1.

The buffer system thus "imposes" its oxidation state on the experimental system, and will continue to do so for some time even if the experimental system consumes hydrogen or oxygen in its reactions. Since then, innumerable buffering systems and experimental arrangements have been devised, the details of which, being all based on the same principle, are not of great interest

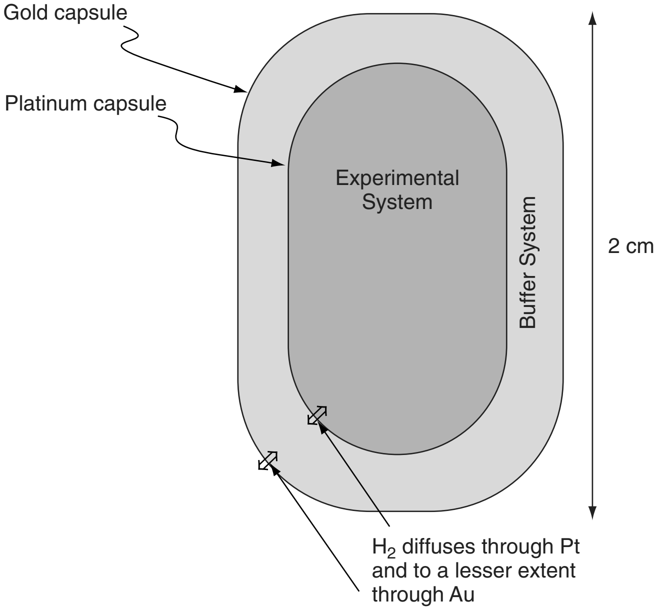


Figure 11.1 Double-capsule technique designed to hold an experimental system in a fixed, known, oxidation state. This arrangement is held in a pressure vessel containing a fluid at high temperature and pressure. The pressure is transmitted to the experimental system by the flexible walls of the gold and platinum tubes. At high temperatures, hydrogen, but not other species, is able to diffuse through the metals.

here. What is interesting is the way that the concept of buffered systems has pervaded geochemistry. It is hardly possible today to discuss a geochemical problem without mentioning a buffering reaction of some sort, whereas before Eugster, such reactions were hardly mentioned, and numerous experiments were performed with results of very limited or zero usefulness because the experiments were not buffered, and the experimental conditions were hence not completely defined.

11.5.3 Buffered natural systems

The buffering capacity of systems with $f \leq 2$ explains only part of their significance in geochemistry. Perhaps even more importance lies in the fact that their properties being fixed, they are thus in principle determinable, and a great deal of effort has been expended in determining these properties. One cannot determine the properties of "granite," only of specific granites, but the (equilibrium) properties of coexisting quartz, albite, and K-spar in the system SiO_2–$NaAlSi_3O_8$–$KAlSi_3O_8$ at a given T, P are fixed and determinable. Buffered systems exist not only in the laboratory, but in nature as well, and other natural systems can be seen to have properties intermediate between those of certain buffered systems, so that the buffering concept provides a framework for thinking about natural systems. For example, hematite and magnetite coexist in many iron formations, and rocks containing magnetite but not hematite or wüstite (FeO) are thought of as lying between the magnetite–hematite and magnetite–wüstite buffers in terms of oxidation state. Similarly syenites containing no quartz and no nepheline can be thought of as lying between the quartz and albite–nepheline buffers for SiO_2.

In view of the fact that the majority of experimental applications of the buffering principle have been in the control of specific gaseous or aqueous species, we should perhaps reiterate that in a buffered system *all* properties are fixed, and that completely solid systems are no exception. Thus for example although the assemblage hematite–magnetite buffers the activity of oxygen through the reaction

$$6Fe_2O_3 = 4Fe_3O_4 + O_2$$

it also buffers the FeO activity through

$$Fe_3O_4 = Fe_2O_3 + FeO$$

the Fe activity through

$$3Fe_3O_4 = 4Fe_2O_3 + Fe$$

the O_3 activity through

$$9Fe_2O_3 = 6Fe_3O_4 + O_3$$

as well as the activity of any other conceivable components of the Fe–O system.

11.5.4 Absolute versus relative buffers

The system having at least as many phases as components can be thought of as a kind of “absolute” buffer. That is, the properties of all phases are absolutely determined once the remaining variables are chosen, and of course once equilibrium is achieved. The system properties, dozens of them, are state variables, are independent of the system’s history, and in principle are determinable, given enough data. We add to this statement once more the cautionary note that in nature a buffering assemblage may not in fact have buffered anything, because of being present in too small a quantity, i.e., the buffering capacity may have been negligible, or because of unfavorable kinetics. Such assemblages are nevertheless always of interest in interpretations of geological history, because they are locations where thermodynamic parameters have a fixed and often known relationship.

Many systems having $f \geq 3$ are also spoken of as buffers. The most common examples are of course the aqueous buffers so commonly used in chemistry, usually to control the pH of a solution. For example, a solution containing sodium acetate and acetic acid will resist changes in pH when an acid or base is added due to the common ion effect. The pH does not change as much as it would if the NaAc and HAc were not there, but it does change, because the system is not an "absolute" buffer in the sense we have been using. Nevertheless, it does resist change and so is buffered. Other systems can be buffered simply due to the large mass involved. Thus the chlorine concentration of the ocean is not controlled by any process or reaction, but pouring salt into the ocean doesn't change it at all; it is buffered by the sheer size of the system.

11.6 Summary

The phase rule is a relationship between the number of components, phases and degrees of freedom in a system at equilibrium. It has an exact analogy with a system of linear equations, because it is derived from such a system. The original phase rule derived by Gibbs (typically, in a few lines, no big deal) applies to intensive variables only; it is independent of the mass of the system. Later, Duhem's theorem extended the phase rule to extensive variables. It is this version which is implicit in most modeling of geochemical processes, because such models must take into account the appearance and disappearance of phases. Understanding either version of the phase rule requires understanding the various usages of the terms "component" and "species." One of the most fruitful deductions from the phase rule is the principle of "absolute" buffering, which has been widely applied in experimental work and in theoretical discussions.

The phase rule is one of the most impressive and, yes, beautiful, results in all of equilibrium thermodynamics. It has proven essential in the study of experimental and natural systems.

12

Redox reactions

12.1 Introduction

Normal seawater contains about 2660 ppm (0.028 m) of sulfur in the form of sulfate (SO_4^{2-}). Sulfur in this form has a valence of +6, meaning that it has six fewer electrons per atom than has native sulfur, which exists (though not in the ocean) as a yellow crystalline solid. In some parts of the ocean, however, sulfur exists in the form of dissolved hydrogen sulfide, $H_2S(aq)$. Sulfur in this form has a valence of −2, meaning that it has two extra electrons compared to native sulfur, and in this form it is a deadly poison. Those parts of the ocean containing this electron-rich form of sulfur contain no living organisms other than a few kinds of bacteria. Obviously, the number of electrons that each sulfur atom has is not a question of interest only to atomic physicists. Changing sulfate-sulfur to H_2S-sulfur or vice versa involves transferring electrons from one to the other, and this electron transfer is the basic element of redox (reduction–oxidation) reactions.

Many naturally occurring elements in addition to sulfur show similar variations in their number of electrons, with similar large differences in their chemical properties. It would be difficult to overemphasize the importance to us of these variations in valence, or numbers of electrons per atom. Biochemistry, for example, is in large part a study of redox reactions. Because natural environments show great variability in their redox state, we need to develop some kind of measurement, an index, which will be useful in characterizing these redox states, much as we use pH as a measurement or index to characterize the acidity of various states, or temperature as a measurement or an index of the hotness of states. In this chapter we develop two such indexes of redox state.

12.2 Electron transfer reactions

You may not have noticed it, but we have considered two kinds of reactions in previous chapters. In some, such as (9.3),

$$SiO_2(s) + 2H_2O = H_4SiO_4(aq)$$

all elements on the right side have the same number of electrons that they have on the left side – there is no change in valence of any element. In others, such as (9.16),

$$CH_4(g) + O_2(g) = CO_2(g) + 2\,H_2(g)$$

there is such a change. For example, the carbon in CH_4 is C^{4-}, and the carbon in CO_2 is C^{4+}. Each carbon atom in methane that changes to a carbon atom in carbon dioxide must get rid of eight electrons – it is *oxidized*. Where do the electrons go? Obviously, they go to the other actors in the reaction. Oxygen in O_2 has a valence of zero (O^0), while in carbon dioxide it is -2 (O^{2-}), so in changing from O_2 to CO_2, two oxygens gain four electrons. The other four electrons go to hydrogen, which has a valence of $+1$ (H^+) in methane and zero in hydrogen gas. Both oxygen and hydrogen are *reduced*, if the reaction goes from left to right as written. Similarly in reaction (9.15),

$$6\,Fe_2O_3(s) = 4\,Fe_3O_4(s) + O_2(g)$$

we see that all of the iron atoms in Fe_2O_3 are ferric iron (Fe^{3+}), while one out of three iron atoms in Fe_3O_4 is ferrous iron (Fe^{2+}). The iron is partially reduced, while some oxygen in Fe_2O_3 is oxidized to $O_2(g)$ – there is a transfer of electrons from iron to oxygen, or from oxygen to iron, depending on which way the reaction goes. Without such electron transfers, these and many other reactions, including many necessary to life processes, could not proceed.

12.3 The role of oxygen

Both of our examples involve oxygen, which is the most common *oxidizing agent* in natural systems. In the presence of oxygen, many elements are oxidized (lose electrons, gain in valence), while oxygen is reduced. You need only think of rusty nails, green staining on copper objects, and burning logs to realize the truth of this. The process of oxidation obviously takes its name from the fact that oxygen is the premier oxidizing agent, but it is actually defined in terms of electron loss, or increase in valence. In other words, the electrons need not come from or go to oxygen; many redox reactions take place without oxygen.

Consider, for example, what happens when you put a piece of iron in a solution of copper sulfate (Figure 12.1). After a while you see the characteristic color of metallic copper forming on the surface of the iron, and the iron gradually crumbles and eventually disappears. Metallic copper precipitates, and iron dissolves. The reaction is essentially

$$Cu^{2+} + Fe \rightarrow Cu + Fe^{2+} \tag{12.1}$$

We need not include the sulfate, because it is not involved in this process – being negatively charged, the SO_4^{2-} ions provide an overall charge balance in

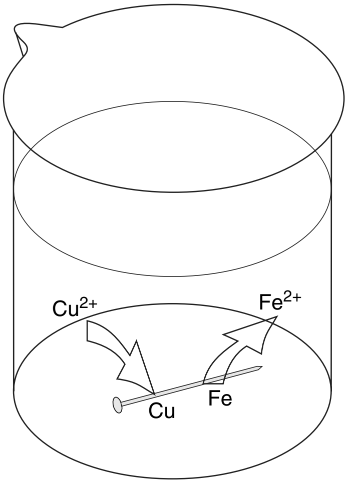


Figure 12.1 An iron nail in a solution of copper sulfate.

the solution. In this example, copper is reduced and iron is oxidized, without the aid of oxygen. (Of course, if we wait long enough, and our solution is open to the atmosphere, oxygen dissolved in the solution will eventually oxidize the copper and the ferrous iron.)

What does thermodynamics tell us about this reaction? Looking up the data from Appendix B, we find

$$\begin{aligned}\Delta_r G^\circ &= \Delta_f G^\circ_{\mathrm{Cu}} + \Delta_f G^\circ_{\mathrm{Fe^{2+}}} - \Delta_f G^\circ_{\mathrm{Cu^{2+}}} - \Delta_f G^\circ_{\mathrm{Fe}} \\ &= 0 + (-78.90) - 65.49 - 0 \\ &= -144.390\,\mathrm{kJ\,mol^{-1}} \\ &= -144\,390\,\mathrm{J\,mol^{-1}}\end{aligned}$$

This means that with both metallic copper and iron present, and cupric and ferrous ions present at 1 molal concentration (and acting ideally), the reaction would proceed spontaneously, as observed. But more interestingly,

$$\begin{aligned}\Delta_r G^\circ &= -RT \ln K \\ -144\,390 &= -(8.3145 \times 298.15) \ln K \\ \log K &= 144\,390/(2.302\,59 \times 8.3145 \times 298.15) \\ &= 25.295\end{aligned}$$

Thus

$$\begin{aligned}K &= \frac{a_{\mathrm{Fe^{2+}}}}{a_{\mathrm{Cu^{2+}}}} \\ &= 10^{25.295}\end{aligned}$$

This means that to reach equilibrium, the activity (≈concentration) of Fe^{2+} would have to be enormously greater than that of the Cu^{2+}, so the reaction will always proceed as written, and cannot be made to go the other way (at least, not by simply adjusting the ion concentrations).

12.4 A simple electrolytic cell

Obviously in this case, and in most natural circumstances, the electron transfer takes place on a molecular level, and we have no control over it. However, what if we could separate the iron and copper, and have the electrons travel through a wire from one to the other? Well, why not? In Figure 12.2, the beaker has been divided into two parts.In one we have a solution of ferrous ions and sulfate ions in contact with a piece of iron (an iron electrode), and in the other we have a solution of cupric ions and sulfate ions in contact with a piece of copper (a copper electrode). Considered separately (and in the absence of oxygen, which would oxidize both electrodes), nothing at all happens. But if we connect a wire between the electrodes, a current begins to flow, because reaction (12.1) wants to occur, and now it can. Iron dissolves, forming more Fe^{2+} in solution, and the electrons, instead of attaching themselves to some immediately adjacent copper ions, must travel through the wire before being able to do that. That is, when they get to the copper electrode, they jump onto some immediately adjacent Cu^{2+} ions, causing copper to precipitate. If the two cells are completely separate, a positive charge would soon build up in the iron solution and a negative charge in the copper solution, stopping the reaction; so we have to provide some kind of connection (a *liquid junction*), which allows sulfate ions to migrate from one solution to the other.

By separating the two parts of the redox reaction, we have caused a current to flow through a wire. We have a simple battery, which we could use to

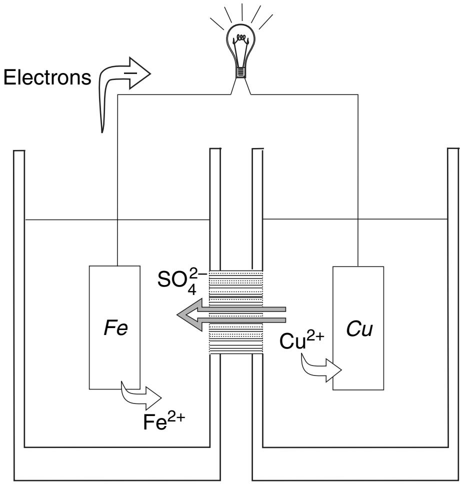


Figure 12.2 An electrolytic cell. Iron dissolves on one side, and copper precipitates on the other. A porous liquid junction allows sulfate to migrate between the solutions. The cell reaction is identical to the reaction in Figure 12.1, Equation (12.1).

power a light bulb, or do other useful things. In most applications of such cells, however, the objective is to obtain thermodynamic data, not to generate electricity, so the cell is operated in a balanced (equilibrium) condition in which no current is allowed to flow, and no changes in the cell compositions take place.

12.4.1 The cell voltage

An electrical circuit delivering direct current such as we have described obviously must have a voltage difference between the two electrodes. If we leave the wire connected to both electrodes, the cell will continue to operate until the iron all dissolves, or until there are no more copper ions to react. That is, the battery will run down, and whatever voltage we had at the beginning decreases to zero during the experiment. The voltage of the cell at any particular point is measured by attaching a voltmeter (which has an extremely high resistance) or a potentiometer (which imposes a voltage in the circuit equal and opposite to the cell voltage) across the electrodes, instead of a piece of wire. Either way, the current flow is stopped, and the voltage measured under equilibrium (no current flow in either direction) conditions. This equilibrium state is, of course, different from the equilibrium state reached when the cell "runs down" and can react no longer. It is a higher energy state that is prevented from reaching a lower energy state by some constraint, in this case an opposing voltage or a high resistance. It is in fact another example of a metastable equilibrium state.

What determines the magnitude of this cell voltage, when it is not zero? Intuitively, we would suspect that it depends a lot on what metals we use to make the electrodes, and if we thought a bit more, we might think that the concentrations of the ions in solution would have an effect too. This is exactly the case, and we must develop an equation relating the voltage to the activities of the reactants and products in the cell reaction.

12.4.2 Half-cell reactions

An oxidation reaction cannot take place without an accompanying reduction reaction – the electrons have to go somewhere – but it is convenient to nonetheless split cell reactions into two complementary "half-cell" reactions. In our copper–iron case, these half-cells are

$$\text{Oxidation}: \quad \mathrm{Fe}(s) = \mathrm{Fe}^{2+} + 2\,\mathrm{e} \tag{12.2}$$

$$\text{Reduction}: \quad \mathrm{Cu}^{2+} + 2\,\mathrm{e} = \mathrm{Cu}(s) \tag{12.3}$$

where e is an electron, and the cell reaction is the sum of these, Equation (12.1).

We also imagine that each half-cell reaction has a half-cell voltage associated with it, and that the cell voltage is the sum of the two half-cell voltages. If we had these half-cell potentials, or voltages,[1] we could tabulate them and mix and match electrodes to calculate the potential of any cell we wanted, much as we tabulate Gibbs energies of compounds so as to be able to calculate the $\Delta_r G^\circ$ of any reaction. Of course, half-cell voltages cannot be measured, just as the G of compounds cannot be measured, but we can get around this, just as we did with $\Delta_f G^\circ$.

With Gibbs energies, we tabulate the difference between the G of a compound and the sum of the Gs of its constituent elements. The elements and their Gibbs energies always cancel out in balanced reactions. With electrodes, we measure and tabulate the difference of every electrode against a standard electrode, and the potential of this standard electrode cancels out in balanced cell reactions. The standard electrode chosen is the *standard hydrogen electrode*, or SHE (Figure 12.3). So if we measure both the copper electrode and the iron electrode and many others in separate experiments against the SHE, we will then be able to calculate the potential of any cell from these tabulated values. Although we can tabulate potentials for every kind of electrode, we wouldn't want to tabulate potentials for every conceivable concentration of product and reactant ions; that's very inefficient. It would be better to tabulate the potentials of each electrode for some standard conditions and to have an equation that could give the potential of the electrode (and of cells constructed from electrodes) at any particular concentrations we are interested in [see equation (12.14)].

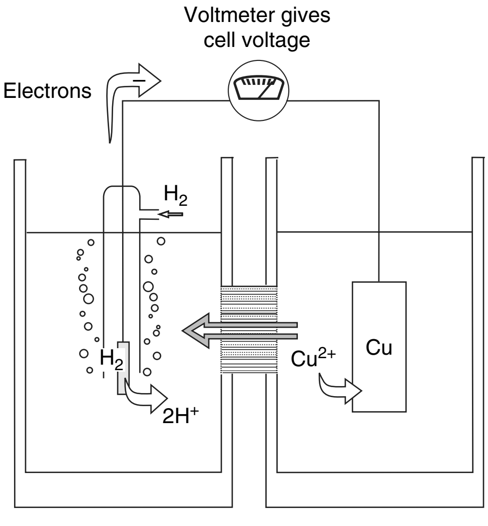


Figure 12.3 An electrolytic cell for measuring the potential of the copper electrode against the standard hydrogen electrode. If the activities of all ions and elements are 1.0, the cell voltage is the standard cell voltage, ε°. The direction of electron flow is indicated for reduction at the copper electrode. When a measurement is made, no current flows.

---

[1] We use the terms *voltage* and *potential* synonymously here. Strictly speaking, voltage is just a particular kind of potential.

12.4.3 Standard electrode potentials

A cell for measuring the potential of the copper–SHE cell is shown in Figure 12.3.The hydrogen electrode is a device for using the reaction

$$2H^+ + 2\,e \rightleftharpoons H_2(g) \tag{12.4}$$

as an electrode.The idea of using a gas as an electrode seems rather bizarre at first. It is accomplished by bubbling hydrogen gas over a specially treated piece of platinum (platinum coated with fine-grained carbon). The platinum serves simply as a source or sink of electrons; the reaction between hydrogen gas and hydrogen ions takes place at the surface of the platinum and is catalyzed by the carbon.

In order that the hydrogen electrode always give the same potential, the activities of both the hydrogen gas and the hydrogen ion must always be the same. These have been standardized at $f_{H_2(g)} = 1$ bar, and $a_{H^+} = 1$, that is, a gas pressure of one bar, and an acid concentration of about 1 molal. A hydrogen electrode operating under these conditions, the standard hydrogen electrode, is assigned a half-cell potential of zero volts by convention.

The half-cell reactions for the cell in Figure 12.3 are

$$\text{Oxidation :} \quad H_2(g) = 2H^+ + 2\,e \tag{12.5}$$

$$\text{Reduction :} \quad Cu^{2+} + 2\,e = Cu(s) \tag{12.6}$$

and the sum of these is the cell reaction

$$Cu^{2+} + H_2(g) = Cu(s) + 2H^+ \tag{12.7}$$

As we mentioned, the cell voltage depends on the activities of all the ions and compounds in the cell reaction; in this case it depends not only on the hydrogen gas pressure and a_{H^+}, but on a_{Cu} and $a_{Cu^{2+}}$ as well. “Standard conditions” is defined as $a = 1$ for all products and reactants in the cell reaction, and so if the hydrogen electrode is operating under SHE conditions ($a_{H_2(g)} = 1$; $f_{H_2(g)} =$ 1 bar), the copper electrode is pure Cu ($a_{Cu(s)} = 1$) and the cupric ion concentration and activity coefficient are adjusted to give $a_{Cu^{2+}} = 1$, the cell voltage will be the *standard* cell voltage, $\mathcal{E}°$.

12.5 The Nernst equation

12.5.1 Work done by cells

In lighting the light bulb or running a small motor, our cell in Figure 12.2 is doing work. We have already seen, in Chapter 4, how much work can be done by a chemical reaction, but you may have forgotten this because we have put so much emphasis on reactions that do *no* work, other than the minimum necessary $P\Delta V$ work. But here we have a chemical reaction that is certainly doing $P\Delta V$ work (the Fe^{2+} and Cu in the cell reaction will have a slightly different molar volume than the Fe and Cu^{2+}, and so some work is done against the atmospheric pressure on the solution), but in addition to this, it is doing work in lighting the bulb. According to §4.10, the maximum amount of work we can get from our cell reaction is given by the $\Delta_r G$ of that reaction, and when $\Delta_r G$ decreases to zero, we reach stable equilibrium and can get no more work from the cell.

The electrical work w required to move a charge of $\mathcal{F}$ coulombs through a potential difference $\mathcal{E}$ volts is

$$-w = \mathcal{F}\mathcal{E}$$

$$(\text{joules} = \text{coulombs} \times \text{volts})$$

where $\mathcal{F}$ is the charge per mole of electrons, so if n is the number of electrons appearing in the reaction as written, there are $n\mathcal{F}$ coulombs of charge, and the work is

$$-w = n\mathcal{F}\mathcal{E} \tag{12.8}$$

Because some convention must be adopted to know whether the voltage $\mathcal{E}$ is positive or negative, you may see Equation (12.8) [as well as (12.9) and (12.10), below] written without the minus sign in some references. The conventions we have adopted (§12.6.1) require the minus sign.

This electrical work is by definition (Chapter 4) the ΔG associated with the process, as long as the electrical work is the only non-$P\Delta V$ work done. Therefore for any process in which $n\mathcal{F}$ coulombs are moved through a potential difference $\mathcal{E}$,

$$\Delta G = -n\mathcal{F}\mathcal{E} \tag{12.9}$$

or

$$\Delta G^\circ = -n\mathcal{F}\mathcal{E}^\circ \tag{12.10}$$

for standard state conditions. As applied to electrochemical cells, these equations are more properly $\Delta\mu = -n\mathcal{F}\mathcal{E}$ and $\Delta\mu^\circ = -n\mathcal{F}\mathcal{E}^\circ$ because many of the individual free energy terms refer to constituents in solution and hence are partial molar terms. These equations connect electrochemistry to the world of thermodynamics. They allow us to calculate the voltage that will be observed in any cell for which we know the cell reaction and the $\Delta_r G$ or the $\Delta_r G^\circ$.

12.5.2 Relation between cell activities and voltage

Consider the general cell reaction

$$b\mathrm{B} + c\mathrm{C} = d\mathrm{D} + e\mathrm{E} \tag{12.11}$$

Let's say that this reaction reaches equilibrium with an external measuring system, giving cell voltage $\mathcal{E}$. If operated under standard conditions, it would give cell voltage $\mathcal{E}^\circ$, and the corresponding Gibbs energies of reaction are

$$\Delta_r\mu = -n\mathcal{F}\mathcal{E} \tag{12.12}$$

and

$$\Delta_r\mu^\circ = -n\mathcal{F}\mathcal{E}^\circ \tag{12.13}$$

From Equation (9.8) we have

$$\Delta_r\mu = \Delta_r\mu^\circ + RT\ln Q \qquad [9.8]$$

where

$$\begin{aligned} Q &= \prod_i a_i^{\nu_i} \\ &= \frac{a_{\mathrm{E}}^e a_{\mathrm{D}}^d}{a_{\mathrm{B}}^b a_{\mathrm{C}}^c} \end{aligned}$$

Recall from Chapter 9 that this activity term is referred to as Q rather than K because it refers to a metastable equilibrium. Substitution of Equations (12.12) and (12.13) gives

$$\mathcal{E} = \mathcal{E}^\circ - \frac{RT}{n\mathcal{F}}\ln Q$$

or

$$\mathcal{E} = \mathcal{E}^\circ - 2.302\,59\frac{RT}{n\mathcal{F}}\log Q \tag{12.14}$$

This is the Nernst equation, after the physical chemist W. Nernst, who derived it at the end of the nineteenth century. As above, n is the number of electrons transferred in the cell reaction (2 in reaction 12.7), $\mathcal{F}$ the Faraday of charge, R the gas constant, and T the temperature (in kelvins). The constant 2.302 59 is used to convert from natural to base 10 logs. At 25 °C the quantity 2.302 59 $RT/\mathcal{F}$ has the value 0.059 16, which is called the Nernst slope. The importance of (12.14) is that it allows calculation of the potentials of cells having nonstandard state concentrations (i.e., real cells) from tabulated values of standard half-cell values or tabulated standard Gibbs energies.

Equation (12.11) could also be considered to represent a half-cell reaction, except that the electron is not shown. So evidently we could use the Nernst equation to calculate half-cell potentials if we knew what value to assign the chemical potential of an electron. It turns out, of course, that because the electrons always cancel out in balanced reactions, we could assign any value we like to the electron Gibbs energy and it would make no difference to our calculated cell potentials. The easiest value to assign is zero, and that is what is done. Therefore, the Nernst equation is used to calculate both half-cell and cell potentials.

12.6 Some necessary conventions

In our discussion so far, we have skipped lightly over some points which are not important to a general understanding, but which if neglected will result in getting the wrong answers in calculations. For example, we said (§12.4.2) that the cell voltage is the *sum* of the two half-cell voltages. Actually, it is a little more complicated. Because half-cell and cell reactions may be written forwards or backwards, and a voltage by itself is not obviously positive or negative, there has to be a set of rules to keep things straight. Unfortunately, there is more than one set of rules. We present here the rules set out by the IUPAC (International Union of Pure and Applied Chemistry), which are followed by most people today. Be warned, however, that several geochemical sources use a different set of rules.

12.6.1 The IUPAC rules

Some of these points have been already discussed. We include all the rules here for completeness.

1. Cell reactions are written such that the left-hand electrode supplies electrons to the outer circuit (i.e., oxidation takes place), and the right-hand electrode accepts electrons from the outer circuit (i.e., reduction takes place).
2. The cell potential is given by

   $$\mathcal{E} = \mathcal{E}_{\text{right electrode}} - \mathcal{E}_{\text{left electrode}}$$

   that is,

   $$\mathcal{E} = \mathcal{E}_{\text{reduction electrode}} - \mathcal{E}_{\text{oxidation electrode}}$$

3. The cell potential is related to the Gibbs energy by

   $$\Delta_r G = -n\mathcal{F}\mathcal{E}$$

4. The electrode potential of a half-cell is equal in magnitude and sign to the potential of a cell formed with the electrode on the right and the standard hydrogen electrode ($\mathcal{E}^\circ = 0$) on the left.
5. Standard half-cell reactions are tabulated and calculated as reductions, for example,

   $$\mathrm{Zn}^{2+} + 2\,\mathrm{e} = \mathrm{Zn}(s) \qquad \mathcal{E}^\circ = -0.763\,\mathrm{V}$$

   However, the half-cell potential is a sign-invariant quantity, that is,

   $$\mathrm{Zn}(s) = \mathrm{Zn}^{2+} + 2\,\mathrm{e} \qquad \mathcal{E}^\circ = -0.763\,\mathrm{V}$$

6. For the reaction

$$bB + cC = dD + eE$$

the Nernst expression is

$$\mathcal{E} = \mathcal{E}^\circ - \frac{RT}{n\mathcal{F}} \ln\left(\frac{a_E^e a_D^d}{a_B^b a_C^c}\right)$$

7. In view of item 5, the Nernst expression for a half-cell is given by

$$\mathcal{E} = \mathcal{E}^\circ - \frac{RT}{n\mathcal{F}} \ln\left(\frac{\text{reduced form}}{\text{oxidized form}}\right)$$

12.6.2 Examples

Let's calculate the potential of the cell in Figure 12.2. By convention 5, both half-cells are written and calculated as reductions, no matter what is happening in the real cell. Thus, for the copper half-cell,

$$\begin{aligned}
&\mathrm{Cu}^{2+} + 2\,\mathrm{e} = \mathrm{Cu}(s)\\
&\Delta_r G^\circ = \Delta_f G^\circ_{\mathrm{Cu}(s)} - \Delta_f G^\circ_{\mathrm{Cu}^{2+}}\\
&\quad = 0 - 65.49\,\mathrm{kJ\,mol^{-1}}\\
&\quad = -65\,490\,\mathrm{J\,mol^{-1}}\\
&\quad = -n\mathcal{F}\mathcal{E}^\circ
\end{aligned}$$

Because two electrons are involved, $n = 2$, so

$$\begin{aligned}
\mathcal{E}^\circ_{\mathrm{Cu\ half\text{-}cell}} &= -\frac{-65\,490}{2 \times 96\,485}\\
&= 0.339\,\mathrm{V}
\end{aligned}$$

For the iron half-cell,

$$\begin{aligned}
&\mathrm{Fe}^{2+} + 2\,\mathrm{e} = \mathrm{Fe}(s)\\
&\Delta_r G^\circ = \Delta_f G^\circ_{\mathrm{Fe}(s)} - \Delta_f G^\circ_{\mathrm{Fe}^{2+}}\\
&\quad = 0 - (-78.90)\,\mathrm{kJ\,mol^{-1}}\\
&\quad = 78\,900\,\mathrm{J\,mol^{-1}}\\
&\quad = -n\mathcal{F}\mathcal{E}^\circ
\end{aligned}$$

$$\begin{aligned}
\mathcal{E}^\circ_{\mathrm{Fe\ half\text{-}cell}} &= -\frac{78\,900}{2 \times 96\,485}\\
&= -0.409\,\mathrm{V}
\end{aligned}$$

For the complete cell,

$$\begin{aligned}
\mathcal{E}^\circ_{\text{Cu-Fe cell}} &= \mathcal{E}^\circ_{\text{reduction half-cell}} - \mathcal{E}^\circ_{\text{oxidation half-cell}} \\
&= \mathcal{E}^\circ_{\text{Cu half-cell}} - \mathcal{E}^\circ_{\text{Fe half-cell}} \\
&= 0.339 - (-0.409) \\
&= 0.748\,\text{V}
\end{aligned}$$

Note that if we wrote the cell backwards,

$$\text{Cu}(s) + \text{Fe}^{2+} = \text{Cu}^{2+} + \text{Fe}(s)$$

both half-cell reactions would still be written and calculated as reductions, but the cell voltage would now be

$$\begin{aligned}
\mathcal{E}^\circ_{\text{Cu-Fe cell}} &= \mathcal{E}^\circ_{\text{reduction half-cell}} - \mathcal{E}^\circ_{\text{oxidation half-cell}} \\
&= \mathcal{E}^\circ_{\text{Fe half-cell}} - \mathcal{E}^\circ_{\text{Cu half-cell}} \\
&= -0.409 - (0.339) \\
&= -0.748\,\text{V}
\end{aligned}$$

Thus the signs of both $\Delta_r G^\circ$ and $\mathcal{E}^\circ$ of the complete cell reaction depend on how the cell is written, but the signs of the half-cell reactions do not.

These are the standard potentials. Suppose the cell is operating under nonstandard (real) conditions. Let's say $a_{\text{Cu}^{2+}}$ is not 1 but 0.1, and $a_{\text{Fe}^{2+}}$ is 1.5. Using the Nernst expression for the complete cell (reaction 12.1) at 25 °C,

$$\begin{aligned}
\mathcal{E} &= \mathcal{E}^\circ - 2.302\,59\frac{RT}{n\mathcal{F}}\log Q \\
&= \mathcal{E}^\circ - \frac{0.059\,16}{n}\log\frac{a_{\text{Fe}^{2+}}}{a_{\text{Cu}^{2+}}} \\
&= 0.748 - \frac{0.059\,16}{2}\log\frac{1.5}{0.1} \\
&= 0.748 - 0.0348 \\
&= 0.713\,\text{V}
\end{aligned}$$

So the real cell with these concentrations would have a potential of 0.713 V, rather than 0.748 V.

The Nernst expression can also be used for the half-cells. Thus for the Cu half-cell,

$$\begin{aligned}\mathcal{E} &= \mathcal{E}^\circ - 2.302\,59\frac{RT}{n\mathcal{F}}\log Q\\ &= \mathcal{E}^\circ - \frac{0.059\,16}{n}\log\frac{a_{\mathrm{Cu}}}{a_{\mathrm{Cu}^{2+}}}\\ &= 0.339 - \frac{0.059\,16}{2}\log\frac{1}{0.1}\\ &= 0.339 - 0.0296\\ &= 0.309\,\mathrm{V}\end{aligned}$$

and for the Fe half-cell,

$$\begin{aligned}\mathcal{E} &= \mathcal{E}^\circ - 2.302\,59\frac{RT}{n\mathcal{F}}\log Q\\ &= \mathcal{E}^\circ - \frac{0.059\,16}{n}\log\frac{a_{\mathrm{Fe}}}{a_{\mathrm{Fe}^{2+}}}\\ &= -0.409 - \frac{0.059\,16}{2}\log\frac{1}{1.5}\\ &= -0.409 + 0.0052\\ &= -0.404\,\mathrm{V}\end{aligned}$$

The cell potential is then

$$\begin{aligned}\mathcal{E}_{\mathrm{Cu-Fe\ cell}} &= \mathcal{E}_{\mathrm{reduction\ half\text{-}cell}} - \mathcal{E}_{\mathrm{oxidation\ half\text{-}cell}}\\ &= \mathcal{E}_{\mathrm{Cu\ half\text{-}cell}} - \mathcal{E}_{\mathrm{Fe\ half\text{-}cell}}\\ &= 0.309 - (-0.404)\\ &= 0.713\,\mathrm{V}\end{aligned}$$

as before.

There certainly are other ways to do these calculations and have them come out right. However, all conventions have their good and bad points, and the IUPAC conventions are the most commonly used.

12.7 Measuring activities

In geochemistry, possibly the greatest interest in galvanic cells is in understanding the concept of *Eh* and its use in determining redox conditions in natural environments (§12.8.1). In chemistry, it is in determining the activities of solutes.

12.7.1 $\mathcal{E}^\circ$ as a source of $\Delta_f G^\circ$

Electrochemical cells are of course of great practical importance in the form of batteries and fuel cells. In thermodynamics, the relationship between cell potential and Gibbs energy is often used the other way around. That is, cell potentials are one of the most accurate and useful sources of information about Gibbs energies of reactions and dissolved substances. Not all Gibbs energy data come from calorimetry. Most data for $\Delta_f G^\circ_{\mathrm{Fe}^{2+}}$, $\Delta_f G^\circ_{\mathrm{Cu}^{2+}}$, and other ionic species come from the measurement of $\mathcal{E}^\circ$ in cells such as shown in Figure 12.3.

Actually the cells in Figures 12.3 and 12.1 are complicated by a factor we have not mentioned. These are cells having a junction between two solutions with different compositions, and this results in an additional source of emf, called a junction potential. It arises because the two solutions must inevitably diffuse into one another to some extent, however small, and ionic concentration gradients are established which create a potential difference. This is a major topic in electrochemistry, but need not be treated in detail here. Suffice it to say that in the case of Figure 12.3, the liquid junction is not necessary. The copper electrode could be placed in the same solution as the hydrogen electrode, as in Figure 12.4, creating a cell without a liquid junction, simplifying the thermodynamic interpretation.

The activity of HCl

As an example of the determination of activity of a solute, we consider HCl. A cell much like that in Figure 12.4 but having a standard calomel or silver chloride reference electrode (§12.8.3) in place of the platinum electrode, and having a solution of HCl rather than a solution of ferrous and ferric ions, can be abbreviated as

$$\mathrm{Pt}|\mathrm{H_2}(g)|\mathrm{HCl}(aq)|\mathrm{AgCl}|\mathrm{Ag}$$

where | indicates a phase boundary. With the half-cell reactions

$$\tfrac{1}{2}\mathrm{H_2}(g) = \mathrm{H^+} + \mathrm{e}$$

and

$$\mathrm{AgCl} + \mathrm{e} = \mathrm{Ag}(s) + \mathrm{Cl^-}$$

the cell reaction is

$$\tfrac{1}{2}\mathrm{H_2}(g) + \mathrm{AgCl} = \mathrm{Ag} + \mathrm{H^+} + \mathrm{Cl^-}$$

or

$$\tfrac{1}{2}\mathrm{H_2}(g) + \mathrm{AgCl} = \mathrm{Ag} + \mathrm{HCl}$$

for which the Nernst equation is

$$\begin{aligned}\mathcal{E} &= \mathcal{E}^\circ - 2.302\,59\frac{RT}{n\mathcal{F}}\log Q \qquad [12.14]\\ &= \mathcal{E}^\circ - 2.302\,59\frac{RT}{n\mathcal{F}}\log\frac{a_{\mathrm{Ag}}a_{\mathrm{HCl}}}{f_{\mathrm{H}_2}^{1/2}a_{\mathrm{AgCl}}}\end{aligned}$$

Because the activity of AgCl and of Ag are fixed at 1.0, the only variables are f_{H_2} and the concentration of HCl, so at a fixed standard state f_{H_2} of 1.0 bars, the cell voltage is a direct measure of the activity of HCl, as long as $\mathcal{E}^\circ$ is known. Determination of $\mathcal{E}^\circ$ requires an extrapolation to infinite dilution, and was the subject of a great deal of research involving the behavior of charged solutes in dilute solutions. For our purposes, it is sufficient to see how cell voltages can be related to solute activities. Some results of these measurements for HCl are shown in Figure 15.3.

12.8 Measuring redox conditions

So far we have only considered cells that we might construct ourselves, in the laboratory. For the scientist interested in natural environments, this is background information. What we really want to know is how to characterize natural environments as being either reducing or oxidizing on some numerical scale. Natural environments don't normally have electrodes sticking out of them, so what is the connection?

In the absence of electrodes and voltages, the redox state of a solution is characterized by the relative concentrations of reduced and oxidized ionic or molecular species in the solution. Thus in a solution containing Fe ions, the solution is relatively reduced if there are more Fe^{2+} ions than Fe^{3+} ions, and vice versa. In a solution containing carbon species, the solution is relatively reduced if there are more CH_4 molecules than CO_2 molecules. In a solution containing sulfur, the solution is relatively reduced if there are more $\mathrm{H}_2\mathrm{S}$ molecules than SO_4^{2-} ions, and vice versa. And so on. For those elements that have more than one valence state in natural environments, the two (or more) states will be present in various ratios, depending on whether the environment is reducing or oxidizing. In a solution containing all of these, the $a_{\mathrm{Fe}^{2+}}/a_{\mathrm{Fe}^{3+}}$, $a_{\mathrm{CH}_4}/a_{\mathrm{CO}_2}$, and $a_{\mathrm{H}_2\mathrm{S}}/a_{\mathrm{SO}_4^{2-}}$ ratios will all be different, but each will be controlled by the same factors (T, P, and the bulk composition of the solution), and so each one should give us the same index of redox conditions, if equilibrium prevails.

But what is this index? We will consider two commonly used ones, *Eh* and f_{O_2}.

12.8.1 Redox potential, *Eh*

Suppose you have a sample solution that contains both ferrous and ferric ions. The ferrous/ferric ratio is a measure of how reduced/oxidized the sample solution is. For any change in this ratio, some reaction involving electron transfer must take place. What we must do is insert an electrode that will supply/absorb these electrons. In other words, we need an electrode that responds to the ferrous/ferric ratio, that is, that has a half-cell potential that varies with this ratio. We could insert this electrode in the solution, connect it to a SHE, and the measured cell potential would depend on the ferrous/ferric ratio in the solution. Finding such an electrode is easier than you might think.

Note that in the Cu and Fe electrodes we have considered, the "reduced form" in both cases is the metal, Cu or Fe. The "oxidized form" is an ion in solution, Cu^{2+} or Fe^{2+}, and the electrode, being made of the metal, is a necessary part of the half-cell. However, in the SHE, both the reduced form (H_2) and the oxidized form (H^+) are in the solution (one as a gas phase); neither is part of the electrode. The platinum electrode itself is nothing but a source or sink for electrons. We are now considering another case where both the reduced form (Fe^{2+}) and the oxidized form (Fe^{3+}) are in the solution, and so all we have to do is provide a source and sink for electrons. All we need is a piece of platinum, as shown in Figure 12.4.

Suppose the solution is quite reduced, with $a_{Fe^{2+}}/a_{Fe^{3+}} = 10$. The half-cell reactions are (both written as reductions)

$$2H^+ + 2\,e = H_2(g) \quad \mathcal{E}^\circ = 0.0\,V$$

and

$$Fe^{3+} + e = Fe^{2+} \quad \mathcal{E}^\circ = 0.769\,V$$

The complete cell is

$$Fe^{3+} + \tfrac{1}{2}H_2(g) = Fe^{2+} + H^+$$

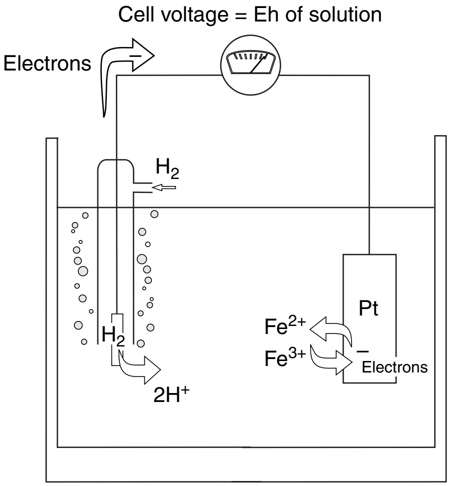


Figure 12.4 How to measure the *Eh* of a solution containing both ferrous and ferric ions. The direction of electron flow is indicated for reduction at the platinum surface. When a measurement is made, no electrons are flowing.

The Nernst equation may be written for either the complete cell or the Fe half-cell, giving the same answers. Thus

$$\begin{aligned}\mathcal{E} = Eh = \mathcal{E}^\circ - \frac{0.059\,16}{1}\log\frac{a_{\mathrm{Fe}^{2+}}}{a_{\mathrm{Fe}^{3+}}}\\ = 0.769 - 0.059\,16 \times \log(10)\\ = 0.710\,\mathrm{V}\end{aligned}$$

So the *Eh* of this relatively reduced solution is 0.710 V. If the solution is quite oxidized, with $a_{\mathrm{Fe}^{2+}}/a_{\mathrm{Fe}^{3+}} = 0.1$,

$$\begin{aligned}\mathcal{E} = Eh = \mathcal{E}^\circ - \frac{0.059\,16}{1}\log\frac{a_{\mathrm{Fe}^{2+}}}{a_{\mathrm{Fe}^{3+}}}\\ = 0.769 - 0.059\,16 \times \log(0.1)\\ = 0.828\,\mathrm{V}\end{aligned}$$

and the *Eh* of this more oxidized solution is 0.828 V.

To summarize, you may think of *Eh* of a solution as either a cell potential or a half-cell potential. It is the potential of a cell having one electrode that responds reversibly to a redox couple (such as $\mathrm{Fe}^{2+}/\mathrm{Fe}^{3+}$) or couples in the solution and the SHE as the other electrode. Or it is the half-cell potential of an electrode responding reversibly to a redox couple or couples in the solution. You may use any kind of electrode as the other side of the cell, as long as you correctly deduce the half-cell potential of the electrode that is responding to conditions in your solution. Half-cell potentials are *defined* in terms of the SHE by our IUPAC conventions.

Therefore, we have an index of redox conditions, only assuming that reduced and oxidized species in solution can readily exchange electrons at a platinum surface. There are some practical difficulties in this respect, as discussed below.

12.8.2 Redox couples other than iron

But suppose our sample solution in Figure 12.4 contains not only the ferrous/ferric redox couple, but also Mn^{2+}–Mn^{4+}, $\mathrm{H_2S}$–$\mathrm{SO_4^{2-}}$, $\mathrm{CH_4}$–$\mathrm{CO_2}$, and others. Even if all these redox couples are at equilibrium, each will have a different activity ratio. To which does our platinum electrode respond? In theory, it responds to all of them simultaneously, and all result in the same *Eh*. Each couple has a different activity ratio, but each also has a different value of $\mathcal{E}^\circ$, and the resulting *Eh* for the solution must be the same for each redox couple if they are at equilibrium, and if each reacts with the platinum electrode.

12.8.3 Some practical difficulties

No unique Eh values

However, in practice the platinum electrode does not respond to all redox couples equally. In fact, the only ones it responds to well are Fe and Mn, and then only if concentrations are high enough. Sulfur, carbon, and many other redox pairs simply do not give up or take up electrons easily at the platinum surface, so measured *Eh* values primarily reflect the ferrous/ferric ratio in solutions. Of course, if equilibrium prevails, this should be enough – the activity ratios of all other couples could be calculated if the ferrous/ferric (and pH) ratio is known. However, at Earth surface conditions in natural environments, equilibrium often (in fact, usually) does not prevail, and the only way to really know the activities of many redox pairs is to analyze the solution for both parts of the pair.

This is strikingly illustrated by the data collected by Lindberg and Runnells (1984), some of which are summarized in Figure 12.5 These authors examined over 150 000 groundwater analyses from the USGS database, plus values from the literature, from which they selected 611 analyses of acceptable quality. The ionic species activities were then calculated (this speciation modeling is discussed in Chapters 16 and 18) and theoretical *Eh* values calculated for several redox couples. Comparison with the measured *Eh* illustrates how unreliable these measurements are.

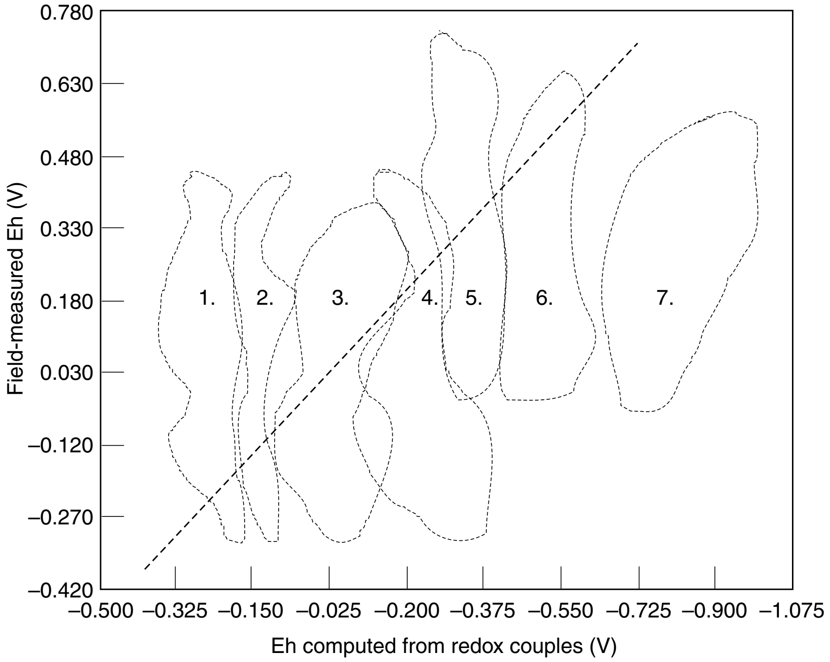


Figure 12.5 *Eh* values, generalized into areas, computed from seven different redox couples compared with the *Eh* value measured in the field. Numbered areas refer to 1. HS^-/SO_4^{2-}; 2. $HS^-/S(s)$; 3. $Fe^{2+}/Fe(OH)_3(s)$; 4. Fe^{3+}/Fe^{2+}; 5. NH_4^+/NO_2^-; 6. NO_2^-/NO_3^-; 7. $O_2(aq)/H_2O$. The dashed line represents perfect agreement. Modified from Lindberg and Runnells (1984).

Natural solutions that have not internally equilibrated cannot be said to have a unique *Eh*. Nevertheless, *Eh* measurements are often useful in a qualitative sense and may be fairly accurate in some situations, such as acid Fe-rich mine drainage systems. But even if *Eh* measurements were not useful at all, the concept of *Eh* is firmly established in the literature of natural environments, where it is used primarily in discussing *models* or hypothetical situations.

Reference electrodes

Another practical difficulty is that the hydrogen electrode is a rather delicate apparatus, really only suitable for laboratory use. How do we get it into the field, to take *Eh* measurements in natural environments? We don't. Other electrodes can be designed that have fixed potentials, independent of what solution they are put into, called reference electrodes. Their potentials with respect to the SHE can be measured, so that field measurements of cells composed of a platinum electrode and a reference electrode can be made, and the readings corrected to what they would have been had a SHE been used, giving the *Eh*.

Two common reference electrodes are the calomel (Hg/Hg_2Cl_2) and the silver/silver chloride (Ag/AgCl) electrodes. In both, the activities of all active parts of the electrode are fixed, so the electrode potential is fixed at a given temperature. For example, the Ag/AgCl electrode half-cell reaction is

$$\mathrm{AgCl} + \mathrm{e} = \mathrm{Ag}(s) + \mathrm{Cl}^-$$

Both Ag and AgCl are present as solid phases so $a_{\mathrm{Ag}} = a_{\mathrm{AgCl}} = 1$, and these are immersed in a solution saturated with solid KCl, which fixes a_{Cl^-} at a constant value (Figure 12.6). With all reactants and products of the half-cell reaction having fixed activities, $\mathcal{E}^\circ_{\mathrm{Ag/AgCl\ half\text{-}cell}}$ has a fixed value, which is 0.222 V. That is, a cell composed of a Ag/AgCl electrode on one side and the

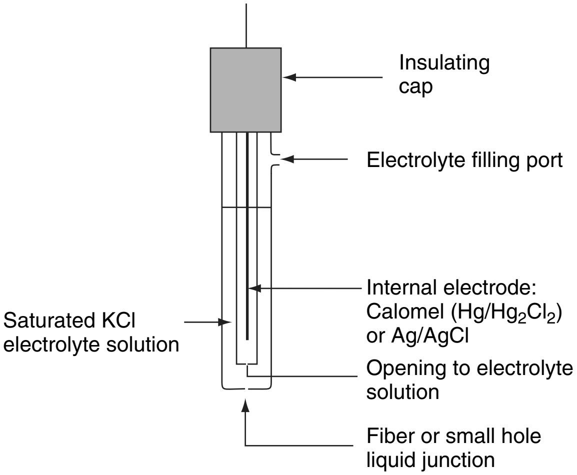


Figure 12.6 A silver/silver chloride reference electrode.

SHE on the other will record a constant potential of 0.222 V. When used as a reference electrode, immersed in a solution containing ferrous and ferric ions and connected to a platinum electrode, for example, the cell reaction is

$$\mathrm{Fe^{3+}} + \mathrm{Ag}(s) + \mathrm{Cl^-} = \mathrm{Fe^{2+}} + \mathrm{AgCl}$$

and the standard cell potential of this cell is

$$\begin{aligned}\mathcal{E}^\circ &= \mathcal{E}^\circ_{\mathrm{Fe^{2+}/Fe^{3+}}} - \mathcal{E}^\circ_{\mathrm{Ag/AgCl}}\\ &= 0.769 - 0.222\\ &= 0.547\,\mathrm{V}\end{aligned}$$

whereas if measured against SHE, $\mathcal{E}^\circ_{\mathrm{Fe^{2+}/Fe^{3+}}}$ is 0.769 V. *Eh* measurements are defined as observed cell voltages using the SHE as reference, so if Ag/AgCl is used instead, 0.222 V must be *added* $(0.547 + 0.222 = 0.769)$ to the observed readings. For calomel reference electrodes, 0.268 V must be added.

12.9 *Eh*–pH diagrams

The only other intensive variable of comparable significance in aqueous systems is pH. It too is a function of the bulk composition at a given T and P, but both are closely related to a large number of important reactions. Therefore, it proves natural to use both as variables in diagrams of systems at fixed T and P, and *Eh*–pH diagrams have become a standard method of displaying and interpreting geochemical data. In the following sections we outline the theoretical basis for calculating these diagrams.

In this section, we will calculate portions of a simple *Eh*–pH diagram for the system Mn–H_2O. This illustrates most of the problems encountered in calculating such diagrams. If you wish to add components such as CO_2 or H_2S, the methods are similar, and details are provided by Garrels and Christ (1965).

12.9.1 General topology of *Eh*–pH diagrams

First, let us examine the completed *Eh*–pH diagram for Mn–H_2O–O_2 in Figure 12.7. There are typically four different types of boundaries shown on these diagrams. The top line, labeled O_2/H_2O, represents conditions for water in equilibrium with O_2 gas at 1 atm. Above this line, a P_{O_2} greater than 1 atm is required for water to exist, so that because the diagram is drawn for a pressure of 1 atm, water is not stable above this line. Similarly, the bottom line H_2O/H_2 represents conditions for water in equilibrium with H_2 gas at 1 atm. Below this line, P_{H_2} values greater than 1 atm are required for water to exist, that is, at 1 atm water is not stable. Therefore, the water stability field is between these two lines.

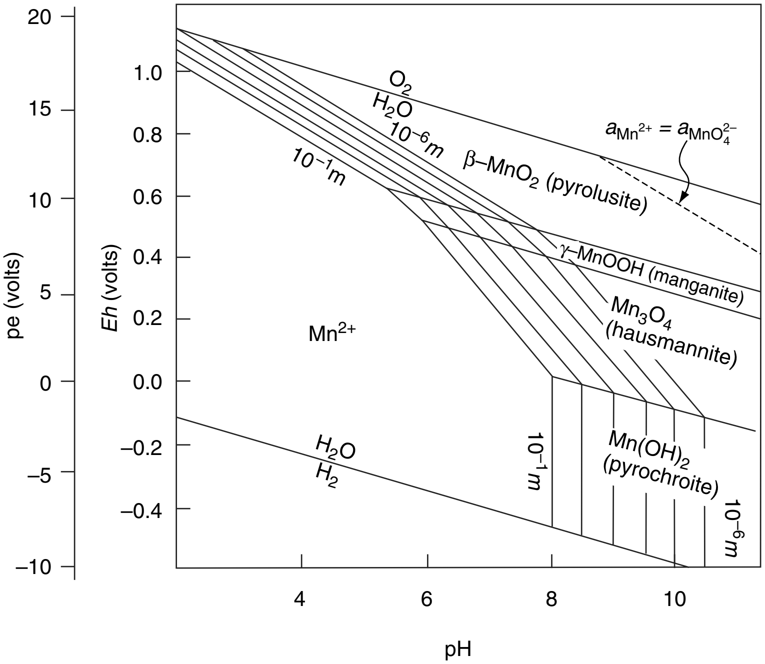


Figure 12.7 *Eh*–pH and pe–pH relations in the system Mn–H_2O–O_2 at 25 °C, 1 bar. Mn^{2+} activities and stability fields of Mn-oxide minerals are included. The pe and *Eh* axes are related by the formulas pe = 5040 *Eh*/*T*.

The second type of boundary separates the stability fields of minerals or solid phases such as hausmannite (Mn_3O_4) and pyrochroite [$Mn(OH)_2$]. These are true phase boundaries: hausmannite is thermodynamically unstable below the hausmannite/pyrochroite boundary and pyrochroite is unstable above it. Thus these first two kinds of boundary represent thermodynamic stability fields for different substances. Notice that on this diagram they all have the same slope (equal to the Nernst slope).

The remaining two kinds of lines are not stability boundaries at all but refer to particular concentrations of dissolved ions. For example, the vertical lines within the pyrochroite stability field represent contoured solubilities of pyrochroite as Mn^{2+} concentrations varying from 10^{-1} to 10^{-6} *m*. Finally, the dashed boundaries between aqueous species, such as that between Mn^{2+} and MnO_4^{2-}, indicate where the activities of the two species are exactly equal. To the right of this line, Mn^{2+} remains present in the solution, but at a lower activity than MnO_4^{2-}, and vice versa.

12.9.2 Sample calculations

It is possible to look up half-cell potentials for many reactions in physical chemistry textbooks and compilations of electrochemical data. However, it is usually a better procedure to choose Gibbs energy data and to use those to calculate *Eh* and $\mathcal{E}°$ for the reactions of interest.

As the first example, we will calculate the two boundaries for the stability field of water.

For the boundary $H_2O(l)$–$O_2(g)$, the half-cell reduction reaction is

$$4H^+ + O_2(g) + 4\ e = 2H_2O(l) \tag{12.15}$$

for which $n = 4$. Using the tabulated $\Delta_f G^\circ$ for water (all others being zero), we find

$$\begin{aligned} \Delta_r G^\circ &= 2(-237.129) - 0 \\ &= -474.258\ \mathrm{kJ\,mol^{-1}} \end{aligned} \tag{12.16}$$

Because

$$\Delta_r G^\circ = -n\mathcal{F}\mathcal{E}^\circ$$

then

$$\begin{aligned} \mathcal{E}^\circ &= -(-474\,258/4 \times 96\,485) \\ &= 1.23\ \mathrm{V} \end{aligned}$$

From the Nernst equation (12.14)

$$Eh = \mathcal{E}^\circ - 2.302\,59\frac{RT}{n\mathcal{F}}\log[1/(f_{O_2} \cdot a_{H^+}^4)]$$

Setting $f_{O_2} = 1$ atm, and recalling that $\mathrm{pH} = -\log a_{H^+}$, gives the equation for the boundary in terms of Eh and pH.

$$Eh = 1.23 - 0.0592\mathrm{pH} \tag{12.17}$$

For the boundary $H_2(g)$–$H_2O(l)$,

$$\begin{aligned} 2H^+ + 2\ e &= H_2(g) \\ \Delta_r G^\circ &= 0 \\ &= -n\mathcal{F}\mathcal{E}^\circ \end{aligned}$$

and

$$Eh = 0 - 2.302\,59\frac{RT}{n\mathcal{F}}\log(f_{H_2}/a_{H^+}^2)$$

or, with $f_{H_2} = 1$ atm, and $n = 2$,

$$Eh = -0.0592\mathrm{pH} \tag{12.18}$$

For boundary $Mn(OH)_2$–Mn_3O_4,

$$\begin{aligned} Mn_3O_4(c) + 2H_2O(l) + 2H^+ + 2\,e &= 3Mn(OH)_2(c) \\ \Delta_r G^\circ &= -94\,140\,\mathrm{J\,mol^{-1}} \\ &= -n\mathcal{F}\mathcal{E}^\circ \\ \mathcal{E}^\circ &= -(-94\,140/2 \times 96\,485) \\ &= 0.488\,\mathrm{V} \end{aligned}$$

and, with $n = 2$,

$$Eh = 0.488 - 0.0592\mathrm{pH} \tag{12.19}$$

For solubility of Mn_3O_4 as Mn^{2+},

$$\begin{aligned} Mn_3O_4(c) + 8H^+ + 2\ e &= 3Mn^{2+} + 4H_2O \\ \Delta_r G^\circ &= -349\,616\,\mathrm{J\,mol^{-1}} \\ \mathcal{E}^\circ &= 352\,711/(2 \times 96\,485) \\ &= 1.812\,\mathrm{V} \end{aligned}$$

$$Eh = \mathcal{E}^\circ - \frac{0.0592}{2}\log\left(\frac{a^3_{\mathrm{Mn^{2+}}}}{a^8_{\mathrm{H^+}}}\right)$$

or

$$Eh = 1.812 - 0.237\mathrm{pH} - 0.0887\log a_{\mathrm{Mn^{2+}}} \tag{12.20}$$

This is plotted for selected values of Mn^{2+} activity ranging from 10^{-1} to 10^{-6} in Figure 12.7.

For equal activity contour of Mn^{2+} and MnO_4^{2-},

$$\begin{aligned} MnO_4^{2-} + 8H^+ + 4\ e &= Mn^{2+} + 4H_2O \\ \Delta_r G^\circ &= -675\,916\ \mathrm{J\,mol^{-1}} \\ \mathcal{E}^\circ &= 672\,369/4 \times 96485 \\ &= 1.751\,\mathrm{V} \end{aligned}$$

$$Eh = 1.751 - 0.1182\mathrm{pH} - 0.0148\log(a_{\mathrm{Mn^{2+}}}/a_{\mathrm{MnO_4^{2-}}})$$

and where the activities of both aqueous species are equal, this reduces to

$$Eh = 1.751 - 0.1182\mathrm{pH} \tag{12.21}$$

This boundary lies at high *Eh* and pH and is illustrated in Figure 12.7.

12.9.3 pe–pH diagrams

There is a second way of calculating the same kinds of diagrams using the alternative variable pe rather than *Eh*. Because this is another way of doing exactly the same thing, it could be argued that the new variable pe is unnecessary and redundant. However, the pe and *Eh* scales differ numerically, and pe calculations are now used about as frequently as *Eh*. It is thus worthwhile discussing the use of this second variable.

The idea was to have an analogy between pH, which refers to hydrated protons, and pe, which would refer to hydrated electrons. Like the proton, the electron is assigned a standard $\Delta_f G^\circ = 0$. Like pH, pe is defined in terms of activity:

$$\mathrm{pH} = -\log a_{\mathrm{H}^+}$$

$$\mathrm{pe} = -\log a_{\mathrm{e}} \tag{12.22}$$

In fact the "*p*" notation is now widely used for various quantities. For example, equilibrium constants are sometimes given in terms of p*K*, where

$$\mathrm{p}K = -\log K$$

In other words if $K = 10^{-6.37}$, then $\mathrm{p}K = 6.37$.

Consider a half-cell reaction

$$a\mathrm{A} + n\,\mathrm{e} = b\mathrm{B}$$

where, at equilibrium

$$K = \frac{a_{\mathrm{B}}^{b}}{a_{\mathrm{A}}^{a}\, a_{\mathrm{e}}^{n}}$$

and

$$\begin{aligned}\log K &= \log \frac{a_{\mathrm{B}}^{b}}{a_{\mathrm{A}}^{a}\, a_{\mathrm{e}}^{n}}\\ &= \log \frac{a_{\mathrm{B}}^{b}}{a_{\mathrm{A}}^{a}} - n\log a_{\mathrm{e}}\\ &= \log \frac{a_{\mathrm{B}}^{b}}{a_{\mathrm{A}}^{a}} + n\,\mathrm{pe}\end{aligned}$$

so

$$\mathrm{pe} = \frac{1}{n}\log K - \frac{1}{n}\log \frac{a_{\mathrm{B}}^{b}}{a_{\mathrm{A}}^{a}}$$

When A and B are in their standard states of unit activity, the last term drops out, and we define pe° as

$$\mathrm{pe}^\circ = \frac{1}{n}\log K$$

Combining equations, we get

$$\begin{aligned}\mathrm{pe} &= \mathrm{pe}^\circ - \frac{1}{n}\log\frac{a_\mathrm{B}^b}{a_\mathrm{A}^a}\\ &= \mathrm{pe}^\circ - \frac{1}{n}\log Q \qquad (12.23)\end{aligned}$$

12.9.4 Comparison of pe and *Eh*

There is a simple, linear relationship between the two variables pe and *Eh*, which is evident in comparing the two defining equations:

$$Eh = \mathcal{E}^\circ - 2.302\,59\frac{RT}{n\mathcal{F}}\log Q \qquad [12.14]$$

$$\mathrm{pe} = \mathrm{pe}^\circ - \frac{1}{n}\log Q \qquad [12.23]$$

Hence (and noting that $\mathrm{pe}^\circ = \mathcal{E}^\circ$),

$$\begin{aligned}\mathrm{pe} &= Eh(\mathcal{F}\ /2.302\,59RT)\\ &= 5040\,Eh/T\\ &= Eh/0.059\,16 \text{ at } 298.15\text{ K} \qquad (12.24)\end{aligned}$$

Thus, if you prefer, you can calculate *Eh* and convert to pe with Equation (12.24), or vice versa. The relationship between the two scales is illustrated in Figure 12.7, where oxidation potential is plotted as both *Eh* and pe. An *Eh*–pH diagram looks exactly like a pe–pH diagram, except that the Y-axis is shifted by the factor $5040/T$.

12.10 Oxygen fugacity

As mentioned in §12.8, there are two methods in common use to represent the same fundamental variable – the oxidation state of a system. It is time to look at the second method, oxygen fugacity. This is a convenient parameter because any reaction that involves a change in oxidation state (any redox reaction) can be written so as to include oxygen as a reactant or product, whether or not oxygen is actually involved in the reaction. We gave an example in Chapter 9 (§9.4.2) of the oxidation of magnetite (in which some of the iron occurs as Fe^{3+} and some as Fe^{2+}) to hematite (in which all of the iron occurs as Fe^{3+}). This reaction often occurs in systems which contain no oxygen molecules at all. Nevertheless, the calculated f_{O_2} ($10^{-68.40}$) for magnetite–hematite equilibrium is a perfectly valid thermodynamic parameter, and the redox reaction involving $O_2(g)$

$$6\,Fe_2O_3(s) = 4\,Fe_3O_4(s) + O_2(g)$$

is simpler than the equivalent reaction in *Eh* mode:

$$3\,Fe_2O_3(s) + 2\,H^+ + 2\ e = 2\,Fe_3O_4(s) + H_2O$$

We will discuss the relative merits of the two methods a little more further on.

12.10.1 Calculation of oxygen fugacity – pH diagrams

Because f_{O_2} and *Eh* are both indicators of the same thing – oxidation state – it is possible to draw $\log f_{O_2}$–pH diagrams that are analogous to the *Eh*–pH calculations we have outlined above. To illustrate this we will construct a $\log f_{O_2}$–pH diagram for the same Mn–H_2O–O_2 system at 25 °C already described. The completed diagram is shown in Figure 12.8 and should be compared with the analogous *Eh*–pH diagram of Figure 12.7. The two diagrams are similar except that most of the phase boundaries on the *Eh*–pH diagram have the Nernst slope, whereas those on the $\log f_{O_2}$–pH diagram have zero slope. Lines on *Eh*–pH diagrams quite typically have nonzero slopes because hydrogen ions and electrons are so commonly involved in half-cell reactions. Reactions balanced with oxygen instead of electrons require H^+ ions much less frequently, and reactions that contain no hydrogen ions have zero slope.

The method of calculating $\log f_{O_2}$–pH boundaries is illustrated with three examples. All other boundaries are derived in the same way. Our examples include the boundaries for water stability and for coexisting minerals, as well

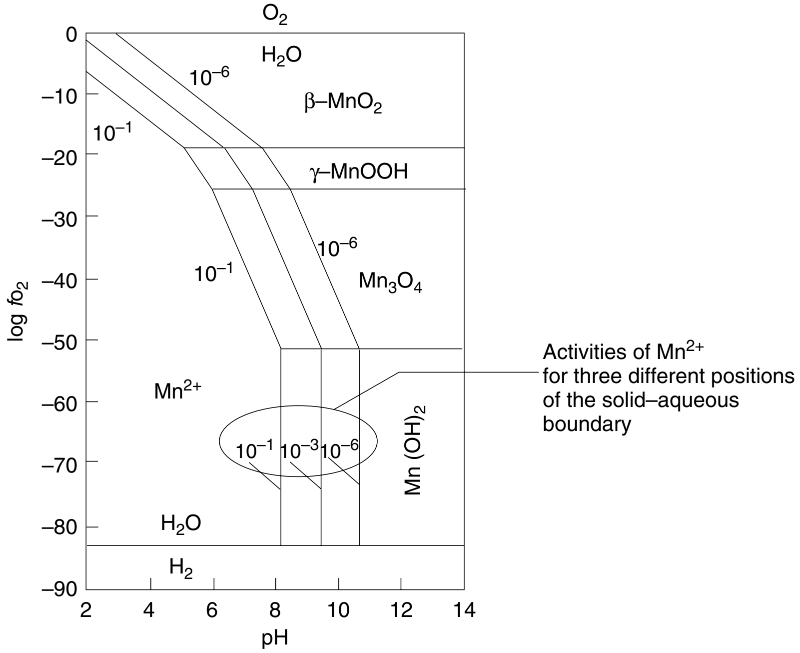


Figure 12.8 $\log f_{O_2}$–pH diagram for the system Mn–H_2O–O_2 at 25 °C, 1 atm.

as the aqueous solubility contours of a mineral. Notice that half-cell reactions are not involved in these calculations.

For the water stability boundaries, the dissociation reaction of water is

$$2H_2O = O_2(g) + 2H_2(g) \tag{12.25}$$

To calculate the equilibrium constant for this reaction,

$$\begin{aligned}\Delta_r G^\circ &= -2(-237\,129)\\ &= 474\,258\ \mathrm{J\,mol^{-1}}\\ &= -RT\ln K\\ K &= 10^{-83.1}\\ &= f_{H_2}^2 f_{O_2}\end{aligned} \tag{12.26}$$

The upper boundary occurs at 1 atm $O_2(g)$ pressure or $\log f_{O_2} = 0$. The lower boundary is at 1 atm $H_2(g)$ pressure; from the equilibrium constant (12.26), this corresponds to $\log f_{O_2} = -83.1$. As noted before, water can exist under conditions outside of these boundaries, but only if the pressure of oxygen or hydrogen is greater than 1 atm.

For the boundary $Mn(OH)_2$–Mn_3O_4,,

$$\tfrac{1}{2}O_2(g) + 3Mn(OH)_2 = Mn_3O_4 + 3H_2O \tag{12.27}$$

$$\begin{aligned}\Delta_r G^\circ &= -143\,093\ \mathrm{J\,mol^{-1}}\\ &= -RT\ \ln(1/f_{O_2}^{\frac{1}{2}})\end{aligned}$$

Hence

$$\log f_{O_2} = -50.14 \tag{12.28}$$

For solubility of Mn_3O_4 as Mn^{2+},

$$3H_2O(l) + 3Mn^{2+} + \tfrac{1}{2}O_2(g) = Mn_3O_4 + 6H^+ \tag{12.29}$$

$$\begin{aligned}\Delta_r G^\circ &= 114\,223\ \mathrm{J\,mol^{-1}}\\ &= -RT\ln\left(\frac{a_{H^+}^6}{f_{O_2}^{\frac{1}{2}} a_{Mn^{2+}}^3}\right)\end{aligned}$$

or

$$\log f_{O_2} = 40 - 12\mathrm{pH} - 6\log a_{Mn^{2+}} \tag{12.30}$$

This is plotted for selected values of Mn^{2+} activity ranging from 10^{-1} to 10^{-6} in Figure 12.8.

12.10.2 Interrelating *Eh*, pH, and oxygen fugacity

The obvious similarity between the *Eh*–pH and $\log f_{O_2}$–pH diagrams of Figures 12.7 and 12.8 suggests that it should be possible to convert directly from one set of coordinates to the other. This can be done using the half-cell reaction (12.15),

$$4H^+ + O_2(g) + 4\,e = 2H_2O(l) \qquad [12.15]$$

and its related Nernst equation,

$$Eh = \mathcal{E}^\circ - \frac{RT}{n\mathcal{F}} \ln[1/(f_{O_2} a^4_{H^+})] \qquad (12.31)$$

This equation can be used to interrelate the three variables *Eh*, pH, and f_{O_2}. The calculations are shown in §12.9.2, leading to an equation (12.17) for the *Eh* of solutions having an f_{O_2} of 1 atm. However, instead of 1 atm, we can just as easily insert any other f_{O_2} value, leading to the more general relation

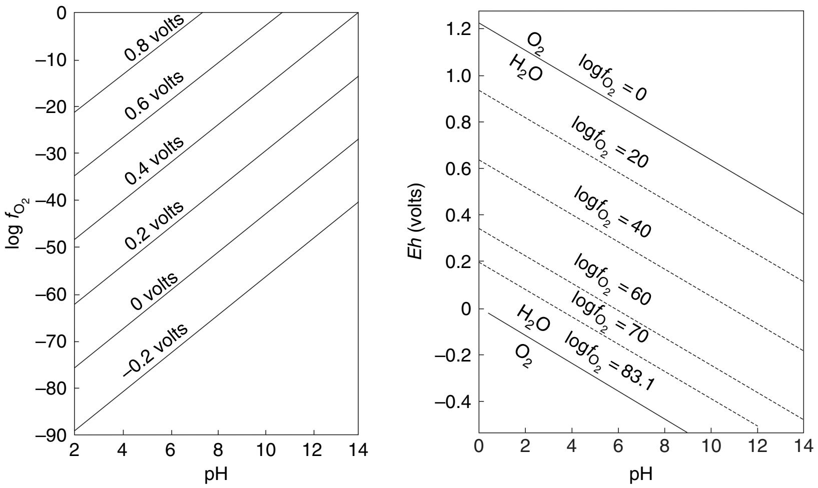
$$Eh = 1.23 + 0.0148 \log f_{O_2} - 0.0592\,\text{pH} \qquad (12.32)$$

Figure 12.9 shows *Eh* contours calculated from (12.32) drawn on a $\log f_{O_2}$–pH diagram, and $\log f_{O_2}$ contours on an *Eh*–pH diagram. The controlling factor is the oxidation state of the system, which in turn is controlled by the bulk composition, and f_{O_2}, f_{H_2}, *Eh*, pe, and all other related variables are simply different ways of quantifying the same thing.

12.11 Redox reactions in organic chemistry

In inorganic reactions we think of hydrogen as always having valence of +1 (H^+) and oxygen as always having valence −2 (O^{2-}) and so we have no difficulty in seeing that carbon in CH_4 is C^{4-}, and carbon in CO_2 is C^{4+}. But what is the oxidation state of carbon in octane, C_8H_{18}, or in acetic acid, CH_3COOH, or in an enzyme containing hundreds of carbon, hydrogen, nitrogen, phosphorous, and other atoms? Applying our "normal" thinking to octane, we get a "nominal" valence for carbon of $18/8 = 2.25$. But how can an atom have a fractional valence? In acetic acid we get $(-4+4)/2 = 0$. But one carbon in acetic acid is bonded to three hydrogens and one carbon, while the other is bonded partly to O and OH. Surely there is some difference in the electron contributions of the two carbons. And in complex entities like large proteins, the idea of counting up carbons and other atoms to get some nominal oxidation state becomes silly.

As pointed out by Helgeson (1991), it is quite possible to assign nominal charge contribution of −1 for each C–H bond, zero for each C–C bond, and +1 for each C–O, C–S, or C–N bond, and arrive at fractional valences that give a consistent accounting of electron transfers in reactions, regardless of the actual



(a) (b)

Figure 12.9 (a) *Eh* contours (in volts) on a $\log f_{O_2}$–pH diagram. (b) $\log f_{O_2}$ contours on an *Eh*–pH diagram. Calculated for 25 °C, 1 atm.

extent to which the shared electrons in the covalent bonds are transferred among the carbon atoms. The fractional valences reflect some average charge on the carbon atoms, which, in the case of the alkanes, changes from -4 in CH_4 to close to -2 as n gets large in $C_nH_{2(n+1)}$. That acetic acid is actually more oxidized than octane is readily seen by writing the reaction

$$2\,C_8H_{18} + 9\,O_2 = 8\,CH_3COOH + 2\,H_2O$$

In fact, any transfer of oxygen or hydrogen in a balanced reaction signifies oxidation–reduction. This can be verified by using a nominal valence scheme as outlined above, but it is always true. Therefore, in looking at reactions between complex molecules, biochemists do not calculate nominal charges to determine if a reaction is an oxidation or a reduction; they simply observe whether oxygen or hydrogen is involved. For example, nicotinamide adenine dinucleotide (AND) exists in two distinct forms, NAD^+ and $NADH + H^+$. The reaction $NAD^+ + H_2 \rightarrow NADH + H^+$ is a reduction, and the reverse reaction is an oxidation. The reaction may involve some molecule much more complex than H_2, but this doesn't matter. We know this reaction is a reduction as written simply because NADH contains one extra hydrogen. Reactions involving electron transfers (redox reactions) are absolutely fundamental to all life processes.

12.12 Summary

On a broad scale the Earth shows a large range of redox conditions, from the highly reduced Ni–Fe core through various silicate layers up into the zone of free water and eventually into the oxygen-rich atmosphere. Therefore, an indicator of the redox state is among the more important of the variables manipulated by geochemists. Like pH, it is an important parameter because it is intimately linked to a large number of reactions of interest to anyone trying to understand the Earth, but again like pH, it is actually no more or less fundamental than any other intensive parameter. For a closed system at a given *T* and *P*, it is completely determined by the bulk composition, as are all intensive parameters, and changes in redox state are accomplished by changing bulk composition.

The measurement of redox conditions by means of a cell voltage, where one electrode has a fixed reference potential and the other is expected to react reversibly with natural systems, is attended by a number of problems. The platinum electrode works well only under certain conditions: It is difficult to get the electrode into reducing environments without allowing some oxidation, and the method is restricted to ambient conditions except in research laboratories. We put up with these problems because there is little choice.

Oxygen fugacity, on the other hand, although a much simpler concept, can be directly measured only at high temperatures, a fact that might seem to rule out its use at Earth-surface conditions, but in fact it does not. Oxygen fugacity can be used (as opposed to *measured*) under any redox conditions, including systems that contain no oxygen molecules whatsoever, as illustrated in Figures 12.10 and 12.11 (they may contain oxygen in combined form, such as H_2O, of course). Oxygen fugacity is an index of redox conditions, not always an approximation to the partial pressure of $O_2(g)$ in a system. It can be

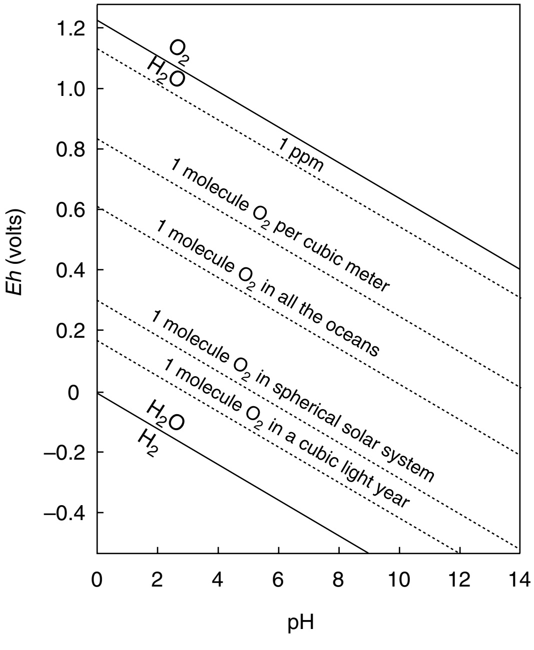


Figure 12.10 An *Eh*–pH diagram with contours of the oxygen concentrations that would result in the redox conditions shown. Obviously, redox conditions over most of the diagram represent solutions that contain no oxygen molecules at all.

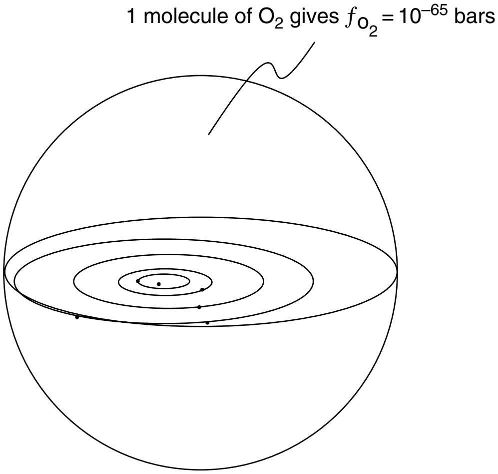


Figure 12.11 The meaning of $f_{O_2} = 10^{-65}$ bars.

calculated because *any* redox reaction can be written so as to include oxygen, whatever is actually happening in the system. In a similar way, it has been pointed out at times that pe is a somewhat fictitious quantity, in that there is good evidence that there are no hydrated electrons in aqueous solutions. In Chapter 1 we emphasized that we would be describing a *model* of chemical reactions, not real reactions. Parameters such as oxygen fugacity illustrate this point. In a sense, we have oxygen and hydrated electrons in our models, though perhaps not in our systems.

Eh has been important in both the measuring and reporting of redox conditions, but an argument can be made that it should be used only in the measurement and not the reporting of redox conditions, that is, that however measurements are *made*, results should be *reported* as f_{O_2} or $\log f_{O_2}$ values. There would be two advantages to this. First, the use of *Eh* entails the use of a relatively complex set of conventions, which are quite difficult to remember unless used continuously. Second, and more importantly, *Eh* is less useful than f_{O_2} because it is so commonly linked in reactions with pH (giving the "Nernst slope"). This means that *a value for Eh without an accompanying value for* pH *is usually meaningless*. This is illustrated by any *Eh*–pH diagram, in which you can see that an *Eh* of 0.0 volts, for example, indicates much more reducing conditions at pH 2 than it does at pH 10. The conversion from an *Eh*–pH point to a $\log f_{O_2}$ value is very simple (equation 12.32). The very small values of f_{O_2} generally obtained at low temperatures should not be a hindrance to its use, which would greatly simplify the reporting and interpretation of any redox conditions.

13

Equations of state

13.1 Introduction

In Chapter 10 we looked at some properties of solutions, derived as deviations from ideal mixing. There was no thought of developing an equation for either (or any) of the substances being mixed, that is, an equation from which we could calculate the properties of each individual compound or phase. If we could do that, we might then extend that approach, and develop equations for the properties of the mixtures. Such equations are equations of state (EoS), and represent a completely different approach to the same problem.

Establishing equations of state for widely occurring substances is a kind of "holy grail" for experimental and theoretical geochemists, because it encapsulates in a compact form all the thermodynamic information about a compound or a system. In this section we consider only EoS for gases and supercritical fluids, except that some EoS also provide the condensation conditions for the gas, i.e., information on the saturation equilibrium. Equations of state in completely different forms also exist for solids and liquids, used extensively in geophysics and astrophysics, but we will not consider these.

We begin by considering the case of the simplest possible substance, the ideal gas.

13.2 The ideal gas

13.2.1 The ideal gas equation

Equations of state are commonly described as equations which relate the P, V and T of a substance. The simplest example is the ideal gas equation,

$$\mathbf{V} = n\frac{RT}{P} \tag{13.1}$$

where n is the number of moles, or

$$V = \frac{RT}{P} \tag{13.2}$$

which expresses the early experimental results of Boyle, Charles, and Gay-Lussac, and can also be derived from statistical mechanics. Thinking of (13.2) as a truncated virial equation is also useful as this immediately implies that it neglects all intermolecular forces, as we will see (§13.5). This equation doesn't work very well (is not very accurate) for real gases, especially at high pressures, but has served as the starting point for a variety of equations which introduce adjustable variables, some of which we will describe below.

13.2.2 Properties of the ideal gas

The thermodynamic properties of the hypothetical ideal gas are of practical interest because, as we will see (§13.2.3, §13.6.1) it is common practice in developing an equation for real systems to first subtract the properties of the ideal gas, which are known, and then deal only with the deviations from these properties. The properties of the ideal gas are in many cases not quite as simple as you might suppose.

Effect of pressure

It is fairly intuitive that many properties of the ideal gas should be independent of pressure, but not independent of temperature. If there is no interaction whatsoever between molecules, which have zero volume, then it should not matter how close together they are (the effect of P). But if we add heat to the gas (we raise the T), that energy cannot disappear, but must be reflected in the thermodynamic properties of the ideal gas.

We can show this analytically without much effort too. In Equation (5.42) we indicated that the effect of pressure on the enthalpy of an ideal gas is zero. That this is also true for the effect of pressure on internal energy, we note that

$$\begin{aligned}\left(\frac{\partial H}{\partial P}\right)_T &= \left(\frac{\partial(U+PV)}{\partial P}\right)_T \\ &= \left(\frac{\partial(U+RT)}{\partial P}\right)_T \\ &= \left(\frac{\partial U}{\partial P}\right)_T \qquad (13.3)\end{aligned}$$

From these relations it is not difficult to show that the effect of P on C_P and C_V is also zero.

On the other hand, Equation (5.41) shows that the effect of pressure on the entropy of ideal gas is not zero, and therefore pressure also affects the ideal gas Gibbs and Helmholtz energies.

Effect of temperature

The effect of temperature on internal energy is the same as its effect on enthalpy, because

$$\begin{aligned}\left(\frac{\partial H}{\partial T}\right)_P &= \left(\frac{\partial(U+PV)}{\partial T}\right)_P \\ &= \left(\frac{\partial U}{\partial T}\right)_P \\ &= C_P \end{aligned} \tag{13.4}$$

The heat capacity of an ideal gas depends on whether the gas is monatomic, diatomic, or whatever, and is derived through statistical mechanics, which we will not go into. For the simplest case, a monatomic ideal gas,

$$\begin{aligned}C_V &= \left(\frac{\partial U}{\partial T}\right)_V \\ &= \tfrac{3}{2}R\end{aligned} \tag{13.5}$$

and because

$$C_P - C_V = R \tag{13.6}$$

then

$$C_P = \tfrac{5}{2}R \tag{13.7}$$

Despite the simplicity of these expressions for the heat capacity of a monatomic ideal gas, expressions for polyatomic molecules such as H_2O can become quite complex. Of course H_2O is not an ideal gas; the question is what is the heat capacity (and other heat capacity-dependent properties) of a gas having molecules with the properties of the H_2O molecule, but with no intermolecular forces? This is a nontrivial problem in statistical mechanics. For H_2O, the results of Wooley (1980) are used. These were fit to an equation by Cooper (1982) which, after some rearranging, was used by Wagner and Pruß (2002) in their equation of state for H_2O (see below). This equation is a summation of simple harmonic oscillator functions, and is

$$\frac{C_P}{R} = b_0 + \sum_{i=1}^{5} b_i \frac{(\beta/T)^2 e^{-\beta_i/T}}{[1 - e^{-\beta_i/T}]^2} \tag{13.8}$$

where b and β are arrays of fit coefficients. Being an equation for heat capacity, it can be integrated to give expressions for Gibbs and Helmholtz energies, as well as enthalpy.

You might reasonably ask at this point, why should we be interested in ideal gas properties, especially when they get so complex, if we want to develop an equation for real gases? The reason is in the form of residual functions.

13.2.3 Residual functions

Residual functions (M^{R}) are defined as

$$M^{\mathrm{R}} = M^{\mathrm{actual}} - M^{\mathrm{ideal\ gas}} \tag{13.9}$$

where M^{actual} is any thermodynamic property of a real system, and $M^{\mathrm{ideal\ gas}}$ is that same property for that system behaving as an ideal gas under the same conditions, either of T and P, or T and ρ.

Because the ideal gas has no intermolecular forces, the various parameters of the ideal gas ($G^{\mathrm{ideal\ gas}}$, $H^{\mathrm{ideal\ gas}}$, etc.) represent those parameters a real fluid would have in the absence of such forces, and the residual functions therefore represent the contributions to those parameters when the intermolecular forces are "turned on" (Prausnitz, et al., 1999, Appendix B). As such they find use in statistical mechanics and in the development of equations of state. Don't confuse residual functions with excess functions.

The functional forms of ideal gas properties and hence of residual functions are well known. For example, the residual enthalpy is

$$H^{\mathrm{R}} = -RT^2 \int_{P_0}^{P} \left(\frac{\partial Z}{\partial T}\right)_P d\ln P$$

the residual entropy is

$$S^{\mathrm{R}} = R \int_{P_0}^{P} \left\{ T \left(\frac{\partial Z}{\partial T}\right)_P + Z - 1 \right\} d\ln P$$

and so on, where Z is the compressibility factor (see below). Ewing and Peters (2000) give a complete list in terms of both (T, P) and (T, ρ). However, knowing the form of the equation is only the beginning. You still have to know how to perform the integration, and this requires real data.

Fugacity coefficient as a residual function

We should note that not only is α a residual function (Equation 8.8), but so in a closely related sense is the fugacity coefficient γ_f. We saw that the Gibbs energy for an ideal gas can be written $dG^{\mathrm{ideal\ gas}} = RT\,d\ln P$ (Equation 8.1), and that for a real gas as $dG = RT\,d\ln f$ (Equation 8.4). So, if $G^{\mathrm{R}} = G - G^{\mathrm{ideal\ gas}}$ is the residual Gibbs energy,

$$dG^{\mathrm{R}} = dG - dG^{\mathrm{ideal\ gas}}$$

$$\int_{P_0}^{P} dG^{\mathrm{R}} = \int_{P_0}^{P} [RT\,d\ln f - RT\,d\ln P]$$

The lower limit of integration is zero on the left and cancels on the right, so

$$\begin{aligned} G^{\mathrm{R}} &= RT \ln \frac{f}{P} \\ &= RT \ln \gamma_f \end{aligned} \tag{13.10}$$

Therefore the fugacity coefficient, in the form $RT \ln \gamma_f$, is a measure of the residual Gibbs energy. Then from (8.13),

$$G^{\mathrm{R}} = RT \int_{P_0}^{P} (Z - 1) d \ln P \tag{13.11}$$

which is a more common way of expressing this function (Ewing and Peters, 2000).

13.2.4 The compressibility factor

J.D. van der Waals noticed, on empirical grounds, that

$$Z = \frac{PV}{RT} \tag{13.12}$$

is the same for most gases, if the comparison is made when each gas has the same relation to its critical point. This is the principle of *corresponding states*. If we define a "reduced temperature"

$$T_{\mathrm{R}} = T/T_{\mathrm{c}}$$

where T is the gas temperature (in kelvins) and T_{c} is its critical temperature (in kelvins), and a "reduced pressure"

$$P_{\mathrm{R}} = P/P_{\mathrm{c}}$$

where P is the gas pressure and P_{c} is its critical pressure, then most gases have about the same Z at a given P_{R}, T_{R}, within certain ranges.

Since van der Waals pointed this out, there have been great advances in our understanding of the fluid state. However, for dense fluids (liquids) these results are quite complex. Ely and Marrucho (2000) point out that

> With this complexity in mind, the most powerful tool available today (just as 25 years ago) for making highly accurate, yet mathematically simple, predictions of the thermophysical properties of fluids and fluid mixtures is the corresponding states principle.

and

> … all modern generalized engineering equations of state are examples of applications of this principle.

van der Waals would be pleased.

The principle of corresponding states is a two-parameter theory and works well only for "simple" molecules, which are the noble gases and a few nonpolar or very slightly polar ones. Two main approaches have been used in expanding its range of applicability. One is to introduce a third parameter, the most successful being the acentric factor, and the other is based on manipulating the intermolecular potentials. We will mention only the acentric factor.

The acentric factor

For molecules more complex than the noble gases, a third parameter is introduced (Pitzer et al. 1955; Pitzer 1955) to form a three-parameter theory of corresponding states. This is the acentric factor ω, which is a function of the acentricity or noncentral nature of the intermolecular forces. It is completely empirical. For simple fluids, it was observed that at a reduced temperature of 0.7, the saturation pressure (i.e., the vapor pressure of the liquid–gas equilibrium) divided by the critical pressure is very close to 1/10, or

$$\frac{P^{\mathrm{s}}}{P_{\mathrm{c}}} = \frac{1}{10}$$

where P^{s} is the saturation pressure. So the acentric factor was defined as

$$\omega = -\log\left(\frac{P^{\mathrm{s}}}{P_{\mathrm{c}}}\right)_{T/T_{\mathrm{c}}=0.7} - 1.0 \tag{13.13}$$

making $\omega = 0$ for simple fluids. This modified theory of corresponding states then says that the properties of gases having the same reduced temperature, reduced pressure and acentric factor are the same.

13.3 Two kinds of EoS

It is important to realize that however valuable *PVT* information is (and it is extremely valuable in the chemical industries), it is not thermodynamically complete information. You cannot calculate a heat capacity from *PVT* data, and this means that you cannot calculate the temperature variation of the Gibbs energy, enthalpy, or entropy. You can calculate the pressure variation of these functions, but you need to start with a baseline showing the variation with T at some pressure. There is a second important class of EoS, sometimes called "thermal EoS," which do provide complete information, and these are equations based on $G(T, P)$ or $A(T, V)$. We will look first at *PVT* equations of state.

Apart from the ideal gas equation, equations relating P, V, and T are usually cubic equations.

13.4 Cubic equations of state

13.4.1 The van der Waals equation

The ideal gas law does not work well for real gases at even moderate pressures. Two of its main problems were recognized by van der Waals, and appropriate corrections were incorporated into his famous equation of state of 1873:

$$P = \underbrace{\frac{RT}{V-b}}_{\text{repulsion}} - \underbrace{\frac{a}{V^2}}_{\text{attraction}} \tag{13.14}$$

where a and b are experimentally derived constants, different for each gaseous compound. The quantity $V-b$ corrects for the volume occupied by molecules of finite volume (molecules of an ideal gas are supposed to be vanishingly small). The parameter b is called the excluded volume per mole, since the volume available for movement of any one molecule is V minus the volume b of all other molecules in the same space. This crowding raises the pressure relative to an ideal gas according to $P = RT/(V-b)$. The second term in (13.14) corrects for intermolecular attractions. The forces between all gas molecules somewhere in the center of a container should be the same in all directions, on the average. However, the average forces on molecules at the outer wall of the container are directed inward; this must reduce the pressure exerted by these outermost molecules on the wall.

The concentration of the outermost zone of molecules is proportional to the gas density $1/V$, as is the concentration of molecules in the next inward zone (presuming concentrations in both zones are the same – a potential source of error). The inward attraction is proportional to the number of molecules in both zones or to $1/V^2$. Introducing the proportionality constant a, this inward pull should modify the outward force or pressure by $-a/V^2$.

The van der Waals equation fits the *PVT* behavior of real gases better than the ideal gas law, but still fails at moderate to high pressures where intermolecular forces are stronger. We would expect it to work a little better for two reasons: First, it is based on a simplistic, but not unreasonable physical picture of a real gas and should therefore have an appropriate mathematical shape. Second, it contains two adjustable parameters a and b which can be fitted to data by regression analysis or other procedures. This gives it flexibility and "personalizes" it for different gases.

The first cubic EoS

But these two correction factors have one other extremely important effect – they transform the simple linear ideal gas equation into a cubic equation. This is perhaps not obvious from the usual presentation in Equation (13.14), but rewriting this as

$$V^3 - \left(b + \frac{RT}{P}\right)V^2 + \frac{a}{P}V - \frac{ab}{P} = 0 \tag{13.15}$$

shows that it is in fact a cubic equation in V. This means that solving the equation for V will produce three roots. The wonderful thing is that given the values of a and b for some gas, solving for V as a function of P will produce one real root and two imaginary roots above and below a certain value of P, and three real, positive roots within a very narrow range of P values. A simple calculation (found by Maxwell in 1875) is used to determine the exact pressure within this range. This pressure represents the two-phase gas–liquid saturation state of the gas. Of the three real roots, the smallest is interpreted

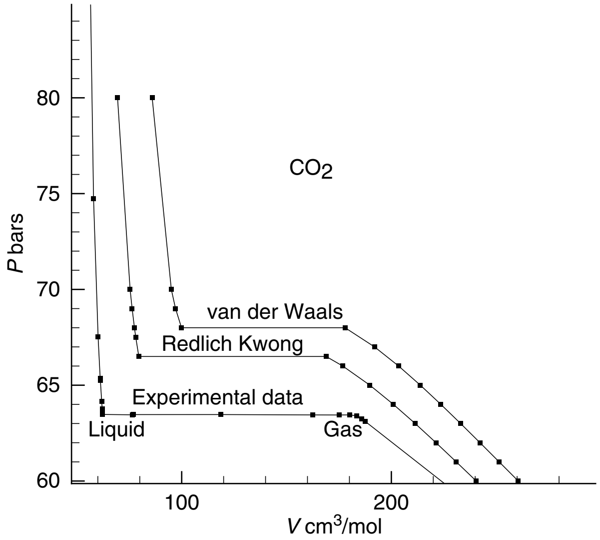


Figure 13.1 Some calculated results using the van der Waals and Redlich–Kwong equations for CO_2 at 25 °C, compared with the experimental data of Michels et al. (1935).

as the liquid volume, the largest as the gas volume, and the intermediate one as a physically meaningless metastable state. Some results for CO_2 are shown in Figure 13.1

The van der Waals equation does remarkably well at low pressures, say up to 10 bars or so, but is not sufficiently accurate for industrial applications. However it obviously has the right form, as shown by the fact that literally hundreds of modifications have been proposed, and the cubic equation of state is a huge topic in some areas of chemical engineering because of the importance of modeling vapor–liquid equilibria (VLE) in some industries. Walas (1985) provides a good introduction, and more recent advances can be found in Orbey and Sandler (1996) and Anderko (2000).

To give some idea of the field we will briefly discuss a couple of the many modifications of the van der Waals equation that have found some use in the Earth sciences. Our objective here is to just to impart a sense of this large and important field of applied thermodynamics. A whole book would be required to cover the field in any detail.

13.4.2 Modifications of the van der Waals equation

The Redlich–Kwong equation

A very widely used modification of the van der Waals equation published by Redlich and Kwong in 1949 is

$$P = \frac{RT}{(V-b)} - \frac{a}{T^{0.5}\,V(V+b)} \tag{13.16}$$

As you can see, this retains the first term of the van der Waals equation but modifies the second term.

Since the RK equation first appeared, it too has undergone many modifications in an attempt to further improve its ability to fit data for gases. Many of these modified Redlich–Kwong equations are summarized by Holloway (1977) and Kerrick and Jacobs (1981), and involve writing the *a* and *b* parameters as a function of *T* and/or *P*. The MRK equation of Kerrick and Jacobs keeps *b* constant but writes *a* as a function of both *P* and *T*:

$$P = \frac{RT(1+y+y^2-y^3)}{V(1-y)^3} - \frac{a(P,T)}{T^{0.5}V(V+b)} \tag{13.17}$$

Redlich–Kwong and van der Waals equations

For CO_2, $T_c = 304.21$ K, $P_c = 73.285$ bars, and at 298.15 K Equations (13.19) and (13.20) give $a = 3655\,953$ and $b = 42.8269$.

Solving Equations (13.14) and (13.16) for *V* as a function of *P* gives the following results:

van der Waals		Redlich–Kwong	
P, bars	*V*, $\text{cm}^3\,\text{mol}^{-1}$	*P*, bars	*V*, $\text{cm}^3\,\text{mol}^{-1}$
60	260.5*	60	240.5*
61	251.4*	61	231.0*
62	242.3*	62	221.3*
63	233.0*	63	211.4*
64	223.6*	64	201.0, 100.0, 86.4
65	213.8*	65	189.8, 109.0, 82.6
66	203.4*	66	176.8, 118.4, 80.4
67	106.2, 114.6, 192.0	67	168.9, 124.3, 79.5
68	99.9, 129.5, 177.9	68	77.5*
69	97.1*	69	76.3*
70	95.1*	70	75.4*
80	85.9*	80	69.4*

* indicates there are two additional complex roots

These results are plotted in Figure 13.1, but only one liquid–gas equilibrium pressure is shown.

Note: If you solve these equations, or any cubic, using a spreadsheet solver, only one root is reported. To see all three roots you must use one of the analytical methods, or a program such as Mathematica or Maple.

where

$$y = b/4V \quad (b \text{ is constant})$$

$$a(P,T) = c(T) + d(T)/V + e(T)/V^2$$

$$z(T) = z_1 + z_2T + z_3T^3 \quad (\text{where } z \text{ is } c,\ d,\text{ or } e).$$

Here you can see that the repulsion term holds b constant but is more complicated (which gives it better flexibility). The a parameter in the attraction term is written as a simple polynomial function of $1/V$ (hence of P) and of T. This is a MRK equation with 10 adjustable parameters, so naturally it fits real data considerably better than the original RK equation, which has only two parameters.

The Peng–Robinson equation

Peng and Robinson (1976) proposed the equation

$$P = \frac{RT}{V-b} - \frac{a}{V(V+b)+b(V-b)} \tag{13.18}$$

with

$$a = (0.457\,235R^2T_c^2/P_c)\,\alpha$$

$$b = 0.077\,796RT_c/P_c$$

$$\alpha = [1 + \kappa(1 - T_R^{0.5})]^2$$

where T_R is the reduced temperature (§13.2.4), and κ is a function of the acentric factor.

This equation is widely used for hydrocarbons and some nonhydrocarbon gases such as CO_2 and N_2, but does not work very well for H_2O and other polar gases. It too has had many modifications, such as those of Stryjek and Vera (1986a,b,c), which greatly improves the results for polar compounds (PRSV equation). This improvement was made, in Stryjek and Vera (1955c), by modifying κ to

$$\kappa = \kappa_0 + [\kappa_1 + \kappa_2(\kappa_3 - T_R)(1 - T_R^{0.5})](1 + T_R^{0.5})(0.7 - T_R^{0.5})$$

where

$$\kappa_0 = 0.378\,893 + 1.489\,7153\omega - 0.171\,318\,48\omega^2 + 0.019\,6544\omega^3$$

and κ_1, κ_2, and κ_3 are adjustable parameters for each compound.

This equation proved useful to Suleimenov and Krupp (1994) in their study of the important system H_2S-H_2O. Stryjek and Vera (1986b, c) also proposed mixing rules for the use of the PRSV equation with mixtures of gases (§13.4.4).

13.4.3 Estimating EoS parameters from critical conditions

The a and b parameters in the RK and van der Waals class of equations are best estimated from a regression fit to an overall *PVT* data set for a gas. However, there is a much quicker way of estimating the same parameters based on the critical behavior of the gas. If these critical data are known with sufficient accuracy and precision, the estimates should be very good, and this approach is used quite frequently when the utmost accuracy is not required.

The first and second derivatives of the critical isotherm must both be zero at the critical point. The van der Waals equation itself should also work at the critical point, so we can write the three following equations:

$$P_c = \frac{RT_c}{V_c - b} - \frac{a}{V_c^2}$$

$$(\partial P/\partial V)_{T_c} = 0 = \frac{-RT_c}{(V_c - b)^2} + \frac{2a}{V_c^3}$$

$$(\partial^2 P/\partial V^2)_{T_c} = 0 = \frac{2RT_c}{(V_c - b)^3} - \frac{6a}{V_c^4}$$

Solving these three equations for the critical constants gives

$$T_c = 8a/27bR$$

$$V_c = 3b$$

$$P_c = a/27b^2$$

and solving for the van der Waals parameters, we find

$$\begin{aligned} a &= 3P_cV_c^2 \\ &= \frac{27R^2T_c^2}{64P_c} \end{aligned} \tag{13.19}$$

$$\begin{aligned} b &= V_c/3 \\ &= \frac{RT_c}{8P_c} \end{aligned} \tag{13.20}$$

where

$$V_c = 3RT_c/8P_c$$

We can apply exactly the same reasoning to the RK equation (13.16) and obtain

$$a_{\mathrm{RK}} = 0.427\,480\,R^2T_c^{2.5}/P_c$$

$$b_{\mathrm{RK}} = 0.086\,640\,RT_c/P_c$$

Thus the a and b constants in both the van der Waals and the RK equations can be calculated quite easily given sufficiently good data on the critical temperature and pressure of a fluid. Alternatively, you can calculate the critical conditions for a gas directly from the a and b parameters for either equation of state.

13.4.4 EoS for mixtures – mixing rules

A natural question to ask is how these equations might be applied to mixtures of gases. To use the van der Waals equation (13.2) or any of its modifications, the a and b parameters must be known for each pure component. If no experimental data are available for a mixture of these components, how do we get values of a and b that could apply to the mixture? Many so-called mixing rules have been devised to do this (see, for example, Prausnitz et al. 1999), but the most popular remain those originally suggested by van der Waals. Since the repulsive b parameter is supposed to be a measure of the volume of the molecules, a simple averaging over all m different gas species might work:

$$b_{\text{mix}}(x_1, x_2, \ldots) = \sum_{j=1}^{m} x_j b_j \tag{13.21}$$

The a (molecular attraction) parameter for the mixture is obtained by averaging over all possible kinds of molecular pairs:

$$a_{\text{mix}}(x_1, x_2, \ldots) = \sum_{j=1}^{m} \sum_{k=1}^{m} x_j x_k a_{jk} \tag{13.22}$$

where

$$a_{jk} = (a_j a_k)^{0.5} \tag{13.23}$$

In (13.22) and (13.23), a_{jk} is supposed to describe the interaction between one molecule of type j and another of type k. For $j = k$, $a_{jk} = a_{jj}$ and is simply the van der Waals a parameter for the pure gas j. The problem is how to find a term $a_{jk}(j \neq k)$ for the interaction between unlike molecules when there are no data on the mixture itself. Berthelot suggested the geometric mean assumption of Equation (13.23) on strictly empirical grounds, and it was used by van der Waals in his own work on gas mixtures. For some cases this assumption can be theoretically justified, but why it works is still a bit of a mystery. As can be imagined, there has been much discussion and controversy over this point, and equations for mixing rules can become quite complex.

A completely different approach to mixing rules is the "one fluid" model, in which it is not a and b which are modified, but the intermolecular potentials. We have a look at this method in §13.7.1.

13.5 The virial equation

In 1901 Kammerlingh-Onnes first used a power series to describe the *PVT* properties of a gas.[1] This is now known as a virial equation of state, and can be written in various related ways:

$$\frac{PV}{RT}=1+\frac{B}{V}+\frac{C}{V^2}+\frac{D}{V^3}+\cdots \tag{13.24}$$

or, because $V=1/\rho$,

$$\frac{P}{RT}=\rho+B\rho^2+C\rho^3+D\rho^4+\cdots \tag{13.25}$$

Often it is more convenient to use P and T as variables, in which case the equation becomes

$$\begin{aligned}\frac{PV}{RT}&=Z\\&=1+B'P+C'P^2+D'P^3+\cdots\end{aligned} \tag{13.26}$$

The coefficients of the terms in P or $1/V$ are known as virial coefficients. The coefficients in the pressure form of Equation (13.26) are related to those in the volume or ρ form by $B'=B/RT$ and $C'=(C-B^2)/(RT)^2$. If (13.26) is written

$$PV=RT+B'P+C'P^2+D'P^3+\cdots \tag{13.27}$$

then $B'=B$, and $C'=(C-B^2)/RT$.

At first sight this looks like nothing more than a polynomial expansion of the ideal gas law. However, it turns out to have real physical significance. Statistical mechanics shows that the second coefficient arises from the interaction of pairs of molecules, the third from the interaction of molecules three at a time, and so on. They can be calculated from known interaction potentials, or used to estimate such potentials from observed *PVT* behavior. The details can be found in most textbooks on statistical mechanics (for example McQuarrie, 2000, Chapter 12), and Prausnitz et al. (1999) give an extensive treatment of various commonly used formulations of these intermolecular forces and their use in equations of state.

Because it is an equation of state, it (at least in principle) allows calculation of all thermodynamic parameters. For example, chemical potentials in terms of virial parameters are (starting with (13.25))

$$\mu-\mu^\circ=RT\ln(\rho RT)+2B\rho+\frac{3C'\rho^2}{2}+\frac{4D'\rho^3}{3}+\cdots \tag{13.28}$$

---

[1] Kammerlingh-Onnes, a colleague of van der Waals, is possibly better known for liquefying helium in 1908, discovering superconductivity in 1911, and receiving the Nobel Prize in Physics in 1913.

Other examples will be cited later on. For a complete summary, see Pitzer (1995, p. 133).

The virial equation approach has also been used for providing an equation of state for aqueous solutes, including electrolytes. In the 1970s, Kenneth Pitzer and his associates developed such equations, which have proven to be quite successful at fitting the behavior of both single- and mixed-salt solutions to high concentrations. This will be treated in more detail in §13.6.3 and Chapter 15.

13.5.1 Fugacities from virial equations

For a pure phase, we can rewrite Equation (13.26) as

$$Z - 1 = B'P + C'P^2 + D'P^3 + \cdots \tag{13.29}$$

and combining this (omitting the D' term) with (8.13), we get

$$\begin{aligned}\ln \gamma_f &= \int_{P_0}^{P} (B'P + C'P^2)\, d\ln P \\ &= \int_{P_0}^{P} \frac{B'P + C'P^2}{P}\, dP \\ &= B'P + \tfrac{1}{2}C'P^2 \end{aligned} \tag{13.30}$$

where the terms referring to the lower limit of integration (P_0) drop out, because P_0 is very small, theoretically zero. Also note that (13.30) is also the residual Gibbs energy G^{R}/RT, from (13.10). From this you can see that in fact all residual thermodynamic properties are also easily expressed in terms of virial coefficients. A complete list is given by Trusler (2000).

The same equations are used for mixtures of gases. In this case the coefficients are combinations of the coefficients of the pure components. They are combined according to "mixing rules," which are also specified by statistical mechanics for any number of components. The second virial coefficient is, for a mixture of m components,

$$B_{\mathrm{mix}} = \sum_{i=1}^{m}\sum_{j=1}^{m} x_i x_j B_{ij} \tag{13.31}$$

which, for a binary system with components 1 and 2 becomes (because $B_{12} = B_{21}$)

$$B_{\mathrm{mix}} = x_1^2 B_{11} + 2x_1 x_2 B_{12} + x_2^2 B_{22} \tag{13.32}$$

so there are in fact three second virial coefficients, which require experimental data for three systems – pure 1, pure 2, and mixtures of 1 and 2. The third coefficient is given by

$$C_{\mathrm{mix}} = \sum_{i=1}^{m}\sum_{j=1}^{m}\sum_{k=1}^{m} x_i x_j x_k C_{ijk} \tag{13.33}$$

Fugacity of water from virial coefficients

Dymond and Smith (1980) present an extensive collection of both second and third coefficients for many compounds at many temperatures from various investigators. For water at 400 °C, we select the data of Kell, McLaurin and Whalley, (1968), who give $B = -73.47\,\mathrm{cm^3\,mol^{-1}}$, $C = 1110\,\mathrm{cm^2\,mol^{-2}}$. Converting these to the pressure form of the equation (§13.5),

$$B' = \frac{-73.47}{(83.1451 \times 673.15)}$$

$$= -0.001\,3127$$

and

$$C' = \frac{1110 - (-73.47^2)}{(83.1451 \times 673.15)^2}$$

$$= -1.368\,81 \times 10^{-6}$$

Inserting these values into Equation (13.30), we get these results, which are plotted in Figure 13.2.

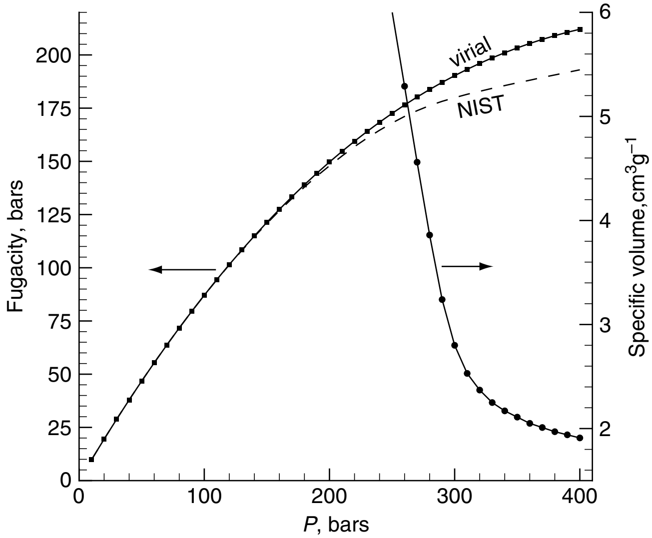


Figure 13.2 The fugacity of water at 400 °C as a function of pressure, calculated from virial coefficients and Equation (13.30). The dashed line represents fugacity from the NIST program STEAM.

(continued)

T °C	P bars	$\ln \gamma_f$ –	f_{H_2O} bars	f_{H_2O} from NIST	sp. vol. $cm^3\,g^{-1}$
400	10	−0.0132	9.87	9.87	306.56
400	100	−0.1381	87.10	87.06	26.43
400	200	−0.2899	149.67	147.90	9.95
400	300	−0.4554	190.26	179.90	2.80
400	400	−0.6346	212.06	193.00	1.91
400	500	−0.8274	218.58	204.50	1.73

As you can see, the results are quite good up to about 200 bars, that is, with H_2O in the gas phase. As water becomes more liquid-like and the critical density of $3.22\,cm^3\,g^{-1}$ is approached the results get progressively worse.

which can be expanded in the same way, giving, for two components and because $C_{112} = C_{121} = C_{211}$, and $C_{122} = C_{212} = C_{221}$

$$C = x_1^3(C_{111} + x_1^2 x_2\,(3C_{112}) + x_1 x_2^2\,(3C_{122}) + x_2^3 C_{222} \tag{13.34}$$

so quite a lot of data are required to use these equations. In fact, well-determined third virial coefficients are extremely rare, and many applications omit this term.

Dropping the third (C) coefficient, Equation (13.29) for a mixture of gases becomes

$$Z = \frac{PV}{RT} = 1 + B'_{\mathrm{mix}}P, \tag{13.35}$$

or

$$\begin{aligned}\frac{P\mathbf{V}}{n_T RT} &= 1 + P\frac{B_{\mathrm{mix}}}{RT}\\ \frac{P\mathbf{V}}{n_T} &= RT + P\sum_{i=1}^{m}\sum_{j=1}^{m} x_i x_j B_{ij}\\ &= RT + \frac{P\sum_{i=1}^{m}\sum_{j=1}^{m} n_i n_j B_{ij}}{n_T^2},\end{aligned}$$

so

$$\mathbf{V} = \frac{n_T RT}{P} + \frac{\sum_{i=1}^{m}\sum_{j=1}^{m} n_i n_j B_{ij}}{n_T} \tag{13.36}$$

where n_T is the total number of moles in the mixture, $x_i = n_i/n_T$, and B_{mix} is given by (13.31) and (13.32). Expressions for the chemical potential and fugacity coefficient are then derived by substituting $\mathbf{V}$ from (13.36) into

$$\left(\frac{\partial \mathbf{G}}{\partial P}\right)_{T,n_i,n_j} = \mathbf{V}$$

and integrating, obtaining (Pitzer, 1995, Equation 13.6),

$$\mathbf{G}_i - \mathbf{G}_i^\circ = RT \ln \frac{P}{P_0} + P\left(2\sum_{j=1}^{m} x_j B_{ij} - B_{\text{mix}}\right) \tag{13.37}$$

where $\mathbf{G}_i^\circ$ is for the mixed ideal gases, and so is $\sum_i \mathbf{G}_i^\circ + RT\sum_i n_i \ln x_i$. Then differentiating (13.37) with respect to n_i you get $\mu_i - \mu_i^\circ$, and from $\ln \gamma_f = (\mu - \mu^{\text{ideal gas}})/RT$ you obtain (Pitzer, 1995, Equation 13.8; Prausnitz et al., 1999, Equation 5.33)

$$\ln \gamma_{f_i} = \frac{P}{RT}\left(2\sum_{j=1}^{m} x_j B_{ij} - B_{\text{mix}}\right) \tag{13.38}$$

which for two components, using (13.32), and changing all x_1 terms to $(1-x_2)$, becomes

$$\ln \gamma_{f_1} = \frac{P}{RT}\left[(1-x_2^2)B_{11} + x_2^2(2B_{12} - B_{22})\right] \tag{13.39}$$

and similarly for component 2.

All these equations can also be written in terms of T and V or ρ, rather than T and P. See Pitzer (1995) and Prausnitz et al. (1999) for a complete summary.

13.5.2 Limitations of the virial EoS

The virial equation of state, first proposed on an empirical basis in the nineteenth century and later found to have a solid theoretical basis in statistical mechanics, is sometimes dismissed by geochemists because it only works well at low to moderate densities (see §13.5.1), and there have not been many direct applications in the geochemical literature – Spycher and Reed (§13.7.3) is an exception. Also it is incapable of representing vapor–liquid equilibria, as do the cubic EoS. However, it is important in any study of the thermodynamics of fluids because references to it are ubiquitous in the literature on equations of state, so an understanding of it and its limitations is fundamental.

From Equation (13.26) we see that the first virial coefficient (that is, omitting the second, third, etc., coefficients) is simply the ideal gas law, and that with two coefficients, $(\partial Z/\partial P)_T = B'$, which is the slope of an isotherm on a plot of Z versus P. A typical plot of Z as a function of P is shown in Figure 13.3, patterned after a similar figure in Pitzer (1991, 1995). This shows that a virial equation truncated at the second coefficient will be quite adequate to 25 bars or so at low T and to much higher pressures at higher T. It also shows that B' (or B) can be positive (repulsive intermolecular forces dominate, higher Ts) or negative (attractive intermolecular forces dominate, lower Ts). At pressures above 100 bars the low-temperature isotherms actually change slope, as shown by the inset, and repulsive forces dominate as the molecules are forced closer together. In this region the virial equation approach fails completely; the intermolecular forces become too complex. Statistical mechanics provides exact equations for the coefficients but experimentally, third and higher coefficients are quite rare, and are really of no practical use.

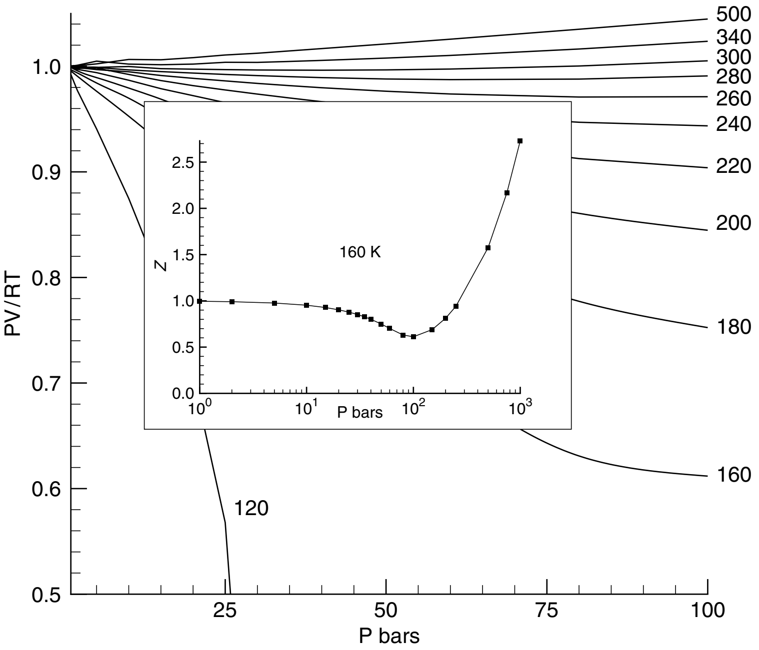


Figure 13.3 Isotherms of the compressibility factor (*PV/RT*) for nitrogen. Temperatures are in kelvins. The inset shows the isotherm for 160 K at higher pressures. Data from Span et al. (2000).

Effect of association

The attractive forces between molecules can lead to association. The association of N_2 molecules into N_4 molecules does not happen, but in other systems it does. For example, the reaction

$$2\,NO_2 = N_2O_4$$

does take place. If nitrogen did associate completely to N_4, the pressure at any T would be reduced by half, as would Z. Figure 13.3 shows that the 120 K isotherm does reach 0.5 at about 25 bars, so in the absence of other information, a possible interpretation would be that N_2 has associated completely to N_4. And, in fact the curvature is negative (the third virial coefficient is negative), so we could assume even larger clusters. However, we note that at higher temperatures, B is positive, which would correspond to negative association, which is impossible.

The point is that it is risky to draw conclusions about the details of chemical reactions on the basis of volumetric data alone, or in fact of any indirect evidence. This point will become important when we discuss the two main interpretations of electrolyte association in Chapter 15. Nevertheless, many substances do show weak to strong association into clusters, and this "chemical interpretation" of gas nonideality is explored extensively in Prausnitz et al. (1999, §5.8).

13.6 Thermal equations of state

As already mentioned, thermal EoS are those capable of directly producing values of all thermodynamic properties, including heat capacity, entropy, enthalpy, and so on. They are based either on a formulation in terms of the Gibbs energy, if the independent variables are T and P, or in terms of the Helmholtz energy if the independent variables are T and V (or ρ). Over much of P–T space, either formulation is quite satisfactory, but when the two-phase vapor–liquid region is involved, the Helmholtz energy (A) is invariably chosen, because for a given T and P, A has two different values for the vapor and liquid states, and so is capable of reporting different values for all parameters for these states. Gibbs energy, on the other hand, has the same value for the coexisting vapor and liquid, and so an equation for G cannot provide thermodynamic values for each of the coexisting phases.

The development of these all-encompassing EoS has become a very large and complex subject, which is more mathematical than thermodynamic. For a recent summary, see Span (2000). It is not unusual for these equations, or rather sets of equations, to have more than 50 adjustable parameters, and even the form of the equations is now adjustable by computer routines. The equations are far too complex to present in detail here, and even programming them from the literature is a nontrivial task. Fortunately many are available in executable form.

We will discuss just a few of these, which are useful to Earth scientists. Thermal EoS have been developed only for pure substances and some binary systems. For more complex systems some form of cubic or virial equation is normally used.

13.6.1 Water substance

Water is generally thought of as a liquid, and the gaseous and solid forms get their own names, so for generality the term “water substance” is used to include all forms of H_2O.

Water is arguably the most important compound on Earth, from many points of view. An equation of state for water was the goal for many researchers for many years, and it proved to be a very difficult task, both experimentally (doing the experiments to obtain the data at high T and P) and theoretically (finding an equation to fit the data). Earth scientists have contributed their share of attempts.

Burnham, Holloway, and Davis

Burnham et al. (1969a) made extensive density measurements at high T and P, and published (1969b) tables of the important thermodynamic properties up to 1000 °C and 10 000 bars. As mentioned earlier, PVT measurements alone cannot provide complete thermodynamic information, so these data were used to expand upward in pressure an earlier published formulation of the thermodynamic properties (Bain, 1964) which extended only to 1000 bars. No attempt was made to include values for ice or water vapor, and despite the apparent simplicity of the isochore shapes (Figure 6.1) they were forced to use three different sets of fit coefficients to cover different subregions of the whole T–P region. Many newer formulations have even more subregions. These data were later superseded, but are sufficiently accurate for most Earth science applications, and are still used at times (e.g., Aranovitch and Newton, 1996, 1997).

Helgeson and Kirkham

Helgeson and Kirkham (1974a) reworked all the data available at that time, and include extensive references to earlier work. They used the data of Burnham et al. as well as other data, and produced equations for many thermodynamic properties not considered by Burnham et al., including the dielectric constant. They covered the region up to 900 °C and 10 kbar.

IAPWS formulations

The International Association for the Properties of Water and Steam (IAPWS; up to 1989 just IAPS) has for many years sponsored the development of increasingly accurate equations of state for water.

Haar, Gallagher and Kell

The formulation by Haar, Gallagher and Kell in 1984 (IAPS-84) was a major step forward in completeness and accuracy, and is incorporated into program SUPCRT92. It is sufficiently accurate for all routine geochemical work, but has now been superseded by IAPWS-95.

Wagner and Pruß

The latest IAPWS formulation is IAPWS-95, and is described by Wagner and Pruß (2002).[2] Software implementing this formulation is available from the National Institute of Standards and Technology (NIST) in the form of program STEAM, written by Harvey, Peskin and Klein (2000). A paper by Hill (1990) contains graphical comparisons of several previous equations of state.

Using the Wagner and Pruß notation, the basic equation has the form

$$f(\rho, T) = f^{\circ}(\rho, T) + f^{\mathrm{r}}(\rho, T) \tag{13.40}$$

where the Helmholtz energy f uses ρ instead of V, and is split into an ideal gas part (f°) and a residual part (f^{r}). This practice is common in the development of equations of state (§13.2.3). The Helmholtz energy is then transformed into

---

[2] Information on the IAPWS formulations is available at www.iapws.org.

a dimensionless form by dividing by RT, and given another symbol, thus $\phi = f/RT$, and

$$\phi(\delta, \tau) = \phi^\circ(\delta, \tau) + \phi^{\mathrm{r}}(\delta, \tau) \tag{13.41}$$

where $\delta = \rho/\rho_c$ is the reduced density, and $\tau = T_{\mathrm{c}}/T$ is the inverse of the reduced temperature.[3]

The equation of state for water is one of the most commonly used equations by geochemists. It is used (albeit in an older version) by SUPCRT92 every time you enter a reaction containing H_2O. Outside of SUPCRT92, properties are easily obtained from program STEAM. So although we have no need to use the equations rather than the programs, it is interesting to see just how complex they are.

The molar Helmholtz energy of an ideal gas, using Wagner and Pruß notation (in which ° signifies ideal gas), is

$$\begin{aligned} f^\circ(\rho, T) &= u^\circ - Ts^\circ(\rho, T) \\ &= [h^\circ - Pv^\circ(\rho, T)] - Ts^\circ(\rho, T) \\ &= h^\circ - RT - Ts^\circ(\rho, T) \end{aligned}$$

When c_p° is inserted into the expression for h° and s°, we get

$$f^\circ(\rho, T) = \left(\int_{T_0}^{T} c_p^\circ dT + h^\circ\right) - RT - T\left[\int_{T_0}^{T} \frac{c_p^\circ - R}{T} - R\ln\left(\frac{\rho}{\rho_0^\circ}\right) + s^\circ\right]$$

where subscript 0 refers to an arbitrary reference state. Next we substitute Equation (13.8) for c_p°, and we get

$$\phi^\circ = \ln\delta + n_1^\circ + n_2^\circ\tau + n_3^\circ\ln\tau + \sum_{i=4}^{8} n_i^\circ \ln\left[1 - e^{-\gamma_i^\circ\tau}\right] \tag{13.42}$$

where the n terms are various combinations of fit coefficients, and other terms are defined with Equation (13.41).

This equation for the ideal gas is then combined with the residual part, which is

$$\begin{aligned} \phi^{\mathrm{r}} = &\sum_{i=1}^{7} n_i\delta^{d_i}\tau^{t_i} + \sum_{i=8}^{51} n_i\delta^{d_i}\tau^{t_i}e^{-\delta^{c_i}} \\ &+ \sum_{i=52}^{54} \delta^{d_i}\tau^{t_i}e^{-\alpha_i(\delta-\epsilon_i)^2-\beta_i(\tau-\gamma_i)^2} + \sum_{i=55}^{56} n_i\Delta^{b_i}\delta\psi \end{aligned} \tag{13.43}$$

---

[3] Using the inverse reduced T actually makes ϕ a Massieu function (Callen, 1960) rather than a Helmholtz function, but this nicety is generally ignored.

with

$$\Delta = \theta^2 + B_i[(\delta - 1)^2]^{a_i}$$

$$\theta = (1 - \tau) + A_i[(\delta - 1)^2]^{1/(2\beta_i)}$$

$$\psi = e^{-C_i(\delta-1)^2 - D_i(\tau-1)^2}$$

and where d and t are arrays of integers, A, B, C, and D are more coefficients, and other symbols are as described above. Equations (13.42) and (13.43) together form the IAPWS-95 equation of state for H_2O.

In addition to this formulation "for general and scientific use," IAPWS also maintains a separate formulation for the special requirements of the steam power industry, the latest of which is IAPWS-IF97 (Wagner et al., 2000). This is also based on the Wagner and Pruß equation, but sacrifices some accuracy for computing speed. However the differences are very small, except in the near vicinity of the critical point. It is divided into five subregions, giving rise to slight discontinuities at the boundaries, whereas IAPWS-95 is in the form of a single equation.

Steam tables

The American Society of Mechanical Engineers (ASME) has produced a book of Steam Tables (Parry et al., 2000) based on IAPWS-IF97, continuing a tradition started by Regnault in 1847. These are detailed tables and charts of the thermodynamic and transport properties of water, which as mentioned above, are perfectly adequate for most conceivable geochemical purposes. Many other organizations in other countries have also published Steam Tables based on the IAPWS-IF97 formulation.

Note, however, that the conventions adopted in Steam Tables are quite different from the "formation from the elements" convention we use in this book, and in all of chemistry and geochemistry. Steam Table values of thermodynamic properties, as well as the compilation of Burnham et al. (1969b) and the NIST program STEAM, report the values of internal energy and entropy as the difference between those properties at the T and P of interest and the same properties of liquid water at the triple point of water. Because these properties are given symbols like u and s, rather than Δu and Δs, this is equivalent to using the convention $u_{\text{triple}} = s_{\text{triple}} = 0$, which, as P_{triple} and v_{triple} are absolute values, also defines the reference values for h, g and a. Steam Table values can be converted to other reference systems by adding the values of s_{triple} and u_{triple} consistent with the desired convention to the Steam Table values. With u and s converted, h, g and a can then be calculated. The values for these quantities are shown in Table 13.1.

The properties of water are available from SUPCRT92, but they are, as mentioned, based on the equation of state of Haar et al (1984). To use the new formulation (IAPWS-95) they must be corrected. Let H_{ST}, S_{ST}, etc. represent

Table 13.1 *Some critical and triple point parameters for water, many from Helgeson and Kirkham (1974a).*

Critical parameters	
T_c	647.096 K
P_c	22.064 MPa
ρ_c	322 kg m^{-3}
Triple point parameters for liquid water	
T_{triple}	273.16 K
P_{triple}	611.657 Pa
V_{triple}	18.018 cm^3 mol^{-1}
$\Delta_f S_{\text{triple}}$	63.31 J mol^{-1} K^{-1}
$\Delta_f U_{\text{triple}}$	−284 039 J mol^{-1}
$\Delta_f H_{\text{triple}}$	−287 721 J mol^{-1}
$\Delta_f G_{\text{triple}}$	−235 517 J mol^{-1}

values obtained from Steam Tables, and $\Delta H_{\text{SUPCRT92}}$, $\Delta S_{\text{SUPCRT92}}$, etc. represent the "apparent" properties that would be calculated by SUPCRT92, as explained in §3.5.5. $\Delta_f H^\circ_{\text{triple}}$, $\Delta_f S^\circ_{\text{triple}}$, etc. are the normal "formation from the elements" properties for liquid water triple point. The corrections for U, H, and S are

$$\left.\begin{aligned} \Delta_a U_{P,T} &= \Delta U_{\text{SUPCRT92}} = U_{\text{ST}} + \Delta_f U^\circ_{\text{triple}} \\ \Delta_a H_{P,T} &= \Delta H_{\text{SUPCRT92}} = H_{\text{ST}} + \Delta_f H^\circ_{\text{triple}} \\ S_{P,T} &= S_{\text{SUPCRT92}} \quad = S_{\text{ST}} + \Delta_f S^\circ_{\text{triple}} \end{aligned}\right\} \tag{13.44}$$

However, G and A are different, because although G_{ST} is simply calculated as $H_{\text{ST}} - T \cdot S_{\text{ST}}$, the value of S_{ST} is not third law entropy. The correction for G is

$$\Delta_a G_{P,T} = \Delta G_{\text{SUPCRT92}} = G_{\text{ST}} + \Delta_f G^\circ_{\text{triple}} - (T - T_{\text{triple}}) \cdot S_{\text{triple}} \tag{13.45}$$

and similarly for A. Values for the molar or specific volumes are absolute properties and need not be changed.

It may seem that while the Steam Table values are different from our standard state "formation from the element" values, *differences* between values at two different T–P states would be the same. This is true for U, V, S, and H, but not true for G and A (see example on page 390).

A small point to note in using Steam Tables as well as program STEAM is that these sources use the International Table calorie, which is defined as 4.1868 joules. Physical chemistry sources generally use the thermochemical calorie, which is 4.184 joules. This makes it difficult to be completely consistent when using calories for aqueous solutions, and is a good reason for using the SI system of units.

13.6.2 The HKF equations

A quite different approach was taken by Helgeson starting in the 1970s to develop an equation of state for solutes at elevated temperatures and pressures. This was presented in a series of four papers by H.C. Helgeson and co-workers between 1974 and 1981 (Helgeson and Kirkham 1974a, 1974b, 1976; Helgeson, Kirkham and Flowers 1981), with fairly extensive modifications added by Tanger and Helgeson (1988). The HKF model is semiempirical, in the sense that it uses a number of empirical parameters within a framework suggested by fundamental physics and thermodynamics, and it is very widely used by geochemists in the form of program SUPCRT92.

Converting steam table values to standard state values

We pick two points at random, for comparison – 900 °C, 5000 bars, and 100 °C, 100 bars, and choose values from IAPS-84 for comparison with SUPCRT92. Also included are the more recent values from IAPWS-95, to show the differences from IAPS-84. Conversion factors used are 4.1868 J/cal and 18.015 268 for the molar mass of water.

	900 °C, 5000 bars				
	V cm^3 mol^{-1}	*H* cal mol^{-1}	*S* cal mol^{-1} K^{-1}	*U* cal mol^{-1}	*G* cal mol^{-1}
IAPWS-95	27.92	16 208	22.81	12 872	−10 552
IAPS-84	27.60	16 209	22.81	12 913	−10 555
SUPCRT92	27.60	−52 550	37.96		−80 475

	100 °C, 100 bars				
	V cm^3 mol^{-1}	*H* cal mol^{-1}	*S* cal mol^{-1} K^{-1}	*U* cal mol^{-1}	*G* cal mol^{-1}
IAPWS-95	18.71	1 837	5.596	1 792	−251.1
IAPS-84	18.71	1 835	5.59	1 791	−250.8
SUPCRT92	18.71	−66 932	20.73		−58 054

Conversion of Steam Table values to "apparent formation from the elements" values using Equations (13.44) and Table 13.1 is straightforward, for example,

$$\begin{aligned}\Delta_a H_{100^\circ,100\,\text{bars}} = \Delta H_{\text{SUPCRT92}} &= H_{\text{ST}} + \Delta_f H^\circ_{\text{triple}} \\ &= 1835 + (-68\,767) \\ &= -66\,932\,\text{cal mol}^{-1}\end{aligned}$$

(continued)

		900 °C, 5 kbar	100 °C, 100 bar	difference	
ΔG	IAPS-84	−10 555	−251	−10 304	
	SUPCRT92	−80 475	−58 054	−22 421	NOT the same
ΔH	IAPS-84	16 209	1835	14 374	
	SUPCRT92	−52 550	−66 932	14 382	same
ΔS	IAPS-84	22.81	5.59	17.22	
	SUPCRT92	37.96	20.73	17.23	same
V	IAPS-84	27.60	18.71	8.89	
	SUPCRT92	27.60	18.71	8.89	same

As you go from one T–P point to another, the change in G in the Steam Tables is not the same as the change in G using the "formation from the elements" convention because the value assigned to the entropy of liquid water at the triple point is different in the two conventions. To change a Steam Tables ΔG to a SUPCRT92 ΔG, subtract the two expressions of Equation (13.45). So changing between our two P, T points,

$$\begin{aligned}\Delta G\ (\text{SUPCRT92}) &= \Delta G\ (\text{Steam Tables}) - (900-100)(63.31/4.1868)\\ &= (-10\,555 + 250.8) - 12\,097\\ &= -22\,401\ \text{cal mol}^{-1}\quad (\text{cf. } -22\,421 \text{ in table above.})\end{aligned}$$

Because it applies mostly to electrolytes, it is discussed in Chapter 15. Briefly, Helgeson models the behavior of solutes by developing equations for the standard state partial molar volume (Helgeson and Kirkham 1976) and standard state partial molar heat capacity (Helgeson et al. 1981) as a function of P and T, with adjustable constants such that they can be applied to a wide variety of solutes. If you know these quantities ($\overline{V}^\circ$, $\overline{C}_P^\circ$), you can calculate the variation of the standard state Gibbs energy, and that leads through fundamental relationships to equilibrium constants, enthalpies, and entropies.

13.6.3 The Pitzer equations

Another approach to modeling aqueous solutes is the Pitzer equations. The contrast with the HKF approach is instructive. As just mentioned, the HKF method is to "build up" an expression for the standard state Gibbs energy of solutes from expressions for standard state partial molar volumes and heat capacities. This can only lead to expressions for other standard state quantities. For application to real systems, activity coefficients must be calculated separately, and although methods for this are discussed in Helgeson et al. (1981), they are not included in program SUPCRT92, and are not in general use.

The Pitzer equations on the other hand are based on an expression for the excess Gibbs energy, not the standard state Gibbs energy, and so fundamental expressions (obtained by differentiating this function) lead to other excess functions including activity coefficients. It also means that all quantities calculated from the Pitzer equations, all of which are derived from the excess Gibbs function, are internally perfectly consistent. This is of course not the same as being accurate, although being carefully tuned to the best experimental data available they tend to work very well.

Another way of summarizing the difference between the HKF and the Pitzer approaches to developing an EoS for solutes is that HKF equations are fit to experimental data after extrapolation to infinite dilution, whereas Pitzer equations are fit to the data themselves, extending to high concentrations. Therefore HKF equations are standard state, Pitzer equations are "real life."

The general topic of excess Gibbs functions for solutes is introduced in §10.4, but the details are discussed, along with the HKF equations, in Chapter 15.

13.7 Other equations of state

There is not much use in summarizing EoS available at the present time, because it is a rapidly developing field, and the summary will quickly become out of date. Nevertheless, here are some references of interest in geochemistry.

13.7.1 Duan et al.

Duan et al. (1992a, b; 1996a) developed an extension of the Lee–Kesler cubic equation of state for pure gases and mixtures in the systems CH_4–CO_2–H_2O and H_2O–H_2S–NaCl to quite high temperatures and pressures. The equation has 15 fit coefficients but is simple and easily programmed in a spreadsheet. The equation is

$$\begin{aligned} Z &= \frac{PV}{RT} \\ &= \frac{P_{\text{R}}V_{\text{R}}}{T_{\text{R}}} \\ &= 1 + \frac{B}{V_{\text{R}}} + \frac{C}{V_{\text{R}}^2} + \frac{D}{V_{\text{R}}^4} + \frac{E}{V_{\text{R}}^5} \\ &\quad + \frac{F}{V_{\text{R}}^2}\left(\beta + \frac{\gamma}{V_{\text{R}}^2}\right)\exp\left(-\frac{\gamma}{V_{\text{R}}^2}\right) \end{aligned} \tag{13.46}$$

where T_{R} and P_{R} are the reduced temperature and pressure, $V_{\text{R}} = RT_c/P_c$, and B, C, D, E, F, β, and γ are simple functions of the fit coefficients and T_{R}.

Duan et al. (1996b) later extended the application of this equation to include polar molecules such as H_2O and mixtures of any of the components H_2O, CO_2, CH_4, N_2, CO, H_2, O_2, H_2S, Ar, and fitted experimental data as well as data from molecular dynamics (MD) simulations up to 2000 K and 22 kbar. Instead of P_{R}, T_{R}, and V_{R}, Z is defined in terms of P_m, T_m and V_m, which are defined in terms of two parameters from the Lennard-Jones formulation of the intermolecular potential. Mixing rules then are simple combinations of these L-J potentials, rather than combinations of the fit coefficients. This equation is also easily programmed in a spreadsheet.

13.7.2 Akinfiev and Diamond

An alternative form of Equation (13.38) is

$$\ln \gamma_{f_i} = \frac{2}{V}\sum_{j=1}^{m} x_j B_{ij} - \ln Z_{\mathrm{mix}} \tag{13.47}$$

where $Z_{\mathrm{mix}} = 1 + B_{\mathrm{mix}}/V$ (Prausnitz et al. 1999, Equation 5.31). Akinfiev and Diamond (2003) use this as a starting point to develop an equation of state for nonelectrolytes, many of which are gaseous under normal conditions. This is of course derived from the virial equation, and has the usual limitation of being valid only at low to moderate densities. Akinfiev and Diamond propose an equation to describe the binary interaction parameter B_{ij} as a function of P and T. Because B_{ij} in the virial equation is a function only of T and is independent of P, the development in Akinfiev and Diamond departs from virial theory, using it only as a starting point. For two components, the B terms in (13.47) become ΔB, called ΔB_ρ, to which they ascribe the temperature dependence

$$2\Delta B_\rho = a + b\left(\frac{10^3}{T}\right)^{0.5}$$

A third fit parameter (in addition to a and b) is a "scaling factor" $(1-\xi)$ which is multiplied by Equation (13.47) written for the pure solvent (water) and then subtracted from the binary (13.47). The final form for the difference in chemical potential for the nonelectrolyte in the aqueous and gaseous standard states is

$$\begin{aligned}\mu^\circ_{2,aq} - \mu^\circ_{2,g} = &- RT\ln N_{\mathrm{w}} + (1-\xi)RT\ln f_i^\circ \\ &+ RT\xi\ln\left(\frac{RT}{M_{\mathrm{w}}}\rho_i^\circ\right) + R\left\{T\rho_i^\circ\left[a + b\left(\frac{10^3}{T}\right)^{0.5}\right]\right\}\end{aligned} \tag{13.48}$$

where N_{w} is the number of moles of water, and M_{w} is the molar mass or gram formulas weight of water. Expressions for partial molar volume, entropy, and heat capacity are then derived by differentiation.

13.7.3 Spycher and Reed

Spycher and Reed (1988) use a more or less "pure" virial equation approach to obtaining fugacities of components in gas mixtures. Their results are therefore limited to geologically moderate temperatures and pressures, and not too near the critical region. They analyzed the available data for H_2O, CO_2, CH_4, and H_2, as well as binary mixtures of these gases. They fit experimental data to an equation in which each virial coefficient is expressed as a polynomial in T,

$$PV/RT = 1 + (a/T^2 + b/T + c)P + (d/T^2 + e/T + f)P^2$$

from which expressions for fugacity coefficients and enthalpies were derived. The emphasis on enthalpies was because a major aim was the ability to calculate the composition of gases during boiling of hydrothermal solutions. These equations and the boiling calculations for fluids under various conditions are incorporated in program CHILLER (Reed, 1982, 1998). Comparing compositions calculated using various methods, they concluded that for modeling gas phase separation in geothermal waters, the assumption of ideal mixing of real gases (Lewis fugacity rule, §8.1.3) is good, but that this assumption is poor for calculations at pressures much larger than the saturation pressure of pure water, where nonideal mixing equations are distinctly better.

13.7.4 Other EoS

In addition to H_2O discussed above, a number of other highly accurate multiparameter EoS for pure gases have been developed (Ar, N_2, CO_2, CH_4, O_2, several organic and refrigerant gases) and are listed in Span (2000). For aqueous solutes, Pitzer (1995) lists those shown in Table 13.2.

Of these, undoubtedly the most interesting for geochemists is that for the system NaCl–H_2O (Archer, 1992; see also Pitzer and Peiper, 1984). This provides data on volumetric properties, solute and solvent activities, enthalpy changes, heat capacities, and solubilities. We looked at some of the data from this equation in Chapter 10, Table 10.1.

Table 13.2 *Some EoS for aqueous binary systems. References in Pitzer (1995, Chapter 18).*

Solute	Max T °C	Max P bars	Max mol kg^{-1}
HCl	375	400	15.6
NaCl	325	1000	saturation
NaBr	325	1500	varies
NaOH	250	400	6.3
KCl	325	500	varies
Na_2SO_4	300	200	saturation
$CaCl_2$	250	400	4.6
$MgSO_4$	200	P_{sat}	saturation

13.8 Summary

Good equations of state are valuable things, because they supply you with needed thermodynamic information, not only more easily than doing experiments or even just looking up numbers in tables, but some kind of averaged numbers from carefully selected data sets. There are two main approaches to developing equations of state (there may be more, but we are keeping things simple here).

One is the many variations of the cubic equation relating P, V, and T, which is very good at predicting vapor–liquid equilibria, but for which no formulation has yet been found to work with all pertinent systems. And although in principle cubic equations can work with any number of components in a mixture, the "mixing rules" are another source of more or less empirical guesswork.

The other main approach is to develop an equation which returns a value of G as a function of system T and P, or A as a function of system T and V. These equations have reached a very high state of development, having 50 or more fit coefficients and a mathematical form which is also a subject of computer fitting procedures. The EoS for water is an outstanding example of this approach. However, the extension of this approach to handle mixtures is a formidable research task. Span (2000) is a good introduction to this subject, and includes many specialized EoS not mentioned here.

When it comes to high temperature and pressures involving aqueous electrolytes, geochemists have little choice at the present time but to use the HKF model in the form of program SUPCRT92, although alternative approaches are being explored. The Pitzer equations are another possibility, but so far are mostly restricted to near ambient conditions. They do, however, perform much better than HKF at high concentrations.

14
Solid solutions

14.1 Introduction

Let's say you are mapping a field area consisting of metamorphic rocks. One unit contains garnet, plagioclase, kyanite, and quartz, with a few other phases. Having studied mineralogy, you know that garnet and plagioclase are not "pure" phases, made up of a single component but are solid solutions of several "end-member" components.[1] You also know that these minerals are related by the "anorthite breakdown reaction," written

$$3\,\mathrm{CaAl_2Si_2O_8^{plag}} = \mathrm{Ca_3Al_2Si_3O_{12}^{garnet}} + 2\,\mathrm{Al_2SiO_5^{kyanite}} + \mathrm{SiO_2^{qtz}} \tag{14.1}$$

for which the equilibrium constant is

$$K = \frac{a_{\mathrm{Ca_3Al_2Si_3O_{12}}}\, a^2_{\mathrm{Al_2SiO_5}}\, a_{\mathrm{SiO_2}}}{a^3_{\mathrm{CaAl_2Si_2O_8}}} \tag{14.2}$$

You analyze your minerals, and the results are reported in terms of oxides. A typical garnet analysis would look like Table 14.1. To make any thermodynamic use of this reaction, you evidently need to somehow extract an activity of $\mathrm{Ca_3Al_2Si_3O_{12}}$ from this information, as well as the activity of $\mathrm{CaAl_2Si_2O_8}$ from a similar analysis of the plagioclase. You know that quartz and kyanite are generally quite pure, so that assuming $a_{\mathrm{SiO_2}} = 1$ and $a_{\mathrm{Al_2SiO_5}} = 1$ will probably be close enough.

How do you do this? And why would it be useful? A complete answer would take a whole book, and is more appropriate to a course in metamorphic petrology, but we can fairly easily see how thermodynamics is applied in such cases. Newton and Haselton (1981) give a summary of the thermodynamics of this reaction, as well as its use as a geobarometer, that is, a mineral assemblage that, when analyzed thermodynamically, will give an estimate of the pressure at the time the minerals equilibrated.

---

[1] Recall that we introduced the concept of components in Chapter 2. Many discussions use the term rather loosely, referring to compositional terms generally. In other cases, such as here, we use the term strictly, as the smallest number of formulas needed to express a system composition (Chapter 11).

Table 14.1 *A typical garnet composition. FeO* = total iron reported as FeO. Sevigny and Ghent (1989), sample S129.*

SiO_2	37.11
Al_2O_3	21.08
FeO*	34.30
MnO	2.46
MgO	3.42
CaO	2.10
Total	100.47

14.2 Solid solutions

Although the term solid solution can include amorphous phases such as glass, in geochemical thermodynamics it generally refers to crystalline solids having a relatively large scale regular lattice structure. We also assume that defects, dislocations, strain and other factors are of negligible importance, although there is a large literature on these topics. The major difference between solid solutions and fluid solutions is of course this regularly- spaced three-dimensional lattice structure. The distribution of atoms on the sites of a given lattice can be quite variable, depending on the properties of the atoms, the type of lattice, and the conditions of formation. This variability primarily affects the entropy of the lattice, and hence the Gibbs energy.

14.2.1 Short and long range ordering

In fluids (gases and liquids) certain particle interactions are often strong enough to result in formation of new particles. For example, in a solution of N_2 and H_2 gases, some NH_3 particles will form through the reaction

$$N_2(g) + 3\,H_2(g) = 2\,NH_3(g)$$

In electrolyte solutions, ionic attraction can result in ion pairs, so that in NaCl solutions there is a strong tendency for Na^+ particles to be closer to Cl^- particles than to other Na^+ particles. These are examples of short-range ordering. Particles are "ordered," their positions determined, over very short distances. In these fluids, there is no possibility of long-range ordering; there is no reason for the nitrogen, hydrogen and ammonia particles or Na^+Cl^- ion pairs to be regularly distributed with respect to each other – they mix randomly at high speeds.[2]

Crystalline solids, however, can have long-range as well as short-range ordering. For example, in disordered Cu_3Au, the gold atoms are arrayed randomly at sites normally occupied by copper atoms in the copper lattice, but given proper heat treatment, all of the gold atoms become located at cube corners and copper atoms at face-centered positions of the unit-cell cube. The Au atoms then show long range order. Silicates are basically an ordered arrangement of oxygen atoms with interstices of various sizes and shapes, depending on the arrangement of the oxygens. Cations will then show preferences for some interstitial sites rather than others. For example, a particular cation may be too big for some sites, and the sites it fits into may be regularly spaced, resulting in long range order for that cation.

If there is a tendency for certain pairs of cations to prefer to be on nearby sites, perhaps to preserve a local charge balance, the result is a short range ordering. This tendency can be weak or strong, analogous to weak or strong complex formation. If the tendency is very strong, the cation pair can be considered to be a particle or molecule, and the distribution of particles over available sites can be calculated using this molecule rather than the individual atoms. Carried to its logical extreme results in "molecular mixing," in which complete mineral formulas are considered to be the mixing units. This is not usually the best approach. Unfortunately, it is difficult to determine the extent of short range ordering in crystals, and taking it into account greatly complicates the thermodynamic treatment. One such approach is the "quasi-chemical" method of Guggenheim (1952). If you need to know more about this, an introduction and applications are given by Saxena (1973). The most common approach, however, is to ignore short range order, and to calculate the mole fractions of solid solution components by considering the distribution of ions on crystallographic sites, a process called mixing-on-sites, or "site-mixing."

The rest of this chapter is in two parts. First, we consider how to calculate mole fractions in solid solutions assumed to have only long range order, and how to combine these mole fractions into Raoultian activities. Then we consider the determination of activity coefficients in (binary) solid solutions, and how regular solution and Margules equations are used to systematize these.

14.2.2 Mixing on sites

Thermodynamic mole fractions

We know that for Raoultian ideal solutions, $a_i = x_i$, or activity is equal to mole fraction, so to calculate the Raoultian activity of solid solution components, all we have to do is calculate their mole fractions. If there is only one

---

[2] Theories having partial long range ordering in electrolyte solutions have been proposed, e.g., Pytkowicz et al. (1977).

crystallographic site on which elements mix randomly, there is no difficulty. An example is the alkali feldspar solid solution, in which Na and K mix randomly on one crystallographic site. The mole fractions of Na and K on this site, are then quite evidently the same as the mole fractions of $NaAlSi_3O_8$ and $KAlSi_3O_8$, because there are no other changes. That is,

$$\begin{aligned} x_{\mathrm{Ab}} &= \frac{m_{\mathrm{NaAlSi_3O_8}}}{m_{\mathrm{NaAlSi_3O_8}} + m_{\mathrm{KAlSi_3O_8}}} \\ &= \frac{m_{\mathrm{Na}}}{m_{\mathrm{Na}} + m_{\mathrm{K}}} \end{aligned}$$

So if you wanted the activity of the albite component in this solid solution, and you didn't mind assuming it was an ideal solution, all you need to do is analyze the feldspar for Na and K.

If there is more than one site on which mixing takes place, the thermodynamics remains simple, but determining the mole fractions becomes more difficult. Equation (7.20),

$$\Delta_{\mathrm{mix}} G_{\mathrm{ideal\ sol'n}} = RT \sum_i x_i \ln x_i \qquad [7.20]$$

gives us the Gibbs energy of mixing on one site. If there are several sites on which mixing is taking place, it gives the energy change on each one separately. To combine these terms to obtain the total molar energy term for the system, we need to know how many of each site there are, which is given by the chemical formulas of the crystal. For example, garnet has the formulas $X_3Y_2Si_3O_{12}$, where the X sites are cubic (8-coordinated), the Y sites are octahedral (6-coordinated), and the Si site is tetrahedral (4-coordinated). If the X sites contain randomly mixed Mg^{2+}, Ca^{2+}, Mn^{2+}, and Fe^{2+}, and the Y sites contain randomly mixed Al^{3+} and Fe^{3+}, there are several possible mineral "end-member" species, such as

pyrope	$Mg_3Al_2Si_3O_{12}$
grossular	$Ca_3Al_2Si_3O_{12}$
almandine	$Fe_3^{2+}Al_2Si_3O_{12}$
spessartine	$Mn_3Al_2Si_3O_{12}$
khoharite	$Mg_3Fe_2^{3+}Si_3O_{12}$
andradite	$Ca_3Fe_2^{3+}Si_3O_{12}$

To simplify, let's consider that Mn^{2+} and Fe^{2+} in the X site have low concentrations and are insignificant, so that the garnet is essentially $(Mg, Ca)_3(Al, Fe)_2Si_3O_{12}$. Including the other elements adds no difficulty, it just adds more terms of the same type, and makes the equations longer.

Now we have four elements mixing on two sites. This looks like a case of four components, pyrope, grossular, andradite and khoharite. However, any of these four formulas can be written as a linear combination of the other three (Spear, 1995, p. 184). We can therefore describe all compositions of this garnet in terms of any three of these four components, and we choose to use pyrope, grossular, and andradite.

The total Gibbs energy of mixing can then be written in two different but equivalent ways. We can use Equation (7.20) with end-member components, or with ionic mixing on individual sites. Using our three components, we have

$$\begin{aligned}\Delta_{\text{mix}} G_{\text{ideal sol'n}} &= RT(x_{\text{Mg}_3\text{Al}_2\text{Si}_3\text{O}_{12}} \ln x_{\text{Mg}_3\text{Al}_2\text{Si}_3\text{O}_{12}} + x_{\text{Ca}_3\text{Al}_2\text{Si}_3\text{O}_{12}} \ln x_{\text{Ca}_3\text{Al}_2\text{Si}_3\text{O}_{12}} \\ &\quad + x_{\text{Ca}_3\text{Fe}_2\text{Si}_3\text{O}_{12}} \ln x_{\text{Ca}_3\text{Fe}_2\text{Si}_3\text{O}_{12}}) \\ &= RT(x_{\text{pyrope}} \ln x_{\text{pyrope}} + x_{\text{grossular}} \ln x_{\text{grossular}} + x_{\text{andradite}} \ln x_{\text{andradite}}) \qquad (14.3)\end{aligned}$$

This applies to ideal mixing. If each log term is corrected with an appropriate activity coefficient, we get

$$\Delta_{\text{mix}} G = RT(x_{\text{pyrope}} \ln a_{\text{pyrope}} + x_{\text{grossular}} \ln a_{\text{grossular}} + x_{\text{andradite}} \ln a_{\text{andradite}}) \qquad (14.4)$$

Considering mixing on individual sites, we can write

$$\begin{aligned}\Delta_{\text{mix}} G_{\text{ideal sol'n}} &= -T\,\Delta_{\text{mix}} S_{\text{ideal sol'n}} \\ &= 3RT(x_{\text{Mg}} \ln x_{\text{Mg}} + x_{\text{Ca}} \ln x_{\text{Ca}}) + 2RT(x_{\text{Al}} \ln x_{\text{Al}} + x_{\text{Fe}^{3+}} \ln x_{\text{Fe}^{3+}}) \\ &= RT(x_{\text{Mg}} \ln x_{\text{Mg}}^3 + x_{\text{Ca}} \ln x_{\text{Ca}}^3 + x_{\text{Al}} \ln x_{\text{Al}}^2 + x_{\text{Fe}^{3+}} \ln x_{\text{Fe}^{3+}}^2) \qquad (14.5)\end{aligned}$$

where, for example,

$$x_{\text{Mg}} = \frac{n_{\text{Mg}}}{n_{\text{Mg}} + n_{\text{Ca}}}$$

in the normal fashion, and for which the general notation is

$$\Delta_{\text{mix}} G_{\text{ideal sol'n}} = RT \sum_{j=1}^{k} n_j \sum_i x_{i,j} \ln x_{i,j} \qquad (14.6)$$

where k is the total number of nonequivalent sites (2 in this example), n_j is the number of times the constituent in question appears in the formulas for the jth site in the solution (in this example, $n_j = 2$ for the X and Y sites), and $x_{i,j}$ is the mole fraction of the ith ion on the jth site (X, or Y site). We could have $k = 3$ by including mixing of Al and Si on the Si site. One thing to watch for in applying this equation is to make sure that a charge balance is maintained. In other words, the substitutions should be independent of one another. See Ulbrich and Waldbaum (1976) for a detailed discussion of this topic.

This result is of limited usefulness. What we really want in most cases is not the total mixing energy, but the activity of some component in a solid solution containing that component. In our example, we want to know the activity of the grossular ($Ca_3Al_2Si_3O_{12}$) component, so as to be able to use our mineral composition in Equation (14.2). To get an expression for the activity of such components, we first transform (14.5) into a form containing the number of moles of our various components. Following Spear (1995, p. 185), we write a mass balance for each cation,

$$\begin{aligned}
n_{\mathrm{Mg}} &= 3\,x_{\mathrm{pyrope}} + 0\,x_{\mathrm{grossular}} + 0\,x_{\mathrm{andradite}} \\
n_{\mathrm{Ca}} &= 0\,x_{\mathrm{pyrope}} + 3\,x_{\mathrm{grossular}} + 3\,x_{\mathrm{andradite}} \\
n_{\mathrm{Al}} &= 2\,x_{\mathrm{pyrope}} + 2\,x_{\mathrm{grossular}} + 0\,x_{\mathrm{andradite}} \\
n_{\mathrm{Fe}} &= 0\,x_{\mathrm{pyrope}} + 0\,x_{\mathrm{grossular}} + 2\,x_{\mathrm{andradite}}
\end{aligned}$$

or (using ⇒ to mean "defined as thermodynamic mole fraction"),

$$\begin{aligned}
n_{\mathrm{Mg}}/3 &= x_{\mathrm{pyrope}} && \Rightarrow \boldsymbol{x}_{\mathrm{Mg}} \\
n_{\mathrm{Ca}}/3 &= x_{\mathrm{grossular}} + x_{\mathrm{andradite}} && \Rightarrow \boldsymbol{x}_{\mathrm{Ca}} \\
n_{\mathrm{Al}}/2 &= x_{\mathrm{pyrope}} + x_{\mathrm{grossular}} && \Rightarrow \boldsymbol{x}_{\mathrm{Al}} \\
n_{\mathrm{Fe}}/2 &= x_{\mathrm{andradite}} && \Rightarrow \boldsymbol{x}_{\mathrm{Fe}}
\end{aligned}$$

giving the relationship between the number of moles of ions on sites and the (thermodynamic) mole fractions ($\boldsymbol{x}$) of our chosen mineral components. Substituting these relations into Equation (14.5) we get

$$\begin{aligned}
\Delta_{\mathrm{mix}} G_{\text{ideal sol'n}} = RT[&x_{\mathrm{pyrope}} \ln \boldsymbol{x}_{\mathrm{Mg}}^3 + (x_{\mathrm{grossular}} + x_{\mathrm{andradite}}) \ln \boldsymbol{x}_{\mathrm{Ca}}^3 \\
&+ (x_{\mathrm{pyrope}} + x_{\mathrm{grossular}}) \ln \boldsymbol{x}_{\mathrm{Al}}^2 + x_{\mathrm{andradite}} \ln \boldsymbol{x}_{\mathrm{Fe}^{3+}}^2]
\end{aligned}$$

and rearranging terms,

$$\begin{aligned}
\Delta_{\mathrm{mix}} G_{\text{ideal sol'n}} = RT[&x_{\mathrm{pyrope}} \ln(\boldsymbol{x}_{\mathrm{Mg}}^3 \boldsymbol{x}_{\mathrm{Al}}^2) + x_{\mathrm{grossular}} \ln(\boldsymbol{x}_{\mathrm{Ca}}^3 \boldsymbol{x}_{\mathrm{Al}}^2) \\
&+ x_{\mathrm{andradite}} \ln(\boldsymbol{x}_{\mathrm{Ca}}^3 \boldsymbol{x}_{\mathrm{Fe}}^2)]
\end{aligned} \tag{14.7}$$

Comparing this with Equation (14.4), we see that our component activity terms are

$$\left.\begin{aligned}
a_{\mathrm{pyrope}} &= \boldsymbol{x}_{\mathrm{Mg}}^3 \boldsymbol{x}_{\mathrm{Al}}^2 \\
a_{\mathrm{grossular}} &= \boldsymbol{x}_{\mathrm{Ca}}^3 \boldsymbol{x}_{\mathrm{Al}}^2 \\
a_{\mathrm{andradite}} &= \boldsymbol{x}_{\mathrm{Ca}}^3 \boldsymbol{x}_{\mathrm{Fe}^{3+}}^2
\end{aligned}\right\} \tag{14.8}$$

where the thermodynamic mole fractions are calculated as shown. An example will help to clarify how this is done (page 402).

All this assumes that we have defined a mole of grossular as $Ca_3Al_2Si_3O_{12}$, and similarly for the other components. If we defined the mole of grossular as $CaAl_{2/3}SiO_4$, that is, based on four oxygens instead of 12, we would have

$$a_{\mathrm{CaAl_{2/3}SiO_4}} = \boldsymbol{x}_{\mathrm{Ca}} \boldsymbol{x}_{\mathrm{Al}}^{2/3}$$

as explained in §9.10.2. If there is no substitution on the Al (octahedral) site, $\boldsymbol{x}_{\rm Al} = 1$, $\boldsymbol{x}_{\rm Ca} = x_{\rm Ca}$, and

$$a_{\rm CaAl_{2/3}SiO_4} = x_{\rm Ca}$$

analogous to the alkali feldspars.

14.2.3 Complications

If that was all there was to determining activities in solid solutions, a large part of metamorphic petrology would be a lot simpler than it is. But there are a number of complicating factors.

- As already mentioned, we need activity coefficients as well as thermodynamic mole fractions. We discuss this in the next section.
- Ionic mixing may be complicated by the need for charge balance. For example in plagioclase, Ca^{2+} cannot substitute for Na^+ unless Al^{3+} substitutes for Si^{4+} as well.
- Mixing can also be complicated by "Al avoidance." Substitution of Al^{3+} for Si^{4+} on tetrahedral sites may take place in such a way as to avoid having Al on adjacent sites, minimizing distortion of the lattice structure.
- It may happen that site-mixing results in a formulas that does not result in $x = 1$ for the pure components. In these cases a "normalization factor" is used, to ensure that end-member mole fractions are 1.0.

Activities from solid solution compositions

We start with the garnet analysis in Table 14.1. The idea is to calculate how many cations of each type there are in the analysis, assuming there are a total of 12 oxygens. Following the method of Powell (1978, p. 75), we recalculate as follows:

oxide	wt.%	A	B	C	D	E
SiO_2	37.11	0.6176	1.2353	0.6176	2.98	5.9534
Al_2O_3	21.08	0.2067	0.6202	0.4135	1.99	2.9893
FeO*	34.30	0.4774	0.4774	0.4774	2.30	2.3009
MnO	2.46	0.0347	0.0347	0.0347	0.17	0.1671
MgO	3.42	0.0848	0.0848	0.0848	0.41	0.4089
CaO	2.10	0.0374	0.0374	0.0374	0.18	0.1805
Total			2.4899			12.000

$F = 12/2.4899$

(continued)

where the columns are calculated as follows:
column A is the weight % divided by the molecular weight
column B is A times the number of oxygens in the formulas
column C is A times the number of cations in the formulas
column D is C times *F*, and is the number of cations per 12 oxygens
column E is D times the ratio of oxygens to cations in the formulas, which is a check on the arithmetic.

The garnet composition can then be expressed as

$$(\mathrm{Mg}_{0.41}\mathrm{Ca}_{0.18}\mathrm{Fe}_{2.30}\mathrm{Mn}_{0.17})\mathrm{Al}_2\mathrm{Si}_3\mathrm{O}_{12},$$

where the difference between 1.99 and 2 for Al and between 2.98 and 3 for Si can be considered as analytical uncertainty, or perhaps that some small amount of substitution has taken place.

The mole fractions on the cubic site are then

$$x_{\mathrm{Fe}} = 2.30/(2.30+0.17+0.41+0.18) = 0.753$$

$$x_{\mathrm{Mn}} = 0.17/(2.30+0.17+0.41+0.18) = 0.055$$

$$x_{\mathrm{Mg}} = 0.41/(2.30+0.17+0.41+0.18) = 0.134$$

$$x_{\mathrm{Ca}} = 0.18/(2.30+0.17+0.41+0.18) = 0.059$$

and the ideal solution activities are (where $x_{\mathrm{Al}} = 1$)

$$a_{\mathrm{almandine}} = x_{\mathrm{Fe}}^3 x_{\mathrm{Al}}^2 = 0.426$$

$$a_{\mathrm{spessartine}} = x_{\mathrm{Mn}}^3 x_{\mathrm{Al}}^2 = 0.00016$$

$$a_{\mathrm{pyrope}} = x_{\mathrm{Mg}}^3 x_{\mathrm{Al}}^2 = 0.0024$$

$$a_{\mathrm{grossular}} = x_{\mathrm{Ca}}^3 x_{\mathrm{Al}}^2 = 0.00021$$

See Powell (1978) for many other examples.

- It often happens that the two end-members of a solid solution series have different crystallographic structures. These will have different energies and entropies, which must be accounted for in calculating phase relations.
- Where independent mixing occurs on more than one site, as in the garnet case, the number of independent components is less than the number of possible end-member components. These are “reciprocal solid solutions,” and this gives rise to the possibility of an additional energy term called “reciprocal Gibbs energy” which must be included in the energy of mixing, and which therefore affects the activity calculations (Wood and Nicholls, 1978).

- We have considered only the case where individual cations mix randomly only on one site. However, rather than being random, they may show some variable degree of "order–disorder," such as the variable Al–Si mixing on the tetrahedral site in feldspars, resulting in the "high" (e.g., sanidine) and "low" (e.g., maximum microcline) varieties. Other minerals exhibit "cross-site mixing," in which an individual cation can mix to a variable degree on two (or more) sites, also resulting in order–disorder. It also results in such solutions being reciprocal solid solutions.
- Implicitly assumed in the thermodynamic analysis of all experimental and natural assemblages is the presence of local equilibrium (§2.3.2). In petrology, an additional problem is knowing at what stage in the long, complex history of the rock the assumed equilibrium was attained.

Most of these complications are considered in some detail by Spear (1995, Chapter 7).

14.3 Activity coefficients in solid solutions

In Chapter 5 we discussed how to get thermodynamic data, and we said the methods should be those based on measuring heat and work, such as calorimetry, and we assumed implicitly that we were dealing with pure substances. Now we are dealing with substances in solution, and the problem is to determine by how much the properties of the pure substance (strictly speaking the substance in some known state) are changed by being in solution. When we know that, we can determine which way reactions involving dissolved substances will go. We have derived a function, the activity, which tells us by how much the Gibbs energy changes in solution. Activities have a composition part and an activity coefficient part, and the hard part in determining the activity is getting an activity coefficient, because we usually have the compositional part (although that can get a bit tricky in solid solutions, as we have just seen). So measuring activity coefficients is fundamental to using solutions in chemical reactions. Failing any knowledge of activity coefficients we can fall back on two alternatives:

- Assume that property changes are those that would occur if our solutions were ideal, that is, we can assume activity coefficients are 1.0, or
- Use some theoretical calculation or estimation of the activity coefficient.

Both these alternatives have drawbacks, but are very commonly used in the Earth sciences, where complex solutions at high temperatures and pressures are commonly of interest. The problem is that determining activity coefficients is difficult, and for the most part confined to binary and ternary systems. In solid solutions, it can be done in various ways, such as

- Comparing the composition of a phase in a buffered assemblage with the ideal composition.
- Calorimetric measurements of enthalpies and entropies of solid solutions, combining these into $\Delta_{\mathrm{mix}} G$ values, and comparing these with the ideal values.
- Equilibrating a solid solution with another phase in which the activity coefficients are known, such as an (almost) ideal gas, a dilute solution, or a phase having previously measured activity coefficients.
- Analyzing solvus data.

We will have a brief look at each of these methods. In the process, we will see the pervasive use of regular solution and Margules equations.

14.3.1 Buffered assemblages

Hensen et al. (1975) experimentally equilibrated the assemblage garnet–anorthite–kyanite–quartz at temperatures of 1000–1300 °C, 15–21 kilobars, together with a source of Mg (enstatite, $MgSiO_3$). Mg does not enter the anorthite, kyanite, or quartz structures, only the garnet, and that to a variable extent, depending on the temperature and pressure. The assemblage anorthite–kyanite–quartz has three components (CaO, Al_2O_3, SiO_2) and three phases, and so is a buffering assemblage at a given T and P (see §11.5). This means that the activity of $Ca_3Al_2Si_3O_{12}$ is fixed at that T and P. It is also equal to the equilibrium constant for the anorthite breakdown reaction, Equation (14.2), because all the other activity terms are fixed at 1.0. That is,

$$K = \frac{a_{\mathrm{Ca_3Al_2Si_3O_{12}}} a^2_{\mathrm{Al_2SiO_5}} a_{\mathrm{SiO_2}}}{a^3_{\mathrm{CaAl_2Si_2O_8}}}$$

$$= a_{\mathrm{Ca_3Al_2Si_3O_{12}}} \quad \text{when } a_{\mathrm{Al_2SiO_5}}, a_{\mathrm{SiO_2}}, \text{ and } a_{\mathrm{CaAl_2Si_2O_8}} = 1.0 \qquad (14.9)$$

which is actually just another way of showing that $a_{\mathrm{Ca_3Al_2Si_3O_{12}}}$ is a constant at T, P. Therefore a knowledge of how the equilibrium constant changes with pressure will allow a calculation of the grossular activity.

Adding Mg to the garnet has the effect of enlarging the garnet stability field, for the same reason that adding salt to water lowers the freezing point. This effect is discussed in Chapter 17 and illustrated in Figure 17.13. In the present case the reaction is not between a solid and a liquid, but between two solid assemblages, but the principle is exactly the same – adding a component lowers the Gibbs energy and the phase field expands as a result. Figure 14.1 shows the equilibrium curve for this reaction calculated by SUPCRT92 based on several sets of experimental data detailed in Helgeson et al. (1978). Also shown are the experimental points of Hensen et al. (1975) at lower pressures, in which the garnet is not pure grossular, but a grossular–pyrope solid solution.

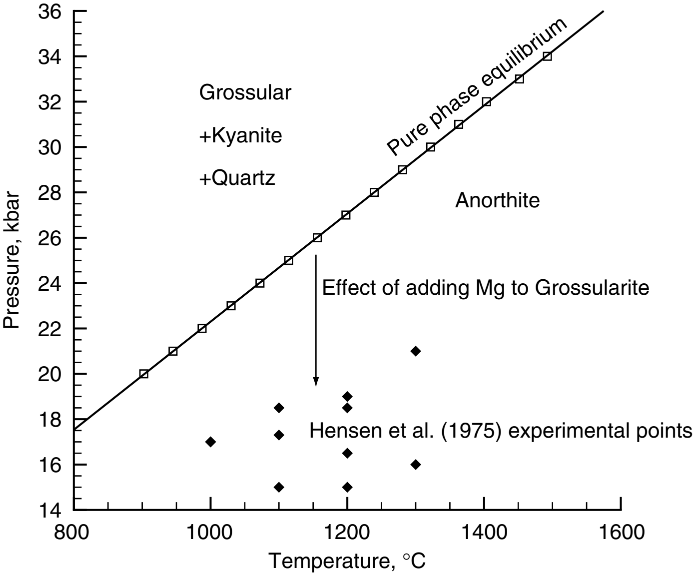


Figure 14.1 The univariant equilibrium curve for the anorthite breakdown reaction, 3 anorthite = grosssular + 2 kyanite + quartz. Calculated with the univariant curve option in SUPCRT92. The effect of adding Mg to the grossular is to lower the equilibrium pressure, as shown by the data of Hensen et al. (1975).

After equilibrating the phases at T and P, the garnet is analyzed, and the ideal Raoultian activity of $Ca_3Al_2Si_3O_{12}$ is calculated from the thermodynamic mole reaction as in Equations (14.8), i.e.,

$$\begin{aligned} a^{\text{ideal}}_{\text{grossular}} &= \boldsymbol{x}_{\text{grossular}} \\ &= \boldsymbol{x}^3_{\text{Ca}}\boldsymbol{x}^2_{\text{Al}} \end{aligned}$$

except that in this case there is only Ca–Mg substitution, so $x_{\text{Al}} = 1$, and $\boldsymbol{x}_{\text{grossular}} = x^3_{\text{Ca}}$. The activity coefficient is then calculated from

$$\gamma_{\text{Ca}_3\text{Al}_2\text{Si}_3\text{O}_{12}} = \frac{a_{\text{grossular}}}{\boldsymbol{x}_{\text{grossular}}}$$

The details of this calculation are shown on page 407.

This study has been included partly because it illustrates the usefulness of the buffering concept, which is the basis of the method. The actual value of the interaction coefficients in garnets are still a subject of some debate, as discussed by Newton and Haselton (1981).

14.3.2 Calorimetric measurements

The enthalpies of solution of several synthetic pyrope–grossular garnets were determined by Newton et al. (1977) by dissolving them in a liquid (2PbO ·

Calculation of grossular activity and activity coefficient

In §9.8 we discussed the effect of pressure on an equilibrium constant, in the special case in which the $\Delta_r V^\circ$ can be considered constant. This equation is

$$\ln K_P = \ln K_{P=1} - \frac{\Delta_r V^\circ}{RT}(P-1) \qquad [9.36]$$

but in that case we assumed that the standard state pressure was 1 bar. The equation can of course be used to calculate the change in $\ln K$ between any two pressures, as long as $\Delta_r V^\circ$ remains constant. In this case, we need the change in $\ln K$ between the pressure of the pure grossular equilibrium curve (call it P°) and the pressure of the grossular–pyrope solid solution curve (call it P). We know that in this case, $P^\circ > P$. We also define the standard state of all minerals involved as the pure phase at P° and T, so that everywhere on the pure grossular equilibrium curve all mineral activities are 1.0, $\Delta_r G^\circ = 0$, $K_{P^\circ,T} = 1$, and $\ln K_{P^\circ,T} = 0$. Equation (9.36) then becomes

$$\begin{aligned}\ln K_{P,T} &= \ln K_{P^\circ,T} - \frac{\Delta_r V^\circ}{RT}(P-P^\circ)\\ &= -\frac{\Delta_r V^\circ}{RT}(P-P^\circ)\end{aligned}$$

For the Hensen et al. experimental point at 18.5 kb, 1200 °C (Table 14.2), $P^\circ = 26\,750$ bars, and $\Delta_r V^\circ = -66\,\mathrm{cm^3\,mol^{-1}}$ or $-22\,\mathrm{cm^3\,mol^{-1}} = -0.526\,\mathrm{cal\,bar^{-1}}$ for the four oxygen form of the formulas, so

Table 14.2 *Results of the experiments of Hensen* et al. *(1975). Numbers in bold are used in the example calculation on page 407. * = corrected from 1581 in the original.*

T °C	P kbar	x_{Ca}	γ	w_G cal mol^{-1}
1300	16	0.11	1.154	565
1300	21	0.22	1.153	731
1200	15	0.10	1.353	1092
1200	16.5	0.12	1.408	1293
1200	**18.5**	**0.18**	**1.260**	**1006**
1200	19	0.19	1.287	1126
1100	15	0.11	1.554	1518*
1100	17.3	0.16	1.527	1637
1100	18.5	0.21	1.418	1527
1000	17	0.20	1.656	1994
1000	17	0.22	1.574	1886

(continued)

$$\begin{aligned}\ln a_{\mathrm{CaAl_{2/3}SiO_4}} &= \frac{0.526(18\,500-26\,750)}{1.987\times 1573.15}\\ &= -1.484\end{aligned}$$

so $a_{\mathrm{CaAl_{2/3}SiO_4}} = 0.227$, and $\gamma_{\mathrm{CaAl_{2/3}SiO_4}} = 0.227/0.18 = 1.26$

Assuming the garnet to be a regular solution, an interaction parameter w_G can be calculated from Equation (10.98), which in this case is

$$\begin{aligned}w_G &= \frac{RT\ln\gamma_{\mathrm{Ca_3Al_2Si_3O_{12}}}}{(1-\boldsymbol{x}_{\mathrm{grossular}})^2}\\ &= \frac{RT\ln\gamma_{\mathrm{Ca_3Al_2Si_3O_{12}}}}{(1-x_{\mathrm{Ca}}^3)^2}\end{aligned}$$

for mixing of Ca_3 and Mg_3 ($Ca_3Al_2Si_3O_{12}$ and $Mg_3Al_2Si_3O_{12}$), or

$$w_G = \frac{RT\ln\gamma_{\mathrm{CaAl_{2/3}SiO_4}}}{(1-x_{\mathrm{Ca}})^2}$$

for mixing of Ca and Mg (i.e., $CaAl_{2/3}SiO_4$ and $MgAl_{2/3}SiO_4$). For our point with $\gamma_{\mathrm{CaAl_{2/3}SiO_4}} = 1.26$,

$$\begin{aligned}w_G &= \frac{1.987\times 1473.15\times\ln 1.26}{(1-0.18)^2}\\ &= 1006\,\mathrm{cal\,mol^{-1}}\end{aligned}$$

B_2O_3) at 970 K (high temperature oxide calorimetry). The results are shown in Table 14.3 and Figure 14.2. The solid line is calculated using a two-parameter Margules equation (Equation 10.103, page 310),

$$G^{\mathrm{EX}}_{\mathrm{sol'n}} = x_1(w_{G_2}x_1x_2) + x_2(w_{G_1}x_1x_2) \qquad [10.103]$$

where grossular is component 1 and pyrope component 2, with the fit parameters $w_{G_1} = 3.82\,\mathrm{kcal\,mol^{-1}}$ and $w_{G_2} = 2.0\,\mathrm{kcal\,mol^{-1}}$. Note that the enthalpies of solution in Figure 14.3a lie below the ideal mixing line; this means that it takes less energy to melt the solid solution than to melt a mixture of the end-members. Therefore the enthalpy of mixing ($= H^{\mathrm{EX}}$, Figure 14.3b) is positive.

Newton et al. (1977) point out that the results of Henson et al. (1975) imply a nonideal entropy of mixing of Mg and Ca in garnets. The equation of the line in Figure 14.3 is $w_G = 7460 - 4.3\,T$. Comparing this with $w_G = w_H - Tw_S$ (Equation 10.91, page 305), we see that their results imply that

Table 14.3 *Enthalpies of solution of pyrope–grossular garnets at 970 K. From Newton* et al., *1977, Table 3. Uncertainty is standard deviation from the mean. Number of calorimeter runs in parentheses.* $H^{\text{EX}} = \Delta_{sol'n}H - \Delta_{ideal\ sol'n}H = \Delta_{\text{mix}}H$ *(Equation 10.20).*

Sample	$\Delta_{\text{sol'n}}H$ kcal mol^{-1}	$\Delta_{\text{sol'n}}H$ kcal mol^{-1}
Pyrope 1	27.73±0.28(5)	
Pyrope 2	27.19±0.24(3)	
Pyrope 3	27.44±0.11(2)	
Pyrope mean value	27.45±0.38(10)	0.0
$Py_{91}Gr_9$	27.11±0.34(5)	1.67
$Py_{82}Gr_{18}$	28.33±0.24(5)	1.78
$Py_{72.5}Gr_{27.5}$	29.29±0.28(4)	2.22
$Py_{53}Gr_{47}$	32.55±0.25(4)	1.84
$Py_{20}Gr_{80}$	38.37±0.19(6)	0.89
$Py_{10}Gr_{90}$	39.75±0.15(4)	0.98
Grossular	42.21±0.41(7)	0.0

$w_H = 7460\,\text{cal mol}^{-1}$ and $w_S = 4.3\,\text{cal mol}^{-1}\,\text{K}^{-1}$. Excess enthalpy and entropy are then

$$\begin{aligned} H^{\text{EX}} &= w_H\, x_{\text{grossular}} x_{\text{pyrope}} \\ &= 7460 x_{\text{grossular}} x_{\text{pyrope}}\ \text{cal mol}^{-1}, \quad \text{and} \\ S^{\text{EX}} &= w_S\, x_{\text{grossular}} x_{\text{pyrope}} \qquad [10.89] \\ &= 4.3 x_{\text{grossular}} x_{\text{pyrope}} \quad \text{cal mol}^{-1}\,\text{K}^{-1} \end{aligned}$$

These values of H^{EX} are not quite the same as the calorimetric results of Newton et al. (1977), but of course are for different temperatures and pressures.

Excess entropy of mixing has also been estimated from cryogenic or low-temperature calorimetry (Chapter 5) by Haselton and Westrum(1980). They determined $S_{298} - S_0$ at 298.15 K for pyrope, grossular, and a single intermdiate composition, $Py_{0.6}$ $Gr_{0.4}$. However, S_{298} for the intermediate (Ca–Mg) garnet requires calculation of the residual entropy "frozen in," i.e., the entropy of the crystal at 0 K.

Residual entropy Assuming ideal mixing of Ca and Mg on the cubic site, the residual entropy for the 12-oxygen formula (mixing of $Ca_3Al_2Si_3O_{12}$ is

$$\begin{aligned} \Delta_{\text{mix}} S_{\text{ideal sol'n}} &= -3R \sum_i xi \ln xi \qquad (14.10) \\ &= 3[-8.31451(0.6\ln(0.6) + 0.4\ln(0.4))] \\ &= 16.788\,\text{Jmol}^{-1}\text{K}^{-1} \end{aligned}$$

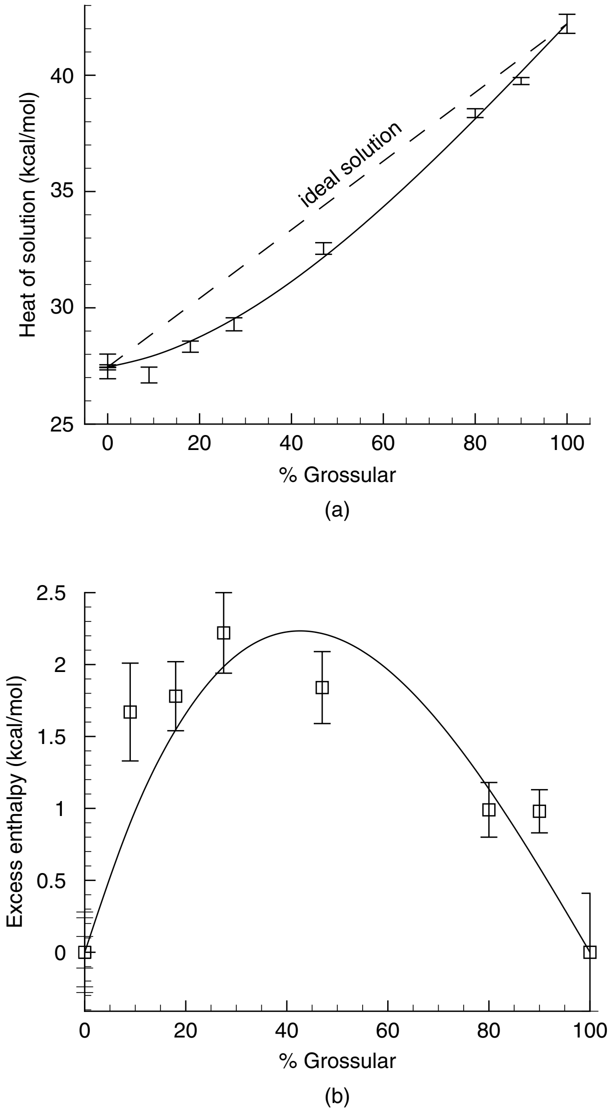


Figure 14.2 (a) Enthalpy of solution of synthetic pyrope–grossular garnets at 970 K. Data from Table 14.3. The solid line is calculated using a two-parameter Margules equation. (b) Measured and calculated H^{EX} from the same data set.

The reason for the 3 on the right side, which makes this equation different from Equation (7.17), is that we are mixing Ca_3 and Mg_3, not Ca and Mg. Perhaps the best way to understand this is to think about the "size of the mole" problem. We showed in §9.10.2 that the molar Gibbs energy G obviously depends on the size of the mole. A mole of N_4 has twice the molar Gibbs energy of N_2, because there is twice the mass. But because $G = H - TS$, the same conclusion holds for the molar enthalpy and the molar entropy. Therefore the molar entropy. $(Ca, Mg)_3Al_2Si_3O_{12}$ is three times the molar entropy of $(Ca, Mg)Al_{2/3}SiO_4$, and hence the 3 in Equation (14.10).

Figure 14.3 Interaction coefficient w_G for the mixing of Ca and Mg on the cubic site in garnet, as determined by Hensen et al. (1975). Data in Table 14.2. The equation of the least squares fit is $w_G = 7460 - 4.3T$.

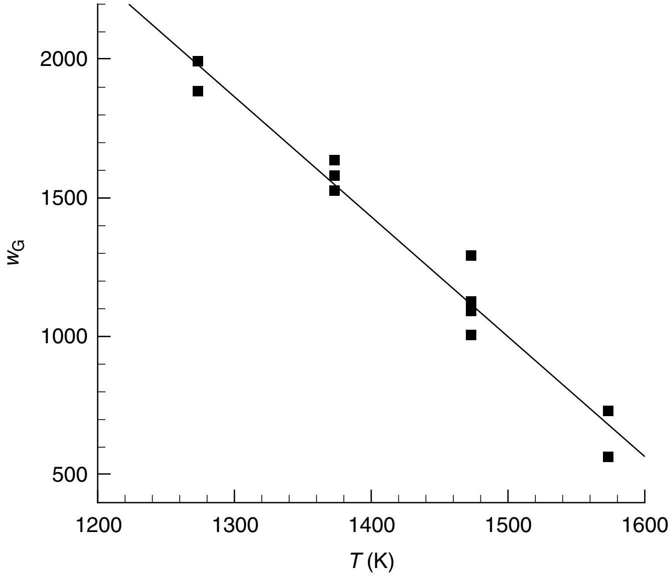
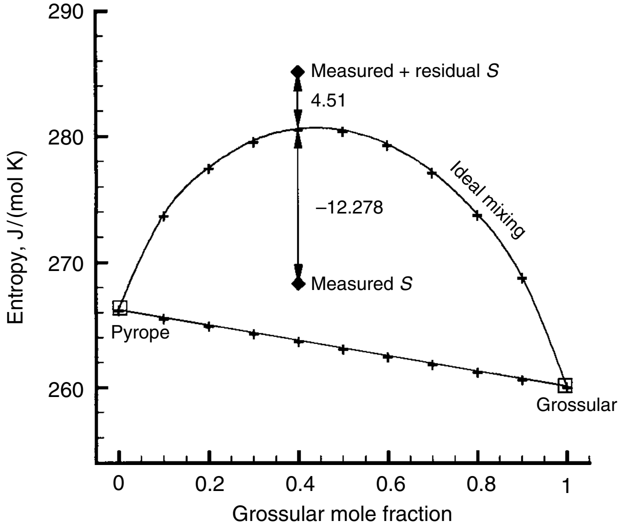


Figure 14.4 The entropy of mixing and the excess entropy of garnets.



The solid solution is assumed to have this residual entropy at 0 K, as well as at higher temperatures so the entropy measured by the heat capacity integral is added to it to get the total molar entropy.these calculations are illustrated on page 410 and in Figure 14.4.

A slightly positive V^{EX} has been obtained from unit call measurements (Haselton and Newton, 1980; Berman, 1990). These data for w_H, w_s and w_v have been combined by Haselton and Newton (1980) into expressions for the activity coefficients of pyrope and grossular.

Implicit in this treatment of Mg–Ca mixing in garnets is the assumption that the presence of other elements on this site or other sites will not affect the calculated parameters. For a more detailed analysis which does not make this assumption, see Berman (1990) and Berman and Koziol (1991).

The mixing properties of garnet components and many other rock-forming solid solutions is an on-going problem, and we can leave resolution of the problems of grossular-pyrope to the experts. The point made here is simply that determinations of excess properties can be made by calorimetry, measuring phase compositions, and X-ray work, and that these can be combined into values of w_G and hence of activity coefficients. None of this is easy to do.

14.3.3 Equilibrating phases

Determination of activity coefficients by equilibrating a phase with another phase having known activity coefficients is well illustrated by work in the system MgO–FeO–SiO_2.

The system MgO–FeO–SiO_2

Phase relations in the system MgO–FeO–SiO_2 were first worked out in detail by Bowen and Schairer in a classic paper in 1935. Dozens, if not hundreds, of studies have been conducted on minerals and mineral assemblages in this system since then. A generalized subsolidus section is shown in Figure 14.5. There are three important solid solution series in this system, two of which, the orthopyroxenes and olivines, are common rock-forming minerals. To examine the determination of activity coefficients in these solutions we will use

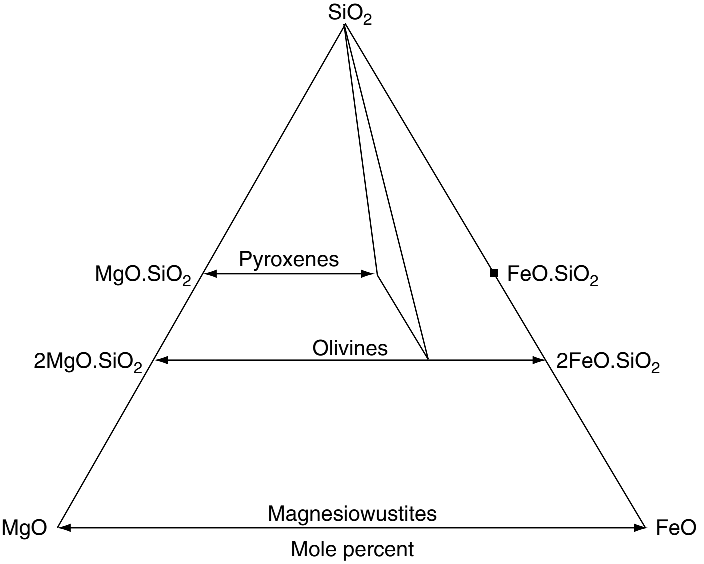


Figure 14.5 Subsolidus phase relations in the system MgO–FeO–SiO_2. This shows that ferrosilite ($FeSiO_3$) is not a stable phase, and that only iron-rich olivines can coexist with a SiO_2 phase.

$\Delta_{\text{mix}}S$ and S^{EX} in Garnets

The third law measurements of Haselton and Westrum (1980) are

Mineral	Composition	S J mol^{-1} K^{-1}	S cal mol^{-1} K^{-1}
Pyrope	$Mg_3Al_2Si_3O_{12}$	266.27	63.64
Grossular	$Ca_3Al_2Si_3O_{12}$	260.12	62.17
$Py_{0.6}Gr_{0.4}$	$(Mg_{0.6}Ca_{0.4})_3Al_2Si_3O_{12}$	268.32	64.13

A linear combination of the two end-members (often called "mechanical mixing") gives

$$\begin{aligned} S_{\text{linear mix}} &= \sum_i x_i S_i^\circ \\ &= (0.6 \times 266.27) + (0.4 \times 260.12) \\ &= 263.81\,\text{J mol}^{-1}\,\text{K}^{-1} \end{aligned}$$

The entropy of ideal mixing of ($CaAl_{2/3}SiO_4$ and $MgAl_{2/3}SiO_4$) is

$$\begin{aligned} \Delta_{\text{mix}} S_{\text{ideal sol'n}} &= -R\sum_i x_i \ln x_i \\ &= -8.31451(0.6\ln(0.6) + 0.4\ln(0.4)) \\ &= 5.596\,\text{J mol}^{-1}\,\text{K}^{-1} \end{aligned} \qquad [7.17]$$

and of ($Ca_3Al_2Si_3O_{12}$ and $Mg_3Al_2Si_3O_{12}$) is $3 \times 5.596 = 16.788$.

The entropy of the solid solution with ideal mixing is then $263.81 + 16.788\,\text{J mol}^{-1}\,\text{K}^{-1} = 280.598\,\text{J mol}^{-1}\,\text{K}^{-1}$. The heat capacity integral gives $268.321\,\text{J mol}^{-1}\,\text{K}^{-1}$, so the entropy of the solid solution is $268.32\,\text{J mol}^{-1}\,\text{K}^{-1}$assuming $S = 0$ at 0 K, or $268.32 + 16.788\,\text{J mol}^{-1}\,\text{K}^{-1}$ assuming a residual entropy of ideal mixing.

The excess entropy including the residual entropy is

$$\begin{aligned} S^{\text{EX}} &= S_{\text{measured}} + S_{\text{residual}} - S_{\text{ideal sol'n}} \\ &= 263.32 + 16.788 - 280.598 \\ &= 4.51\,\text{J mol}^{-1}\,\text{K}^{-1} \end{aligned} \qquad [18.23]$$

but would be $268.32 - 280.598 = -12.278\,\text{J mol}^{-1}\,\text{K}^{-1}$ without it.

The w_S parameter is then

$$\begin{aligned} w_S &= S^{\text{EX}}/x_{\text{Py}}x_{\text{Gr}} \\ &= 4.51/(0.6 \times 0.4) \\ &= 18.79\,\text{J mol}^{-1}\,\text{K}^{-1} \end{aligned} \qquad [10.89]$$

data from the work of Hahn and Muan (1962) and Nafziger and Muan (1967). Despite the fact that there has been quite a bit of experimental work on this system since 1967 the data we have chosen are sufficient for our purpose.

Both orthopyroxene and olivine have two nonequivalent sites over which Mg and Fe are distributed. The two sites could be called X and Y as in garnets, but historically they are called M_1 and M_2, so that pyroxene $[(Mg,Fe)SiO_3]$, for example, can be written $M_1M_2Si_2O_6$, and Fe–Mg distribution over the two sites may be ordered (most Fe in one site, most Mg in the other) or completely disordered (random occupancy of the two sites), and similarly for olivine. The degree of ordering is a strong function of temperature (Ganguly, 1982), and substantial ordering may remain even at 1200 °C, but Navrotsky (1971) showed that substantial deviations from random mixing on the two different sites leads to almost no deviation from the ideal mixing relationships. In other words, despite the fact that the solutions are not "regular" in the original sense, the regular solution formulations are still useful.

Experimental method

This work was done by the classical method of suspending fine-grained mineral assemblages in a furnace at high temperatures (mostly 1200 °C in this work) for 1–14 days, then allowing the assemblage to drop out of the furnace, so that the rapid cooling would "quench" the minerals and so retain the compositions they had in the furnace. The assemblage was then ground up again, and the heating/quenching process repeated two or three times to ensure that the minerals had reached their equilibrium compositions. Because the charges contained iron the oxidation state was a major concern, and this was controlled by causing a mixture of CO_2 and H_2 gases to flow through the furnace. These gases react at high temperatures very quickly, according to

$$CO_2(g) + 2\,H_2(g) = CH_4(g) + O_2(g) \tag{14.11}$$

so that the oxygen atmosphere in the furnace (the f_{O_2}) could be controlled by adjusting the CO_2/H_2 ratio of the gases flowing into the furnace.

The authors did not have an electron microprobe to determine the compositions of the minerals after the experiments. The compositions were determined by X-ray diffraction, comparing lattice parameter measurements with those from standards of known composition. For the pyroxenes, this was not sufficiently precise, so compositions were determined graphically by projecting tie lines from known coexisting olivine compositions through the bulk composition to the pyroxene boundary. This explanation may not be clear if you have no experience with phase diagrams; suffice it to say that the whole process is difficult and time-consuming.

Pyroxene solid solutions

The activity of the $FeSiO_3$ component of the $MgSiO_3$–$FeSiO_3$ (enstatite–ferrosilite) solid solution was obtained by using as experimental charges compositions in this series plus silica plus Fe metal, in the furnace atmosphere having the controlled f_{O_2} as described above. The relevant reaction is

$$(FeSiO_3)^{\text{in px ss}} = Fe + SiO_2 + \tfrac{1}{2}O_2(g) \tag{14.12}$$

for which the equilibrium constant is

$$\frac{a_{Fe}a_{SiO_2}a_{O_2}^{1/2}}{a_{FeSiO_3}} = K$$

With a_{Fe} and a_{SiO_2} fixed at 1.0 by the presence of the pure solid phases (SiO_2 is undoubtedly in the form of tridymite, but is not explicitly mentioned), this can be rewritten

$$\tfrac{1}{2}\ln f_{O_2} - \ln(x_{FeSiO_3}\gamma_{FeSiO_3}) = \frac{-\Delta_r G^\circ}{RT} \tag{14.13}$$

Using the α function (Equation 10.116) the activity coefficient of $FeSiO_3$ can be written

$$\ln(\gamma_{FeSiO_3}) = \alpha\, x_{MgSiO_3}^2$$

Inserting this into Equation (14.13),

$$\tfrac{1}{2}\ln f_{O_2} - \ln(x_{FeSiO_3}) = \frac{-\Delta_r G^\circ}{RT} - \alpha\, x_{MgSiO_3}^2 \tag{14.14}$$

If $\Delta_r G^\circ$ is known, measured pyroxene compositions and f_{O_2} values allow calculation of α. If $\Delta_r G^\circ$ is to be determined as well, we note that (14.14) is an equation in the form $y = ax + b$, where

$$y = \tfrac{1}{2}\ln f_{O_2} - \ln(x_{FeSiO_3})$$
$$x = x_{MgSiO_3}^2$$

Therefore a plot of $\frac{1}{2}\ln f_{O_2} - \ln(x_{FeSiO_3})$ versus $x_{MgSiO_3}^2$ will give a slope of α and an intercept of $-\Delta_r G^\circ/RT$.

The results showed that a_{FeSiO_3} and x_{FeSiO_3} were the same, within experimental precision. More exactly, they found that $\alpha = 0.07 \pm 0.12$, which is not significantly different from the ideal value of zero.

Magnesiowüstite solid solutions

The activities of FeO in solid solutions of FeO and MgO were obtained at 1100° and 1300 °C by Hahn and Muan (1962). The method was similar to that for the pyroxene solid solutions described above. Solid solutions of FeO and

MgO were equilibrated with metallic iron in a controlled atmosphere furnace. The reaction is

$$\text{FeO}^{\text{in sol'n}} = \text{Fe}^{\text{metal}} + \tfrac{1}{2}\text{O}_2(g) \tag{14.15}$$

Wüstite has a defect structure in which the Fe/O ratio varies with f_{O_2} and T. However in equilibrium with iron, the composition of pure wüstite (often written as "FeO" to indicate that the composition is not stoichiometric) is nearly constant at the composition $\text{Fe}_{0.95}\text{O}$. Therefore that composition is the standard state for the activities reported by Hahn and Muan.

Because the activity of pure iron is 1.0, the equilibrium constant for reaction (14.15) is

$$\frac{f_{\text{O}_2}^{1/2}}{a_{\text{FeO}}} = K \tag{14.16}$$

If the f_{O_2} for equilibrium of pure wüstite and pure iron is first determined, and then the f_{O_2} for equilibrium between wüstite in solution with MgO and pure iron is determined, the activity of FeO (relative to the $\text{Fe}_{0.95}\text{O}$ composition) is just (the square root of) the ratio of these two f_{O_2} values. Thus

$$a_{\text{FeO}} = \left(\frac{f_{\text{O}_2}^{\text{FeO in solution}}}{f_{\text{O}_2}^{\text{pure FeO}}} \right)^{1/2}$$

The results for a_{FeO} at 1100 and 1300 °C are shown in Figure 14.6.

The curve through the FeO activity data was drawn by assuming a regular solution, and determining w_G by plotting $RT \ln \gamma_{\text{FeO}}$ versus x_{MgO}^2 as in

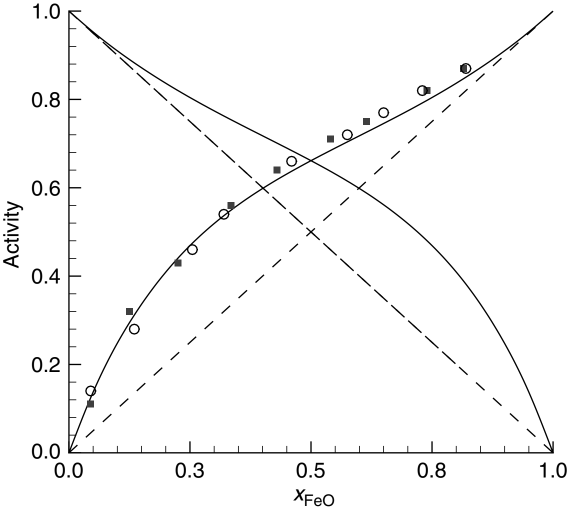


Figure 14.6 Activity of FeO and MgO in magnesiowüstite solid solutions. Data from Hahn and Muan (1962). Squares: 1100 °C. Circles: 1300 °C. Curves calculated as discussed in the text.

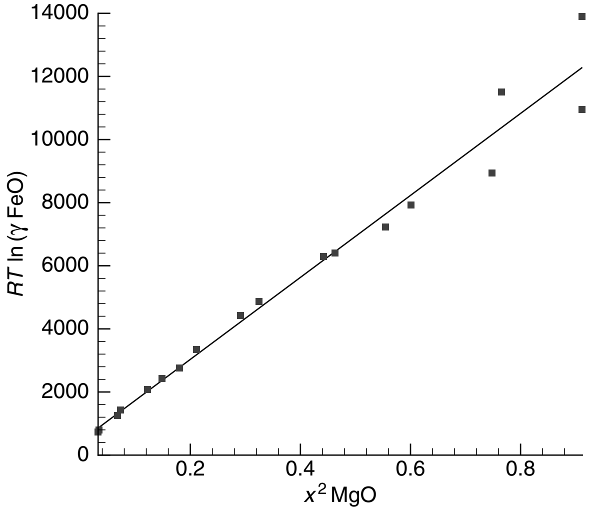


Figure 14.7 A plot of $RT \ln \gamma_{\text{FeO}}$ versus x^2_{MgO}. Least squares fit gives a value of $w_G = 12\,990\,\text{J mol}^{-1}$.

Figure 14.7. Values of γ_{FeO} are of course obtained by dividing each a_{FeO} measurement by its x_{FeO} value. The least-squares slope gives $w_G = 12\,990\,\text{J mol}^{-1}$, which was then used to calculate the a_{FeO} curve in Figure 14.6. If the least squares slope is forced to go through the origin, as it theoretically should, w_G changes to $13\,703\,\text{J mol}^{-1}$, but this hardly affects the position of the calculated a_{FeO} curve.

Because only FeO activities were measured, activities of MgO must be calculated using the Gibbs–Duhem relationship. Having a value for w_G, we can just put this into Equation (10.120), and the resulting curve for a_{MgO} is shown in Figure 14.6. Using the longer procedure of calculating α_{FeO} for each measurement and using Equation (10.119) results in a very similar curve.

Olivine solid solutions

The activities of the olivine components $\text{FeSi}_{0.5}\text{O}_2$ and $\text{MgSi}_{0.5}\text{O}_2$ were determined in three ways, that is, by using three different assemblages:

Olivine + SiO_2

Iron-rich olivines (fayalites) were equilibrated with SiO_2 and known gas phase compositions, exactly as the pyroxene case above. The reaction is

$$(\text{FeSi}_{0.5}\text{O}_2)^{\text{in ol ss}} = \text{Fe} + \tfrac{1}{2}\text{SiO}_2 + \tfrac{1}{2}\text{O}_2(\text{g}) \tag{14.17}$$

and activities of $\text{FeSi}_{0.5}\text{O}_2$ determined in the same way. This is only possible for a very limited range of iron-rich compositions, as shown in Figure 14.5.

Olivine + pyroxene

Assemblages of olivine and pyroxene were equilibrated, and Fe and Mg distributed themselves between the two solid solutions until equilibrium compositions were reached. Because the elements are exchanged between the two phases, the reaction is called an *exchange reaction*, written

$$\mathrm{FeSiO_3}^{\text{in px ss}} + \mathrm{MgSi_{0.5}O_2}^{\text{in ol ss}} = \mathrm{MgSiO_3}^{\text{in px ss}} + \mathrm{FeSi_{0.5}O_2}^{\text{in ol ss}} \tag{14.18}$$

and the equilibrium constant expression is

$$K = \frac{a_{\mathrm{MgSiO_3}}\, a_{\mathrm{FeSi_{0.5}O_2}}}{a_{\mathrm{FeSiO_3}}\, a_{\mathrm{MgSi_{0.5}O_2}}} \tag{14.19}$$

Because the pyroxenes are ideal, this can be written

$$K = \left[\frac{x_{\mathrm{MgSiO_3}}\, x_{\mathrm{FeSi_{0.5}O_2}}}{x_{\mathrm{FeSiO_3}}\, x_{\mathrm{MgSi_{0.5}O_2}}}\right] \frac{\gamma_{\mathrm{FeSi_{0.5}O_2}}}{\gamma_{\mathrm{MgSi_{0.5}O_2}}} \tag{14.20}$$

Now because $(\gamma_{\mathrm{FeSi_{0.5}O_2}}/\gamma_{\mathrm{MgSi_{0.5}O_2}})$ must be constant, and the two terms are also related through the Gibbs–Duhem equation (§4.14.2), they can in principle be calculated. If we let C represent the part of (14.20) within the square brackets, Wagner (1952, Chapter 7) showed that

$$\ln \gamma_{\mathrm{FeSi_{0.5}O_2}} = -x_{\mathrm{MgSi_{0.5}O_2}} \ln C - \int_0^{\mathrm{MgSi_{0.5}O_2}} \ln C \, dx_{\mathrm{MgSi_{0.5}O_2}} \tag{14.21}$$

and similarly for $\ln \gamma_{\mathrm{MgSi_{0.5}O_2}}$. Therefore determining the coexisting olivine–pyroxene compositions at equilibrium over a range of compositions is sufficient to determine the activity coefficients, and hence the activities, in one solid solution series (in this case the olivines), if those in the other series (the pyroxenes) are known. The results are shown in Figure 14.8 and Table 14.4.

Olivine + magnesiowüstite

Solid solutions of olivine and magnesiowüstite were equilibrated, and Fe and Mg exchanged according to the reaction

$$\mathrm{FeO}^{\text{in oxide}} + \mathrm{MgSi_{0.5}O_2^{\text{in olivine}}} = \mathrm{MgO}^{\text{in oxide}} + \mathrm{FeSi_{0.5}O_2^{\text{in olivine}}}$$

and the results analyzed exactly as in the olivine + pyroxene experiments. The only difference is that in the pyroxene solutions the activity coefficients were 1.0, while in the magnesiowüstite solutions the activity coefficients had other values.

Activities and mole fractions in Table 14.4 are combined to give activity coefficients and plotted versus $x^2_{\mathrm{MgSi_{0.5}O_2}}$ and $x^2_{\mathrm{FeSi_{0.5}O_2}}$, respectively, just as in Figure 14.7. The slope of the least squares fit is $w_G = 4785.2\,\mathrm{J\,mol^{-1}}$. Note that $x_{\mathrm{FeSi_{0.5}O_2}}$ and $x_{\mathrm{MgSi_{0.5}O_2}}$ are not independent, so w_G should be the same using one or the other, and in fact they are, within experimental precision. Using both together, as in Figure 14.9, gives a weighted average of the two.

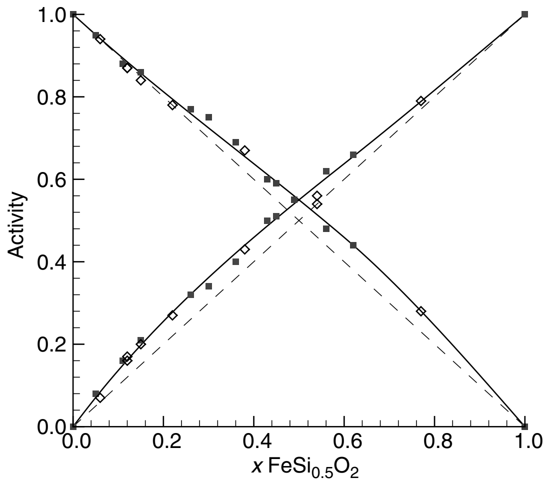


Figure 14.8 The binary system $\mathrm{MgSi_{0.5}O_2}$–$\mathrm{FeSi_{0.5}O_2}$ at 1200 °C. Squares represent data from olivine + pyroxene experiments, circles represent data from olivine + magnesiowüstite experiments (Table 14.4). The curves are drawn using activity coefficients calculated with Equation (10.98) using $w_G = 4785.2\ \mathrm{J\,mol^{-1}}$. Data from Nafziger and Muan (1967).

Table 14.4 *Olivine mole fractions and activities. Data from Nafziger and Muan (1967).*

Olivine + pyroxene equilibria				Olivine + magnesiowüstite equilibria			
$x_{\mathrm{FeSi_{0.5}O_2}}$	$x_{\mathrm{MgSi_{0.5}O_2}}$	$a_{\mathrm{FeSi_{0.5}O_2}}$	$a_{\mathrm{MgSi_{0.5}O_2}}$	$x_{\mathrm{FeSi_{0.5}O_2}}$	$x_{\mathrm{MgSi_{0.5}O_2}}$	$a_{\mathrm{FeSi_{0.5}O_2}}$	$a_{\mathrm{MgSi_{0.5}O_2}}$
0	1	0	1	0.06	0.94	0.07	0.94
0.05	0.95	0.08	0.95	0.12	0.88	0.16	0.87
0.11	0.89	0.16	0.88	0.12	0.88	0.17	0.87
0.15	0.85	0.21	0.86	0.15	0.85	0.20	0.84
0.26	0.74	0.32	0.77	0.22	0.78	0.27	0.78
0.30	0.70	0.34	0.75	0.38	0.62	0.43	0.67
0.36	0.64	0.40	0.69	0.54	0.46	0.56	0.54
0.43	0.57	0.50	0.60	0.77	0.23	0.79	0.28
0.45	0.55	0.51	0.59				
0.49	0.51	0.55	0.55				
0.56	0.44	0.62	0.48				
0.62	0.38	0.66	0.44				
1.00	0.00	1.00	0.00				

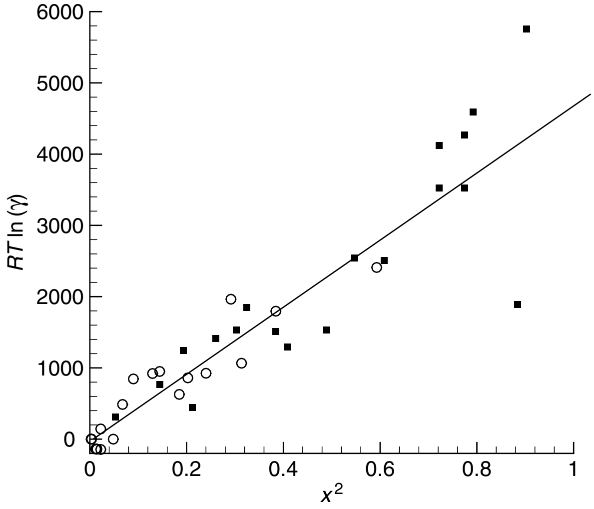


Figure 14.9 A plot of $RT \ln \gamma$ versus x^2 for olivines. The linear least squares fit has a slope of $w_G = 4785.2\,\mathrm{J\,mol^{-1}}$. Data calculated from Table 14.4.

14.3.4 Fitting Margules equations to solvus data

If a binary system exhibits a solvus (a composition region in which two solid solutions coexist at equilibrium; see Figures 17.12, 17.21 for examples), a convenient way of determining w_G values is to use the fact that the chemical potential of each component is the same in every pair of compositions at the same T (and P) on each side of the solvus. For example if we know that compositions x'_A and x''_A on a symmetrical solvus are at equilibrium, then from Equation (10.94) we write

$$RT\ln x'_\mathrm{A} + w_G(x'_\mathrm{B})^2 = RT\ln x''_\mathrm{A} + w_G(x''_\mathrm{B})^2$$

Rearranging and using $x_2 = 1 - x_1$, we can solve for w_G:

$$w_G = \frac{RT\ln(x'_1/x''_1)}{(1-x''_1)^2 - (1-x'_1)^2} \tag{14.22}$$

Using (14.22) we can estimate w_G from the known compositions of two coexisting phases at each temperature. If we solve for w_G at a number of temperatures along the solvus, and write w_G as a function of T, then we can get w_S from

$$(\partial w_G/\partial T)_P = -w_S$$

and w_H from (see page 305)

$$w_G = w_H - Tw_S$$

With luck, there might be data for the solvus at various pressures. We could then find w_G as a function of P, and calculate w_V,

$$(\partial w_G/\partial P)_T = w_V$$

Alternatively, we could estimate V^{EX} from X-ray data on the single-mineral solid solution just above the solvus.

The more usual case, however, is that the solvus is not symmetrical, but is skewed toward one of the components. In this case we need to solve for two Margules parameters, in order to use Equation (10.103),

$$G^{\mathrm{EX}}_{\mathrm{sol'n}} = x_1(w_{G_2}x_1x_2) + x_2(w_{G_1}x_1x_2) \qquad [10.103]$$

so we need two equations. Following the same procedure as before, but this time not converting x_2 into $(1-x_1)$ we get

$$\begin{aligned} RT\ln x_1' &+ (2w_{G_2} - w_{G_1})(x_2')^2 + 2(w_{G_1} - w_{G_2})(x_2')^3 \\ &= RT\ln x_1'' + (2w_{G_2} - w_{G_1})(x_2'')^2 + 2(w_{G_1} - w_{G_2})(x_2'')^3 \end{aligned} \qquad (14.23)$$

$$\begin{aligned} RT\ln x_2' &+ (2w_{G_1} - w_{G_2})(x_1')^2 + 2(w_{G_2} - w_{G_1})(x_1')^3 \\ &= RT\ln x_2'' + (2w_{G_1} - w_{G_2})(x_1'')^2 + 2(x_{G_2} - x_{G_1})(x_1'')^3 \end{aligned} \qquad (14.24)$$

Given the compositions of the two coexisting phases, the above two equations can be solved for w_{G_1} and w_{G_2}. To do this you could use one of many generalized equation solvers, or use rearranged versions of (14.23) and (14.24), which give w_{G_1} and w_{G_2} directly in terms of composition (see Thompson, 1967, p. 355 and Eugster et al., 1972, p. 164, for examples). For additional details and examples of applications to mineral systems, see Thompson and Waldbaum (1968, 1969), and Waldbaum and Thompson (1968, 1969).

14.4 Summary

We need to know which way reactions will go when solutions are involved, such as minerals dissolving/precipitating, or electrolytes dissociating. We have a thermodynamic potential (the Gibbs energy) which fills this role for pure phases, so we just need to know how to determine it for dissolved substances. In addition, we need to know how it changes with T, P, and composition of the solution, which involves knowing how to determine the temperature, pressure, and compositional derivatives of this quantity.

Excess properties, the difference between the property in a real solution and in an ideal solution, are generally expressed as a *relative* or *relative partial molar* properties, such as the relative enthalpy, $\mathbf{L}$, or *relative partial molar enthalpy*, $\overline{L}$. The Gibbs energy is treated differently. The fact that $\mathbf{G}_{T,P}$ is a thermodynamic potential leads naturally to the definition of a relative partial molar Gibbs energy $(\mu - \mu^\circ)$ which is not the difference from an ideal solution ($\mu - \mu^\circ$ is not zero even for an ideal solution) but the difference from a *standard state*, which in this chapter is a pure phase, but may also be some hypothetical state. The form of the equation relating $\mu - \mu^\circ$ to composition then leads naturally to the definition of the *activity*, an important quantity which essentially is what we were looking for – a compositional variable directly related to the Gibbs energy of a dissolved substance.

For ideal solutions, the activity is just the mole fraction, but for non-ideal solutions, a correction factor, the activity coefficient, is introduced. Fairly simple regular solution and Margules equations are then found which can express the variation of the excess Gibbs energy or the activity coefficients as a function of concentration for nonideal solutions. These are of limited usefulness for complex solid solutions such as the amphiboles, and are best at systematizing relatively simple binary solutions, fluid and solid, and when data are sufficient, ternary solutions. Beyond this, the notation gets very cumbersome, and the data requirements become very great. Nevertheless, within its range of usefulness, the Margules approach is very useful indeed, as witnessed by the large literature on the subject. Grover (1977) lists 40 publications, with brief descriptions of their contents, which use Margules equations in mineral systems.

15

Electrolyte solutions

15.1 Introduction

In dealing with the thermodynamic properties of ions we have one difficulty in addition to those encountered in dealing with compounds and elements. For compounds and elements we found that although we could measure absolute values for some properties, others such as enthalpy and the other energy terms contained an undetermined constant. We got around this by using the concept of "formation from the elements." It would of course be very convenient to also have thermodynamic properties of individual ions, but because positively and negatively charged ions cannot be separated from each other to any significant extent, their individual properties cannot be measured. To get around this, we need an additional convention, while retaining the formation from the elements convention.

In the following section we discuss the problems of activities of ionic species. Following that we discuss the conventions used to obtain numerical values for the state variables of individual ions, and we discuss the theory underlying the two major approaches to systematizing the data on electrolytes, the HKF and the Pitzer models. Because these are essentially equations of state, we introduced them in Chapter 13 (§13.6.2 and §13.6.3).

15.2 Activities of electrolyte components

15.2.1 Fugacity versus concentration for electrolytes

An electrolyte is a compound that splits up into charged particles when dissolved in water. For example, halite, NaCl, splits up into Na^+ and Cl^- particles (ions) in solution. What would you expect to be the consequences of this for activity and fugacity? To examine this let's consider the compound AB which splits up into particles A and B in solution. But before doing that, we should think about what we expect the relationship to be.

What to expect

In §8.2.3 we defined the standard state for solute i as the ideal ($\gamma_i^\circ = 1$) one molal ($m_i^\circ = 1$) solution. One consequence of this is that as $m_i \rightarrow 0$, $\gamma_i \rightarrow 1$, and $a_i \rightarrow m_i$. To simplify, we just say that in dilute solutions, $a_i = m_i$. If we were able to determine the fugacity of i in these solutions, we might not bother with this kind of standard state; we could use $a_i = f_i/f_i^\circ$, as we do for gases. But just because f_i/f_i° is not useful doesn't mean it is not true. If f_i° is the fugacity of i in an ideal one molal solution, and f_i is the fugacity of i in some other solution at the same T, then

$$\begin{aligned} a_i &= f_i/f_i^\circ \\ &= (m_i\gamma_i)/(m_i^\circ\gamma_i^\circ) \end{aligned}$$

so that in general

$$f_i = f_i^\circ a_i$$

and in very dilute solutions where $\gamma_i/(m_i^\circ\gamma_i^\circ) = 1$,

$$f_i = f_i^\circ m_i$$

In other words, f_i° is the (Henry's law) constant of proportionality for $f_i \propto a_i$, or $f_i \propto m_i$ in very dilute solutions. So we expect fugacity to be directly proportional to concentration in dilute solutions.

What we get

Now consider electrolyte AB, which dissociates into particles A and B. For the moment it doesn't matter whether they are electrically charged or not. In general, there will be an equilibrium constant for the reaction $\text{AB}^\circ = \text{A} + \text{B}$, and we can write an equilibrium constant for this (dispensing with activity coefficients for the moment, which just complicate but do not change the point we are getting at),

$$\begin{aligned} K_1 &= \frac{m_\text{A} \cdot m_\text{B}}{m_{\text{AB}^\circ}} \quad \text{and, because } m_\text{A} = m_\text{B}, \\ &= \frac{m_\text{A}^2}{m_{\text{AB}^\circ}} \end{aligned}$$

where we use m_AB for the analytical concentration, that is, the result of analyzing the solution for AB (in this case, this is *component* AB), and m_{AB° for the concentration of *species* AB in solution, where, because $m_\text{A} = m_\text{B}$, $m_\text{AB} = m_{\text{AB}^\circ} + m_\text{A}$ (or $m_\text{AB} = m_{\text{AB}^\circ} + m_\text{B}$). We can also suppose that in the vapor phase above the solution, the vapor pressure or fugacity of AB, f_AB, is proportional to m_{AB° (not to m_AB, because some of AB has become A and B). Thus Henry's law gives

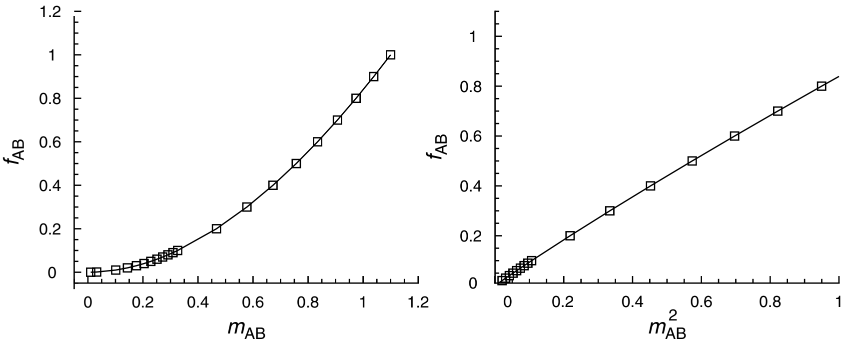
$$\begin{aligned} f_\text{AB} &\propto m_{\text{AB}^\circ} \\ &= K_2 \cdot m_{\text{AB}^\circ} \end{aligned}$$

If both K_1 and K_2 are equal to 10.0 and we consider a range of fugacities, we find the relations shown in Table 15.1 and Figure 15.1.

Clearly, fugacity is directly proportional to the square of concentration m^2_{AB}, in this case approaching a limiting value of 1.0. Plotting fugacity versus concentration m_{AB} gives a limiting slope of zero, meaning that increasing fugacity does not increase concentration, violating Henry's law. This relationship is perfectly reasonable, because fugacity should be related not to component AB but to species AB°.

Table 15.1 *The relationship between fugacity and* m_{AB} *and* m^2_{AB}

f_{AB}	m_{AB°	m_A	m_B	$m_{AB^\circ}+m_A = m_{AB}$	$(m_{AB^\circ}+m_A)^2 = m^2_{AB}$	f_{AB}/m_{AB}	f_{AB}/m^2_{AB}	K_1 and K_2
1	0.1	1.00	1.00	1.10	1.21	0.91	0.83	10
0.8	0.08	0.89	0.89	0.97	0.95	0.82	0.84	10
0.6	0.06	0.77	0.77	0.83	0.70	0.72	0.86	10
0.4	0.04	0.63	0.63	0.67	0.45	0.59	0.88	10
0.2	0.02	0.45	0.45	0.47	0.22	0.43	0.92	10
0.09	0.009	0.30	0.30	0.31	0.10	0.29	0.94	10
0.08	0.008	0.28	0.28	0.29	0.08	0.28	0.95	10
0.07	0.007	0.26	0.26	0.27	0.07	0.26	0.95	10
0.06	0.006	0.24	0.24	0.25	0.06	0.24	0.95	10
0.05	0.005	0.22	0.22	0.23	0.05	0.22	0.96	10
0.04	0.004	0.20	0.20	0.20	0.04	0.20	0.96	10
0.03	0.003	0.17	0.17	0.18	0.03	0.17	0.97	10
0.02	0.002	0.14	0.14	0.14	0.02	0.14	0.97	10
0.01	0.001	0.10	0.10	0.10	0.01	0.10	0.98	10
0.001	0.0001	0.03	0.03	0.03	0.00	0.03	0.99	10
0.0001	0.00001	0.01	0.01	0.01	0.00	0.01	1.00	10



(a) (b)

Figure 15.1 Illustration of the relationship between fugacity and concentration for an electrolyte AB. (a) f_{AB} versus m_{AB}. (b) f_{AB} versus m^2_{AB}. Data in Table 15.1.

The activity of HCl (l)

To show that this behavior is not confined to components like AB, but is in fact representative of real electrolytes, consider the case of HCl, used by Pitzer and Brewer (1961). Pure HCl is a gas that on dissolving in water dissociates to a very large degree into H^+ and Cl^-. The solubility of HCl in water is so great that even with very sensitive techniques, the concentration of HCl reaches about 4 molal before the vapor pressure of HCl above the solution becomes large enough to be measurable. Table 15.2 shows some values for the vapor pressure of HCl for solutions from 4 to 10 molal in HCl. These solutions are not sufficiently dilute to illustrate the dilute solution behavior we want, but Pitzer and Brewer show how to estimate the fugacities down to much lower values, and activity coefficients are known (Robinson and Stokes, 1959). The results are shown in Figure 15.2. Clearly, the behavior is the same as for component AB, and the explanation is likely the same as well.

Table 15.2 *HCl vapor pressure data from Pitzer and Brewer, (1961), p. 312. Activity coefficients from Robinson and Stokes (1959), Appendix 8.*

m	$P \times 10^4$ atm	$P \times 10^4$ bars	$\gamma_\pm$	$a = m\gamma_\pm$
4	0.2395	0.2427	1.762	7.048
5	0.6974	0.6884	2.38	11.90
6	1.842	1.866	3.22	19.32
7	4.579	4.640	4.37	30.59
8	11.10	11.25	5.9	47.20
9	25.39	25.73	7.94	71.46
10	55.26	55.99	10.44	104.4

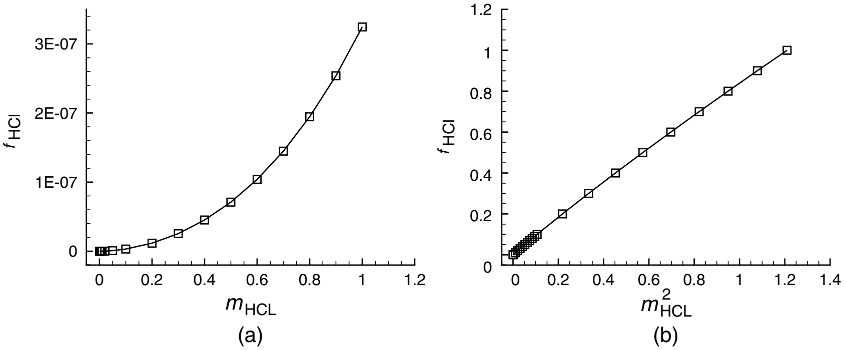


Figure 15.2 The calculated fugacity of HCl (a) as a function of m_{HCl}, and (b) as a function of m^2_{HCl}. Data from Pitzer and Brewer (1961), p. 312.

15.2.2 The activity of electrolytes

So we find that $f \propto m^2$ for a binary electrolyte like HCl or NaCl. In dilute solutions then,

$$f_{\mathrm{HCl}} = f^{\circ}_{\mathrm{HCl}} m^2_{\mathrm{HCl}}$$

or in general,

$$f_{\mathrm{HCl}} = f^{\circ}_{\mathrm{HCl}} (\gamma_{\pm} m^2_{\mathrm{HCl}})$$

and

$$\begin{aligned} a_{\mathrm{HCl}} &= f_{\mathrm{HCl}} / f^{\circ}_{\mathrm{HCl}} \\ &= \gamma_{\pm} m^2_{\mathrm{HCl}} \end{aligned} \tag{15.1}$$

where $\gamma_{\pm}$ is a new kind of activity coefficient to be discussed below.

If we went through the same procedure for an electrolyte which dissociates into three ions (e.g., Na_2SO_4) we would find $f \propto m^3$, and for four ions (e.g., $AlCl_3$) we would find $f \propto m^4$, and so on, so that for example,

$$a_{\mathrm{AlCl_3}} = \gamma_{\pm} m^4_{\mathrm{AlCl_3}}$$

In very dilute solutions, HCl is completely dissociated – the concentration of HCl° is undetectable, and $\gamma_{\pm} = 1$, so

$$a_{\mathrm{HCl}} = m^2_{\mathrm{HCl}}$$

Obviously we could get the same results, mathematically, not by squaring the concentration, but by taking the square root of the activity,

$$a^{1/2}_{\mathrm{HCl}} = m_{\mathrm{HCl}}$$

This raises an interesting question.

15.2.3 Choice of solute component

Before going on to define single ion activities and activity coefficients, let's pause to reflect on the similarity between the case considered here (a completely dissociated electrolyte in water), and the olivine solid solution case considered in Chapter 9 (§9.10.2). The physical systems are completely different, but the thermodynamic problem is almost identical, the only significant difference being that in the olivine case the concentrations were measured by mole fractions and ideality consisted in conforming to Raoult's law, while here concentrations are measured in molality and ideality is represented by Henry's law. Apart from that, the problem in both cases consists in choosing a solute component that is appropriate to the situation.

In Chapter 9 we showed the purely formal relation between component activities

$$a_i = (a_{0.5i})^2$$

and showed that in the case of Mg_2SiO_4–Fe_2SiO_4 solid solutions, the activity of component Mg_2SiO_4 would plot as a parabolic curve versus mole fraction, while component $MgSi_{0.5}O_2$ would come close to plotting as a straight line. In the HCl–H_2O case we find that component HCl is strongly curved on a plot of f (or a) versus m, but gives a straight line when plotted versus m^2. We are entitled to ask why, in the case of HCl, did we not behave consistently and plot component $H_{0.5}Cl_{0.5}$ versus m instead of changing the concentration axis from m to m^2? Obviously $a_{H_{0.5}Cl_{0.5}}$ versus m_{HCl} would result in a straight line too, because $a_{H_{0.5}Cl_{0.5}} = a_{HCl}^{1/2}$.

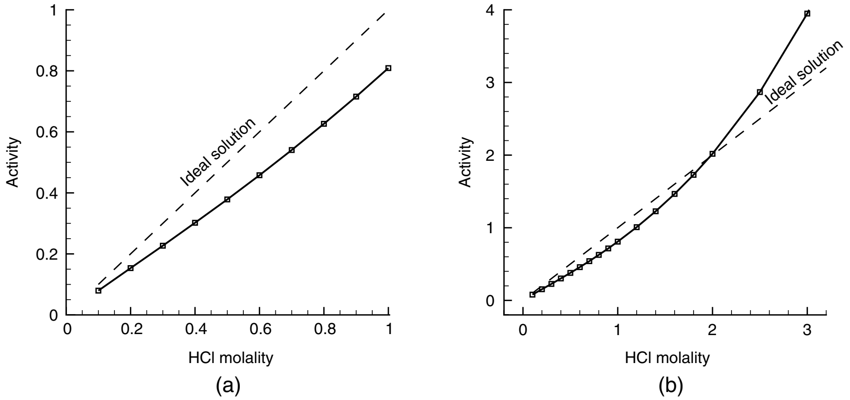
We can answer this in two ways. First, we *will* introduce and use component $H_{0.5}Cl_{0.5}$ (and analogous components for other compounds), but it is given a special symbol ($a_{\pm}$) for reasons that will become apparent. The other way to answer the question is to point out that although we need no knowledge of the molecular nature of a system to be able to apply thermodynamics to it, we would be silly to ignore such knowledge when it is available. Our rationale for using m^2 rather than m is based on overwhelming experimental evidence that many binary electrolytes split completely into two particles on dissolving in water. This provides the *reason* why a_{HCl} versus m_{HCl}^2 works better than a_{HCl} versus m_{HCl}, just as the observation that Mg and Fe effectively occupy a single crystallographic site in olivine provides the reason why $a_{MgSi_{0.5}O_2}$ versus x works better than $a_{Mg_2SiO_4}$ versus x. Thermodynamics doesn't care which component we use, but points out that if we are pleased by a_{HCl} versus m_{HCl}^2, we can get the same results from $a_{H_{0.5}Cl_{0.5}}$ versus m_{HCl}.

In general then, we see again that the choice of components is an important part of any thermodynamically based analysis or theory. Certain choices will "work better" or be more appropriate than others, and there will always be a reason for this. Although the reason *may* be that the component corresponds closely to the actual chemical species (N_2 certainly works better than N_4 in gas mixture equations), this is not necessarily the case as we have seen in both the olivine and HCl–H_2O examples.

15.2.4 Activity of HCl (II)

To develop the relationships shown in Figure 15.2 we relied on some rather hypothetical fugacity calculations. This helps to understand the relationships, but of course real, measured activities show the same thing. In Figure 15.3 is

Figure 15.3 Activity of aqueous HCl.
(a) Concentrations up to 1 molal.
(b) Concentrations up to 3 molal. Data from Robinson and Stokes (1959).



shown the measured activity of HCl in aqueous solution. These measurements can be made in various ways, such as isopiestic (§5.8.4), freezing point lowering, or conductivity, but the most accurate is by measuring the voltage of galvanic cells, as discussed in §12.7.1.

HCl is a typical strong electrolyte, virtually completely dissociated into ions at ambient conditions. Nevertheless, careful work can determine an equilibrium constant for the reaction

$$\mathrm{HCl}^\circ = \mathrm{H}^+ + \mathrm{Cl}^-$$

where, as for component AB, we use HCl° for *species* HCl. According to SUPCRT92, K for this reaction is $10^{5.13}$, meaning that there are very few HCl particles in dilute solutions.[1] Obviously then, the m_{HCl} referred to on the x-axis of Figures 15.3 does not refer to HCl°, it refers to the analytical concentration of HCl in the solution, which just happens to be present largely as ions. As before, this is *component* HCl. And, because the x-axis concentration is m and not m^2, the activity term must refer to $\mathrm{H}_{0.5}\mathrm{Cl}_{0.5}$, as discussed above, otherwise the activities would not be asymptotic to ideal behavior.

This activity term ($a_{\mathrm{HCl}}^{1/2}$ or $a_{\mathrm{H}_{0.5}\mathrm{Cl}_{0.5}}$) is called $a_{\pm,\mathrm{HCl}}$ in the electrolyte literature. Appropriately, the activity coefficients measurable as the ratio $a_\pm/m$ on this diagram are called $\gamma_{\pm,\mathrm{HCl}}$. The reason for using the notation $a_{\pm,\mathrm{HCl}}$ for $a_{\mathrm{H}_{0.5}\mathrm{Cl}_{0.5}}$ can be seen by continuing the molecularly based reasoning we were pursuing before we stopped to consider the similarity to the olivine case. Going back to the relation

$$a_{\mathrm{HCl}} = m_{\mathrm{HCl}}^2 \tag{15.2}$$

---

[1] Referring to HCl "particles" is not particularly realistic. They are ion pairs, the meaning of which has generated considerable discussion. For our purpose here, the point is simply the distinction between ion pair HCl and component HCl.

which holds for infinitely dilute solutions, we can next write

$$a_{\mathrm{HCl}} = m_{\mathrm{H^+}} m_{\mathrm{Cl^-}} \tag{15.3}$$

for the same conditions, because each mole of HCl breaks down completely to H^+ and Cl^-. At this point we introduce the concept of activity coefficients for individual ions, and define them such that they approach 1.0 in infinitely dilute solutions, just as we did with the activity coefficients of undissociated solutes. They are a useful mental concept in spite of not being measurable. At infinite dilution, then,

$$\gamma_{\mathrm{H^+}} = \gamma_{\mathrm{Cl^-}} = 1.0$$

and under other conditions they take on values such that

$$a_{\mathrm{HCl}} = (m_{\mathrm{H^+}}\gamma_{\mathrm{H^+}})(m_{\mathrm{Cl^-}}\gamma_{\mathrm{Cl^-}}) \tag{15.4}$$

remains true. Therefore

$$a_{\mathrm{HCl}} = a_{\mathrm{H^+}} a_{\mathrm{Cl^-}} \tag{15.5}$$

or

$$a_{\mathrm{H^+}} a_{\mathrm{Cl^-}} / a_{\mathrm{HCl}} = 1.0 \tag{15.6}$$

This defines the equilibrium constant for the reaction between *component* HCl and the ions H^+ and Cl^-. It follows then that for the reaction

$$\mathrm{HCl}\ (\textit{not}\ \mathrm{HCl}^\circ) = \mathrm{H^+} + \mathrm{Cl^-}; \quad \Delta_r G^\circ = 0 \tag{15.7}$$

As usual, it is best to see the truth of a relationship by understanding it rather than by seeing no fault with its derivation. In this case this can be accomplished by realizing that in the ideal one molal standard state to which $\Delta_r G^\circ$ refers, the solute component HCl consists in solution entirely as H^+ and Cl^-, therefore the G of component HCl has no choice but to be identical to $G_{\mathrm{H^+}} + G_{\mathrm{Cl^-}}$, from which it follows that $\Delta_r G^\circ = 0$. Pitzer (1995, p. 206) discusses a similar case, the hydration of aqueous NH_3 to NH_4OH, which happens to a small and unknown extent, i.e.,

$$\mathrm{NH_3}(aq) + \mathrm{H_2O}(l) = \mathrm{NH_4OH}(aq)$$

The activity of species NH_4OH° is unknown, so NH_4OH is considered a component, which of course is equal to the known amount of NH_3 in solution, so as before, $K = 1$ and $\Delta_r G^\circ = 0$. Another case is the hydration of dissolved CO_2, discussed in §9.10.

In the solution chemistry literature it is usual to refer to the solvent in a binary system as component 1, and to the solute (as normally written) as component 2. Therefore in this case a_{HCl} could also be referred to as a_2.

We can now see why $a_{\mathrm{H}_{0.5}\mathrm{Cl}_{0.5}}$ is called $a_{\pm,\mathrm{HCl}}$. It comes from the concept of individual ion activities, because if

$$a_{\mathrm{HCl}} = a_{\mathrm{H}^+} a_{\mathrm{Cl}^-}$$

then

$$\begin{aligned} a_{\mathrm{HCl}}^{1/2} &= (a_{\mathrm{H}^+} a_{\mathrm{Cl}^-})^{1/2} \\ &= a_{\pm,\mathrm{HCl}} \end{aligned} \tag{15.8}$$

which means that $a_\pm$ refers to the *geometric mean* of the activities of the two particles + and − resulting from dissociation, which in turn gives rise to its other name, the *stoichiometric mean ionic activity*. Similarly,

$$\begin{aligned} \gamma_{\mathrm{HCl}}^{1/2} &= (\gamma_{\mathrm{H}^+} \gamma_{\mathrm{Cl}^-})^{1/2} \\ &= \gamma_{\pm,\mathrm{HCl}} \end{aligned} \tag{15.9}$$

the *stoichiometric mean ionic activity coefficient*. The measurable quantity $\gamma_{\pm,\mathrm{HCl}}$ is considered to be the geometric mean of two unmeasurable quantities, γ_{H^+} and γ_{Cl^-}. The word *stoichiometric* is there because it doesn't really matter if dissociation is complete or not in the real solutions. The treatment is the same regardless. Similarly, we define the geometric mean of the ion concentrations as

$$m_{\pm,\mathrm{HCl}} = (m_{\mathrm{H}^+} m_{\mathrm{Cl}^-})^{1/2}$$

so that

$$a\pm = m_\pm \gamma_\pm$$

The great convenience of being able to think in terms of individual ionic properties means that we always look at ionic solutions this way, rather than as a simple changing of solute component.

15.2.5 Unsymmetrical electrolytes

The discussion so far has focused on HCl, but all symmetrical electrolytes will have the same relationships between m, a, and γ. By symmetrical electrolytes we mean those in which both ions have the same charge, so that on dissolution equal numbers of positive and negative ions result. Other examples are

NaCl, $MgSO_4$, HNO_3, and so on. However some complications develop for unsymmetrical electrolytes, i.e., those giving unequal numbers of positive and negative ions, such as Na_2SO_4, $AlCl_3$, and so on. To begin with f, a_2 and $a_{\pm}$, the definitions parallel those for the symmetrical case. As mentioned, for Na_2SO_4, $f \propto m^3$, while for $AlCl_3$, $f \propto m^4$, giving rise to the expressions

$$a_2 = (a_{\mathrm{Na^+}})^2(a_{\mathrm{SO_4^{2-}}}) \text{ for } \mathrm{Na_2SO_4}$$

and

$$a_2 = (a_{\mathrm{Al^{3+}}})(a_{\mathrm{Cl^-}})^3 \text{ for } \mathrm{AlCl_3}$$

which lead, for the same reasons as before, to

$$a_{\pm,\mathrm{Na_2SO_4}} = \left[(a_{\mathrm{Na^+}})^2(a_{\mathrm{SO_4^{2-}}})\right]^{1/3}$$

and

$$a_{\pm,\mathrm{AlCl_3}} = \left[(a_{\mathrm{Al^{3+}}})(a_{\mathrm{Cl^-}})^3\right]^{1/4}$$

which gives

$$a_{\mathrm{Na_2SO_4}} = a^3_{\pm,\mathrm{Na_2SO_4}}$$

and

$$a_{\mathrm{AlCl_3}} = a^4_{\pm,\mathrm{AlCl_3}}$$

Klotz (1964) points out that these definitions of activity for unsymmetrical salts imply new and rather strange standard states for these electrolytes. If we insist on having, for example,

$$a_{\mathrm{Na_2SO_4}} = (a_{\mathrm{Na^+}})^2(a_{\mathrm{SO_4^{2-}}})$$

it follows that in the binary system H_2O–Na_2SO_4 at very dilute concentrations of Na_2SO_4,

$$\begin{aligned} a_{\mathrm{Na_2SO_4}} &= (2m_{\mathrm{Na_2SO_4}})^2(m_{\mathrm{Na_2SO_4}}) \\ &= 4m^3_{\mathrm{Na_2SO_4}} \end{aligned}$$

This means that we cannot say, as before, that

$$\lim_{m\to 1}(f/m^3) = f^\circ$$

because this leads to

$$a_{\mathrm{Na_2SO_4}} = m^3_{\mathrm{Na_2SO_4}}$$

unless we let $\gamma_{\mathrm{Na_2SO_4}} \to 1/4$ as $m^3 \to 0$.

In order to preserve the convenience of having $\gamma \to 1$ as $m \to 0$ for all electrolytes, the standard state is chosen such that its fugacity, $f°$, is 1/4 of $h_{\mathrm{Na_2SO_4}}$, where $h_{\mathrm{Na_2SO_4}}$ is the Henry's law constant for $\mathrm{Na_2SO_4}$.

This results in $\gamma_{\pm,\mathrm{Na_2SO_4}}$ being defined as

$$\gamma_{\pm,\mathrm{Na_2SO_4}} = \frac{a_\pm}{4^{1/3}m} \tag{15.10}$$

and to retain the expression

$$\gamma_\pm = \frac{a_\pm}{m_\pm}$$

we define the *stoichiometric mean ionic molality* as

$$\begin{aligned} m_\pm &= \left[(m_{\mathrm{Na^+}})^2(m_{\mathrm{SO_4^{2-}}})\right]^{1/3} \\ &= \left[(2m_{\mathrm{Na_2SO_4}})^2(m_{\mathrm{Na_2SO_4}})\right]^{1/3} \\ &= 4^{1/3}m_{\mathrm{Na_2SO_4}} \end{aligned}$$

This is all quite confusing on the first run-through, but is quite logical. See Klotz (1964) for additional discussion. Table 15.3 contains a summary of these relationships.

Activity coefficient corrections (that is, correction to the concentration to get the activity) of one to two orders of magnitude such as these are not at all uncommon in aqueous systems.

Table 15.3 *Stoichiometric activity and activity coefficient expressions for single dissolved electrolytes. Expanded from Klotz (1964, Table 21.1)*

		NaCl	$\mathrm{Na_2SO_4}$	$\mathrm{AlCl_3}$	$\mathrm{MgSO_4}$	$\mathrm{A_{\nu_+}B_{\nu_-}}$
f_2	$\propto$	m^2	m^3	m^4	m^2	$m^{\nu_++\nu_-}$
a_2	=	a_+a_-	$a_+^2a_-$	$a_+a_-^3$	a_+a_-	$a_+^{\nu_+}a_-^{\nu_-}$
	=	$\gamma_\pm^2m^2$	$4\gamma_\pm^3m^3$	$27\gamma_\pm^4m^4$	$\gamma_\pm^2m^2$	$(\nu_+^{\nu_+}\nu_-^{\nu_-})\gamma_\pm^{(\nu_++\nu_-)}m^{(\nu_++\nu_-)}$
$a_\pm$	=	$(a_+a_-)^{1/2}$	$(a_+^2a_-)^{1/3}$	$(a_+a_-^3)^{1/4}$	$(a_+a_-)^{1/2}$	$(a_+^{\nu_+}a_-^{\nu_-})^{1/(\nu_++\nu_-)}$
a_+	=	$\gamma_\pm m$	$2\gamma_\pm m$	$\gamma_\pm m$	$\gamma_\pm m$	$\gamma_\pm\nu_+m$
a_-	=	$\gamma_\pm m$	$\gamma_\pm m$	$3\gamma_\pm m$	$\gamma_\pm m$	$\gamma_\pm\nu_-m$
$m_\pm$	=	m	$4^{1/3}m$	$27^{1/4}m$	m	$m(\nu_+^{\nu_+}\nu_-^{\nu_-})^{1/(\nu_++\nu_-)}$
$\gamma_\pm$	=	$(\gamma_+\gamma_-)^{1/2}$	$(\gamma_+^2\gamma_-)^{1/3}$	$(\gamma_+\gamma_-^3)^{1/4}$	$(\gamma_+\gamma_-)^{1/2}$	$(\gamma_+^{\nu_+}\gamma_-^{\nu_-})^{1/(\nu_++\nu_-)}$
	=	$a_\pm/m_\pm$	$a_\pm/m_\pm$	$a_\pm/m_\pm$	$a_\pm/m_\pm$	$a_\pm/m_\pm$

Activity of $CaCl_2$

From Robinson and Stokes (1959, p. 478) we find that in a 2.0 *m* solution, $\gamma_{\pm,\mathrm{CaCl_2}}$ at 25 °C is 0.792. Therefore

$$a_{\mathrm{Ca^{2+}}} = 2 \times 0.792 = 1.584$$

$$a_{\mathrm{Cl^-}} = 2 \times 2 \times 0.792 = 3.168$$

$$a_{\mathrm{CaCl_2}} = 1.584 \times 3.168^2 = 15.897,$$

or

$$a_{\mathrm{CaCl_2}} = 4 \times 0.792^3 \times 2^3 = 15.897$$

$$a_{\pm,\mathrm{CaCl_2}} = 15.897^{1/3} = 2.514$$

15.2.6 General relationships

The same relationships can be derived in a more formal (and more usual) way by beginning with the equations for the chemical potential of electrolyte solutes. Because the total Gibbs energy of the solute must equal the sum of its parts (cations + anions),

$$\left.\begin{aligned} \mu_2 &= \nu_+\mu_+ + \nu_-\mu_- \\ \mu_2^\circ &= \nu_+\mu_+^\circ + \nu_-\mu_-^\circ \end{aligned}\right\} \tag{15.11}$$

Expanding the top equation in (15.11),

$$\mu_2 = \nu_+(\mu_+^\circ + RT\ln a_+) + \nu_-(\mu_-^\circ + RT\ln a_-)$$

Combining this with

$$\mu_2 = \mu_2^\circ + RT\ln a_2$$

gives

$$a_2 = a_+^{\nu_+} a_-^{\nu_-} \tag{15.12}$$

where

$$\left.\begin{aligned} a_+ &= \gamma_+ m_+ \\ a_- &= \gamma_- m_- \end{aligned}\right\}$$

and

$$\left.\begin{aligned} m_+ &= \nu_+ m \\ m_- &= \nu_- m \end{aligned}\right\}$$

Dividing by ν (where $\nu = \nu_+ + \nu_-$) gives an *average* μ_2 and μ_2°, so

$$\mu_2/\nu = \mu_2^\circ/\nu + RT\ \ln(a_+^{\nu+} a_-^{\nu-})^{1/\nu}$$

Then defining

$$\left.\begin{aligned} \mu_\pm &= (\mu_+^{\nu+} \cdot \mu_-^{\nu-})^{\frac{1}{\nu}} \\ \mu_\pm^\circ &= (\mu_+^{\circ\nu+} \cdot \mu_-^{\circ\nu-})^{\frac{1}{\nu}} \\ a_\pm &= (a_+^{\nu+} \cdot a_-^{\nu-})^{\frac{1}{\nu}} \end{aligned}\right\} \tag{15.13}$$

we get

$$\mu_\pm = \mu_\pm^\circ + RT\ \ln a_\pm \tag{15.14}$$

Similarly $\gamma_\pm$ and $m_\pm$ are defined as

$$\left.\begin{aligned} \gamma_\pm &= (\gamma_+^{\nu_+} \cdot \gamma_-^{\nu_-})^{\frac{1}{\nu}} \\ m_\pm &= (m_+^{\nu_+} \cdot m_-^{\nu_-})^{\frac{1}{\nu}} \end{aligned}\right\} \tag{15.15}$$

The rest of the equations in Table 15.3 follow directly.

HKF notation

These relations can also be expressed generally using the notation of Helgeson et al. (1981). The terms m and $(\nu_+ + \nu_-)$ in the table above are equivalent to m_k and ν_k respectively in the following equations:

$$\begin{aligned} \gamma_{\pm,k} &= \left(\prod_j \gamma_j^{\nu_{j,k}}\right)^{1/\nu_k} \\ a_k &= \prod_j \left(\gamma_j \nu_{j,k} m_k\right)^{\nu_{j,k}} \\ &= \gamma_{\pm,k}^{\nu_k} m_k^{\nu_k} \prod_j \nu_{j,k}^{\nu_{j,k}} \\ a_{\pm,k} &= a_k^{1/\nu_k} \\ &= \gamma_{\pm,k} m_k \left(\prod_j \nu_{j,k}^{\nu_{j,k}}\right)^{1/\nu_k} \end{aligned}$$

where j represents the ions of component k, ν_j represents the stoichiometry of ion j, and

$$\nu_k = \sum_j \nu_{j,k}$$

For example, if k is $CaCl_2$,

$$\nu_{Ca^{2+}} = 1, \quad \nu_{Cl^-} = 2, \quad \nu_k = 3$$

and

$$\begin{aligned} a_{\mathrm{CaCl_2}} &= (\gamma_{\mathrm{Ca^{2+}}} m_{\mathrm{CaCl_2}})^1 (\gamma_{\mathrm{Cl^-}} 2m_{\mathrm{CaCl_2}})^2 \\ &= 4(\gamma_{\mathrm{Ca^{2+}}} \gamma_{\mathrm{Cl^-}}^2) m_{\mathrm{CaCl_2}}^3 \end{aligned}$$

For mixtures of electrolytes, with and without common ions, the notation becomes more complex. Helgeson et al. (1981) give the most comprehensive set of equations available.

15.2.7 Stoichiometric versus ionic properties

The word "stoichiometric" is included in the names given $a_\pm$, $\gamma_\pm$, and $m_\pm$ to indicate that they are derived according to the formulas given, which is more or less equivalent to the supposition that the electrolyte dissociates completely into ions. If the electrolyte dissociates only partly into ions, these quantities may still be derived, but they won't represent what they have been described as representing, i.e., the geometric mean of the properties of the free ions.

Therefore other classes of activities and activity coefficients are defined ("mean ionic activities" and "mean ionic activity coefficients," without the "stoichiometric"), which are related to the stoichiometric quantities by the degree of dissociation, α, where

$$\alpha = m_j/m$$

Thus

$$\gamma_j = \alpha \cdot \overline{\gamma}_j$$

where $\overline{\gamma}_j$ is the mean ionic activity coefficient of the jth ion, and for component k,

$$\gamma_j \cdot m = \overline{\gamma}_j \cdot m_j$$

$$\overline{\gamma}_{\pm,k} = \left(\prod_j \overline{\gamma}_j^{\nu_{j,k}}\right)^{\frac{1}{\nu_k}}$$

and similarly for the other properties.

We have introduced the factor α to account for the degree of dissociation, but as pointed out quite eloquently by Pitzer and Brewer (1961) in their revision of the classic text by Lewis and Randall (1923), the value of α varies with the method used to measure it; a situation not uncommon in studies of complex phenomena controlled by molecular interactions. Therefore it is to a large extent up to the investigator whether to use stoichiometric or ionic properties. Stoichiometric properties may be used whether or not ionic association is important, and have the advantage of not requiring estimates of α, or the amount of complexing and ion-pairing. As we will see later in this chapter, the Pitzer and HKF approaches to modeling electrolyte solutions exemplify these two different methods.

15.3 Numerical values for single-ion properties

We now take a look at how numerical values are assigned to the thermodynamic properties of single ions. There is more than one method, of course. Most physical chemistry courses for example will mention the MacInnes convention, which postulates that because the K^+ and Cl^- ions have similar properties, their activity coefficients should be identical. This means that

$$\gamma_{\pm,\mathrm{KCl}} = \gamma_{\mathrm{K}^+} = \gamma_{\mathrm{Cl}^-}$$

and from this, other single ion activity coefficients can be determined. For example,

$$\gamma_{\pm,\mathrm{CaCl_2}} = (\gamma_{\mathrm{Ca}^{2+}}\gamma^2_{\mathrm{Cl}^-})^{1/3}$$

so that

$$\gamma_{\mathrm{Ca}^{2+}} = \frac{\gamma^3_{\pm,\mathrm{CaCl_2}}}{\gamma^2_{\pm,\mathrm{KCl}}}$$

However, this method runs into problems, such as not working well in mixed electrolytes, and not giving coefficients compatible with Debye–Hückel coefficients, so it not widely used.

The usual practice is to give up any hope of determining individual ion activities or activity coefficients, and to use conventional ones. The convention as generally stated is that we assume all properties of the hydrogen ion to be zero. This, combined with another convention, that the properties of the elements are zero, results in the numbers we always use. But assuming these properties to be zero is not really necessary. What is calculated, of course, is the *difference* of an ion property from that of the hydrogen ion. Assuming the property of the hydrogen ion to be zero only simplifies the notation, not the concept.

In the following discussion of single ion properties, we do not at first make any assumptions about the properties of the elements or the hydrogen ion, but we do define properties in terms of differences. Then we introduce the conventions and see how this simplifies notation. The point is that thermodynamics is not dependent on any property being zero, only the notation is. In essence, the procedure followed is a slight variation of the "formation from the elements" procedure discussed in Chapter 3.

It will be convenient to continue with HCl in water as our example, so that the question becomes how do we arrive at numerical values for all the thermodynamic properties of the ions H^+ and Cl^-? The answer begins with the equation for the formation of H^+ and Cl^- from the elements,

$$\frac{1}{2}\mathrm{H_2}(g) + \frac{1}{2}\mathrm{Cl_2}(g) = \mathrm{H}^+ + \mathrm{Cl}^- \tag{15.16}$$

The standard Gibbs energy change for this reaction is found by combining data from three other reactions. First, the association of hydrogen and chlorine to give HCl gas has the properties

$$\frac{1}{2}\mathrm{H_2}(g)+\frac{1}{2}\mathrm{Cl_2}(g)=\mathrm{HCl}(g)$$

$$\log K=16.70;\quad \Delta_r G^\circ=-95\,299\,\mathrm{J\,mol^{-1}}$$

(data from Wagman et al., 1982) and the partitioning of HCl between liquid and vapor phases gives

$$\mathrm{HCl}(g)=\mathrm{HCl}(aq)$$

$$\log K=6.29;\quad \Delta_r G^\circ=-35\,929\,\mathrm{J\,mol^{-1}}$$

Combining these gives

$$\frac{1}{2}\mathrm{H_2}(g)+\frac{1}{2}\mathrm{Cl_2}(g)=\mathrm{HCl}(aq)$$

$$\log K=22.99;\quad \Delta_r G^\circ=-131\,228\,\mathrm{J\,mol^{-1}}$$

We have already shown that for the reaction

$$\mathrm{HCl}(aq)=\mathrm{H^+}+\mathrm{Cl^-}$$

$$\log K=1;\quad \Delta_r G^\circ=0$$

so it follows that for the reaction we want,

$$\frac{1}{2}\mathrm{H_2}(g)+\frac{1}{2}\mathrm{Cl_2}(g)=\mathrm{H^+}+\mathrm{Cl^-}$$

$$\log K=22.99;\quad \Delta_r G^\circ=-131\,228\,\mathrm{J\,mol^{-1}}$$

At this point we can go no further without making some arbitrary decision, i.e., without formulating a convention for ionic properties. The decision is basically to define the Gibbs energy of formation of the chloride ion as

$$\begin{aligned}\Delta_f\overline{G}^\circ_{\mathrm{Cl^-}}&=\overline{G}^\circ_{\mathrm{Cl^-}}+\overline{G}^\circ_{\mathrm{H^+}}-\tfrac{1}{2}G^\circ_{\mathrm{H_2}}-\tfrac{1}{2}G^\circ_{\mathrm{Cl_2}}\\&=-131\,228\,\mathrm{J\,mol^{-1}}\end{aligned}\qquad(15.17)$$

Once this decision is made for one ion, such as $\mathrm{Cl^-}$, the logjam is broken and values for all other ions can be derived. For example, we know that

$$\mathrm{Na}(s)+\tfrac{1}{2}\mathrm{Cl_2}(g)=\mathrm{Na^+}+\mathrm{Cl^-}$$

$$\Delta_r G^\circ=-393\,133\,\mathrm{J\,mol^{-1}}$$

so that

$$\overline{G}^\circ_{\mathrm{Na^+}}+\overline{G}^\circ_{\mathrm{Cl^-}}-G^\circ_{\mathrm{Na}}-\tfrac{1}{2}G^\circ_{\mathrm{Cl_2}}=393\,133\,\mathrm{J\,mol^{-1}}$$

Combining this with

$$\overline{G}^\circ_{\mathrm{Cl}^-} - \tfrac{1}{2}G^\circ_{\mathrm{Cl}_2} = -131\,228 + \tfrac{1}{2}G^\circ_{\mathrm{H}_2} - \overline{G}^\circ_{\mathrm{H}^+}$$

we get

$$\begin{aligned}\overline{G}^\circ_{\mathrm{Na}^+} - G^\circ_{\mathrm{Na}} &= -393\,133 - (\overline{G}^\circ_{\mathrm{Cl}^-} - \tfrac{1}{2}G^\circ_{\mathrm{Cl}_2})\\ &= -393\,133 - (-131\,228 + \tfrac{1}{2}G^\circ_{\mathrm{H}_2} - \overline{G}^\circ_{\mathrm{H}^+})\end{aligned}$$

so that

$$\overline{G}^\circ_{\mathrm{Na}^+} - G^\circ_{\mathrm{Na}} + \tfrac{1}{2}G^\circ_{\mathrm{H}_2} - \overline{G}^\circ_{\mathrm{H}^+} = -261\,905\,\mathrm{J\,mol^{-1}} \tag{15.18}$$

and this quantity is called $\Delta_f\overline{G}^\circ_{\mathrm{Na}^+}$.

So in general, just as the Gibbs energy of formation of a compound is the absolute Gibbs energy of the compound *minus* the sum of the Gibbs energies of the constituent elements, so the Gibbs energy of formation of an ion is the absolute Gibbs energy of the ion *minus* the Gibbs energy of the corresponding element, *plus* the quantity $Z_j(\tfrac{1}{2}G^\circ_{\mathrm{H}_2} - \overline{G}^\circ_{\mathrm{H}^+})$, where Z_j is the ionic charge. This is generalized formally as

$$\Delta_f\overline{\Xi}_{j,\mathrm{conventional}} = \overline{\Xi}_{j,\mathrm{absolute}} - \frac{1}{\nu_i}\Xi_i + Z_j(\tfrac{1}{2}\Xi_{\mathrm{H}_2} - \overline{\Xi}_{\mathrm{H}^+}) \tag{15.19}$$

where Ξ represents G, H, A, or U, and i represents the pure element, having stoichiometry ν_i, corresponding to ion j. Just as with the "formation from the elements" convention for compounds, the "excess baggage" [in this case $-\frac{1}{\nu_i}\Xi_i + Z_j(\tfrac{1}{2}\Xi_{\mathrm{H}_2} - \overline{\Xi}_{\mathrm{H}^+})$] always cancels out in balanced equations.

The same rules apply for all other thermodynamic properties of ions. For instance,

$$\begin{aligned}\Delta_f\overline{V}^\circ_{\mathrm{Ca}^{2+}} &= \overline{V}^\circ_{\mathrm{Ca}^{2+}} - V^\circ_{\mathrm{Ca}} + V^\circ_{\mathrm{H}_2} - 2\overline{V}^\circ_{\mathrm{H}^+}\\ \Delta_f\overline{C_P}^\circ_{\mathrm{Al}^{3+}} &= \overline{C_P}^\circ_{\mathrm{Al}^{3+}} - C^\circ_{P\mathrm{Al}} + \tfrac{3}{2}C^\circ_{P\mathrm{H}_2} - 3\overline{C_P}^\circ_{\mathrm{H}^+}\\ \Delta_f\overline{S}^\circ_{\mathrm{O}^{2-}} &= \overline{S}^\circ_{\mathrm{O}^{2-}} - \tfrac{1}{2}S^\circ_{\mathrm{O}_2} - S^\circ_{\mathrm{H}_2} + 2\overline{S}^\circ_{\mathrm{H}^+}\end{aligned}$$

Of course, for $\overline{V}^\circ$, $\overline{S}^\circ$ and $\overline{C_P}^\circ$ for which absolute values for the elements are available, "formation from the element" properties are not usually tabulated or used. Thus for instance

$$\begin{aligned}\Delta_f\overline{S}^\circ_{\mathrm{Na}^+} &= \overline{S}^\circ_{\mathrm{Na}^+} - S^\circ_{\mathrm{Na}} + \tfrac{1}{2}S^\circ_{\mathrm{H}_2} - \overline{S}^\circ_{\mathrm{H}^+}\\ &= 73.132\,\mathrm{J\,mol^{-1}\,K^{-1}}\end{aligned}$$

But S°_{Na} is known to be $51.210\,\mathrm{J\,mol^{-1}\,K^{-1}}$ and $S^\circ_{\mathrm{H_2}}$ is $130.684\,\mathrm{J\,mol^{-1}\,K^{-1}}$, so

$$\begin{aligned}\overline{S}^\circ_{\mathrm{Na^+,conventional}} &= \overline{S}^\circ_{\mathrm{Na^+,absolute}} - \overline{S}^\circ_{\mathrm{H^+}}\\ &= 73.132 + 51.210 - 130.684/2\\ &= 59.0\,\mathrm{J\,mol^{-1}\,K^{-1}}\end{aligned}$$

Therefore just as for compounds, the properties of the elements are not involved in the definition of the entropies, volumes, and heat capacities of the ions. The formal statement is

$$\overline{\Xi}_{j,\mathrm{conventional}} = \overline{\Xi}_{j,\mathrm{absolute}} - Z_j\overline{\Xi}_{\mathrm{H^+}} \tag{15.20}$$

where $\overline{\Xi}$ represents $\overline{S}$, $\overline{V}$, or $\overline{C_P}$.

15.3.1 Additional conventions

We can effect some simplification in the equations defining the thermodynamic properties of the ions by introducing the usual conventions mentioned above. If, for example, we say that the absolute Gibbs energies and enthalpies of all pure elements are set at zero, then the defining equation for Gibbs energies and enthalpies (Equation 15.19) becomes the same as that for $\overline{S}$, $\overline{V}$, and $\overline{C_P}$ (Equation 15.20). If in addition we define all properties of the hydrogen ion as zero, then the conventional ionic properties become the same as the corresponding absolute properties, and we could have stopped at Equation (15.17).

Another possibility is to assume that

$$\tfrac{1}{2}G^\circ_{\mathrm{H_2}} - \overline{G}^\circ_{\mathrm{H^+}} - \overline{G}^\circ_{\mathrm{e}} = 0 \tag{15.21}$$

where G°_{e} is the Gibbs energy of the electron. This is universally adopted in electrochemistry, so as to derive values for half-cell potentials. If (15.21) is adopted, but not necessarily the assumption of zero Gibbs energy for the elements, (15.17) becomes

$$\Delta_f\overline{G}^\circ_{\mathrm{Cl^-}} = \overline{G}_{\mathrm{Cl^-}} - \tfrac{1}{2}G_{\mathrm{Cl_2}} - \overline{G}_{\mathrm{e}}$$

which is the Gibbs energy change for the reaction

$$\tfrac{1}{2}\mathrm{Cl_2} + \mathrm{e} = \mathrm{Cl^-}$$

which is the simplest and most straightforward representation of the formation of the chloride ion from the element. Straightforward, that is, if you overlook the fact that the thermodynamic properties of electrons are unknown, and that it cannot serve as a basis for measurement. Nevertheless, (15.21) is a convenient and often-used convention.

It is important to realize that thermodynamics is in no way dependent on the assumption that the energies or enthalpies of the elements are zero, which is obviously untrue, and is one of the many factors lending a veil of uncertainty over thermodynamic proceedings. Nor is it dependent on the truth of Equation (15.21), which was discussed in Chapter 12.

15.4 The Debye–Hückel theory

In 1923 Peter Debye and Erich Hückel published two remarkable papers that described an a priori method of calculating the activity coefficient of electrolytic solutes in dilute solution. Without doubt this was one of the major breakthroughs in electrolyte solution theory.

The problem addressed by Debye and Hückel in 1923 was to describe an ionic solution using the primitive model of a solvent as an ideal structureless dielectric fluid, with solute ions as spherically uniform particles with charges located at their centers. Primary interactions are considered to be long range Coulombic forces. In a nonionic solution, thermal motions cause all particles to be randomly distributed. However, in ionic solutions, long range Coulombic forces cause each ion to be surrounded by a fluctuating group of ions of opposite charge, forming an "ionic atmosphere." Without the thermal motion that is always present, the ions in an electrolyte solution would actually assume the ordered structure of an ionic crystal. The model Debye and Hückel used included the competing effects of thermal motion (causing disorder) and Coulombic interaction (promoting order).

The complete derivation of the Debye–Hückel (D–H) theory is readily available in many standard references (e.g., Harned and Owen, 1958; Monk, 1961; Robinson and Stokes, 1959; Bockris and Reddy, 1970; as well as in an English translation of the original papers, Debye 1954). It is helpful in using the D–H theory to understand the physical model and assumptions used and the corresponding limitations, but it is too lengthy for inclusion here.

The Debye–Hückel equation in terms of stoichiometric mean ionic activity coefficients is (where γ_{H} indicates it is a Henryan activity coefficient)

$$\begin{aligned}\log \gamma_{\mathrm{H}} &= \log \gamma_{\pm} \\ &= \frac{-|z_+ z_-| A\sqrt{I}}{1 + \mathring{a} B\sqrt{I}} \qquad (15.22)\end{aligned}$$

where

$$\begin{aligned}A &= \frac{1.8248 \times 10^6 \rho_0^{1/2}}{(\epsilon_0 T)^{3/2}} \qquad (15.23) \\ &= 0.5091 \text{ at } 25\,^{\circ}\mathrm{C}\end{aligned}$$

and

$$B = \frac{50.292\rho_0^{1/2}}{(T\epsilon_0)^{1/2}} \tag{15.24}$$

$$= 0.3283 \text{ at } 25\,^\circ\text{C}$$

The parameters ρ_0 and ϵ_0 refer to the density and dielectric constant of pure water at the T and P of interest, z_+ and z_- are the valences of the cation and anion constituents of the salt, and å is the D–H distance of closest approach in units of angstroms. The product åB usually approximates unity. The D–H A and B parameters, calculated over a wide range of P and T, are tabulated by Helgeson and Kirkham (1974b, pp. 1202 and 1256). I is the molal ionic strength, defined by the following sum over all anions and cations,

$$I = \frac{1}{2}\sum_i m_i z_i^2 \tag{15.25}$$

The D–H equation (15.22), as we have seen, takes into account only long range electrostatic interactions between ions, and thus cannot be expected to work in solutions above a certain limiting concentration. It may easily be calculated that only two or three solvent molecules separate individual ions in a 1 molar solution. In practice the D–H equation works reasonably well to concentrations of approximately 10^{-2} to $10^{-1}m$ but no higher.

At very low concentrations the second term in the denominator of (15.22) approaches zero, and the equation reduces to

$$\log\gamma_\pm = -|z_+z_-|A\sqrt{I} \tag{15.26}$$

This is called the Debye–Hückel limiting law and has the advantage of being simpler and not including the adjustable å parameter; however, because of this, it cannot be used at concentrations above approximately 10^{-3} m. Its main use is in extrapolating experimental data on activity coefficients back to infinite dilution, because in the very dilute region, the extrapolated values must have the D–H slope. Well-known relations exist between activity coefficients and other thermodynamic parameters, so modifications of (15.26) (temperature and pressure derivatives) are used in extrapolating many kinds of data, not just activity coefficients.

The å parameter is adjustable and is usually determined by regression-fitting the D–H equation (15.22) to experimental data for individual salts. These have been tabulated for many compounds by Kielland (1937) and Butler (1964, pp. 434–5). These compilations give å for both anion and cation; where these values are not the same, the mean of the å values for cation and anion provides the best fit with observed activities (Butler, 1964, p. 436).

A well-known problem with using an adjustable å parameter for individual salts or ions is that if this is done for the individual solutes of a multicomponent solution, dG_{solution} will not be an exact differential, because the cross-differentiation criterion applied to the activity terms of the total differential (§C.2.1) will not be satisfied. To satisfy this criterion, the å parameter, or the åB product, in the denominator of the D–H equation should be the same for all solute components of a mixed electrolyte solution. This is one reason why several modifications and extensions of the D–H equation use a constant (often 1.0) for the åB term (Guggenheim, 1935).

15.4.1 The D–H extended term

Activity coefficients predicted by the D–H equation decrease monotonically with solute concentration. Measured activity coefficients typically decrease at first, but then increase at higher concentrations. This indicates that the simple coulombic model used by the D–H theory is inadequate in more concentrated solutions, which is not surprising. There have been many theoretical attempts to model the additional interactions that occur at high concentrations. Detailed summaries are given by Friedman (1962) and Helgeson et al. (1981), and a brief summary is in Nordstrom and Munoz (1994). Two of these are the $\dot{B}$ method and ion hydration method, but they are rather similar in effect, in that they add a more or less linear positive term to the right side of Equation (15.22).

The $\dot{B}$ term

The most general approach (Pitzer and Brewer, 1961, p. 326) would be to start with the D–H limiting law (Equation 15.26) and add a power series of virial coefficients, as for gases. A simpler approach, begun by Scatchard in 1936, and used by Pitzer and Brewer (1961, pp. 326, 578 and Appendix 4), is to define a deviation function $\dot{B}$ (called $\dot{B}$ to distinguish it from the first virial coefficient, which in a sense it replaces) as the difference between observed and predicted activity coefficients for an electrolyte such as NaCl. This is

$$\begin{aligned}\dot{B} &= \log \gamma_{\text{observed}} - \log \gamma_{\text{D-H}} \\ &= \frac{\log \gamma_{\text{observed}} + A|z_+z_-|\sqrt{I}/(1+\text{å}B\sqrt{I})}{I}\end{aligned} \tag{15.27}$$

The resulting modification of (15.22) is

$$\log \gamma_{\pm} = \frac{-|z_+z_-|A\sqrt{I}}{1+\text{å}B\sqrt{I}} + \dot{B}I \tag{15.28}$$

$\dot{B}$ was redefined by Helgeson (1969) by replacing the stoichiometric ionic strength I with the "true ionic strength" $\overline{I}$ which includes a correction for ion-pairing and complexing. Various methods of calculating $\overline{I}$ are described

Calculating å and $\dot{B}$

Calculating values of å and $\dot{B}$ with modern computing tools is easy. For example, most spreadsheets these days have a "solver" option. Billo (2001) shows how this can be used to provide the least-squares values to any number of coefficients in an equation. His book is addressed to Excel® users, but will work with any spreadsheet having a solver.

In the following example, data from Archer (1992) on the stoichiometric mean ionic activity of NaCl at 25 °C are fitted to Equation (15.28), with å and $\dot{B}$ as unknowns.

A molality	B $\gamma_\pm$	C Equation (15.28)	D (difference)^2
0	1	1.0000	0
0.01	0.9023	0.9011	1.444E – 06
0.025	0.8597	0.8570	7.133E – 06
0.05	0.8205	0.8160	1.984E – 05
0.1	0.7771	0.7702	4.774E – 05
0.2	0.7326	0.7231	8.943E – 05
0.4	0.6917	0.6812	1.104E – 04
0.6	0.6724	0.6628	9.223E – 05
0.8	0.6622	0.6544	6.097E – 05
1.0	0.6572	0.6515	3.209E – 05
1.5	0.6582	0.6578	1.982E – 07
2.0	0.6708	0.6748	1.606E – 05
2.5	0.6912	0.6985	5.292E – 05
3.0	0.7179	0.7270	8.324E – 05
3.5	0.7501	0.7596	9.095E – 05
4.0	0.7874	0.7959	7.244E – 05
4.5	0.8298	0.8357	3.433E – 05
5.0	0.8774	0.8788	2.012E – 06
5.5	0.9302	0.9254	2.298E – 05
6.0	0.9887	0.9755	1.744E – 04
		sum of squares ⇒	0.00101097

Column A is the NaCl concentration, Column B the activity coefficient from Archer (1992), Column C the activity coefficient calculated from Equation (15.28), and Column D is the square of the difference between Columns C and D. At the bottom of Column D is the sum of squares, and the solver finds values of å and $\dot{B}$ which minimize this value.

(continued)

Column C contains formulas like

=EXP((((-A*SQRT(A73))/(1+B*anot*SQRT(A73)))+Bdot*A73)*2.30259)

where A, B, anot and Bdot are the names of cells in the spreadsheet. A73 is a cell in column A. Cells named A and B contain fixed values of 0.5091 and 0.3283, and cells named anot and Bdot have values which are changed by the solver to minimize the sum of squares. Initial guesses as to the values of å and $\dot{B}$ must be provided, and in this case they must be reasonably close to the final, best values. In other problems, initial values are less important.

The results are å $= 3.426$ Å, and $\dot{B} = 0.0536$, and these values were used in drawing the curve through the diamond symbols in Figure 15.4. å and $\dot{B}$ are highly correlated, and should be determined together unless å is accurately known.

by Brimhall and Crerar (1987, p. 280). Values of $\dot{B}$ for several electrolytes at 25 °C as a function of concentration are shown in Figure 15.5. The constancy of these values at higher concentrations is remarkable, and influenced Pitzer in his development of his ion-interaction theory (§15.7).

Helgeson (1969) derived values of $\dot{B}$ at elevated temperatures by fitting equation (15.27) to activity coefficient data for NaCl (after determining å values using the hydration approach, below) and suggested that these values could be used for all solute components in hydrothermal solutions, based on the fact that NaCl is the dominant solute in most such solutions. Testifying to the impact of this 1969 work is the fact that these $\dot{B}$ values are still widely used today, despite

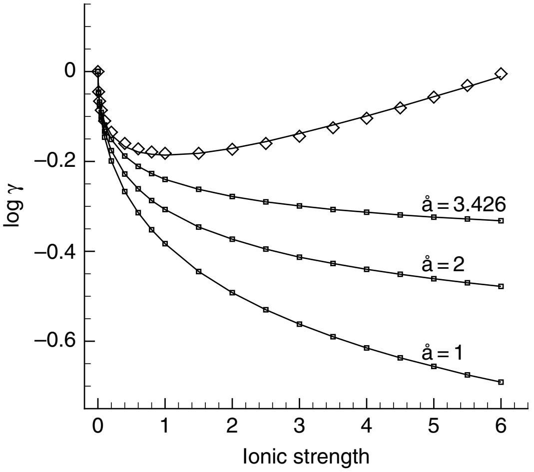


Figure 15.4 Curves labeled å = 1, etc., are calculated from Equation (15.22). Diamond symbols are stoichiometric mean ionic activity coefficients for NaCl at 25 °C from the EoS of Archer (1992) as a function of (stoichiometric) ionic strength. The curve through the diamond symbols is calculated using Equation (15.28)

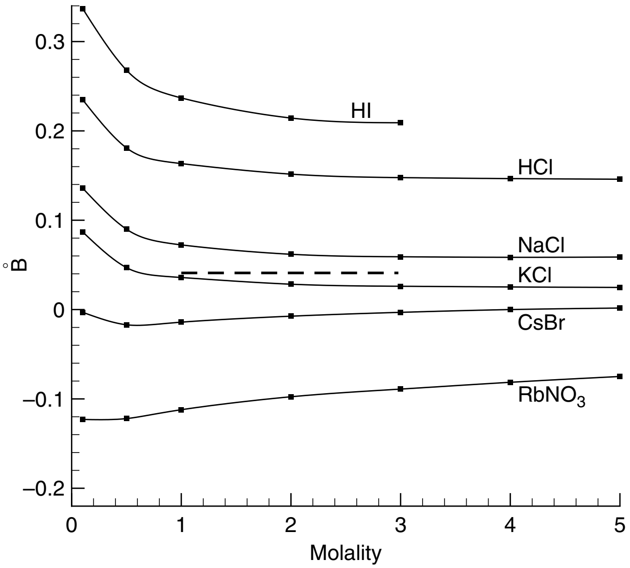


Figure 15.5 $\dot{B}$ versus molality for several salts. Data from Pitzer and Brewer (1961).

that fact that revised values (called b_γ) for NaCl and several other salts, and for NaCl as a function of P as well as T, were provided by Helgeson et al. (1981).

Helgeson et al. (1981) also revised Equation (15.28) so that the $\dot{B}$ parameter explicitly includes a hydration parameter derived from the Born equation and a second parameter for remaining short range interactions.

The ion hydration approach

An alternative to the $\dot{B}I$ term is a combination of terms representing ion-hydration (Stokes and Robinson, 1948),

$$\log \gamma_{\pm,i} = \frac{-A|z_+z_-|\sqrt{I}}{1+\mathring{a}B\sqrt{I}} - \frac{h}{\nu}\log a_\mathrm{w} - \log[1+0.018(\nu-h)m] \qquad (15.29)$$

which serves the same purpose, i.e., that of increasing $\log\gamma$ at high concentrations. In this equation, a_w is the activity of water, h is a hydration number, and ν is the number of ions in the electrolyte (e.g., 2 for NaCl). Equation (15.29) is also capable of successfully fitting activity coefficients of some electrolytes to 4 to 5 molal, where both å and h are adjustable parameters. This equation was used in the influential work of Helgeson (1969) to determine å and h values as a function of temperature, by fitting the equation to NaCl activity coefficients.

Nesbitt (1982) pointed out that using hydration numbers which are independent of solution concentration is unreasonable. For example, at about 4 molal MgCl$_2$, and using the Stokes and Robinson hydration number (h) of 13.7, virtually all the water in solution would reside in hydration shells, leaving none as solvent. Nesbitt (1982) suggested a minor variation in the hydration scheme, in which water is partitioned between solvent and hydration shells, such that the hydration number decreases as the solvent activity decreases. The resulting equation (15.30) fits stoichiometric activity coefficients to considerably higher ionic strengths than the original equation (15.29).

$$\log \gamma_{\pm,i} = \frac{-A|z_+z_-|\sqrt{I}}{1+B\mathring{a}\sqrt{I}} - \frac{kh^*(a_w-1.0)}{2.303\nu} - \log[1+0.018(\nu-kh^*a_w)m] \qquad (15.30)$$

In this equation, h^* is the maximum number of waters in the hydration shell (at low ionic strength) and k is the distribution coefficient for water between solvent and hydration shell. Fitting activity coefficient data then results in values of å and kh^*. For many chloride 1:1 and 2:1 electrolytes, the fit is identical to that from Equation (15.29) at low ionic strength, but remains good to much higher ionic strengths. Wolery and Jackson (1990) have proposed other modifications to the hydration approach.

Comment on extended terms

The Debye–Hückel theory is a cornerstone of electrolyte theory. It is always used in extrapolating data to infinite dilution, and must be embedded in any generalized treatment of activity coefficients as a function of concentration, as it is in the Pitzer equations. However, at concentrations beyond the validity of the limiting law (Equation 15.26), all attempts at predicting electrolyte behavior at higher concentrations are more or less empirical.

The two approaches to this problem current in geochemistry are the Pitzer and the HKF models. Extended terms are not a problem for Pitzer; his ion-interaction parameters handle all concentrations above the limiting law. But the fact is that for several reasons, the HKF model is by far the most extensively used in geochemical calculations today, and will be for the foreseeable future. This model is essentially a standard state model. Using it, one can calculate useful equilibrium constants over a wide range of T and P. However, to use the equilibrium constants in real situations, you need activities and hence activity coefficients, and by far the most common method of obtaining these is to use the "B-dot method", Equation (15.28). This equation is built in to all widely used geochemical modeling programs such as PHREEQC and The Geochemist's Workbench™.

For example, the database (llnl.dat) for program PHREEQC, an excellent and versatile program from the USGS, contains the following lines:

```
-bdot
    0.0374  0.0410  0.0438  0.0460
    0.0470  0.0470  0.0340  0.0000
```

where `-bdot` is the $\dot{B}$ deviation parameter for eight temperatures from 0 to 300 °C. They are fit with an equation for intermediate temperatures during program operation, and are used to calculate the activity coefficient for every species using Equation (15.28). These values are from Helgeson (1969),

Table 2. The database used by The Geochemist's Workbench™ has identical values.

The problem with this is that although appropriate values of å and $\dot{B}$ can fit activity coefficients to high concentration, as shown in Figure 15.4, they must be determined for individual sets of data. In addition, they are highly correlated, so that values of $\dot{B}$ are very sensitive to changes in å. The databases mentioned use a single, fixed value of å for each species at all temperatures, obtained from generalized sources such as Kielland (1937). Then there are lesser problems, such as that Helgeson (1969) used "true ionic strength" in determining $\dot{B}$, whereas the å values from most sources were determined using stoichiometric ionic strength.

What happens to these terms in complex multisalt solutions at high temperatures and pressures is generally unknown. But then, many geochemical calculations give results for situations in which no one knows what to expect, so errors introduced by activity coefficients go unrecognized. Where comparisons with experimental data are made, the results can be shockingly bad, as shown in Figure 15.6.

Apart from the problem of lack of data, especially kinetic data, this situation probably constitutes the most serious problem in the application of thermodynamics to geochemical problems involving hydrothermal solutions. The development of an algorithm which is more accurate than the B-dot method as presently used, but which could still be used for any species at any T and P would be a significant advance in geochemical modeling practice.

15.5 Activity coefficients of neutral molecules

Activity coefficients of electrolytes are determined in a number of ways, but at higher temperatures the problems become more difficult. A pioneering effort was the direct determination of water activity in NaCl solutions by a high-temperature type of isopiestic apparatus by Liu and Lindsay (1972), and the Gibbs–Duhem calculation of NaCl activity coefficients. More recently, high-temperature flow calorimetry determinations of excess enthalpy and heat capacity, plus density measurements, are combined to give excess Gibbs energies and hence activity coefficients.

The determination of the activity coefficients of species that exist dominantly as neutral molecules, such as $SiO_2(aq)$, $H_2S(aq)$, and $CO_2(aq)$, is much simpler. In these cases it is usually possible to establish a two-phase equilibrium between the substance in its pure state (solid or gaseous) and the substance in its aqueous or dissolved state. This leads to a simple and rigorous determination of the activity coefficient in solutions of varying composition.

For example, consider H_2S gas in equilibrium with $H_2S(aq)$. The first ionization constant of H_2S is about 10^{-7}, so that species other than molecular H_2S can be neglected in this connection. The reaction of interest is

$$H_2S(g) = H_2S(aq)$$

for which the equilibrium constant is

$$\begin{aligned} K &= a_{\mathrm{H_2S}(aq)}/a_{\mathrm{H_2S}(g)} \\ &= m_{\mathrm{H_2S}}\gamma_{\mathrm{H_2S}}/f_{\mathrm{H_2S}} \end{aligned}$$

$f_{\mathrm{H_2S}}$ can either be taken as equal to the partial pressure of H_2S or calculated from this quantity, and $m_{\mathrm{H_2S}}$ is measurable as the total sulfur content of the solution. Then if K is known, $\gamma_{\mathrm{H_2S}}$ can be obtained in any solution simply by measuring $f_{\mathrm{H_2S}}$ and $m_{\mathrm{H_2S}}$. For solids such as SiO_2 which dissolve to neutral species, the situation is even simpler, as the activity of the solid can be taken as 1.0. Normally, K is obtained by performing the experiment in pure water, and assuming that in this case, $\gamma_i = 1$. Therefore, activity coefficients obtained in this way are actually values of

$$\frac{\gamma_i(\text{ in solution of interest, i.e., a salt solution})}{\gamma_i(\text{ in pure water}).}$$

If the "solution of interest" is a solution of neutral salt B in water, having concentration m_{B}, eliminating the first term in Equation (15.28) gives

$$\log \gamma_i = k_{\mathrm{S}} m_{\mathrm{B}}$$

where k_{S} replaces $\dot{B}$ and is called the Setchénow coefficient, and m replaces I. In some treatments, the I or $\overline{I}$ term is retained. It is common practice in modeling studies (e.g., as in virtually all commonly available speciation programs) to assume that the activity coefficient of all uncharged species is 1.0, or to have a single fixed Setchénow coefficient, usually 0.1, which is a common value at 25 °C. However, measured activity coefficients of aqueous H_2S, CO_2, SiO_2, H_3BO_3, etc., show a much more complex behavior, being quite different in different salt solutions, with positive values in some and negative in others. The Pitzer equations provide a more detailed and accurate treatment of uncharged species in the form of measured λ coefficients (Equations 15.35–15.39). Barta and Bradley (1985) derived Pitzer interaction coefficients from the solubility of H_2S, CO_2 and CH_4 in NaCl solutions from 25 to 350 °C, and 0–6 molal. This was used by Barrett et al. (1988) to derive Setchénow coefficients for these gases to 300 °C. Also note the completely different approach using regular solution theory by Nesbitt (§10.6.1).

An extensive review of neutral solutes in aqueous salt solutions is given by Randell and Failey (1927a, b, c). See also Long and McDevit (1952) and Oelkers and Helgeson (1991). Note that we have only spoken of neutral species of the type that can be obtained as the *dominant* species in a solution; activity coefficients for the neutral species of weak electrolytes and other neutral species in a matrix of charged particles constitute a more difficult problem. Their activity coefficients are usually assumed to be 1.0, or are taken as equal to those of some other neutral species such as H_2S or CO_2 under the same conditions, without much justification.

The activity coefficients of neutral species in electrolyte solutions at elevated temperatures and pressures is a subject of continued study (Oelkers and Helgeson, 1991).

15.6 Ion association, ion pairs, and complexes

Soon after the appearance of the Debye–Hückel theory, it was found that the theory did not work well for many electrolytes. In 1926, Bjerrum suggested that electrostatic attraction between pairs of oppositely charged ions resulted in "ion-pairs," which would account for the lower measured activity coefficients in these solutions. The problem then, as now, was how best to define and measure the extent of ion-pairing. How close must two ions be to become an ion-pair? What is the difference between an ion-pair and a complex? Is it really necessary to know this to use thermodynamics? Helgeson (1981) notes that "The distinction between ion association and short-range ionic interaction is nebulous at best."

A simple view is that in an ion-pair the two ions are separated by one or more water molecules, while in a complex the ions are in contact, often covalently bonded. However, the distinction is not easily made, and there appear to be all gradations between the two. One common distinction, however, is that at higher temperatures, there is an increase in ion-pairing, but a decrease in complex formation. The reason for increased ion-pairing is due mainly to the decrease in the dielectric constant of water, as discussed in §6.2.4, and the decrease in complexing is due to the increased thermal energy of the particles, tending to break them apart.

The decrease in dielectric constant of water, resulting in increased electrostatic attraction between oppositely charged ions, means that many "strong" electrolytes become "weak" at high temperatures, and the identification of solute species becomes easier, or even possible, by various spectroscopic methods. That is, there are sufficient particles, and they stay together long enough, to produce absorption of radiation. This difference between solutions at room temperature and those at high temperature results in a corresponding difference in opinions about how best to think about them.

For most electrolyte solutions at or near room temperature, there have been two main approaches, exemplified by the Pitzer and HKF models to be described below. (These two approaches were also mentioned on page 304.) In one, referred to as ion-interaction theory and embodied in the Pitzer equations, no attempt is made (except if there are very strong complexes present) to identify species. Components are treated stoichiometrically, and all ion interactions are accounted for in the form of fit coefficients in some form of equation. Advocates of this approach point to the rather uncertain nature of our knowledge of aqueous species. Molecular dynamics simulations of these solutions (e.g., Driesner et al, 1998) show that the "real" situation is probably very complicated. So the answer to one question is no, you do not have to know anything about what is happening in the solution to use thermodynamics. In the other approach, called ion-association, and embodied in the HKF model and program SUPCRT92, ion-pair and complex species are identified and their activities calculated using activity coefficients calculated from D–H theory. The fact that such species become more abundant and more easily measured at high temperature therefore has an effect on one's thinking about how to approach this problem.

Pitzer, in a discussion following Helgeson's (1981) paper outlining the HKF model, put the situation clearly. The "alternative competitive formulations" referred to are the two approaches mentioned above.

> As I see it, the domain in which alternative competitive formulations are more-or-less equally effective is roughly the region of water density greater than 0.7 to 0.8. Clearly, when you get water densities much less than 0.7, you're going into the domain where strong electrolytes become weak electrolytes. I fully understand Dr. Helgeson's reasons for wanting to maintain a single type of treatment continuously across that boundary. On the other hand, there are advantages, if you're working primarily in this relatively high dielectric – high density domain, of dealing with the materials as if they were completely dissociated. Again, I emphasize that the amount of dissociation in that territory is at least somewhat arbitrary.

This pretty well sums up the present situation. In problem areas dealing with concentrated solutions at or near the Earth's surface, that is, in the high dielectric constant, high density domain, the ion-interaction or Pitzer equation approach is dominant. In problem areas dealing with hydrothermal, metamorphic or aqueous solutions of igneous origin, the ion-pairing, species identification approach is dominant, and at the moment the HKF model is the most widely used. Each approach is weak in the other's domain. Pitzer equations do not yet have many parameters available for high temperatures and pressures, and although ion-pairing and B-dot activity coefficients can be used for concentrated, low temperature solutions, the results can be disastrous (Figure 15.6).

It is interesting to note that considerations of ion-pairs versus complexes are problems in physical chemistry, not in thermodynamics. Whether there are ion-pairs or complexes, one uses the available data on equilibrium constants or ion-interaction constants and makes the necessary calculations. In fact, in many geochemical discussions, no distinction is made between ion-pairs and complexes. All species are commonly referred to as complexes – if an equilibrium constant or an ion-interaction constant is available, it doesn't matter what it is called.

15.7 The Pitzer equations

In the 1970s, Kenneth Pitzer and his associates developed a theoretical model for the excess Gibbs energy of electrolyte solutions which combines the Debye–Hückel equation with additional terms in the form of a virial equation. This has proven to be extraordinarily successful at fitting the behavior of both single- and mixed-salt solutions to high concentrations. Summary articles outlining this approach, with references to the primary sources, are Pitzer (1981, 1987, 1991), and Pitzer (1995) presents a comprehensive treatment of thermodynamics which integrates this approach with other subjects.

Much of the theory underlying the Pitzer equations was developed for gases and extended to electrolyte solutions largely by Joseph Mayer (Mayer and Mayer, 1940; Mayer, 1950), and especially McMillan and Mayer (1945). For a brief summary see McQuarrie (2000, Chapter 15) and for a comprehensive summary see Friedman, (1962). If you do consult these references, your knowledge of statistical mechanics had better be pretty good. Mazo and Mou (1991) describe the technical details in McMillan and Mayer as “rather intricate.”

The model relates the excess total Gibbs energy of a system to an equation similar in principle to the virial equation we saw for gases in §13.5, in which the first term is not the ideal gas expression, but a simplified form of the Debye–Hückel equation. The general equation used by Pitzer (from which many others are derived by differentiation) is

$$\frac{\mathbf{G}^{\mathrm{EX}}}{\mathrm{kg_w}RT} = f(I) + \sum_i \sum_j \lambda_{ij}\, m_i m_j + \sum_i \sum_j \sum_k \mu_{ijk}\, m_i m_j m_k \tag{15.31}$$

where $\mathrm{kg_w}$ is the number of kilograms of water. This means we consider the total Gibbs energy of a system containing 1 kg of water, which in turn means all concentration will be molalities. $\mathbf{G}^{\mathrm{EX}}/\mathrm{kg_w}$ is not the same as G^{EX}.

The first virial coefficient $f(I)$ is some function of the ionic strength and is not 0 as it would be for an ideal solution, but is in fact a version of the Debye–Hückel equation, which represents departure from ideality in very dilute solutions. The following term is a function of the interactions of all pairs of ions, and the third term a function of the interactions of ions taken three at a time. The second coefficient, λ_{ij}, is a function of ionic strength, but the third coefficient μ_{ijk} is considered to be independent of ionic strength and equals zero if i, j, and k are all anions or cations. Later extensions to the model published by Pitzer and co-workers allow for an ionic strength dependence to the third coefficient. Pitzer (1987) and Harvie and Weare (1980) note that higher virial coefficients are required only for extremely concentrated solutions, so the series is stopped at the third coefficient.

To relate this to activity coefficients we know, from (10.53) and (10.54), that the partial derivatives of $\mathbf{G}^{\text{EX}}$ in (15.31) are

$$\begin{aligned}\left(\frac{\partial \mathbf{G}^{\text{EX}}/(\text{kg}_{\text{w}}RT)}{\partial m_i}\right)_{P,T,n_{\text{H}_2\text{O}}} &= \mu_i^{\text{EX}}/RT \\ &= \ln \gamma_i\end{aligned}$$

and

$$\phi - 1 = -\frac{(\partial \mathbf{G}^{\text{EX}}/\partial \text{kg}_{\text{w}})_{n_i}}{RT\sum_i m_i}$$

Equation (15.31) can therefore be rewritten in terms of activity coefficients for specific ions and the osmotic coefficient of the solvent by forming the appropriate derivatives of the virial terms on the right hand side. The final form taken by these terms was a subject of both theoretical considerations and empirical trial-and-error, and is described in Pitzer (1973). Omitting a great deal of such development, the final working equations are as follows, using the notation of Harvie et al. (1984).

The equations for a binary electrolyte solution are relatively simple. For mixed electrolytes they get a bit more complicated. We look first at the equations for a binary electrolyte.

15.7.1 Binary electrolytes *MX*

First, the parametric equation for excess Gibbs energy is

$$\begin{aligned}\frac{\mathbf{G}^{\text{EX}}}{\text{kg}_{\text{w}}RT} &= -A^{\phi}\left(\frac{4I}{b}\right)\ln(1+b\sqrt{I}) \\ &\quad + 2\nu_M\nu_X\left[m^2 B_{MX} + m^3\nu_M z_M C_{MX}\right]\end{aligned} \tag{15.32}$$

where

$$B_{MX} = \beta^{(0)} + 2\beta^{(1)}_{MX}\left[1-(1+\alpha\sqrt{I})\exp(-\alpha\sqrt{I})\right]/\alpha^2 I$$

$$C_{MX} = C^{\phi}_{MX}/2\sqrt{|z_M z_X|}$$

The corresponding equations for the osmotic and activity coefficients are

Osmotic coefficient

$$\begin{aligned}\phi - 1 &= -|z_M z_X|A^{\phi}\frac{\sqrt{I}}{1+b\sqrt{I}} \\ &\quad + m\frac{2\nu_M\nu_X}{\nu}\left[\beta^{(0)}_{MX} + \beta^{(1)}_{MX}\exp(-\alpha\sqrt{I})\right] \\ &\quad + m^2\frac{2(\nu_M\nu_X)^{3/2}}{\nu}C^{\phi}_{MX}\end{aligned} \tag{15.33}$$

In these equations, $\nu = \nu_M + \nu_X$, where ν is the number of ions; z is the charge on the ion, I is the (stoichiometric) molal ionic strength, and other terms are explained below.

Stoichiometric mean ionic activity coefficient

$$\begin{aligned}\ln\gamma_{\pm} = {} & -|z_M z_X| A^{\phi}\left[\frac{\sqrt{I}}{1+b\sqrt{I}} + \frac{2}{b}\ln(1+b\sqrt{I})\right] \\ & + m\frac{2\nu_M\nu_X}{\nu}\left\{2\beta_{MX}^{(0)} + \frac{2\beta_{MX}^{(1)}}{\alpha^2 I}\left[1-\left(1+\alpha\sqrt{I}-\frac{\alpha^2 I}{2}\right)\exp(-\alpha\sqrt{I})\right]\right\} \\ & + \frac{3m^2}{2}\left[\frac{2(\nu_M\nu_X)^{3/2}}{\nu}C_{MX}^{\phi}\right]\end{aligned} \tag{15.34}$$

15.7.2 Mixed electrolytes

Osmotic coefficient

$$\begin{aligned}\phi - 1 = {} & \frac{2}{\sum_i m_i}\left[\frac{-A^{\phi}I^{3/2}}{1+bI^{1/2}} + \sum_{c=1}^{Nc}\sum_{a=1}^{Na} m_c m_a (B_{ca}^{\phi} + ZC_{ca})\right. \\ & + \sum_{c=1}^{Nc-1}\sum_{c'=c+1}^{Nc} m_c m_{c'}\left(\Phi_{cc'}^{\phi} + \sum_{a=1}^{Na} m_a \Psi_{cc'a}\right) \\ & + \sum_{a=1}^{Na-1}\sum_{a'=a+1}^{Na} m_a m_{a'}\left(\Phi_{aa'}^{\phi} + \sum_{c=1}^{Nc} m_c \Psi_{aa'c}\right) \\ & \left. + \sum_{n=1}^{Nn}\sum_{a=1}^{Na} m_n m_a \lambda_{na} + \sum_{n=1}^{Nn}\sum_{c=1}^{nc} m_n m_c \lambda_{nc}\right]\end{aligned} \tag{15.35}$$

The activity coefficient for water is then

$$\ln a_{\mathrm{H_2O}} = -\frac{18.0154}{1000}\left(\sum_{i}^{Nc+Na} m_i\right)\phi \tag{15.36}$$

Activity coefficient for cation X

$$\begin{aligned}\ln\gamma_X = {} & z_X^2 F + \sum_{c=1}^{Nc} m_c (2B_{cX} + ZC_{cX}) + \sum_{a=1}^{Na} m_a\left(2\Phi_{Xa} + \sum_{c=1}^{Nc} m_c \Psi_{Xac}\right) \\ & + \sum_{c=1}^{Nc-1}\sum_{c'=c+1}^{Nc} m_c m_{c'}\Psi_{cc'X} + |z_X|\sum_{c=1}^{Nc}\sum_{a=1}^{Na} m_c m_a C_{ca} + \sum_{n=1}^{Nn} m_n (2\lambda_{nX})\end{aligned} \tag{15.37}$$

Activity coefficient for anion M

$$\ln \gamma_M = z_M^2 F + \sum_{a=1}^{Na} m_a (2B_{Ma} + ZC_{Ma}) + \sum_{c=1}^{Nc} m_c \left(2\Phi_{Mc} + \sum_{a=1}^{Na} m_a \Psi_{Mca} \right)$$
$$+ \sum_{a=1}^{Na-1} \sum_{a'=a+1}^{Na} m_a m_{a'} \Psi_{Maa'} + |z_M| \sum_{c=1}^{Nc} \sum_{a=1}^{Na} m_c m_a C_{ca} + \sum_{n=1}^{Nn} m_n (2\lambda_{nM}) \quad (15.38)$$

Activity coefficients for neutral species N

$$\ln \gamma_N = \sum_{c=1}^{Nc} m_c (2\lambda_{Nc}) + \sum_{a=1}^{Na} m_a (2\lambda_{Na}) \quad (15.39)$$

In these equations, Nc is the number of cations in the solution, m_c is the molality of cation c; M is the cation for which the activity coefficient is being calculated with charge z_M; and c and c' refer to all cations and all cations other than M respectively. Na, m_a, z_X, X, a and a' represent anions in the same way, and N refers to neutral species. No explicit account of ion association is taken, so all molalities m_i used in the above expressions refer to the total analytic concentration of species i. The function Z is defined as

$$Z = \sum_{i}^{Nc+Na} |z_i| m_i$$

The Debye–Hückel parameter A^ϕ used in these equations is numerically slightly different from that in the usual D–H equation:

$$A^\phi = \frac{1}{3} \left(\frac{2\pi N_A \rho_0}{1000} \right)^{1/2} \frac{e^3}{(k\epsilon_0 T)^{\frac{3}{2}}} \quad (15.40)$$

where N_A is Avogadro's constant, e is the electronic charge, k is the Boltzmann constant, and ρ_0 and ϵ_0 are the density and dielectric constant of pure water.

F is a function summing the D–H equation and additional terms:

$$F = -A^\phi \left(\frac{I^{1/2}}{1 + bI^{1/2}} + \frac{2}{b} \ln(1 + bI^{1/2}) \right)$$
$$+ \sum_{c=1}^{Nc} \sum_{a=1}^{Na} m_c m_a B'_{ca} + \sum_{c=1}^{Nc-1} \sum_{c'=c+1}^{Nc} m_c m_{c'} \Phi'_{cc'} \quad (15.41)$$
$$+ \sum_{a=1}^{Na-1} \sum_{a'=a+1}^{Na} m_a m_{a'} \Phi'_{aa'}$$

At 25 °C, $A^\phi = 0.392$ and b ($= \mathring{a}B$ in D–H) is taken as 1.2. The quantities B and C are adjustable parameters derived from measured activity coefficients on single salt + water solutions. The parameters Φ and Ψ come from data on aqueous systems containing two salts. Referring to the original virial equation (15.31), B and Φ are combinations of the second virial coefficients λ_{ij}, and C and Ψ are combinations of the third coefficients μ_{ijk}. These parameters are defined explicitly below.

All coefficients B_{MX} in the Pitzer equations (15.35)–(15.37) and (15.41) vary with ionic strength I. For electrolytes of valence 1–1 and 1–2 (e.g., NaCl and Na_2SO_4) they are written in terms of two regression parameters specific to the electrolyte, $\beta^{(0)}$ and $\beta^{(1)}$, a parameter α which depends on the type of electrolyte (for 1–1, 1–2 and 2–1 salts, $\alpha = 2.0$), and the ionic strength:

$$B^{\phi}_{MX} = \beta^{(0)}_{MX} + \beta^{(1)}_{MX} e^{-\alpha I^{1/2}}$$

$$B_{MX} = \beta^{(0)}_{MX} + \beta^{(1)}_{MX} g(\alpha I^{1/2})$$

$$B'_{MX} = \beta^{(1)}_{MX} g'(\alpha I^{1/2})/I$$

The functions g and g' in the above expressions are:

$$g(x) = 2\,(1 - (1 + x)e^{-x})\,/x^2$$

$$g'(x) = -2\left(1 - (1 + x + \tfrac{1}{2}x^2)e^{-x}\right)/x^2$$

where $x = \alpha I^{1/2}$. The parameters $\beta^{(0)}_{MX}$ and $\beta^{(1)}_{MX}$ are derived from regression fits to osmotic coefficient data for aqueous solutions of single salts.

With electrolytes of higher valence such as 2–2 (e.g., $CaSO_4$), the B^{ϕ}_{MX} expressions are augmented with an additional term (which is required to account for increased ion interaction)

$$B^{\phi}_{MX} = \beta^{(0)}_{MX} + \beta^{(1)}_{MX} e^{-\alpha_1 I^{1/2}} + \beta^{(2)}_{MX} e^{-\alpha_2 I^{1/2}}$$

$$B_{MX} = \beta^{(0)}_{MX} + \beta^{(1)}_{MX} g(\alpha_1 I^{1/2}) + \beta^{(2)}_{MX} g(\alpha_2 I^{1/2})$$

$$B'_{MX} = \beta^{(1)}_{MX} g'(\alpha_1 I^{1/2})/I + \beta^{(2)}_{MX} g'(\alpha_2 I^{1/2})/I$$

For these higher-valence electrolytes $\alpha_1 = 1.4$ and $\alpha_2 = 12.0$.

The parameters C_{MX} in Equations (17.40)–(17.42) are related to tabulated parameters derived from data on aqueous single-salt systems

$$C_{MX} = \frac{C^{\phi}_{MX}}{2\sqrt{|z_M z_X|}}$$

The parameters Φ^{ϕ}_{ij} and Ψ_{ijk} are derived from two-salt data and correct for interactions between ions of the same sign. The Ψ_{ijk} parameters are assumed independent of ionic strength and apply when i and j are different anions and k is a cation, or when i and j are different cations and k is an anion. The second virial coefficient terms Φ^{ϕ}_{ij} are defined by

$$\Phi^{\phi}_{ij} = \Theta_{ij} + {}^{E}\Theta_{ij}(I) + I \cdot {}^{E}\Theta'_{ij}(I)$$

$$\Phi_{ij} = \Theta_{ij} + {}^{E}\Theta_{ij}(I)$$

$$\Phi'_{ij} = {}^{E}\Theta'_{ij}(I)$$

The terms ${}^E\Theta_{ij}(I)$ and ${}^E\Theta'_{ij}(I)$ are functions solely of the ionic strength, and account for mixing for unsymmetrical electrolytes when the ions i and j have differing amounts of charge of the same polarity ($+$ or $-$). These parameters are zero when i and j have the same charge. Harvie and Weare observed that including these terms significantly improved agreement with real data for complex systems such as Na–Ca–Cl–SO_4–H_2O. Equations used to calculate these terms are given by Pitzer (1975, 1987) and summarized by Harvie and Weare (1980); they do not require fitting of experimental data as do the other terms, but are purely theoretical. Pitzer (1987, Appendix A) shows these terms graphically as a function of I, which is useful for checking your calculations.

15.7.3 Determining Pitzer parameters

The parameters $\beta^{(0)}$, $\beta^{(1)}$, $\beta^{(2)}$ and C^{ϕ} are determined by regressing activity coefficient or osmotic coefficient data. As an example, we use activity coefficient data for $NiSO_4$, which was investigated by Reardon (1989), as part of a study of the system $NiSO_4$–H_2SO_4–H_2O. We get the data for $\ln\gamma_{\pm}$ for $NiSO_4$ from Robinson and Stokes (1958), and perform the regression in a spreadsheet, using the solver tool, as we did on page 442 for the Debye–Hückel parameters. The idea is not to promote the use of spreadsheets, but to show that the process is quite simple, despite the formidable appearance of the equations. Given a small amount of programming experience, many other approaches using freely available regression subroutines are possible.

In a binary electrolyte solution such as this one, terms containing λ, Θ, or ψ are zero, since these involve interactions with two dissimilar anions or cations. In most such cases, the $\beta^{(2)}$ parameter is unnecessary, because it is invoked to account for exceptionally strong ion–ion interactions. In fact, Pitzer shows that $\beta^{(2)}$ should approach $-K/2$ in the limit of infinite dilution, where K is the association constant for the ion-pair. The work of Harvie and Weare (1980), Eugster, Harvie and Weare (1980), and Harvie, Eugster and Weare (1982), who modeled solubility equilibria in the multicomponent oceanic salt system is considered a milestone in the application of the Pitzer equations, and the set of parameters in Harvie, Mller and Weare (1984) is considered a sort of standard for modeling of seawater evaporitic systems.

To use the model at higher temperatures, derivatives of these coefficients with respect to T and P are determined by fitting other kinds of data, such as density, heat of dilution, or heat capacity. A discussion of recent developments is Voight (2001).

Determination of Pitzer parameters

For a 2:2 electrolyte such as $NiSO_4$, Equation (15.7.1) simplifies to

$$\begin{aligned}\ln\gamma_{\pm} &= -4A^{\phi}\left(\frac{\sqrt{I}}{1+1.2\sqrt{I}}+\frac{2}{1.2}(1+1.2\sqrt{I})\right)\\ &\quad +2m\boldsymbol{\beta}^{(0)}+\left(\frac{4m^2g'(\alpha_1\sqrt{I})}{I}+2mg(\alpha_1\sqrt{I})\right)\boldsymbol{\beta}^{(1)} \qquad (15.42)\\ &\quad +\left(\frac{4m^2g'(\alpha_2\sqrt{I})}{I}+2mg(\alpha_2\sqrt{I})\right)\boldsymbol{\beta}^{(2)}+\frac{3}{2}m^2\boldsymbol{C}^{\phi}\end{aligned}$$

where the required regression parameters are in bold symbols.

The numbers in Table 15.4 are copied from a spreadsheet. As before, the solver tool is used to minimize the sum of the (difference)^2 column by adjusting the values in four cells holding the coefficient values. The results are

	Table 15.4	Reardon (1989)
$\beta^{(0)}$	0.1658	0.1594
$\beta^{(1)}$	2.892	2.926
$\beta^{(2)}$	−48.97	−42.76
C^{ϕ}	0.0289	0.0406

The differences between these results and those of Reardon (1989) are due to the fact that he used different (but very similar) values for the activity coefficients. These results fit the measured data very closely, as shown in Figure 15.7. This is perhaps not surprising as there are four fit coefficients for this fairly simple curve. The significance, however, is that these coefficients can be used for the same Ni^{2+}-SO_4^{2-} interactions in multicomponent systems in which all the other two- and three-ion interaction coefficients have been determined.

If you do use a spreadsheet, using named cells for constants and coefficients makes complicated equations much more readable.

15.7.4 Gypsum solubility

To illustrate what some of these equations with double summations look like when expanded in a real case, consider the problem of calculating the solubility of gypsum in solutions of NaCl. To calculate this you need, among other

Table 15.4 *Spreadsheet values for the regression of* $ln\gamma_\pm$ *values for* $NiSO_4$.

$m_{\mathrm{NiSO_4}}$	I	$\ln\gamma_\pm$	$g(\alpha_1\sqrt{I})$	$g'(\alpha_1\sqrt{I})$	$g(\alpha_2\sqrt{I})$	$g'(\alpha_2\sqrt{I})$	Equation (15.42)	(difference)^2
0.1	0.4	−1.8971	0.5668	−0.1542	0.0345	−0.0340	−1.8944	7.130E − 06
0.2	0.8	−2.2537	0.4542	−0.1683	0.0173	−0.0173	−2.2554	2.859E − 06
0.3	1.2	−2.4757	0.3855	−0.1697	0.0115	−0.0115	−2.4790	1.084E − 05
0.4	1.6	−2.6408	0.3370	−0.1668	0.0086	−0.00868	−2.6424	2.396E − 06
0.5	2.0	−2.7693	0.3002	−0.1621	0.0069	−0.0069	−2.7709	2.572E − 06
0.6	2.4	−2.8788	0.2711	−0.1568	0.0057	−0.0057	−2.8762	6.485E − 06
0.7	2.8	−2.9661	0.2473	−0.1513	0.0049	−0.0049	−2.9644	2.909E − 06
0.8	3.2	−3.0407	0.2275	−0.1458	0.0043	−0.0043	−3.0392	2.169E − 06
0.9	3.6	−3.1055	0.2106	−0.1404	0.0038	−0.0038	−3.1031	5.965E − 06
1.0	4.0	−3.1582	0.1961	−0.1353	0.0034	−0.0034	−3.1577	2.830E − 07
1.2	4.8	−3.2441	0.1723	−0.1257	0.0028	−0.0028	−3.2439	6.289E − 08
1.4	5.6	−3.3022	0.1536	−0.1171	0.0024	−0.0024	−3.3048	6.729E − 06
1.6	6.4	−3.3438	0.1384	−0.1095	0.0021	−0.0021	−3.3448	9.664E − 07
1.8	7.2	−3.3667	0.1259	−0.1026	0.0019	−0.0019	−3.3669	4.0230E − 08
2.0	8.0	−3.3726	0.1154	−0.0964	0.0017	−0.0017	−3.3734	7.799E − 07
2.5	10	−3.3326	0.0954	−0.0834	0.0013	−0.0013	−3.3314	1.330E − 06
							sum of squares ==>	5.352E-05

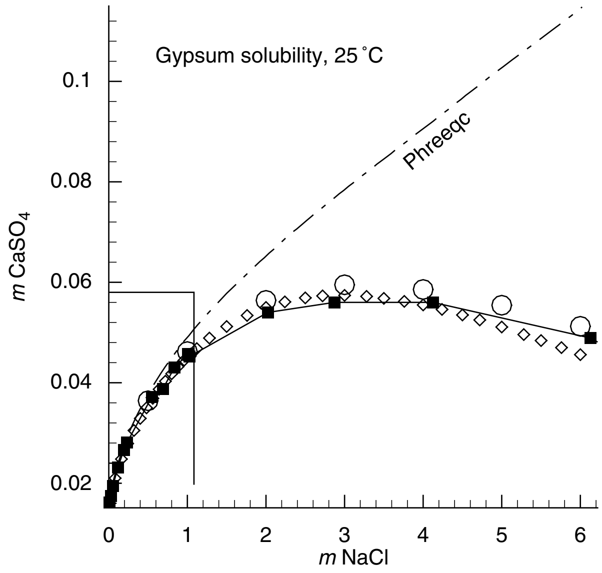


Figure 15.6 The solubility of gypsum in NaCl solutions. Solid squares: data from Marshall and Slusher (1966). Diamonds: calculated by program PHRQPITZ. Circles: calculated by program GYPSUM (see Appendix E). Dash-dot line: calculated by program PHREEQC.

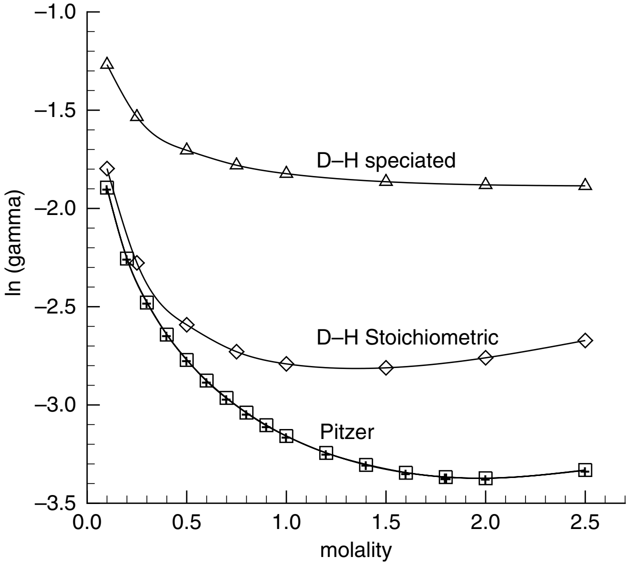


Figure 15.7 Calculated and measured values of $\ln \gamma_\pm$ for $NiSO_4$. + symbols: data from Robinson and Stokes (1959). Squares: calculated from Equation (15.42). D–H curves discussed in the text.

things, the activity coefficient of Ca^{2+}. In a solution with Na^+, Cl^-, Ca^{2+}, and SO_4^{2-} ions, Equation (15.38) becomes

$$\begin{aligned}\ln \gamma_{\mathrm{Ca}^{2+}} = {} & 4F + m_{\mathrm{Cl}}(2B_{\mathrm{CaCl}} + ZC_{\mathrm{CaCl}}) + m_{\mathrm{SO_4}}(2B_{\mathrm{CaSO_4}} + ZC_{\mathrm{CaSO_4}}) \\ & + m_{\mathrm{Na}}(2\Phi_{\mathrm{CaNa}} + m_{\mathrm{Cl}}\Psi_{\mathrm{CaNaCl}} + m_{\mathrm{SO_4}}\Psi_{\mathrm{CaNaSO_4}}) + m_{\mathrm{Cl}}m_{\mathrm{SO_4}}\Psi_{\mathrm{CaClSO_4}} \\ & + 2(m_{\mathrm{Na}}m_{\mathrm{Cl}}C_{\mathrm{NaCl}} + m_{\mathrm{Na}}m_{\mathrm{SO_4}}C_{\mathrm{NaSO_4}} + m_{\mathrm{Ca}}m_{\mathrm{Cl}}C_{\mathrm{CaCl}} + m_{\mathrm{Ca}}m_{\mathrm{SO_4}}C_{\mathrm{CaSO_4}})\end{aligned} \tag{15.43}$$

The equation giving $\ln \gamma_{\mathrm{SO_4^{2-}}}$ (15.37) looks quite similar. The function F as defined in Equation (15.41) becomes

$$\begin{aligned}F = {} & -A^\phi\left(\frac{I^{1/2}}{1+bI^{1/2}} + \frac{2}{b}\ln(1+bI^{1/2})\right) \\ & + m_{\mathrm{Na}}m_{\mathrm{Cl}}B'_{\mathrm{NaCl}} + m_{\mathrm{Na}}m_{\mathrm{SO_4}}B'_{\mathrm{NaSO_4}} + m_{\mathrm{Ca}}m_{\mathrm{Cl}}B'_{\mathrm{CaCl}} + m_{\mathrm{Ca}}m_{\mathrm{SO_4}}B'_{\mathrm{CaSO_4}} \\ & + m_{\mathrm{Na}}m_{\mathrm{Ca}}\Phi'_{\mathrm{NaCa}} + m_{\mathrm{Cl}}m_{\mathrm{SO_4}}\Phi'_{\mathrm{ClSO_4}}\end{aligned} \tag{15.44}$$

A more complete development of these equations is shown in the Fortran 90 subroutines actually used to calculate this quantity, listed in Appendix E. The results from this subroutine, called GYPSUM, plus results from the USGS program PHRQPITZ and program PHREEQC, are shown in Figure 15.6. Program PHRQPITZ is a complete implementation of the Pitzer equations, while subroutine GYPSUM was written specifically to illustrate this particular problem, with an effort to keep the code as clear as possible. Both reproduce the data very well. Small differences may be due to slightly different values of the parameters. Program PHREEQC, on the other hand, is an implementation of the ion-pairing model, and embodies the Debye–Hückel equation (15.28). The results from PHREEQC are obviously useless above about 1 molal.

As a further illustration of the difference between the Pitzer and the ion-pairing approaches, Figure 15.7 shows some calculated values of $\ln \gamma_{\pm}$ for $NiSO_4$. Experimental data as compiled by Robinson and Stokes (1959) are reproduced perfectly by the Pitzer equations, with the parameters we derived on page 457. The other two curves are calculated using the "B-dot" method, Equation (15.28). In one case, we use stoichiometric ionic strength, in which all solutes are assumed to be completely dissociated. In the other, we use "true ionic strength" or "speciated ionic strength," in which the ionic strength is calculated after speciation of the solution. Speciation is discussed in more detail in Chapter 16, but the essence here is that we first calculate the concentration of the neutral ion pair $NiSO_4^{\circ}$, as well as many other species, and the ionic strength is calculated from the ions remaining after this process. Because there is quite a lot of $NiSO_4^{\circ}$, this reduces the ionic strength considerably, and hence $\ln \gamma_{\pm}$ becomes less negative. Data for this figure are shown in Table 15.5. As mentioned earlier, the fact that å and $\dot{B}$ values are determined from stoichiometric data such as the Robinson and Stokes data in Figure 15.7 and then are used in ion-pair speciation models such as implemented by PHREEQC and The Geochemist's Workbench™, is one of the problems with using the B-dot method.

Table 15.5 *Activity coefficients for $NiSO_4$ calculated from the Pitzer equations and from Equation (15.28), using the B-dot method.*

Molality	Stoichiometric ionic strength	Speciated ionic strength	D–H stoichiometric $\ln \gamma_{\pm}$	D–H speciated $\ln \gamma_{\pm}$	Pitzer $\ln \gamma_{\pm}$
0.1	0.4	0.241	−1.797	−1.268	−1.894
0.3	1.0	0.538	−2.277	−1.534	−2.479
0.5	2.0	0.958	−2.592	−1.704	−2.771
0.8	3.0	1.315	−2.729	−1.780	−3.039
1.0	4.0	1.627	−2.791	−1.823	−3.158
1.5	6.0	2.158	−2.811	−1.864	—
2.0	8.0	2.599	−2.760	−1.880	−3.373
2.5	10.0	2.978	−2.672	−1.885	−3.331

		Debye–Hückel parameters used by PHREEQC			
A	B	$\dot{B}$	å $NiSO_4$	å Ni^{2+}	å SO_4^{2-}
0.5114	0.3288	0.041	3	6	4

15.8 The HKF model for aqueous electrolytes

The other major contribution to the systematization of our knowledge of aqueous electrolyte solutions at elevated temperatures and pressures takes a completely different approach. This was presented in a series of four papers by H.C. Helgeson and co-workers between 1974 and 1981, with fairly extensive modifications added by Tanger and Helgeson (1988). We present here an outline of this model, with some explanation and comparison with the Pitzer model. We refer to it as the HKF or revised HKF model, after the three authors of Part IV of the series of papers just mentioned, Helgeson, Kirkham, and Flowers.

15.8.1 Overall structure of the HKF model

The HKF model is semiempirical, in the sense that it uses a number of empirical parameters within a framework suggested by fundamental physics and thermodynamics. The variation of the Gibbs energy of individual ions with T, P, and composition can be represented by writing the total differential of the (partial molar) Gibbs energy of the jth ion, giving

$$d\overline{G}_j = -\overline{S}_j dT + \overline{V}_j dP + \left(\frac{\partial \mu_j}{\partial dn_j}\right)_{T,P} dn_j$$

where j is an ion in an aqueous solution of any composition. Changes in $\overline{G}_j$ due to changes in T, P or n_j are found by integrating this equation,

$$\int_{T_\mathrm{r},P_\mathrm{r},n_0}^{T,P,n} d\overline{G}_j = -\int_{T_\mathrm{r}}^{T} \overline{S}_{j,P_\mathrm{r},n_0} dT + \int_{P_\mathrm{r}}^{P} \overline{V}_{j,T,n_0} dP + \int_{n_0}^{n} d\mu_{j,T,P} \tag{15.45}$$

This is illustrated in Figure 15.8. The first two integrals on the right hand side take place at concentration n_0. If we equate this with the standard state (ideal one molal solution), then $\overline{S}_j$ and $\overline{V}_j$ become $\overline{S}_j^\circ$ and $\overline{V}_j^\circ$ respectively. The third integral takes care of departures from standard state conditions (change of composition) at T and P. Integration of $\int_{T_\mathrm{r}}^{T} \overline{S}_j^\circ\, dT$ requires knowledge of $\overline{C_{P}}_j^\circ$, and integration of $\int_{n_0}^{n} d\mu_j\ (= \mu_j - \mu_j^\circ = RT \ln a_j)$ requires knowledge of γ_j, so that the minimum information needed to know how the Gibbs energy of an ion j varies with T, P, and composition is how $\overline{C_{P}}_j^\circ$ and $\overline{V}_j^\circ$ vary with T and P respectively, and how γ_j varies with composition.

15.8.2 The Born functions

The Born function was introduced in §6.2.4. It is qualitatively accurate but quantitatively inaccurate. However the fact that it succeeds as well as it does suggests that it contains a large part of the truth, and might serve as the basis of a more satisfactory model. In fact it serves as the basis of the HKF model, which uses the Born function to provide the Gibbs energy of solvation (or hydration in

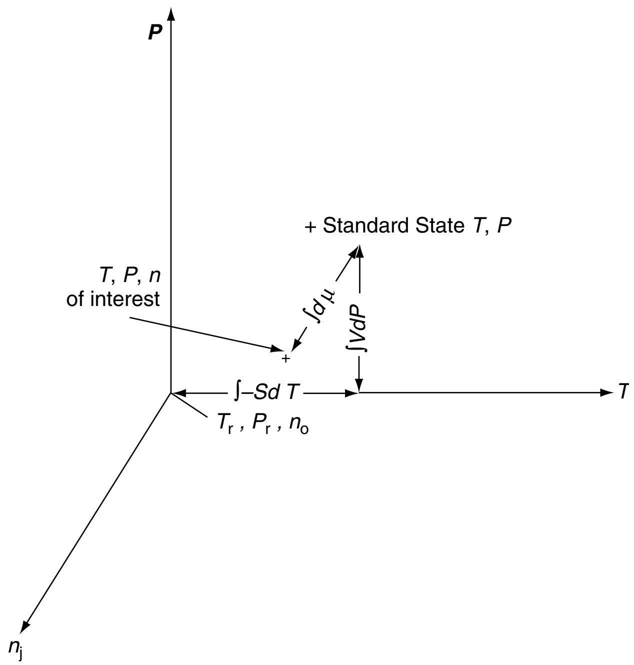


Figure 15.8 Representation of the integration of $d\overline{G}_j$ from reference conditions to conditions of interest, where j is an ion in an aqueous solution of any composition. The path from T_r, P_r, n_0 to T, P, n is represented by Equation (15.45).

aqueous solvent) in a strict sense (not including the energy effects of disrupting the water structure in the neighborhood of the hydrated ion). In other words, the Born function is taken as providing $\Delta\overline{G}$ for the process

$$\left\{\begin{array}{c} j \text{ in water with disrupted local} \\ \text{structure but before} \\ \text{formation of inner sphere} \\ \text{of strongly attached water molecules} \end{array}\right\} \rightarrow \left\{\begin{array}{c} j \text{ in water with inner} \\ \text{hydration sphere formed,} \\ \text{plus local disrupted} \\ \text{structure} \end{array}\right\}$$

A number of theoretical difficulties in equating the Born function with this process are believed to be accommodated in the $r_{e,j}$ parameter, which in the Born model is the ion radius, but in the HKF model is an adjustable parameter called the *effective* ionic radius. The $r_{e,j}$ parameters were originally related to crystallographic ionic radii $(r_{x,j})$ and ionic charge z_j in a simple linear fashion in the HKF model, and were independent of T and P

$$\begin{array}{ll} r_{e,j} = r_{x,j} + 0.94|z_j| & \text{for cations} \\ r_{e,j} = r_{x,j} & \text{for anions} \end{array}$$

In the revised model, $r_{e,j}$ became a function of T and P (Tanger and Helgeson, 1988), and is discussed below. An empirical term having several adjustable parameters to account for all other effects of adding an ion to water is then added to complete the model.

The Born function for individual ions is calculable theoretically, but as properties of individual ions are not measurable, it could never be tested experimentally. Two conventions are required to get around this problem. First, the properties of anions are taken to be equivalent to those of their corresponding acids, so that for instance the properties of Cl^- are equivalent to those of HCl(aq). Therefore we define a "conventional" electrostatic Born parameter as

$$\begin{aligned}\omega_j &= \frac{N_A(z_j e)^2}{2r_{e,j}} - \left(\frac{N_A(z_{H^+}e)^2}{2r_{e,H^+}}\right)\cdot z_j \\ &= \frac{N_A(z_j e)^2}{2r_{e,j}} - 0.5387 z_j \end{aligned} \tag{15.46}$$

where 0.5387 is the value of $[N_A(z_j e)^2]/(2r_{e,j})$ for the H^+ ion (Helgeson and Kirkham, 1976). The Born function then becomes

$$\Delta\overline{G}^\circ_{s,j} = \omega_j\left(\frac{1}{\epsilon} - 1\right) \tag{15.47}$$

which is the difference between the $\Delta\overline{G}$ of solvation of ion j (j in vacuum $\rightarrow j$ in water) and the $\Delta\overline{G}$ of solvation of H^+ (H^+ in vacuum $\rightarrow H^+$ in water). As described above, in the HKF model, this is equated with the $\Delta\overline{G}$ of formation of the inner hydration sphere only. This is sometimes called the *conventional* $\Delta\overline{G}$ of solvation, although since it is the only one we can deal with in a practical way, it is more often just called the $\Delta\overline{G}$ of solvation.

The second convention required is that the properties of individual ions are related to the properties of electrolytes by the additivity rule

$$\overline{\Xi}^\circ_k = \sum_j \nu_{j,k}\overline{\Xi}^\circ_j \tag{15.48}$$

where $\overline{\Xi}$ is any partial molar property and $\nu_{j,k}$ represents the stoichiometry of the jth ion in the kth electrolyte. An example would be $\overline{V}^\circ_{CaCl_2} = \overline{V}^\circ_{Ca^{2+}} + 2\overline{V}^\circ_{Cl^-}$. Thus for example, because the properties ($\overline{\Xi}$) of the Cl^- ion are identified with those of HCl they are "known," and $\overline{\Xi}$ for the Ca^{2+} ion is then obtained from

$$\overline{\Xi}_{Ca^{2+}} = \overline{\Xi}_{CaCl_2} - 2\overline{\Xi}_{Cl^-}$$

In summary, because properties are additive, and the properties of anions are known by convention, then all ionic properties are knowable, in the conventional sense.

It follows from the additivity convention that we can write an analogous Born function for aqueous electrolytes, k, as

$$\Delta\overline{G}^\circ_{s,k} = \omega_k\left(\frac{1}{\epsilon} - 1\right)$$

where

$$\omega_k = \sum_j \nu_{j,k}\omega_j$$

In other words, ω_k for electrolytes is simply the sum of its ω_j terms, e.g.,

$$\omega_{\mathrm{CaCl}_2} = \omega_{\mathrm{Ca}^{2+}} + 2\omega_{\mathrm{Cl}^-}$$

(Note that in this summation, the properties of the hydrogen ion cancel out.) It follows too that the fit parameters in the model to be described are also additive, so that the procedure is to develop the fit parameters for data on electrolytes, then to split them into additive parameters for individual ions. In the following equations, if there is no j or k subscript, the equation applies to both, as long as the appropriate parameters are used.

15.8.3 The $r_{e,j}$ term

The (conventional) entropy and volume changes of solvation according to the Born model are then derived by differentiation, which in the original HKF model was a simple procedure because $r_{e,j}$ was independent of T, P, which meant that ω_j was also independent of T, P. Thus in the original model,

$$\begin{aligned}\Delta\overline{S}^\circ_s &= -\left(\frac{\partial\Delta\overline{G}^\circ_s}{\partial T}\right)\\ &= -\omega\frac{\partial}{\partial T}\left(\frac{1}{\epsilon}-1\right)\\ &= \omega Y\end{aligned} \tag{15.49}$$

and

$$\begin{aligned}\Delta\overline{V}^\circ_s &= \left(\frac{\partial\Delta\overline{G}^\circ_s}{\partial P}\right)\\ &= \omega\frac{\partial}{\partial P}\left(\frac{1}{\epsilon}-1\right)\\ &= -\omega Q\end{aligned} \tag{15.50}$$

and

$$\begin{aligned}\Delta\overline{C_{P_s}}^\circ &= T\left(\frac{\partial\Delta\overline{S}^\circ_s}{\partial T}\right)\\ &= \omega T\frac{\partial Y}{\partial T}\\ &= \omega T X\end{aligned} \tag{15.51}$$

where Y, Q, and X are Born coefficients, i.e., temperature and pressure derivatives of the dielectric constant of water, derived and calculated by Helgeson and Kirkham (1974a). However, in the revised model, $r_{e,j}$ becomes a rather complex function of T and P, giving rather more complex expressions for the solvation terms, containing partial derivatives of ω_j. It is helpful in seeing the overall structure of this model to look first at the simple forms of the solvation contributions (above), and then to regard the more complex forms as unfortunate complications.

In the revised model,

$$\left.\begin{array}{ll} r_{e,j} = r_{x,j} + |z_j|(0.94 + g) & \text{for cations} \\ r_{e,j} = r_{x,j} + |z_j|g & \text{for anions} \end{array}\right\} \tag{15.52}$$

where, as before, z_j is the ionic charge and $r_{x,j}$ is the crystallographic ionic radius. The added term g is a complex function of T and P obtained by regression of volume and heat capacity data for aqueous NaCl. This results of course in more complex expressions for the solvation terms, which are (Tanger and Helgeson 1988, Appendix G)

$$\begin{aligned} \Delta\overline{S}^\circ_s &= -\left(\frac{\partial\Delta\overline{G}^\circ_s}{\partial T}\right)_P \\ &= \omega Y - \left(\frac{1}{\epsilon} - 1\right)\left(\frac{\partial\omega}{\partial T}\right)_P \end{aligned} \tag{15.53}$$

$$\begin{aligned} \Delta\overline{V}^\circ_s &= \left(\frac{\partial\Delta\overline{G}^\circ_s}{\partial P}\right)_T \\ &= -\omega Q + \left(\frac{1}{\epsilon} - 1\right)\left(\frac{\partial\omega}{\partial P}\right)_T \end{aligned} \tag{15.54}$$

and

$$\begin{aligned} \Delta\overline{C_{P_s}}^\circ &= T\left(\frac{\partial\Delta\overline{S}^\circ_s}{pT}\right) \\ &= \omega TX + 2TY\left(\frac{\partial\omega}{\partial T}\right)_P - T\left(\frac{1}{\epsilon} - 1\right)\left(\frac{\partial^2\omega}{\partial T^2}\right)_P \end{aligned} \tag{15.55}$$

Note that in regions where $g = 0$ (essentially where $T < 150\,^\circ\mathrm{C}$ or $P > 2000$ bars), the T and P derivatives of ω in these expressions become zero (because $r_{e,j}$ is independent of T, P when $g = 0$), and the solvation terms resume their original fairly simple forms.

The $r_{e,j}$ term in the HKF model essentially takes the place of the å term in the D–H model, and because it is different for each ion, the overall model suffers from the problem mentioned in §15.4, i.e., that dG_{solution} is not an exact differential. However, the authors consider that the error introduced is acceptable in view of other sources of uncertainty.

15.8.4 The empirical part

After defining the Born function as described above, comparison of experimental values of $\Delta\overline{V}^\circ$ and $\Delta\overline{C_P}^\circ$ with calculated values of $\Delta\overline{V}^\circ_s$ and $\Delta\overline{C_P}^\circ_s$ showed that the discrepancies could be fitted with functions of the form

$$\Delta\overline{V}^\circ_n = a_1 + a_2 f(P) + a_3 f_1(T) + a_4 f(P) f_1(T) \tag{15.56}$$

and

$$\Delta\overline{C_P}^\circ_n = c_1 + c_2 f_2(T) \tag{15.57}$$

where subscript n stands for nonsolvation, and

$$f_1(T) = 1/(T-\Theta)$$
$$f_2(T) = 1/(T-\Theta)^2$$
$$f(P) = 1/(\Psi + P)$$

In the original HKF model, Θ was a fit parameter for each ion having values usually ranging from about 200 to 260 K. Studies of supercooled water reviewed by Angell (1982, 1983; references in Tanger and Helgeson, 1988) however show that 228 ± 3 K is a singular temperature at which several properties approach $\pm\infty$, and in the revised model Θ takes on the fixed value of 228 K. The Ψ parameter is also fixed at 2600 bars.

15.8.5 Expressions for $\Delta\overline{V}^\circ$ and $\Delta\overline{C_P}^\circ$

Combining the Born and empirical parts of the model gives

$$\begin{aligned}\overline{V}^\circ &= \Delta\overline{V}^\circ_n + \Delta\overline{V}^\circ_s \\ &= \underbrace{a_1 + \frac{a_2}{\Psi + P} + \frac{a_3}{T-\Theta} + \frac{a_4}{(\Psi+P)(T-\Theta)}}_{\text{nonsolvation part}} \\ &\quad \underbrace{-\,\omega Q + \left(\frac{1}{\epsilon} - 1\right)\left(\frac{\partial\omega}{\partial P}\right)_T}_{\text{solvation part}}\end{aligned} \tag{15.58}$$

for the (conventional) standard partial molar volume of ion j or electrolyte k as a function of T and P, and

$$\begin{aligned}\overline{C_P}^\circ &= \Delta\overline{C_P}^\circ_n + \Delta\overline{C_P}^\circ_s \\ &= \underbrace{c_1 + \frac{c_2}{(T-\Theta)^2}}_{\text{nonsolvation part}} \\ &\quad \underbrace{+\,\omega T X + 2TY\left(\frac{\partial\omega}{\partial T}\right)_P - T\left(\frac{1}{\epsilon} - 1\right)\left(\frac{\partial^2\omega}{\partial T^2}\right)_P}_{\text{solvation part}}\end{aligned} \tag{15.59}$$

for the (conventional) standard partial molar heat capacity of ion j or electrolyte k as a function of T only.

No extra effort need be expended to determine the effect of pressure on $\Delta\overline{C_P}^{\circ}$ because this information is included in the expression for partial molar volume already obtained. That is, because

$$\left(\frac{\partial \overline{S}}{\partial P}\right) = -\left(\frac{\partial \overline{V}}{\partial T}\right) \tag{15.60}$$

it follows that

$$\left(\frac{\partial \overline{C_P}^{\circ}}{\partial P}\right) = -T\left(\frac{\partial^2 \overline{V}^{\circ}}{\partial T^2}\right) \tag{15.61}$$

This gives

$$\overline{C_P}^{\circ}{}_{P} - \overline{C_P}^{\circ}{}_{P_\mathrm{r}} = \int_{P_\mathrm{r}}^{P} -T\left(\frac{\partial^2 \overline{V}^{\circ}}{\partial T^2}\right) dP \tag{15.62}$$

or

$$\overline{C_P}^{\circ}{}_{P,T} = \overline{C_P}^{\circ}{}_{P_\mathrm{r},T} + \int_{P_\mathrm{r}}^{P} -T\left(\frac{\partial^2 \overline{V}^{\circ}}{\partial T^2}\right) dP \tag{15.63}$$

which on integration turns out to be

$$\begin{aligned}
\overline{C_P}^{\circ}{}_{P,T} &= \overline{C_P}^{\circ} \\
&= \underbrace{c_1 + \frac{c_2}{(T-\Theta)^2} - \left(\frac{2T}{(T-\Theta)^3}\right)\left[a_3(P-P_\mathrm{r}) + a_4 \ln\left(\frac{\Psi+P}{\Psi+P_\mathrm{r}}\right)\right]}_{\text{nonsolvation part}} \\
&\quad \underbrace{+\,\omega T X + 2TY\left(\frac{\partial \omega}{\partial T}\right)_P - T\left(\frac{1}{\epsilon}-1\right)\left(\frac{\partial^2 \omega}{\partial T^2}\right)_P}_{\text{solvation part}}
\end{aligned} \tag{15.64}$$

15.8.6 Expressions for $\Delta\overline{S}^{\circ}$, $\Delta\overline{H}^{\circ}$ and $\Delta\overline{G}^{\circ}$

Having expressions for the temperature and pressure effects on $\overline{C_P}^{\circ}$ and $\overline{V}^{\circ}$, straightforward, if somewhat lengthy, integration gives expressions for $\Delta\overline{S}^{\circ}$, $\Delta\overline{H}^{\circ}$, and $\Delta\overline{G}^{\circ}$, which can refer either to an ion j or an electrolyte k, depending on the fit parameters used in the expression. Thus

$$\begin{aligned}
\overline{S}^{\circ}_{P,T} - \overline{S}^{\circ}_{P_\mathrm{r},T_\mathrm{r}} &= \int_{T_\mathrm{r}}^{T} \frac{\overline{C_P}^{\circ}}{T} dT - \int_{P_\mathrm{r}}^{P} \left[\left(\frac{\partial \overline{V}^{\circ}}{\partial T}\right)_P\right]_T dP \\
&= c_1 \ln\left(\frac{T}{T_\mathrm{r}}\right) - \frac{c_2}{\Theta}\left[\left(\frac{1}{T-\Theta}\right) - \left(\frac{1}{T_\mathrm{r}-\Theta}\right) + \frac{1}{\Theta}\cdot \ln\left(\frac{T_\mathrm{r}(T-\Theta)}{T(T_\mathrm{r}-\Theta)}\right)\right]
\end{aligned}$$

$$
\begin{aligned}
&+\left(\frac{1}{T-\Theta}\right)^2\left[a_3(P-P_\mathrm{r})+a_4\,\ln\left(\frac{\Psi+P}{\Psi+P_\mathrm{r}}\right)\right]\\
&+\omega Y-\left(\frac{1}{\epsilon}-1\right)\left(\frac{\partial\omega}{\partial T}\right)_P-\omega_{P_\mathrm{r},T_\mathrm{r}}Y_{P_\mathrm{r},T_\mathrm{r}}
\end{aligned}
\tag{15.65}
$$

$$
\begin{aligned}
\overline{H}^\circ_{P,T}-\overline{H}^\circ_{P_\mathrm{r},T_\mathrm{r}}&=\int_{T_\mathrm{r}}^{T}\overline{C_P}^\circ dT+\int_{P_\mathrm{r}}^{P}\left[\overline{V}^\circ-T\left(\frac{\partial\overline{V}^\circ}{\partial T}\right)_P\right]_T dP\\
&=c_1(T-T_\mathrm{r})-c_2\left[\left(\frac{1}{T-\Theta}\right)-\left(\frac{1}{T_\mathrm{r}-\Theta}\right)\right]\\
&\quad+a_1(P-P_\mathrm{r})+a_2\ln\left(\frac{\Psi+P}{\Psi+P_\mathrm{r}}\right)\\
&\quad+\left(\frac{2T-\Theta}{(T-\Theta)^2}\right)\left[a_3(P-P_\mathrm{r})+a_4\,\ln\left(\frac{\Psi+P}{\Psi+P_\mathrm{r}}\right)\right]\\
&\quad+\omega\left(\frac{1}{\epsilon}-1\right)+\omega TY-T\left(\frac{1}{\epsilon}-1\right)\left(\frac{\partial\omega}{\partial T}\right)_P\\
&\quad-\omega_{P_\mathrm{r},T_\mathrm{r}}\left(\frac{1}{\epsilon_{P_\mathrm{r},T_\mathrm{r}}}-1\right)-\omega_{P_\mathrm{r},T_\mathrm{r}}T_\mathrm{r}Y_{P_\mathrm{r},T_\mathrm{r}}
\end{aligned}
\tag{15.66}
$$

$$
\begin{aligned}
\overline{G}^\circ_{P,T}-\overline{G}^\circ_{P_\mathrm{r},T_\mathrm{r}}&=-\overline{S}^\circ_{P_\mathrm{r},T_\mathrm{r}}(T-T_\mathrm{r})+\int_{T_\mathrm{r}}^{T}\overline{C_P}^\circ dT-\int_{T_\mathrm{r}}^{T}\frac{\overline{C_P}^\circ}{T}dT+\int_{P_\mathrm{r}}^{P}\overline{V}^\circ dP\\
&=-\overline{S}^\circ_{P_\mathrm{r},T_\mathrm{r}}(T-T_\mathrm{r})-c_1\left[T\ln\left(\frac{T}{T_\mathrm{r}}\right)-T+T_\mathrm{r}\right]\\
&\quad+a_1(P-P_\mathrm{r})+a_2\,\ln\left(\frac{\Psi+P}{\Psi+P_\mathrm{r}}\right)\\
&\quad-c_2\left(\left[\left(\frac{1}{T-\Theta}\right)-\left(\frac{1}{T_\mathrm{r}-\Theta}\right)\right]\left(\frac{\Theta-T}{\Theta}\right)-\frac{T}{\Theta^2}\ln\left(\frac{T_\mathrm{r}(T-\Theta)}{T(T_\mathrm{r}-\Theta)}\right)\right)\\
&\quad+\left(\frac{1}{T-\Theta}\right)\left[a_3(P-P_\mathrm{r})+a_4\,\ln\left(\frac{\Psi+P}{\Psi+P_\mathrm{r}}\right)\right]+\omega\left(\frac{1}{\epsilon}-1\right)\\
&\quad-\omega_{P_\mathrm{r},T_\mathrm{r}}\left(\frac{1}{\epsilon_{P_\mathrm{r},T_\mathrm{r}}}-1\right)+\omega_{P_\mathrm{r},T_\mathrm{r}}Y_{P_\mathrm{r},T_\mathrm{r}}(T-T_\mathrm{r})
\end{aligned}
\tag{15.67}
$$

At this point we have shown how the HKF model develops expressions for the standard state parameters $\overline{V}^\circ$ and $\overline{C_P}^\circ$ and hence $\overline{S}^\circ$, $\overline{H}^\circ$, and $\overline{G}^\circ$ at high temperatures and pressures. The standard state universally used is the ideal one molal solution, which means that these parameters refer to the properties of ions or electrolytes in infinitely dilute solutions. You might suppose that therefore they would not be of much use to geochemists interested in natural solutions, which are often quite concentrated, but you would be wrong. The standard state properties allow the calculation of the equilibrium constant for reactions involving ions at high T, P, and thus permit the general nature of many important processes to be understood, even in cases where activity coefficients are unknown. Of course for quantitative calculation of ionic concentrations and mass transfers in such cases, activity coefficients are also required.

15.8.7 Contributions of the solvation and nonsolvation terms

A striking feature of the partial molar volumes and heat capacities of aqueous electrolytes is their inverted-U shape as a function of temperature. Experimental data that cover a sufficiently large range of temperature invariably exhibit a maximum, generally somewhere between 50 and 100 °C. This was illustrated in Figures 10.6 and 10.12, which show data for the partial molar volume and heat capacity of NaCl. The existence of singular temperatures for water at −45 °C (228 K, Angell, 1982, 1983) and 374 °C (the critical temperature) makes it seem entirely reasonable that thermodynamic parameters of solutes in water should approach $\pm\infty$ at these limits, and therefore reasonable that they should exhibit extrema (or inflection points) between these temperatures.[2]

The revised HKF model is constructed such that the nonsolvation contribution to $\overline{V}^\circ$ and $\overline{C_P}^\circ$ dominates at low temperatures and becomes $-\infty$ at 228 K, and the solvation contribution dominates at high temperatures. The contributions of the solvation and nonsolvation parts of the partial molar volume of Na^+ are compared in Figure 15.9, and in Figure 15.10 the solvation and nonsolvation contributions to the partial molar heat capacity of HCl are shown as a function of temperature. Note how the shapes of the two contributions combine to give the inverted-U shape of the measured heat capacity. This illustrates quite nicely how the two contributions combine to produce a maximum, and it can easily be imagined how the shape of the combined curve is controlled by the fit parameters of the two contributions. Of course, the two contributions do not always cross in such a pedagogically convenient way.

In preparing Figures 15.9, 15.10 and 15.11, the original (pre-Tanger and Helgeson) equations were used. For example, the solvation contribution to $\overline{C_P}^\circ$ is Equation (15.52), and the nonsolvation contribution is

$$\overline{C_P}^\circ = c_1 + \frac{c_2}{(T - \Theta)}$$

which is the first term in Equation (15.64) except for the square exponent in the denominator. The calculations are simpler, and the differences from the later equations are not important in this context.

---

[2] Don’t confuse the properties of *component* NaCl in Figures 10.6 and 10.12 with the properties of *species* NaCl° (named “NaCl,aq”) in program SUPCRT92 (Johnson et al., 1992).

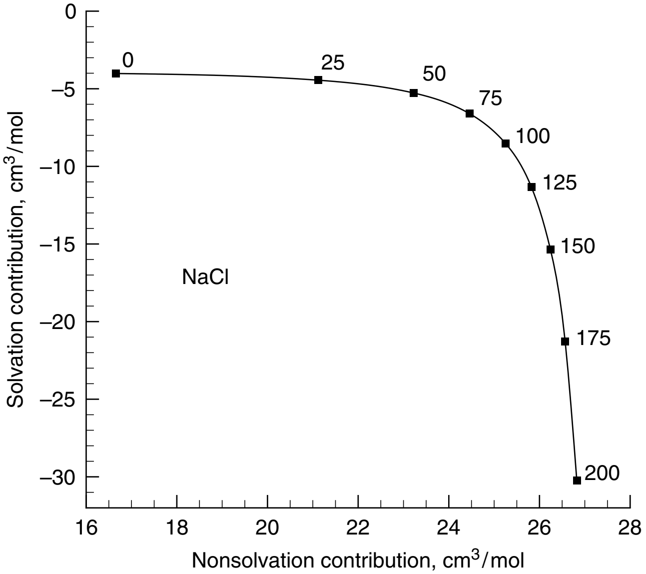


Figure 15.9 Solvation versus nonsolvation contribution to the standard partial molar volume of NaCl at saturation pressure, according to the HKF model. Labels are T °C.

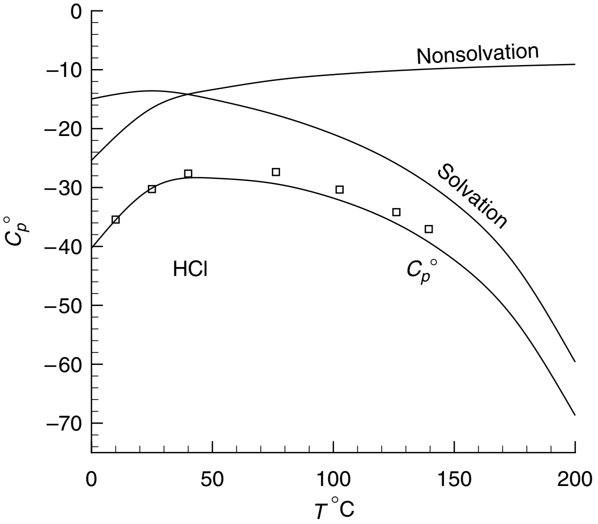


Figure 15.10 The standard partial molar heat capacity of aqueous HCl as a function of temperature. Squares are the experimental data of Tremaine et al. (1986), and the solvation and nonsolvation contributions, which add to the line fitting the data, are from the HKF model.

15.8.8 Equivalence of the properties of aqueous HCl and Cl$^-$

The availability of the HKF model makes simple a demonstration of a feature of single ion properties that we have explained previously. By convention (the convention that the properties of the hydrogen ion are zero) the properties of aqueous anions are taken to be those of the corresponding strong acid, unless there are complicating factors. This means, for example, that the properties of

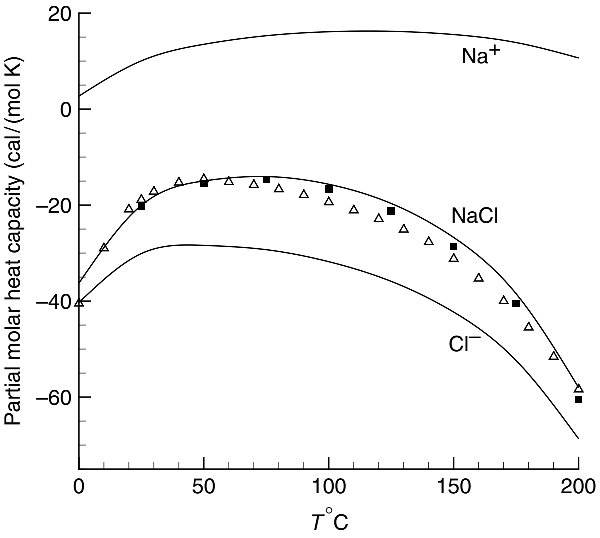


Figure 15.11 The standard partial molar heat capacities of aqueous NaCl, Na+, and Cl− as a function of temperature. Triangles: data from Gardner et al. (1969). Solid squares: data from Helgeson et al. (1984, Table 15).

the chloride ion are those of aqueous HCl, and the properties of the sulfate ion are those of aqueous H_2SO_4. The conventions involved in this were discussed in §15.3.1.

In Figure 15.11 we show the HKF representation of the partial molar heat capacities of Na^+, Cl^-, and NaCl, together with the experimental data. The curve shown for the chloride ion is in fact the same as that shown for HCl in Figure 15.10, because of course no data for the chloride ion itself exist. The heat capacity of aqueous NaCl is from Figure 10.12. The curve for the partial molar heat capacity of Na^+ is derived from the two lower curves from Equation (15.48), which in this case is

$$\begin{aligned}\overline{C_P}^{\circ}_{\mathrm{NaCl}} &= \overline{C_P}^{\circ}_{\mathrm{Na^+}} + \overline{C_P}^{\circ}_{\mathrm{Cl^-}}\\ &= \overline{C_P}^{\circ}_{\mathrm{Na^+}} + \overline{C_P}^{\circ}_{\mathrm{HCl}}\end{aligned}$$

This illustrates how “data” for Na^+ are derived.

15.9 Comparison of Pitzer and HKF models

These two models present the two main approaches to modeling solutions, which have been in evidence since the beginning, as discussed on page 304. Pitzer stands back, as it were, from the details of ionic interactions, and builds up an empirical model of complex solutions from data on the simpler binary systems of which they are composed. All aspects of ionic interaction are buried in the magnitudes of the parameters. This empirical model works very well at predicting the macroscopic properties of complex solutions, apparently because the form of the equations he has chosen to use suits the problem very well.

Helgeson, on the other hand, prefers to think about standard states, using standard thermodynamic relationships augmented by empirical fit parameters in an ion-pairing context to represent standard state parameters of aqueous species to high temperatures and pressures. The Pitzer equations are limited at present for the most part to ambient temperatures, but are capable of accurate representation of phase relationships and compositions up to the limit of component solubilities. The HKF equations are better suited to extrapolation, are used up to high temperatures and pressures, and deal with many components for which there are as yet no Pitzer parameters. However, being standard state parameters, the model relies on activity coefficients for application to real systems, and this is a continuing problem.

There are advantages and disadvantages to both approaches. The achievements of Harvie and Weare and others in modeling evaporitic brines with the Pitzer model is at the moment well beyond the range of the HKF model, but in providing standard thermodynamic parameters for individual ions to high temperatures and pressures the HKF model is reaction-oriented, and hence fits the way most geochemists think. It has therefore been more widely used at present.

16

Rock–water systems

16.1 Real problems

We have now completed our survey of the thermodynamic principles required to model natural systems. It only remains to gain practice in formulating problems involving natural systems in thermodynamic terms. Quite often, that is the hardest part. Once the problem is set up in terms of relevant reactions and components, the equations can be solved by anyone who has absorbed the previous chapters. However, choosing the appropriate components and setting up the relevant balanced reactions only comes from experience. In this chapter we explore a few situations that have been investigated by thermodynamic methods.

16.2 Is the sea saturated with calcium carbonate?

If you have ever been to Florida or the Bahamas, you may be aware that there are vast areas adjacent to the coasts where the sea bottom at shallow levels is a white mud, which turns out to be made of almost pure aragonite. Carbonate muds extend well out to the deep sea as well; in fact, a fairly large proportion of the sea bottom is composed of calcium carbonate. There are also countless calcitic atolls and reefs throughout the tropical zones of the world. Given this amount of contact between the sea and calcium carbonate, both calcite and aragonite, plus the fact that there are vigorous oceanic currents stirring things up constantly, plus the fact that things have not changed drastically for millions of years, you would think that there would be little doubt that the system consisting of the oceans plus their bottom sediments must have reached equilibrium by now. If these were the only factors involved, perhaps they would have, but the situation is quite a bit more complicated. Why would anyone want to know? Reactions involving carbonate in the oceans are fundamental to an understanding of the global CO_2 cycle, which in turn is linked to global warming and other things we would like to understand.

How do you tell if a solution is saturated?

To explore this problem further, we must first find out how to determine whether a solution is saturated, undersaturated, or supersaturated with a given mineral or compound. One answer would be to just observe the solution in contact with the mineral. If the mineral dissolves, the solution is undersaturated. If the mineral grows in size, the solution is supersaturated. If nothing happens, the solution is saturated – it is at equilibrium. This method and variations of it are used, but it is very difficult for a number of reasons. We would like to be able to predict the state of saturation for a sample of water without performing difficult experiments on it. We would like simply to determine the chemical composition of the solution and calculate theoretically the state of saturation.

In other words, we want a thermodynamic answer. Having just spent several chapters developing a method for determining which way a reaction will go, we should be able to put it to use here. The reaction could be written

$$\text{solid mineral} = \text{dissolved mineral} \tag{16.1}$$

If this reaction goes to the right, the solution is undersaturated. If it goes to the left, the solution is supersaturated. All we need to do is to determine the molar Gibbs energy of the dissolved mineral and compare it to the molar Gibbs energy of the pure mineral, and the question is answered.

16.2.1 Solubility products

But dissolution reactions that result in uncharged solutes such as $H_4SiO_4(aq)$ and $H_2CO_3(aq)$ are unusual. Most solutes are ionized to some extent, that is, they break up into charged particles, called ions. In other words, we write the dissolution reaction not as in (10.1), but as

$$\text{solid mineral} = \text{aqueous ions} \tag{16.2}$$

For example, calcium carbonate (calcite or aragonite), when it dissolves, breaks up into calcium and carbonate ions (Figure 16.1), written as

$$\mathrm{CaCO_3}(s) = \mathrm{Ca^{2+}} + \mathrm{CO_3^{2-}} \tag{16.3}$$

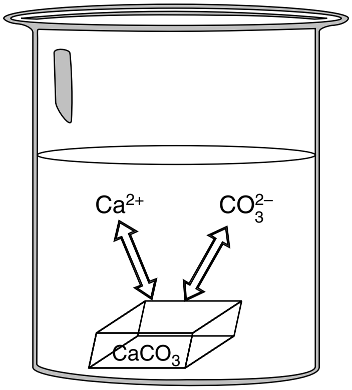


Figure 16.1 When calcite dissolves, the solute consists of electrically charged ions.

The equilibrium constant for reaction (16.3) can be found in our routine way:

$$\begin{aligned}\Delta_r G^\circ &= \Delta_f G^\circ_{\mathrm{Ca}^{2+}} + \Delta_f G^\circ_{\mathrm{CO}_3^{2-}} - \Delta_f G^\circ_{\mathrm{CaCO}_3(s)}\\ &= -553.58 + (-527.81) - (-1128.79)\\ &= 47.40\,\mathrm{kJ\,mol^{-1}}\end{aligned}$$

for calcite, or

$$\begin{aligned}&= -553.58 + (-527.81) - (-1127.75)\\ &= 46.36\,\mathrm{kJ\,mol^{-1}}\end{aligned}$$

for aragonite. This gives

$$\begin{aligned}\Delta_r G^\circ &= -RT\ln K\\ 47\,400 &= -(2.30259 \times 8.3145 \times 298.15)\log K\\ \log K &= -8.304\end{aligned}$$

for calcite, or -8.122 for aragonite.

This equilibrium constant is

$$K = \frac{a_{\mathrm{Ca}^{2+}}\, a_{\mathrm{CO}_3^{2-}}}{a_{\mathrm{CaCO}_3(s)}}$$

Now, if, as in Figure 16.2, we are dealing with the solubility of pure calcium carbonate its activity is 1.0, so the equilibrium constant becomes

$$K_{\mathrm{sp}} = a_{\mathrm{Ca}^{2+}}\, a_{\mathrm{CO}_3^{2-}}$$

and is called a *solubility product* constant, or just a solubility product. Whereas in the case of quartz solubility we found the equilibrium constant to be equal

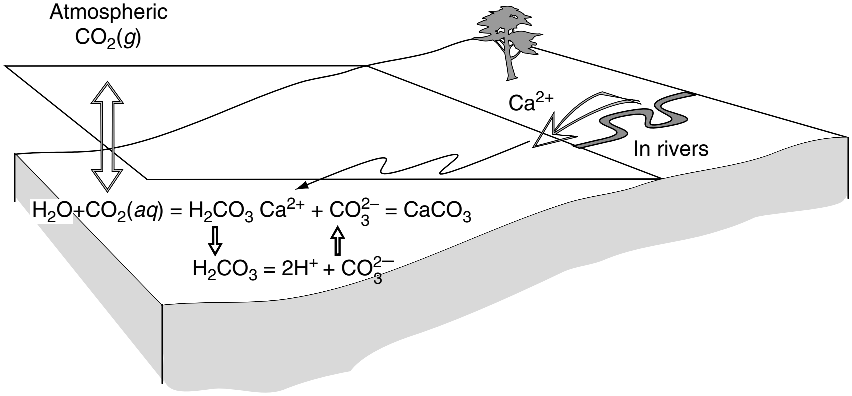


Figure 16.2 Calcite may precipitate in natural bodies of water, but the calcium and the carbonate may come from different sources.

to the solubility itself (§9.4.1), here we find the equilibrium constant to be equal to a product of two ion activities. Therefore, "the solubility of calcite" has a somewhat ambiguous meaning. If it refers to the concentration (activity) of calcium in solution, this obviously depends on how much carbonate ion is in solution, and vice versa. It is the *combination* of calcium ion and carbonate ion activities that determines whether calcite is over- or undersaturated in a solution. For example, suppose we have determined that the activity of CO_3^{2-} in a solution (say the one in Figure 16.1) is 10^{-5}. A solution having $a_{Ca^{2+}}\, a_{CO_3^{2-}} = 10^{-8.304}$ will be in equilibrium with calcite, so in this case the equilibrium activity of the calcium ion is $a_{Ca^{2+}} = 10^{-3.304}$. In a solution with $a_{CO_3^{2-}} = 10^{-6}$, the equilibrium value of $a_{Ca^{2+}}$ is $10^{-2.304}$, and so on.

Having now learned about the solubility product, please do not tack the "sp" subscript on to every equilibrium constant you calculate. There are equilibrium constants for many types of reactions. The solubility product is an equilibrium constant for a reaction having a solid mineral or compound on the left side and its constituent ions on the right.

16.2.2 IAP, K_{sp}, Ω, and SI

Of course natural solutions, such as seawater, are not necessarily at equilibrium. In Figure 16.2 we see a river carrying dissolved material, including calcium and carbonate ions, entering the sea. Carbonate ions are already there, because the sea is in contact with the atmosphere, which contains carbon dioxide, and when CO_2 dissolves it produces carbonate and bicarbonate ions. Because calcium and carbonate are being added, there may be a tendency for them to increase beyond the equilibrium value, and for calcite to precipitate as a result. The product of the calcium and carbonate ion activities *which are actually present in a solution*, regardless of any theory, is called the *ion activity product* (IAP) for that solution. It follows that when IAP > K_{sp}, calcite will precipitate, and when IAP < K_{sp} calcite will dissolve. The IAP/K_{sp} ratio is called Ω, and the logarithm of the ratio is called the *saturation index* (SI), so that when SI > 0 calcite precipitates, and when SI < 0 calcite dissolves (Table 16.1).

These relationships should be fairly intuitive, but if you have difficulty convincing yourself, consider Equation (9.8).

$$\Delta_r\mu = \Delta_r\mu^\circ + RT\ln Q \qquad [9.8]$$

Table 16.1 *Relations between IAP, K_{sp}, and SI.*

IAP, K_{sp}	Ω	SI $\left(= \log \frac{\text{IAP}}{K_{sp}}\right)$	Result
IAP < K_{sp}	< 1	negative	mineral dissolves
IAP > K_{sp}	> 1	positive	mineral precipitates
IAP = K_{sp}	1	0	equilibrium

In the case we are considering – calcite and its ions, reaction (16.3) – when $Q = K_{\text{sp}}$, $\Delta_r\mu = 0$. Therefore, if $Q = \text{IAP} > K_{\text{sp}}$, $\Delta_r\mu > 0$, and reaction (16.3) will go to the left (calcite precipitates). And if $Q = \text{IAP} < K_{\text{sp}}$, $\Delta_r\mu < 0$, reaction (16.3) will go to the right (calcite dissolves).

16.3 Determining the IAP – speciation

Now to answer our question as to whether the sea is saturated with calcite, we need only determine its IAP and compare it with K_{sp} for calcite. Easier said than done. Apart from some specialized electrochemical techniques, it is generally not possible to analyze a solution for the concentration of specific ions such as Ca^{2+} or CO_3^{2-}. Analyses are made for the *total* calcium, or the *total* carbonate in the solution. There are many kinds of ions and uncharged molecular species such as ion-pairs in a solution, and so the total calcium concentration would consist of the sum of the concentrations of all the ionic and molecular species containing Ca, and similarly for carbonate. Thus

$$m_{\text{Ca,total}} = m_{\text{Ca}^{2+}} + m_{\text{CaHCO}_3^+} + m_{\text{CaSO}_4^\circ} + \cdots$$

$$m_{\text{CO}_3\text{,total}} = m_{\text{CO}_3^{2-}} + m_{\text{HCO}_3^-} + m_{\text{H}_2\text{CO}_3} + \cdots$$

Here, CaSO_4° represents an electrically neutral species resulting from the joining of a Ca^{2+} ion and a SO_4^{2-} ion in solution. There are several more species in each of these summations in real seawater which we won't mention. But however many there are, the concentrations and activities of all of them can be calculated if for every species there is a known equilibrium constant relating it to other species and/or minerals, and if there is a suitable equation for calculating the activity coefficients (γ_{H}) of each of the species.

What it amounts to is the well-known fact that if in a set of equations you have the same number of equations as you have variables, it is possible to solve for every variable. Given the total concentrations of the various constituents of seawater (not only calcium and carbonate, but however many you are interested in), plus an equilibrium constant for each, plus the knowledge that the total positive charges must equal the total negative charges, it is always possible to achieve this goal. When more than a few ions are involved, the procedure can be carried out only on a computer, and many programs are now available for doing this.

It is also worth pointing out that the process of speciation is mathematically equivalent to finding the system configuration which minimizes the Gibbs energy. This can be shown formally, but if you think of reactions which have not equilibrated as being somewhere above the minimum on a U-shaped Gibbs energy versus progress variable curve (see Figure 18.8 for an example), and being at the minimum when at equilibrium, it should be fairly intuitive. When all possible reactions achieve their minimum Gibbs energy, then so does the system.

16.3.1 Speciation using a calculator

Let's start with a simple problem we can do "by hand." What are the species activities in a dilute solution of acetic acid? To calculate these activities, you need to know

- what ions there are in solution, and
- the equilibrium constants for the formation of these ions, at the temperature and pressure you are considering.

This is a simple case, in which there are only four ions: H^+, OH^-, HAc, and Ac^-, where we let HAc stand for $CH_3COOH(aq)$ and Ac^- for CH_3COO^-. There are two equilibrium constants,

$$a_{H^+} a_{OH^-} = a$$

$$\frac{a_{H^+} a_{Ac^-}}{a_{HAc}} = b$$

one charge balance,

$$m_{H^+} = m_{OH^-} + m_{Ac^-}$$

and one mass balance,

$$m_{HAc} + m_{Ac^-} = c$$

The equilibrium constants require activities, but we start with the assumption that all activity coefficients are 1.0. With four equations and four unknowns, a solution is possible. Some manipulation of the equations then results in

$$m_{H^+}^3 + b\,m_{H^+}^2 - (a + bc)\,m_{H^+} + ab = 0 \tag{16.4}$$

This is what the equation looks like with all activity coefficients equal to 1.0. The real version includes several activity coefficients, which of course are generally nonzero. This cubic equation has three roots, but only one is reasonable. With $a = 10^{-14}$, $b = 10^{-4.76}$, and $c = 0.1$, the results are

$$m_{H^+} = 10^{-2.88}$$

$$m_{OH^-} = 10^{-11.12}$$

$$m_{HAc} = 0.0987$$

$$m_{Ac^-} = 0.0013$$

To include activity coefficients, you need to calculate an ionic strength from the results of this calculation, calculate an activity coefficient for each ion from an appropriate expression of the Debye–Hückel equation, solve the exact form of Equation (16.4) again, and iterate until all the answers don't change. You can readily imagine that with a few more species this procedure would become impossible without a computer. In fact with more species there is no closed-form, exact solution like Equation (16.4), and the equations must be solved by an iterative method.

This procedure is called *speciation* – calculating the activity and concentration of every known ion and complex or ion-pair in a solution of a given bulk composition and at a given T and P. It is basically a problem in numerical analysis, and has not much to do with thermodynamics. In Chapter 15 we saw that strictly speaking we need not bother with species at all, we could assume all solutes such as acetic acid are completely ionized, except that in this particular case that approach does not work very well; HAc is too strongly associated. Nevertheless, speciation in the sense of this chapter is now an everyday occurrence in geochemistry.

16.3.2 Speciation using a computer

There are now hundreds of programs designed to compute such species in complex, multicomponent systems. Thorough reviews of both methods and programs include Van Zeggeren and Storey (1970); Nordstrom et al. (1979); Wolery (1979, 1983); Smith and Missen (1982); Reed (1982); Nordstrom and Ball (1984); DeCapitani and Brown (1987).

Program PHREEQC

A good program which is freely available at the USGS web site is PHREEQC. This program is capable of many modeling operations besides speciation, but let's see what it does with the same problem. To tell the program what to do, you prepare an input file:

```
DATABASE       llnl.dat
TITLE          Acetic acid 0.1 molal
SOLUTION 1
       units          mol/kgw
       temp           25
       pH             7       charge
       Acetate        0.1
END
```

The command `charge` tells the program to change the concentration of H^+ until the charges balance. After nine iterations, an output file is produced which includes these lines:

```
                                                  Log       Log       Log
          Species        Molality     Activity  Molality  Activity   Gamma
          H+           1.344e-003   1.292e-003    -2.872    -2.889  -0.017
          OH-          7.756e-012   7.442e-012   -11.110   -11.128  -0.018
          H2O          5.553e+001   9.983e-001    -0.001    -0.001   0.000
Acetate       1.000e-001
          HAcetate     9.866e-002   9.866e-002    -1.006    -1.006   0.000
          Acetate-     1.344e-003   1.290e-003    -2.872    -2.889  -0.018
H(0)          2.650e-017
          H2           1.325e-017   1.325e-017   -16.878   -16.878   0.000
O(0)          0.000e+000
          O2           0.000e+000   0.000e+000   -58.442   -58.442   0.000
```

The program has found these results:

$$m_{\mathrm{H^+}} = 10^{-2.872}$$

$$m_{\mathrm{OH^-}} = 10^{-11.110}$$

$$m_{\mathrm{HAc}} = 0.098\,66$$

$$m_{\mathrm{Ac^-}} = 0.001\,344$$

These are very close to the results from Equation (16.4) because in this case the activity coefficients are in fact close to 1.0. PHREEQC finds them to be

$$\gamma_{\mathrm{H+}} = 0.962$$

$$\gamma_{\mathrm{OH^-}} = 0.959$$

$$\gamma_{\mathrm{HAc}} = 1.0 \quad \text{(the value assumed for all uncharged species)}$$

$$\gamma_{\mathrm{Ac-}} = 0.959$$

Program SPECIES

Almost all computer programs which perform speciation calculations use a database containing all the necessary information about each reaction involved, and they usually have some method of calculating the change in these values with temperature, up to 300 or 350 °C. Without modification, none of these database-using programs are capable of performing speciation at supercritical temperatures and pressures. However, equilibrium constants for many reactions of interest at high T and P are available from program SUPCRT92.

Useful in this regard is program EQBRM, which is listed in Anderson and Crerar (1993), and which uses the method of Crerar (1975). This program has no database – all necessary equilibrium constants are entered in the input file, so the T and P is irrelevant as far as the program is concerned. This type of program is also useful in many research situations, where small model problems can be investigated using data and species of the user's choice, rather than data and species provided by a database.

Program EQBRM has been rewritten in MATLAB®, a convenient programming environment, and is listed in Appendix E. Although there are many variations in numerical methods of speciation, the method of Crerar (1975) is relatively simple, and is worth studying to see how it is done.

The MATLAB® program is named SPECIES. It uses as input a matrix of numbers which are prepared in a spreadsheet, then copied into MATLAB®. For the acetic acid problem, the output looks like this

```
all speciated K values are OK

No. of iterations:
   39
   'Species'     'Molalities'    'Gammas'    'Activities'
   'H+ '       [    0.0013629]  [0.96169]  [   0.0013107]
   'OH-'       [ 7.9536e-012]  [0.95928]  [7.6298e-012]
   'HAc'       [     0.098637]  [ 1.0001]  [     0.09865]
   'Ac-'       [    0.0013629]  [0.95974]  [    0.001308]
pH is:
     2.8825
```

which is again essentially the same as the other methods.

16.3.3 The predominance diagram for carbonate species

One of the more useful diagrams in geochemistry is the predominance diagram. In this, species activities are plotted versus pH, or often as pH versus $\log f_{O_2}$, and one of the more informative of these is the predominance diagram for aqueous CO_2. There are three carbonate species,[1] H_2CO_3, HCO_3^-, and CO_3^{2-} which would result from dissolving carbon dioxide gas in water. We have already calculated the equilibrium constant for one of the relevant ionic equilibria (§9.3.1):

$$H_2CO_3(aq) = HCO_3^- + H^+; \quad K = 10^{-6.37} \tag{16.5}$$

Another is

$$HCO_3^- = CO_3^{2-} + H^+; \quad K = 10^{-10.33} \tag{16.6}$$

And let's say that our total carbonate is $0.10\,m$, so that

$$m_{CO_3,\text{total}} = m_{CO_3^{2-}} + m_{HCO_3^-} + m_{H_2CO_3} = 0.10\,m \tag{16.7}$$

These three equations contain four variables, $m_{CO_3^{2-}}$, $m_{HCO_3^-}$, $m_{H_2CO_3}$, and m_{H^+}, so we need another equation, which is the charge balance,

$$m_{H^+} = m_{HCO_3^-} + 2\,m_{CO_3^{2-}} \tag{16.8}$$

---

[1] We use H_2CO_3 in the conventional sense discussed in §9.10.1.

An additional four unknowns are the activity coefficients of each of the four species for which we could write four more equations, making a total of eight equations and eight unknowns. This is a routine speciation problem. Program SPECIES gives

```
No. of iterations:
   38

   'Species'    'Molalities'   'Gammas'    'Activities'
   'H+    '    [2.0976e-004]  [0.9838]  [2.0636e-004]
   'OH-   '    [4.9278e-011]  [0.9834]  [4.8459e-011]
   'H2CO3'     [     0.0998]  [1.0000]  [     0.0998]
   'HCO3-'     [2.0976e-004]  [0.9835]  [2.0629e-004]
   'CO3-2'     [5.0009e-011]  [0.9350]  [4.6757e-011]
pH is:
   3.6854
```

so

$$a_{\mathrm{H^+}} = 10^{-3.685}$$

$$a_{\mathrm{H_2CO_3}} = 0.0998$$

$$a_{\mathrm{HCO_3^-}} = 2.063 \times 10^{-4}$$

$$a_{\mathrm{CO_3^{2-}}} = 4.678 \times 10^{-11}$$

You can easily check these results. For example we can check that

$$\begin{aligned}\frac{a_{\mathrm{HCO_3^-}}\, a_{\mathrm{H^+}}}{a_{\mathrm{H_2CO_3}}} &= \frac{2.063 \times 10^{-4} \cdot 10^{-3.685}}{0.0998} \\ &= 10^{-6.37}\end{aligned}$$

as it should.

Much more interesting than this single example, however, would be to see how the activities or concentrations of H_2CO_3, HCO_3^-, and CO_3^{2-} vary as a function of pH. Natural solutions contain many components in addition to CO_2 and water, and so the pH can be quite different from the one we have just calculated. To do this, we simply choose specific pH values from 0 to 14 and solve for the activities of the three carbonate species. The result is shown in Figure 16.3. This diagram makes it easy to see which species is dominant (has the largest concentration) at any given pH. For example, in seawater, with a pH of about 8.1, carbonate is present almost entirely as the bicarbonate species.

An interesting feature of this diagram is the fact that the intersection of the lines representing $a_{\mathrm{H_2CO_3}}$ and $a_{\mathrm{HCO_3^-}}$ occurs at a pH of 6.37, which is the p*K* value of the first ionization constant of H_2CO_3, and similarly for the intersection of the HCO_3^- and CO_3^{2-} lines, as shown in Figure 16.4. The reason for this is easy to see when you look at the equilibrium constant expressions, for example

$$\frac{a_{\mathrm{HCO_3^-}}\, a_{\mathrm{H^+}}}{a_{\mathrm{H_2CO_3}}} = 10^{-6.37}$$

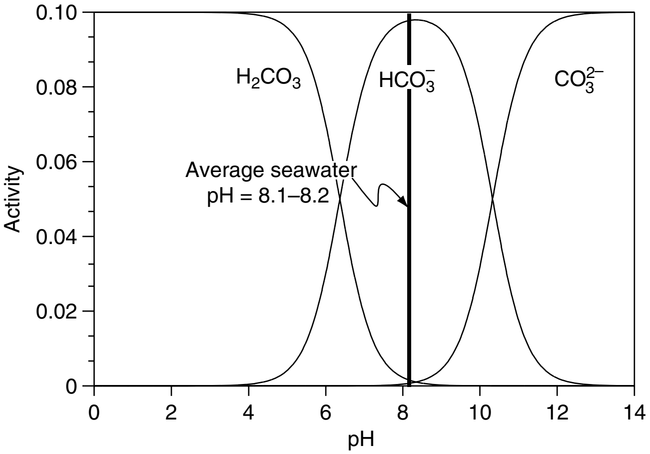


Figure 16.3 Activities of $H_2CO_3(aq)$, and bicarbonate and carbonate ions as a function of pH, for a total concentration of 0.10 *m*.

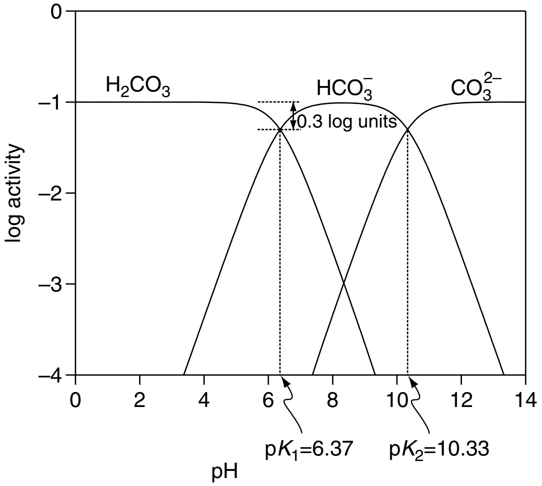


Figure 16.4 Same as Figure 16.3, but with a log activity scale. The pH values of the crossover points are the p*K* values of carbonic acid ionization.

At the crossover or intersection point, $a_{\mathrm{HCO}_3^-} = a_{\mathrm{H}_2\mathrm{CO}_3}$, so that

$$a_{\mathrm{H}^+} = 10^{-6.37}$$

which means that the intersection of the lines representing $a_{\mathrm{H}_2\mathrm{CO}_3}$ and $a_{\mathrm{HCO}_3^-}$ occurs at a pH of 6.37.

Another interesting fact is that the crossover points occur at an activity of 0.05 (Figure 16.3), because at each of the two crossover points, the activity of the third species (the one not involved in the crossover) is negligibly small, so that the two "crossing species" make up virtually the total activity or concentration, and therefore they both have a value of one-half the total concentration. Because $\log \frac{1}{2} = -0.30$, this means that on the log scale the crossovers occur 0.3 log units below the plateau representing the total concentration, as shown

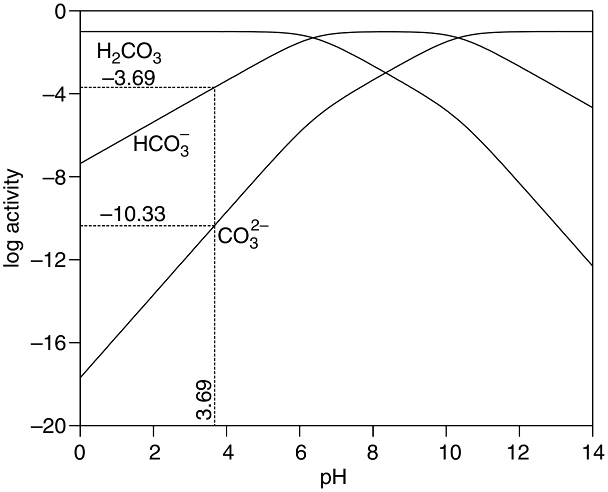


Figure 16.5 Same as Figure 16.4, but with an expanded activity axis. At a pH of 3.69, the HCO_3^- and CO_3^{2-} activities are those calculated in the text.

in Figure 16.4. A third interesting fact is that the slopes of the lines representing the activities of the species are either $+1$ or -1, just below their intersections (although they may change to $+2$ and -2 farther down, as shown in Figure 16.5). The combination of these three properties of these diagrams (called predominance diagrams) makes it very easy to rapidly sketch such a diagram, given some pK values.

Predominance diagrams are quite useful in seeing (and remembering) the relationships between species in dissociation reactions. For example, in Figure 16.5 you can see that $a_{CO_3^{2-}}$ goes to some very low values in acid solutions, a fact that explains (when you look at the solubility product) why carbonate minerals have such high solubilities in acid solutions. You also see that the activities of dissolved species never go to zero, at least in the model. In reality, of course, very low activities may mean that that species does not exist in the system.

As another example, Figure 16.6 shows the distribution of phosphate species at the same total concentration of 0.10 m. Such diagrams are clearly useful for any solute that can exist in a variety of species, differing only by the number of hydrogen ions (protons) they have.

16.4 Combining the IAP and the K_{sp}

So *speciation* allows us to determine the concentrations and activities of all known species in any chemically analyzed solution. Applied to seawater, it allows us to determine the IAP of calcite and of many other minerals. This is now done routinely on oceanographic research vessels, and a certain amount of variability is found in the composition of seawater from various locations. However, the composition of *average* seawater is quite well known and is

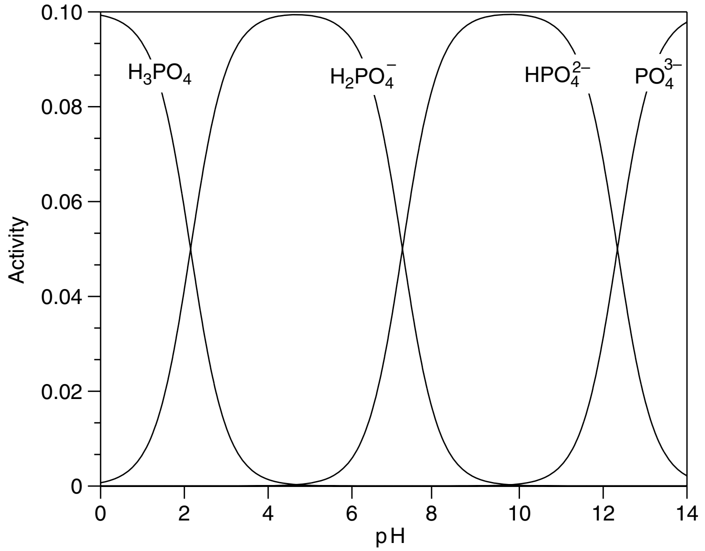


Figure 16.6 The activity distribution diagram for phosphoric acid. The p*K* values are 2.15, 7.21, and 12.34.

not greatly different from that proposed in the classic work of Garrels and Thompson (1962), who first applied this method, from which we get the data in Table 16.2.

From these data, we find

$$\begin{aligned}\mathrm{IAP} &= a_{\mathrm{Ca}^{2+}}\, a_{\mathrm{CO}_3^{2-}}\\ &= 0.0025 \times 0.000\,0054\\ &= 1.35 \times 10^{-8}\\ &= 10^{-7.870}\end{aligned}$$

Table 16.2 *Properties of the major ions in near-surface seawater.*

Major ions	Concentration m	Amount occurring as free ions %	Activity coefficient γ_H	Activity a
Na^+	0.475	99	0.76	0.357
Mg^{2+}	0.054	87	0.36	0.017
Ca^{2+}	0.010	91	0.28	0.0025
K^+	0.010	99	0.64	0.0063
Cl^-	0.56	100	0.64	0.36
SO_4^{2-}	0.028	54	0.12	0.0018
HCO_3^-	0.0024	69	0.68	0.0011
CO_3^{2-}	0.0003	9	0.20	0.0000054

Table 16.3 *IAP, Ω, and SI for calcium carbonate in average seawater.*

	IAP	K_{sp}	Ω	SI
seawater	13.5×10^{-9}			
calcite		4.97×10^{-9}	2.72	0.43
aragonite		7.55×10^{-9}	1.79	0.25

and we found (§16.2.1) that for calcite $K_{sp} = 10^{-8.304}$, and for aragonite $K_{sp} = 10^{-8.122}$. Therefore, average seawater is slightly supersaturated with calcium carbonate (Table 16.3).

This is a reasonably interesting result, as far as it goes. It means that marine organisms should have no difficulty in precipitating their carbonate shells, and once precipitated, they should not redissolve. This is consistent with the vast amounts of aragonitic mud on the Florida and Bahamas coastlines. These muds are made up almost exclusively of the shells of microscopic marine organisms, which sink to the bottom when the organisms die and do not redissolve.

16.4.1 What part of the sea is saturated with $CaCO_3$?

So far, thermodynamics and observations fit together fairly well. However, we know from oceanographic surveys that although vast areas of the sea are underlain by these carbonate muds, especially in water depths less than 3–4 km, the deepest parts of the ocean basins (4–6 km depth) have little or no carbonate in the bottom muds. Down to a variable depth, but usually between 3 and 4 km, the bottom muds are close to 100% calcium carbonate. Then within a relatively short increase in depth, the percentage of carbonate in the muds drops off rapidly, becoming zero or close to zero at another variable depth, but usually 4–5 km. The depth at which the rapid increase in carbonate dissolution begins is called the lysocline, and the depth below which there is little or no carbonate is called the carbonate compensation depth (CCD). Carbonate-secreting organisms are active at the surface virtually everywhere, and their carbonate shells are settling down through the water column everywhere, not just in shallow water. But while they coat the ocean floor at shallow depths, they never reach depths greater than 5 km or so – they dissolve completely at these depths. So obviously the oceans are saturated with $CaCO_3$ at and near the surface but undersaturated at great depths.

The exact explanation for this is one of the many continuing problems of chemical oceanography, but from our point of view it illustrates two things.

- The oceans, like most natural systems, are not at chemical, thermal, or mechanical equilibrium, but in spite of this, our equilibrium thermodynamic model is quite useful if we know how to apply it. We have shown how it is useful at the ocean surface; it is also useful at great depth. But obviously it would not be useful to apply it to the oceans as a whole system – they are too far from equilibrium.

- The explanation for the CCD and its variations is complex, involving the kinetics of carbonate dissolution, variations in ocean chemistry, temperature, and pressure, worldwide circulation patterns, and other factors. Equilibrium chemical thermodynamics does not suffice for an understanding of this natural system, but it is invariably the starting point for all other types of investigation. You must have an understanding of the equilibrium state before you can understand the departures from this state.

16.5 Mineral stability diagrams

The problem of calcite saturation in the sea is only one of a large number of problems in oceanography, geology, soil chemistry, and many other areas of science that involve solid$\rightleftharpoons$fluid reactions, and a variety of diagrams have been used to illustrate the thermodynamic relationships involved. Humans find two-dimensional diagrams much easier to understand than multidimensional systems of equations.

16.5.1 The reaction of feldspar with water

One of the most common minerals on Earth is feldspar, and its reaction with water to form other minerals such as clays is of great interest in several fields. In soils, feldspars (solid solutions of $NaAlSi_3O_8$, $KAlSi_3O_8$, and $CaAl_2Si_2O_8$) react to form clay minerals, helping to control soil acidity. In petroleum reservoirs, the same reaction forming clay minerals can have serious effects on the rock permeability and oil recovery. K-feldspars often have appreciable amounts of Pb substituting for K in their structures, and reaction of K-feldspar with formation fluids is thought to release this Pb to the fluid, which may then go on to form a lead ore-body elsewhere on its travels through the Earth's crust. During metamorphism, when rocks containing feldspars are subjected to high temperatures and pressures deep in the crust, feldspars participate in a variety of reactions with fluids and with other minerals, all of which are of interest to geologists studying the history of the Earth.

The question for us is, how do we apply thermodynamics to these reactions? The first thing we must do is write a reaction that seems interesting. This is one of the most difficult steps, but one that is rarely discussed because there are no rules to guide us. There is a very large number of reactions that could be written involving K-feldspar, depending on what other things are in the system, but only a few of these are useful. Only experience and scientific insight can distinguish between what might turn out to be useful and what will not. The reactions that appear in texts have, of course, proven to be useful.

If you were to write a reaction between K-feldspar and kaolinite with no previous experience or prejudices, but just putting them on opposite sides of an equal sign and adding other compounds to balance, you would likely wind up with something like this:

$$KAlSi_3O_8(s) + \tfrac{3}{2}H_2O(l) = \tfrac{1}{2}Al_2Si_2O_5(OH)_4(s) + 2\,SiO_2 + KOH \qquad (16.9)$$

We have not specified whether SiO_2 is quartz or $SiO_2(aq)$ – you have your choice. If you choose quartz, and quartz is present in the system, its activity will be 1.0; if you choose $SiO_2(aq)$, its activity will be (a dimensionless number equal to) the molality of silica in solution. Pure KOH is a solid, but it ionizes completely in solution, and you again have a choice of the (s) or the (aq) forms of data. If we choose to use quartz, $KOH(aq)$, and the maximum microcline form of $KAlSi_3O_8$ (or K-feldspar), the equilibrium constant turns out to be $10^{-7.88}$, and, because all the products and reactants except KOH have $a = 1$, at least if the water is reasonably pure, this turns out to be the molality of KOH in our system. What system? We have not been very specific about what our system consists of, other than that it contains K-feldspar, kaolinite, and water, at 25 °C, 1 bar. We might be thinking about a soil with groundwater.

But, you say, groundwater contains organic solutes, CO_2, and lots of other things. Why are they not in my system? Well, you can put them in if you wish, in balanced reactions, but you are under no obligation to do so. Your system is a model system, a simplified version of the real thing, and what goes in is under your control. Here we are choosing to look at the predicted KOH concentration. It may be that we have calculated something that is not useful, but that is our fault, not the fault of thermodynamics. Or we may have overlooked some factor which will invalidate our result, but that remains to be seen. We will see examples of such pitfalls shortly.

Knowing the KOH concentration that would equilibrate with K-feldspar and kaolinite is interesting, as far as it goes, but it is not the best way to look at this system. With a little insight, we can see that by subtracting OH^- from each side, and changing from quartz to $SiO_2(aq)$, we get

$$KAlSi_3O_8(s) + \tfrac{1}{2}H_2O(l) + H^+ = \tfrac{1}{2}Al_2Si_2O_5(OH)_4(s) + 2\,SiO_2(aq) + K^+ \qquad (16.10)$$

In this reaction, we have three potentially measurable things – the silica concentration $[SiO_2(aq)]$, the K^+ concentration, and the pH $(-\log a_{H^+})$. Given any two, we could predict the third, using the thermodynamic model. Or we could construct a three-dimensional diagram using these three parameters as axes. That's a bit too ambitious for us, so we will combine the a_{H^+} and a_{K^+} parameters into a ratio and plot this against $a_{SiO_2}(aq)$. This is now a standard procedure.

The way it is done is worth remembering, because it is used with various kinds of reactions. Write the equilibrium constant, take the logarithm of both sides, put your y-axis parameter on the left, the x-axis parameter on the right, and combine the rest of the terms. The equilibrium constant is

$$K = \frac{a^{1/2}_{\mathrm{Al_2Si_2O_5(OH)_4}} a^2_{\mathrm{SiO_2}(aq)} a_{\mathrm{K^+}}}{a_{\mathrm{KAlSi_3O_8}} a^{1/2}_{\mathrm{H_2O}(l)} a_{\mathrm{H^+}}} \tag{16.11}$$

If kaolinite, K-feldspar, and water are reasonably pure, their activities are 1, and

$$\log K = \log \frac{a_{\mathrm{K^+}}}{a_{\mathrm{H^+}}} + 2 \log a_{\mathrm{SiO_2}(aq)}$$

or

$$\log \frac{a_{\mathrm{K^+}}}{a_{\mathrm{H^+}}} = -2 \log a_{\mathrm{SiO_2}(aq)} + \log K \tag{16.12}$$

or

$$y = ax + b$$

where y is $\log(a_{\mathrm{K^+}}/a_{\mathrm{H^+}})$, x is $\log a_{\mathrm{SiO_2}(aq)}$, $a = -2$, and $b = \log K$. Thus Equation (16.12) is the equation of a straight line having a slope of -2, if we plot $\log(a_{\mathrm{K^+}}/a_{\mathrm{H^+}})$ against $\log a_{\mathrm{SiO_2}(aq)}$. To get K, we apply our routine method,

$$\begin{aligned}\Delta_r G^\circ = {} & \tfrac{1}{2}\Delta_f G^{\circ\ \text{kaolinite}}_{\mathrm{Al_2Si_2O_5(OH)_4}} + 2\,\Delta_f G^\circ_{\mathrm{SiO_2}(aq)} + \Delta_f G^\circ_{\mathrm{K^+}} \\ & - \Delta_f G^{\circ\ \text{microcline}}_{\mathrm{KAlSi_3O_8}} - \tfrac{1}{2}\Delta_f G^\circ_{\mathrm{H_2O}(l)} - \Delta_f G^\circ_{\mathrm{H^+}}\end{aligned}$$

Getting numbers from the tables,

$$\begin{aligned}\Delta_r G^\circ &= \tfrac{1}{2}(-3799.7) + 2(-833.411) + (-283.27) \\ &\quad - (-3742.9) - \tfrac{1}{2}(-237.129) \\ &= 11.523\,\mathrm{kJ\,mol^{-1}} \\ &= 11\,523\,\mathrm{J\,mol^{-1}}\end{aligned}$$

Then

$$\begin{aligned}\Delta_r G^\circ &= -RT \ln K \\ 11\,523 &= -(8.3145 \times 298.15) \ln K_{298}\end{aligned}$$

so

$$\begin{aligned}\log K_{298} &= -11\,523/(2.302\,59 \times 8.3145 \times 298.15) \\ &= -2.019\end{aligned}$$

Therefore,

$$\log \frac{a_{K^+}}{a_{H^+}} = -2 \log a_{SiO_2(aq)} - 2.019 \qquad (16.13)$$

This is a line on a plot of $\log(a_{K^+}/a_{H^+})$ versus $\log a_{SiO_2(aq)}$ having a slope of -2 and a y-intercept (the value of y when $x = 0$) of -2.019, as shown in Figure 16.7. The line we have just calculated is the thick line.

We now have a line on a graph. What does it mean? The meaning is implicit in the methods we used to get the line. We put the activities of the minerals and water equal to 1, and we used the equilibrium constant. That means that the line is the locus of solution conditions (a_{K^+}/a_{H^+}) and $a_{SiO_2(aq)}$ for which the pure minerals and water are in equilibrium with each other. For any values of (a_{K^+}/a_{H^+}) and $a_{SiO_2(aq)}$ that do not lie on the line, our solution cannot be in equilibrium with both minerals, although it might be in equilibrium with one or the other. Applying LeChatelier's principle to reaction (16.10), we see that increasing $a_{SiO_2(aq)}$ or increasing (a_{K^+}/a_{H^+}) favors the formation of K-feldspar, so a field of K-feldspar lies to the right and above our line, and a field of kaolinite lies to the left and below.

The next problem is that having chosen a fairly complex system like this, there are more possible reactions than the one we have chosen. K-feldspar can react not only to kaolinite, but also to muscovite, and kaolinite can also react

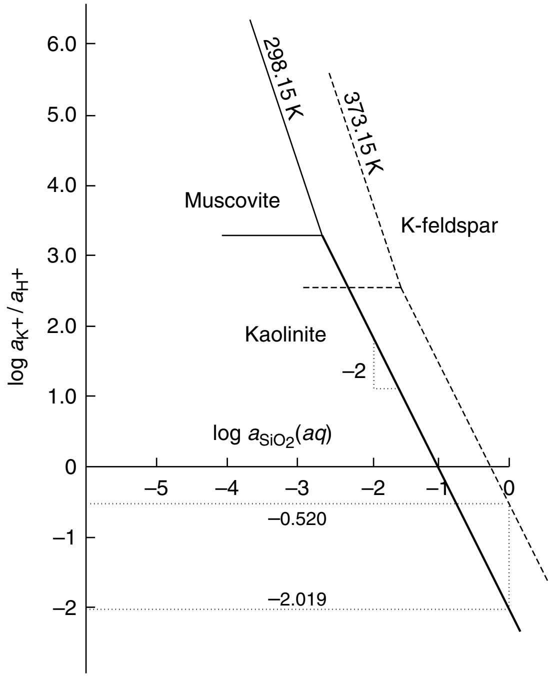


Figure 16.7 A plot of $\log(a_{K^+}/a_{H^+})$ versus $\log a_{SiO_2(aq)}$ at 298.15 K and 373.15 K.

to form muscovite. These reactions are

$$\tfrac{3}{2}\mathrm{KAlSi_3O_8}(s) + \mathrm{H^+} = \tfrac{1}{2}\mathrm{KAl_3Si_3O_{10}(OH)_2}(s) + 3\,\mathrm{SiO_2}(aq) + \mathrm{K^+} \tag{16.14}$$

and

$$\mathrm{KAl_3Si_3O_{10}(OH)_2}(s) + \tfrac{3}{2}\mathrm{H_2O} + \mathrm{H^+} = \tfrac{3}{2}\mathrm{Al_2Si_2O_5(OH)_4}(s) + \mathrm{K^+} \tag{16.15}$$

Using the same methods as before, we find that reaction (16.14) has a slope of -3 on our graph, and an intercept ($\log K$) of -4.668, and reaction (16.15) has a slope of 0 (it is independent of $a_{\mathrm{SiO_2}}$) and an intercept of 3.281. It can be shown (using the phase rule) that these three lines must intersect at a point, and so another way to draw them is to calculate the point of intersection, which is $\log a_{\mathrm{SiO_2}(aq)} = -2.650$, $\log(a_{\mathrm{K^+}}/a_{\mathrm{H^+}}) = 3.281$, and draw lines with slopes 0, -2, and -3 through this point. We now have a kind of phase diagram, showing which minerals are stable, not as a function of T and P, but of the composition of a solution in equilibrium with the minerals.

Effect of temperature

There are other mineral phases to be added to our diagram, but first let's look at the effect of temperature. If we want the same diagram for a temperature of 100 °C, we must calculate K at this temperature. To do this, we can use (9.20) or (9.22). We'll use both in the following calculations. For reaction (16.10), $\Delta_r H^\circ = 43\,437\,\mathrm{J\,mol^{-1}}$ and $\Delta_r S^\circ = 106.449\,\mathrm{J\,mol^{-1}}/K$. Therefore

$$\begin{aligned}\Delta_r G^\circ_{373} &= \Delta_r H^\circ_{298} - T\,\Delta_r S^\circ_{298}\\ &= 43\,437 - 373.15 \times 106.449\\ &= 3715.556\,\mathrm{J\,mol^{-1}}\end{aligned}$$

$$\begin{aligned}\log K &= -3715.556/(2.302\,59 \times 8.3145 \times 373.15)\\ &= -0.520\\ &= \log\frac{a_{\mathrm{K^+}}}{a_{\mathrm{H^+}}} + 2\,\log a_{\mathrm{SiO_2}(aq)}\end{aligned}$$

giving

$$\log\frac{a_{\mathrm{K^+}}}{a_{\mathrm{H^+}}} = -2\,\log a_{\mathrm{SiO_2}(aq)} - 0.520$$

as the equation of the K-feldspar–kaolinite boundary at 100 °C.

For reaction (16.14), $\Delta_r H^\circ = 74\,473\,\mathrm{J\,mol^{-1}}$ and $\log K_{298} = -4.668$, and so

$$\begin{aligned}\log K_{373} &= \log K_{298} - \frac{\Delta_r H^\circ_{298}}{2.302\,59\,R}\left(\frac{1}{T} - \frac{1}{298.15}\right)\\ &= -4.668 - \frac{74\,473}{2.302\,59 \times 8.3145}\left(\frac{1}{373.15} - \frac{1}{298.15}\right)\\ &= -2.046\end{aligned}$$

For reaction (16.15), $\Delta_r H^\circ = -18\,635\,\mathrm{J\,mol^{-1}}$, $\Delta_r S^\circ = -1.165\,\mathrm{J\,mol^{-1}}/K$, and so

$$\begin{aligned}\log K_{373} &= \frac{-\Delta_r H^\circ_{298}}{2.302\,59\,RT} + \frac{\Delta_r S^\circ_{298}}{2.302\,59\,R}\\ &= \frac{18\,635}{2.302\,59 \times 8.3145 \times 373.15} + \frac{-1.165}{2.302\,59 \times 8.3145}\\ &= 2.548\end{aligned}$$

These three lines intersect at $\log(a_{\mathrm{K}^+}/a_{\mathrm{H}^+}) = 2.55$, $\log a_{\mathrm{SiO}_2(aq)} = -1.53$, as shown in Figure 16.7.

Effect of choice of data

This business of looking up data and plotting by hand gets tedious. Surely there are computer programs to do all this for us? There are many, but none of them relieves the user of a fundamental responsibility – you are responsible for choosing your data, and of course for understanding what you are doing. People learning how to use thermodynamics generally assume that the data they find at the back of their text, such as those in Appendix B, are "true" in some absolute sense, and this does no harm while learning the subject. However, when you begin to apply the subject in some area of interest, you sooner or later discover that there are a number of sources of data, that they often disagree, and that the choice of data can affect your results considerably.

One such source of data is SUPCRT92 (Johnson et al., 1992). In Figure 16.8 are shown the mineral boundaries calculated using K values from SUPCRT92 for 373.15 K, in comparison with those from Figure 16.7. This comparison is made to impress on you the fairly large differences that can result from using different sources of data. Most compilations of data these days have been subjected to some sort of quality and consistency analysis, but methods vary. The refinement and improvement of thermodynamic data is a never-ending process, and the choice of thermodynamic data is an integral part of applying thermodynamics to natural systems. The data in SUPCRT92 are especially good for most geochemical applications, because they have been derived using many high pressure–temperature experiments involving minerals that are not always considered by chemists. In other words, they are "tuned" to mineral equilibria. On the other hand, it does not contain many compounds that do not occur in nature which are of great interest to chemists.

The differences between the data in Appendix B and those in SUPCRT92 are relatively small (Table 16.4), yet they result in quite large differences in phase diagrams like Figure 16.8. Let this be a lesson in choosing data – the best are none too good, and no one knows for sure what “the best” data are.

While we are using SUPCRT92, we might as well add two final reactions to our diagram. These relate kaolinite and muscovite to gibbsite:

$$\mathrm{Al_2Si_2O_5(OH)_4}(s) + \mathrm{H_2O}(l) = 2\,\mathrm{Al(OH)_3}(s) + 2\,\mathrm{SiO_2}(aq) \tag{16.16}$$

and

$$\mathrm{KAl_3Si_3O_{10}(OH)_2}(s) + 3\,\mathrm{H_2O}(l) + \mathrm{H^+} = 3\,\mathrm{Al(OH)_3}(s) + 3\,\mathrm{SiO_2}(aq) + \mathrm{K^+} \tag{16.17}$$

These two reactions also meet at a point with the muscovite–kaolinite reaction, as shown in Figure 16.8.

Table 16.4 *The difference between the data in Appendix B and in* SUPCRT92.

Mineral	SUPCRT92 $\Delta_f G°$, kJ mol^{-1}	Appendix B $\Delta_f G°$, kJ mol^{-1}	App. B – SUPCRT92 kJ mol^{-1}	% Difference
albite	−3708.313	−3711.5	−3.2	0.08
K-feldspar	−3746.245	−3742.9	−3.3	0.08
muscovite	−5591.083	−5608.4	−17.3	0.30
quartz	−856.239	−856.64	−0.4	0.05

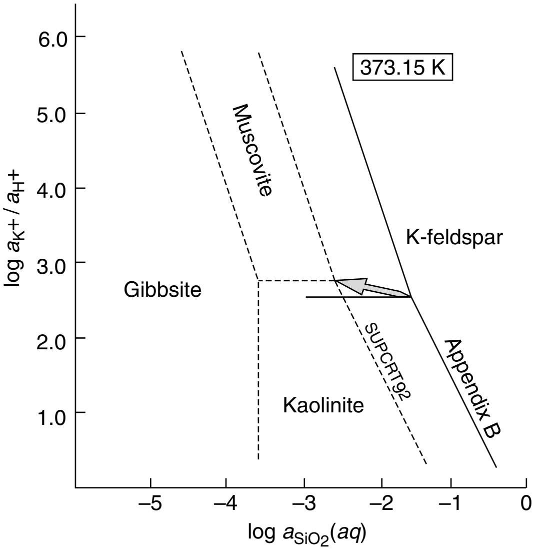


Figure 16.8 A plot of $\log(a_{K^+}/a_{H^+})$ versus $\log a_{SiO_2(aq)}$ at 373.15 K, showing the effect of changing from the data in Appendix B to data in program SUPCRT92.

The use of program SUPCRT92 is of no fundamental importance here. We could have illustrated the point that the choice of data is important by using data from any number of other sources.

The problem of metastable phases

One of the pitfalls in choosing data and drawing diagrams, such as we have done, is that we may not have considered all the possible reactions in our system. That is, there may be phases that are more stable than the ones we have chosen – we may have chosen metastable phases. This is illustrated in our system by the fact that there is another aluminosilicate phase, pyrophyllite, which has more silica in it than does kaolinite and so is stable at higher values of $a_{\mathrm{SiO}_2(aq)}$. Considering the reactions

$$\mathrm{KAlSi_3O_8}(s) + \mathrm{H}^+ = \tfrac{1}{2}\,\mathrm{Al_2Si_4O_{10}(OH)_2}(s) + \mathrm{SiO_2}(aq) + \mathrm{K}^+ \tag{16.18}$$

and

$$\tfrac{1}{2}\mathrm{Al_2Si_4O_{10}(OH)_2} + \tfrac{1}{2}\mathrm{H_2O}(l) = \tfrac{1}{2}\mathrm{Al_2Si_2O_5(OH)_4}(s) + \mathrm{SiO_2}(aq) \tag{16.19}$$

results in the two boundaries shown in Figure 16.9, which completely enclose the K-feldspar–kaolinite boundary. This means that at 100 °C, and according to the SUPCRT92 data, K-feldspar and kaolinite are not stable together in the

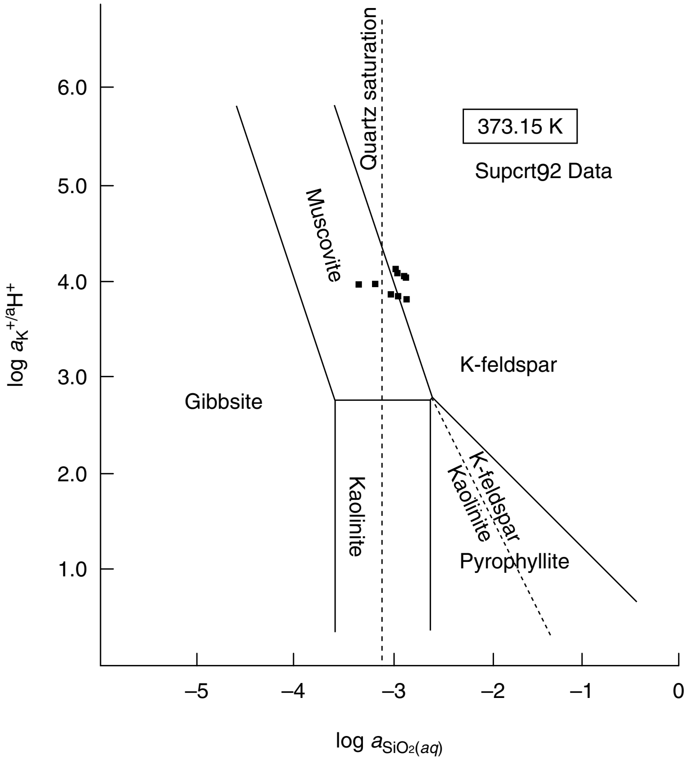


Figure 16.9 A plot of $\log(a_{\mathrm{K}^+}/a_{\mathrm{H}^+})$ versus $\log a_{\mathrm{SiO}_2(aq)}$ at 373.15 K, using data from SUPCRT92, showing the stability field of pyrophyllite and the quartz saturation line. Each black square represents a brine from the Kettleman North Dome oil field (Merino, 1975).

presence of water. They should react to form pyrophyllite. In nature, of course, they may not. K-feldspar may well react directly to kaolinite, in spite of thermodynamics, but it would be a metastable reaction; that is, a reaction involving metastable phases.

One other thing. The solubility of quartz, calculated by the method in §9.4.1, is $10^{-3.078}\,m$ at 100 °C according to SUPCRT92. This is shown in Figure 16.9 as a vertical line. A solution in equilibrium with quartz at 100 °C must lie on this line. Solutions to the right of it are supersaturated, and solutions to the left are undersaturated, with quartz. Therefore, if your system contains quartz, as most soils do, then not only is the assemblage K-feldspar + kaolinite metastable with respect to pyrophyllite, but pyrophyllite is itself metastable. According to our diagram, pyrophyllite can only exist in an aqueous solution if that solution is supersaturated with silica. If the silica were to precipitate as quartz, pyrophyllite should break down to form kaolinite, releasing silica to the solution, by reaction (16.19). Again, nature may not do this. It is quite common for natural solutions to be supersaturated with silica, even in the presence of quartz, which is one reason that we use $\log a_{\mathrm{SiO_2}(aq)}$ as a variable in our diagrams.

Plotting solution data

Also shown on Figure 16.9 are a few data points taken at random from Merino (1975). Each one is from a different well in the Kettleman North Dome oil field in California. The numbers for $a_{\mathrm{K^+}}$, $a_{\mathrm{H^+}}$, and $a_{\mathrm{SiO_2}(aq)}$ are obtained from the chemical analyses in Merino (1975) by speciation. It appears that the solution concentrations are clustering around the K-feldspar–muscovite boundary, near the quartz saturation line, and are likely being controlled by these phases, which are known to be in the sandstones containing the fluids sampled.

The albite–K-feldspar diagram

There are, of course, a great many diagrams we might plot, depending on our interests. To investigate the possibility that the K/Na ratio in the fluids at Kettleman North Dome is controlled by the coexistence of albite and K-feldspar, we could write and plot the reaction

$$\mathrm{NaAlSi_3O_8}(s) + \mathrm{K^+} = \mathrm{KAlSi_3O_8}(s) + \mathrm{Na^+} \tag{16.20}$$

Following the procedures outlined above, we find that $\Delta_r G^\circ_{373} = -7052.9\,\mathrm{J\,mol^{-1}}$ (Appendix B) or $-14\,945\,\mathrm{J\,mol^{-1}}$ (SUPCRT92). If pure albite and pure K-feldspar are present in the rocks, the equilibrium constant for this reaction becomes

$$\begin{aligned} K_{373} &= \frac{a_{\mathrm{Na^+}}}{a_{\mathrm{K^+}}} && (16.21)\\ &= 10^{0.987} && \text{Appendix B}\\ \text{or}\quad &= 10^{2.092} && \text{SUPCRT92} \end{aligned}$$

and we could plot a_{Na^+} versus a_{K^+}, with a line separating a field of albite from a field of K-feldspar. However, if we want also to have muscovite on the same diagram, it would be convenient to use the $a_{\mathrm{K}^+}/a_{\mathrm{H}^+}$ parameter from our previous diagrams. So we simply divide both numerator and denominator on the right side of (16.21) by a_{H^+}, giving us

$$\begin{aligned} K_{373} &= \frac{(a_{\mathrm{Na}^+}/a_{\mathrm{H}^+})}{(a_{\mathrm{K}^+}/a_{\mathrm{H}^+})} \qquad (16.22)\\ &= 10^{0.987}\ (\text{App. B}) \quad \text{or} \quad 10^{2.092}\ (\textsc{supcrt}92) \end{aligned}$$

To add a boundary between muscovite and albite, we write a reaction between these two minerals involving the same ions and SiO_2, which turns out to be

$$\mathrm{KAl_3Si_3O_{10}(OH)_2}(s) + 3\,\mathrm{Na^+} + 6\,\mathrm{SiO_2}(s) = 3\,\mathrm{NAlSi_3O_8}(s) + 2\,\mathrm{H^+} + \mathrm{K^+} \qquad (16.23)$$

However, here we choose to use quartz as our SiO_2, rather than $SiO_2(aq)$, because the solutions appear to have their silica content controlled by quartz, or near enough, and besides, this diagram has no silica variable. If you prefer, you could calculate an average $a_{\mathrm{SiO_2}(aq)}$ of the data points and use this value (or any other value) for our new diagram – the choice is yours. The equilibrium constant for (10.22) at 100 °C is $10^{-16.282}$ (Appendix B) or $10^{-14.948}$ (SUPCRT92), and, like (10.20), needs a little manipulation to get it into a form we can use.

$$\begin{aligned} K_{373} &= \frac{a_{\mathrm{K}^+}a^2_{\mathrm{H}^+}}{a^3_{\mathrm{Na}^+}}\\ &= \frac{a_{\mathrm{K}^+}a^2_{\mathrm{H}^+}}{a^3_{\mathrm{Na}^+}} \cdot \frac{a_{\mathrm{H}^+}}{a_{\mathrm{H}^+}}\\ &= \frac{a_{\mathrm{K}^+}/a_{\mathrm{H}^+}}{(a_{\mathrm{Na}^+}/a_{\mathrm{H}^+})^3} \qquad (16.24) \end{aligned}$$

so

$$\begin{aligned} \log K_{373} &= \log\frac{a_{\mathrm{K}^+}}{a_{\mathrm{H}^+}} - 3\log\frac{a_{\mathrm{Na}^+}}{a_{\mathrm{H}^+}}\\ \log\frac{a_{\mathrm{Na}^+}}{a_{\mathrm{H}^+}} &= \tfrac{1}{3}\log\frac{a_{\mathrm{K}^+}}{a_{\mathrm{H}^+}} + 16.282/3 \quad \text{or} \quad +14.948/3 \qquad (16.25) \end{aligned}$$

and (10.24) is the equation of the muscovite–albite boundary on a plot of $a_{\mathrm{Na}^+}/a_{\mathrm{H}^+}$ versus $a_{\mathrm{K}^+}/a_{\mathrm{H}^+}$, shown in Figure 16.10. The black squares in Figure 16.10 represent the same samples as shown in Figure 16.9. They cluster near the muscovite–albite–K-feldspar intersection, which would be expected if the solution compositions are being more or less controlled by the host rocks. However, the data are definitely offset from this intersection into the muscovite field and appear to lie along the metastable extension of the albite–K-feldspar boundary. This may well be an example of fluid compositions being controlled by a metastable assemblage, as was mentioned in connection with pyrophyllite.

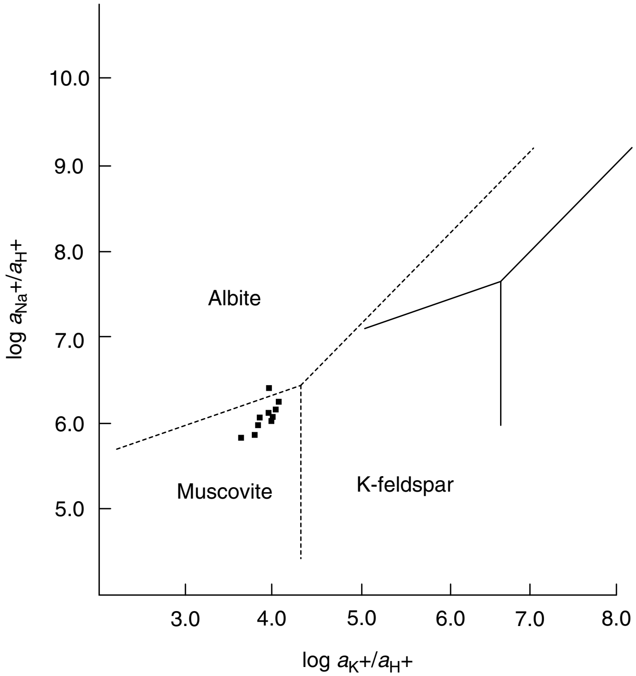


Figure 16.10 A plot of $\log(a_{Na^+}/a_{H^+})$ versus $\log(a_{K^+}/a_{H^+})$ at 373.15 K for solutions saturated with quartz. Dashed lines from program SUPCRT92, solid lines from Appendix B. A few points from Merino (1975) are shown as black squares.

16.6 Summary

In this chapter we have seen how a knowledge of equilibrium constants and activities can be used to construct diagrams that relate fluid compositions to the minerals in equilibrium with those fluids. Real fluid compositions (activities) can then be plotted on the diagrams to see how well the hypothesis of equilibrium holds, or, assuming equilibrium, to predict what phases are coexisting with fluids, and what will happen if we change the fluid composition. There are a large number of other applications of equilibrium thermodynamics, to be sure, but if you understand equilibrium constants and activities thoroughly, you will have no problem in understanding most other applications.

In the case of the Kettleman North Dome fluids, the mineralogy of the rocks in contact with the fluids is known. You might say, well, if we know what minerals are there, why go to all this trouble to show that the fluid compositions reflect this? Would you not expect them to reflect the mineralogy of their host rocks? For one thing, we didn't know whether the composition of oil field brines or any other kind of natural fluid was controlled by host rocks or not, until this kind of test was done. For all we knew, perhaps disequilibrium reigned supreme. Furthermore, we're still learning.

Why is it important? This question gets at the whole idea of using science to try to understand natural processes. Until you have a quantitative model that can simulate or account for natural data, such as shown in Figures 16.9

and 16.10, you cannot hope to change or control the situation to your benefit. For example, the petroleum geologists at Kettleman North Dome might want to inject something to change the fluid characteristics. Without a chemical model, they would have no way of predicting what would happen. The equilibrium model has its limitations, but it is a good place to start.

17

Phase diagrams

17.1 What is a phase diagram?

A phase diagram in the general sense is any diagram that shows what phase or phases are stable as a function of some chosen system variable or variables. Therefore, the *Eh*–pH, log f_{O_2}–pH and activity–activity diagrams we have been looking at are a kind of phase diagram. However, if you mention the subject of phase diagrams to a petrologist, a metallurgist, or a ceramic scientist, they will immediately think of a particular type of diagram that is of great usefulness in these subjects. In these sciences, the compositions of phases and their relationships during phase changes, particularly solid→liquid and liquid→solid changes, are of particular importance, so diagrams that depict this information as a function of temperature and pressure have come to be the subject of "phase diagrams."

17.1.1 Thermodynamics and phase diagrams

Though it is true that phase relations can always be described in terms of the thermodynamic principles and equations we have been discussing, and that any phase diagram can in principle be calculated given the appropriate data, the emphasis in this chapter changes from one of calculating what we want to know from numbers in tables of data, to one of simply representing experimentally derived facts in diagrammatic form. The reason for this is that once we get into systems more complex than a single component, and especially when high-temperature melt phases are involved, the calculations are often not possible because the data are not available, and even if they are available, are not very accurate, because they are very sensitive to small inaccuracies in the data. Therefore, in this book, although we will show the relationship between functions such as G and our diagrams, this will be in an illustrative rather than a quantitative way.

17.1.2 Phase diagrams as models

Metallurgists and ceramicists quite often deal with simple two- and three-component systems and use phase diagrams to represent their experimental results on the phase relations in these systems. The diagrams therefore truly represent their systems. Petrologists, on the other hand, are interested in the origins of natural rocks, which commonly have ten or more important components. Systems this complex cannot be represented in simple diagrams and, in fact, can hardly even be thought about in a quantitative way. Experiments can and have been done using natural rocks, but the results are complex and may not be generally applicable. Therefore, petrologists use simpler systems such as those having two and three components to better understand the principles involved and to investigate simple models of the complex systems in nature.

Phase diagrams represent *equilibrium* relationships. Once these are depicted, simple *processes* such as melting and crystallization can be considered, but because as represented on diagrams these involve continuous successions of equilibrium states, they are *reversible processes* in the sense of §2.6.2.

17.2 Unary systems

Figure 17.1 shows a typical although hypothetical unary (one component) diagram for compound α (α stands for the formulas of some compound, such as NaCl or $CaCO_3$). Although the diagram shows three different *phases* (solid, liquid, and gas), all three have the same composition (whatever the

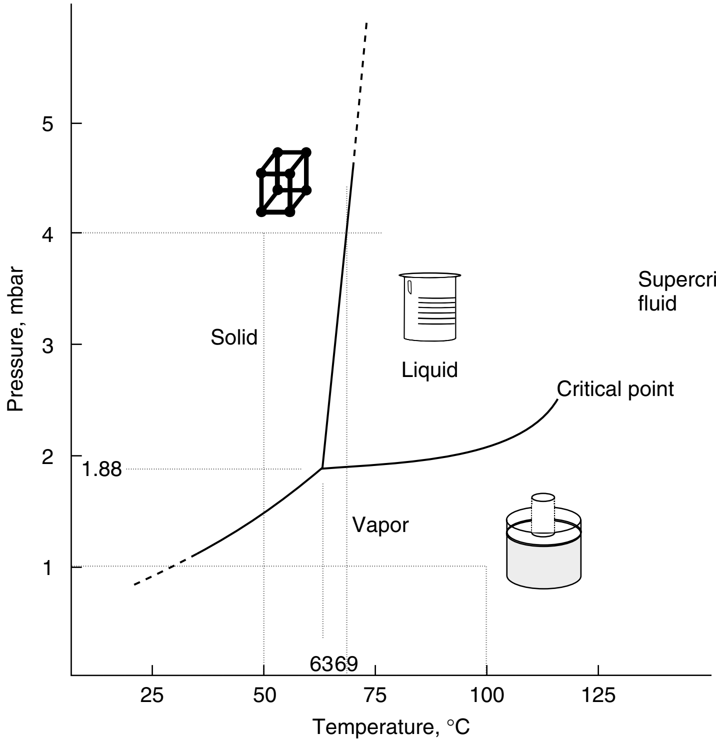


Figure 17.1 Phase diagram for hypothetical compound α.

chemical composition of α is), and so the system is unary. This simple diagram contains a surprising amount of information, but you must know how to "read" the diagram. First, note that the diagram contains labeled *areas*, *lines* that separate the areas, and *points*. Every location on the diagram has a pair of x–y coordinates, that is, a pressure and a temperature. For example, a pressure of 4 mbar and a temperature of 50 °C are the coordinates of a point in the area marked Solid. Under these conditions, the stable form of compound α is observed to be a crystalline solid. If the pressure on α is reduced to 1 mbar, and the temperature increased to 100 °C, the stable form of α is gaseous. Similarly, for any combination of pressures and temperatures within the area marked Liquid, the stable form of α is liquid. The phase diagram is in fact a record of these experimental observations about the form of α under various conditions of P and T. As mentioned earlier, the vast majority of phase diagrams record the results of experiments – they are not usually the result of theoretical calculations. They are more often a *source* of thermodynamic data than the result of using such data.

Obviously, within these areas, P and T could be changed considerably without changing the nature of the phase, although the *properties* of the phase (its density or heat capacity, say) would certainly change with P and T. It appears, then, that for α, and for any pure compound, we must choose *two* variables in order to define the state of the compound. Thus to answer the question "what is the density (or heat capacity, refractive index, entropy,…) of α?" we must first specify two variables – the P and the T we are interested in. One is not enough – at 4 mbar, α can have quite a range of densities, but at 4 mbar, 50 °C, its density is fixed and determinable, as are all its other properties. So we say that in each of its three forms – solid, liquid, and gas – α has two *degrees of freedom* – two variables must be specified before *all* are specified. These two variables are in practice usually T and P, but in principle any two would do. The phase rule summarizes all this discussion by simply saying

$$
\begin{aligned}
f &= c - p + 2 \\
&= 1 - 1 + 2 \\
&= 2
\end{aligned}
$$

With α at 4 mbar and 50 °C, consider that we raise the temperature gradually. Nothing much happens, except that the properties of α change continuously, until we reach 69 °C, which is the temperature of the boundary between the solid and liquid fields. At this T, solid α is observed to begin to melt, and at this T, any proportions of solid and liquid α are possible (i.e., almost all solid α with a drop of liquid; or almost all liquid α with a tiny amount of solid; or anything in between). However, if the temperature is held very slightly above the melting T, α becomes completely liquid. The solid–liquid boundary line then is a locus of T–P conditions that permit the coexistence of solid and liquid α. It records the melting temperature of α as a function of pressure.

Note too that because it is a line rather than an area, or because there are two coexisting phases rather than one, we now have only one degree of freedom. In other words, at 4 mbar we now have no choice of temperature. If solid and liquid coexist at equilibrium, the temperature must be 69 °C – it is chosen for us. We can still choose whatever P we like (within certain limits), but once we have exercised our one degree of freedom and chosen a pressure, the temperature and all properties of the two phases of α are fixed. Again, we note that the one degree of freedom can be any property, not just T or P. We might choose a certain value for the entropy of solid α, for example; we would then find that there was only one T and P where solid α with this particular S_α could coexist with liquid α. The phase rule agrees, saying

$$\begin{aligned} f &= c - p + 2 \\ &= 1 - 2 + 2 \\ &= 1 \end{aligned}$$

Similar comments apply to the boundary between the fields of Liquid and Vapor, which records the *boiling temperatures* of α, and the boundary between the Solid and Vapor fields, which records the *sublimation temperatures* of α. Where these three boundaries come together at about 63 °C, 1.88 mbar (a *triple point*), the three phase fields come together, and solid, liquid and gaseous α can coexist in any proportions at this particular T and P. Note that for the coexistence of these three phases, we have lost another degree of freedom. In fact we have no choice at all – if we want three phases to coexist, the T and P must be 63 °C and 1.88 mbar. As the number of coexisting phases increases, the number of degrees of freedom decreases. Negative degrees of freedom are not possible, so in a one-component system the phase rule predicts that the *maximum* number of phases at equilibrium is three:

$$\begin{aligned} f &= c - p + 2 \\ &= 1 - 3 + 2 \\ &= 0 \end{aligned}$$

17.2.1 Gibbs energy sections

Despite the fact, mentioned in §17.1.1, that phase diagrams are for the most part experimentally derived, they are controlled by and must conform to fundamental thermodynamic relationships. Understanding phase diagrams is enhanced by examining the relationships between the diagrams and the underlying thermodynamics.

From our study of thermodynamics in previous chapters, we know that the stable state of a system under given conditions is that state having the lowest value of the Gibbs energy, **G** (or G). If a system does not have the lowest possible value of **G**, a spontaneous process will take place (according to our model) until this lowest value is achieved. Also, we know that if two phases are in equilibrium in a unary system, the Gibbs energy of the component is the same in each phase (§6.3.1; Figure 6.8). Therefore, the phase boundaries in Figure 17.1 are places where G_α is the same in two phases, as shown in Figure 17.2. Note too that we may calculate and plot the Gibbs energy (and other properties) of a liquid phase in regions where it is not the stable phase. When we say, for example, that at 4 mbar, 50 °C in the solid stability field, $G_\alpha^{\text{solid}} < G_\alpha^{\text{liquid}}$, we imply that if liquid α could exist at 4 mbar, 50 °C, its G would be greater than that of G_α^{solid}. We could, in fact, plot the values of G for all possible phases over all parts of the diagram. If we did so and looked at a part near the solid–liquid boundary, we would see something like Figure 17.3.

G–T sections

Figure 17.4 shows a section through Figure 17.1 at a pressure of 2 mbar. At temperatures below 64 °C at 2 mbar pressure, α is solid, and the Gibbs energy of this solid (G_α^{solid}) is shown by the line labeled *solid*. Naturally, as we don't *know* the absolute Gibbs energy of any substance, we cannot place any absolute

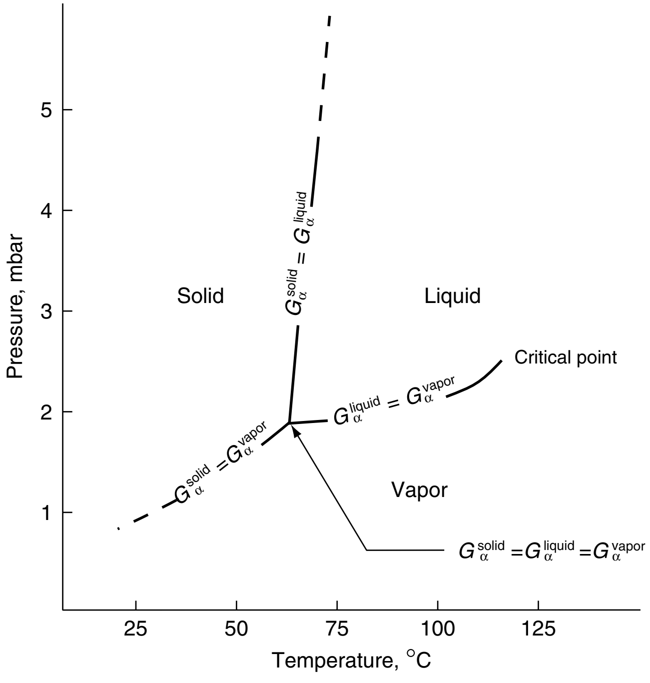


Figure 17.2 Gibbs energy relationships in the phase diagram for compound α.

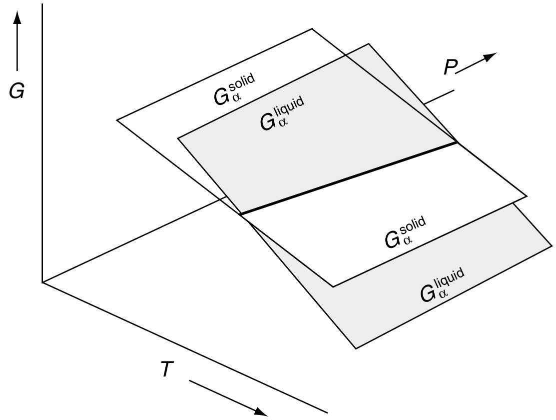


Figure 17.3 *G–T–P* diagram for part of Figure 17.1. The heavy line at the intersection of the G_α^{solid} and G_α^{liquid} surfaces is the melting curve.

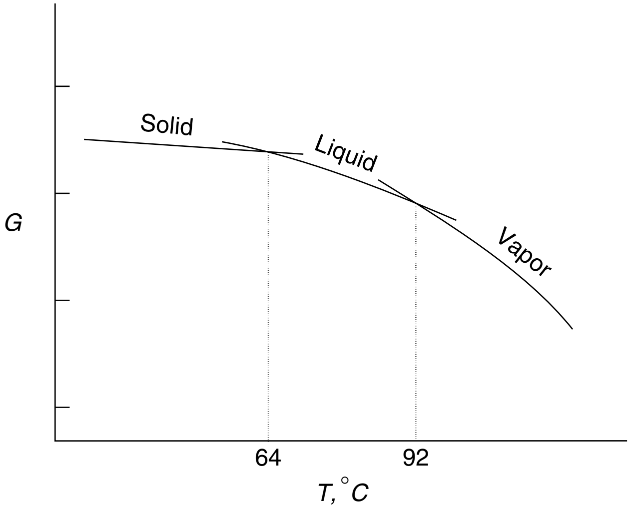


Figure 17.4 *G–T* section through Figure 17.1 at 2 mbar.

numbers on the G-axis. However, we *do* know the slope of this line [the slope is $(\partial G_\alpha/\partial T)_P = -S_\alpha$, and we know S for most compounds], and so we could establish some arbitrary energy divisions on the G-axis and plot a line with the correct slope. This line would have a gentle downward curvature because S gradually increases with T, but to a first approximation it is a straight line. This line continues to the melting temperature, 64 °C, at which point it intersects another line giving the values of G_α^{liquid}. This line has a steeper slope, because the entropy of a liquid is always greater than the entropy of a solid of the same composition. At the intersection, $G_\alpha^{\text{solid}} = G_\alpha^{\text{liquid}}$, as required by phase equilibrium theory (§6.3.1, §6.4).

The G_α^{liquid} line then continues with a gentle downward curvature through the liquid stability region at 2 mbar until it reaches another phase boundary, the boiling curve, at 92 °C. Here it intersects the G_α^{vapor} curve, which has a still steeper slope, because the entropy of gases is always much greater than that of liquids.

Note the similarity of this diagram to Figure 6.14, where we considered *G*–*T* sections in a quantitative way, to calculate the positions of phase boundaries.

G–*P* sections

Figure 17.5 shows a *G*–*P* section through Figure 17.1 at a temperature of 69 °C. At pressures below 1.89 mbar, α is gaseous, and the Gibbs energy of this gas is shown by the line labeled *vapor*. The slope of the line is $(\partial G_\alpha/\partial P)_T = V_\alpha$, and as the molar volume of gases is large, the line has a steep slope. This line intersects another line, giving the values of G_α^{liquid}, having a smaller positive slope, because $V_\alpha^{\text{liquid}} < V_\alpha^{\text{gas}}$. This line continues, again with slight downward curvature because the molar volume of the liquid decreases slightly with increasing pressure, until it reaches the freezing curve at 4 mbar, where it intersects the line giving G_α^{solid}. Note the similarity between this diagram and Figure 6.9.

17.2.2 Some important unary systems

Substances whose phase relations are interesting for various reasons include carbon (C), iron (Fe), water (H_2O), silica (SiO_2), aluminum silicate (Al_2SiO_5), and calcium carbonate ($CaCO_3$).

H_2O

The phase diagram for water at relatively low pressures is shown in Figure 17.6. Water is a most unusual substance. It is one of the very few compounds that expands when it freezes, meaning that ice floats. Most substances have solid

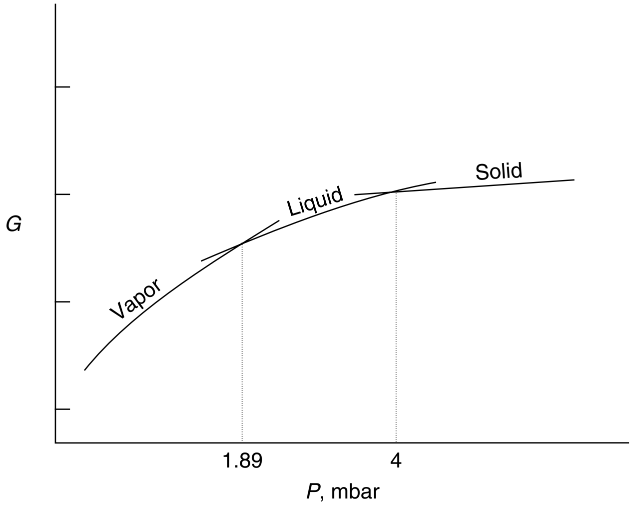
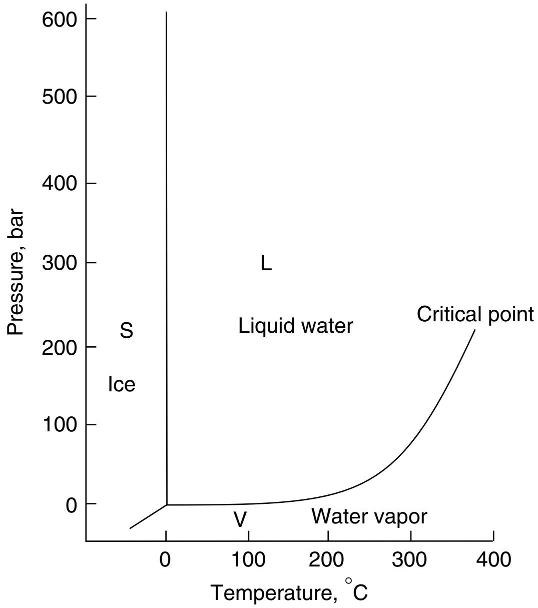


Figure 17.5 *G*–*P* section through Figure 17.1 at 69 °C.

Figure 17.6 Phase diagram for H_2O at relatively low pressures. The solid–liquid boundary is very steep, but in fact has a negative slope.



forms that are denser than their corresponding liquids, and hence will sink during freezing. The fact that ice floats in water is shown in Figure 17.6 by the fact that the liquid–solid boundary (the freezing/melting curve) has a negative slope. In our “typical” unary system (Figure 17.1), this curve has a positive slope. In both cases (Figures 17.1 and 17.6), the denser phase lies at higher pressures, as required by LeChatelier’s principle. The unusual thing is that in the H_2O system, the denser phase is the liquid.

The other term in the slope expression ($dP/dT = \Delta S/\Delta V$) is ΔS, which is invariably greater in the liquid than in the solid; therefore, the volume change, ΔV, determines whether dP/dT will be positive or negative.

Figure 17.7 shows the same system over a much greater range of pressures. The striking thing about this diagram is the large number of polymorphs of ice, each with its own stability field. These polymorphs give rise to several *triple points*, showing that the solid–liquid–vapor triple point shown in Figure 17.1, which every unary system has, is often not the only one. We came across this phenomenon (a triple point generated by solid polymorphs) previously (§6.5, Figure 6.13). Note the fact that liquid water will freeze (to ice VII) at about 24 kbar at the boiling temperature (100 °C). Note too that the negative slope of the freezing curve (between Ice I and Liquid) extends to only about 2 kbar.

SiO_2

Silica, one of the most common compounds on Earth, has a number of interesting and complex phase relations, shown in Figure 17.8.

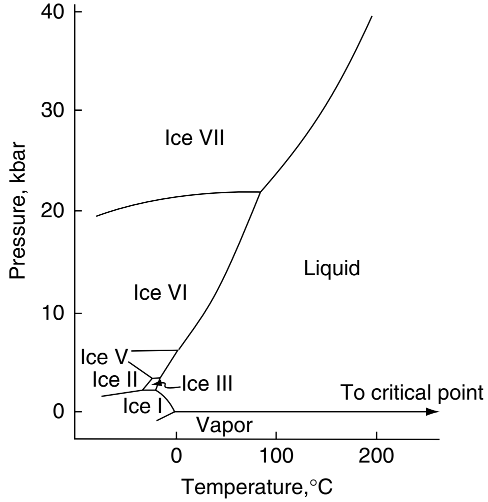


Figure 17.7 Phase diagram for H_2O at high pressures. Ice IV, not shown, is a metastable form of ice in the region of Ice V.

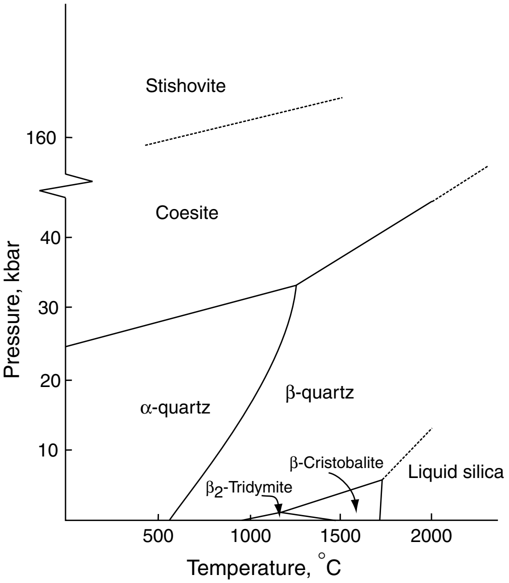


Figure 17.8 Phase diagram for SiO_2.

17.3 Binary systems

17.3.1 Types of diagrams

When we consider the phase relations in systems having two components instead of one, we add one dimension to our diagrams. That is, in unary diagrams all phases have the same composition, and so we don't need an axis showing compositions – we can use both dimensions available on a sheet of paper for physical parameters, and we choose *T* and *P*. With two components, we find that phases commonly contain different proportions of these components – they have different compositions. Since this is of great interest, we use one dimension for composition, leaving only one other for either T or P. Most commonly temperature variations are of more interest, and so diagrams showing phase relations on a T–X diagram are very common.[1] This relationship is illustrated in Figure 17.9. In this book, we concentrate on T–X sections, but it is well to realize that there are other varieties of diagrams. Not only do we have P–X sections, but P–T or isoplethal (constant composition) sections, as well as various projections.

17.3.2 The melting relations of two components

Suppose that you now understand unary phase relations very well, but have never encountered binary systems, and you are given the following problem. There are two minerals, A and B. We know the melting point of each mineral, $T_{\mathrm{m_A}}$ and $T_{\mathrm{m_B}}$ at atmospheric pressure. We grind samples of A and B together in various proportions, say 25% A, 75% B; 50% A, 50% B; and 75% A, 25% B, and we perform experiments to determine the melting temperature of these mixtures. Your job is to draw a diagram *predicting* the most likely results. The diagram should show temperature as the vertical axis, and composition as the horizontal axis, and of course the known melting temperatures of the pure minerals A and B should be plotted on the vertical axes at each end of the composition axis.

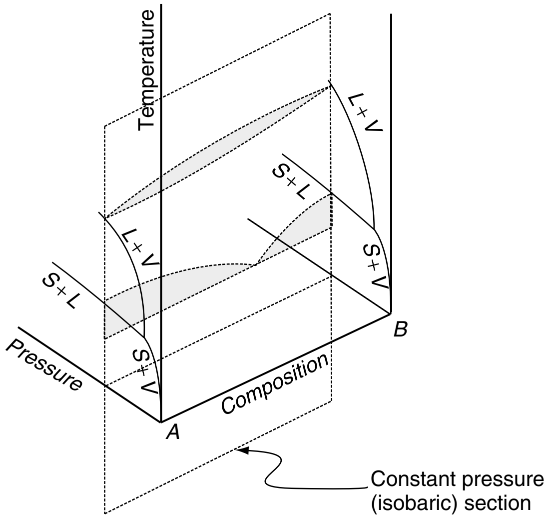


Figure 17.9 The P–T–X box. Most binary phase diagrams are T–X sections through this box. The phase relations will, of course, vary with the pressure chosen for the section. $P = 1$ bar is the commonest choice.

---

[1] We use the "X" in the expressions T–X or P–T–X to mean "composition" generally, whether measured as mole fractions or weight percent, or in some other way.

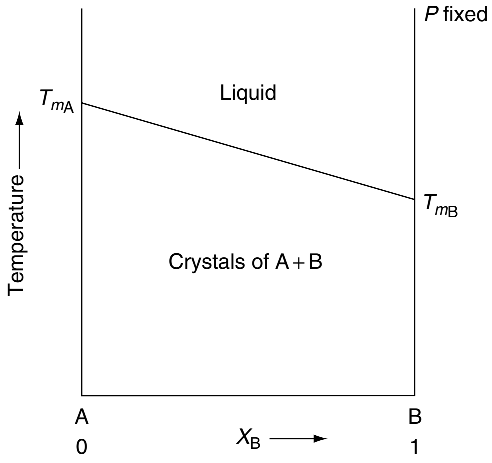


Figure 17.10
An uneducated guess as to the melting temperatures of mixtures of minerals A and B.

It seems very likely that your guess would look like Figure 17.10. In other words, you would probably suppose that the melting temperature of mixtures of A and B would be some kind of average of the melting temperatures of the pure compounds, much in the way that volumes are averages as shown in Figure 10.2a. But binary systems are not quite that simple. Figure 17.10 is thermodynamically impossible, even if A and B were not separate phases but formed a solid solution, but we will not bother to prove this. Suffice it to say that experiments on hundreds of binary systems have never given results consistent with Figure 17.10.

What *does* happen depends on what compounds A and B actually are. Let's suppose that A is the component $CaMgSi_2O_6$, and B is the component $CaAl_2Si_2O_8$. The stable forms of these components at ordinary temperatures are the minerals diopside ($CaMgSi_2O_6$) and anorthite ($CaAl_2Si_2O_8$), and so we will represent component $CaMgSi_2O_6$ by the symbol Di and component $CaAl_2Si_2O_8$ by the symbol An. Diopside melts at 1392 °C, and anorthite melts at 1553 °C. We perform the experiments mentioned above, that is, we grind up both samples into fine powders, then mix the powders in various proportions and heat them up in separate experiments and observe what happens at various temperatures. What a surprise – we find that *all mixtures begin to melt at the same temperature!* And when we analyze the composition of the first liquid to form, we find that *the first liquid to form in all mixtures has the same composition!* The temperature is 1274 °C (called the *eutectic temperature*), and the composition is 42% An, 58% Di (called the *eutectic composition*). On heating to still higher temperatures, another surprise awaits us. For those mixtures having *more* than 42% An, temperatures above 1274 °C result in disappearance of all diopside in the mixtures – we are left with only liquid plus anorthite crystals. For those mixtures having *less* than 42% An, temperatures above 1274 °C result in the disappearance of all anorthite in the mixtures – only liquid plus diopside crystals are left. So *below* 1274 °C, only two phases coexist – crystals of anorthite and diopside. *Above* 1274 °C, again only two phases coexist – either liquid and anorthite, or liquid and diopside. Only at 1274 °C exactly are three phases observed to coexist at equilibrium – anorthite, diopside, and liquid. And note that in the binary system, we have melting far below the melting temperature of either of the pure components. These relationships are summarized in Figure 17.11. They may seem strange at first, but as we will see, they are one of a rather small set of relationships that satisfy the phase rule.

The isobaric phase rule

But, first, we must mention a slight modification of the regular phase rule, Equation (11.1). As shown in Figure 17.9, the experiments we are discussing at a fixed pressure of 1 bar can be represented on a plane or section through *P–T–X* space. The general phase rule (11.1) applies to this *P–T–X* space. The fact that we confine ourselves to a fixed *P* plane within this space means that we have "used" one of our degrees of freedom – we have chosen $P = 1$ bar, and the same would be true for any other constant *P* section (or constant *T* section, for that matter). Therefore *on our T–X plane* the phase rule is

$$f = c - p + 1 \tag{17.1}$$

This shows that the maximum number of phases that can coexist at equilibrium in a binary system at an arbitrarily chosen pressure (or temperature) is three ($p = 3$ for $c = 2$, $f = 0$), which is consistent with our observations.

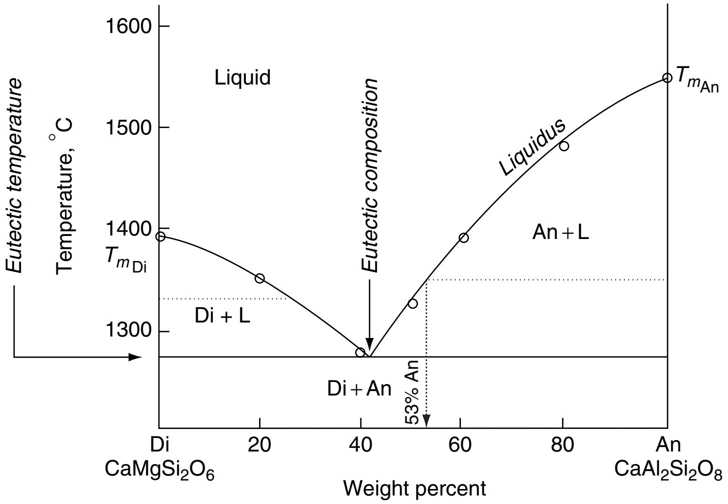


Figure 17.11 The system Di–An at 1 bar pressure. Two representative tie-lines are shown.

17.3.3 Reading the binary diagram

The main features of the phase relations in Figure 17.11 follow directly from this fact. During the heating of our mixture of Di and An crystals, we have two phases, and

$$\begin{aligned} f &= c - p + 1 \\ &= 2 - 2 + 1 \\ &= 1 \end{aligned}$$

This means that to fix all the properties of both kinds of crystals, we need only choose the temperature (pressure being already fixed at 1 bar). However, when the first drop of liquid forms, $p = 3$ (diopside crystals, anorthite crystals, and liquid), and $f = 0$. Another word for $f = 0$ is *invariant*. When $p = 3$ on an isobaric plane, we have *no* choice as to T, P, or the compositions of the phases – they are all fixed. This explains why all mixtures begin to melt at the same temperature, and why the liquid formed is always the same composition no matter what the proportions of the two kinds of crystals. No other arrangement would satisfy the phase rule.

A line on a phase diagram joining points representing phases that are at equilibrium with each other is called a *tie-line*. Each of the two-phase regions in Figure 17.11 (Di + L; An + L; Di + An) is filled with imaginary tie-lines joining liquid and solid compositions, or two solid compositions, that are at equilibrium. Only two of these tie-lines are shown. Consider the tie-line at 1350 °C in the region labeled An + L. One end of the line is on the curved line representing liquid compositions (called the *liquidus*), and the other end is on the vertical line representing 100% An composition. The composition scale across the bottom of the diagram applies at any temperature, so we can get the liquid composition by dropping a perpendicular from the liquidus to the composition scale, showing that the liquid composition at 1350 °C in equilibrium with pure anorthite crystals is 53% component An, 47% component Di. The composition of the solid phase is given by the other end of the line, which is at 100% An. In each of these two-phase regions, such as An + L, $f = 1$, which means that once we have chosen the temperature, say 1350 °C, all properties of all phases are fixed. Therefore, all proportions of Di and An in this region will have the same liquid and solid compositions. In other words, *any* starting mixture of diopside and anorthite crystals having more than 53% anorthite would, when heated to 1350 °C, consist of a liquid of composition 53% An, 47% Di, plus crystals of pure anorthite. Mixtures having between about 20% An and 53% An would be completely liquid at this temperature, and mixtures having 0–20% An would consist of pure diopside crystals plus a liquid of composition 20% An, 80% Di.

By imagining tie-lines across the An + L region at successively higher temperatures, we see that the composition of the liquid in equilibrium with anorthite crystals gets progressively richer in component An. Similarly, the tie-lines in the Di + L region show that the liquid gets richer in Di as temperature increases.

Because the temperature of the three-phase tie-line is fixed, it follows that both above and below this temperature there must be regions having only two phases. We already know that below the three-phase line the two phases are Di and An. Above the three-phase line one of the phases must be liquid, because melting has started. Therefore, the liquid can coexist with only one other phase, obviously in this case either Di or An, but not both. As T increases, the proportion of liquid must increase, eventually becoming 100% liquid. This simple analysis is sufficient to explain the main features of the diagram. "Reading" binary diagrams consists largely of distinguishing between one-phase regions, which have no tie-lines (e.g., the Liquid region), two-phase regions, which have tie-lines joining two phases at equilibrium, and three-phase tie-lines, which separate two-phase regions, and join three phases at equilibrium.

17.3.4 A more general example

The system Di–An is misleadingly simple in two respects. For one thing, the diagram shows that both diopside and anorthite remain pure while heated in contact with the other component until the melting temperature is reached (1392 °C for Di, 1553 °C for An). Actually, phases (in theory) never remain perfectly pure when in contact with other phases – some mutual solution always takes place, although as in the case of Di and An it is sometimes small and does not show on the diagram. A more realistic case is shown in Figure 17.12. The diagram is essentially the same as the Di–An diagram, except that there is a field of A_{ss} and of B_{ss}, where subscript "ss" stands for solid solution.

The other respect in which the Di–An diagram is misleading is that, in fact, it is not strictly speaking a true binary system. This somewhat surprising

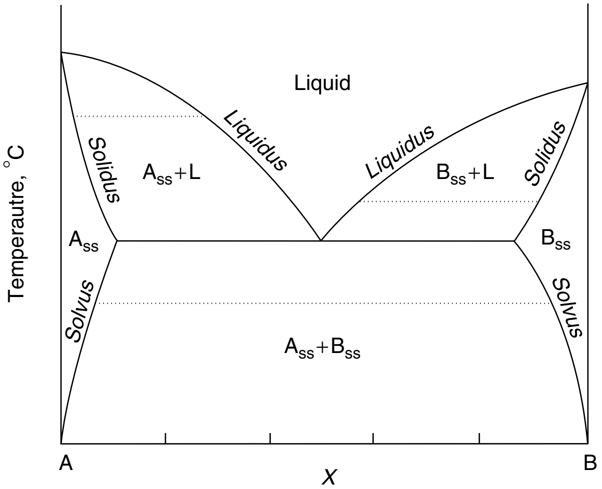


Figure 17.12 A more representative binary system. The difference is that both components show solid solution fields. Three representative two-phase tie-lines are shown.

statement cannot be fully explained without discussing ternary systems. Suffice it to say that just because you choose two components does not necessarily mean that you have a binary system. *To be truly binary, all compositions of all phases must lie on the plane of the diagram.* This must be the case in simple systems such as Cu–Au and with single solution phases such as liquids; but with complex components such as Di and An, although the *bulk composition* must lie on the plane of the diagram, the compositions of coexisting *phases* may each lie "off the plane" of the diagram. Careful work has shown that in the system Di–An, diopside crystals are not pure but contain some Al. This means that, because bulk compositions lie on the Di–An plane, phases coexisting with diopside must be somewhat deficient in Al. To portray this in a diagram, one needs a three-component triangle. Just remember that not all choices of two components are binary systems – some are planes within ternary systems.

Solid solutions

There is no difference in principle between a solid solution and a liquid or gaseous solution. Substances dissolve into one another, like sugar into tea, or like oxygen into nitrogen, because the Gibbs energy change of such a process is negative – they are spontaneous processes. Consider the system Di–An at a temperature of 1600 °C (Figure 17.11). At this temperature, both pure Di and pure An are liquid phases. If one gram of Di liquid and one gram of An liquid were mixed together, they would dissolve into one another to form a homogeneous liquid solution, represented by a point in the middle of the diagram on the 1600 °C isotherm. If pure diopside crystals are mixed with pure anorthite crystals at 1200 °C, on the other hand, nothing happens – they do not dissolve into each other.

Components A and B in Figure 17.12, on the other hand, behave differently. Liquid A and liquid B still mix to form a homogeneous solution, but when solid A and solid B are mixed together, they dissolve into one another to a limited extent. Salt will dissolve into water, but not without limit – it will dissolve only until the water becomes saturated. Similarly, solid B will dissolve into solid A, but not without limit. It dissolves into A until A is saturated with B, and at the same time A dissolves into B until B is saturated with A. The saturation limits of each component in the other is shown by a line called the *solvus*. The existence of a solvus shows that A and B exhibit *limited miscibility* in the solid state. They exhibit *complete miscibility* in the liquid state. "Miscibility" does not really mean "mixability," although they sound similar. "Mixability," if it is a word, just means things can be mixed together – mutual dissolution is not implied. "Miscibility" means the ability to dissolve into something else.

Figure 17.12 shows a eutectic, but the two solid phases in equilibrium with the liquid are not pure A and pure B; A contains some B in solid solution (A_{ss}) and B contains some A in solid solution (B_{ss}). Similarly, at temperatures above the eutectic, the liquid is not in equilibrium with pure A or pure B, but with

A_{ss} and B_{ss}. The compositions of the solid solutions in equilibrium with liquid are given by lines called the *solidus*.

17.3.5 Freezing point depression

Figure 17.11 shows that mixtures of diopside and anorthite become completely liquid at temperatures lower than the melting temperatures of either pure diopside or pure anorthite. This is also shown by the more general system in Figure 17.12 and is, in fact, an extremely common feature of binary systems. It is called *freezing (*or *melting) point depression* and is, in fact, why we put salt on icy roads in winter. The melting temperature of ice is lowered in the presence of the second component (salt, NaCl), and so the ice melts and the resulting salty water corrodes our cars. But why is the freezing point depressed?

The answer is found in the basic thermodynamic relationships between the phases. Figure 17.4 shows the absolute Gibbs energies of the solid, liquid, and vapor phases of our compound α as a function of temperature at a pressure of 2 mbar. Let's call compound α component A (just as we called component $CaMgSi_2O_6$ Di). Figure 17.4 therefore shows G_A^{solid}, G_A^{liquid}, and G_A^{vapor} as a function of T. If we now add a second component B to A, what happens to these Gibbs energies? To start with the simplest case, we will suppose that B does not dissolve into solid A or into vapor A, but it does dissolve into liquid A (if you add NaCl to H_2O, the salt will not dissolve into ice or into steam, but it will dissolve into liquid water). Therefore, the curves for G_A^{solid} and G_A^{vapor} will not change, because A^{solid} and A^{vapor} are unchanged in the presence of B. But what is the Gibbs energy of component A in a liquid containing both A and B?

The answer is shown in Figure 7.4, which shows that when B dissolves into A, the molar Gibbs energy of component A in the solution at a given concentration (which we call μ_A), is *lower* than the molar Gibbs energy of pure A (G_A°). This relationship is quite general and without exception, because otherwise A and B would not form a solution. We will be mentioning this relationship at various points throughout this chapter. The consequence of the fact that G_A^{liquid} is lowered but G_A^{solid} is not is shown in Figure 17.13. The shaded surface in this figure represents the free energy curve from Figure 7.4, extended into a range of temperatures. It shows the lowering of the total Gibbs energy of the liquid phase as component B is added. At the arbitrary amount of 10% B, a tangent surface to the free energy surface extends back to the 0% B plane, analogous to the tangent at $X_B = 0.4$ in Figure 7.4. The trace of this tangent surface on the G–T section for component A gives μ_A, the molar Gibbs energy of A in the solution containing 10% B, 90% A. It of course lies below the curve of G_A^{liquid} for pure A. But because the curve for G_A^{solid} has not moved, the *intersection* of the G_A^{solid} and μ_A^{liquid} curves is moved to a lower temperature. The intersection of the μ_A^{liquid} and G_A^{solid} curves is the point where these two

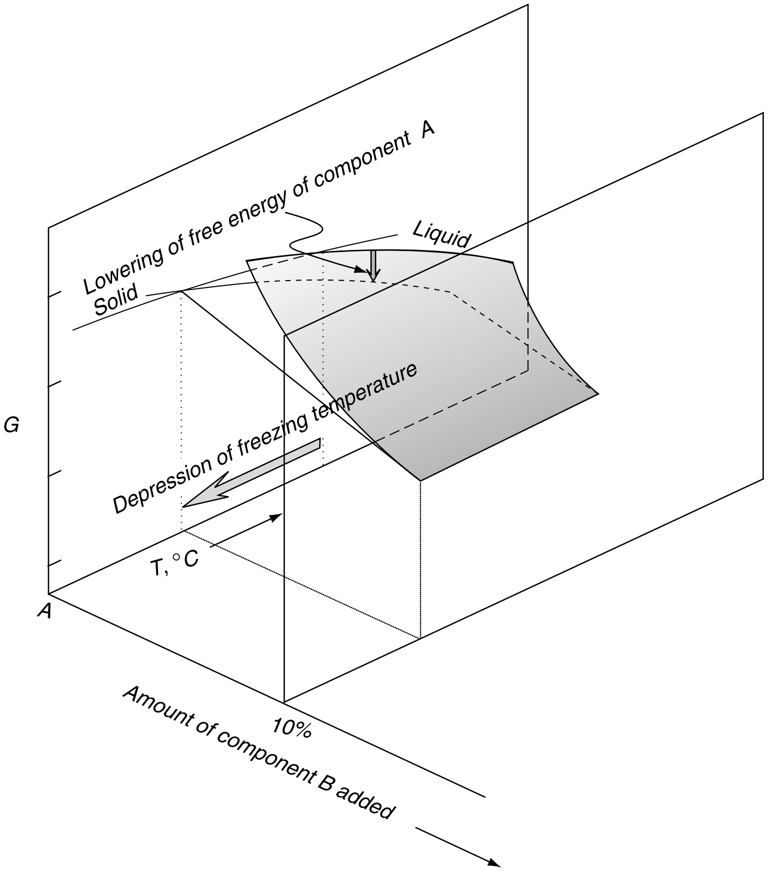


Figure 17.13 The *G*–*T* plane is taken from Figure 17.14. Component B enters the liquid phase and causes a lowering of μ_A^{liquid}, which in turn causes a depression of the freezing temperature.

quantities are equal, and it defines the temperature at which A in the solid state and A in the liquid state are in equilibrium. For pure A, this is the melting or freezing temperature; for the system A–B, it defines a point on the liquidus of A and is the result of freezing point depression.

This relationship is shown in again in Figure 17.14, this time including the vapor curve. If the vapor curve does not move (no B dissolves into vapor A), depression of G_A^{liquid} results in a raising of the boiling temperature as well as a lowering of the freezing temperature. This is also an extremely common effect.

17.3.6 Freezing point elevation

But suppose our simplifying assumption that no B enters the solid phase is not true. There is no difference in principle between the thermodynamics of solid and liquid solutions, and so if B dissolves into solid A the curve for G_A^{solid} will be lowered for the reasons just discussed. Normally, B is less soluble in solid A than in liquid A, and so the amount of lowering is less for the solid

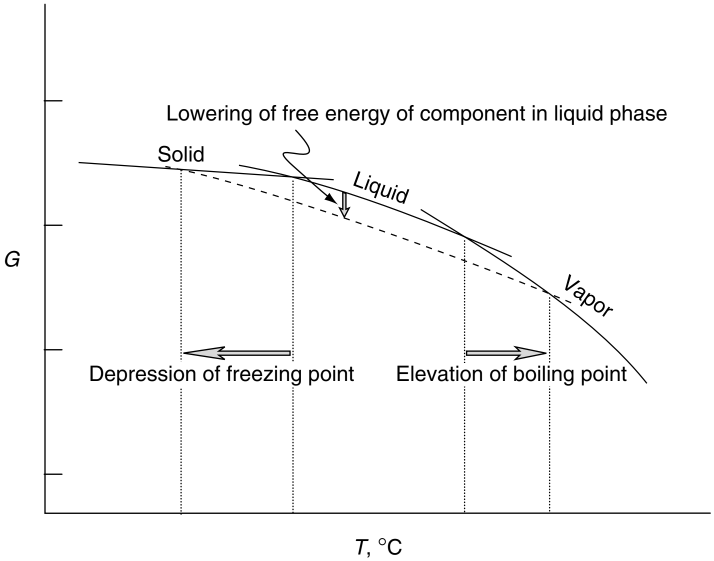


Figure 17.14 *G–T* section, showing lowering of the Gibbs energy of A in the liquid phase, causing depression of the freezing point and elevation of the boiling point.

phase, and the freezing point is still lowered. This is shown by systems like that in Figure 17.12, where the liquidus of A slopes downward, even though B is shown as entering both the liquid and the solid phases of A.

However, what of the possibility that the G_A^{solid} curve might be lowered *more* than the G_A^{liquid} curve? This would happen if more B dissolved into solid A than into liquid A and would result in a *freezing point elevation* as shown in Figure 17.15. This explains an important feature of many binary systems.

17.3.7 Systems having complete solid miscibility

A and B in Figure 17.12 show limited solid miscibility, but some important systems show *complete* miscibility in the solid state, giving rise to a diagram that looks quite different, as shown in Figure 17.16. In a sense, it is simpler than the ones we have looked at so far – in fact it looks rather like Figure 17.10, except that the “melting line” in Figure 17.10 is a *melting loop* in Figure 17.16. But the most important difference is that in Figure 17.16 A and B *dissolve completely into one another in the solid state*. This takes some getting used to. We are quite familiar with sugar dissolving into tea, but the idea of placing two solid objects together and observing one disappear into the other is not something in our experience. But this is just another example of something that thermodynamics says *should* happen but in fact does not, because of energy barriers. The thermodynamic model does not consider these barriers, and so does not always work. These solid solutions do exist, however, because they do not form from solids dissolving into one another at low temperatures. They form at high temperatures, sometimes over long periods of time, and then cool down in their mutually dissolved state. Many important alloys and mineral

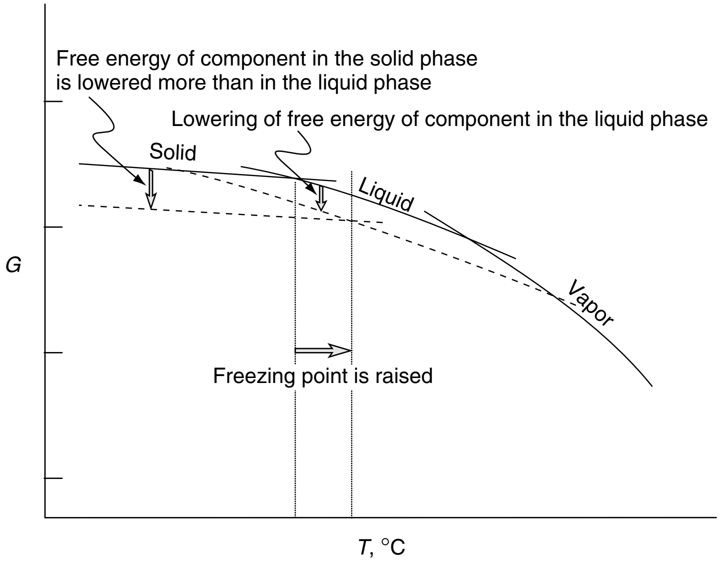


Figure 17.15 *G*–*T* section, showing a greater lowering of the solid Gibbs energy than liquid Gibbs energy, causing elevation of the freezing point.

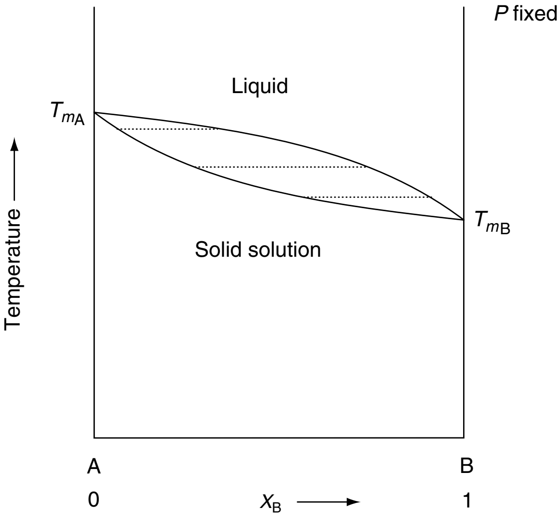


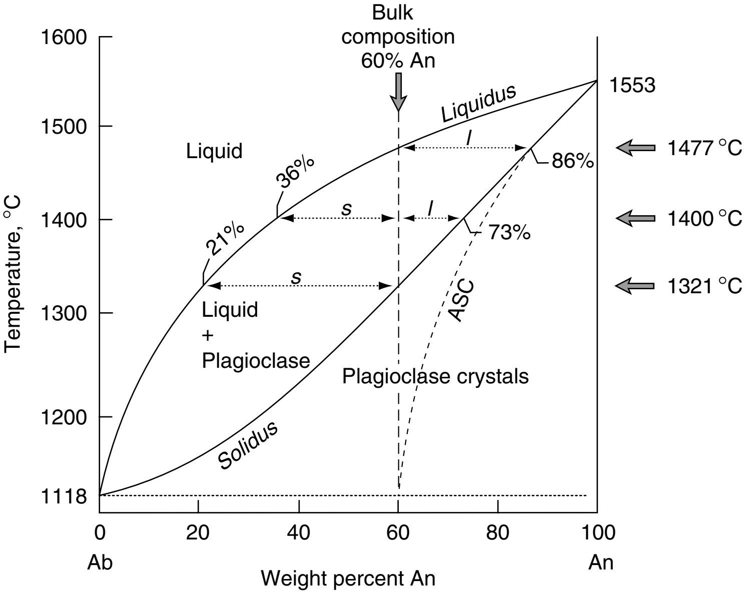
Figure 17.16 A binary system A–B showing complete miscibility in both the solid and liquid states. Three representative tie-lines are shown.

groups are such complete solid solutions, including the feldspars, olivines, and some pyroxenes and amphiboles.

Note that in Figure 17.16, T_{m_A} is lowered by adding B, but T_{m_B} is raised by adding A. This is because more B enters liquid A than solid A, but more A enters solid B than liquid B, and the free energy consequences of this are shown in Figures 17.14 and 17.15, respectively.

The most important mineralogical example of this type of system is the plagioclase feldspar system, shown in Figure 17.17. Plagioclase is a mineral whose

Figure 17.17
The plagioclase feldspar system at 1 bar pressure. The curve labeled ASC is the average solid composition during fractional crystallization of the 60% An bulk composition.



composition may vary from virtually pure albite ($NaAlSi_3O_8$), or component Ab, to almost pure anorthite ($CaAl_2Si_2O_8$), or component An, depending on the composition of the liquid from which it crystallizes. The melting behavior of a complete solid solution such as this is a simple melting loop – a combined liquidus and solidus that goes from one pure component over to the other. The melting loop is filled with imaginary horizontal tie-lines, three of which are shown in Figure 17.17. They indicate the compositions of liquids, on the liquidus, and the compositions of plagioclase crystals, on the solidus, which are in equilibrium with each other.

17.3.8 Equilibrium versus fractional cooling and heating

Binary phase diagrams show phase compositions that are at equilibrium – they show what you would obtain if you heated a bulk composition to a certain temperature and waited long enough for equilibrium to be attained. The time required to reach equilibrium after a change in temperature or pressure varies greatly with the system, but equilibrium is *never* achieved instantaneously. Therefore, if we use the diagram to consider what would happen during continuous cooling or heating a given bulk composition, we cannot be considering what would *really* happen in our system during cooling or heating; we are considering *model* processes, as usual.

There are any number of models of processes we could devise involving phase changes in binary systems, but two are especially common – complete equilibrium (reversible) processes, and "surface equilibrium" (perfect fractional) processes. We will discuss only cooling processes. Heating processes are the exact reverse of cooling processes in the equilibrium case, but not always the exact reverse in the case of fractional processes.

Perfect equilibrium crystallization

Suppose we had a liquid having a composition of 60% An, 40% Ab at a temperature of about 1600 °C (Figure 17.17). On cooling this liquid, nothing much happens (except that the properties of the liquid, such as its density, refractive index, entropy, Gibbs energy, etc., change) until it reaches a temperature of 1477 °C, the liquidus temperature for this composition. At this point, the bulk composition is still 100% liquid, but the first tiny crystal of plagioclase appears. Its composition, given by the solidus, is 86% An, 14% Ab. As cooling continues, plagioclase crystals continue to form, and previously formed crystals change their composition so that all crystals always have the equilibrium composition, with no compositional gradients. When the temperature reaches 1400 °C, the liquid has composition 36% An, and the crystals 73% An. (These compositions are obtained by dropping a perpendicular line from the point of interest to the compositional axis at the bottom of the diagram.) When the composition of the solids reaches the bulk composition of 60% An, the liquid must disappear, and this happens at a temperature of 1321 °C. Further cooling results in no further changes in composition of the crystals.

Perfect fractional crystallization

Maintaining perfect equilibrium while cooling is one end of a complete spectrum of possibilities. The other end of the spectrum is that crystals form, but always completely out of equilibrium. This end of the spectrum involves an infinite number of cases and so is rather difficult to discuss in a finite number of words. A subset of these possibilities is the case where crystallization produces crystals in equilibrium with the liquid, as required by the diagram, but after forming, they do not react with the liquid in any way. This is called surface equilibrium (because the liquid is at all times in equilibrium with the surface of the crystals) or fractional crystallization, and is a model process just as much as is equilibrium crystallization. It is also used in connection with liquid–vapor processes (fractional distillation; fractional condensation), as well as isotope fractionation processes.

There are two ways of imagining a process of perfect fractional crystallization.

- As soon as a tiny crystal forms, it is removed from the liquid. This might be by reaching into the liquid with a pair of tweezers and physically removing the crystal, or the crystal might immediately sink to the bottom or float to the top of the liquid, where it becomes covered by other crystals and is removed from contact with the liquid.

- As soon as a tiny crystal forms with a composition given by the solidus, it is covered by a layer of another composition, given by the solidus at a slightly lower temperature. Successive layers are formed, each controlled by the position of the solidus, but after forming, the various layers do not homogenize in the slightest. The result is a compositionally zoned crystal.

Note that crystals are removed from contact with the liquid in both cases. This is the essential element of fractional crystallization.

Considering the same bulk composition, 60% An, the cooling history is the same as before until the first tiny crystal forms at 1477 °C, having a composition of 86% An. On further cooling, the liquid composition follows the liquidus, as before, and any *new* crystals that form have compositions given by the solidus at that temperature; but previously formed crystals, being removed from contact with the liquid, do not change their original compositions. The net result is that at any temperature below 1477 °C, the *average* composition of all solids formed is more An-rich than would be the case in equilibrium crystallization, that is, more An-rich than the solidus at that temperature. Because of this, at each temperature below 1477 °C, there must be a larger proportion of liquid of Ab-rich composition to balance the solid composition, that is, to give the known bulk composition. Therefore, whereas in equilibrium crystallization the last drop of liquid must disappear at 1321 °C, in fractional crystallization it does not, and in fact liquids continue to exist right down to pure Ab composition, where the last liquid crystallizes pure albite. This is the important aspect of fractional crystallization from a petrological point of view – that a given bulk composition can generate a much wider range of liquid compositions, and hence a wider range of igneous rocks, than can equilibrium crystallization.

It is possible to calculate the average composition of the solids during fractional crystallization, but we will not do this. Just note that a curve indicating the average composition of all solids generated must begin at 1477 °C on the solidus, and it must end at 1118 °C at a bulk composition of 60% An, when the last liquid disappears. This curve is shown in Figure 17.17, labeled “ASC.” For equilibrium cooling, the “ASC” curve is, of course, the same as the solidus.

17.3.9 The lever rule and mass balances

Phase diagrams contain information not only about phase compositions and their temperatures and pressures, but about the *proportions* of phases for a given bulk composition. This is done using what is called the lever rule. Look at the three tie-lines we have just been discussing in Figure 17.17. Consider first the line extending from the liquidus (36% An) to the solidus (73% An), at 1400 °C. This line is composed of two parts. One part, labeled l, represents the proportion of liquid, and the other part, labeled s, represents the proportion of solids. The fraction (by weight) of liquid in the bulk composition is thus $l/(l+s)$, and the fraction of solids is $s/(l+s)$. The easiest way to measure the lengths of l and s

is probably by comparing the compositions of the end points of the tie-line at the liquidus and solidus with the bulk composition. Thus the *s* portion of the tie-line has a length of $60-36=24\%$, and the *l* portion of the tie-line has a length of $73-60=13\%$. The total length of the tie-line is $73-36=37\%$. Therefore the proportion or fraction of solid in the total bulk composition is $24/37=0.65$, and the fraction of liquid is $13/37=0.35$, and of course $0.65+0.35=1.0$. If we had a bulk composition weighing 10 g, then at 1400 °C, 1 bar, it would be made up of $0.65\times 10=6.5$ g of crystals (73% An composition), and $0.35\times 10=3.5$ g of liquid (composition 36% An). This lever rule can be used in any two-phase region, given the bulk composition, and the lengths of the lines can be measured in % composition as we have done, or in millimeters, or inches, or any other units.

Note that at intersection of the bulk composition line and the liquidus, the *s* portion of the line reduces to zero, because there is 100% liquid, and at the intersection of the bulk composition line and the solidus, the *l* portion of the line reduces to zero. This provides a way of remembering which side of the tie-line represents which phase.

Mass balances

Because we know the proportions and compositions of the phases, it is a simple matter to combine these to calculate the bulk composition. But, as we *know* the bulk composition (we needed it to get the phase proportions), the calculation is circular. Nevertheless, it is a useful check on our reading and construction of diagrams. For example, at 1400 °C in Figure 17.17, the mass balance is

$$\begin{gathered}\text{(solid fraction}\times\text{solid composition)}\\+(\text{liquid fraction}\times\text{liquid composition})=\text{bulk composition}\\(0.65\times 73)+(0.35\times 36)=60\end{gathered}$$

where 60 is the bulk composition in % An.

17.3.10 Binary *G–X* sections

The fact that the Gibbs energy of solutions is represented by a convex-downward curve (a "festoon") was introduced in §10.6 and Figure 7.4. Both solid solutions and liquid solutions are represented by such curves, and understanding of binary diagrams is increased by constructing such curves on *G–X* sections. Each *G*-curve moves upward with decreasing temperature ($\partial G/\partial T=-S$), but the liquid curve moves upward faster than the solid curve, because the entropy of liquids is greater than solids. The stable phase for any bulk composition is always indicated by the lowest *G*, either on a solid or liquid curve or on a tangent joining two such curves.

A binary eutectic system

Consider first the Di–An system in Figure 17.18, an example of a simple binary eutectic system. The T–X section is shown at the bottom of the diagram, and G–X sections through the system at various temperatures are shown above it. In understanding this diagram, it is important to remember that the G of a mixture of crystals that are completely immiscible (show no mutual solid solution) is simply the weighted average of the G of the two pure end-members, which in this case appears as a straight line joining G_{Di} and G_{An}. This straight line appears on all sections, whether the solids are stable phases or not. The line labeled "mixture of Di and An crystals" in Figure 17.18 represents this situation.

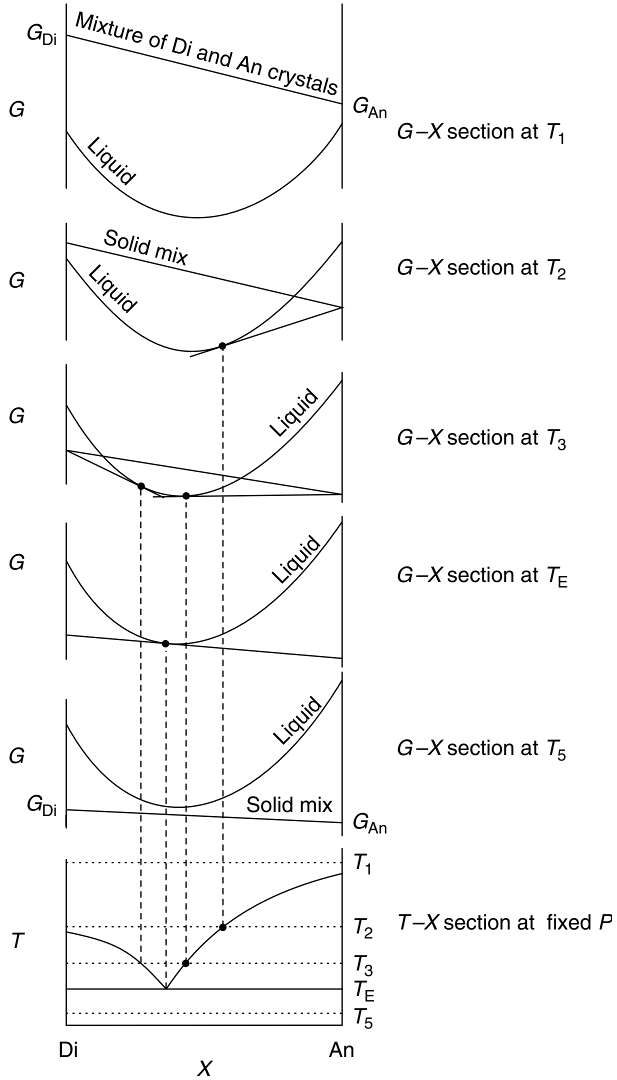


Figure 17.18
G–X sections through a binary eutectic diagram.

When components *do* mutually dissolve to form a solution, Equation (10.4) still works for volumes if the solution is ideal, but even for ideal solutions, it does not work for G. The Gibbs energy of the solution must be less than the weighted average of the G of the two pure end-members for mutual solution to take place, and it is represented by the "festoon," or convex-downward loop. Therefore, the liquid solution formed when liquid Di and liquid are mixed together is represented by such a loop. It is shown in all sections, even when the liquid is not the stable phase.

Understanding these G–X sections is helped by realizing that

- The line representing the G of the mixture of solid Di and solid An is shown in every section regardless of whether diopside or anorthite is or is not stable. If they are not stable, the line represents *metastable* Gibbs energies.
- The line (festoon) representing the G of the complete liquid solution between Di and An is shown in all sections regardless of whether the liquid is stable or not. If the liquid is not stable, its curve represents *metastable* Gibbs energies.
- The stable phase or phases in each section must have the lowest free energy available.

Starting with the G–X section at T_1 at the top of the diagram, we see that for every bulk composition between Di and An, the liquid Gibbs energy is everywhere *lower* than the Gibbs energy of a mixture of solid Di and solid An. In other words, the stable phase across the diagram is liquid. On the T–X section, note that at T_1, we are above the melting temperatures of both components, and in the field of liquid at all compositions.

As we cool from T_1 to T_2, the Gibbs energy of liquids and solids (whether stable or metastable) increases, but that of liquids increases more, so that at T_2, the G of liquid An has become greater than that of solid An, but the G of liquid Di remains less than that of solid Di. At some temperature between T_1 and T_2 we must have passed a point where $G_{\mathrm{An}}^{\mathrm{liquid}} = G_{\mathrm{An}}^{\mathrm{solid}}$, that is, the melting temperature of An. At T_2, the stable form of pure Di is liquid, but the stable form of pure An is solid. The lowest Gibbs energy available to the system as we go from Di toward An is liquid, but just after passing the minimum on the curve, the lowest Gibbs energy available is neither liquid nor a mixture of crystals, but a mixture of liquid and An crystals. In this mixture, the Gibbs energy of component An in the crystals must be the same as the Gibbs energy of component An in the liquid.

Recall from Figure 7.4 that the tangent to a Gibbs energy curve of a solution has intercepts giving the chemical potential of each component in the solution. Therefore, a tangent to the liquid curve that has an intercept on the An axis at the Gibbs energy of solid An will indicate that liquid composition in which $\mu_{\mathrm{An}}^{\mathrm{liquid}} = \mu_{\mathrm{An}}^{\mathrm{solid}}$. That tangent point is, of course, at the composition of the liquidus at that temperature, as shown by the dotted line joining the G–X section at T_2 with the T–X section. As the temperature falls below T_2, that tangent, rooted on the An axis at the free energy of solid An ($G_{\mathrm{An}}^{\mathrm{solid}}$), moves to greater Di compositions, because the liquid loop is moving up with respect to the $G_{\mathrm{An}}^{\mathrm{solid}}$ point.

At T_3, solid diopside is now the stable phase on the Di side of the diagram, and the tangent situation described above holds for both components. With falling temperature, the two tangent points move toward each other, becoming one tangent at T_{E}, the eutectic temperature. Note that at T_{E} there must be only one tangent because μ_{Di} and μ_{An} must be the same in all three break phases.[2]

On further cooling, the tangent breaks away from the liquid curve and becomes a straight line below the liquid curve, giving the Gibbs energy of a mixture of diopside and anorthite crystals just as in the section at T_1. The difference is that now it is completely *below* the liquid curve, and therefore a mixture of crystals is the stable configuration of the system.

A melting loop system

The story is rather similar for G–X sections through a melting loop diagram at various temperatures (Figure 17.19). However, instead of dealing with the intersection of a solution curve, or festoon, and a straight line, we have the intersection of two solution curves – one for the liquid solution and one for the solid solution. If this seems confusing, go back to §10.6 and recall why a Gibbs energy curve for a solution must be convex downward. Then remember that this applies whether the solution is solid, liquid, or gaseous. Finally, remember that in these sections (Figure 17.19) we plot the positions of both solution curves in every section, regardless of whether the solution is the stable phase or not. The point is to determine which parts of which curves give, or combine to give, the lowest Gibbs energy available to the system at each composition across the system.

At the top of the diagram, the section is drawn at the melting temperature of anorthite crystals. Therefore, the liquid and solid curves join at the An axis, because $G_{\mathrm{An}}^{\mathrm{solid}} = G_{\mathrm{An}}^{\mathrm{liquid}}$. Going toward component Ab, the liquid curve is everywhere below the solid curve, showing that liquid is everywhere the stable phase at this temperature. Similarly, the section at $T_{m_{\mathrm{Ab}}}$ shows that $G_{\mathrm{Ab}}^{\mathrm{solid}} = G_{\mathrm{Ab}}^{\mathrm{liquid}}$, and that at compositions toward component An, the solid curve is below the liquid curve, showing that at this temperature, a solid solution of Ab and An is the stable phase. At intermediate temperatures, the two curves intersect. The liquid curve is lower on one side, and the solid curve is lower on the other side. Intermediate compositions have the lowest possible Gibbs energy only by being a mixture of solid and liquid, and because the chemical potentials of both Ab and An must be the same in both phases, the compositions of the two solutions at equilibrium must be given by the only common tangent to the two curves at each temperature.

---

[2] If you think about this statement, and look at Equations (7.26), you will see why we said (§17.3.4) that phases never remain absolutely pure when heated together, at least according to our model.

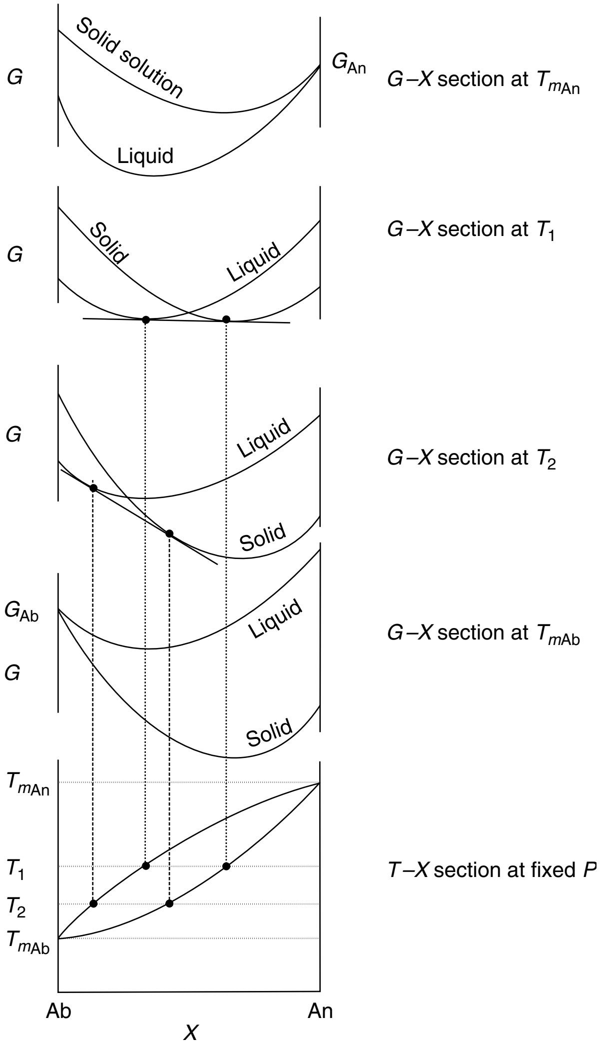


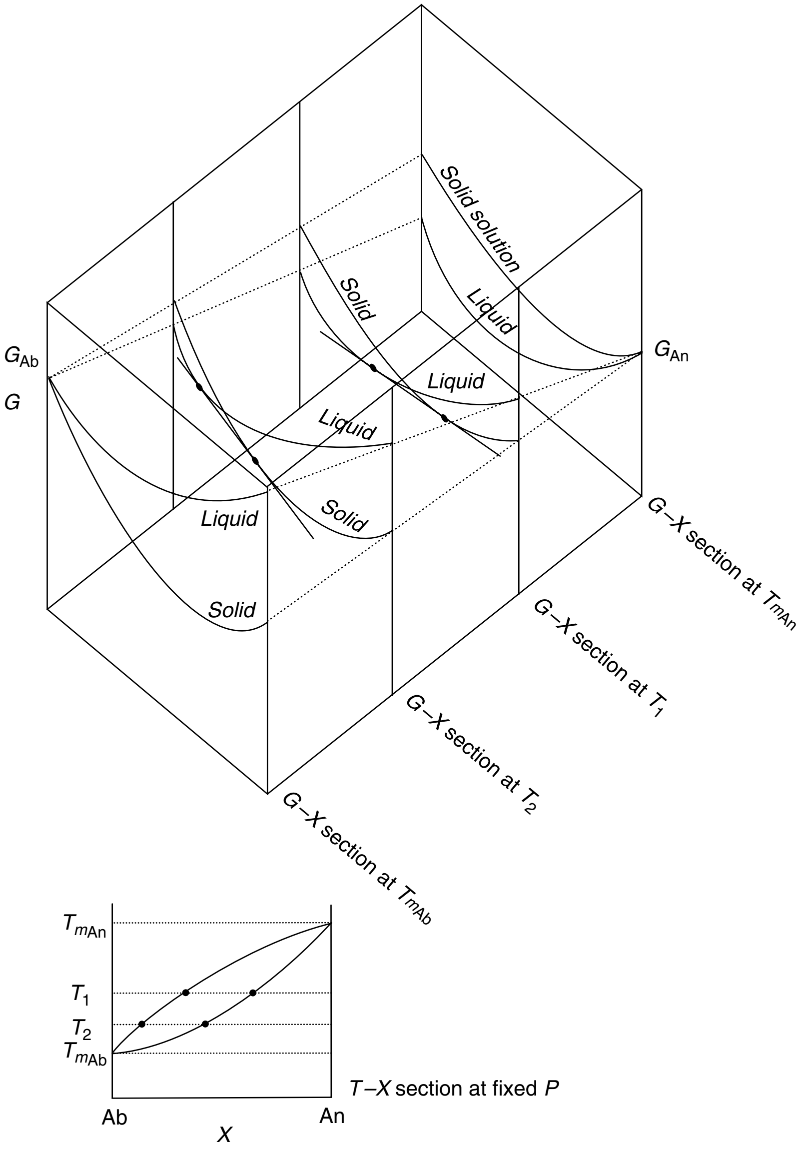
Figure 17.19 G–X sections through a melting loop diagram.

The sections in Figure 17.19 have been reassembled into a three-dimensional view in Figure 17.20. The only major advantage of this is that you can now see the relative rises of the liquid and solid loops with decreasing temperature, as indicated by the dotted lines on the sides of the box. The line representing liquids has a steeper slope than that representing solids for the same reasons as in Figures 17.4 and 6.14.

17.3.11 Binary diagram elements

Binary phase diagrams can become quite complex, but the complexities are nothing but the elements of simpler diagrams, combined in such a way as to

Figure 17.20
A perspective view of G–X sections through a melting loop diagram.



satisfy the phase rule. There are essentially only two elements (Figure 17.21), both of which contain two-phase tie-lines, and hence are elements controlling the compositions of coexisting phases.

- The phase transition loop, which separates two different kinds of solutions. This can be a melting loop as shown previously in Figure 17.16, separating a liquid solution from a solid solution, the end points being melting temperatures. But it can also be a boiling loop, separating a gas or vapor from a liquid, the end points being

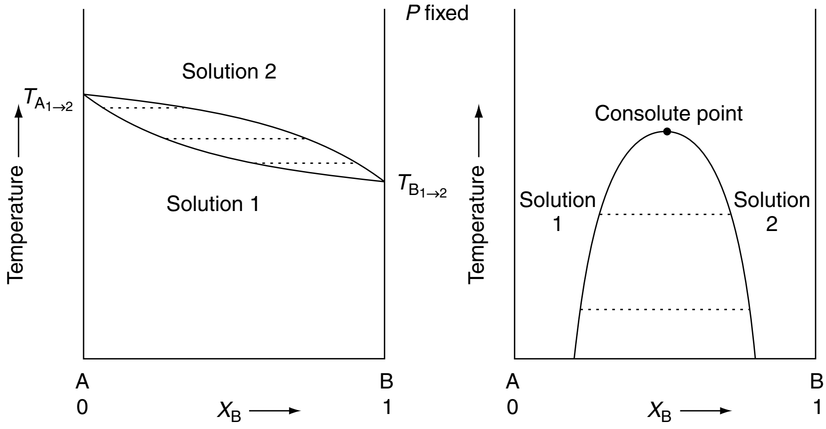


Figure 17.21 The two basic binary diagram elements. In the phase transition loop (left diagram) solution 1 and solution 2 can be solid and solid, solid and liquid, or liquid and vapor, respectively. $T_{A_{1\to 2}}$ and $T_{B_{1\to 2}}$ are melting temperatures, boiling temperatures, or polymorphic phase transition temperatures for pure A and B respectively. Three representative tie-lines are shown. In the solvus (right diagram), solution 1 and solution 2 can be two solids or two liquids. Two representative tie-lines are shown.

boiling temperatures, and it can also be a polymorphic or solid–solid phase transition loop separating two solid solutions having different structures, the end points being polymorphic transition temperatures. Phase transition loops occur simply because solutions cannot change to other solutions with no change in composition.[3]

- The solvus, which separates two solutions of the same kind, such as two solid solutions, or two liquid solutions (Figure 17.21). Increasing the temperature normally increases the solubility of one in the other, and the two phases can become identical (the solvus closes) at an *upper consolute point*. As this name implies, solvi can sometimes close downward (with decreasing temperature), but this is rare. Normally the solvus keeps widening downward.

Figure 17.22 shows some examples of how these two elements combine. For example, a binary minimum melting loop (top of Figure 17.22) can be considered to be produced by combining two simple melting loops. A simple *peritectic* can be considered to be what happens when a solvus intersects a simple melting loop, and a eutectic what happens when a solvus intersects a binary minimum melting loop. You can try to make these intersections with other topologies, but they will generally not obey the phase rule. (The difference between a peritectic and a eutectic is illustrated in Figure 17.23. In both, three phases exist

---

[3] This can be proven using thermodynamics, but we will not bother. The exception to this rule is a solution having a maximum or minimum in temperature, where it can melt or boil to another solution having the same composition (see Figure 17.23, lower left, upper right.)

Figure 17.22 Some simple binary "elements," and how they combine.

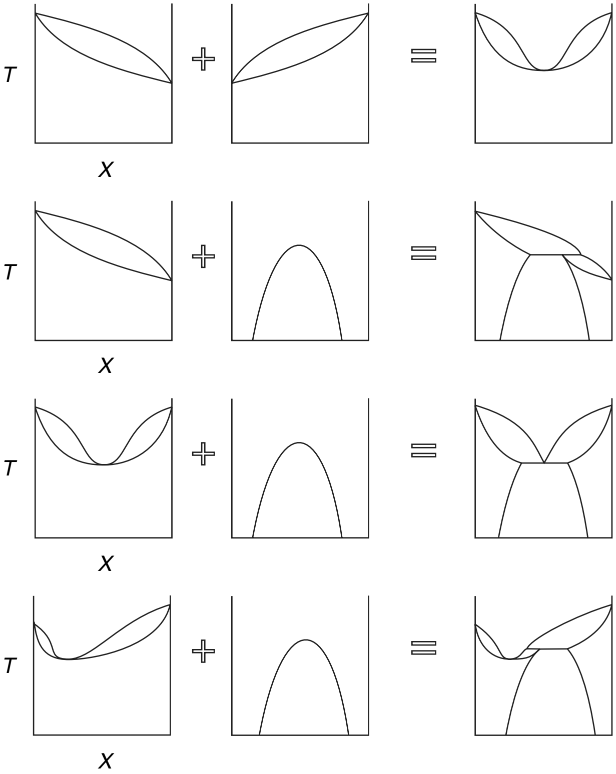
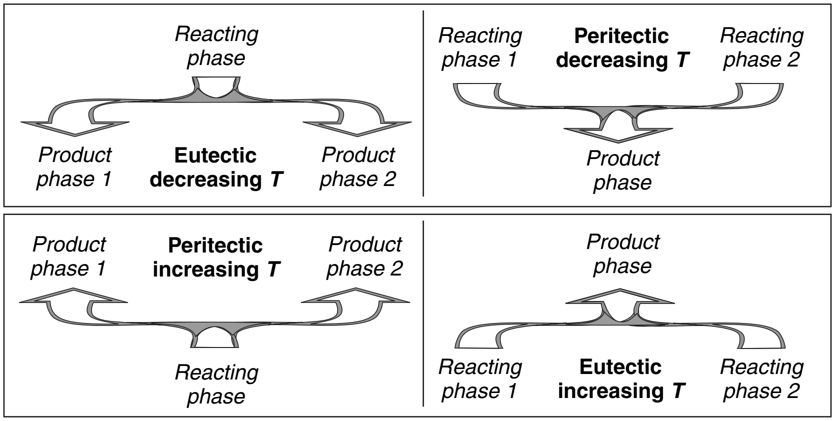


Figure 17.23
The difference between a peritectic and a eutectic.



together at equilibrium, and so both are represented by a three-phase tie-line. However, the reaction relationships are exactly reversed. What happens at a eutectic during cooling is the same as what happens at a peritectic on heating. Which phases are solids and which are liquids is immaterial.)

Similarly, as conditions change (say, increasing pressure), these configurations can become more complex, but without introducing any new features. For example, at the top of Figure 17.24 is an attempt to portray the effect produced by a solvus moving upward, due to changing conditions. At low pressures, the solvus intersects the melting loop, but does not go through it. If the solvus moves upward in temperature faster than does the melting loop, it must eventually poke its way through the top of the melting loop, as shown. Every such intersection must be accomplished with no more than three coexisting phases, so three-phase tie-lines are produced. Try to imagine the top right diagram on Figure 17.24 as a cross-cutting melting loop and solvus, perhaps with dotted lines completing the individual elements. Then try to satisfy the phase rule in some other way; you will find it difficult.

Finally, note that as far as the phase rule is concerned, one phase transition is much like another. Thus an α–β polymorphic transition in a single component behaves just like a melting point or a boiling point when a second component is added. Because one polymorph will in general have a greater capacity for the second component than the other polymorph, a polymorphic phase transition loop is created in the binary system, exactly analogous to boiling and melting loops. These phase transition loops may extend from side to side of the diagram in a completely miscible solid solution, but more likely they will intersect a solvus, as shown at the bottom of Figure 17.24. As before, a three-phase tie-line is created by every such intersection.

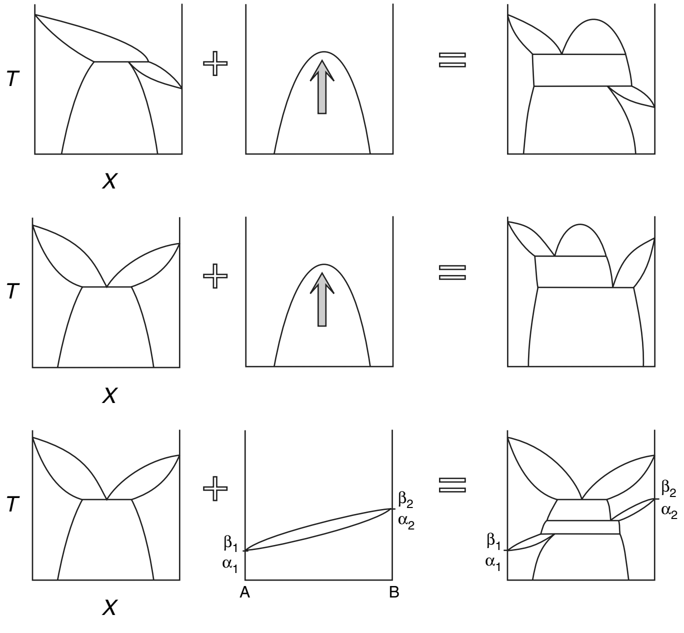


Figure 17.24 What happens when a solvus moves upward through melting loops, and the effect of adding a polymorphic transition.

It is possible to have a four-phase tie-line in a binary system, but this could only be at a unique temperature and pressure, just like a triple point in a unary system. Binary sections are not usually drawn for such unique conditions. That is, when we chose our pressure for our *T–X* section in §17.3.2 (the isobaric phase rule), it would be extremely unlikely for this choice to be just the pressure needed for four-phase equilibrium, and so three-phase tie-lines are the norm in binary sections.

17.3.12 Cooling curves

Temperature–time curves are often used as a means of experimentally determining the temperatures of phase changes, and thinking about them can add to your understanding of phase relations. In looking at the four different cooling curves in Figure 17.25, you should imagine that you are in a laboratory, conducting an investigation into the system A–B. One way to proceed would be to prepare a number of different bulk compositions (thoroughly mix various proportions of A and B together), then heat each bulk composition to a number of different temperatures, wait long enough to achieve equilibrium, then observe what phases are present, and measure their compositions. Figure 17.25 (left side) would then represent the results you obtained from a large number of such experiments.

Another way to proceed would be to heat each bulk composition to a temperature sufficiently high to produce a homogeneous liquid, and then to cool slowly while observing the temperature. When the liquidus temperature is reached, the latent heat of crystallization is released, resulting in a slower rate of cooling. When the eutectic temperature is reached, cooling will cease completely, while three phases coexist. When the liquid disappears, cooling will resume. By observing the inflection points and plateaus in the temperature–time curves, you may deduce the positions of the liquidus and eutectic for the various compositions. You would still need to do more experiments to determine phase compositions, but the cooling-curve method can often give the general shape

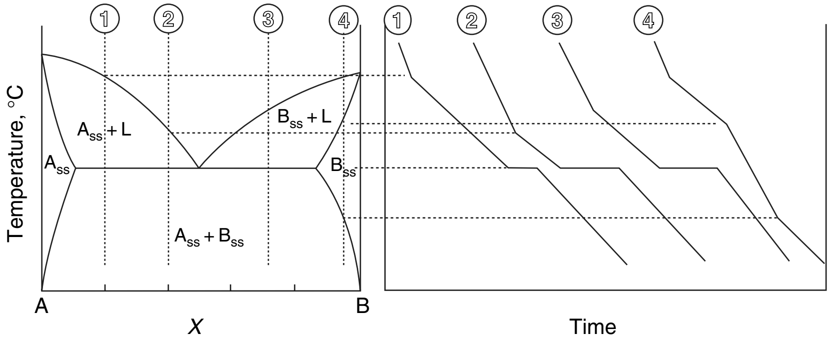


Figure 17.25
Temperature–time curves through a binary eutectic system.

of a diagram in a relatively short time. Note that composition number 4 does not pass through the eutectic, and so does not show a temperature plateau.

17.3.13 Intermediate compounds

In all the binary systems we have considered so far, no compounds are formed *between* the compounds A and B. That is, there are no compounds AB, or A_2B, or A_2B_3, and so on. What happens if these do exist? Consider the binary system A–B that contains the binary compound AB. The simplest possibility is that both A–AB and AB–B are binary systems of the same type, such as simple eutectic systems. Then the two systems are "glued together," as in Figure 17.26.[4]

Another common possibility is that the liquidus for one of the end-member compounds extends completely over the intermediate compound, as in Figure 17.27. When this happens, compound A–B does not melt to a liquid of its own composition – it breaks down at the peritectic temperature to a

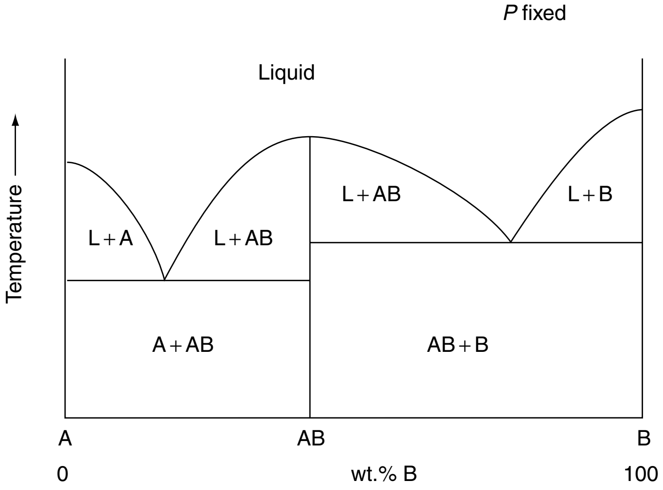


Figure 17.26 Intermediate compound AB divides the binary system A–B into two similar parts. Note that if the composition axis were in mole fraction or mole %, AB would appear midway between A and B, but not if the axis is in weight %.

---

[4] An interesting point about diagrams containing intermediate compounds such as Figure 17.26 is the observation that the liquidus (and the solidus, if one is present) is always "flat" (has a zero tangent) at the intermediate composition, as shown in the diagram, whereas normally, addition of a second component (in this case, A or B added to AB) lowers the liquidus temperature, as explained in §17.3.5. A proof sometimes mentioned [and attributed by Pitzer (1995, p. 171) to Lorentz and Stortenbecker in 1892] shows that this will result if compound AB dissociates into two or more species in the liquid phase, but that a "cusp" (a nonzero tangent of the liquidus on both sides of the intermediate compound) will develop if the compound does not dissociate. However, Berndt and Diestler (1968) contend that dissociation of the intermediate compound is not required, only that the melting point is a first-order transition, as it normally is. We leave this argument to phase diagram aficionados.

Figure 17.27 Intermediate compound AB displays incongruent melting.

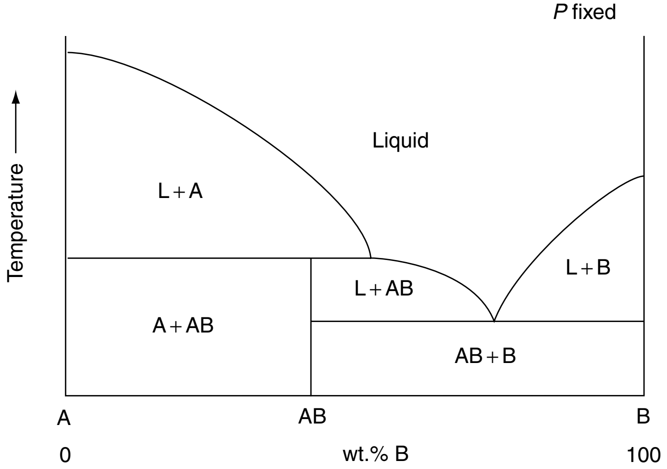
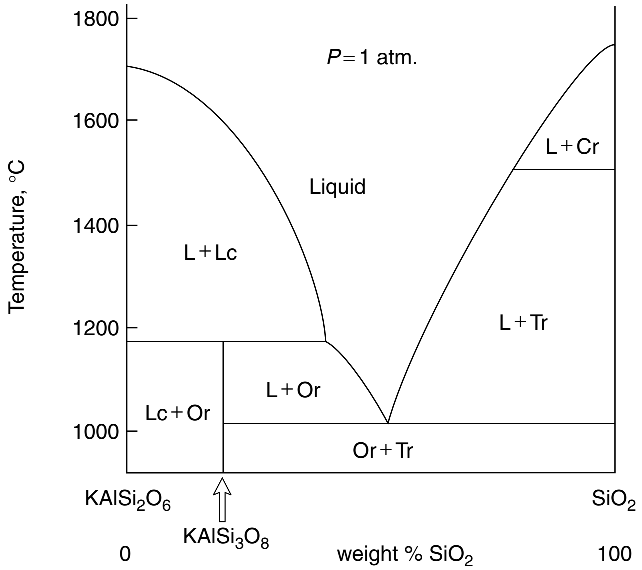


Figure 17.28 The system $KAlSi_2O_6$–SiO_2, with intermediate compound $KAlSi_3O_8$. Lc-leucite. Or-orthoclase (K-feldspar). Tr-tridymite. Cr-cristobalite. L-liquid.



different compound plus a liquid. This is known as *incongruent melting*. A good example of this is the system $KAlSi_2O_6$–SiO_2 (leucite–silica), which contains the intermediate compound $KAlSi_2O_6 \cdot SiO_2$, or $KAlSi_3O_8$, K-feldspar, shown in Figure 17.28. The large liquidus surface extending over the intermediate compound in these diagrams will often "shrink" with increasing pressure, leading ultimately to the "glued together" type of system (Figure 17.26). In other words, AB may melt incongruently at low pressures and congruently at high pressures.

17.4 Ternary systems

17.4.1 Ternary compositions

With the addition of a third component, we now need two dimensions to display all possible compositions in a system, and so we lose the ability to display composition and temperature simultaneously. We can only display compositions at a chosen *T* and *P* on a section, or we can *project* compositions from various conditions onto a single plane, as discussed below. The method of depicting compositions within a triangle is shown in Figure 17.29. Each apex of the triangle represents 100% of one of the components. The proportion of each component in a ternary composition is measured by the distance of a point from the side of the triangle opposite the component in question, as shown. The triangles are usually isometric, but not necessarily. Right angled

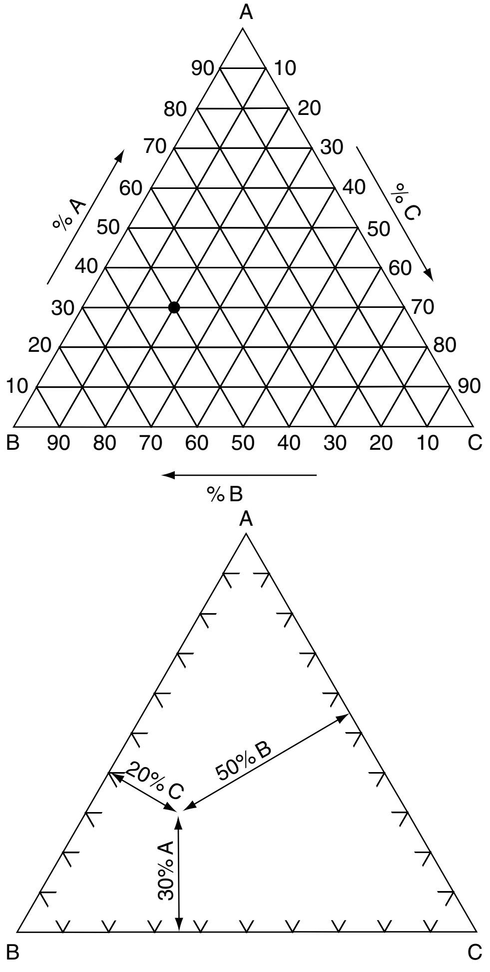


Figure 17.29 Representation of a ternary composition. The dot in the upper triangle represents a composition of 30% A, 50% B, and 20% C, as shown in the lower triangle.

triangles are also used in some circumstances, and other shapes could be used if desired.

17.4.2 Sections and projections

In discussing binary systems we used only binary *sections*, although we mentioned that various kinds of *projections* could be used. We must now expand on this statement. As shown in Figure 17.9, a section shows what you would see after slicing through a *P–T–X* box, as though with a knife. A projection, on the other hand, is what you see by peering through the box at some angle, usually parallel to one of the axes, and seeing the curves on all sides of the box at the same time. For example, Figure 17.30 shows a *P–T projection* of Figure 17.9. You see the unary phase diagrams for both components superimposed on one another. In a more complete projection, you would also see various curves projected from *within* the box, as well as the curves on the faces of the box. For example, there is a curve joining the critical points of each pure component, which crosses through the box, showing the critical points of binary compositions, which is not shown.

Ternary projections

In looking at ternary systems, we will start with the projection. To best see the meaning of ternary projections, we start with the oblique view of a simple ternary eutectic system, Figure 17.31. In this figure we see that each side of the compositional triangle has a binary isobaric *T–X* section constructed on it, perpendicular to the compositional triangle. We see, too, that the liquidus lines of the binaries are joined into surfaces that extend across the ternary space. Each binary eutectic point becomes a ternary *cotectic line* extending into the ternary, down to a *ternary eutectic* point. Points and lines on these surfaces are projected onto a plane surface, as depicted in Figure 17.31. The projection

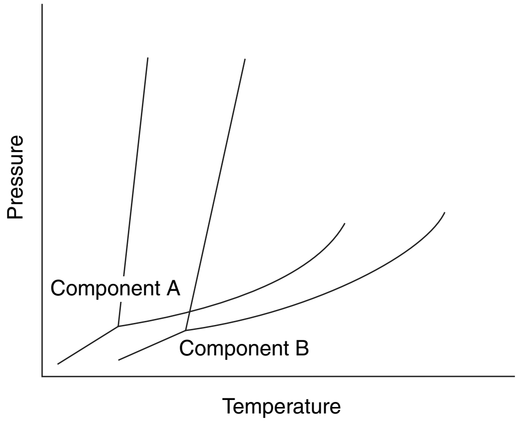


Figure 17.30 A *P–T* projection of Figure 17.9.

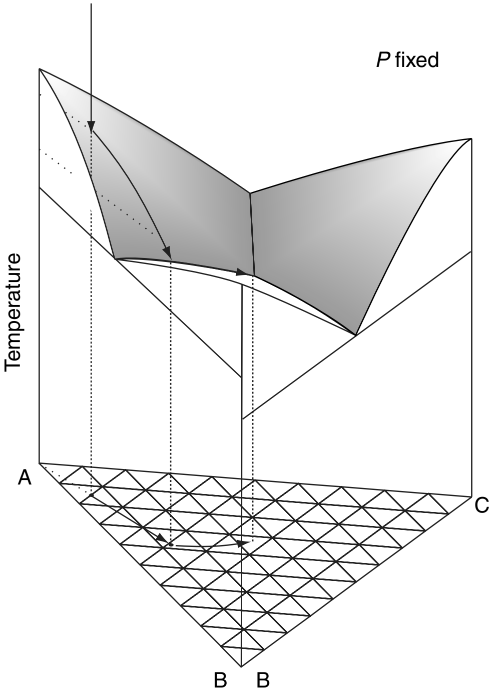


Figure 17.31 A ternary eutectic system ABC in an oblique view. The cooling path of a liquid of composition 80% A, 15% B, 5% C is shown.

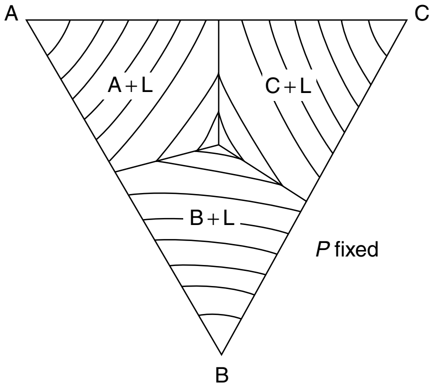


Figure 17.32
A polythermal projection of the liquidus surfaces in Figure 17.31.

is what you would see if you looked straight down on the three-dimensional object in Figure 17.31, parallel to the temperature axis. The results of such projection are shown in Figures 17.32 and 17.33.

In Figure 17.32 the cotectic lines and *isothermal contours* on the liquidus surfaces have been projected onto a plane surface. The triangle is divided by the cotectic lines into three areas, labeled A + L, B + L, and C + L (where A stands for solid A, etc.), because bulk compositions in these areas will consist of these phases at temperatures below those of the liquidus surfaces. This will

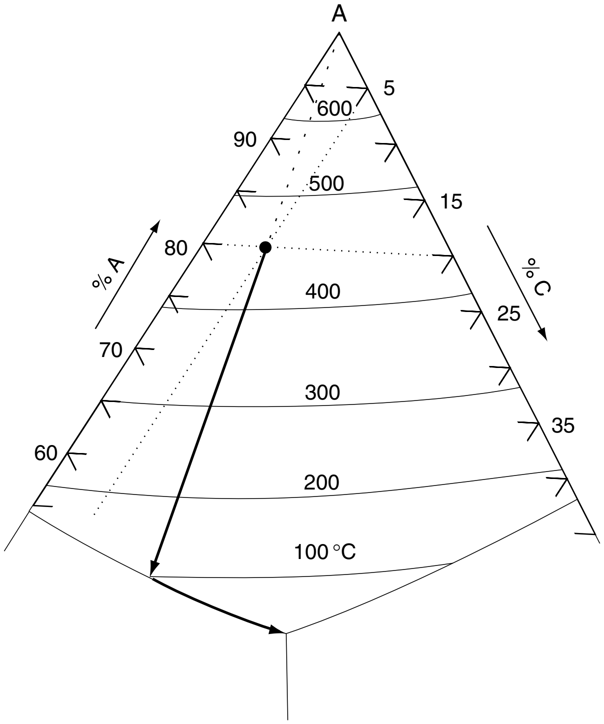


Figure 17.33 Polythermal projection of the liquidus surface of component A in Figure 17.31.

become clearer by examining what happens to a liquid of composition 80% A, 15% B, 5% C, as shown in Figures 17.31 and 17.33.

Starting at a temperature well above the liquidus surface, as shown in Figure 17.31, the liquid cools vertically downward until it hits the liquidus surface at a temperature of 450 °C (Figure 17.33). This liquidus surface is the locus of points indicating the first appearance of crystals of composition A, and so crystals of A start to separate from the liquid. A tie-line joins the liquid composition to the A-axis. Because composition A is being subtracted from the liquid, the liquid composition *must move directly away from* A, as shown in Figure 17.33. The composition of the liquid stays on the liquidus surface, always on the continuation of a straight line through composition A and the point [80% A, 15% B, 5% C]. A continuous series of tie-lines join the liquid composition to the A-axis (two of which are shown in Figure 17.31). This continues until the liquid composition hits the cotectic line (temperature 100 °C, Figure 17.33), which joins the liquidus surfaces of A and B. On this line, the liquid is simultaneously in equilibrium with crystals of A and crystals of B, and so B starts to precipitate. On further cooling, both A and B precipitate, and the liquid composition moves down the cotectic line until it hits the ternary eutectic. At this point, C starts to precipitate, and all three solids precipitate until the liquid is used up. At the ternary eutectic, the number of phases is 4 (solid A, solid B, solid C, and L), and

$$\begin{aligned} f &= c - p + 1 \\ &= 3 - 4 + 1 \\ &= 0 \end{aligned}$$

The ternary eutectic is thus an isobaric invariant point, and no temperature change can take place until the liquid is all used up, at which time the crystals will resume cooling.

Ternary sections

Consider a temperature midway between the melting points of the pure components and the three eutectic temperatures in system ABC, as shown in Figure 17.34. An isothermal plane at this temperature will cut through all three liquidus surfaces. Near each apex, the plane lies below each liquidus surface, so it shows an area of solid plus liquid filled with tie-lines. In the center of the diagram, the plane is everywhere above the liquidus surfaces, and so it shows a blank "field of liquid." Sections at successively lower temperatures would show the two-phase fields expanding and coalescing, leaving smaller and smaller liquid fields.

17.4.3 The granite system

The simple ternary eutectic system discussed above represents just a beginning to the general subject of ternary phase diagrams. Features such as solid

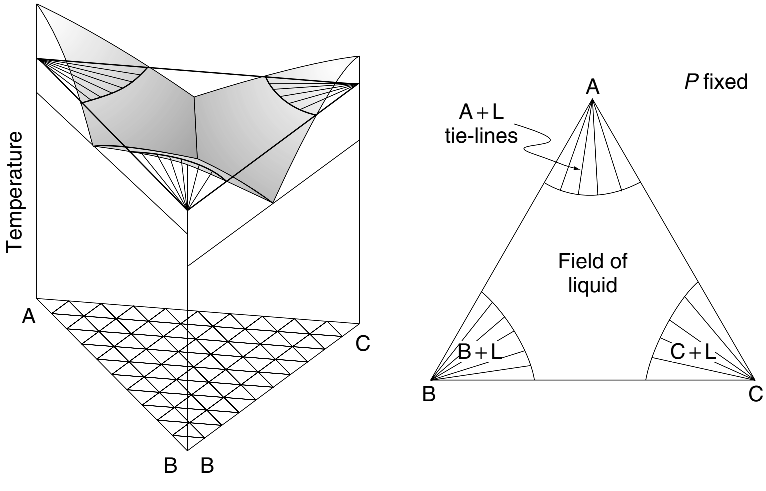


Figure 17.34
An isothermal section through system ABC.

solutions, peritectics, intermediate compounds, and so on all introduce complications that we will not discuss. However, as an example of the simple ternary eutectic diagram, let's have a look at the system SiO_2–$NaAlSi_3O_8$–$KAlSi_3O_8$–H_2O, which is used as a model for natural granites. Component SiO_2 is called q, component $NaAlSi_3O_8$ is called ab, and component $KAlSi_3O_8$ is called or. This is because the most common form of SiO_2 is quartz (q), pure $NaAlSi_3O_8$ is the mineral albite (ab), and one of the varieties of $KAlSi_3O_8$ is the mineral orthoclase (or). Quartz is abundant in granites, as is plagioclase, made up mostly of ab, and K-feldspar, made up mostly of or. Granites usually contain several minerals in addition to quartz, plagioclase, and K-feldspar, but these minerals (that is, quartz, albite-rich plagioclase and K-feldspar) often account for 80–90% of a granite, so the system q–ab–or is quite a useful model in trying to understand the crystallizing or melting histories of granites in general.

The granite system at 10 kbar

The granite system diagram is shown as a polythermal projection in Figure 17.35.

Compare this diagram with Figures 17.8 and 17.17. In Figure 17.8 you see that the melting point of solid SiO_2 (as β-cristobalite) is about 1700 °C at 1 bar, whereas in Figure 17.35 it is somewhat less than 1100 °C. The melting point of albite at 1 bar in Figure 17.17 is 1118 °C, whereas in Figure 17.35 it is just over 700 °C. Raising the pressure from 1 bar to 10 000 bar increases the melting points, so that cannot be the explanation. What is going on?

The difference lies in the fact that the granite system is not the "dry" system q–ab–or, but the "wet" system q–ab–or–H_2O. The presence of water at high

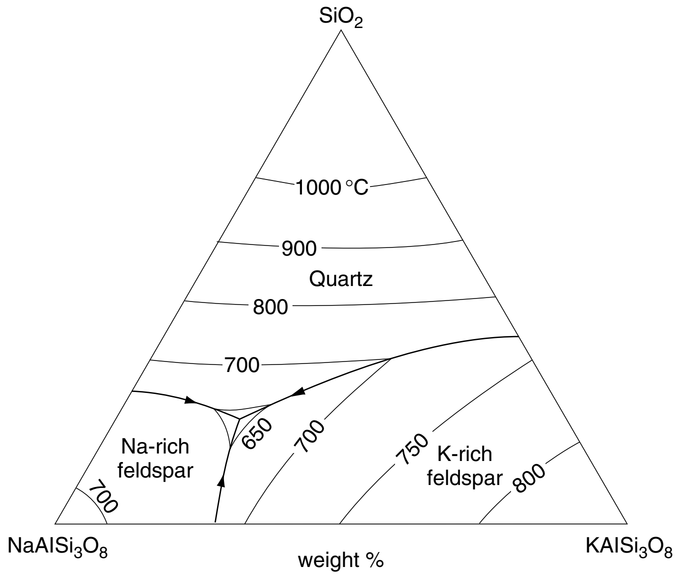


Figure 17.35 The granite system at 10 kilobars.

pressures has the effect of substantially lowering the melting temperatures of the pure minerals. Thus Figure 17.35 is not really a ternary projection, but a quaternary projection. All liquids in the diagram have not ternary compositions, but quaternary compositions; that is, they are all saturated with water, and supercritical water is an extra phase that is not shown in the diagram. The presence of the water brings the liquidus temperatures down well into the range of temperatures found in the Earth's crust and means that water is an important component in the history of real granites. However, as water is an extra phase as well as an extra component, we can treat Figure 17.35 exactly like a ternary eutectic and essentially forget about the water.

According to Figure 17.35, crystallization of any bulk composition within the system will generate a final liquid composition at the ternary eutectic at just under 650 °C. Note that the composition of this final liquid (or initial liquid on heating) lies quite close to the albite corner – ternary eutectics do not always occur near the center of the diagram. At other pressures, the position of this eutectic changes. At lower pressures, it moves "northeast," roughly directly away from the albite composition, as shown in Figure 17.36.

A complicating factor here is that the system only shows a eutectic at high pressures. Below about 5 kbar, the eutectic changes to a ternary minimum, as indicated by the change from the circle to plus signs in Figure 17.36. The reason for this is shown in the lower two sequences of diagrams in Figure 17.23. In one of these, a binary eutectic is generated when a more-or-less symmetrical melting loop intersects a solvus. This corresponds to the ab–or–H_2O system at high pressures. In the other, the melting loop has a minimum temperature offset to one side, so that even after the solvus intersection takes place, the minimum is preserved. This corresponds to the system ab–or–H_2O at low pressures. This difference between a eutectic and a minimum is preserved in

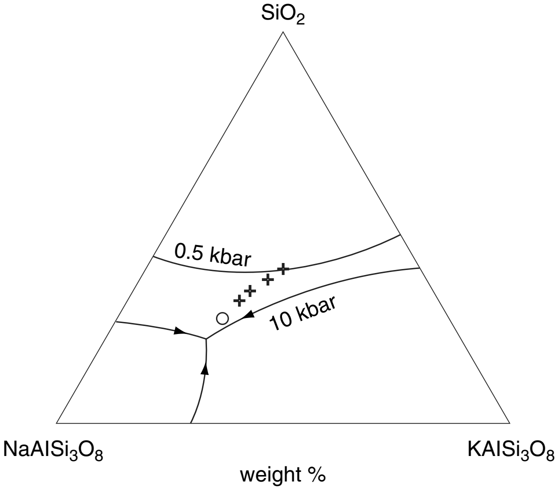


Figure 17.36 The granite system at pressures from 0.5 to 10 kbar. The plus signs show the position of the ternary minimum at 0.5, 2, 3, and 4 kbar. The circle shows the position of the ternary eutectic at 5 kbar.

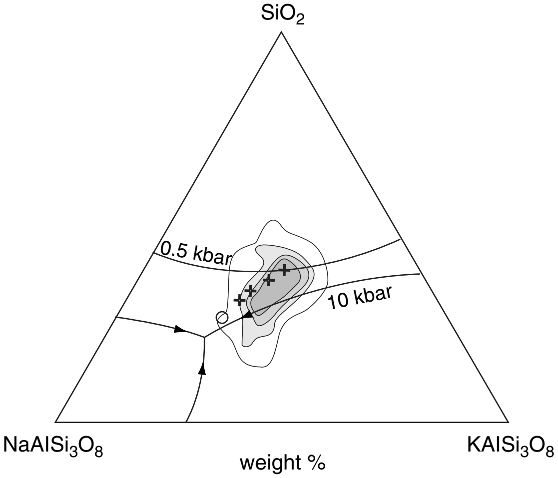


Figure 17.37 The compositions of natural granites superimposed on Figure 17.36. Several hundred individual points are contoured, with the darker colors indicating a higher frequency of points.

the ternary. However, from the point of view of liquid compositions generated during cooling in ternary systems, there is little difference between a ternary eutectic and a ternary minimum. Both represent the final liquid composition for many bulk compositions in the system.

17.4.4 Granite compositions

This brief explanation is not sufficient for you to understand all the details of this system, but it is sufficient to understand one important result. When the compositions of natural granites are normalized to q–ab–or and plotted in the q–ab–or triangle, a remarkable coincidence of compositions and the ternary minima and eutectics results, as shown in Figure 17.37. As natural granites have undoubtedly crystallized at a variety of pressures, the compositions would be expected to be strung out along the track defined by the ternary minima and eutectics, as they are, *providing that natural granites actually form by crystallizing from silicate liquids*. The demonstration by Tuttle and Bowen (1958) that this was the case provided strong evidence for the magmatic origin of granites. The slight offset of the highest frequency of natural compositions toward the KAlSi$_3$O$_8$ apex, as well as other aspects of the diagram, have been the subject of much discussion.

17.5 Summary

Phase diagrams are a kind of concise representation of the equilibrium relationships between phases as a function of chosen intensive variables, such as temperature, pressure, composition, pH, oxidation potential, activity ratio, and so on. They are extremely useful, not only in representing what is known about a system, but in thinking about processes involving phase changes. Most of the diagrams in Chapters 12 and 14 are phase diagrams, although the term is most often used in relation to the *P–T–X* type of diagram discussed in this chapter.

Although in principle phase diagrams can be calculated from thermodynamic data, in complex systems the relationships are generally determined experimentally. The diagrams must nevertheless obey rules established by equilibrium thermodynamics. Therefore an understanding of the material in all the previous chapters is a prerequisite to a real understanding of the simple points, lines, and surfaces found in phase diagrams.

The following extract is from Bridgman (1958: 423). It may help to dispel any notion that determining phase relations is a simple and straightforward exercise.

Bridgman was a professor at Harvard University and in 1946 received the Nobel Prize in physics for his research on materials at high pressures.

> The conditions under which the different modifications of ice appear are somewhat capricious, and often inconvenient manipulation is necessary to arrive in the part of the phase diagram desired. The behavior is particularly striking in the neighborhood of the V-VI-liquid triple point, say, between 0 °C and −10 °C and between 5000 and 6000 bar. With the ordinary form of apparatus, water in a steel piezometer with pressure transmitted to it by mercury, it is very difficult to produce the modification V. For example, if liquid water at −10 °C is compressed across the melting curve, the liquid will persist in the sub-cooled condition for an indefinite time, without freezing to V, the stable form. But if the compression is carried further, to the unstable prolongation of the liquid-VI line, freezing to the form VI will take place almost at once on crossing the line, in spite of the fact that VI is unstable with respect to V. Now in order to make V appear, VI may be cooled 30° or 40°, when it will spontaneously change to V, the stable form. If pressure is now released back to 5000, say, and temperature is raised to the melting line of V, melting takes place at once and so the coordinates of the melting line may be found. Suppose now after the melting is completed the liquid water be kept in the neighborhood of the melting line for several days, and then the pressure increased again across the melting line at −10 °C; it will be found that the instant the melting line of V is crossed the liquid freezes to V. This suggests that there has persisted in the liquid phase some sort of structure, not detected by ordinary large-scale experiments, favorable to the formation of V. It is now known, of course, from X-ray analysis, that structures are possible in the liquid; this experiment suggests a specificity in these structures that might well be the subject of further study.
>
> The formation of the nucleus of V may be favored by the proper surface conditions. Thus if there is any glass in contact with the liquid, either a fragment of glass wool purposely introduced into the liquid, or by enclosing the water in a glass instead of a steel bulb, freezing to V takes place at once without the slightest hesitation immediately on carrying the virgin liquid across the freezing line.

18
Process modeling

18.1 Introduction

In this chapter we will have a brief look at how equilibrium thermodynamics is used in dealing with *processes*, that is, reacting systems, rather than just systems at equilibrium. Realistically, this should mean including the science of chemical kinetics with our thermodynamics, and we should also include other factors, such as fluid flow, temperature and pressure gradients, and surface reactions, to build increasingly realistic models of complex natural phenomena involving the movement and chemical reactions of fluids in soils and rocks in the Earth's crust.

That is all a bit too ambitious for this book. Computational models including all these subjects are at the forefront of research in several areas. What we will attempt in this chapter is first, to outline the rudiments of chemical kinetics, and then show how equilibrium thermodynamics simulates chemically reacting systems *without* using kinetics. The reason for discussing kinetics is that process modeling in thermodynamics has some points in common with kinetics, because both sciences consider the problem of chemical reactions proceeding from start to finish, and both use the progress variable. In simulating reactions in thermodynamics, however, we use the progress variable but not the real time variable, with which it is closely connected in kinetics. The use of a real time variable is what distinguishes kinetics from thermodynamics.

18.1.1 Quasi-static processes

In the following discussions, it will be useful to introduce the term *quasi-static*. This term has various meanings as used by various authors, but in this chapter it refers to an irreversible process carried out in a very large number of very small steps. The difference between a quasi-static and a reversible process is that the reversible process refers to a series of *stable* equilibrium states, while the quasi-static process is a series of *metastable* equilibrium states. After every step of a reversible process, the system is at stable equilibrium with only two constraints. After every step of a quasi-static process, the system is at a metastable equilibrium and has at least three constraints. This concept and the need for it will become clear by considering some examples.

18.2 Kinetics

18.2.1 The myth of equilibrium

If you look around at the world we live in at the surface of planet Earth, it seems to be characterized by constant change; virtually nothing is permanent – even the solid rocks of the crust are weathering and being washed into the sea. And organic substances, including us, are among the most evanescent of objects – here today, gone tomorrow. Where are those states of equilibrium that are so important in our thermodynamic model? It is in fact a tribute to the creativity of the human mind that scientists were able to "see through" the constant flux around them and create what is in essence a model of energy relationships in a world that does not exist – the equilibrium world.

As we have seen in the last few chapters, this mathematical model of energy relationships is tremendously useful in the real world, basically because even though the real world is in a constant state of flux or change, there are many situations in which it approaches fairly closely a state of equilibrium, and even in cases where it does not, it is changing towards some equilibrium state, and our equilibrium models are useful in many ways. But obviously we would like to know more about the state of flux itself. How fast do the changes we see take place, and what controls this rate of change? We enter the world of chemical kinetics.

18.2.2 The progress variable

Consider a generalized chemical reaction

$$a\mathrm{A} + b\mathrm{B} = c\mathrm{C} + d\mathrm{D} \tag{18.1}$$

where A, B, C, and D are chemical formulas, and a, b, c, d are the stoichiometric coefficients. We pointed out in §9.3 that when this reaction reaches equilibrium,

$$c\mu_\mathrm{C} + d\mu_\mathrm{D} = a\mu_\mathrm{A} + b\mu_\mathrm{B}$$

and

$$\begin{aligned}\Delta_r\mu &= c\mu_\mathrm{C} + d\mu_\mathrm{D} - a\mu_\mathrm{A} - b\mu_\mathrm{B} \\ &= 0\end{aligned} \tag{18.2}$$

Equation (18.1) can be generalized to

$$\sum_i \nu_i M_i = 0 \tag{18.3}$$

where M_i are the chemical formulas and ν_i are the stoichiometric coefficients, with the stipulation that ν_i is positive for products and negative for reactants. Equation (18.2) can then be generalized to

$$\sum_i \nu_i \mu_i = 0 \tag{18.4}$$

Up to now, we have been most interested in reactions that reach equilibrium. Now let's look at what happens before that point is reached, that is, while the reaction is taking place. Let's say that the reaction proceeds from left to right as written. It doesn't matter for the moment whether all the reactants and products are in the same phase (a *homogeneous* reaction) or in different phases (a *heterogeneous* reaction). During the reaction, A and B disappear and C and D appear, but the *proportions* of A:B:C:D that appear and disappear are fixed by the stoichiometric coefficients. If the reaction is

$$\mathrm{A}+2\,\mathrm{B}\rightarrow 3\,\mathrm{C}+4\,\mathrm{D} \tag{18.5}$$

then for every mole of A that reacts (disappears), two moles of B must also disappear, while three moles of C and four moles of D must appear. This is simply a mass balance, independent of thermodynamics or kinetics, and can be expressed as

$$\frac{dn_{\mathrm{A}}}{\nu_{\mathrm{A}}}=\frac{dn_{\mathrm{B}}}{\nu_{\mathrm{B}}}=\frac{dn_{\mathrm{C}}}{\nu_{\mathrm{C}}}=\frac{dn_{\mathrm{D}}}{\nu_{\mathrm{D}}}=\frac{dn_i}{\nu_i} \tag{18.6}$$

where, if all reactants and products are pure phases, the differentials dn_{A}, dn_{B}, and so on, can refer to a change in the amount of A, B, and so on, of any convenient magnitude, not necessarily an infinitesimal change. We can then represent every term in (18.6) by a single variable $d\xi$ so

$$\frac{dn_{\mathrm{A}}}{\nu_{\mathrm{A}}}=\frac{dn_{\mathrm{B}}}{\nu_{\mathrm{B}}}=\frac{dn_{\mathrm{C}}}{\nu_{\mathrm{C}}}=\frac{dn_{\mathrm{D}}}{\nu_{\mathrm{D}}}=d\xi \tag{18.7}$$

from which it appears that

$$\frac{dn_{\mathrm{A}}}{d\xi}=\nu_{\mathrm{A}};\quad \frac{dn_{\mathrm{B}}}{d\xi}=\nu_{\mathrm{B}};\quad \ldots\ \frac{dn_i}{d\xi}=\nu_i \tag{18.8}$$

where our new variable ξ is called the reaction progress variable, and in this case represents an arbitrary number of moles. Equation (18.8) says that in reaction (18.5), $dn_{\mathrm{A}}/d\xi=-1$, $dn_{\mathrm{B}}/d\xi=-2$, $dn_{\mathrm{C}}/d\xi=3$, and so on, which simply means that for every mole of A that disappears, two moles of B also disappear, three moles of C appear, and so on.

18.2.3 The reaction rate

Having defined reaction increments $d\xi$, we can now define the *rate of reaction* as

$$\frac{d\xi}{dt}=\frac{1}{\nu_{\mathrm{A}}}\frac{dn_{\mathrm{A}}}{dt}=\frac{1}{\nu_{\mathrm{B}}}\frac{dn_{\mathrm{B}}}{dt}=\ldots=\frac{1}{\nu_i}\frac{dn_i}{dt} \tag{18.9}$$

where dt is an increment of time, and $d\xi/dt$ is the derivative of ξ with respect to t and is an expression of the amount of progress of the reaction as a function of time, or simply the rate of reaction.

This expression (18.9) is written in terms of the absolute number of moles of A, B, and so on, (n_{A}, n_{B},…), but by considering a fixed volume we could change these to concentration terms. Thus

$$\frac{d\xi}{dt} = \frac{1}{\nu_{\mathrm{A}}}\frac{dC_{\mathrm{A}}}{dt} = \frac{1}{\nu_{\mathrm{B}}}\frac{dC_{\mathrm{B}}}{dt} = \ldots = \frac{1}{\nu_i}\frac{dC_i}{dt} \tag{18.10}$$

where C is some unit of concentration such as $\mathrm{mol\,cm^{-3}}$.

So evidently, the rate of reaction can be determined by measuring the concentration of *any* of the reactants or products as a function of time. With one important stipulation.

Elementary and overall reactions

If you actually measure the rate of change of concentration of products and reactants in many ordinary chemical reactions, you find that the relationship in (18.10) is often not obeyed. This is because the reaction does not actually proceed *as written*, at the molecular level. For example, reaction (18.5), taken literally, indicates that a molecule of A reacts with two molecules of B, and at that instant, three molecules of C and four molecules of D are formed. But this might not be what happens at all, and in view of the improbability of three molecules ($\mathrm{A} + 2\,\mathrm{B}$) meeting at a single point, it probably is not in this case. The reaction as written may well represent the *overall* result of a series of *elementary* reactions. Thus A and B may in fact react to form a number of *intermediate species* such as X and Y, which then react with each other or with A or B to form C and D. In thermodynamics, the existence of such intermediate species is not important to the study of the overall reaction, as long as equilibrium is attained, but in kinetics, they contribute to the overall rate of reaction and may actually be *rate-controlling*, even though their concentrations may be small.

Of course, it is also possible that intermediate species do form, but they achieve a *steady-state* concentration, that is, they break up just as rapidly as they form. In this case, Equation (18.10) would be obeyed, even though it did not represent what actually happens at the molecular level.

From now on in this section, we will use = in overall reactions, and → in elementary reactions (those that actually proceed as written).

18.2.4 Rate laws

A rate law is a statement about how the rate of a reaction depends on the concentrations of the participating species. If one thinks about chemical reactions as something that happens at the molecular level when molecules collide with one another, it makes sense that the number of collisions, and hence the rate of reaction, should depend on how many molecules of each type there are; that is, their concentrations.[1] In most cases, a simple power function of concentrations is found to apply. For reaction (18.1) it is

$$\text{rate of reaction} = \frac{d\xi}{dt} = k \cdot C_{\mathrm{A}}^{n_{\mathrm{A}}} C_{\mathrm{B}}^{n_{\mathrm{B}}} C_{\mathrm{C}}^{n_{\mathrm{C}}} C_{\mathrm{D}}^{n_{\mathrm{D}}} \tag{18.11}$$

The constant of proportionality, k, is called the *rate constant*. The exponents $n_{\mathrm{A}} \ldots n_{\mathrm{D}}$ are often integers, but can be fractional or decimal numbers, especially in heterogeneous reactions where adsorption and other surface-related effects can influence reaction rates. They define the *order* of the reaction. If n_{A} is 2, the reaction is said to be second order in A. The sum of the exponents gives the overall order of the reaction.

The rate laws for chemical reactions are expressions that best fit experimental data, and the order of reaction is the sum of the experimentally determined exponents.

Rate laws for elementary reactions

Rate laws are determined by analyzing one or more reactant or product species as a function of time as a reaction proceeds, and then inspecting the results to see what theoretical form best fits the data.

First order

Rate laws for elementary reactions are for the most part what one would expect. For example, a simple molecular (or nuclear) decomposition,

$$\mathrm{A} \rightarrow \text{products}$$

proceeds at a rate that depends only on the concentration of A; the more A, the more decomposition per unit time. The rate law is

$$\frac{d\xi}{dt} = -\frac{dC_{\mathrm{A}}}{dt} = k \cdot C_{\mathrm{A}} \tag{18.12}$$

and the reaction is first order. The decay of radioactive elements is an example of such reactions.

If we simplify C_{A} to C and let $C = C^{\circ}$ at time $t = 0$, integration of (18.12) gives

$$\int_{C^{\circ}}^{C} \frac{dC}{C} = -k \int_{0}^{t} dt \tag{18.13}$$

$$\ln \frac{C^{\circ}}{C} = kt \tag{18.14}$$

$$\ln C = \ln C^{\circ} - kt \tag{18.15}$$

$$C = C^{\circ} e^{-kt} \tag{18.16}$$

---

[1] In thermodynamics, we must use "corrected" concentrations, or activities. In kinetics it is the actual concentrations that are important.

These equations suggest various ways of plotting data to see if they fit a first order rate law. For example, a plot of $\ln(C^\circ/C)$ versus t will give a straight line with a slope equal to the rate constant for concentration data from a first order reaction.

Second order

The most common type of elementary reaction results from bimolecular collisions:

$$\mathrm{A} + \mathrm{B} \rightarrow \text{products} \tag{18.17}$$

Here we expect the frequency of reaction to be proportional to the concentrations of the reactants and the concentration of products to have no effect, and so the rate law is

$$\frac{d\xi}{dt} = -\frac{dC_\mathrm{A}}{dt} = -\frac{dC_\mathrm{B}}{dt} = k \cdot C_\mathrm{A}^1 C_\mathrm{B}^1 \tag{18.18}$$

and the reaction is second order.

If the initial concentrations of A and B are C_A° and C_B°, the stoichiometry of (18.17) requires that

$$C_\mathrm{A}^\circ - C_\mathrm{A} = C_\mathrm{B}^\circ - C_\mathrm{B}$$

Solving this for C_B and substituting this result in (18.18) gives

$$-\frac{dC_\mathrm{A}}{dt} = k \cdot C_\mathrm{A}(C_\mathrm{A} - C_\mathrm{A}^\circ + C_\mathrm{B}^\circ) \tag{18.19}$$

Integration of (18.19) then gives

$$\ln\left(\frac{C_\mathrm{A}^\circ C_\mathrm{B}}{C_\mathrm{B}^\circ C_\mathrm{A}}\right) = (C_\mathrm{B}^\circ - C_\mathrm{A}^\circ)kt \tag{18.20}$$

Therefore, a plot of

$$\left(\frac{1}{C_\mathrm{B}^\circ - C_\mathrm{A}^\circ}\right)\ln\left(\frac{C_\mathrm{A}^\circ C_\mathrm{B}}{C_\mathrm{B}^\circ C_\mathrm{A}}\right)$$

versus time t will result in a straight line with a slope equal to the rate constant for concentrations taken from a second order reaction. Similar equations can be derived for reactions with different stoichiometric coefficients.

There are a number of other rate laws, but this will suffice to give an idea of the procedures involved. However, it should be emphasized that most chemical reactions are "overall" reactions, and that their understanding in terms of their fundamental elementary reactions is a goal not often and not easily achieved.

18.2.5 Temperature dependence of rate constants

In §9.6 we saw that the temperature dependence of the equilibrium constant K could be expressed as

$$\frac{d\ln K}{d(1/T)} = -\frac{\Delta_r H^\circ}{R}$$

or, alternatively

$$\frac{d\ln K}{dT} = \frac{\Delta_r H^\circ}{RT^2}$$

In 1889, Arrhenius proposed a similar equation for the temperature effect on rate constants,

$$\frac{d\ln k}{d(1/T)} = -\frac{E_\mathrm{a}}{R} \tag{18.21}$$

or

$$\frac{d\ln k}{dT} = \frac{E_\mathrm{a}}{RT^2} \tag{18.22}$$

where E_a is called the Arrhenius activation energy, or just the activation energy, and turns out to be closely related to the "energy barrier" between products and reactants in chemical reactions. The general form of this equation has been shown to be derivable from statistical mechanics.

Experimental data for a great many reactions over a large range of temperatures shows that the Arrhenius equation is usually closely obeyed, showing that E_a is either a constant or a weak function of temperature, and so we can integrate the equation to give

$$\ln k = \ln A - \frac{E_\mathrm{a}}{RT} \tag{18.23}$$

or

$$k = Ae^{-E_\mathrm{a}/RT} \tag{18.24}$$

where A, which enters (18.23) as a constant of integration, is called the *pre-exponential* factor.

The activation energy of an overall reaction is made up of the individual contributions of the elementary reactions making up the overall reaction. The magnitude of the activation energy can vary from virtually zero to hundreds of kJ per mole and, besides controlling the temperature dependence of the rate constant, provides clues as to the nature of the reaction mechanisms, because the energies involved in many types of diffusion, electron exchange, and bond-breaking processes are known.

Calculation of activation energies for elementary gas phase reactions from relatively simple molecular considerations has been fairly successful, but heterogeneous reactions are still far from such theoretical treatment.

18.2.6 An example: kinetics of acetate decomposition

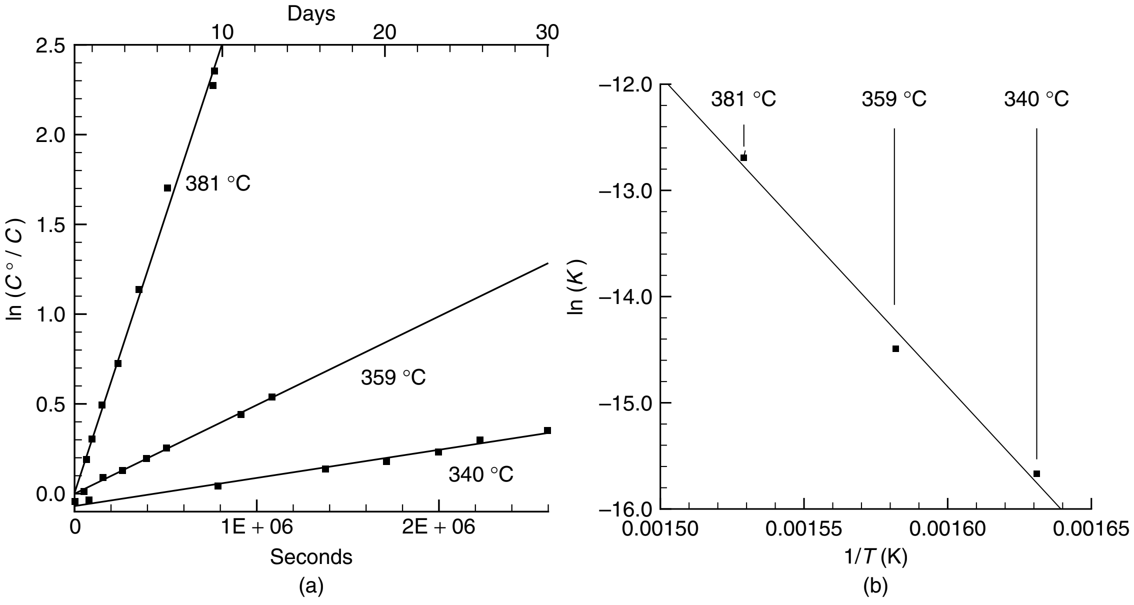
An example of the use of the rate constant and activation energy in a homogeneous reaction is the work of Palmer and Drummond (1986) on the kinetics of the decomposition of acetate compounds such as acetic acid and sodium acetate at elevated temperatures. Organic compounds are generally quite unstable at elevated temperatures, but some persist for surprisingly long times. Just how long is an important geological question, because these compounds may be involved in many important processes, such as the generation of natural gas and transport of base metals. To find out how long acetates could be expected to persist at various temperatures, Palmer and Drummond measured the concentration of acetate remaining after various times at several temperatures, using several different containing surfaces, because it is found that surfaces play a catalytic role. That is, the rate of reaction depends on the surface available to the reaction. Some of their data are shown in Table 18.1 and Figure 18.1.

Table 18.1 *Data from Palmer and Drummond (1986) on the breakdown of sodium acetate on titanium surfaces at three temperatures.*

Time			
hours	seconds	Acetate molality	$\ln(C^o/C)$
Experiment 26, 340 °C			
22.0	79 200	0.04E+010	−0.036
219.0	788 400	9.59E−01	0.042
382.5	1 377 000	8.71E−01	0.138
475.5	1 711 800	8.37E−01	0.178
554.5	1 996 200	7.92E−01	0.233
618.0	2 224 800	7.42E−01	0.298
721.0	2 595 600	7.03E−01	0.352
Experiment 27, 359 °C			
14.0	50 400	9.89E−01	0.011
43.5	156 600	9.15E−01	0.089
73.0	262 800	8.78E−01	0.130
109.5	394 200	8.22E−01	0.196
140.5	505 800	7.75E−01	0.255
253.5	912 600	6.43E−01	0.442
301.0	1 083 600	5.84E−01	0.538

Table 18.1 (cont.)

Time			
hours	seconds	Acetate molality	$\ln(C^o/C)$
Experiment 28, 381 °C			
1.0	3600	1.05E+00	−0.044
18.5	66600	8.28E−01	0.189
26.5	95400	7.38E−01	0.304
41.5	149400	6.11E−01	0.493
66.0	237600	4.84E−01	0.726
98.0	352800	3.21E−01	1.136
142.0	511200	1.82E−01	1.704
211.0	759600	1.03E−01	2.273
213.5	768600	9.50E−02	2.354



(a) (b)

Figure 18.1 (a) The results of experiments by Palmer and Drummond (1986) on the decomposition of sodium acetate at three temperatures. (b) Rate constants derived from the slopes of the lines in (a), in an Arrhenius plot. Data in Tables 18.1 and 18.2.

18.3 Using the progress variable

Figure 4.9b shows the **G**TP surfaces for the stable and a metastable forms of a system. In Figure 18.2 we show the same thing, but the surfaces represent (metastable) reactants and (stable) products. The Reactants and Products can

Table 18.2 *Rate constants derived from the slopes of the curves in Figure 18.1.*

Expmt	T(°C)	Rate const. (s^{-1})	1/T (K)	ln *K*
26	340	1.570E − 07	0.001 631	−15.764
27	359	4.937E − 07	0.001 582	−14.332
28	381	3.087E − 06	0.001 529	−12.778

represent any chemical reaction whatsoever. An irreversible reaction between these two states is represented by $A' \rightarrow A$ at a fixed T and P. At A′, which is metastable because of some unspecified constraint, we release this constraint momentarily and allow the reaction to proceed irreversibly by an amount $\Delta\xi$ (measured in moles), forming some Products. Then we reapply the constraint, and the system settles down into its new metastable state between A′ and A, with both Reactants and a little bit of Products coexisting. Then we release the constraint momentarily again, another $\Delta\xi$ of reaction occurs, and we reapply the constraint again. Note the difference between this and a reversible reaction. The reversible reaction is also a continuous succession of equilibrium states, but they are stable equilibrium states, having only the normal two constraints, such as T and P, with no third constraint. In Figure 18.2 a reversible reaction would lie entirely on the Products surface, or on the Reactants surface.

Well, that's fine in the abstract, but how do we represent this arrow in thermodynamics? And why would we want to do it? How we do it is simplicity itself and illustrates once again the difference between reality and our model or simulation of reality.

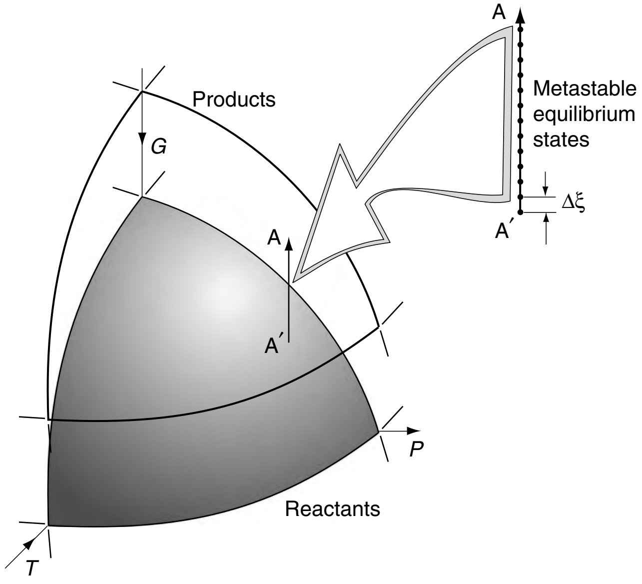


Figure 18.2 A sequence of metastable equilibrium states for the reaction $A' \rightarrow A$ at constant T, P. The progress variable is ξ. Compare with Figure 4.9.

18.3.1 Aragonite–calcite example

Let's consider one of the simplest kinds of reaction, a polymorphic change such as aragonite → calcite. Aragonite on museum shelves actually does not change to calcite at all, but we can do it mathematically with ease. From (18.7) we have

$$\frac{dn_\mathrm{A}}{\nu_\mathrm{A}} = \frac{dn_\mathrm{B}}{\nu_\mathrm{B}} = d\xi \qquad [18.7]$$

where, if A is aragonite and B is calcite, then $\nu_\mathrm{A} = -1$ and $\nu_\mathrm{B} = 1$. Thus

$$dn_\mathrm{aragonite} = -d\xi \qquad (18.25)$$

$$dn_\mathrm{calcite} = d\xi \qquad (18.26)$$

where $n_\mathrm{aragonite}$ is some number of moles of aragonite, and similarly for calcite. If n° is the number of moles of each to start with, then integrating these equations from n° to some new value of n gives

$$\int_{n^\circ}^{n} dn_\mathrm{aragonite} = -\int d\xi$$

$$n_\mathrm{aragonite} - n^\circ_\mathrm{aragonite} = -\Delta\xi \qquad (18.27)$$

and similarly

$$n_\mathrm{calcite} - n^\circ_\mathrm{calcite} = \Delta\xi \qquad (18.28)$$

Equations (18.27) and (18.28) can also be rewritten

$$n_\mathrm{aragonite} = n^\circ_\mathrm{aragonite} - \Delta\xi \qquad (18.29)$$

$$n_\mathrm{calcite} = n^\circ_\mathrm{calcite} + \Delta\xi \qquad (18.30)$$

which shows that whatever amounts of each mineral we have to start with, this amount is decreased by $\Delta\xi$ moles for aragonite and increased by $\Delta\xi$ moles for calcite, every time we allow the reaction (the integration, really) to proceed by $\Delta\xi$. What could be simpler? If we let $n^\circ_\mathrm{calcite} = 0$ and $n^\circ_\mathrm{aragonite} = 1$, and we proceed from pure aragonite to pure calcite in four steps of $\Delta\xi = 0.25$ moles, then after one step $n_\mathrm{aragonite} = 0.75$ moles, $n_\mathrm{calcite} = 0.25$ moles, and so on, and the result can be diagramed as in Figure 18.3. We could, of course, use as many small steps as we like, changing aragonite into calcite quasi-statically, although this never happens in nature.

Iron oxidation example

But we are not restricted to such simple reactions, or to only one reaction. Let's next consider a case where we have two simultaneous reactions,

$$12\,\mathrm{Fe} + 8\,\mathrm{O_2}(g) = 4\,\mathrm{Fe_3O_4} \qquad (18.31)$$

$$4\,\mathrm{Fe_3O_4} + \mathrm{O_2}(g) = 6\,\mathrm{Fe_2O_3} \qquad (18.32)$$

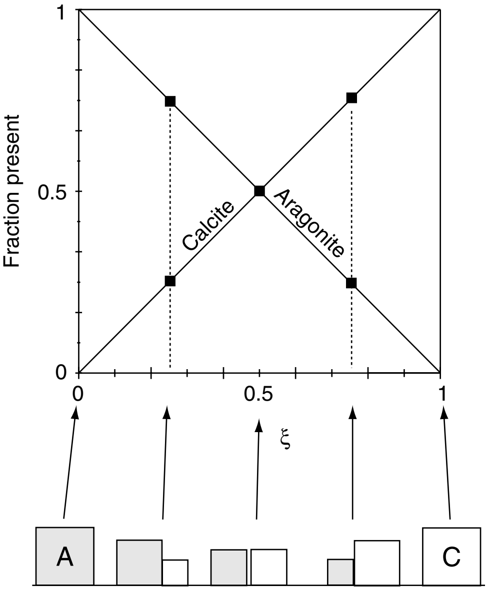


Figure 18.3
The irreversible reaction aragonite (A)→calcite (C) considered as a function of the progress variable ξ.

This is the oxidation of native iron, first to the intermediate stage of magnetite, then to the final product, hematite. In the presence of abundant oxygen at 25 °C, thermodynamics tells us that the stable equilibrium state is hematite plus oxygen; no Fe or Fe_3O_4 should remain. But what happens during the reaction? Does magnetite form and then change to hematite as the equations imply, or is magnetite bypassed completely? Well, what *actually* happens is a matter for experimentation – you must bring iron and oxygen together under various conditions and see what happens. Right now we cannot do that. But what happens in our model of iron oxidation is entirely under our control.

The way we have written the reactions in (18.31) and (18.32), every mole of magnetite that forms eventually gets transformed into hematite. That is, if you add the two reactions together, you get

$$12\,\mathrm{Fe} + 9\,\mathrm{O_2}(g) = 6\,\mathrm{Fe_2O_3} \tag{18.33}$$

and so magnetite does not appear. Another way to show this is to write the reaction progress equations, analogous to (18.25) and (18.26). This time, however, we have two reactions, and $\Delta\xi$ need not be the same for each. For the moment, we will keep them separate, as $\Delta\xi_{18.31}$ and $\Delta\xi_{18.32}$. Thus

$$\left.\begin{aligned} n_{\mathrm{Fe}} &= n^\circ_{\mathrm{Fe}} - 12\,\Delta\xi_{18.31} \\ n_{\mathrm{O_2}} &= n^\circ_{\mathrm{O_2}} - 8\,\Delta\xi_{18.31} - \Delta\xi_{18.32} \\ n_{\mathrm{Fe_3O_4}} &= n^\circ_{\mathrm{Fe_3O_4}} + 4\,\Delta\xi_{18.31} - 4\,\Delta\xi_{18.32} \\ n_{\mathrm{Fe_2O_3}} &= n^\circ_{\mathrm{Fe_2O_3}} + 6\,\Delta\xi_{18.32} \end{aligned}\right\} \tag{18.34}$$

showing that if $\Delta\xi_{18.31} = \Delta\xi_{18.32}$, $n_{\mathrm{Fe_3O_4}}$ stays constant at $n^{\circ}_{\mathrm{Fe_3O_4}}$, or zero if we start with only Fe and oxygen.

But suppose (18.31) proceeds faster than does (18.32), that is, has a greater reaction rate. The reaction rate is defined as $d\xi/dt$, which we could call r, so that the rates for (18.31) and (18.32) are

$$r_{18.31} = \frac{d\xi_{18.31}}{dt} \tag{18.35}$$

$$r_{18.32} = \frac{d\xi_{18.32}}{dt} \tag{18.36}$$

and

$$\frac{r_{18.31}}{r_{18.32}} = \frac{d\xi_{18.31}/dt}{d\xi_{18.32}/dt}$$

Now if reaction (18.31) actually proceeds twice as fast as (18.32), we say $r_{18.31} = 2\,r_{18.32}$, and

$$\begin{aligned} d\xi_{18.31} &= \frac{d\xi_{18.32} r_{18.31}}{r_{18.32}} \\ &= 2\,d\xi_{18.32} \end{aligned} \tag{18.37}$$

and after integration,

$$\Delta\xi_{18.31} = 2\,\Delta\xi_{18.32} \tag{18.38}$$

So note that our thermodynamic simulation can accommodate *relative* reaction rates, but not "real time" rates.

Substituting (18.38) into the progress reactions (18.34), and letting the resulting $\Delta\xi_{18.32}$ be simply $\Delta\xi$, we get

$$\left.\begin{aligned} n_{\mathrm{Fe}} &= n^{\circ}_{\mathrm{Fe}} - 24\,\Delta\xi \\ n_{\mathrm{O_2}} &= n^{\circ}_{\mathrm{O_2}} - 17\,\Delta\xi \\ n_{\mathrm{Fe_3O_4}} &= n^{\circ}_{\mathrm{Fe_3O_4}} + 4\,\Delta\xi \\ n_{\mathrm{Fe_2O_3}} &= n^{\circ}_{\mathrm{Fe_2O_3}} + 6\,\Delta\xi \end{aligned}\right\} \tag{18.39}$$

So that this time magnetite does appear and continues to coexist with hematite, at least until the Fe is all used up.

This result should be fairly intuitive. If magnetite appears and disappears at the same rate, the amount present will not change. But if it forms faster than it disappears, it will accumulate, along with hematite. [We could have got the same result by multiplying all the stoichiometric coefficients in (18.31) by two.] To take a specific case, we could let $n^{\circ}_{\mathrm{Fe}} = 480$ moles, $n^{\circ}_{\mathrm{O_2}} = 510$

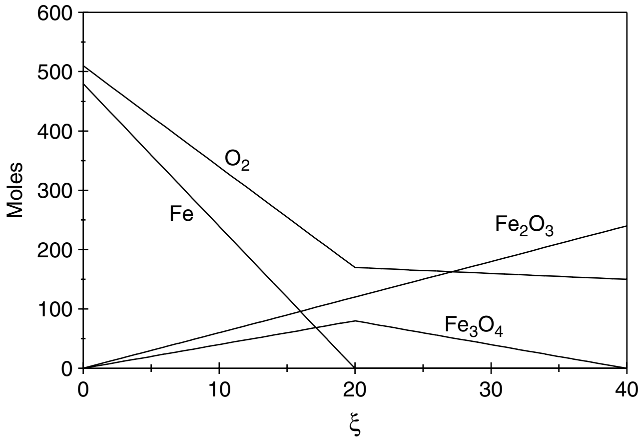


Figure 18.4 Two simultaneous iron oxidation reactions as a function of ξ.

moles, $n^\circ_{Fe_3O_4} = 0$, and $n^\circ_{Fe_2O_3} = 0$. The numbers for Fe and oxygen are chosen so that oxygen is in excess of that required to oxidize all the Fe to hematite, and also so that the Fe will be used up at an even number of $\Delta\xi$ steps. This is not necessary – any starting numbers can be used. The resulting graph of moles versus ξ is shown in Figure 18.4. Note that after all the Fe is used up at $\xi = 20$ $(480 - 20 \times 24 = 0)$, reaction (18.31) is no longer available, and so the progress equations (18.39) change to reflect (18.32) only. Note too that when the magnetite has finally disappeared, we have used 480 moles of Fe and $510 - 150 = 360$ moles of O_2 to produce 240 moles of Fe_2O_3, or

$$480\,\mathrm{Fe} + 360\,\mathrm{O_2}(g) = 240\,\mathrm{Fe_2O_3}$$

which is consistent with (18.33). The final equilibrium state is independent of how we get there, but the *reaction path* depends greatly on kinetics.

A solution example

The iron oxidation example is actually a special case of two simultaneous reactions, where the reactants and products do not change activities as the reaction proceeds. In the more general case of gaseous or aqueous solutions, two simultaneous reactions involving gases or solutes will be a bit more complex because the concentrations or activities of both reactants and products will change during the reaction, and therefore the rate of each reaction will change continuously (except in the special case of zeroth order reactions). For example, consider the following simultaneous reactions,

$$\mathrm{A} \rightarrow \mathrm{B} \qquad (18.40)$$

$$\mathrm{B} \rightarrow \mathrm{C} \qquad (18.41)$$

where A, B, and C can be gaseous compounds or aqueous solutes, and where the rate constants are $k_{18.40}$ for A → B, and $k_{18.41}$ for B → C. The rate equations are [cf. (18.12)]

$$-\frac{dC_\mathrm{A}}{dt} = k_{18.40}C_\mathrm{A} \tag{18.42}$$

$$\frac{dC_\mathrm{B}}{dt} = k_{18.40}C_\mathrm{A} - k_{18.41}C_\mathrm{B} \tag{18.43}$$

$$\frac{dC_\mathrm{C}}{dt} = k_{18.41}C_\mathrm{B} \tag{18.44}$$

If the initial concentrations are C_A°, $C_\mathrm{B}^\circ = 0$, and $C_\mathrm{C}^\circ = 0$, (18.42) gives

$$C_\mathrm{A} = C_\mathrm{A}^\circ e^{-k_{18.40}t}$$

Equation (18.43) then becomes

$$\frac{dC_\mathrm{B}}{dt} = k_{18.40}C_\mathrm{A}^\circ e^{-k_{18.40}t} - k_{18.41}C_\mathrm{B} \tag{18.45}$$

With a little ingenuity, this is integrated to give

$$C_\mathrm{B} = \frac{k_{18.40}C_\mathrm{A}^\circ}{(k_{18.41} - k_{18.40})}\left(e^{-k_{18.40}t} - e^{-k_{18.41}t}\right) \tag{18.46}$$

Then C_C is obtained from the mass balance,

$$C_\mathrm{C} = C_\mathrm{A}^\circ - C_\mathrm{A} - C_\mathrm{B} \tag{18.47}$$

The result, for $k_{18.40} = 1, k_{18.41} = 0.5$, is shown in Figure 18.5. Note the general similarity to Figure 18.4, in the sense that the intermediate compound (B, Fe_3O_4) accumulates, then disappears as equilibrium is approached. The amount of intermediates formed is controlled entirely by the kinetics of the reactions.

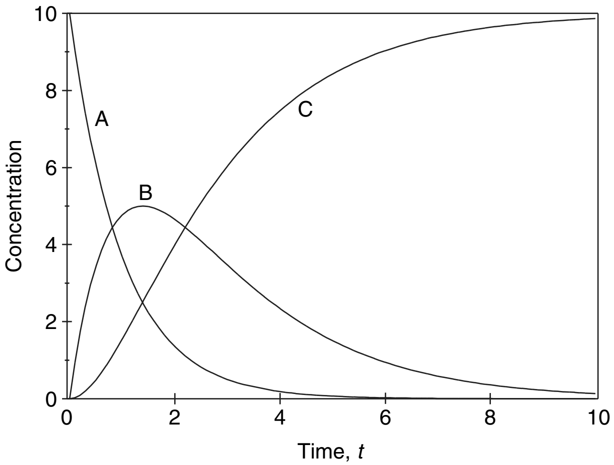


Figure 18.5
Two simultaneous gas reactions as a function of time, *t*.

18.3.2 The reaction of K-feldspar with water

Finally, having discussed some of the basic concepts, we get to an example that involves some interesting mineralogical results. We have seen that during the course of a reaction, intermediate compounds may appear and disappear, depending on the reaction kinetics. Actually, this may happen in more complex systems even if kinetics are not considered, or more exactly, if the rates of all reactions involved are the same.

The dissolution of K-feldspar in water is now a classic example of the usefulness of the reaction path or process modeling approach, using quasi-static reactions. It was first used by Helgeson and his colleagues in the 1960s when the methods were developed. A useful summary is Helgeson (1979). In this case, the problem is to unravel all the intermediate reactions that might occur during the reaction of K-feldspar with water. We have already considered this system in Chapter 16, where we saw that several reactions were possible, involving muscovite, kaolinite, and other minerals. However, in Chapter 16 we did not explicitly consider how the solution might come to have a certain pH or $a_{\mathrm{K}^+}/a_{\mathrm{H}^+}$ ratio. These were considered to be controlled by outside influences. This time we will start with the metastable system K-feldspar + water (metastable because the two reactants are separated) and follow the reactions that occur when the two react quasi-statically. Because we will be considering very small increments of ξ, the process can be imagined as dropping tiny grains of K-feldspar into a large tank of water, as in Figure 18.6. After each grain is added, we wait for the water in the tank to reach equilibrium, and then we add another grain, and so on, until K-feldspar is in equilibrium with the solution.

The reaction path

A surprising number of things happen when the reaction is considered in this way. When the first few grains drop in and dissolve completely, the dissolution

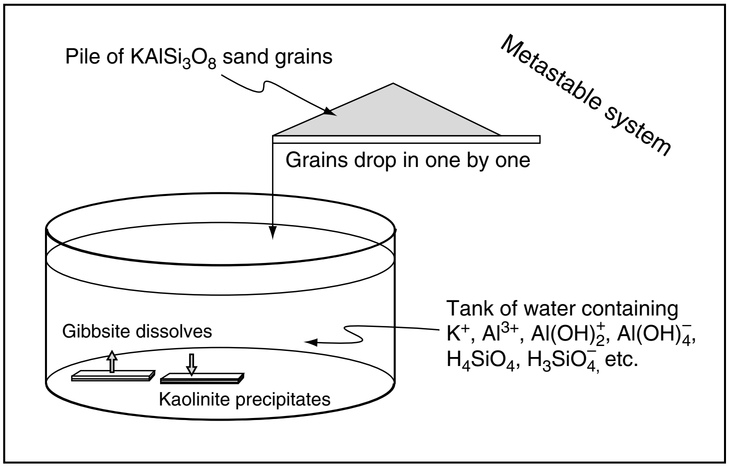


Figure 18.6 Reaction path model for the reaction of K-feldspar with water. Gibbsite-dissolving, kaolinite-precipitating is represented by path B→C in Figure 18.7.

must be *congruent* (i.e., the solute produced by K-feldspar dissolution has the same stoichiometry as the feldspar), but of course we must know the nature and thermodynamic properties of the solute species. According to our present knowledge of this system there are quite a few, including K^+, Al^{3+}, $Al(OH)^{2+}$, $Al(OH)_4^-$, H_4SiO_4, and $H_3SiO_4^-$. There are a few others that we can omit without serious error. If you perform a speciation calculation of the kind discussed in §16.3 on a solution having K, Al, and Si in the proportions 1:1:3 (as in K-feldspar) at very low concentration (after the first grain has dropped in), you find that the dominant species are K^+, $Al(OH)_4^-$, and H_4SiO_4. Essentially, $Al(OH)_4^-$ must dominate the Al species to maintain a charge balance with K^+. Therefore, the dissolution reaction of K-feldspar can be approximated by

$$KAlSi_3O_8 + 8H_2O \rightarrow K^+ + Al(OH)_4^- + 3H_4SiO_4 \tag{18.48}$$

and the three species on the right will steadily increase in concentration as more and more grains of K-feldspar dissolve. However, this does not tell the whole story. Also increasing during dissolution of the feldspar are all the other species produced [Al^{3+}, $Al(OH)^{2+}$, and $H_3SiO_4^-$], though at lower concentrations. Initially, the concentrations of all these species are so small that the concentrations of H^+ and OH^- remain constant at 10^{-7}. The species all increase from zero, maintaining the overall 1:1:3 stoichiometry, and stop when the solution becomes saturated with K-feldspar. However, *before* that happens, the solution may become saturated with other minerals, which will precipitate as K-feldspar continues to dissolve. This results in *incongruent* dissolution, because the stoichiometry of K:Al:Si in the solution will no longer be the same as in feldspar.

To find out whether the solution has become saturated with another mineral, the solubility products of all minerals in the system considered (i.e., all minerals that contain any combination of the elements in the system) must be compared against the corresponding ion activity product (IAP, §16.2.2) in the solution after each increment of dissolution $\Delta\xi$. This can be literally hundreds of minerals in large model systems. In the relatively simple K-feldspar case, there are only a few minerals that could possibly form. The first of these is gibbsite. The solubility product for gibbsite is

$$Al(OH)_3(s) = Al^{3+} + 3OH^-; \quad K_{sp} = a_{Al^{3+}} a_{OH^-}^3 \tag{18.49}$$

Saturation in gibbsite occurs when its activity product exceeds the solubility product for gibbsite:

$$a_{Al^{3+}} a_{OH^-}^3 > K_{sp} \tag{18.50}$$

When this happens, the a_{K^+}, a_{H^+}, and $a_{SiO_2(aq)}$ values of the solution result in it plotting at point A in Figure 18.7. Now if K-feldspar continues to dissolve (or,

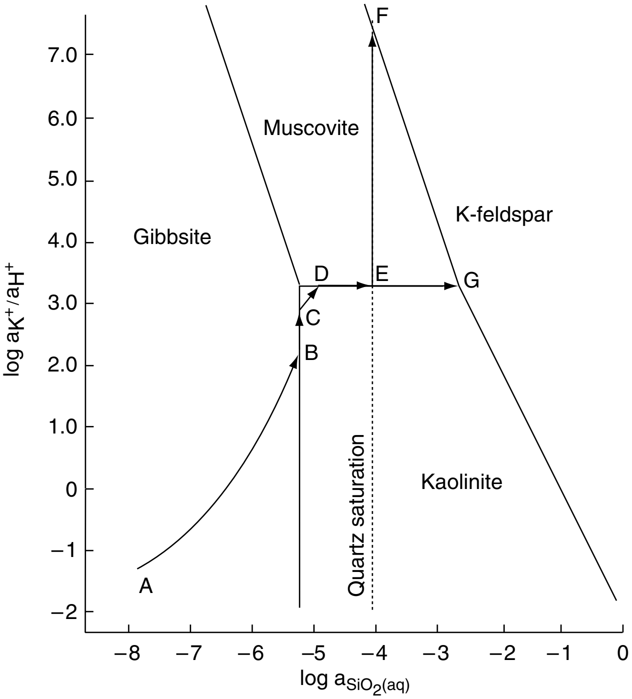


Figure 18.7 The reaction path from Figure 18.6 plotted in $\log a_{K+}/a_{H+}$ versus $\log a_{SiO_2(aq)}$ space at 25 °C, 1 atm.

if you prefer, we continue to perform speciation calculations for solutions in which K, Al, and Si continue to increase by $\Delta\xi$ in the ratio 1:1:3), gibbsite will continue to precipitate, and the remaining solution will have compositions that follow the path A→B in Figure 18.7. During this process, the silica content of the solution continues to increase until, at point B, a silica-bearing mineral (kaolinite) becomes stable. The coexistence of gibbsite and kaolinite buffers the activity of silica according to

$$2\mathrm{Al(OH)_3}(s) + 2\mathrm{SiO_2}(aq) = \mathrm{Al_2Si_2O_5(OH)_4}(s) + \mathrm{H_2O}$$

Therefore, as K-feldspar continues to dissolve, aqueous SiO_2 does not increase, but is used to convert previously precipitated gibbsite into kaolinite. K^+ continues to increase, and the net result is the path B→C. This is the part of the path shown in Figure 18.6. At C, all gibbsite is used up, and the solution composition can resume increasing in silica content, following a path C→D roughly parallel to its original path (A→B), only this time precipitating kaolinite rather than gibbsite. Along C→D, K, Al, and Si are all increasing in solution until, at D, K-mica (muscovite) begins to precipitate, because its solubility product is exceeded. Again, coexistence of minerals buffers a solution parameter, this time (a_{K^+}/a_{H^+}), through the relation

$$\mathrm{KAl_3Si_3O_{10}(OH)_2} + \tfrac{3}{2}\mathrm{H_2O} + \mathrm{H^+} = \tfrac{3}{2}\mathrm{Al_2Si_2O_5(OH)_4} + \mathrm{K^+}$$

Because the ratio (a_{K^+}/a_{H^+}) is fixed, but SiO_2 continues to increase as K-feldspar dissolves, kaolinite reacts to form muscovite, and the solution follows path D→E, at which point the solution becomes saturated with quartz, and if equilibrium is maintained, quartz will begin to precipitate. With four components (K_2O, Al_2O_3, SiO_2, H_2O), a maximum of four phases can coexist at our arbitrarily chosen T and P (25 °C, 1 atm) according to the phase rule. With quartz, muscovite, kaolinite, and water, this number has now been reached and cannot be exceeded (K-feldspar doesn't count; it is being used as a source of solutes, and has not yet equilibrated with the solution). Therefore, if we continue to add K_2O, Al_2O_3, and SiO_2 from the K-feldspar to the solution, the solution will stay at point E while kaolinite reacts with the solution to form muscovite, and quartz continues to precipitate. When kaolinite is all used up, additional dissolution of K-feldspar will drive the solution composition along E→F, with the SiO_2 content of the solution buffered by the presence of quartz. At point F, K-feldspar finally becomes stable.

Alternatively, at point E, if quartz does not precipitate, the solution composition could continue from E to G, where K-feldspar would also become stable, but this time in a solution oversaturated with quartz. It would coexist metastably with muscovite and kaolinite, rather than stably with muscovite and quartz. In nature, quartz quite often does not precipitate at low temperatures, and in computer calculations that simulate equilibrium, it can be prevented from "precipitating" by removing it from the list of minerals available to the program.

The numerical model

The description of the reaction path given above is not just imagined; it is the result of reaction path calculations similar in principle to the aragonite→calcite and iron oxidation examples we have already considered. However, the K-feldspar dissolution reaction is obviously a bit more complicated. For one thing, the dissolution reaction considered in the calculations is not (18.48), but the complete reaction involving all known species believed to be significant. This reaction obviously changes from time to time, as minerals appear and disappear. For the path A→B in Figure 18.7, gibbsite is considered to be in equilibrium with the aqueous solution, all dissolved aqueous species are assumed to be equilibrated, and the dissolution reaction is

$$\begin{aligned}
&\overline{\nu}_{\mathrm{KAlSi_3O_8}}\mathrm{KAlSi_3O_8} + \overline{\nu}_{\mathrm{H_2O}}\mathrm{H_2O} \\
&\to \overline{\nu}_{\mathrm{Al(OH)_3}}\mathrm{Al(OH)_3}(\mathit{gibbsite}) \\
&\qquad + \overline{\nu}_{\mathrm{K^+}}\mathrm{K^+} + \overline{\nu}_{\mathrm{Al^{3+}}}\mathrm{Al^{3+}} + \overline{\nu}_{\mathrm{Al(OH)^{2+}}}\mathrm{Al(OH)^{2+}} + \overline{\nu}_{\mathrm{Al(OH)_4^-}}\mathrm{Al(OH)_4^-} \\
&\qquad + \overline{\nu}_{\mathrm{H_4SiO_4}}\mathrm{H_4SiO_4} + \overline{\nu}_{\mathrm{H_3SiO_4^-}}\mathrm{H_3SiO_4^-} + \overline{\nu}_{\mathrm{H^+}}\mathrm{H^+} + \overline{\nu}_{\mathrm{OH^-}}\mathrm{OH^-}
\end{aligned} \tag{18.51}$$

where the quantities $\overline{\nu}_i$ above are the stoichiometric coefficients of each species, but with the overbar added to indicate that they are not constants as before, but variables to be calculated. Although the stoichiometric coefficients in each elementary or individual reaction such as (18.49) are always integers, the overall dissolution reactions such as (18.51) are made up of numerous simultaneous individual reactions, and the coefficients become nonintegral. Calculating them is part of the problem.

A convenient way to start would be to let the system contain 1000 g water and assign $\overline{\nu}_{\mathrm{KAlSi_3O_8}} = -1$. Each $\overline{\nu}_i$ then represents the rate of change in the molality of the subscripted species i with a reaction step $d\xi$:

$$\frac{dm_i}{d\xi} = \overline{\nu}_i \tag{18.52}$$

which is analogous to Equation (18.8). To solve for each individual $\overline{\nu}_i$ we need 10 equations in the 10 unknowns $\overline{\nu}_i$, and these are provided by a combination of equilibrium constant expressions and mass balances for K, Al, Si, O, and H. Solving these 10 equations using matrix algebra results in the 10 values of $\overline{\nu}_i$ in (18.51), recalling that $\overline{\nu}_{\mathrm{KAlSi_3O_8}} = -1$. Then using

$$m_i = m_i^\circ + \overline{\nu}_i \Delta\xi \tag{18.53}$$

which is equivalent to (18.29) and (18.30) in our aragonite→calcite example, and (18.39) in our iron-oxidation example. New concentrations of all aqueous species are calculated after each reaction increment, as well as new quantities of any solids that are dissolving or precipitating. Inclusion of information on the relative reaction rates, often using the approximation afforded by transition state theory for reactions with unknown rates, is easily done as shown in the iron-oxidation example. These calculations, plus calculation of all individual activity coefficients, is repeated over and over (iteratively) until the final equilibrium state is reached.

A useful introduction to modeling reaction paths is Steinmann et al. (1994). They show that the reaction path for K-feldspar dissolution as projected in $\log a_{\mathrm{K^+}}/a_{\mathrm{H^+}}$ versus $\log a_{\mathrm{SiO_2}(aq)}$ space is highly dependent on the starting solution composition, and that with certain assumptions, the path can be calculated using a spreadsheet.

18.3.3 Comment

Clearly there is no theoretical limit to the complexity of the reactions that might be considered in this way. In addition, it is quite possible to *couple* this type of calculation with other types, such as fluid flow, heat flow, pressure changes, diffusion, permeability changes, deformation, and so on, because these other model calculations also are carried out iteratively in a large series of small steps. Thus, for example, after carrying out a $\Delta\xi$ step in a reaction path model, we could then take a small step in a heat flow model, then a small step in a fluid flow model, and then return to the reaction model, and so on. The heat flow calculation would depend on the results of the reaction path, because the reaction would be exo- or endothermic, and the fluid flow calculation would depend on any permeability changes caused by the chemical reactions, as well as on viscosity changes caused by changes in temperature. Each model result is dependent on all the others in a coupled fashion. Naturally, the demands on data, scientific insight, programming talent, and computer resources grows enormously with each additional factor to be modeled. This area is one of considerable activity at the present time.

Why would we want to perform such complex calculations, especially when some of them may be quite unrealistic? For example, we know that quasi-static reactions are unrealistic, yet they are computed everywhere these days. Well, you could easily say that the equilibrium thermodynamic model is unrealistic too, yet it is quite useful. Reaction path models are useful because they help us shape our ideas about what might be happening in natural processes. You need not believe the results of such models in an absolute sense, but they usually reflect some or many aspects of what is really happening in the natural situation being simulated.

Besides, the game is not over. The comparison of model results with natural situations, and the improvement of the models to better simulate nature, is itself an iterative process, which will continue as long as we continue to be interested in understanding the world around us. Most scientists believe that if you can make a mathematical model of a situation, no matter how crude, you will understand that situation at a deeper level than if you just speculate about it. Poincaré (1952) said “It is far better to predict without certainty, than never to have predicted at all.” Model building is thus important, but it must be combined with careful observation of nature, whether in the laboratory or in the field, otherwise it will be quite inappropriate and perhaps useless. It is quite possible to calculate exact answers to real problems that are quite meaningless, no matter how clever the mathematics is.

18.4 Affinity and the progress variable

A striking feature of the reaction path calculations we have been considering is that in a sense they are independent of the Gibbs energy, which we know controls which way these reactions proceed. In other words, by changing the sign in equations like (18.25) and (18.26) we can use exactly the same methods to change calcite into aragonite, graphite into diamond, or to form K-feldspar from the dissolved species in solution, like running a movie backwards. To investigate this further, we must link the progress variable with the Gibbs energy. The K-feldspar case is a little too complex, and the calcite–aragonite case is too simple. We choose a system that has an equilibrium constant for one reaction, and a manageable number of species, the system N_2–H_2.

18.4.1 Components and species again

Consider the two-component system N_2–H_2. Components N_2 and H_2 refer to the result you would get if you analyzed the system for total nitrogen and total hydrogen. The actual molecular form taken by each element in the system is irrelevant. However, N_2 and H_2 might also refer to the *species* in the system, i.e., diatomic nitrogen molecules and diatomic hydrogen molecules. In real systems, species N_2 and H_2 combine to form ammonia,

$$\mathrm{N_2}(g) + 3\,\mathrm{H_2}(g) = 2\,\mathrm{NH_3}(g) \tag{18.54}$$

so that there are at least three major species. In many systems, there may be only a few components, but dozens or hundreds of species.

Equation (4.65)

$$d\mathbf{G} = -\mathbf{S}dT + \mathbf{V}dP + \sum_i^c \mu_i dn_i \qquad [4.65]$$

was derived for systems of c independent components. For example, in the two-component system N_2–H_2, $c = 2$, and the final term on the right side would be

$$\mu_{\mathrm{N_2}} dn_{\mathrm{N_2}} + \mu_{\mathrm{H_2}} dn_{\mathrm{H_2}}$$

However, the equation can also be used with i representing some or all of the *species* in the system, rather than independent components, as long as the species are related to one another in balanced chemical reactions, such that the number of independent compositional parameters remains equal to c. Thus the final term on the right side of (4.65) could also read

$$\mu_{\mathrm{N_2}} dn_{\mathrm{N_2}} + \mu_{\mathrm{H_2}} dn_{\mathrm{H_2}} + \mu_{\mathrm{NH_3}} dn_{\mathrm{NH_3}}$$

where N_2 and H_2 are now species and not components. Although we now have three compositional terms, we also have an equation (18.54) relating them (assuming equilibrium), so there are still only two independent compositional terms.

Changing from components to species in this way provides for some flexibility we shall take advantage of shortly. In a closed system, we cannot change the chemical potentials of components N_2 and H_2, and the last term in (4.65) is zero. However, the chemical potentials of the species in (18.54) *can* change in a closed system, if the reaction progresses to the left or the right from some metastable state towards the stable equilibrium state. The last term in (4.65) (which would look a bit different in having $\sum_{i=1}^{s}$ rather than $\sum_{i=1}^{c}$, that is, s species rather than c components) would not be zero, even in a closed system, if we were considering chemical reactions progressing towards equilibrium. Furthermore, in such cases this last term must always be negative, no matter in which direction the reaction proceeds, to be consistent with $d\mathbf{G}_{T,P} \leq 0$. One other thing to note is that, although in this case there is one reaction and three species, in other cases you may be considering one out of many simultaneous reactions, and many species. In these cases, all possible reactions must be proceeding towards equilibrium at each increment of ξ, not just the one you happen to be considering.

18.4.2 The affinity

Considering a system at constant T and P, Equation (4.65) now shows that

$$d\mathbf{G}_{T,P} = \sum_{i=1}^{s} \mu_i dn_i \tag{18.55}$$

where the i are not the c independent components, but the s species we are considering which form from those components. In our ammonia example, Equation (18.54), $s = 3$. Into this we substitute the relation $dn_i = \nu_i d\xi$ from (18.8), to get

$$d\mathbf{G}_{T,P} = \sum_{i=1}^{s} (\nu_i \mu_i) d\xi \tag{18.56}$$

Comparing this to (4.46),

$$d\mathbf{G} = -\mathbf{S}\,dT + \mathbf{V}\,dP - \mathcal{A}\,d\xi \qquad [4.46]$$

we see that another definition of the affinity is

$$\mathcal{A} = -\sum_{i}^{s} \nu_i \mu_i \tag{18.57}$$

Recalling that our ν_i are positive for products and negative for reactants, the quantity $\sum_i \nu_i \mu_i$ is simply the difference in partial molar Gibbs energy between products and reactants. For (18.5) this is

$$\underbrace{(3\,\mu_{\mathrm{C}} + 4\,\mu_{\mathrm{D}})}_{\text{products}} - \underbrace{(\mu_{\mathrm{A}} + 2\,\mu_{\mathrm{B}})}_{\text{reactants}} \tag{18.58}$$

If this sum is zero, the reaction is at stable equilibrium, there is no third constraint, the affinity is zero, the third term on the right side of (4.46) disappears, and both

$$d\mathbf{G} = -\mathbf{S}\,dT + \mathbf{V}\,dP$$

and (4.40)

$$dG = -S\,dT + V\,dP \qquad [4.40]$$

apply to the (closed) system. If some constraint prevents reaction (18.5) from proceeding to equilibrium, then the sum in (18.58) is not zero, but may be positive or negative, depending on whether (18.5) wants to go to the left or right. If the sum $\sum_i \nu_i \mu_i$ is positive ($\mathcal{A}$ negative), the reaction wants to go to the left as written, $d\xi$ is negative, and $\mathcal{A}\,d\xi$ is positive. If the sum is negative ($\mathcal{A}$ positive), the reaction proceeds to the right, $d\xi$ is positive, and $\mathcal{A}\,d\xi$ is positive. So $\mathcal{A}\,d\xi$ is inherently positive (or zero), and $d\mathbf{G}_{T,P}$ is inherently negative for a spontaneous reaction, consistent with (from 4.46)

$$d\mathbf{G}_{T,P} = -\mathcal{A}d\xi \tag{18.59}$$

Equation (18.57) shows that the affinity is a $\Delta\mu_{T,P}$ term, giving the "distance" in **G** between stable and metastable equilibrium surfaces, or states, and the amount of useful work that a chemical reaction can do as it reaches equilibrium. It is in fact represented by the vector $A \to A'$ in Figures 4.9b and 18.2. It might be worth noting, too, that the units of ξ are moles, of $\mathcal{A}$ are J mol^{-1}, and so the Gibbs energy in equations such as (18.59) is the total, not the molar, Gibbs energy. As mentioned above, a common modeling practice is to define the system as containing one kilogram of water, so that all mole numbers (n_i) become molalities (m_i).

18.4.3 Ammonia formation example

Now we can consider the formation of ammonia from hydrogen and nitrogen (reaction 18.54) patterned after a similar presentation in Denbigh (1981, §4.3). Consider a system at T and one bar consisting of one mole of pure N_2 and three moles of pure H_2 kept separate from one another. This system has four constraints. In addition to T and P (the only constraints needed to define stable equilibrium), a third constraint is the separation of the gases, and the fourth is a constraint on the reaction to form NH_3. We release one constraint by allowing the gases to mix, resulting in a metastable gaseous solution of N_2 and H_2. Then we allow reaction (18.54) to take place in the forward direction in 10 increments, from pure $N_2 + 3\,H_2$ to two moles of pure NH_3. Thus $dn_{NH_3}/d\xi = 2$, and ξ increases from 0 to 1.0 during the reaction. After each increment of reaction, $n_{N_2} = (1-\xi)$, $n_{H_2} = 3(1-\xi)$, and $n_{NH_3} = 2\xi$. The total number of moles in the system at any stage is

$$\begin{aligned} n_{N_2} + n_{H_2} + n_{NH_3} &= (1-\xi) + (3-3\xi) + (2\xi) \\ &= 4 - 2\xi \end{aligned}$$

and the mole fractions are

$$\left.\begin{aligned} x_{N_2} &= \frac{1-\xi}{4-2\xi} \\ x_{H_2} &= \frac{3-3\xi}{4-2\xi} \\ x_{NH_3} &= \frac{2\xi}{4-2\xi} \end{aligned}\right\} \tag{18.60}$$

The goal is to minimize an expression for the Gibbs energy of this solution, so starting with expressions from Chapter 7,

$$\Delta_{\text{mix}} G_{\text{ideal sol'n}} = G_{\text{ideal sol'n}} - \sum_i x_i G_i^\circ \qquad [7.19]$$

$$= RT \sum_i x_i \ln x_i \qquad [7.20]$$

we write

$$G_{\text{ideal sol'n}} = \sum_i x_i G_i^\circ + RT \sum_i x_i \ln x_i \qquad (18.61)$$

In the present case, this becomes (for pure substances $G^\circ = \mu^\circ$)

$$\begin{aligned} G_{\text{ideal sol'n}} &= x_{N_2}\mu^\circ_{N_2} + x_{H_2}\mu^\circ_{H_2} + x_{NH_3}\mu^\circ_{NH_3} \\ &\quad + RT[x_{N_2}\ln x_{N_2} + x_{H_2}\ln x_{H_2} + x_{NH_3}\ln x_{NH_3}] \end{aligned} \qquad (18.62)$$

To convert molar G to the total $\mathbf{G}$ of the solution, we multiply both sides by the denominator of the mole fraction term, $(4-2\xi)$, so

$$\begin{aligned} \mathbf{G}_{\text{ideal sol'n}} &= n_{N_2}\mu^\circ_{N_2} + n_{H_2}\mu^\circ_{H_2} + n_{NH_3}\mu^\circ_{NH_3} \\ &\quad + RT[n_{N_2}\ln x_{N_2} + n_{H_2}\ln x_{H_2} + n_{NH_3}\ln x_{NH_3}] \end{aligned} \qquad (18.63)$$

Substituting for n_i,

$$\begin{aligned} \mathbf{G}_{\text{ideal sol'n}} &= (1-\xi)\mu^\circ_{N_2} + (3-3\xi)\mu^\circ_{H_2} + 2\xi\mu^\circ_{NH_3} \\ &\quad + RT[(1-\xi)\ln x_{N_2} + (3-3\xi)\ln x_{H_2} + 2\xi \ln x_{NH_3}] \\ &= (\mu^\circ_{N_2} + 3\mu^\circ_{H_2}) + \xi(2\mu^\circ_{NH_3} - \mu^\circ_{N_2} - 3\mu^\circ_{H_2}) \\ &\quad + RT[(1-\xi)\ln x_{N_2} + (3-3\xi)\ln x_{H_2} + 2\xi \ln x_{NH_3}] \end{aligned} \qquad (18.64)$$

so

$$\begin{aligned} \mathbf{G}_{\text{ideal sol'n}} - (\mu^\circ_{N_2} + 3\mu^\circ_{H_2}) &= \xi(\Delta_r\mu^\circ) + RT[(1-\xi)\ln x_{N_2} \\ &\quad + (3-3\xi)\ln x_{H_2} + 2\xi\ln x_{NH_3}] \qquad (18.65) \\ &= \xi(\Delta_r\mu^\circ) + \Delta_{\text{mix}}\mathbf{G} \end{aligned}$$

where $\mathbf{G}$ is the total Gibbs energy of the system containing $(4-2\xi)$ moles of nitrogen, hydrogen and ammonia gases, assuming ideality, $\Delta_r\mu^\circ$ is the standard Gibbs energy of reaction (18.54) at T from SUPCRT92, and ξ is the progress variable which can have any value between 0 and 1.0. Note that the $RT[\ldots]$ term is simply the total energy form of (7.20), and can be called $\Delta_{\text{mix}}\mathbf{G}$, the total Gibbs energy of mixing. All we have done is to convert Equation (18.61)

Reaction deltas – *d*, Δ and ▲

On page 29 we looked at the meaning of $\Delta_r V°$, the value of which changes depending on how you write the reaction, and in §3.4.1, page 38, we used differential notation where we could easily have used delta notation. For example, we wrote $\boldsymbol{w} = mg \cdot dh$ where we could just as easily have written $\boldsymbol{w} = mg \cdot \Delta h$. Because there is a persistent tendency to see differentials as infinitesimals, we made the point that dh is not *necessarily* infinitesimal, so in a sense we have not changed anything in switching from Δ to d. What we said there is true enough, but it is not the whole story.

Δ is a scalar difference between any two quantities. It could be the difference in weight of two rock specimens. A differential dx, on the other hand, although it need not be infinitesimal (see Figure C.1), is not *just* the difference between any two quantities called x. It means that there is a mathematical function for which it represents some increment in the x variable. Therefore $dG_{T,P} = 0$ has a significantly different meaning than $\Delta G_{T,P} = 0$. $dG_{T,P} = 0$ signifies the extremum of a mathematical function, while $\Delta G_{T,P} = 0$ does not.

Another point that is sometimes made about "delta notation" is that it is commonly used in two different ways (MacDonald, 1990; Spencer, 1973). One way is the scalar difference just mentioned, as in describing the difference in V, G, etc., between two equilibrium states; for example, before and after the expansion of a gas. The other is in using the progress variable, as in

$$\begin{aligned}\left(\frac{\partial \mathbf{G}}{\partial \xi}\right)_{T,P} &= -a\mu_{\mathrm{A}} - b\mu_{\mathrm{B}} + c\mu_{\mathrm{C}} + d\mu_{\mathrm{D}} \\ &= \Delta\mu \text{ or, sometimes, } \Delta G\end{aligned}$$

which is the application of Equation (18.59) to reaction (18.1). The point is that in this case Δ refers to the slope of the **G** versus ξ curve, the instantaneous rate of change of **G**. To eliminate the supposed confusion, a different symbol is recommended for this usage, such as ▲ or $\tilde{\Delta}$, so that Equation (9.8) would become

$$\blacktriangle G = \blacktriangle G° + RT \ln Q \quad \text{[MacDonald, 1990, Equation (4)]}$$

This appears to be an idea whose time has not yet come.

into Equation (18.65) by switching to total Gibbs energy and introducing our variable mole fractions (18.60).

Evidently we can plot values of $[\mathbf{G} - (\mu^\circ_{\mathrm{N_2}} + 3\mu^\circ_{\mathrm{H_2}})]$ as a function of reaction progress at various Ts, and because the two $\mu°$ terms are constants, the curve will show the true shape of the Gibbs energy variation and the minimum at the equilibrium value of ξ. Recall that $\xi = 0$ means a solution of one mole

Table 18.3 *Data from Equation (18.65) for $T = 200\,°C$.*

ξ	$\xi(\Delta_r\mu°)$ J mol^{-1}	$\Delta_{\text{mix}}\mathbf{G}°$ J mol^{-1}	$\mathbf{G} - (\mu°_{N_2} + 3\mu°_{H_2})$ J mol^{-1}
0.0	0	−8849	−8849
0.1	392	−11047	−10655
0.2	784	−12020	−11236
0.3	1176	−12427	−11252
0.4	1567	−12389	−10821
0.5	1959	−11937	−9977
0.6	2351	−11062	−8711
0.7	2743	−9714	−6971
0.8	3135	−7780	−4645
0.9	3527	−4989	−1462
1.0	3918	0	3918

N_2 and three moles H_2, and $\xi = 1$ means two moles of pure NH_3. Somewhere in between there is an equilibrium composition where reaction (18.54) is at equilibrium, and $\mu_{N_2} + 3\mu_{H_2} = 2\mu_{NH_3}$. Table 18.3 shows the values of the terms in Equation (18.65) at 200 °C, and they are plotted in Figure 18.8. The values of $\Delta_r\mu°(\equiv \Delta_r G°)$ are obtained from a least squares fit of data from SUPCRT92, which is $\Delta_r\mu° = -97\,116.2 + 213.536\,T(\text{K})$.

Note how the mixing term alone gives a minimum at $\xi = 0.34$, but the curve is "tilted" to lower ξ values by the contribution of the $\Delta_r\mu°$ term. Also note that there is a large energy drop at $\xi = 0$, before any ammonia has formed. This is due entirely to the mixing of N_2 and H_2 before the reaction starts. At $\xi = 1$, there is zero energy of mixing, and the system (pure NH_3) has a higher Gibbs energy than pure $N_2 + 3\,H_2$.

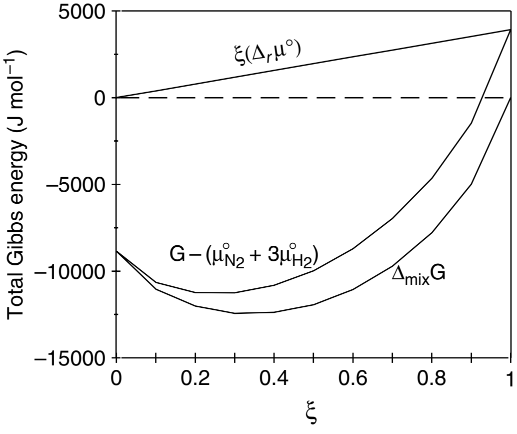


Figure 18.8 The energy of mixing and the reaction energy combine to give a minimum in system **G** at $\xi = 0.252$ at 200 °C.

Equation (18.65) can be differentiated with respect to ξ, and the resulting expression equated to zero to solve for the minimum value of ξ. You won't want to do this without a program that does symbolic algebra because the differentiated expression is quite lengthy, but the result at 200 °C is $\xi = 0.252$.

Performing the same calculations at temperatures from 25 to 300 °C reveals that the equilibrium ξ value for this reaction changes considerably with temperature, from almost pure product at 25 °C to almost pure reactants at 300 °C, as shown in Figure 18.9.

Ammonia speciation

The mixing curve in Figure 18.8 is useful to show that the **G** function actually does have a minimum, and how the progress variable can be used to simulate stages in an irreversible reaction. In fact in progressing past the equilibrium composition, we actually drove the reaction "backwards" all the way to pure NH_3. In that sense it is no different from our previous examples, except that now we can see which direction is spontaneous and which is "backwards."

But apart from reaction path models, we are usually interested in the stable equilibrium state of a system, rather than "artificially" constructed mixing

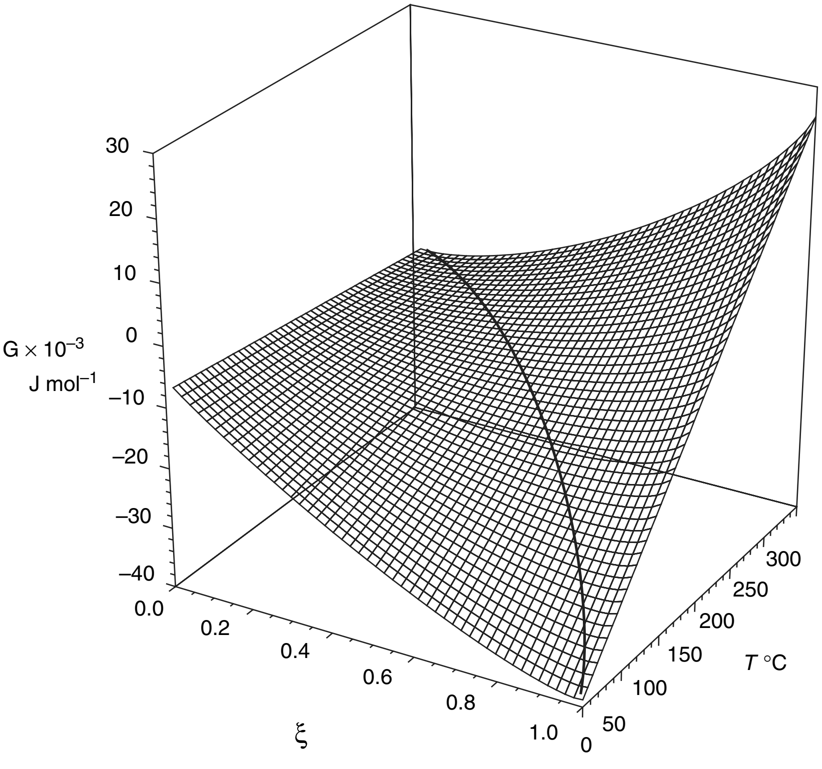


Figure 18.9
The $[\mathbf{G} - (\mu^\circ_{N_2} + 3\mu^\circ_{H_2})]$–$T$–$\xi$ surface for reaction (18.54). The approximate location of the equilibrium ξ (Table 18.4) is shown by the heavy line.

curves like Figure 18.8. A more usual thing to do with reactions like this would be to use the equilibrium constant,

$$\frac{a_{\mathrm{NH}_3}^2}{a_{\mathrm{N}_2} a_{\mathrm{H}_2}^3} = K_{18.54} \tag{18.66}$$

values for which are also obtainable from SUPCRT92. We will assume ideal gas conditions (where $f_i/f_i^\circ = x_i$), so the activity terms in (18.66) are the mole fractions in Equations (18.60).

Because the mole fractions are all a function of a single variable ξ, Equation (18.66) then contains a single unknown, and can be solved for ξ and hence the mole numbers and mole fractions of the three species which satisfy the equilibrium constant. Not surprisingly, these results agree with the **G**-minimization calculation, and are shown in Table 18.4 as a function of temperature.

Finally, we can have a look at the affinity values. The easiest way to calculate $\mathcal{A}$ is to first multiply Equation (8.26) by ν_i, then sum over all species, resulting in

$$\sum_{i=1}^{s} \nu_i \mu_i = \sum_{i=1}^{s} \nu_i \mu_i^\circ + \sum_{i=1}^{s} \nu_i RT \ln a_i \tag{18.67}$$

Combining this with (18.57) we get

$$-\mathcal{A} = \Delta_r G^\circ + RT \ln Q \tag{18.68}$$

where $Q = \prod_i a_i^{\nu_i}$, and where the a_i are not necessarily the (stable) equilibrium values. Then substituting $-RT \ln K$ for $\Delta_r G^\circ$ from

$$\Delta_r G^\circ = -RT \ln K \qquad [9.11]$$

we get

$$\mathcal{A} = RT \ln(K/Q) \tag{18.69}$$

Table 18.4 *Species mole numbers which satisfy $K_{18.54}$ and Equations (18.60).*

T °C	$\log K_{18.54}$	n_{N_2} mol	n_{H_2} mol	n_{NH_3} mol	ξ
25	5.764	0.03177	0.09530	1.9365	0.9682
50	4.508	0.06535	0.1961	1.8693	0.9347
100	2.467	0.2074	0.6223	1.5851	0.7926
150	0.8740	0.4684	1.4059	1.0627	0.5314
200	−0.4061	0.7425	2.2275	0.5150	0.2575
250	−1.4611	0.8975	2.6923	0.2051	0.1026
300	−2.3460	0.9591	2.8772	0.08190	0.04095

Note the similarity of this expression to the saturation index (§16.2.2), since for a simple solubility product expression, Q becomes the same as IAP.

Values of the mole fractions (activities) of each of the three species are calculated from the mole numbers in Table 18.4, and combined into values of Q at each ξ. The results are shown in Table 18.5 and Figure 18.10.

The affinity tells you how far the reaction is from equilibrium in J mol^{-1}, and is positive or negative for the reaction proceeding right or left, respectively. Combined with a positive or negative $d\xi$, respectively, $\mathcal{A}d\xi$ is always positive, as mentioned earlier.

Table 18.5 *Calculation of the affinity at $T = 200\,°C$ from Equation (18.69).*

ξ	Q	K/Q	$\mathcal{A}$ J mol^{-1}[a]
0.05	0.006916	56.7648	15889
0.10	0.03260	12.0408	9789
0.20	0.1875	2.09387	2907
0.2575	0.3925	1.000	0
0.30	0.6420	0.6116	−1935
0.40	1.8729	0.2096	−6147
0.50	5.3333	0.07361	−10263
0.60	16.333	0.0240	−14667
0.70	60.583	0.00648	−19823
0.80	341.33	0.00115	−26625
0.90	5805.0	0.000068	−37774

[a] Actually, joules per 2 moles of NH_3

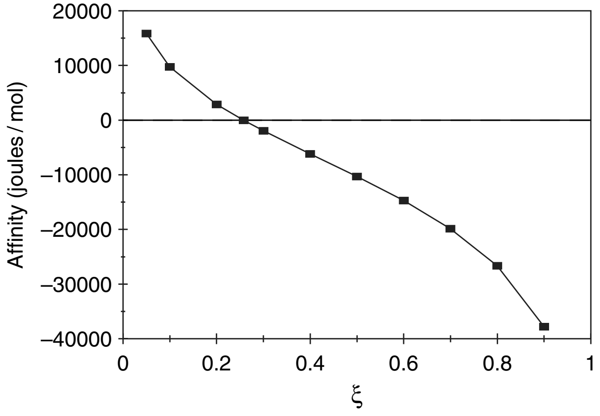


Figure 18.10 Affinity of reaction (18.54) at 200 °C. Data in Table 18.5.

In systems having several simultaneous reactions, affinities and progress variables for each reaction as well as for the whole system may be calculated. Unfortunately, affinities of individual reactions in such systems are not related in any simple way to the order in which those reactions will reach equilibrium, due to the effects of reaction coupling (Helgeson, 1979). Similarly, although the affinity is commonly used in theoretical expressions for reaction rate constants, it is never the only determining factor. Nevertheless, it is always a central concept in thinking about irreversible processes.

18.5 Final comment

We have emphasized in various places a point of view about thermodynamics which is fairly philosophical, which may also seem of not much use to someone interested in complex natural phenomena. That is the idea that in doing our thermodynamic calculations, we are not really calculating the properties of natural systems, but properties of simplified models of these systems. The models are mathematical, and the properties and processes in the model include some that have no counterpart in the real world. Nevertheless, model results are useful in understanding the real systems, if the models are properly or appropriately constructed. The calculations must satisfy stringent mathematical relationships, often giving them a gloss of certainty to the untrained eye, but actually they are of use only if the model they constitute is appropriate, i.e., if it has some similarity to the problems of interest.

One reason that it is a good idea to make the distinction between our models and reality is that our "explanations" of thermodynamics (and other exact sciences) are often in terms of mathematical planes, surfaces, tangents, and other even more abstract concepts. Students look at these explanations with understanding of a sort, but also with an underlying but usually unexpressed bafflement as to what these shining, unblemished, perfect, mathematical constructs have to do with anything real. It is best to come to terms with this problem by admitting that thermodynamics is in fact all about these mathematical constructs. It is not difficult to understand thermodynamics as the mathematics of certain planes, surfaces, tangents, etc. The more difficult problem is to understand why these mathematical constructs have such direct relevance to our universe, but that is best left to the philosophers.

How do we assure ourselves that our models are appropriate? There is of course no way to be sure. The construction of models useful in understanding nature is the essence of science and relies as much upon creativity and imagination as any painting or musical composition. Unfortunately, the models use the language of mathematics, rather than shapes and colors or musical notes and are hence not understandable to anyone who does not know the language. This is unfortunate, because mathematics and mathematical models have their own kind of beauty, which easily rivals that of the arts. Although we don't know why the universe should be such that mathematics is so useful in describing it, there is no doubt that it is useful. It seems clear that the more mathematical tools you have in your repertoire, the more adept you will be at fitting mathematics to your observations.

As mentioned in §18.3.3, the models being developed to simulate complex natural phenomena are becoming very complex. Not only is it difficult to create such models, and to compile enough basic data to enable them to work, but it becomes increasingly difficult to know how well they work. There is a great deal of uncertainty about what nature is actually doing, so that it's often hard to know just how well model results coincide with natural observations. This means that in understanding natural systems, insightful field observations are just as important as model construction. And model results may turn out to be reasonably accurate, based on entirely incorrect ideas. Just because a model "works" does not mean it is right. All this means that there is plenty of scope for thermodynamics and the tools built upon it in the years to come, in our quest to understand our world and how it works.

Appendix A

Constants and Numerical values

The SI (Système International) units

Physical quantity	SI unit	Symbol for SI unit	Unit in terms of base units	Unit in terms of other SI units
Base (fundamental) Units				
length	meter	m	—	—
mass	kilogram	kg	—	—
time	second	s	—	—
electric current	ampere	A	—	—
temperature	kelvin	K	—	—
amount of substance	mole	mol	—	—
Derived SI Units				
velocity (speed)			m/s	
acceleration			$\mathrm{m/s^2}$	N/kg
force	newton	N	$\mathrm{kg\,m/s^2}$	J/m
pressure	pascal	Pa	$\mathrm{kg/(m\,s^2)}$	$\mathrm{N/m^2}$
energy	joule	J	$\mathrm{kg\,m^2/s^2}$	N m
entropy	joule per kelvin	S	$\mathrm{kg\,m^2/(s^2\,K)}$	J/K
power	watt	W	$\mathrm{kg\,m^2/s^3}$	J/s
momentum			kg m/s	
frequency	hertz	Hz	$\mathrm{s^{-1}}$	
electric charge	coulomb	C	A s	V F
voltage (emf)	volt	V	$\mathrm{kg\,m^2/(A\,s^3)}$	W/A; C/F
electric resistance	ohm	Ω	$\mathrm{kg\,m^2/(A^2\,s^3)}$	V/A
capacitance	farad	F	$\mathrm{A^2\,s^4/(kg\,m^2)}$	C/V

Fundamental physical constants[a]

Quantity	Symbol in this text	Value	Units
speed of light in vacuum	c	299 792 458	$\mathrm{m\,s^{-1}}$
constant of gravitation	g	6.672 59	$10^{-11}\,\mathrm{m^3\,kg^{-1}\,s^{-2}}$
elementary charge	e	1.602 177 33	$10^{-19}\,\mathrm{C}$
Planck constant	h	6.626 0755	$10^{-34}\,\mathrm{J\,s}$
Avogadro constant	N_A	6.022 136	$10^{23}\,\mathrm{mol^{-1}}$
Faraday constant	$\mathcal{F}$	96 485.309	$\mathrm{C\,mol^{-1}}$
molar gas constant	R	8.314 510	$\mathrm{J\,mol^{-1}K^{-1}}$
Boltzmann constant, R/N_A	k	1.380 658	$10^{-23}\mathrm{J\,K^{-1}}$
molar volume[b]	V	0.022 414 10	$\mathrm{m^3\,mol^{-1}}$

[a] Cohen and Taylor (1988).

[b] The volume per mole of ideal gas at 101325 Pa and 273.15 K

Miscellaneous useful conversions and older units

ln 10	2.302 585
$\ln x$	$\ln 10 \times \log_{10} x$
1 cal	4.184 J[a]
R	$1.987\,216\ \mathrm{cal\,K^{-1}\,mol^{-1}}$
$\mathcal{F}$	$96\,485.309\ \mathrm{J\ V^{-1}\ mol^{-1}}$
	$23\,060.542\ \mathrm{cal\ V^{-1}mol^{-1}}$
$RT/\mathcal{F}$	0.025 692 73 V ($T = 298.15\,\mathrm{K}$)
$2.302585\,RT/\mathcal{F}$	0.059 1597 V ($T = 298.15\,\mathrm{K}$)
1 bar	10^5 pascal
	14.504 psi
	$0.10\ \mathrm{J\,cm^{-3}}$
	$0.023\,9006\ \mathrm{cal\,cm^{-3}}$
1 atm	1.013 25 bar
	101 325 pascal
	14.696 psi
1 $\mathrm{cm^3}$	$0.10\ \mathrm{J\,bar^{-1}}$
	$0.0239\,006\ \mathrm{cal\,bar^{-1}}$
1 Å	1 angstrom = 10^{-8} cm

[a] This is the thermochemical calorie, used in most of physical chemistry. The International Table calorie used in Steam Tables is 4.1868 J (see page 388).

Appendix B

Standard state thermodynamic properties of selected minerals and other compounds

Part 1. Inorganic substances

Data from Wagman et al. (1982); with a few additions from other sources – Al species from Drever (1988); silica species and all volume data from SUPCRT92 (Johnson et al., 1992).

Formulas	Form	Mol. wt.	$\Delta_f H^\circ$	$\Delta_f G^\circ$	S°	C_P°	V°
		g mol^{-1}	kJ mol^{-1}		J mol^{-1} K^{-1}		cm^3 mol^{-1}
Aluminum							
Al	s	26.9815	0	0	28.33	24.35	
Al^{3+}	aq	26.9815	−531.0	−485.0	−321.7	—	−45.3
$Al(OH)^{2+}$	aq		−767.0	−693.7	—	—	
$Al(OH)_2^+$	aq		−1010.7	−901.4	—	—	
$Al(OH)_3^\circ(aq)$	aq		−1250.4	−1100.7	—	—	
$Al(OH)_4^-$	aq	95.0111	−1490.0	−1307.0	102.9	—	45.60
Al_2O_3	α, corundum	101.9612	−1675.7	−1582.3	50.92	79.04	25.575
$Al_2O_3 \cdot H_2O$	boehmite	119.9766	−1980.7	−1831.7	96.86	131.25	39.07
$Al_2O_3 \cdot H_2O$	diaspore	119.9766	−1998.91	−1841.78	70.67	106.19	35.52
$Al_2O_3 \cdot 3H_2O$	gibbsite	156.0074	−2586.67	−2310.21	136.90	183.47	63.912
$Al_2O_3 \cdot 3H_2O$	bayerite	156.0074	−2576.5	—	—	—	
$Al(OH)_3$	amorphous	78.0037	−1276.0	—	—	—	
Al_2SiO_5	andalusite	162.0460	−2590.27	−2442.66	93.22	122.72	51.53
Al_2SiO_5	kyanite	162.0460	−2594.29	−2443.88	83.81	121.71	44.09
Al_2SiO_5	sillimanite	162.0460	−2587.76	−2440.99	96.11	124.52	49.90
$Al_2Si_2O_7 \cdot 2H_2O$	kaolinite	258.1616	−4119.6	−3799.7	205.0	246.14	99.52
$Al_2Si_2O_7 \cdot 2H_2O$	halloysite	258.1616	−4101.2	−3780.5	203.3	246.27	99.30
$Al_2Si_2O_7 \cdot 2H_2O$	dickite	258.1616	−4118.3	−3795.9	197.1	239.49	99.30
$Al_6Si_2O_{13}$	mullite	426.0532	−6816.2	−6432.7	255.0	326.10	—
$Al_2Si_4O_{10}(OH)_2$	pyrophyllite	360.3158	−5642.04	−5268.14	239.41	294.34	126.6

Table (cont.)

Formulas	Form	Mol. wt. g mol^{-1}	$\Delta_f H°$ kJ mol^{-1}	$\Delta_f G°$ kJ mol^{-1}	$S°$ J mol^{-1} K^{-1}	$C_P°$ J mol^{-1} K^{-1}	$V°$ cm^3 mol^{-1}
Barium							
Ba	s	137.3400	0	0	62.8	28.07	
Ba^{2+}	aq	137.3400	−537.64	−560.77	9.6	—	−12.9
BaO	s	153.3394	−553.5	−525.1	70.42	47.78	
BaO_2	s	169.3388	−634.3	—	—	66.9	
BaF_2	s	175.3368	−1207.1	−1156.8	96.36	71.21	
BaS	s	169.4040	−460.0	−456.0	78.2	49.37	
$BaSO_4$	barite	233.4016	−1473.2	−1362.2	132.2	101.75	52.10
$BaCO_3$	witherite	197.3494	−1216.3	−1137.6	112.1	85.35	45.81
$BaSiO_3$	s	213.4242	−1623.60	−1540.21	109.6	90.00	
Calcium							
Ca	s	40.0800	0	0	41.42	25.31	
Ca^{2+}	aq	40.0800	−542.83	−553.58	−53.1	—	−18.4
CaO	s	56.0794	−635.09	−604.03	39.75	42.80	
$Ca(OH)_2$	portlandite	74.0948	−986.09	−898.49	83.39	87.49	
CaF_2	fluorite	78.0768	−1219.6	−1167.3	68.87	67.03	24.542
CaS	s	72.1440	−482.4	−477.4	56.5	47.40	
$CaSO_4$	anhydrite	136.1416	−1434.11	−1321.79	106.7	99.66	45.94
$CaSO_4 \cdot 2H_2O$	gypsum	172.1724	−2022.63	−1797.28	194.1	186.02	
$Ca_3(PO_4)_2$	β,whitlockite	310.1828	−4120.8	−3884.7	236.0	227.82	
$Ca_3(PO_4)_2$	α	310.1828	−4109.9	−3875.5	240.91	231.58	
$CaCO_3$	calcite	100.0894	−1206.92	−1128.79	92.9	81.88	36.934
$CaCO_3$	aragonite	100.0894	−1207.13	−1127.75	88.7	81.25	34.150
$CaSiO_3$	wollastonite	116.1642	−1634.94	−1549.66	81.92	85.27	39.93
$CaSiO_3$	pseudowollastonite	116.1642	−1628.4	−1544.7	87.36	86.48	
$CaAl_2SiO_6$	Ca-Al pyroxene	218.1254	−3298.2	−3122.0	141.4	165.7	
$CaAl_2Si_2O_8$	anorthite	278.2102	−4227.9	−4002.3	199.28	211.42	100.79
$CaTiO_3$	perovskite	135.9782	−1660.6	−1575.2	93.64	97.65	
$CaTiSiO_5$	sphene	196.0630	−2603.3	−2461.8	129.20	138.95	
$CaMg(CO_3)_2$	dolomite	184.4108	−2326.3	−2163.4	155.18	157.53	64.365
$CaMgSi_2O_6$	diopside	216.5604	−3206.2	−3032.0	142.93	166.52	66.090
Carbon							
C	graphite	12.0112	0	0	5.740	8.527	5.298
C	diamond	12.0112	1.895	2.900	2.377	6.113	3.417
CO_3^{2-}	aq	60.0094	−677.149	−527.81	−56.9	—	−6.1
HCO_3^-	aq	61.0174	−691.99	−586.77	91.2	—	24.2
CO	g	28.0106	−110.525	−137.168	197.674	29.142	24465.6

Table (cont.)

		Mol. wt.	$\Delta_f H^\circ$	$\Delta_f G^\circ$	S°	C_P°	V°
Formulas	Form	g mol^{-1}	kJ mol^{-1}		J mol^{-1} K^{-1}		cm^3 mol^{-1}
CO_2	g	44.0100	−393.509	−394.359	213.74	37.11	24465.6
CO_2	aq	44.0100	−413.80	−385.98	117.6	—	32.8
H_2CO_3	aq	62.0254	−699.65	−623.08	187.4	—	
CH_4	g	16.0432	−74.81	−50.72	186.264	35.309	24465.6
C_2H_6	g	30.0704	−84.68	−32.82	229.60	52.63	24465.6
CN	g	26.0179	437.6	407.5	202.6	29.16	
CN^-	aq	26.0179	150.6	172.4	94.1	—	
HCN	g	27.0259	135.1	124.7	201.78	35.86	
HCN	aq	27.0259	107.1	119.7	124.7	—	
Chlorine							
Cl_2	g	70.9060	0	0	233.066	33.907	24465.6
Cl^-	aq	35.4530	−167.159	−131.228	56.5	−136.4	17.3
HCl	aq	36.4610	−167.159	−131.228	56.5	−136.4	17.3
HCl	g	36.4610	−92.307	−95.299	186.908	29.12	24465.6
Copper							
Cu	s	63.5400	0	0	33.15	24.435	
Cu^+	aq	63.5400	71.67	49.98	40.6	—	
Cu^{2+}	aq	63.5400	64.77	65.49	−99.6	—	
CuO	tenorite	79.5394	−157.3	−129.7	42.63	42.30	
Cu_2O	cuprite	143.0794	−168.6	−146.0	93.14	63.64	
$CuSO_4 \cdot 3H_2O$	bonattite	213.6478	−1684.31	−1399.36	221.3	205.0	
$CuSO_4 \cdot 5H_2O$	calcanthite	249.6786	−2279.65	−1879.745	300.4	280.0	
CuS	covellite	96.6040	−53.1	−53.6	66.5	47.82	
Cu_2S	chalcocite	159.1440	−79.5	−86.2	120.9	76.32	
Fluorine							
F_2	g	37.9968	0	0	202.78	31.30	
HF	g	20.0064	−271.1	−273.2	173.779	29.133	
HF	aq	20.0064	−320.08	−296.82	88.7	—	
F^-	aq	18.9984	−332.63	−278.79	−13.8	−106.7	
Hydrogen							
H_2	g	2.0160	0	0	130.684	28.824	24465.6
H^+	aq	1.0080	0	0	0	0	0
OH^-	aq	17.0074	−229.994	−157.244	−10.75	−148.5	
H_2O	l	18.0154	−285.830	−237.129	69.91	75.291	18.068
H_2O	g	18.0154	−241.818	−228.572	188.825	33.577	24465.6

Table (cont.)

Formulas	Form	Mol. wt. g mol^{-1}	$\Delta_f H^\circ$ kJ mol^{-1}	$\Delta_f G^\circ$ kJ mol^{-1}	S° J mol^{-1} K^{-1}	C_P° J mol^{-1} K^{-1}	V° cm^3 mol^{-1}
Iodine							
I_2	s	253.8088	0	0	116.135	54.438	
I^-	aq	126.9044	−55.19	−51.57	111.3	−142.3	
HI	aq	127.9124	−55.19	−51.57	111.3	—	
IO_3^-	aq	174.9026	−221.3	−128.0	118.4	—	
IO_4^-	aq	190.9020	−155.5	−58.5	222.0	—	
Iron							
Fe	s	55.8470	0	0	27.28	25.10	
Fe^{2+}	aq	55.8470	−89.1	−78.90	−137.7	—	
Fe^{3+}	aq	55.8470	−48.5	-4.7	−315.9	—	
$Fe_{0.947}O$	wüstite	68.8865	−266.27	−245.12	57.49	48.12	
Fe_2O_3	hematite	159.6922	−824.2	−742.2	87.40	103.85	
Fe_3O_4	magnetite	231.5386	−1118.4	−1015.4	146.4	143.43	
FeO(OH)	goethite	88.8538	−559.0	(−487.02)	(60.25)	—	
$Fe(OH)_2$	s	89.8618	−569.0	−486.5	88.0	—	
$Fe(OH)_3$	s	106.8692	−823.0	−696.5	106.7	—	
FeS	troilite	87.9110	−100.0	−100.4	60.29	50.54	
FeS_2	pyrite	119.9750	−178.2	−166.9	52.93	62.17	
$FeCO_3$	siderite	115.8564	−740.57	−666.67	92.9	82.13	
Fe_2SiO_4	fayalite	203.7776	−1479.9	−1379.0	145.2	132.88	
Lead							
Pb	s	207.1900	0	0	64.81	26.44	
Pb^{2+}	aq	207.1900	−1.7	−24.43	10.5	—	
PbO	yellow	223.1894	−217.32	−187.89	68.70	45.77	
PbO	red	223.1894	−218.99	−188.93	66.5	45.81	
PbF_2	s	245.1868	−664.0	−617.1	110.5	—	
$PbCl_2$	s	278.0960	−359.41	−314.10	136.0	—	
PbS	galena	239.2540	−100.42	−98.7	91.2	49.50	
$PbSO_4$	anglesite	303.2516	−919.94	−813.14	148.57	103.207	
$PbCO_3$	cerussite	267.1994	−699.1	−625.5	131.0	87.40	
$PbSiO_3$	s	283.2742	−1145.70	−1062.10	109.6	90.04	
Magnesium							
Mg	s	24.3120	0	0	32.68	24.89	
Mg^{2+}	aq	24.3120	−466.85	−454.8	−138.1	—	
MgO	periclase	40.3114	−601.70	−569.43	26.94	37.15	
$Mg(OH)_2$	brucite	58.3268	−924.54	−833.51	63.18	77.03	
MgF_2	sellaite	62.3088	−1123.4	−1070.2	57.24	61.59	

Table (cont.)

		Mol. wt.	$\Delta_f H^\circ$	$\Delta_f G^\circ$	S°	C_P°	V°
Formulas	Form	g mol^{-1}	kJ mol^{-1}		J mol^{-1} K^{-1}		cm^3 mol^{-1}
MgS	s	56.3760	−346.0	−341.8	50.33	45.56	
$MgCO_3$	magnesite	84.3214	−1095.8	−1012.1	65.7	75.52	28.018
$MgCO_3 \cdot 3H_2O$	nesquehonite	138.3676	—	−1726.1	—	—	
$MgSiO_3$	enstatite	100.3962	−1549.00	−1462.09	67.74	81.38	
Mg_2SiO_4	forsterite	140.7076	−2174.0	−2055.1	95.14	118.49	
Manganese							
Mn	s	54.9380	0	0	32.01	26.32	
Mn^{2+}	aq	54.9380	−220.75	−228.1	−73.6	50.0	
MnO_4^-	aq	118.9356	−541.4	−447.2	191.2	−82.0	
MnO_4^{2-}	aq	118.9356	−653.0	−500.7	59.0	—	
MnO	manganosite	70.9374	−385.22	−362.90	59.71	45.44	
Mn_3O_4	hausmannite	228.8116	−1387.8	−1283.2	155.6	139.66	
Mn_2O_3	s	157.8742	−959.0	−881.1	110.5	107.65	
MnO_2	pyrolusite	86.9368	−520.03	−465.14	53.05	54.14	
$Mn(OH)_2$	amorphous	88.9528	−695.4	−615.0	99.2	—	
MnS	alabandite	87.0020	−214.2	−218.4	78.2	49.96	
$MnCO_3$	rhodochrosite	114.9474	−894.1	−816.7	85.8	81.50	
$MnSiO_3$	rhodonite	131.0222	−1320.9	−1240.5	89.1	86.44	
Mn_2SiO_4	tephroite	201.9596	−1730.5	−1632.1	163.2	129.87	
Mercury							
Hg	l	200.5900	0	0	76.02	27.983	
Hg	g	200.5900	61.317	31.820	174.96	20.786	
Hg^{2+}	aq	200.5900	171.1	164.4	−32.2	—	
Hg_2^{2+}	aq	401.1800	172.4	153.52	84.5	—	
HgS_2^{2-}	aq	264.7180	—	41.9	—	—	
$HgCl_4^{2-}$	aq	342.4020	−554.0	−446.8	293.0	—	
Hg_2Cl_2	s	472.0860	−265.22	−210.745	192.5	—	
HgO	s, red	216.5894	−90.83	−58.539	70.29	44.06	
HgO	s, yellow	216.5894	−90.46	−58.409	71.1	—	
HgS	cinnabar	232.6540	−58.2	−50.6	82.4	48.41	
HgS	metacinnabar	232.6540	−53.6	−47.7	88.3	—	
Molybdenum							
Mo	s	95.9400	0	0	28.66	24.06	
MoO_3	s	127.9388	−745.09	−667.97	77.74	74.98	
MoS_2	molybdenite	160.0680	−235.1	−225.9	62.59	63.55	

Table (cont.)

Formulas	Form	Mol. wt.	$\Delta_f H^\circ$	$\Delta_f G^\circ$	S°	C_P°	V°
		g mol^{-1}	kJ mol^{-1}		J mol^{-1} K^{-1}		cm^3 mol^{-1}
Nickel							
Ni	s	58.7100	0	0	29.87	26.07	
Ni^{2+}	aq	58.7100	−54.0	−45.6	−128.9	—	
NiO	bunsenite	74.7094	−239.7	−211.7	37.99	44.31	
NiS	s	90.7740	−82.0	−79.5	52.97	47.11	
Nitrogen							
N_2	g	28.0134	0	0	191.61	29.125	
NO	g	30.0061	90.25	86.55	210.761	29.844	
NO_2	g	46.0055	33.18	51.31	240.06	37.20	
N_2O	g	44.0128	82.05	104.2	219.85	38.45	
N_2O_4	l	92.0110	−19.50	97.54	209.2	142.7	
N_2O_4	g	92.0110	9.16	97.89	304.29	77.28	
N_2O_5	s	108.0104	−43.1	113.9	178.2	143.1	
N_2O_5	g	108.0104	11.3	115.1	355.7	84.5	
NH_3	g	17.0307	−46.11	−16.45	192.45	35.06	
NO_3^-	aq	62.0049	−205.0	−108.74	146.45	−86.6	
NH_4^+	aq	18.0837	−132.51	−79.31	113.4	79.9	
NH_4OH	aq	35.0461	−366.12	−263.63	181.21	—	
Oxygen							
O_2	g	31.9988	0	0	205.138	29.355	
O_2	aq	31.9988	−11.7	16.4	110.9	–	
OH^-	aq	17.0074	−229.994	−157.244	−10.75	−148.5	
H_2O	l	18.0154	−285.830	−237.129	69.91	75.291	18.068
H_2O	g	18.0154	−241.818	−228.572	188.825	33.577	24465.6
Potassium							
K	s	39.1020	0	0	64.18	29.58	
K^+	aq	39.1020	−252.38	−283.27	102.5	21.8	9.0
KCl	sylvite	74.5550	−436.747	−409.14	82.59	51.30	
$KAlSi_3O_8$	sanidine	278.3367	−3959.7	−3739.9	232.88	204.51	
$KAlSi_3O_8$	microcline	278.3367	−3968.1	−3742.9	214.22	202.38	108.741
$KAlSiO_4$	kaliophilite	158.1671	−2121.3	−2005.3	133.1	119.79	
$KAlSi_2O_6$	leucite	218.2519	−3034.2	−2871.4	200.08	164.14	
$KAl_3Si_3O_{10}OH_2$	muscovite	398.3133	−5984.4	−5608.4	306.3	–	14.087
Silicon							
Si	s	28.0860	0	0	18.83	20.00	
SiO_2	α-quartz	60.0848	−910.94	−856.64	41.84	44.43	22.688

Table (cont.)

SiO_2	α-cristobalite	60.0848	-909.48	-855.43	42.68	44.18	
SiO_2	α-tridymite	60.0848	-909.06	-855.26	43.5	44.60	25.740
SiO_2	coesite	60.0848	-906.31	-851.62	40.376	43.51	20.641
SiO_2	amorphous	60.0848	-903.49	-850.70	46.9	44.4	
SiO_2	aq	60.0848	-877.699	-833.411	75.312	318.40	16.1
H_4SiO_4	aq		-1449.359	-1307.669	215.132	468.98	
$HSiO_3^-$	aq		-1125.583	-1013.783	41.84	-137.24	9.5
Silver							
Ag	s	107.8700	0	0	42.55	25.351	
Ag^+	aq	107.8700	105.579	77.107	72.68	21.8	
Ag_2O	s	231.7394	−31.05	−11.20	121.3	65.86	
AgCl	cerargyrite	143.3230	−127.068	−109.789	96.2	50.79	
Ag_2S	acanthite	247.8040	−32.59	−40.67	144.01	76.53	
Ag_2S	argentite	247.8040	−29.41	−39.46	150.6	–	
Sodium							
Na	s	22.9898	0	0	51.21	28.24	
Na^+	aq	22.9898	−240.12	−261.905	59.0	46.4	−1.2
NaCl	halite	58.4428	−411.153	−384.138	72.13	50.50	27.015
Na_2SiO_3	s	122.0638	−1554.90	−1462.80	113.85	—	
$NaAlSiO_4$	nepheline	142.0549	−2092.8	−1978.1	124.3	—	54.16
$NaAlSi_3O_8$	low albite	262.2245	−3935.1	−3711.5	207.40	205.10	100.07
$NaAlSi_2O_6$	jadeite	202.1397	−3030.9	−2852.1	133.5	—	60.40
Sulfur							
S	orthorhombic	32.0640	0	0	31.80	22.64	
S^{2-}	aq	32.0640	33.1	85.8	−14.6	—	
HS^-	aq	33.0720	−17.6	12.08	62.8	—	
SO_4^{2-}	aq	96.0616	−909.27	−744.53	20.1	−293.0	
HSO_4^-	aq	97.0696	−887.34	−755.91	131.8	−84.0	
S_2	g	64.1280	128.37	79.30	228.18	32.47	
H_2S	g	34.0800	−20.63	−33.56	205.79	34.23	
H_2S	aq	34.0800	−39.7	−27.83	121.0	—	
SO_2	g	64.0628	−296.830	−300.194	248.22	39.87	
SO_3	g	80.0622	−395.72	−371.06	256.76	50.67	
Titanium							
Ti	s	47.9000	0	0	30.63	25.02	
TiO	s	63.8994	−519.7	−495.0	50.0	39.96	
TiO_2	anatase	79.8988	−939.7	−884.5	49.92	55.48	
TiO_2	brookite	79.8988	−941.8	—	—	—	
TiO_2	rutile	79.8988	−944.7	−889.5	50.33	55.02	

Table (cont.)

Uranium						
U	s	238.0290	0	0	50.21	27.665
UO_2	uraninite	270.0278	−1084.9	−1031.7	77.03	63.60
UO_3	orthorhombic	286.0272	−1223.8	−1145.9	96.11	81.67
U^{3+}	aq	238.0290	−489.1	−475.4	192.0	—
U^{4+}	aq	238.0290	−591.2	−531.0	410.0	—
UO_2^{2+}	aq	270.0278	−1019.6	−953.5	−97.5	—
Zinc						
Zn	s	65.3700	0	0	41.63	25.40
Zn^{2+}	aq	65.3700	−155.89	−147.06	−112.1	46.0
ZnO	zincite	81.3694	−348.28	−318.30	43.64	40.25
ZnS	wurtzite	97.4340	−192.63	—	—	—
ZnS	sphalerite	97.4340	−205.98	−201.29	57.7	46.0
$ZnCO_3$	smithsonite	125.3794	−812.78	−731.52	82.4	79.71
Zn_2SiO_4	willemite	222.8236	−1636.74	−1523.16	131.4	123.34

Part 2. Organic substances

N.B.: columns for $\Delta_f G^\circ$ and $\Delta_f H^\circ$ are reversed from Part 1, and $\Delta_f G^\circ$ and $\Delta_f H^\circ$ are in J rather than kJ. Note that a large database of data on organic compounds is being assembled by Everett Shock, and is freely available at http://webdocs.asu.edu. Data from Shock and Helgeson (1990).

			$\Delta_f G^\circ$	$\Delta_f H^\circ$	S°	C_P°	V°
Formulas	Form	Name	J mol^{-1}		J mol^{-1} K^{-1}		cm^3 mol^{-1}
n-Alkanes							
CH_4	aq	methane	−34 451	−87 906	87.82	277.4	37.30
CH_4	g	methane	−50 720	−74 810	186.26	35.31	24 465.6
C_2H_6	aq	ethane	−16 259	−103 136	112.17	369.4	51.20
C_3H_8	aq	propane	−8213	−127 570	141.00	462.8	67.00
C_4H_{10}	aq	*n*-butane	151	−151 586	167.44	560.2	82.80
C_5H_{12}	aq	*n*-pentane	8912	−173 887	198.74	640.2	98.60
C_6H_{14}	aq	*n*-hexane	18 493	−198 322	221.33	733.0	114.40
C_7H_{16}	aq	*n*-heptane	27 070	−221 543	251.04	821.7	130.20
C_8H_{18}	aq	*n*-octane	35 899	−248 571	266.94	910.4	146.00
1-Alkenes							
C_2H_4	aq	ethylene	81 379	35 857	120.08	261.5	45.50
C_3H_6	aq	1-propene	74 935	−1213	153.55	350.2	61.30
C_4H_8	aq	1-butene	84 977	−23 577	181.59	438.9	77.10
C_5H_{10}	aq	1-pentene	94 014	−46 861	209.62	527.6	92.90
C_6H_{12}	aq	1-hexene	101 964	−71 233	237.65	616.3	108.70
C_7H_{14}	aq	1-heptene	110 667	−94 851	265.68	705.0	124.50
C_8H_{16}	aq	1-octene	120 164	−117 654	293.72	793.7	140.30
Alkylbenzenes							
C_6H_6	aq	benzene	133 888	51 170	148.53	361.1	83.50
$C_6H_5CH_3$	aq	toluene	126 608	13 724	183.68	430.1	97.71
$C_6H_5C_2H_5$	aq	ethylbenzene	135 729	−10 460	208.36	504.2	113.80
Alcohols							
CH_3OH	aq	methanol	−175 937	−246 312	134.72	158.2	38.17
C_2H_5OH	aq	ethanol	−181 293	−287 232	150.21	260.2	55.08
C_6H_5OH	aq	phenol	−52 656	−153 302	191.63	315.1	86.17
Ketones							
C_3H_6O	aq	acetone	−161 084	−258 236	185.77	241.4	66.92
Carboxylic acids							
HCOOH	aq	formic acid	−372 301	−425 429	162.76	79.5	34.69
CH_3COOH	aq	acetic acid	−396 476	−485 762	178.66	169.7	52.01

Table (cont.)

		Name	$\Delta_f G^\circ$	$\Delta_f H^\circ$	S°	C_P°	V°
Formulas	Form	g mol^{-1}	J mol^{-1}		J mol^{-1} K^{-1}		cm^3 mol^{-1}
C_2H_5COOH	aq	propanoic acid	−390 911	−512 414	206.69	234.3	67.90
C_3H_7COOH	aq	butanoic acid	−381 539	−535 343	234.72	336.8	84.61
C_4H_9COOH	aq	pentanoic acid	−373 288	−559 359	262.76	432.2	100.50
$C_5H_{11}COOH$	aq	hexanoic acid	−364 343	−582 789	292.46	523.8	116.55
$C_6H_{13}COOH$	aq	heptanoic acid	−356 268	−607 015	318.82	612.5	132.30
$C_7H_{15}COOH$	aq	octanoic acid	−348 946	−631 993	346.85	701.2	148.10
Carboxylate anions							
$HCOO^-$	aq	formate	−350 879	−425 429	90.79	−92.0	26.16
CH_3COO^-	aq	acetate	−369 322	−486 097	86.19	25.9	40.50
$C_2H_5COO^-$	aq	propanoate	−363 046	−513 084	110.88	129.3	54.95
$C_3H_7COO^-$	aq	butanoate	−354 008	−535 259	133.05	186.2	70.30
$C_4H_9COO^-$	aq	pentanoate	−345 598	−562 371	160.25	329.7	86.31
$C_5H_{11}COO^-$	aq	hexanoate	−336 603	−585 300	189.54	418.4	102.21
$C_6H_{13}COO^-$	aq	heptanoate	−327 984	−609 023	217.57	469.4	118.60
$C_7H_{15}COO^-$	aq	octanoate	−319 407	−632 746	242.67	558.1	134.40
Amino acids							
$C_2H_5NO_2$	aq	glycine	−370 778	−513 988	158.32	39.3	43.25
$C_3H_7NO_2$	aq	alanine	−371 539	−552 832	167.36	141.4	60.45
$C_5H_{11}NO_2$	aq	valine	−356 895	−616 303	178.24	302.1	90.79
$C_6H_{13}NO_2$	aq	leucine	−343 088	−632 077	215.48	397.9	107.57
$C_6H_{13}NO_2$	aq	isoleucine	−343 925	−631 366	220.92	383.3	105.45
$C_3H_7NO_3$	aq	serine	−510 866	−714 627	194.56	117.6	60.62
$C_4H_9NO_3$	aq	threonine	−502 080	−749 354	222.59	210.0	76.86
$C_4H_7NO_4$	aq	aspartic acid	−721 322	−947 132	229.28	127.2	71.79
$C_5H_9NO_4$	aq	glutamic acid	−723 832	−970 688	294.97	177.0	89.36
$C_4H_8N_2O_3$	aq	asparagine	−538 272	−780 985	230.96	125.1	77.18
$C_5H_{10}N_2O_3$	aq	glutamine	−529 694	−804 709	258.99	187.0	94.36
$C_9H_{11}NO_2$	aq	phenylalanine	−207 108	−460 575	221.33	384.1	121.92
$C_{11}H_{11}N_2O_2$	aq	tryptophan	−112 550	−409 195	153.13	420.1	144.00
$C_9H_{11}NO_3$	aq	tyrosine	−365 263	−658 562	190.37	299.2	123.00
$C_5H_{11}NO_2S$	aq	methionine	−502 917	−743 078	274.89	292.9	105.30
Peptides							
$C_4H_8N_2O_3$	aq	diglycine	−489 612	−734 878	226.77	158.99	319.11
$C_5H_{10}N_2O_3$	aq	alanylglycine	−488 398	−778 684	212.13	252.30	398.40
$C_8H_{16}N_2O_3$	aq	leucylglycine	−462 834	−847 929	303.76	497.06	608.10
$C_4H_6N_2O_2$	aq	diketopiperazine	−240 329	−415 471	223.84	71.13	321.04

Appendix C

Some mathematics

C.1 Essential mathematics

C.1.1 Differentials and derivatives

Figure C.1 illustrates the usual definition of a derivative. For any function $y = y(x)$, the *derivative* of y, is a function $y'(x)$ where

$$y'(x) = \lim_{\Delta x \to 0} \left(\frac{y(x + \Delta x) - y(x)}{\Delta x} \right) \tag{C.1}$$

As shown in Figure C.1, the quantity

$$\frac{y(x + \Delta x) - y(x)}{\Delta x}$$

is the slope of a line that intersects the function $y(x)$ at two points, (x_1, y_1) and (x_0, y_0), and $\Delta x = x_1 - x_0$, $\Delta y = y_1 - y_0$. As Δx gets smaller, x_1 and y_1 approach x_0 and y_0, and in the limit as $\Delta x \to 0$, the line of intersection becomes the tangent to $y(x)$ at (x_0, y_0). The notation $y'(x)$ indicates that the derivative, or the slope of the tangent, is a new function of x, quite distinct from the original function $y(x)$.

If we let $\Delta x = dx$ and define dy, the *differential of* y, as

$$dy = y'(x)\, dx \tag{C.2}$$

then

$$y'(x) = \frac{dy}{dx} \tag{C.3}$$

and the symbol dy/dx is often used to represent the derivative. dx has already been defined as equal to Δx, that is, any increment of x, and dy is a linear function of dx as shown in Figure C.2. Obviously neither dx nor dy is necessarily an infinitesimal quantity. It is an unfortunate fact that because dy/dx is equal to the derivative, the concept of which involves allowing dx to become infinitesimally small, many students of thermodynamics get the idea that differentials are infinitesimal quantities, and this is a stumbling block to the intuitive grasp of the many equations involving differentials. During integration, of course, differentials can and do take on infinitesimal values.

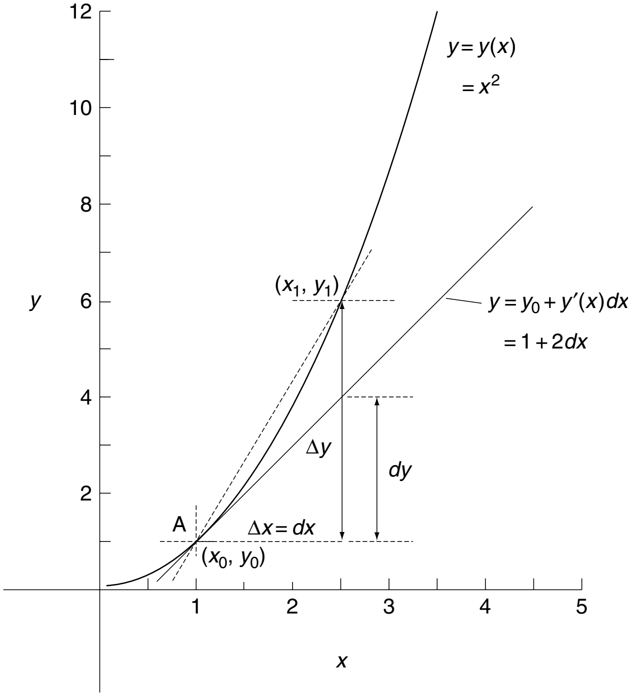


Figure C.1 The meaning of *dy*/*dx*.

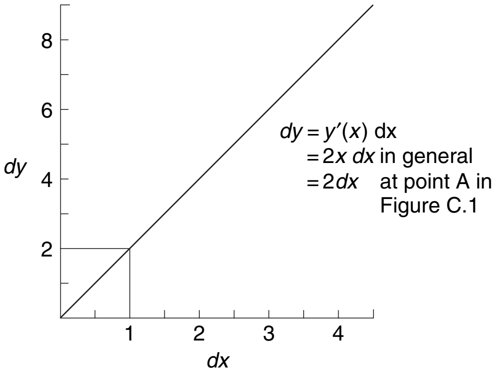


Figure C.2 *dy* is a homogeneous linear function of *dx*.

Of particular interest in thermodynamics is the extremum value of certain functions, that is, the maximum or minimum point. According to the calculus, this is the point where the derivative passes through zero, or $dy/dx = 0$. In Figure C.3, $dy/dx = 2x$ (or $dy = 2x\,dx$), which equals zero at $x = 0$. In differential form, we say the minimum occurs at $dy = 0$. This means that at the minimum, where the tangent is horizontal, y will not change $(dy = 0)$ no matter what the size of dx.

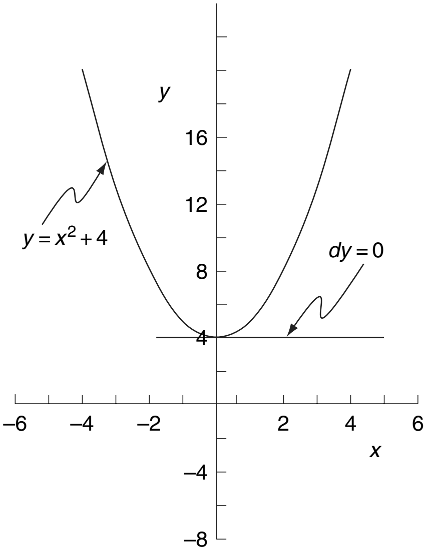


Figure C.3 The differential of y is zero at the minimum in the curve.

C.1.2 Partial derivatives and total differentials

A function having several variables may be differentiated with respect to one of the variables, keeping all the others at fixed values. Thus the function

$$z = z(x, y)$$

can be differentiated with respect to x, keeping y constant, thus evaluating $(\partial z/\partial x)_y$, and it can also be differentiated with respect to y, keeping x constant, evaluating $(\partial z/\partial y)_x$. These quantities are termed *partial derivatives*. The new shape of the "*d*" symbol is to remind us of the partial nature of the differentiation process, and the subscripts remind us which variables are being held constant. In cases where there is no likelihood of confusion, the subscripts are often omitted. For example, if

$$z = 2x^2 + 4y^3$$

then

$$\frac{\partial z}{\partial x} = \left(\frac{\partial z}{\partial x}\right)_y = 4x$$

and

$$\frac{\partial z}{\partial y} = \left(\frac{\partial z}{\partial y}\right)_x = 12y^2$$

The *total differential* of z, dz, is defined as

$$dz = \left(\frac{\partial z}{\partial x}\right)_y dx + \left(\frac{\partial z}{\partial y}\right)_x dy \tag{C.4}$$

For example, if the function $V = V(T, P)$ is

$$V = \frac{RT}{P} \tag{C.5}$$

where R is a constant, then

$$dV = \left(\frac{\partial V}{\partial T}\right)_P dT + \left(\frac{\partial V}{\partial P}\right)_T dP \tag{C.6}$$

$$= \frac{R}{P} dT + \frac{-RT}{P^2} dP \tag{C.7}$$

Equation (C.4) has a very straightforward geometrical meaning, discussed in connection with the fundamental equation in Chapter 4 (§4.6). Thermodynamics commonly deals with continuous changes in multivariable systems. For this reason, total differentials are frequently used, and it is essential to have a clear idea of their meaning.

C.1.3 Integration

Integration is the inverse of differentiation. That is, the problem is to find a function when its rate of change is known. It is performed by summing up (functions of) differentials that are chosen to be very small (infinitesimals). This can be done either in the general case, giving indefinite integrals, or between specified limits, giving definite integrals. For example,

$$\int y'(x)\, dx = y(x) + \text{constant} \tag{C.8}$$

is the general case, since differentiation of $y(x)$ plus any constant will give $y'(x)$, the derivative of a constant being zero.

$$\int_a^b y'(x)\, dx = y(b) - y(a)$$

is the definite integral between the limits a and b and can be thought of as the area under the curve $y'(x)$ in the x–y plane, between the limits $x = a$ and $x = b$. Both methods of integration have been used in applications of thermodynamics in the Earth sciences, but generally speaking the definite integral is better suited to quantitative calculation. Variables are often said to be known "within a constant" or some similar phrase, which refers to a variable obtained from the indefinite integration of some function, as in Equation (C.8).

It is perhaps worthwhile to note here that the concept of infinitesimals, infinitely small increments, etc., remained mathematically unsatisfactory for a long time after Newton and Leibniz invented the calculus, and has been largely abandoned in the teaching of the calculus. These ideas have been superseded by the concept of limits. The continued and widespread use of the term infinitesimal in the science literature seems to be a kind of shorthand way of referring to the process of limit-taking. It seems that scientists are not much bothered by many mathematical niceties which are of great concern to mathematicians.

C.2 Nonessential mathematics

C.2.1 Exact and inexact differentials

Exact differentials are those obtained by differentiating some continuous, single-valued function. Since the function can be obtained again by integration, exact differentials are expressions whose integrals will be the same no matter which path of integration is chosen. Accordingly, the total differentials we have just discussed are also exact differentials. They are the “normal” kind of differential. What then is an inexact differential?

Obviously if one always starts with a function and derives differentials from it, one will never generate an inexact differential. However, in the physical sciences one is apt to come across differential expressions of the type

$$M\,dx + N\,dy + \cdots \tag{C.9}$$

where M and N may or may not be functions of x and y. The concept of inexact differentials arises because not every expression of this type (called a Pfaff differential expression) is equal to the differential of some function $z = z(x, y)$. For example, $(x\,dy + y\,dx)$ is clearly exact, being equal to $d(xy)$, but $(x\,dy - y\,dx)$ is not equal to the differential of any function of x and y, and is inexact.

Expressions that are not exact (as well as those that are) are very important in thermodynamics. (Note that we may speak of either dz or of $M\,dx + N\,dy$ as being an exact differential, or of being exact.)

If M and/or N are not functions of x and y, the expression is inexact. If M and N are functions of x and y, one may determine whether the expression is exact (whether it is equal to the differential of some function of x and y) by determining whether M and N are the partial derivatives of some function with respect to x and y. If they are, then for the expression to be exact it is necessary and sufficient that

$$\left(\frac{\partial M}{\partial y}\right)_x = \left(\frac{\partial N}{\partial x}\right)_y \tag{C.10}$$

for all pairs of terms in the expression.

Equation (C.10) is known as the *reciprocity relation* or cross-differentiation identity. It follows from the fact that the order of differentiation of our original function $z = z(x, y)$ with respect to x and y is immaterial. Mathematically this is written

$$\left(\frac{\partial}{\partial y}\left(\frac{\partial z}{\partial x}\right)_y\right)_x = \left(\frac{\partial}{\partial x}\left(\frac{\partial z}{\partial y}\right)_x\right)_y \tag{C.11}$$

If, in Equation (C.9), $M = (\partial z/\partial x)_y$ and $N = (\partial z/\partial y)_x$ for some function $z = z(x, y)$, then Equation (C.10) must follow from (C.11).

As an example of the use of this relation, let us test Equation (C.7) for exactness. Applying the reciprocity relation gives

$$\left(\frac{\partial}{\partial T}\left(\frac{-RT}{P^2}\right)\right)_P = \left(\frac{\partial}{\partial P}\left(\frac{R}{P}\right)\right)_T$$
$$= \frac{-R}{P^2}$$

Hence Equation (C.6) has been shown to be an exact differential expression, as we knew it must be because we derived it by differentiating the function $V = RT/P$.

Frankly, it is rarely of interest in thermodynamics to test expressions for exactness. The expressions in common use are generally well known to be exact or inexact. However, the reciprocity relationship is useful in deriving relationships between thermodynamic variables.

Getting back to the inexact differentials, suppose we have an expression

$$X\,dx + Y\,dy + Z\,dz \tag{C.12}$$

where X, Y, and Z are functions of x, y, and z but reciprocity is not satisfied,[1] and the expression is not exact. We may assign another variable to represent this expression, e.g.

$$\Phi = X\,dx + Y\,dy + Z\,dz$$

and this variable (Φ) is called an inexact differential. To remind ourselves that it represents a differential expression rather than an ordinary algebraic expression, it is usually combined with some kind of letter "*d*" such as D or δ or $đ$– any form except d, which is reserved for exact differentials. Using δ, we write

$$\delta\Phi = X\,dx + Y\,dy + Z\,dz$$

Inexact differentials such as $\delta\Phi$ are thus simply a shorthand notation for a differential expression such as (C.12) and are not the differential of any function. It follows that they cannot be integrated without specifying a path.

Of special interest is the simplest of inexact differentials, $y\,dx$. When y and x are both independent variables, $y\,dx$ is clearly inexact, since we may write the equivalent expression $y\,dx + 0\,dy$ and applying the reciprocity relationship,

$$\left(\frac{\partial y}{\partial y}\right)_x = 1 \neq \left(\frac{\partial 0}{\partial x}\right)_y$$

Clearly there is no function the differential of which is equal to $y\,dx$. Similarly, not being exact, the integral of $y\,dx$ is not independent of the path chosen. But

---

[1] In the case of three or more terms, the reciprocity relation must hold between any two.

expressions like $y\,dx$ are important in thermodynamics and must be integrated. The most common examples are $P\,dV$ and $T\,dS$.

Expressions that are not exact can only be integrated after making them exact, and this is done by choosing a functional relationship between x and y, i.e., by making

$$y = y(x)$$

Now y and x are no longer independent of one another, and the expression

$$y(x)\,dx$$

is exact and may be integrated unambiguously. Reciprocity is satisfied since we may write $y(x)\,dx$ as

$$y(x)\,dx + 0\,dy$$

and

$$\begin{aligned} \left(\frac{\partial y(x)}{\partial y}\right)_x &= 0 \\ &= \left(\frac{\partial 0}{\partial x}\right)_y \end{aligned}$$

The manner in which y is transformed to $y(x)$ is arbitrary, but is commonly done by applying an "integrating denominator," or an "integrating factor." Thus, if the differential expression

$$\delta\Phi = M\,dx + N\,dy$$

is not exact, it can frequently be made so by dividing through by some particular $\lambda = \lambda(x, y)$. λ is thus an integrating denominator. (Of course, if there is such a function λ for $M\,dx + N\,dy$, finding out what it is may be another matter.) It happens that Pfaff differential expressions in two variables will always have integrating denominators, while with more than two they may or may not. This topic is discussed in detail by Reiss (1965, Chapter 2), who gives the conditions necessary to have integrating denominators.

Probably the most important example we shall encounter on this topic is the relationship between entropy, heat, and temperature:

$$dS = \delta q/T \tag{C.13}$$

As we shall see, S is a function of state, hence dS is exact. Hence, T is an integrating denominator for the Pfaff differential δq. That is, dividing the inexact differential δq by T produces the exact differential dS. δq represents some quantity of heat (which is made to become very, very small upon integration), and in this text (Chapter 3) we just call it q. It is in fact a differential term as shown by Tunell (1932). Analogous remarks hold for $dV = \delta w/P$.

C.2.2 Homogeneous functions

Functions are said to be homogeneous and of degree n if we can multiply every variable in the function by a constant, then factor out the constant. For example, $f(x, y, z)$ is homogeneous in degree n if

$$f\,(kx, ky, kz) = k^n f\,(x, y, z) \tag{C.14}$$

Thus, $x^2 + y^2 + z^2$ is homogeneous second degree because

$$(kx)^2 + (ky)^2 + (kz)^2 = k^2\,(x^2 + y^2 + z^2)$$

Clearly, homogeneity is concerned with the exponents of the variables in a function. Homogeneity is not always quite so obvious, however. For example

$$\frac{x^2y^3}{z} + xyz^2$$

is homogeneous in the fourth degree.

Some functions are homogeneous in only some of the variables and not in others, which is the usual case in thermodynamics. Thus

$$x^2y^3 + x^2y^4$$

is homogeneous second degree in x, but not in y, and not overall, because

$$(kx)^2y^3 + (kx)^2y^4 = k^2(x^2y^3 + x^2y^4)$$

Extensive properties are defined as those that depend on the mass (or number of moles) of the system considered, such as volume or total energy. Mathematically speaking then, extensive properties are homogeneous in the first degree in the masses of the components. For a two-component system for example,

$$\mathbf{V}(T, P, kn_1, kn_2) = k^1\,\mathbf{V}(T, P, n_1, n_2)$$

where n_1 and n_2 are the masses of components 1 and 2, and P, T and $\mathbf{V}$ are the pressure, temperature, and total volume of the system.

Intensive properties are defined as those that do not depend on the mass of the system considered. They are thus homogeneous in the zeroth degree in the masses of the components. For example, for the density

$$\rho(T, P, kn_1, kn_2) = k^0\,\rho(T, P, n_1, n_2)$$

where ρ is the density. Temperature and pressure are also intensive variables, since they do not depend on what mass of system is considered, but the order of homogeneity depends on what other variables are imposed.

The usefulness of the fact that thermodynamic functions are homogeneous in the first or zeroth degree is due mainly to Euler's theorem regarding homogeneous functions.

C.2.3 Euler's theorem for homogeneous functions

This theorem says that for a function such as $f(x, y, z)$ that is homogeneous in degree n,

$$n \cdot f(x, y, z) = x\left(\frac{\partial f}{\partial x}\right) + y\left(\frac{\partial f}{\partial y}\right) + z\left(\frac{\partial f}{\partial z}\right) \tag{C.15}$$

thus relating a function to its partial derivatives in a way somewhat reminiscent of the relation between the total differential and the partial derivatives. Euler's theorem for homogeneous nth order functions is obtained by differentiating (C.14) with respect to k (using the chain rule on the left hand side):

$$\frac{\partial f}{\partial kx} \cdot \frac{\partial kx}{\partial k} + \frac{\partial f}{\partial ky} \cdot \frac{\partial ky}{\partial k} + \frac{\partial f}{\partial kz} \cdot \frac{\partial kz}{\partial k} = n \cdot k^{n-1} \cdot f(x, y, z)$$

or, because $(\partial kx/\partial k) = x$, and similarly for y and z,

$$x\frac{\partial f}{\partial kx} + y\frac{\partial f}{\partial ky} + z\frac{\partial f}{\partial kz} = n \cdot k^{n-1} \cdot f(x, y, z) \tag{C.16}$$

Because this result must hold for arbitrary values of k, we may set $k = 1$ in (C.16) and obtain

$$x\left(\frac{\partial f}{\partial x}\right) + y\left(\frac{\partial f}{\partial y}\right) + z\left(\frac{\partial f}{\partial z}\right) = n \cdot f(x, y, z) \tag{C.17}$$

This simple relationship is quite useful. For example, taking the volume of a two-component system

$$\mathbf{V} = \mathbf{V}(T, P, n_1, n_2)$$

at constant T, P and with $\mathbf{V}$ homogeneous in degree 1 in n_1, n_2, we can immediately write, from (C.17),

$$\mathbf{V} = n_1\left(\frac{\partial \mathbf{V}}{\partial n_1}\right)_{T,P,n_2} + n_2\left(\frac{\partial \mathbf{V}}{\partial n_2}\right)_{T,P,n_1} \tag{C.18}$$

$$= n_1\overline{V}_1 + n_1\overline{V}_2 \tag{C.19}$$

or, dividing by $(n_1 + n_2)$ (C.20)

$$V = x_1\overline{V}_1 + x_2\overline{V}_2 \tag{C.21}$$

In other words, the total system volume is the sum of its parts – the moles of each component times its partial molar volume, summed over all components, Equation (C.18), or a similar relation for the molar volume, Equation (C.21), which is in fact Equation (10.6).

Similarly, because **G** is homogeneous degree 1 in the masses of the components, we know immediately that, if there are c components,

$$\mathbf{G} = n_1\left(\frac{\partial \mathbf{G}}{\partial n_1}\right) + n_2\left(\frac{\partial \mathbf{G}}{\partial n_2}\right) + \cdots + n_c\left(\frac{\partial \mathbf{G}}{\partial n_c}\right) \tag{C.22}$$

$$= \sum_{i=1}^{c} n_i \mu_i \tag{C.23}$$

which is Equation (4.66).

Naturally, it is not essential to understand or to use Euler's theorem in learning or using thermodynamics. But some people find these more mathematical relationships satisfying and illuminating, while others wonder why anyone would bother.

C.2.4 Legendre transforms

The Legendre transform allows one to change a function to a different function having as independent variables the partial derivatives of the original function, without losing any information. This description in words is more difficult than the operation itself. To see its usefulness in thermodynamics, one simply needs to realize that fundamentally the first and second laws of thermodynamics give us a criterion of system stability in terms of entropy (S), volume (V), and energy (U). In other words, we have some very useful relationships beginning with the function

$$U = U(S, V) \tag{C.24}$$

If you then realize that

$$T = (\partial U/\partial S)_V$$

and

$$P = -(\partial U/\partial V)_S$$

and look at the description we have just given of the Legendre transform, you will see that it will allow us to define a new function that is just as useful as (C.24) but that uses T and P as independent variables instead of S and V. The development of thermodynamics does not depend on the Legendre transform, but it is elegant and concise. It illustrates quite beautifully the underlying unity and symmetry among the thermodynamic state functions and their independent variables.

The following geometrical approach is modified from Callen (1960). Given the function

$$y = y(x_1, x_2, x_3, \cdots, x_n)$$

we want a method whereby the derivatives

$$p_i = \partial y/\partial x_i$$

are used as independent variables in a new function containing the same information as the original. To begin, consider a function of a single independent variable

$$y = y(x)$$

Geometrically, $p = dy/dx$ is a tangent (Figure C.4a). We might be tempted to simply eliminate x and find $y = y(p)$, but this would lose some information since knowing y as a function of the slope does not give us y as a function of x (Figure C.4b).

For example, let

$$y = x^3$$

then

$$\begin{aligned} p &= dy/dx \\ &= 3x^2 \end{aligned}$$

and

$$y = (p/3)^{1.5}$$

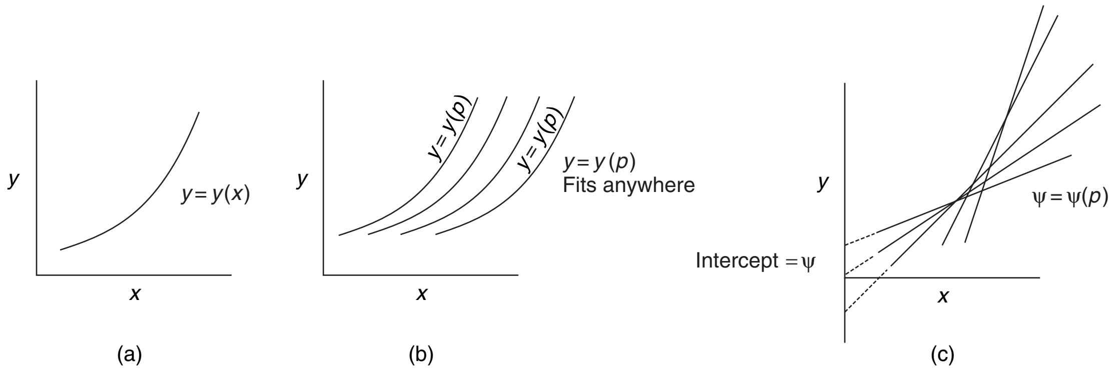


Figure C.4 Illustration of why one must know the slope (p) as a function of the y-intercept (Ψ) to have the same information as one has in the function $y = y(x)$. (a) The function $y = y(x)$. y is known for any x. (b) y is known as a function of the slope ($p = dy/dx$) of $y = y(x)$. This does not fix the position of the curve with respect to the x-axis. (c) The slope p as a function of the y-intercept of the slope (Ψ). This defines an infinite set of tangents that outline the original function $y = y(x)$.

We now have y as a function of dy/dx, but it is not what we want. However, if we knew the *intercept* as a function of the slope, we would have the same information we started with, since the original curve $y = y(x)$ can be considered as being defined or outlined by an infinite number of tangents, each uniquely defined by a slope and intercept (Figure C.4c).

Thus, if Ψ is the intercept, then

$$\Psi = \Psi(p)$$

is the relation we want.

Now since, as shown in Figure C.5,

$$p = \frac{y - \Psi}{x - 0}$$

then

$$\Psi = y - px$$

and, in case you didn't notice, the Legendre transform has been found. It can be shown that in the general case

$$y = y(x_1, x_2, x_3, \cdots, x_n)$$

the Legendre transform is

$$\begin{aligned}\Psi &= y - p_1x_1 - p_2x_2 - p_3x_3 \cdots - p_nx_n \\ &= y - \sum_i p_ix_i \qquad \text{(C.25)}\end{aligned}$$

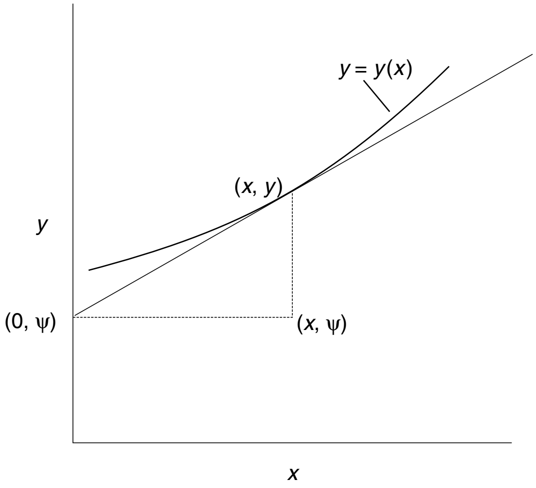


Figure C.5 The function $y = y(x)$ and the tangent to the function at an arbitrary point. Ψ is the y-intercept of the tangent, and the slope $p = (y - \Psi)/(x - 0)$.

That is, to form the Legendre transform of a function, subtract from the original function the products of each variable to be changed and the derivative of the function with respect to that variable. After that, one can proceed to tidy up by eliminating y in the new function by differentiating. Thus in the case of $y = y(x)$,

$$\Psi = y - px$$

$$\begin{aligned} d\Psi &= dy - p\,dx - x\,dp \\ &= -x\,dp \quad (dy = p\,dx \text{ by definition}) \end{aligned}$$

or

$$\frac{d\Psi}{dp} = -x$$

and in the general case

$$d\Psi = -\sum_i x_i dp_i$$

For example, if

$$y = x^3$$

then

$$p = 3x^2$$

and

$$\begin{aligned} \Psi &= y - x \cdot 3x^2 \\ &= x^3 - 3x^3 \\ &= -2x^3 \end{aligned}$$

or since

$$x = (p/3)^{\frac{1}{2}}$$

$$\Psi = -2(p/3)^{\frac{3}{2}}$$

and

$$\begin{aligned} d\Psi/dp &= -(p/3)^{\frac{1}{2}} \\ &= -x \end{aligned}$$

Consider another example. If we have

$$x = y^3 - 3z^2$$

we know x as a function of the two independent variables y and z. If we need a function not of y and z but y and $(\partial x/\partial z)$, we transform one variable, as above. If we need a function not of y and z or y and $(\partial x/\partial z)$ but of $(\partial x/\partial y)$ and $(\partial x/\partial z)$, we transform both variables, and invent the new function h, such that

$$h = x - y(\partial x/\partial y)_z - z(\partial x/\partial z)_y$$

For thermodynamic purposes, that goes far enough, but we can demonstrate that h is a function of $(\partial x/\partial y)$ and $(\partial x/\partial z)$. Thus, since

$$x = y^3 - 3z^2$$

let

$$\begin{aligned} p_1 &= (\partial x/\partial y)_z \\ &= 3y^2 \end{aligned}$$

and

$$\begin{aligned} p_2 &= (\partial x/\partial z)_y \\ &= -6z \end{aligned}$$

Then

$$\begin{aligned} h &= y^3 - 3z^2 - y(3y^2) - z(-6z) \\ &= -2y^3 + 3z^2 \\ &= -2(p_1/3)^{\frac{3}{2}} + 3(-p_2/6)^2 \\ &= (-2/\sqrt{27})p_1^{\frac{3}{2}} + p_2^2/12 \end{aligned}$$

Thus h is a function of p_1 and p_2, and

$$(\partial h/\partial p_1)_{p_2} = -y$$

and

$$(\partial h/\partial p_2)_{p_1} = -z$$

Application to thermodynamics

The Legendre transform is not trivial, but its application to thermodynamics is. For example, start with $U = U(S, V)$, which simply says we have a function U with independent variables S and V. This functional relationship is Equation (4.9), $dU = T\,dS - P\,dV$. Following the recipe, we subtract from U the product of each independent variable times the derivative of the function with respect to that variable. Let's call the new function G. Thus

$$G = U - S\left(\frac{\partial U}{\partial S}\right)_V - V\left(\frac{\partial U}{\partial V}\right)_S$$

and, because $(\partial U/\partial S)_V = T$ and $(\partial U/\partial V)_S = -P$, we get

$$G = U - TS + PV$$

It is not easy to see from this definition that G is in fact a function of T and P. To show this, simply write the differential of G,

$$dG = dU - T\,dS - S\,dT + P\,dV + V\,dP$$

and then expand dU by inserting Equation (4.9),

$$\begin{aligned} dG &= T\,dS - P\,dV - T\,dS - S\,dT + P\,dV + V\,dP \\ &= -S\,dT + V\,dP \end{aligned} \tag{C.26}$$

From (C.26) we see that the independent variables for G are T and P.

By transforming only S we obtain the definition of the Helmholtz energy, and by transforming only V we get the definition of enthalpy. For more information, see Alberty (2001).

Appendix D
How to use SUPCRT92

SUPCRT92 is a FORTRAN program written originally by students and associates of Prof. H.C. Helgeson at the University of California, Berkeley, and in its present form by J.W. Johnson, Lawrence Livermore National Laboratory, California. Anyone can use it, free of charge. It uses a large database of minerals, gases, and aqueous species in which most of the mineral data are from Helgeson et al. (1978), known as HDNB, which was the first large-scale attempt at obtaining internally consistent data for minerals by including phase equilibrium data as well as calorimetric and other sources. It uses the Maier–Kelley heat capacity formulation for minerals, and the revised HKF model for aqueous solutes. It will produce thermodynamic properties to 1000 °C and 5000 bars.

The data file is revised periodically by Everett Shock, and available at

http://geopig.asu.edu/supcrt_data.html.

The data file comes in two forms, sequential and direct access. The readable sequential file must be converted into the direct access form that SUPCRT92 uses by using CPRONS92, a program in the SUPCRT92 group. Modifications to the database are made using MPRONS92. As mentioned in Chapter 13, SUPCRT92 encodes the Haar et al. (1984) equation of state for water, augmented by Levelt Sengers et al. (1983) in the critical region. Dielectric constants are derived with the method of Johnson and Norton (1991).

The program is interactive, with a number of options at each step. The following instructions are sufficient to get you started, but do not encompass all possibilities. If you get lost, hit Ctrl-C (hold down the Ctrl key and hit letter C), and start again. In the following, program responses which you see on screen are in italics.

Change to the SUPCRT92 directory, and type SUPCRT.

```
Welcome to SUPCRT92 (James W. Johnson, author).
last revised 17 July, 1991.
would you like to use the default thermodynamic database? (y/n)
```

The default database is DPRONS92.DAT, and is out of date. Either rename the latest version to this, or type “n”, ENTER, enter the name of the database you want to use, then hit ENTER.

choose file option for specifying reaction-independent parameters.
1= select one of three default files.
2= select an existing nondefault file.
3= build a new file.

SUPCRT92 wants to know what values of T and P you need answers for. The three options under no. 1 are often sufficient (see below), and if so, type “1” and hit ENTER. If you want a range of T and P not included in the choices under “1” (for example 25 to 1000 °C at 1 and 2 bars), type “3” and hit ENTER. Note that under this option, you also have the choice of having SUPCRT92 report the temperatures of a univariant reaction ($\log K = 0$) at specified pressures, isochoric properties, and other options.

input solvent phase region
1 = one-phase region.
2 = liq-vapor saturation curve.
3 = EQ3/6 one phase/sat grid.

Option 1 gives you a table of answers at every 100 °C from 0 to 1000 °C, at every pressure from 500 bars to 5000 bars, at 500 bar intervals. Option 2 gives you answers at 25 °C intervals from 0 to 350 °C along the liquid–vapor curve for water. Option 3 is for use with programs EQ3 and EQ6. If these options do not suit you, start again and “build a new file” at the previous step. You can specify a range of T and P, and the intervals.

choose file option for specifying reactions:
1 = use an existing reaction file.
2 = build a new reaction file.

You must now specify the reaction (or single mineral, gas, etc.) you are intersted in. If you have done this before, and saved it in a file, choose 1, and then give the file name when prompted. If this is something new, type “2”, and hit ENTER.

specify number of reactions to be processed.

You can enter up to 50 separate reactions, one after the other. In the following, I assume you entered “1”.

input title for reaction 1 of 1.

This title will appear on the output for convenience. It does not affect the calculations.

enter [coeff species] pairs, separated by blanks, one pair per line, for reaction 1 (conclude with [0 done]).

Enter the desired reaction with each stoichiometric coefficient and species, separated by a blank, on a separate line. Use negative coefficients for reactants and positive coefficients for products. SUPCRT92 does not check if the reaction is balanced, or if it makes any sense. Each species must be written exactly as it appears in the database, because SUPCRT92 uses the name to find the data. Minerals must be typed in capital letters, aqueous species in capitals, lower case, numerals, and + or − as indicated, and so on.

For example, the reaction

$$\mathrm{Ca_3Fe_2Si_3O_{12}} + 4\mathrm{H_2S}(aq) + 2\mathrm{Cu^+} + 3\mathrm{CO_2}(g) = 3\mathrm{SiO_2}(qtz) + 3\mathrm{CaCO_3} + 2\mathrm{CuFeS_2} + 2\mathrm{H^+} + 3\mathrm{H_2O}$$

would be entered as follows:

```
−1 ANDRADITE
−4 H2S,aq
−2 Cu+
−3 CO2,g
3 QUARTZ
3 CALCITE
2 CHALCOPYRITE
2 H+
3 H2O
0
```

You don't need to type "done".

Note that if you want the properties of a single mineral, ion, or gas, just enter that mineral or ion as a product, with no reactants. For example, for tables of the thermodynamic properties of halite (including $\Delta_a G^\circ$ and $\Delta_a H^\circ$), just enter

```
1 HALITE
0
```

reaction 1 stoichiometry. Is this correct?

The program repeats your input. If you made an error, type "n", and repeat. If the program detects an error (a mineral misspelled, or not in the database) it will tell you, and you need only repeat that entry.

would you like to save these reactions to a file (y/n)

If you will be using this reaction repeatedly, type "y". You can then use this file at the earlier prompt, *choose file option for specifying reactions.*

specify name for tabulated output file:

This is the name of the file your results will be in.

would you like to generate x-y plot files for logK, G, H, S, Cp, and V of reaction? (y/n)

If you say “y”, you get a whole bunch of files with your data arranged in a way more convenient for plotting. The format is such that with minimal editing they can be read and used by spreadsheets and plotting programs. You get all or none of them. They have extensions .kxy, .gxy, .hxy, and so on.

reaction in progress...

…a few more messages…

…execution completed

Your results should now be in the file with the name you specified.

Appendix E
Computer programs

E.1 Fortran

The Pitzer equations have quite a formidable appearance, and although Harvie and Weare (1980) claim that implementation is "rather simple," it can be quite a chore, especially if many components and potential solids are considered. The equations have many double summation terms, and it can be hard to visualize what they would look like when implemented. To help readers understand these equations, I have written a FORTRAN program to calculate the solubility of gypsum in mixed NaCl–Na_2SO_4 solutions at 25 °C. Shown here are two subroutines, which implement the Pitzer approach for this system. The complete program contains additional subroutines which solve for the gypsum content and test whether the solubility product is satisfied, but they have nothing to do with the Pitzer model, and are not included here. The program can be downloaded from http://www.cambridge.org/0521847729. If you do have to copy it, omit the line numbers.

E.1.1 Subroutine Pitzer

Given the concentrations of Na, Ca, Cl, and SO_4 (as some combination of NaCl and Na_2SO_4), this subroutine calculates the activity coefficients of Ca and SO_4 and the activity of water, using the Pitzer equations and the parameters from Harvie and Weare (1980). The expression $a_{\mathrm{Ca}^{2+}} a_{\mathrm{SO}_4^{2-}} a_{\mathrm{H_2O}}^2$ can then be tested against the solubility product.

Even readers not familiar with FORTRAN should be able to see how the many summation signs in the Pitzer equations are expanded into actual expressions, and how the many parameters fit in. The equations are broken up into term1, term2, etc., to correspond with the text in §15.7.2. Here are some comments which may help to understand the program.

- Statements beginning with ! are comments, and are ignored by the program. Symbol & indicates that the statement is continued on the next line.
- Many computer languages, including FORTRAN, require that every variable be initially declared as to its type, whether character, real number, integer, logical, etc. Lines 9–29 do

this. All variables in this case are real (decimal) numbers. Specifying "PARAMETER" simply ensures that the number cannot be changed accidentally during execution.

- Lines 33–53 just define some commonly used things like ionic strength.
- Calculation of the variables ${}^E\theta_{ij}$ and ${}^E\theta'_{ij}$ is more complicated, and so is done in a separate subroutine (called in line 56).
- Lines 59–61 then define the parameters that use these terms.
- Most of the parameter names should be more or less self-evident, but the parameters $B^{(0)}$, $B^{(1)}$, and $B^{(2)}$ have been combined into three-element vectors. For example, variable BNaCl has three elements, in which BNaCl(0) $= B^{(0)}$, BNaCl(1) $= B^{(1)}$, and BNaCl(2) $= B^{(2)}$.
- The rest of the lines then do the real work of calculating the activity coefficients of Ca^{2+} and SO_4^{2-} and the activity of water.
  - The Debye–Hückel term in lines 67–79.
  - $\gamma_{Ca^{2+}}$ in lines 81–95.
  - $\gamma_{SO_4^{2-}}$ in lines 97–110.
  - The osmotic coefficient and a_{H_2O} in lines 12–127.
  - Functions $g(x)$ and $g'(x)$ are defined in lines 131–141.
- Note that the use of the $B^{(0)}$, $B^{(1)}$, and $B^{(2)}$ terms is different depending on which ions are concerned. Thus in term1, term2, term6, and termb, the line concerning Ca and SO_4 is different from the other lines. This is because both ions are doubly charged, resulting in stronger attraction, requiring different fitting parameters.

To make the program easier to read, Equation (15.38) in Chapter 15 is split into a number of terms which have separate statements in the program. These are as follows.

$$
\begin{aligned}
\ln \gamma_M = & z_M^2 F \\
& + \underbrace{\sum_{a=1}^{Na} m_a (2B_{Ma} + ZC_{Ma})}_{\text{term2}} \\
& + \underbrace{\sum_{c=1}^{Nc} m_c \left(2\Phi_{Mc} + \sum_{a=1}^{Na} m_a \Psi_{Mca} \right)}_{\text{term3}} \\
& + \underbrace{\sum_{a=1}^{Na-1} \sum_{a'=a+1}^{Na} m_a m_{a'} \Psi_{Maa'}}_{\text{term4}} \\
& + \underbrace{|z_M| \sum_{c=1}^{Nc} \sum_{a=1}^{Na} m_c m_a C_{ca}}_{\text{term5}} \\
& + \sum_{n=1}^{Nn} m_n (2\lambda_{nM})
\end{aligned}
$$

Equation (15.37) is split up as follows.

$$
\begin{aligned}
\ln \gamma_X = {} & z_X^2 F \\
& + \underbrace{\sum_{c=1}^{Nc} m_c (2B_{cX} + ZC_{cX})}_{\text{term6}} \\
& + \underbrace{\sum_{a=1}^{Na} m_a \left(2\Phi_{Xa} + \sum_{c=1}^{Nc} m_c \Psi_{Xac} \right)}_{\text{term7}} \\
& + \underbrace{\sum_{c=1}^{Nc-1} \sum_{c'=c+1}^{Nc} m_c m_{c'} \Psi_{cc'X}}_{\text{term8}} \\
& + \underbrace{|z_X| \sum_{c=1}^{Nc} \sum_{a=1}^{Na} m_c m_a C_{ca}}_{\text{term9}} \\
& + \sum_{n=1}^{Nn} m_n (2\lambda_{nX})
\end{aligned}
$$

Equation (15.35) is split as follows.

$$
\begin{aligned}
\phi - 1 = \frac{2}{\sum_i m_i} \times \Bigg[& \underbrace{\frac{-A^{\phi} I^{3/2}}{1 + bI^{1/2}}}_{\text{terma}} + \underbrace{\sum_{c=1}^{Nc} \sum_{a=1}^{Na} m_c m_a (B_{ca}^{\phi} + ZC_{ca})}_{\text{termb}} \\
& \underbrace{+ \sum_{c=1}^{Nc-1} \sum_{c'=c+1}^{Nc} m_c m_{c'} \left(\Phi_{cc'}^{\phi} + \sum_{a=1}^{Na} m_a \Psi_{cc'a} \right)}_{\text{termc}} \\
& + \underbrace{\sum_{a=1}^{Na-1} \sum_{a'=a+1}^{Na} m_a m_{a'} \left(\Phi_{aa'}^{\phi} + \sum_{c=1}^{Nc} m_c \Psi_{aa'c} \right)}_{\text{termd}} \\
& + \sum_{n=1}^{Nn} \sum_{a=1}^{Na} m_n m_a \lambda_{na} + \sum_{n=1}^{Nn} \sum_{c=1}^{nc} m_n m_c \lambda_{nc} \Bigg]
\end{aligned}
$$

Equation (15.41) is split as follows.

$$
\begin{aligned}
F = {} & \underbrace{-A^{\phi} \left(\frac{I^{1/2}}{1 + bI^{1/2}} + \frac{2}{b} \ln(1 + bI^{1/2}) \right)}_{\text{f\_gamma}} + \underbrace{\sum_{c=1}^{Nc} \sum_{a=1}^{Na} m_c m_a B'_{ca}}_{\text{term1}} \\
& + \underbrace{\sum_{c=1}^{Nc-1} \sum_{c'=c+1}^{Nc} m_c m_{c'} \Phi'_{cc'}}_{\text{term1a}} + \underbrace{\sum_{a=1}^{Na-1} \sum_{a'=a+1}^{Na} m_a m_{a'} \Phi'_{aa'}}_{\text{term1b}}
\end{aligned}
$$

```
SUBROUTINE pitzer(mNa,mCl,mCa,mSO4,gamma_Ca,gamma_SO4,aH2O)

!Uses the Pitzer equations to calculate the activity coefficients of Ca and SO4
!and the activity of H2O in Ca-Na-Cl-SO4 solutions. Illustrates the use of the Pitzer equations.
!Equation numbers from Harvie & Weare (1980) are indicated by HW().

IMPLICIT none

REAL,INTENT(in) :: mNa, mCl, mSO4, mCa
REAL,INTENT(out):: gamma_Ca, gamma_SO4, aH2O
REAL            :: v, w, y, f_gamma, Z, F, IS, ISp
REAL            :: phi_1 NaCa, phip_NaCa, phiphi_NaCa, Etheta, Ethetap
REAL            :: phi_ClSO4, phip_ClSO4, phiphi_ClSO4
REAL            :: term1, term1a, term1b, term2, term3
REAL            :: term4, term5, term6, term7, term8, term9
REAL            :: terma, termb, termc, termd
REAL            :: CNaCl, CNaSO4, CCaCl, CCaSO4
REAL            :: ln_gamma_Ca, ln_gamma_SO4, osmotic, sum_m
!parameter data are from Harvier & Weare (1980).
REAL,PARAMETER  :: Aphi = 0.392, b = 1.2
REAL,PARAMETER  :: alpha = 2.0, alpha1 = 1.4, alpha2 = 12.0
REAL,PARAMETER  :: theta_NaCa = 0.07, theta_ClSO4 = 0.02
REAL,PARAMETER  :: psi_NaCaCl = -0.014, psi_NaCaSO4 = -0.023
REAL,PARAMETER  :: psi_ClSO4Na = 0.0014, psi_ClSO4Ca = 0.0
REAL,PARAMETER  :: CphiNaCl=0.00127, CphiNaSO4=0.00497, CphiCaCl=-0.00034, CphiCaSO4=0.0
REAL,DIMENSION(3),PARAMETER :: BNaCl  = (/ 0.07650, 0.2664, 0.0 /)
REAL,DIMENSION(3),PARAMETER :: BNaSO4 = (/ 0.01958, 1.1130, 0.0 /)
REAL,DIMENSION(3),PARAMETER :: BCaCl  = (/ 0.31590, 1.6140, 0.0 /)
REAL,DIMENSION(3),PARAMETER :: BCaSO4 = (/ 0.20000, 2.6500, -57.70 /)

EXTERNAL mixing

!SOME PRELIMINARIES:
!===================

!sum of all m terms
sum_m = mNa+mCa+mCl+mSO4

!Ionic strength functions
IS = (mNa+mCl+4*mCa+4*mSO4)/2
ISp = sqrt(IS)

!some commonly used parameters
v = alpha *ISp
w = alpha1*ISp
y = alpha2*ISp
Z = mNa+mCl+2*mCa+2*mSO4

!convert Cphi to C                                                      HW(9)
CNaCl  = CphiNaCl /2
CNaSO4 = CphiNaSO4/(2*sqrt(2.0))
CCaCl  = CphiCaCl /(2*sqrt(2.0))
CCaSO4 = CphiCaSO4/4

!calculate unsymmetrical mixing terms Etheta, Ethetap
```

```
CALL mixing(IS,Etheta,Ethetap)

!calculate phi terms: interactions between ions of like sign
phiphi_NaCa = theta_NaCa + Etheta + IS*Ethetap                                  HW(10)
phi_NaCa = theta_NaCa + Etheta
phip_NaCa = Ethetap

phiphi_ClSO4 = theta_ClSO4 + Etheta + IS*Ethetap
phi_ClSO4 = theta_ClSO4 + Etheta
phip_ClSO4 = Ethetap

!D-H TERM:
!=========

term1 = mNa*mCl *(BNaCl(2) *gp(v)/IS                 )           &
     + mNa*mSO4*(BNaSO4(2)*gp(v)/IS                 )           &
     + mCa*mCl *(BCaCl(2) *gp(v)/IS                 )           &
     + mCa*mSO4*(BCaSO4(2)*gp(w)/IS + BCaSO4(3)*gp(y)/IS)

term1a = mNa*mCa *phip_NaCa
term1b = mCl*mSO4*phip_ClSO4

f_gamma = -Aphi*((ISp/(1+b*ISp)) + (2/b)*(log(1+b*ISp)))
F = f_gamma + term1 + term1a + term1b                                            HW(3)

!ACTIVITY COEFFICIENT OF Ca:
!===========================

term2 = mCl *(2*(BCaCl(1) + BCaCl(2) *g(v)                    ) + Z*CCaCl)    &
       + mSO4*(2*(BCaSO4(1)+ BCaSO4(2)*g(w) + BCaSO4(3)*g(y)) + Z*CCaSO4)

term3 = mNa*(2*phi_NaCa + mCl*psi_NaCaCl + mSO4*psi_NaCaSO4)

term4 = mCl*mSO4*psi_ClSO4Ca

term5 = 2*(mNa*mCl*CNaCl + mNa*mSO4*CNaSO4                                      &
       + mCa*mCl*CCaCl + mCa*mSO4*CCaSO4)

ln_gamma_Ca = 4*F + term2 + term3 + term4 + term5                              HW(2b)
gamma_Ca = exp(ln_gamma_Ca)

!ACTIVITY COEFFICIENT OF SO4:
!============================

term6 = mNa*(2*(BNaSO4(1) + BNaSO4(2)*g(v)                   ) + Z*CNaSO4)    &
       + mCa*(2*(BCaSO4(1) + BCaSO4(2)*g(w) + BCaSO4(3)*g(y)) + Z*CCaSO4)

term7 = mCl*(2*phi_ClSO4 + mNa*psi_ClSO4Na + mCa*psi_ClSO4Ca)

term8 = mNa*mCa*psi_NaCaSO4

term9 = term5

ln_gamma_SO4 = 4*F + term6 + term7 + term8 + term9
gamma_SO4 = exp(ln_gamma_SO4)                                                  HW(2c)
```

```
!OSMOTIC COEFFICIENT:
!====================

terma = -Aphi*IS**(1.5)/(1+b*ISp)

termb = mNa*mCl   * (BNaCl(1) + BNaCl(2) *exp(-v)                + Z*CNaCl ) &
        + mNa*mSO4 * (BNaSO4(1)+ BNaSO4(2)*exp(-v)               + Z*CNaSO4) &
        + mCa*mCl  * (BCaCl(1) + BCaCl(2) *exp(-v)               + Z*CCaCl ) &
        + mCa*mSO4 * (BCaSO4(1)+ BCaSO4(2)*exp(-w) + BCaSO4(3)*exp(-y) + Z*CCaSO4)

termc = mNa*mCa*(phiphi_NaCa + mCl*psi_NaCaCl + mSO4*psi_NaCaSO4)

termd = mCl*mSO4*(phiphi_ClSO4 + mNa*psi_ClSO4Na + mCa*psi_ClSO4Ca)

osmotic = 1+(2/sum_m)*(terma + termb + termc + termd)                       HW(2a)
aH2O = exp(-osmotic*sum_m/55.51)

CONTAINS

function g(x)
      IMPLICIT none
      REAL :: g,x
      g = 2*(1-(1+x)*exp(-x))/x**2
END function g

function gp(x)
      IMPLICIT none
      REAL :: gp,x
      gp = -2*(1-(1+x+x**2/2)*exp(-x))/x**2
END function gp

END SUBROUTINE pitzer
```

E.1.2 Subroutine Mixing

Subroutine Mixing is called by subroutine Pitzer, and calculates the variables ${}^E\theta_{ij}$ and ${}^E\theta'_{ij}$, given the ionic strength. These variables were defined by Pitzer (1975, 1983), and account for the interaction of ions of like but different charge, that is, where i and j are both cations or both anions (such as Na^+ and Ca^{2+}, or Cl^- and SO_4^{2-} in this case). The term ${}^E\theta_{ij}$ was derived by Pitzer from the statistical mechanics theory of Friedman (1962). Its evaluation involves integrals [functions J_0, J_1 in the notation of Harvie and Weare (1980)] which have no analytical solution, and Pitzer (1975, 1995) gives approximate methods. Harvie (1981) and Harvie and Weare (1980) used a more accurate scaled Chebyshev expansion method, also described in Pitzer (1987, 1991), which is used in this subroutine.

Harvie (1981) fit numerically derived values of J_0 with two Chebyshev polynomial approximations, one for $x \le 1$ (lines 47–53) and one for $x > 1$ (lines 55–60), in which arrays *bk* and *dk* are calculated. Then using these, J_0 and its x-derivative J_1 are calculated from the formulas in lines 63–72. There are three varieties of each *J*-function, because they depend on the charges on the ions as well as the ionic strength. The differences are established in lines 37–39. The subroutine uses double precision throughout, using the symbol `dp`, established in line 11.

```
SUBROUTINE mixing(I,Etheta,Ethetap)                                                              1
                                                                                                 2
!Evaluates unsymmetrical mixing terms Etheta, Ethetap, using the Chebyshev approximation         3
!mentioned in Harvie & Weare (1980), Appendix.                                                   4
                                                                                                 5
IMPLICIT none                                                                                    6
                                                                                                 7
REAL,INTENT(in)              :: I                                                                8
REAL,INTENT(out)             :: Etheta,Ethetap                                                   9
INTEGER                      :: m,k                                                             10
INTEGER,PARAMETER            :: dp = kind(1.0D0)                                                11
REAL(DP)                     :: x,xMN,xMM,xNN,z,dzdx,JMN,JMM,JNN,JpMN,JpMM,JpNN                 12
REAL(DP),PARAMETER           :: Aphi = 0.392                                                    13
REAL(DP),DIMENSION(0:20,2)   :: ak                                                              14
REAL(DP),DIMENSION(0:22)     :: bk,dk                                                           15
                                                                                                16
!array ak values are from Pitzer (1991) Table B-1, and copied from file phrqpitz.for            17
!in the USGS phrqptz distribution                                                               18
ak(0:20,1) = (/1.925154014814667_dp, -0.060076477753119_dp, -0.029779077456514_dp,     &        19
             -0.007299499690937_dp,  0.000388260636404_dp,  0.000636874599598_dp,     &        20
              0.000036583601823_dp, -0.000045036975204_dp, -0.000004537895710_dp,     &        21
              0.000002937706971_dp,  0.000000396566462_dp, -0.000000202099617_dp,     &        22
             -0.000000025267769_dp,  0.000000013522610_dp,  0.000000001229405_dp,     &        23
             -0.000000000821969_dp, -0.000000000050847_dp,  0.000000000046333_dp,     &        24
              0.000000000001943_dp, -0.000000000002563_dp, -0.000000000010991_dp/)             25
                                                                                                26
ak(0:20,2) = (/0.628023320520852_dp,  0.462762985338493_dp,  0.150044637187895_dp,     &        27
             -0.028796057604906_dp, -0.036552745910311_dp, -0.001668087945272_dp,     &        28
              0.006519840398744_dp,  0.001130378079086_dp, -0.000887171310131_dp,     &        29
             -0.000242107641309_dp,  0.000087294451594_dp,  0.000034682122751_dp,     &        30
             -0.000004583768938_dp, -0.000003548684306_dp, -0.000000250453880_dp,     &        31
              0.000000216991779_dp,  0.000000080779570_dp,  0.000000004558555_dp,     &        32
             -0.000000006944757_dp, -0.000000002849257_dp,  0.000000000237816_dp/)             33
                                                                                                34
!zCa = +2; zNa = +1 and zSO4 = -2; zCl = 1                                                      35
!so if M is Ca or SO4 and N is Na or Cl,                                                        36
xMN = 6*2*Aphi*sqrt(I)    ! i.e., 6*zCa*zNa*0.0392*sqrt(I); 6*2*zCl*zSO4*0.0392*sqrt(I)    HW(A1)
xMM = 6*4*Aphi*sqrt(I)    ! i.e., 6*zCa*zCa*0.0392*sqrt(I); 6*4*zSO4*zSO4*0.0392*sqrt(I)       38
xNN = 6*1*Aphi*sqrt(I)    ! i.e., 6*zNa*zNa*0.0392*sqrt(I); 6*1*zCl*zCl*0.0392*sqrt(I)         39
                                                                                                40
!this DO loop evaluates the J functions as described in Pitzer (1991) appendix B.               41
DO k=1,3                                                                                        42
      IF (k==1) x=xMN                                                                           43
      IF (k==2) x=xMM                                                                           44
      IF (k==3) x=xNN                                                                           45
                                                                                                46
      IF (x <= 1)            THEN                                                               47
                 z = 4.0_dp * x**(0.2_dp) - 2.0_dp                                              48
                 dzdx = (0.8_dp) * x**(-0.8_dp)                                                 49
                 DO m=20,0,-1                                                                   50
```

```
                    bk(m) = z*bk(m+1) - bk(m+2) + ak(m,1)
                    dk(m) = bk(m+1) + z*dk(m+1) - dk(m+2)
          END DO
ELSE
          z = (40.0_dp/9.0_dp) * x**(-0.1_dp) - 22.0_dp/9.0_dp
          dzdx = -(40.0_dp/90.0_dp) * x**(-1.1_dp)
          DO m=20,0,-1
                    bk(m) = z*bk(m+1) - bk(m+2) + ak(m,2)
                    dk(m) = bk(m+1) + z*dk(m+1) - dk(m+2)
          END DO
END IF

IF (k==1) THEN
          JMN = 0.25_dp*x -1.0_dp + 0.5_dp*(bk(0) - bk(2))
          JpMN = 0.25_dp + 0.5_dp*(dzdx)*(dk(0) - dk(2))
ELSE IF (k==2) THEN
          JMM = 0.25_dp*x -1.0_dp + 0.5_dp*(bk(0) - bk(2))
          JpMM = 0.25_dp + 0.5_dp*(dzdx)*(dk(0) - dk(2))
ELSE
          JNN = 0.25_dp*x -1_dp + 0.5_dp*(bk(0) - bk(2))
          JpNN = 0.25_dp + 0.5_dp*(dzdx)*(dk(0) - dk(2))
END IF
END DO

!finally, calculation of the Etheta terms.
Etheta = (2.0_dp/(4.0_dp*I)) * (JMN - 0.5_dp*JMM - 0.5_dp*JNN)                    ! HW(A2)
Ethetap=-(Etheta/I)+(2.0_dp/(8.0_dp*I**2))*(xMN*JpMN-0.5_dp*xMM*JpMM-0.5_dp*xNN*JpNN)! HW(A3)

END SUBROUTINE mixing
```

E.2 MATLAB

E.2.1 Program SPECIES

This is essentially the program EQBRM (Anderson and Crerar, 1993, Appendix E) written in MATLAB®. It uses input prepared by a spreadsheet and is easily modified to handle multiple inputs.

```
% SPECIES.M
% speciation program, patterned after EQBRM.
% like EQBRM, can only operate on one composition at a time.
% all input data (except species labels) are in a single matrix.
% m is number of species; n is number of rxns; (m-n-1) is number of components.

% input matrix size is (m,m+5)
% matrix format:
% ==============
% row 1   : m,n,all other columns zero.
% row 2   : stoichiometric coeffs for rxn 1.
%  :
% row n+1 : stoichiometric coeffs for rxn n.
% row n+2 : mass balance  for component 1.
%  :
% row m   : mass balance for component (m-n-1).
% row m+1 : (m-n-1) compositions, all other columns zero.
```

```
% row m+2 : n K values, all other columns zero.
% row m+3 : m valences.
% row m+4 : m initial guesses.
% row m+5 : DHA, DHB, all other columns zero.
% row m+6 : m values of anot, if any (not required).
% row m+7 : m values of bdot, if any (not required).

% written February 2003 by GMA.

% data input
s=input('enter vector of species names (not required)> ');
in=input('enter matrix of data> ');
m=in(1,1);
n=in(1,2);
coeffs=in(2:n+1,:);
mass=in(n+2:m,:);
B=in(m+1,1:(m-n-1));
K=in(m+2,1:n);
z=in(m+3,:);
a=in(m+4,:);
DHA=in(m+5,1);
DHB=in(m+5,2);
if (s=='')
    s=[1:m]';
end

%use Davies or expanded DH?

if (DHB~=0)
  anot=in(m+6,:);
  bdot=in(m+7,:);
end

%expand a into nxm matrix and (n-1)xm matrix

a_n  = repmat(a,n,1);
a_n_1 = repmat(a,m-n-1,1);

%clear any old values
clear gam_n test k b;

%initial gamma values
gam_n(1:n,1:m)=1;

%START ITERATION HERE

test(1:m)=1;
j=0;
while any(test>=1e-10)
  j=j+1;

%construct Jacobian matrix of derivatives C

dK=0.43429.*coeffs./a_n;
C=[z;mass;dK];

%construct Y vector

zm=sum(z.*a)';
i=0;
while i<(m-n-1)
  i=i+1;
```

```
  b(i)=-B(i)+sum(mass(i,:).*a_n_1(i,:));
end
bm=b’;
i=0;
while i<n
  i=i+1;
  k(i)=-K(i)+sum(coeffs(i,:).*log10(a_n(i,:)))+sum(coeffs(i,:).*log10(gam_n(i,:)));
end
km=k’;
Y=-[zm;bm;km];

%evaluate deltaa matrix

deltaa=C\Y;

%improve estimated compositions

aold=a;
i=0;
while i<m
  i=i+1;
  a(i)=aold(i)+deltaa(i);
%prevent negative compositions
  if(a(i))<=0
    a(i)=1e-20;
  end
end

test=abs((a-aold)./aold).*100;

%expand new a values

a_n =repmat(a,n,1);
a_n_1=repmat(a,m-n-1,1);

%calculate new gammas
IS=0.5*sum(a.*z.ˆ2);
if (DHB˜=0)      %if DHB not zero use extended DH
  loggam=((-DHA.*z.ˆ2.*ISˆ.5)./(1+DHB*anot.*ISˆ.5))+bdot.*IS;
else             %else use Davies
  loggam=((-DHA.*z.ˆ2.*ISˆ.5)./(1+ISˆ.5))+0.2*DHA*IS.*z.ˆ2;
end

gam=exp(loggam.*2.30259);

%expand gamma matrix

gam_n=repmat(gam,n,1);

%should converge in much less than 100 iterations

if j>=100
  disp(’Iterated 100 times without convergence.’)
  break
end

end  %END ITERATION

%this section calculates K for each reaction from calculated species activities
%and compares these with input K values as a check.

logact=log10(a.*gam);
```

```
k=0;
  while k<size(coeffs,1)
       k=k+1;
     K1=sum(coeffs(k,:).*logact);
       if k==1
            K2=K1;
        elseif k>1
            K2=[K2,K1];
        end
    end

%K2 is now a vector of calculated K values.

K3 = [K2;K];

%K3 is a matrix of calculated (row 1) and input (row 2) K values.
%if the two rows are identical, speciation is OK.

msg=0;
K4=1-K2(:)'./K;
     if any(abs(K4)>0.0001)
        msg=1
        disp('K values in rxn NG')
  end

%output
disp(' ')
disp(datestr(now))
disp(' ')
if msg==0
     disp('all speciated K values are OK')
else
     disp('some speciated K values are NG')
end
disp(' ')
disp('No. of iterations:')
disp(j)
%uncomment next two lines to see calculated and input K values
%disp('check K values:')
%disp(K3)
output=cell(m+1,4);
output(1,1)={'Species'};
output(1,2)={'Molalities'};
output(1,3)={'Gammas'};
output(1,4)={'Activities'};
for i=1:m
     output(i+1,1)={s(i,:)};
     output(i+1,2)={a(i)};
     output(i+1,3)={gam(i)};
     output(i+1,4)={a(i)*gam(i)};
end
disp(output)

%edit these lines to use column number for H+.
disp('pH is:')
disp(-log10(a(1)*gam(1)))
```

The spreadsheet input for the CO_2 speciation in Chapter 16 looks like this:

	1	2	3	4			
	H	OH	H2CO3	HCO3	CO3		Row
	5	3	0	0	0	m,n	1
rxn 1	-1	-1	0	0	0	H + OH = H2O	2
2	-1	0	1	-1	0	H + HCO3 = H2CO3	3
3	-1	0	0	1	-1	H + CO3 = HCO3	4
	0	0	1	1	1	CO3 mass balance	5
	0.1	0	0	0	0	CO3 composition	6
	14.0	6.37	10.33	0	0	K values	7
	1	-1	0	-1	-2	valence	8
	1E-07	1E-07	0.01	0.001	0.0001	initial guess	9
	0.5114	0.3288	0	0	0	DH A,B	10
	9	3.5	3	4.5	3	anot	11
	0.041	0.041	0.0	0.041	0.041	bdot	12

Only the number matrix is copied to Matlab; the labels are for clarity.

The program assumes that H^+ is the species in column 1. If it is not, edit the last line of the program so that the reported pH uses a different column number.

Rows 11 and 12 are optional. If they are included, the program uses the extended form of the Debye–Hückel equation. If only one number is entered (in column 1) in row 10, it is intepreted as the *A* term in the Davies equation, and that is used, as in the original EQBRM.

Appendix F
Symbols used

F.1 Variables

a	activity.
$\mathbf{A}$	total Helmholtz free energy; total Helmholtz energy.
A	molar Helmoltz free energy, molar Helmholtz energy.
B, C,…	virial coefficients.
f	fugacity.
$\mathbf{G}$	total Gibbs free energy; total Gibbs energy.
G	molar Gibbs free energy; molar Gibbs energy.
$\overline{G}$	partial molar Gibbs energy.
$\mathbf{H}$	total enthalpy.
H	molar enthalpy.
$\overline{H}$	partial molar enthalpy.
K	equilibrium constant.
m	molality.
M	molarity.
n	mole number; number of moles.
$\boldsymbol{q}$	an unspecified amount of heat energy.
q	heat energy per mole of reaction.
$\mathbf{S}$	total entropy.
S	molar entropy.
$\overline{S}$	partial molar entropy.
T or T K	temperature in kelvins.
T °C	temperature in degrees Celsius.
P	pressure.
$\mathbf{U}$	total internal energy.
U	molar internal energy.
$\mathbf{V}$	total volume.
V	molar volume.
$\overline{V}$	parial molar volume.
$\boldsymbol{w}$	an unspecified amount of work energy.
w	work energy per mole of reaction.

x	mole fraction in a liquid or solid phase.
y	mole fraction in a gas phase.
α	residual volume, in calculation of fugacity.
	function used in the Gibbs–Duhem integration.
γ	activity coefficient.
γ_{R}	Raoultian activity coefficient.
γ_{H}	Henryan activity coefficient.
γ_f	fugacity coeffient.
μ	chemical potential.

F.2 Superscripts

$^\circ$	standard state.
$^\infty$	infinite dilution.
pure	a pure phase, e.g., pure gaseous CO_2.
halite	a pure phase, made explicit.
R	residual, e.g., $V^{\mathrm{R}} = V - V^{\text{ideal gas}}$

F.3 Subscripts

f	formation from the elements, as in $\Delta_f G^\circ$, $\Delta_f H^\circ$.
a	apparent formation, as in $\Delta_a G^\circ$, $\Delta_a H^\circ$.
r	reaction, as in $\Delta_r G$, $\Delta_r H$.
r	reference, as in T_{r}, P_{r}.
R	reduced, as in reduced temperature, T_{R}.
rev	reversible.
i	any specified species or component, as in μ_i, the chemical potential of i.
$\hat{i}$	all specified species or components *except* i.

F.4 Comments on non-IUPAC usage

t	is recommended for Celsius temperatures.
T	is recommended for kelvin temperatures.
p	is recommended for pressure.
	I prefer T °C, T or T K, and P.
*	is recommended for a pure phase.
	I use it so seldom it seemed unwise to introduce yet another symbol.
ϕ	is recommended for fugacity coefficient.

f	is recommended for Raoultian activity coefficient.
γ_m	is recommended for Henryan activity coefficient, using molalities. I prefer γ_f, γ_R, and γ_H to indicate the essential similarity of these quantities.
V, H, etc.	is assumed for total volume, total enthalpy, etc.
V_m, H_m, etc.	is recommended for molar volume, molar enthalpy, etc. although lower case v, h, etc., is becoming common too.
V', H', etc.	is recommended for partial molar quantities. The distinction between total and molar quantities is of course vital to understanding thermodynamics. I prefer to emphasize this difference by using bold symbols for total quantities, as shown above. I have previously used roman and italic letters (e.g., V versus V) for this distinction, but now believe this is an insufficient distinction. The overbar notation for partial molar quantities (e.g., $\overline{V}$) is widely used, and I prefer it.

Appendix G
A short history of thermodynamic constraints

G.1 Introduction

In this text, particularly in Chapter 4, I have tried to make clear the importance of the concept of constraints. Here I present what little I know about the historical role of this concept. There are only a few texts which use the constraint concept in a more than perfunctory way. It is just between this usage is so uncommon that I present thissummary.

G.2 Schottky et al.(1929)

It appears that the concept of constraints in thermodynamics was first clearly stated by Schottky et al. (1929). Dr. Edgar Froese has kindly provided a translation of some key passages. Italics are in the original. A comment by Dr. Froese is in bold text.

The measure of irreversibility and the concept of constraint (p. 54–56).

A consideration of the heat exchanges accompanying an irreversible process $1 \longrightarrow 2$ and a reversible process $2 \longrightarrow 1$ to restore the system to state 1 gives

$$-\int_2^1 \frac{dq_{\text{rev}}}{T} - \int_1^2 \frac{dq_{\text{irrev}}}{T} > 0$$

and since the first integral is $S_1 - S_2$,

$$(S_2 - S_1) - \int_1^2 \frac{dq_{\text{irrev}}}{T} > 0$$

It is necessary to consider an apparent paradox in connection with this relationship. The reversible change $2 \rightarrow 1$ assumes that the system passes through a succession of equilibrium states and, in particular, that state 2 and state 1 are both equilibrium states. On the other hand, we postulate an irreversible process $1 \rightarrow 2$ without changing the external conditions. It appears as if an equilibrium state changed spontaneously to a different state. Thus it appears not to have been in equilibrium.

The solution to this paradox, which, in fact, is necessary to make the previous deductions valid, is that state 1 is a constrained system. It does not proceed irreversibly to state 2 because it is hindered to do so by some constraints. Were these constraints absolute, never to be released, the previous considerations would have been pointless; the application is rather limited to constraints which can be arbitrarily removed (without thermodynamic means) thus permitting the irreversible process 1 → 2. Such constraints can be, e.g. produced by:

- locking a piston in a cylinder separating two gases at different pressures;
- resistance to deformation of a solid;
- constraints in chemical change like nonreacting cold H_2 and O_2;
- constraints in heat exchange between different parts of a system.

In all these cases, a further condition must be fulfilled in order to assign a definite entropy increase to the irreversible changes. It must be possible not only to release the constraints and thereby induce the change to state 2, but also, after the release of the constraints and after completion of the process, to bring about *a reversible restoration of the change by means of only external thermodynamic effects*. After all, only in such case is a thermodynamic definition and determination of entropy possible.

The independent chemical changes as constrained changes (p. 135; p. 144)

Changes here considered are characterized by the fact that they can occur at fixed external conditions. Therefore, they are constrained changes. We are dealing with different equilibrium states at the same physical conditions (same temperature and either same P or V). That is only possible if these are constraints to the progress of these considered changes. Example: the reaction $2H_2 + O_2 \rightarrow 2H_2O$, at ordinary temperatures, is generally a constrained reaction. A catalyst or heating will cause the system to leave its apparently, stable equilibrium state. We can say, generally, that *chemical thermodynamics is thermodynamics of constrained equilibrium states*. There is only one state of a system which will not change upon removal of the constraint; that is the state of complete, unconstrained equilibrium.

The multiplicity of the inner (chemical) state, which is independent of external conditions, is made possible only by the fact that the transition of inner states, from one to another, is a constrained process.

Schottky et al. (1929) proceeded to derive the extent of reaction ξ (p. 151), apparently independently of De Donder (1920). There is no definite statement that ξ is used as a possible constraint, but I think it is implied. They derive

$$\left(\frac{\partial \mathbf{F}}{\partial \xi}\right)_{V,T} = \left(\frac{\partial \mathbf{G}}{\partial \xi}\right)_{P,T} = \sum \nu_i \mu_i$$

For this, they use a special symbol, $\Re$, so their $\Re = -\mathcal{A}$, where $\mathcal{A}$ is the De Donder affinity, again without reference to De Donder. So it seems to me that his greatest contribution is the clear recognition of constrained equilibria as a necessity if one wants to compare two different states at the same external conditions. This, of course, requires that thermodynamic properties in constrained states are defined.

G.3 Tisza (1966)

This is a collection of articles written between 1949 and 1965. Tisza says (p. 41)

> Gibbs stated the extremum principle in two versions: *In an isolated system the entropy tends to a maximum at constant energy, or alternatively, the energy tends to a minimum at constant entropy.*
>
> Although these statements undoubtedly express important truths, they lack precision to the point of being paradoxical. If an isolated system is not at equilibrium, we can associate no entropy to it, and if it is in equilibrium, its entropy can no longer increase. Many authors have grappled with this dilemma until a satisfactory solution was found in terms of the *composite system.*

Tisza does not make much use of the term *constraint*, but was of course aware that the composite system implies the existence of constraints. He says (p. 120–121)

> The nature of this problem is illustrated by the following paradox: How are we to give a precise meaning to the statement that entropy tends toward a maximum, whereas entropy is defined only for systems in equilibrium? Thus in an isolated system, the entropy is constant, if it is defined at all.
>
> This difficulty is resolved in a natural way by the artifice of composite systems that enable us to deal with more or less constrained equilibria.

G.4 Callen (1960)

Callen acknowledges being greatly influenced by Tisza. He says (1960)

> The basic problem of thermodynamics is the determination of the equilibrium state that eventually results after the removal of internal constraints in a closed composite system

Implicit in this statement is that the state of the system *before* the removal of the constraint was also an equilibrium state. Callen describes a composite system as a system made up of two or more *simple* systems, defined in turn as

> systems that are macroscopically homogeneous, isotropic, uncharged, and chemically inert, that are sufficiently large that surface effects can be neglected, and that are not acted on by electric, magnetic or gravitational fields.

Constraints are defined as

> Constraints that prevent the flow of energy, volume, or matter among the simple systems constituting the composite system are known as *internal* constraints.

An example, would be the cylinder containing a gas and having an internal piston that we considered in Chapter 4 (Figure 4.1). In this text I identify such states having an extra constraint as metastable states, but Callen does not use the term metastable in this connection.

Using his definition of a simple system, all manner of third constraints can be accommodated [although the film example, Equation (4.13), would be excluded]. Thus if the simple system is isotropic, then the stretched system (Equation 4.12) is a composite system. If the simple system is a crystal of K-feldspar in water, then it is inert, unreactive, and a composite system. Releasing the "unreactive" constraint in steps of $d\xi$ allows the dissolution to proceed (Equation 4.14). This formulation has the advantage of stressing that the constraint and its release are in fact conceptual, not necessarily real.

I suggest (in Chapter 4) that for geochemical purposes it would be simpler instead to define Callen's composite system as one having any kind of third constraint. Thus we could think of the dissolving K-feldspar case, and indeed all chemical reactions, as cases where the third constraint is simply separation of the reactants, and removing the constraint in steps of $d\xi$ is visualized by removing the separation for small amounts of reactants (Figure 18.7). This is preferable to the usual interpretation in terms of partial equilibrium. The concept of composite systems then becomes superfluous, but the quotation from Callen above lends support to the idea that releasing a (third) constraint is in fact the central element in most thermodynamic problems involving chemical reactions.

G.5 Reiss (1965)

The concept of constraints plays a central role in the text by Reiss (1965), and was very influential in my thinking about thermodynamics. However there is a significant difference in our treatments of metastability.

G.5.1 Constraints

To begin with, Reiss (1965) has the conventional definition of a thermodynamic system (p. 3):

> A thermodynamic system is an arbitrary geometrical portion of the universe with fixed or movable boundaries which may contain matter or energy or both.

Reiss (1965, p. 11) then says (italics in the original),

> Every state of equilibrium is subject to certain constraints which are imposed upon the system. In fact, constraints and variables of constraint are in one-to-one correspondence.

and (p. 13),

> Just as constraints and variables are in one-to-one correspondence with one another, they are each in one-to-one correspondence with the *kinds* of work which a system may perform on its environment or vice versa.

and (p. 14),

> In this connection, it should be pointed out that a system comes to equilibrium *subject* to certain constraints.

From these excerpts it is clear that the concept of constraints in Reiss (1965) and in this text are virtually the same. The only difference is that I emphasize the difference betwen real systems and thermodynamic systems, and that the constraint variables are mathematical in nature, because thermodynamics is a mathematical theory which attempts to simulate nature.

G.5.2 Metastability

Reiss (1965, p. 13) says

> A thermodynamic system is in equilibrium when none of its thermodynamic properties are changing with time at a measurable rate.

and (p. 17),

> The concept of metastability seems to have been employed mainly to indicate the degree of control the experimenter has over the constraint. … However, from the thermodynamic point of view, differences in degree of control are meaningless.

and (p. 18),

> Thus in the interest of clarity it is preferable to describe all equilibrium states which can be achieved by a system as stable, the proper number of variables always being specified.

Thus Reiss prefers to consider all states of equilibrium on an equal basis, as long as the number of constraints is specified. He attaches no special significance to the third constraint, and considers metastability in the conventional sense (very slow reaction rates). However, it seems to me that the concept of metastability, so common in geochemistry, is clarified by defining it in terms of a third constraint, rather than in terms of reaction rates.

G.6 Weinreich (1968)

Constraints are emphasized by Weinreich (1968) in a chapter entitled "Constraints and Spontaneous Processes." His treatment has some similarities to that of Reiss and Callen, but does not completely coincide with that presented in this text. His definition (p. 135) is

> A variable of constraint is any nonthermal variable which is part of the specification of the equilibrium state of a system, but which does not occur in the calculation of work done by the system.

which does not coincide with our definition, and (p. 137)

> A process which results from the release of a constraint is called a *spontaneous* process.

which does.

He also says that metastable states " have a good deal in common" with constrained states, but the only example he gives is of a supercooled vapor phase which "requires no physical constraints, but rather some good luck and/or a sense of humor" to be treated as an equilibrium state. He does not distinguish between real states and thermodynamic states. In addition, his statements about the applicability of the fundamental equation do not take into account all possibilities.

In summary, Weinreich is one of the few who recognize that constraints can clarify thermodynamic concepts, but his usage is not completely in accord with that of this text.

G.7 Summary

The concept of thermodynamics as theory of constrained equilibrium has been clearly stated by several authors, but not used by the majority of other authors.

There is no doubt about the “correctness” of this approach, so this must be because many authors do not consider it useful. I use it in this text because I consider it not only useful, but because it helps to make the distinction between real and thermodynamic systems, which I consider important to the understanding of thermodynamics.

The concept of constraints as used in this book differs from that of Schottky, Callen, and Reiss, in that

- I identify metastable states with the third and higher constraints, and
- I recognize the progress variable ξ as a type of constraint.

References

Akinfiev, N. N. and Diamond, L. W., 2003, Thermodynamic description of aqueous nonelectrolytes at infinite dilution over a wide range of state parameters. *Geochim. et Cosmochim. Acta*, **67**: 613–27.

Alberty, R. A., 2001, Use of Legendre transforms in chemical thermodynamics (IUPAC Technical Report). *Pure Appl. Chem.*, **73**: 1349–80.

Anderko, A., 2000, Cubic and generalized equations of state, *in:* J. V. Sengers, R. F. Kayser, C. J. Peters, and Jr. H. J. White, eds., *Equations of State for Fluids and Fluid Mixtures*, Amsterdam: Elsevier Science B. V., Chapter 4, pp. 75–126.

Anderson, G. M., Castet, S., Mesmer, R. E., and Schott, J., 1991, The density model for estimation of thermodynamic parameters of reactions at high temperatures and pressures. *Geochim. et Cosmochim. Acta*, **55**: 1769–79.

Anderson, G. M., 1995, Is there alkali-aluminum complexing at high temperatures and pressures? *Geochim. et Cosmochim. Acta*, **59**: 2155–61.

1996, *Thermodynamics of Natural Systems.* New York: Wiley.

Anderson, G. M. and Crerar, D. A., 1993, *Thermodynamics in Geochemistry – The Equilibrium Model*. New York: Oxford University Press.

Aranovitch, L. Ya. and Newton, R. C., 1996, H_2O activity in concentrated NaCl solutions at high pressures and temperatures measured by the brucite-periclase equilibrium. *Contrib. Mineral. Petrol.*, **125**: 200–12.

1997, H_2O activity in concentrated KCl and KCl-NaCl solutions at high pressures and temperatures measured by the brucite-periclase equilibrium. *Contrib. Mineral. Petrol.*, **127**: 261–71.

Archer, D. G., 1992, Thermodynamic properties of the $NaCl + H_2O$ system II. Thermodynamic properties of NaCl(aq), $NaCl \cdot 2H_2O$(cr), and phase equilibria. *J. Phys. Chem. Ref. Data*, **21**: 793–829.

Bain, R. W., 1964, *Steam Tables 1964.* National Engineering Lab., Edinburgh: Her Majesty's Stationery Office.

Barrett, T. J., Anderson, G. M., and Lugowski, J., 1988, The solubility of H_2S in NaCl solutions to 95 °C and the fugacity coefficients of several gases in brines to 300 °C. *Geochim. et Cosmochim. Acta*, **52**: 807–11.

Barta, L. and Bradley, D. J., 1985, Extension of the specific interaction model to include gas solubilities in high temperature brines. *Geochim. et Cosmochim. Acta*, **49**: 195–203.

Bassett, R. L. and Melchior, D. C., 1990, Chemical modeling of aqueous systems: an overview, In: D. C. Melchior and R. L. Bassett, eds., Chemical Modeling of Aqueous

Systems. ACS Symposium Series 416. Washington, DC: American Chemical Society, pp. 1–14.

Benson, S. W., 1968, *Thermochemical Kinetics*. New York: Wiley.

Berman, R. G., 1988, Internally consistent thermodynamic data for minerals in the system Na_2O–K_2O–CaO–MgO–FeO–Fe_2O_3–Al_2O_3–SiO_2–TiO_2–H_2O–CO_2. *J. Petrology*, **29**: 445–522.

1990, Mixing properties of Ca-Mg-Fe-Mn garnets. *Amer. Mineralogist*, **75**: 328–44.

Berman, R. G. and Brown, T. H., 1984, A thermodynamic model for multicomponent melts, with application to the system CaO–Al_2O_3–SiO_2. *Geochim. et Cosmochim. Acta*, **48**: 661–78.

1985, Heat capacity of minerals in the system Na_2O–K_2O–CaO–MgO–FeO–Fe_2O_3–Al_2O_3–SiO_2–TiO_2–H_2O–CO_2: representation, estimation, and high temperature extrapolation. *Contrib. Mineral. Petrol.*, **89**: 168–83. Erratum: Contrib. Mineral. Petrol., 1986, **94**, 262.

Berman, R. G. and Koziol, A. M., 1991, Ternary excess properties of grossular-pyrope-almandine garnet and their influence in geothermobarometry. *Amer. Mineralogist*, **76**: 1223–31.

Berndt, A. F. and Diestler, D. J., 1968, On the slope of the liquidus and solidus curves at the melting point of a congruently melting compound. *J. Phys. Chem.*, **72**: 2263–5.

Billo, E. J., 2001, *EXCEL™ for Chemists*, 2nd edition. New York: Wiley-VCH.

Bockris, J. O'M. and Reddy, A. K. N., 1970, *Modern Electrochemistry*, vol. I. New York: Plenum Publishing Co. (paperback edition 1973).

Born, M., 1920, Volumen und Hydrationswärme der ionen. *Z. Phys.* **1**: 45–8.

Bowen, N. L. and Schairer, J. F., 1935, The system MgO–FeO–SiO_2. *Amer. J. Sci.*, **29**: 151–217.

Bridgman, P. W., 1958, *The Physics of High Pressure*. London: G. W. Bell and Sons.

Brimhall, G. H. and Crerar, D. A., 1987, Ore fluids: magmatic to supergene, in: I. S. E. Carmichael and H. P. Eugster, eds., *Thermodynamic Modeling of Geological Materials: Minerals, Fluids and Melts*, Reviews in Mineralogy, 17. Washington DC: Mineralogical Society of America, pp. 235–322.

Brown, T. H., Berman, R. G., and Perkins, E. H., 1988, Ge0-Calc: Software package for calculation and display of pressure-temperature-composition phase diagrams using an IBM or compatible personal computer. *Computers and Geosciences*, **14**: 279–89.

Burnham, C. W., Holloway, J. R., and Davis, N. F., 1969a, The specific volume of water in the range 1000 to 8900 bars, 20° to 900 °C. *Amer. J. Sci.*, Schairer vol. **267-A**: 70–95.

1969b, Thermodynamic Properties of Water to 1000 °C and 10,000 Bars. Special Paper 132, Boulder, CO: Geological Society of America.

Busey, R. H., Holmes, H. F., and Mesmer, R. E., 1984, The enthalpy of dilution of aqueous sodium chloride to 673 K using a new heat-flow and liquid-flow microcalorimeter. Excess thermodynamic parameters and their pressure coefficients. *J. Chem. Thermodynamics*, **16**: 343–72.

Butler, J. N., 1964, *Ionic Equilibrium, a Mathematical Approach*. Reading, MA: Addison-Wesley.

Callen, H. B., 1960, *Thermodynamics*. New York: Wiley.

Carmichael, D. M., 1986, Induced stress and secondary mass transfer: thermodynamic basis for the tendency toward constant volume constraint in diffusion metsomatism, in: *H.C. Chemical Transport in Metasomatic Processes*. H. C. Helgeson, ed., NATO ASI Series C, 218. Dordrecht: Reidel, pp. 239–64.

Carmichael, I. S. E., Turner, F. J., and Verhoogen, J., 1974, *Igneous Petrology*. New York: McGraw-Hill.

Chase, M. W. Jr., 1998, *NIST-JANAF Thermochemical Tables*, 4th edition. J. Phys. Chem. Ref. Data, Monograph 9, Parts I and II, 1951. Washington DC: National Bureau of Standards.

Chou, I. M., Seal, R. R., and Hemingway, B. S., 2002, Determination of melanterite-rozenite and calcanthite-bonattite equilibria by humidity measurements at 0.1 MPa. *Amer. Mineralogist*, **87**: 108–14.

Cohen, E. R. and Taylor, B. N., 1988, The 1986 CODATA recommended values of the fundamenhal physical constants. *J. Phys. Chem. Ref. Data*, **17**: 1795–1803.

Cooper, J. R., 1982, Representation of the ideal-gas thermodynamic properties of water. *Int. Jr. Thermophys.*, **3**: 35–43.

Crawford, W. A. and Hoersh, A. L., 1972, Calcite-aragonite equilibrium from 50 to 150 °C. *Amer. Mineralogist*, **57**: 995–8.

Crerar, D. A., 1975, A method for computing multicomponent chemical equilibria based on equilibrium constants. *Geochim. et Cosmochim. Acta.*, **39**: 1375–84.

Criss, C. M. and Cobble, J. W., 1961, The thermodynamic properties of high temperature aqueous solutions. I. Standard partial molar heat capacities of sodium chloride and barium chloride from 0 to 100 °C. *Jr. Amer. Chem. Soc.*, **83**: 3223–8.

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

Debye, P. J., 1954, *The Collected Papers of P. J. Debye.* New York: Interscience.

DeCapitani, C., and Brown, T. H., 1987, The computation of chemical equilibrium in complex systems containing nonideal solutions. *Geochim. et Cosmochim. Acta*, **51**: 2639–52.

De Donder, Th. (with F. H. van den Dungen and G. van Lerberghe), 1920, *Leçons de thermodynamique et de chimie-physique*. Paris: Gauthiers-Villars.

de Heer, J., 1986, *Phenomenological Thermodynamics*. Englewood Cliffs, NJ: Prentice-Hall.

Denbigh, K., 1966, *The Principles of Chemical Equilibrium.* Cambridge: Cambridge University Press.

Dickerson, R. E., 1969, *Molecular Thermodynamics*. New York: W. A. Benjamin.

Douglas, T. B. and King. E. G., 1968, High-temperature drop calorimetry, in: J. P. McCullough and D. W. Scott, eds., *Experimental Thermodynamics*, vol. 1. London: Butterworths, pp. 293–332.

Drever, J. I., 1988, *The Geochemistry of Natural Waters*, 2nd edition. Englewood Cliffs, NJ: Prentice-Hall.

Driesner, T., Seward, T. M., and Tironi, I. G., 1998, Molecular dynamics simulation study of ionic hydration and ion association in dilute and 1 molal solutions from ambient to supercritical conditions. *Geochim. et Cosmochim. Acta*, **62**: 3095–107.

Duan, Z., M∅ller, N., and Weare, J. H., 1992a, An equation of state for the CH_4–CO_2–H_2O system: I. Pure systems from 0 to 1000 °C and 0 to 8000 bar. *Geochim. et Cosmochim. Acta*, **56**: 2605–17.

1992b, An equation of state for the CH_4–CO_2–H_2O system: II. Mixtures from 50 to 1000 °C and 0 to 1000 bar. *Geochim. et Cosmochim. Acta*, **56**: 2619–31.

1996a, Prediction of the solubility of H_2S in NaCl aqueous solution: an equation of state approach. *Chem. Geol.*, **130**: 15–20.

1996b, A general equation of state for supercritical fluid mixtures and molecular dynamics simulation of mixture *PVTX* properties. *Geochim. et Cosmochim. Acta,* **60**: 1209–16.

2000, Accurate prediction of the thermodynamic properties of fluids in the system H_2O–CO_2–CH_4–N_2 up to 2000 K and 100 kbar from a corresponding states/one fluid equation of state. *Geochim. et Cosmochim. Acta*, **64**: 1069–75.

Dymond, J. C. and Smith, E. B., 1980, *The Second Virial Coefficients of Gases: A Critical Compilation.* Oxford: Clarendon Press.

Ely, J. F. and Marrucho, I. M. F., 2000, The corresponding states principle. Chapter 8, pp. 289–320, *in*: J. V. Sengers, R. F. Kayser, C. J. Peters, and H. J. Jr. White, eds., Equations of State for Fluids and Fluid Mixtures. Elsevier Science B. V., Amsterdam.

Eugster, H., Albee, A., Bence, A., Thompson, J. B., Jr., and Waldbaum, D., 1972, The two-phase region and excess mixing properties of paragonite-muscovite crystalline solutions: *J. Petrol*, **13**: 147–79.

Eugster, H. P., Harvie, C. E., and Weare, J. H., 1980, Mineral equilibria in a 6–7 component seawater system, Na–K–Mg–Ca–SO_4–Cl–H_2O, at 25 °C. *Geochim. et Cosmochim. Acta*, **44**: 1335–47.

Ewing, M. B. and Peters, C. J., 2000, Fundamental considerations. Chapter 1, pp. 5–34, *in*: J. V. Sengers, R. F. Kayser, C. J. Peters, and H. J. Jr. White, eds., Equations of State for Fluids and Fluid Mixtures. Part I, pp. 1–434; Part II, pp. 435–885. Elsevier Science B. V., Amsterdam.

Franck, E. U., 1956, Hochverdichteter Wasserampf II. lonendissociation von KCl in H_2O bis 750 °C. *Z. Phys. Chem.*, **8**: 107–26.

1961, Überkritisches Wasser als electrolytisches Lösungsmittel. *Angew. Chem.*, **73**: 309–22.

Feynman, R. P., Leighton, R. B., and Sands, M., 1963, *The Feynman Lectures on Physics*, vol. 1. New York: Addison Wesley, Chapter 4, p. q.

Friedman, H. L., 1962, *Ionic Solution Theory*. New York: Interscience.

Ganguly, J., 1982, Mg-Fe order-disorder in ferromagnesian silicates. II. Thermodynamics, kinetics, and geological applications, in: S.K. Saxena, ed., *Advances in Geochemistry*, vol. 2, New York: Springer-Verlag, pp. 58–99.

Gardner, W. L., Mitchell, R. E., and Cobble, J. W., 1969, The thermodynamic properties of high-temperature aqueous solutions. XI. Calorimetric determination of the standard partial molar heat capacity and entropy of sodium chloride solutions from 100 to 200 °C. *J. Phys. Chem.*, **73**: 2025–32.

Garrels, R. M. and Christ, C. L., 1965, *Solutions, Minerals, and Equilibria.* New York: Harper and Row.

Garrels, R. M. and Thompson, M. E., 1962, A chemical model for seawater at 25 °C and one atmosphere total pressure. *Amer. J. Sci.*, **260**: 57–66.

Gates, J. A., Wood, R. H., and Quint, J. R., 1982, Experimental evidence for the remarkable behavior of the partial heat capacity at infinite dilution of aqueous electrolytes at the critical point. *J. Phys. Chem.*, **86**: 4948–51.

Greenspan, L., 1976, Humidity fixed points of binary saturated aqueous solutions. *J.Res. of National Bureau of Standards – A. Physics and Chemistry*, **81A**: 89–96.

Gibbs, J. W., 1875, On the equilibrium of heterogeneous substances. Reprinted in: *The Scientific Papers of J. Willard Gibbs*, vol.1: *Thermodynamics*. New York: Dover Publications Inc., 1961, pp. 55–349.

Grover, J., 1977, Chemical mixing in multicomponent systems: an introduction to the use of Margules and other thermodynamic excess functions to represent non-ideal behavior, in: D. G. Fraser, ed., *Thermodynamics in Geology*. Boston: Reidel, pp. 67–97.

Gu, Y., Gammons, C. H., and Bloom, M. S., 1994, A one-term extrapolation method for estimating equilibrium constants of aqueous reactions at elevated temperatures. *Geochim. et Cosmochim. Acta*, **58**: 3545–60.

Guggenheim, E. A., 1935, The specific thermodynamic properties of aqueous solutions of strong electrolytes. *Phil. Mag*., **19**: 588–643.

1952, *Mixtures.* Oxford: Clarendon Press.

Haar, L., Gallagher, J. S., and Kell, G. S., 1984, *NBS/NRC Steam Tables*. Washington, DC: Hemisphere Publishing Corp.

Hahn, W. C. and Muan, A., 1962, Activity measurements in oxide solid solutions. The system "FeO"-MgO in the temperature interval 1100° to 1300 °C. *Trans. Metallurgical Society of AIME*, **224**: 416–20.

Harned, H. S. and Owen, B. B., 1958, *The Physical Chemistry of Electrolyte Solutions*, 3rd edition. New York: Reinhold.

Harvey, A. H., Peskin, A. P., and Klein, S. A, 2000, *NIST/ASME Steam Properties*. NIST Standard Reference Database 10, Version 2.2 (National Institute of Standards and Technology, Gaithersburg, MD, 2000). Available from the NIST website, http://www.nist.gov/srd/nist10.

Harvie, C. E., 1981, Theoretical investigations in geochemistry and atom surface scattering. Unpublished Ph.D. Thesis, University of California, San Diego.

Harvie, C. E. and Weare, J. H., 1980, The prediction of mineral solubilities in natural waters: the Na–K–Mg–Ca–Cl–SO_4–H_2O system from zero to high concentration at 25 °C. *Geochim. et Cosmochim. Acta*, **44**: 981–7.

Harvie, C. E., Eugster, H. P., and Weare, J. H., 1982, Mineral equilibria in the six component seawater systems, Na–K–Mg–Ca–SO_4–Cl–H_2O at 25 °C. II. Compositions of the saturated solutions. *Geochim. et Cosmochim. Acta*, **46**: 1603–18.

Harvie, C. E., MØller, N., and Weare, J. H., 1984, The prediction of mineral solubilities in natural waters: the Na–K–Mg–Ca–H–Cl–SO_4–OH-HCO_3–CO_3–CO_2–H_2O system to high ionic strengths at 25 °C. *Geochim. et Cosmochim. Acta.*, **48**: 723–51.

Haselton, H. T. and Newton, R. C., 1980, Thermodynamics of pyrope-grossular garnets and their stabilities at high temperatures and pressures. *J. Geophys. Res*., **85**: 6973–82.

Haselton, H. T. and Westrum, E. F., 1980, Low-temperature heat capacities of synthetic pyrope, grossular, and pyrope_{60} grossular_{40}. *Geochim. et Cosmochim. Acta*, **44**: 701–9.

Helgeson, H. C., 1969, Thermodynamics of hydrothermal systems at elevated temperatures and pressures. *Amer. J. Sci.*, **267**: 729–804.

1979, Mass transfer among minerals and hydrothermal solutions, in: H. L. Barnes, ed., *Geochemistry of Hydrothermal Ore Deposits*, 2nd edition. New York: Wiley-Interscience, pp. 568–610.

1981, Prediction of the thermodynamic properties of electrolytes at high pressures and temperatures, in: D. T. Rickard and F. E. Wickman, eds., *Chemistry and Geochemistry*

of Solutions at High Temperatures and Pressures. Proceedings of a Nobel Symposium. Physic and Chemistry of the Earth, 13 and 14. Oxford: Pergamon Press, pp. 133–77.
1991, Organic/inorganic reactions in metamorphic processes. *Can. Mineralogist*, **29**: 707–39.

Helgeson, H. C. and Kirkham, D. H., 1974a, Theoretical prediction of the thermodynamic behavior of aqueous electrolytes at high temperatures and pressures: I. Summary of the thermodynamic/electrostatic properties of the solvent. *Amer. Jr. Sci.*, **274**: 1089–198.
1974b, Theoretical prediction of the thermodynamic behavior of aqueous electrolytes at high pressures and temperatures: II. Debye–Hückel parameters for activity. *Amer. J. Sci.*, **274**: 1199–261.
1976, Theoretical prediction of the thermodynamic properties of aqueous electrolytes at high pressures and temperatures: III. Equation of state for aqueous species at infinite dilution. *Amer. J. Sci.*, **276**: 97–240.

Helgeson, H. C., Delany, J. M., Nesbitt, H. W., and Bird, D. K., 1978, Summary and critique of the thermodynamic properties of rock-forming minerals. *Amer. J. Sci.*, **278A**: 1–229.

Helgeson, H. C., Kirkham, D. H., and Flowers, G. C., 1981, Theoretical prediction of the thermodynamic properties of aqueous electrolytes at high pressures and temperatures: IV. Calculation of activity coefficients, osmotic coefficients, and apparent molal and standard and relative partial molal properties to 600 °C and 5 kb. *Amer. J. Sci.*, **281**: 1249–493.

Hemingway, B. S. and Robie, R. A., 1977, Enthalpies of formation of low albite, $NaAlSi_3O_8$, gibbsite, $(Al(OH)_3)$, and $NaAlO_2$; revised values for $H^o_{f,298}$ and $G^o_{f,298}$ of some aluminosilicate minerals. *US Geological Survey J. Res.*, **5**: 413.

Hemingway, B. S., Haas, J. L. Jr., and Robinson, G. R. Jr., 1982, *Thermodynamic Properties of Selected Minerals in the System* Al_2O_3*–CaO–*SiO_2*–*H_2O *at 298.15 K and 1 bar (*10^5 *Pascals) Pressure and at Higher Temperatures*. US Geological Survey Bulletin 1544. Washington, DC: Government Printing Office.

Henley, R. W., Truesdell, A. H., Barton, Jr., P. B., and Whitney, J. A., 1984, *Reviews in Economic Geology*, vol. 1: *Fluid–Mineral Equilibria in Hydrothermal Systems*. El Paso, TX: Economic Geology Publishing Co.

Hensen, B. J., Schmid, R., and Wood, B. J., 1975, Activity-composition relationships for pyrope-grossular garnet. *Contrib. Mineral. Petrol.*, **51**: 161–6.

Hildebrand, J. H. and Scott, R. L., 1964, *The Solubility of Nonelectrolytes.* New York: Dover Publications. A reprint of the third edition of the original text, published in 1950.

Hill, P. G., 1990, A unified fundamental equation for the thermodynamic properties of H_2O. *J. Phys. Chem. Ref. Data*, **19**: 1233–74.

Höhne, G. W. H., Hemminger, W., and Flammersheim, H.-J., 1996, *Differential Scanning Calorimetry: An Introduction for Practitioners*. Berlin: Springer-Verlag.

Holloway, J., 1977, Fugacity and activity of molecular species in supercritical fluids, in: D. G. Fraser, ed., *Thermodynamics in Geology*. Dordrecht: Reidel, pp. 161–81.

Jaynes, E. T., 1988, The evolution of Carnot's principle, in: *Maximum Entropy and Bayesian Methods in Science and Engineering*, vol. 1: Foundations. G. J. Erickson and C. R. Smith, eds., Dordrecht: Kluwer Academic Publishers.
1991, The Gibbs paradox, in: C. R. Smith et al., eds., *Maximum Entropy and Bayesian Methods*, Seattle: Kluwer Academic Publishers, pp. 1–21.

Jerz, J. K. and Rimstidt, J. D., 2003, Efflorescent iron sulfate minerals: paragenesis, relative stability, and environmental impact. *Amer. Mineralogist*, **88**: 1919–32.

Johnson, J. W. and Norton, D., 1991, Critical phenomena in hydrothermal systems: state, thermodynamic, electrostatic, and transport properties of H_2O in the critical region. *Amer. J. Sci.* **291**: 541–648.

Johnson, J. W., Oelkers, E. H., and Helgeson, H. C., 1992, SUPCRT92, a software package for calculating the standard molal thermodynamic properties of minerals, gases, aqueous species as functions of temperature and pressure. *Computers Geosci.*, **18**: 899–947.

Kell, G. S., McLaurin, G. E., and Whalley, E., 1968, PVT properties of water. II. Virial coefficients in the range 150°–450 °C without independent measurement of vapor volumes. *J. Chem. Phys.*, **48**: 3805–13.

Kelley, K. K., 1960, *Contributions to the Data on Theoretical Metallurgy*, XIII: *High-Temperature Heat-Content, Heat Capacity, and Entropy Data for the Elements and Inorganic Compounds.* US Bureau of Mines Bulletin 584. Washington, DC: Government Printing Office.

Kerrick, D. and Jacobs, G., 1981, A modified Redlich–Kwong equation for H_2O, CO_2, and H_2O–CO_2 mixtures at elevated pressures and temperatures. *Amer. J. Sci.*, **281**: 735–67.

Kharaka, Y. K., Gunter, W. D., Aggarwal, P. K., Perkins, E. H., and DeBraal, J. D., 1988, SOLMINEQ88: a computer program for geochemical modeling of water–rock interactions. Menlo Park, CA: US Geological Survey, Water Resource Investigations, Report 88–4227.

Kielland, J., 1937, Individual activity coefficients of ions in aqueous solutions. *J. Amer. Chem. Soc.*, **59**: 1675–8.

Kivelson, D., and Oppenheim, I., 1966, Work in irreversible expansions. *J. Chem. Edu.* **43**: 233–5.

Klotz, I. M., 1964, *Chemical Thermodynamics*, revised edition. New York: W.A. Benjamin, Inc.

Knapp, R. A., 1989,. Spatial and temporal scales of local equilibrium in dynamic fluid-rock systems. *Geochim. et Cosmochim. Acta*, **53**: 1955–64.

Korzhinskii, D. S., 1959, *Physicochemical Basis of the Paragenesis of Minerals*. New York: Consultant's Bureau, Inc.

Kress, V. C., 2003, On the thermodynamics of associated solutions. *Amer. J. Sci.*, **303**: 708–22.

Krupka, K. M., Robie, R. A., and Hemingway, B. S., 1979, High-temperature heat capacities of corundum, periclase, aanorthite, $CaAl_2Si_2O_8$ glass, muscovite, pyrophyllite, $KAlSi_3O_8$ glass, grassular, and $NaAlSi_3O_8$ glass. *Amer. Mineralogist*, **64**: 86–101.

Lambert, F. L., 1999, Shuffled cards, messy desks, and disorderly dorm rooms – examples of entropy increase? Nonsense! *J. Chem. Edu.* **76**: 1385–7.

Lasaga, A. C. and Kirkpatrick, R. J., 1981, *Kinetics of Geochemical Processes. Reviews in Mineralogy*, vol. 8. Washington, DC: Mineralogical Society of America.

Leatherdale, W. H., 1974, *The Role of Analogy, Model and Metaphor in Science.* Amsterdam: North-Holland Publishing Co.

Levelt Sengers, J. M. H., Kamgar-Parsi, B., Balfour, F. W., and Sengers, J. V., 1983, Thermodynamic properties of steam in the critical region. *J. Phys. Chem. Data*, **12**(**1**): 1–28.

Lewis, G. N. and Randell, M., 1914, The free energy of the various forms of elementary sulfur. *J. Amer. Chem. Soc.*, **36**: 2468–75.

1923, *Thermodynamics and the Free Energy of Chemical Substances*, 1st edition. New York: McGraw-Hill.

1961, *Thermodynamics*, 2nd edition. Revised by K. S. Pitzer and L. Brewer. New York: McGraw-Hill.

Lindberg, R. D. and Runnells, D. D., 1984, Groundwater redox reactions: an analysis of equilibrium state applied to *Eh* measurements and geochemical modeling. *Science*, **225**: 925–7.

Lindsay, W. T., Jr., 1980, Estimation of concentration quotients for ionic equilibria in high temperature water. *41st. International Water Conference*, Pittsburgh, PA, pp. 284–94.

Liu, C. T. and Lindsay, W. T., 1972, Thermodynamics of sodium chloride solutions at high temperatures. *J. Sol. Chem.*, **1**: 45–69.

Long, F. A. and McDevit, W. F., 1952, Activity coefficients of nonelectrolytes in aqueous electrolyte solutions. *Chem. Rev.*, **51**: 119–69.

Lupis, C. H. P., 1983, *Chemical Thermodynamics of Materials*. New York: Elsevier Publishing Co.

MacDonald, J. J., 1990, Equilibrium, Gibbs energy, and entropy. Rates and differences. *J. Chem. Edu.*, **67**: 380–2.

McMillan, W. G., Jr. and Mayer, J. E., 1945, The statistical thermodynamics of multicomponent systems. *J. Chem. Phys.*, **13**: 276–305.

Maier, C. G. and Kelly, K. K., 1932, An equation for the representation of high temperature heat content data. *Amer. Chem. Soc. J.*, **54**: 3243–6.

Majer,V., Hui, L., Crovetto, R., and Wood, R. H., 1991, Volumetric properties of aqueous 1–1 electrolyte solutions near and above the critical temperature of water. 1. Densities and apparent molar volumes of NaCl(aq) from 0.0025 mol kg^{-1} to 3.1 mol kg^{-1}, 604.4 K to 725.5 K, and 18.5 MPa to 38.0 MPa. *J. Chem. Thermodynamics*, **23**: 213–29.

Margules, M., 1895, Über die Zusammensetzung der gestättigten Dämpfe von Michungen. Sitzungsberichte der Math.-Naturwiss. Classe Akademie der Wissenschaften, Wien, v. 104, Abtheilung IIa, pp. 1243–78.

Marshall, W. L. and Franck, E. U., 1981, Ion product of water substance, 0–1000ºC, 1–10,000 bars, new international formulation and its background. *J. Phys. Chem. Ref. Data*, **10**: 295–304.

Marshall, W. L. and Slusher, R., 1966, Thermodynamics of calcium sulfate dihydrate in aqueous sodium chloride solutions, 0–110°. *J. Phys. Chem.*, **70**: 4015–27.

Mayer, J. E., 1950, The theory of ionic solutions. *J. Chem. Phys.*, **18**: 1426–36.

Mayer, J. E. and Mayer, M., 1940, *Statistical Mechanics*. New York: Wiley.

Mazo, R. M. and Mou, C. Y., 1991, Introduction to the statistical mechanics of solutions, in: K. S. Pitzer, ed., *Activity Coefficients in Electrolyte Solutions*, 2nd edition. Boca Raton, FL: CRC Press, pp. 29–73.

McGlashan, M. L., 1979, *Chemical Thermodynamics*. London: Academic Press.

Merino, E., 1975, Diagenesis in Tertiary sandstones from Kettleman North Dome, California – II. Interstitial solutions: distribution of aqueous species at 100 °C and chemical relation to the diagenetic mineralogy. *Geochim. et Cosmochim. Acta*, **39**: 1629–45.

McQuarrie, D. A., 2000, *Statistical Mechanics*. Sausalito, CA: University Science Books.

Mesmer, R. E., 1985, A model for estimation of thermodynamic quantities for reactions – uncertainties from such predictions. Presented at the Symposium on Hydrothermal Reactions, Pennsylvania State University, State College, PA.

1986, Aqueous chemistry and thermodynamics to high temperatures and pressures. Presented at the 41st Annual Calorimetry Conference, Somerset, NJ.

Mesmer, R. E. and Baes, C. F., Jr., 1974, Phosphoric acid dissociation equilibria in aqueous solutions to 300 °C. *J. Sol. Chem.*, **3**: 307–21.

Mesmer, R. E., Marshall, W. L., Palmer, D. A., Simonson, J. M., and Holmes, H. F., 1987, Chemical equilibria and thermodynamics of reactions at high temperatures and pressures. Presented at the International Conference on Thermodynamics of Aqueous Systems with Industrial Applications, 10–14 May, Airlie House, Warenton, VA.

1988, Thermodynamics of aqueous association and ionization constants at high temperatures and pressures. *J. Sol. Chem.*, **17**: 699–718.

Mesmer, R. E., Patterson, C. S., Busey, R. H., and Holmes, H. F., 1989, Ionization of acetic acid in NaCl(aq) media: a potentiometric study to 573 K and 130 bar. *J. Phys. Chem.*, **93**: 7483–90.

Michels, A., Blaisse, B., and Michels, C., 1935, The isotherms of CO_2 in the neighbourhood of the critical point and round the coexistence line. *Proc. Roy. Soc. London*, **A160**: 358–75.

Millero, F. J., 1971, The molal volumes of electrolytes. *Chem. Rev.*, **71**: 147–76.

1996, *Chemical Oceanography*, 2nd edition. Boca Raton, FL: CRC Press.

Mills, I., Civitas, C., Homann, K., Kallay, N., and Kuchitsu, K., eds., 1993, *Quantities, Units and Symbols in Physical Chemistry* (the IUPAC Green Book), 2nd edition. Oxford: Blackwell Science.

Monk, C. B., 1961, *Electrolytic Dissociation*. London: Academic Press.

Morel, F. M. M. and Hering, J. G., 1993. *Principles and Applications of Aquatic Chemistry*. New York: Wiley.

Morey, G. W., Fournier, R. O., and Rowe, J. J., 1962, The solubility of quartz in the temperature interval from 25° to 300 °C. *Geochim. et Cosmochim. Acta*, **26**: 1029–43.

Nafziger, R. H. and Muan, A., 1967, Equilibrium phase compositions and thermodynamic properties of olivines and pyroxenes in the system MgO–FeO–SiO_2. *Amer. Mineralogist*, **52**: 1364–85.

Nahon, D. and Merino, E., 1997, Pseudomorphic replacement in tropical weathering: evidence, geochemical consequences, and kinetic-rheological origin. *Amer. J. Sci.*, **297**: 393–417.

Nash, L. K., 1972, *Elements of Statistical Thermodynamics*. Reading, MA: Addison-Wesley.

Navrotsky, A., 1971, The intracrystalline cation distribution and the thermodynamics of solid solution formation in the system $FeSiO_3$–$MgSiO_3$. *Amer. Mineralogist*, **56**: 201–11.

Nesbitt, H. W., 1982, The Stokes and Robinson hydration theory: A modification with application to concentrated electrolyte solutions. *J. Sol. Chem.*, **11**: 415–22.

1984, Calculation of the solubility of CO_2 in NaCl-rich hydrothermal solutions using regular solution equations. *Chem. Geol.*, **43**: 319–30.

Newton, R. C. and Haselton, H. T., 1981, Thermodynamics of the garnet-plagioclase-Al_2SiO_5-quartz geobarometer, in: R. C. Newton, A. Navrotsky, and B. J. Wood, eds., *Thermodynamics of Minerals and Melts*, New York: Springer-Verlag, pp. 131–47.

Newton, R. C., Charlu, T. V., and Kleppa, O. J., 1977, Thermochemistry of high pressure garnets and clinopyroxenes in the system CaO–MgO–Al_2O_3–SiO_2. *Geochim. et Cosmochim. Acta*, **41**: 369–77.

Nicholson, R. V., 1994, Iron sulfide oxidation mechanisms: laboratory studies, in: J. L. Jambor and D. W. Blowes, eds., *Short Course Handbook on Environmental Geochemistry of Sulfidic Mine-Wastes*. Nepean, Ontario: Mineralogical Association of Canada, Short Course Handbook, vol. **22**, pp. 163–83.

Nordstrom, D. K. and Alpers, C. N., 1999, Geochemistry of acid mine waters. in: G. S. Plumlee and M. J. Logsdon, eds., *The Environmental Geochemistry of Mineral Deposits*. Reviews in Economic Geology: vol. 6A. Littleton, CO: Society of Economic Geologists, pp. 133–160.

Nordstrom, D. K. and Ball, J. W., 1984, Chemical models, computer programs and metal complexation in natural waters, in: C. J. M. Kramer and J. C. Duinker, eds., *Complexation of Trace Metals in Natural Waters*. The Hague: Nijhoff and Junk, pp. 149–64.

Nordstrom, D. K. and Munoz, J. L., 1994, *Geochemical Thermodynamics*, 2nd edition. Boston: Blackwell Scientific Publications.

Nordstrom, D. K., Plummer, L. N., Wigley, T. M. L., Wolery, T. J., and Ball, J. W., 1979, A comparison of computerized chemical models for equilibrium calculations in aqueous systems, in: E. A. Jenne, ed., *Chemical Modelling in Aqueous Solutions*. Symposium Series 93. American Chemical Society, pp. 857–92.

Oelkers, E. H. and Helgeson, H. C., 1991, Calculation of activity coefficients and degrees of formation of neutral ion pairs in supercritical electrolyte solutions. *Geochim. et Cosmochim. Acta*, **55**: 1235–51.

Orbey, H. and Sandler, S. I., 1998, *Modeling Vapor–Liquid Equilibria. Cubic Equations of State and their Mixing Rules*. Cambridge: Cambridge University Press.

Palmer, D. A. and Drummond, S. E., 1986. Thermal decarboxylation of acetate; Part 1, The kinetics and mechanism of reaction in aqueous solution. *Geochim. et Cosmochim. Acta*, **50**: 813–23.

Pankratz, L. B., 1964, *High-Temperature Heat Contents and Entropies of Muscovite and Dehydrated Muscovite*. US. Bureau of Mines, Report of Investigations 6371. Washington DC: Government Printing Office.

Parry, W. T., Bellows, J. C., Gallagher, J. S., and Harvey, A. H., 2000, *ASME International Steam Tables for Industrial Use*. New York: ASME Press.

Peng, D.-Y. and Robinson, D. B., A new two-constant equation of state. *Ind. Eng. Chem. Fundamen.*, **15**: 59–64.

Pitzer, K. S., 1955, The volumetric and thermodynamic properties of fluids. I. Theoretical basis and virial coefficients. *J. Amer. Chem. Soc.*, **77**: 3427–43.

1973, Thermodynamics of electrolytes. I. Theoretical basis and general equations. *J. Phys. Chem.*, **77**: 268–77.

1975, Thermodynamics of electrolytes. V. Effects of higher-order electrostatic terms. *J. sol. Chem.*, **4**: 249–65.

1979, Theory: ion interaction approach, in: R. M. Pytkowicz, ed., *Activity Coefficients in Electrolyte Solutions*, vol. 1. Boca Raton, FL: CRC Press, pp. 157–208.

1981, Characteristics of very concentrated solutions, in: F. Wickmann and D. Rickard, eds., *Proceedings, Nobel Symposium on the Chemistry and Geochemistry of Solutions at High Temperatures and Pressures* (Stockholm, Royal Swedish Acad. Sci.), vols. 13/14 of *Physic and Chemistry of the Earth*. New York: Pergamon Press.

1983, Thermodynamics of unsymmetrical electrolyte mixtures. Enthalpy and heat capacity. *J. Phys. Chem.*, **87**: 2360–4.

1987, A thermodynamic model for aqueous solutions of liquid-like density, in: I. S. E. Carmichael and H. P. Eugster, eds., *Thermodynamic Modeling of Geological Materials: Minerals, Fluids and Melts*. Reviews in Mineralogy, Washington DC: vol. 17. Mineralogical Society America, pp. 97–142.

1991, Ion interaction approach: theory and data correlation, in: K. S. Pitzer, ed., *Activity Coefficients in Electrolyte Solutions*, 2nd edition. Boca Raton, FL: CRC Press, pp. 75–153.

1995, *Thermodynamics*, 3rd edition. New York: McGraw-Hill.

Pitzer, K. S. and Brewer, L., 1961, Revised edition of *Thermodynamics* by G. N. Lewis and M. Randall. 2nd edition: New York: McGraw-Hill.

Pitzer, K. S. and Peiper, 1984, Thermodynamic properties of aqueous sodium chloride solutions. *J. Phys. Chem. Ref. Data*, **13**: 1–102.

Pitzer, K. S., Peiper, J. C., and Busey, R. H., 1984, Thermodynamics properties of aqueous sodium chloride solutions. *J. Phys. Chem. Ref. Data*, **13**: 1–102. Published by the American Chemical Society and the American Institute of Physics for the National Bureau of Standards.

Pitzer, K. S., Lippman, D. Z., Curl, Jr., R. F., Huggins, C. M. and Petersen, D. E., 1955, The volumetric and thermodynamic properties of fluids. II. Compressibility factor, vapor pressure, and entropy of vaporization. *J. Amer. Chem. Soc.*, **77**: 3433–40.

Poincaré, H., 1952, *Science and Hypothesis*. New York: Dover Publications. Republication of the first English translation published in 1905.

Powell, R., 1978, *Equilibrium Thermodynamics in Petrology*. New York: Harper and Row.

Prausnitz, M., Lichtentaler, R. N., and de Azevedo, E. G., 1999, *Molecular Thermodynamics of Fluid-Phase Equilibria*, 3rd edition. Upper Saddle River, NJ: Prentice Hall PTR.

Prigogine, I. and Defay, R., 1954, *Chemical Thermodynamics*. London: Longman Green.

Prue, J. E., 1969, Ion pairs and complexes: Free energies, enthalpies, and entropies. *J. Chem. Edu.*, **46**: 12–16.

Purrington, R. D., 1997, *Physics in the Nineteenth Century*. Rutgers, NJ: Rutgers University Press.

Pytkowicz, R. M., Johnson, K., and Curtis, C., 1977, Long-range order model of aqueous electrolyte solutions. *Geochem J.* **11**: 1–11.

Ramberg, H., 1971, Temperature changes associated with adiabatic decompression in geological processes. *Nature*, **234**: 539–40.

Randell, M. and Failey, C. F., 1927a, The activity coefficients of gasses in aqueous salt solutions. *Chem. Rev.*, **4**: 271–84.

Rard, J. A. and Platford, R. F., 1991, Experimental methods: isopiestic, in: K. S. Pitzer, ed., *Activity Coefficients in Electrolyte Solutions*, 2nd edition. Boca Raton, FL: CRC Press, pp. 209–77.

Reardon, E. J., 1989, Ion interaction model applied to equilibria in the $NiSO_4$–H_2SO_4–H_2O system. *J. Phys. Chem.*, **93**: 4630–6.

Redlich, O., and Kwong, J., 1949, The thermodynamics of solutions. V. An equation of state. Fugacities of gaseous solutions. *Chem. Rev.*, **44**: 233–44.

Reed, M. H., 1982. Calculation of multicomponent chemical equilibria and reaction processes in systems involving minerals, gases and an aqueous phase. *Geochim. et Cosmochim. Acta*, **46**: 513–28.

1998, Calculation of simultaneous chemical equilibria in aqueous-mineral-gas systems and its application to modeling hydrothermal processes, in: J. P. Richards and P. B. Larson, eds., *Techniques in Hydrothermal Ore Deposits Geology*. Reviews in Economic Geology, vol. 10. El Paso, TX: Economic Geology Publishing Co., pp. 109–24.

Reiss, H., 1965, *Methods of Thermodynamics*. New York: Blaisdell.

Robie, R. A., 1987, Calorimetry, in: G. C. Ulmer and H. L. Barnes, eds., *Hydrothermal Experimental Techniques*. New York: Wiley, pp. 389–422.

Robie, R. A. and Hemingway, B. S., 1972, *Calorimeters for Heat of Solution and Low-Temperature Heat Capacity Measurements*. Washington DC: Government Printing Office, US Geological Survey Prof. Paper 755.

Robie, R. A., Hemingway, B. S., and Fisher, J. R., 1978, *Thermodynamic Properties of Minerals and Related Substances at 298.15 K and 1 bar (*10^5 *Pascals) Pressure and at Higher Temperatures*. US Geological Survey Bulletin, 1452. Washington, DC: London.

Robinson, R. A. and Stokes, R. H., 1959, *Electolyte Solutions*, 2nd edition. Butterworths.

Rumble, D., 1982, The role of perfectly mobile components in metamorphism. *Ann. Rev. Earth Planet. Sci.*, **10**: 221–33.

Saxena, S. K., 1973, *Thermodynamics of Rock-forming Crystalline Solutions*. New York: Springer-Verlag.

Schottky, W., 1929, Thermodynamik, in *Gemeinschaft,* ed. H. Ulich und C. Wagner. Berlin: Verlag von Julius Springer.

Sengers, J. V., Kayser, R. F., Peters, C. J., and White, H. J., Jr., eds., 2000, *Equations of State for Fluids and Fluid Mixtures*. Amsterdam: Elsevier Science B.V.

Sevigny, J. H. and Ghent, E. D., 1989, Pressure, temperature and fluid composition during amphibolite facies metamorphism of graphitic metapelites, Howard Ridge. *British Columbia J. Metamorphic Petrol.*, **7**: 499–505.

Shannon, C., 1949, *The Mathematical Theory of Communication*. Urbana, IIL: University of Illinois Press.

Singer, P. C. and Stumm, W., 1968, Acidic mine drainage: The rate-determining step. *Science*, **167**: 1121–3.

Smith, W. R. and Missen, R. W., 1982, *Chemical Reaction Equilibrium Analysis: Theory and Algorithms*. New York: Wiley.

Span, R., 2000, *Multiparameter Equations of State*. Berlin: Springer-Verlag.

Span, R., Lemmon, E. W., Jacobsen, R. T., Wagner, W., Yokozeki, A., 2000, A reference equation of state for the thermophysical properties of nitrogen for temperatures from 63.151 to 1000 K and pressures to 2200 MPa. *J. Phys. Chem. Ref. Data*, **29**(**6**): 1361–1433.

Spear, F. S., 1995, *Metamorphic Phase Equilibria and Pressure-Temperature-Time Paths*. Washington DC: Mineralogical Society of America.

Spencer, J. N., 1973, ΔG and $\partial G/\partial\xi$. *J. Chem. Edu.*, **51**: 577–9.

Spycher, N. F. and Reed, M. H., 1988, Fugacity coefficients of H_2, CO_2, CH_4, H_2O, and of H_2O–CO_2–CH_4 mixtures: a virial equation treatment for moderate pressures and temperatures applicable to calculations of hydrothermal boiling. *Geochim. et Cosmochim. Acta*, **52**: 739–49.

Steinmann, P., Lichtner, P. C., and Shotyk, W., 1994, Reaction path approach to mineral weathering reactions. *Clays and Clay Minerals*, **42**: 197–206.

Stokes, R. H. and Robinson, R. A., 1948, Ionic hydration and activity in electrolyte solutions. *J. Amer. Chem. Soc.*, **70**: 1870–8.

Stryjek, R. and Vera, J. H., 1986a, PRSV – An improved Peng–Robinson equation of state for pure compounds and mixtures. *Can. J. Chem. Engng.*, **64**: 323–33.

1986b, PRSV – An improved Peng–Robinson equation of state with new mixing rules for strongly non-ideal mixtures. *Can. J. Chem. Engng.*, **64**: 334–40.

1986c, PRSV2 – A cubic equation of state for accurate vapor-liquid equilibria calculations. *Can. J. Chem. Engng.*, **64**: 820–6.

Suleimenov, O. M. and Krupp, R. E., 1994, Solubility of hydrogen sulfide in pure water and in NaCl solutions from 20 to 320 °C and at saturation pressures. *Geochim. et Cosmochim. Acta*, **58**: 2433–44.

Tanger, J. C., IV and Helgeson, H. C., 1988, Calculation of the thermodynamics and transport properties of aqueous species at high pressures and temperatures: revised equations of state for the standard partial molal properties of ions and electrolytes. *Amer. J. Sci.*, **288**: 19–98.

Thompson, J. B. Jr., 1967, Thermodynamic properties of simple solutions, *in*: P. H. Ableson, ed., *Researches in. Geochemistry* vol. 2. New York: Wiley, pp. 340–61.

1970, Geochemical reaction and open systems. *Geochim. et Cosmochim. Acta*, **34**: 529–51.

Thompson, J. B. Jr. and Waldbaum, D. R, 1968, Mixing properties of sanidine crystalline solutions: I. Calculations based on ion-exchange data. *Amer. Mineralogist*, **53**: 1965–99.

1969, Mixing properties of sanidine crystalline solutions: III. Calculations based on two-phase data. *Amer. Mineralogist*, **54**: 811–38.

Tisza, L, 1966, *Generalized Thermodynamics*. Cambridge, MA: MIT Press.

Tremaine, P. R., Sway, K., and Barbero, J. A., 1986, The apparent molar heat capacity of aqueous hydrochloric acid from 10 to 40 °C. *J. Sol. Chem.*, **15**: 1–22.

Trusler, J. M., 2000, The virial equation of state, in: J. V. Sengers, R. F. Kayser, C. J. Peters, and Jr. H. J. White, eds., *Equations of State for Fluids and Fluid Mixtures*. Amstredam: Elsevier Science B. V. pp. 34–74.

Tunell, G., 1932, On the history and analytical expression of the first and second laws of thermodynamics, and the rôle of the differentials dW and dQ. *J. Phys. Chem.*, v. 36, pp. 1744–1771.

Tuttle, O. F. and Bowen, N. L., 1958, *Origin of Granite in the Light of Experimental Studies in the System* $NaAlSi_3O_8$–$KAlSi_3O_8$–SiO_2–H_2O. Boulder, CO: Geological Society of America, Memoir 74.

Ulbrich, H., and Waldbaum, D., 1976, Structural and other contributions to the third law entropies of silicates. *Geochim. et Cosmochim. Acta*, **40**: 1–24.

Ulmer, G. C. and Barnes, H. L., 1987, *Hydrothermal Experimental Techniques*. New York: Wiley.

Van Zeggeren, F. and Storey, S. H., 1970, *The Computation of Chemical Equilibria*. Cambridge: Cambridge University Press.

Voight, W., 2001, Solubility equilibra in multicomponent oceanic salt systems from $t = 200$ to 200 °C. Model parameterization and databases. *Pure Appl. Chem.*, v. 73, pp. 831–844.

Wagman, D. D., Evans, W. H., Parker, V. B., Schumm, R. H., Halow, I., and Bailey, S. M., 1982, The NBS tables of chemical thermodynamic properties. *J. Phys. Chem. Ref. Data*, 2 Supplement no. 2. Washington DC: American Chemical Society.

Wagner, C., 1952, *Thermodynamics of Alloys*. Reading, MA: Addison-Wesley Publishing Co.

Wagner, W., et al., 2000, The IAPWS industrial formulation 1997 for the thermodynamic properties of water and steam. *ASME J. Eng. Gas Turbines and Power*, **122**: 150–222.

Wagner, W., and A. Pruß, 2002, The IAPWS formulation 1995 for the thermodynamic properties of ordinary water substance for general and scientific use. *J. Phys. Chem. Ref. Data*, **31**: 387–535.

Walas, S. M., 1985, *Phase Equilibria in Chemical Engineering*. London: Butterworth.

Waldbaum, D. R. and Thompson, J. B., Jr., 1968, Mixing properties of sanidine crystalline solutions: II. Calculations based on volume data. *Amer. Mineralogist*, **53**: 2000–15.

1969, Mixing properties of sanidine crystalline solutions: IV. Phase diagrams from equations of state. *Amer. Mineralogist*, **54**: 1274–98.

Weare, J. H., 1987, Models of mineral solubility in concentrated brines with application to field observations, in: I. S. E. Carmichael and H. P. Eugster, eds., *Thermodynamic Modeling of Geological Materials: Minerals, Fluids and Melts*, Reviews in Mineralogy, vol. 17. Washington DC: Mineralogical Society of America, pp. 143–76.

Weinreich, G., 1968, *Fundamental Thermodynamics*. Reading, MA: Addison-Wesley Publishing Co.

Wilks, J., 1961, *The Third Law of Thermodynamics*. Oxford: Oxford University Press.

Williamson, M. A. and Rimstidt, J. D., 1994, The kinetics and electrochemical rate-determining step of aqueous pyrite oxidation. *Geochim. et Cosmochim. Acta*, **58**: 5443–54.

Wolery, T. J., 1979, *Calculation of Chemical Equilibrium between Aqueous Solution and Minerals. The EQ3/6 Software Package*. Livermore, CA Lawrence Livermore National Lab.

1983, *EQ3NR, A Computer Program for Geochemical Aqueous Speciation-Solubility Calculations. User's Guide and Documentation*. Livermore, CA Lawrence Livermore National, Lab.

Wolery, T. J. and Jackson, K. J., 1990, Activity coefficients in aqueous salt solutions. Hydration theory equations, in: (eds., D. C. Melchior and R. L. Bassett). *Chemical Modeling of Aqueous Systems II.* ACS Symposium Series 416. Washington, DC: American Chemical Society, pp. 16–29.

Wood, B. J. and Nicholls, J., 1978, The thermodynamic properties of reciprocal solid solutions. *Contrib. Mineral. Petrol.*, **66**: 389–400.

Wood, R. H., 1989, Flow calorimetry and densitometry at high temperatures. *Thermochimica Acta*, **154**: 1–11.

Wood, S. A. and Samson, I. M., 1998, Solubility of ore minerals and complexation of ore metals in hydrothermal solutions, in: J. P. Richards and P. B. Larson, eds., *Techniques in Hydrothermal Ore Deposits Geology*. Reviews in Economic Geology, vol. 10. El Paso, TX: Economic Geology Publishing Co., pp. 33–80.

Wood, S. A. and Spera, F. J., 1984, Adiabatic decompression of aqueous solutions: Applications to hydrothermal fluid migration in the crust. *Geology*, **12**: 707–10.

Wooley, H. W., 1980, Thermodynamic properties for H_2O in the ideal gas state, *in: Water and Steam – Their Properties and Current Industrial Applications. Proceedings of the 9th International Conference on the Properties of Steam*, J. Straub and K. Scheffler, eds., Oxford: Pergamon, pp. 166–80.

Zhu, C. and Anderson, G. M., 2002, Environmental Applications of Geochemical Modeling. Cambridge: Cambridge University Press.

Zimmerman, G. H., Gruszkiewicz, M. S., and Wood, R. H., 1995, New apparatus for conductance measurements at high temperatures: conductance of aqueous solutions of LiCl, NaCl, NaBr, and CsBr at 28 Mpa and water densities from 700 to 260 kg m^{-3}. *J. Chem. Phys.*, **29**: 11612–25.

Index