STATISTICAL MECHANICS OF SOLIDS

LOUIS A. GIRIFALCO

STATISTICAL MECHANICS OF SOLIDS

MONOGRAPHS ON THE
PHYSICS AND CHEMISTRY
OF MATERIALS #58

STATISTICAL MECHANICS OF SOLIDS

Louis A. Girifalco

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Oxford New York
Auckland Bangkok Buenos Aires Cape Town Chennai
Dar es Salaam Delhi Hong Kong Istanbul Karachi Kolkata
Kuala Lumpur Madrid Melbourne Mexico City Mumbai Nairobi
São Paulo Shanghai Taipei Tokyo Toronto

Published by Oxford University Press, Inc.
198 Madison Avenue, New York, New York 10016

www.oup.com

First issued as an Oxford University Press paperback, 2003

Library of Congress Cataloging-in-Publication Data
Girifalco, L.A. (Louis A.)
Statistical mechanics of solids / Louis A. Girifalco.
p. cm.
Includes bibliographical references.
ISBN 0-19-511965-7; 0-19-516717-1 (pbk.)
1. Solids—Mathematical models. 2. Statistical mechanics.
I. Title.
QC176.G54 2000
530.4'1—dc21 99-33191

9 8 7 6 5 4 3 2 1

Printed in the United States of America
on acid-free paper

Preface

Statistical mechanics is one of the greatest achievements of the human mind, combining theoretical elegance with powerful methods for studying an enormous range of specific systems and phenomena. It connects the properties of the material world to the basic constituents of matter at the deepest level. It lays bare the atomic and molecular content of thermodynamics and shows how to obtain new and important relations among the properties of matter, how to organize enormous amounts of scientific information, and how to understand it. Furthermore, it is a dynamic discipline in that it continues to grow and enhance our understanding of the physical world. Its study is essential for those who want to probe deeply into the nature of things.

Statistical mechanics can be studied from a variety of viewpoints and at different levels of theoretical sophistication. It is an important discipline for the study of matter in every form and at every level of aggregation. The number of available texts on the subject is therefore large, and they range from elementary to quite advanced treatments. However, modern texts that include topics on solids generally focus on such phenomena as phase transitions at critical points, universality, and those commonalities that permit similar treatments of solids and liquids. Aside from my earlier book (*Statistical Physics of Materials*, 1973), there has been no satisfactory text that concentrates solely on the statistical mechanics of solids from the point of view of the material properties of the solid state. This is such an important subject with such far-reaching applications that the neglect of this approach is somewhat surprising. My major objective was to fill this gap.

This book grew out of a course in statistical mechanics of solids I give to students of materials science at the University of Pennsylvania. It bears some resemblance to my previous book but includes much new and updated material. My intention is to provide a useful reference as well as a basis for course work.

This book is intended for those whose primary interests are in the *properties* of solids rather than in the statistical mechanical theory itself. One of my objectives is to aid students in developing a physical insight that relates properties to statistical models and theories. Such insight is extremely valuable for those studying the application of statistical mechanics to condensed matter. I have therefore included some discussion of the physics of the systems and phenomena presented along with the statistical mechanical developments and their physical interpretations.

Another objective is to make the subject matter accessible to as wide an audience as possible among those whose background is equivalent to the first- or second-year graduate school level. This includes students with undergraduate

degrees in physics, chemistry, and materials science, and their differences must be recognized. The variation in their mathematical training, their knowledge of the solid state, and, surprisingly, their grasp of thermodynamics poses a special problem for the teacher. I address this problem by starting at what should be a common base and trying not to assume too much in the way of specialized background.

A third objective is to give a representative sample of those aspects of materials and solid state science that benefit the most from the application of statistical mechanics. The field has grown enormously since my earlier effort in 1973 and includes many new important subjects. The contents of this book reflect this.

These objectives have determined both the orientation and the content of the book. Regarding the fundamentals of statistical mechanics, I have tried to present a careful exposition of the basic concepts without excessive formalism. Thus, although statistical mechanics is developed as a tool for understanding properties, rather than for its own sake, I have given a fairly detailed discussion of the nature of ensembles and their relation to thermodynamics. In fact, chapters 1–3 should be excellent preparation for advanced study of both the fundamentals of statistical mechanics and its theoretical formalism.

The book starts with a brief review of thermodynamics despite the fact that students should have had at least one course in the subject. My experience has been that the review is needed, particularly regarding the origin and meaning of entropy and the second law. I do not agree with some of the modern approaches of first teaching thermodynamics as either being derived from statistical mechanics or as a purely axiomatic structure that connects to experiment only after considerable formal development. Both these approaches are valid and interesting, but should be preceded by a knowledge of the origin of thermodynamics from experimental data.

Chapters 4 and 5 present the statistical thermodynamics of simple crystals and include a discussion of both harmonic and anharmonic properties. The traditional theory of the harmonic crystal is still of great value and has not changed much in the past decades. But modern developments make it possible to deal with anharmonicity in a more rigorous and transparent way than before, while still maintaining much of the simplicity of the original Gruneisen method.

Chapters 6 and 7 work out the consequences of the free electron theory of metals and semiconductors, for both equilibrium and transport properties.

The emphasis on properties and physical insight has led to a sequence on cooperative phenomena in which order–disorder alloys and magnetic ordering are treated separately in chapters 8 and 9 (although the commonalities of these two kinds of phenomena are stressed). The overall theory of phase transformations in chapter 10 and of critical points in chapter 11 is taken up only after the presentation of these two topics. I believe this ordering keeps a closer connection with real materials than does starting with the more general theory.

Because of the ever-growing importance of surface phenomena, chapter 12 is devoted to the thermodynamics and statistical mechanics of surfaces.

The theory of random flight has a large number of uses and is given in chapter 13. For the study of materials, the most important applications are to polymers and diffusion. Chapter 14 presents the statistical mechanics of polymer chains, which draws heavily on the theory in chapter 13.

Chapters 15–17 develop the statistical mechanics of point defects and diffusion. These are of great importance for the properties of solids and can be given an elegant treatment, at least for simple crystals, that well illustrates the power of the statistical theory.

The range of properties that have a statistical basis is so large and so important that no selection of topics can be complete. However, it is obvious that there are some subjects that must be presented because they are so basic, and others that are part of the mainstream of modern interests in material properties. These considerations, along with my personal interests, have determined my selection of specific applications of statistical mechanics. I have included a number of appendices on topics that are important for understanding statistical mechanics in solids, but that would interrupt the flow and be out of place in the body of the book.

The exercises at the end of each chapter are designed to enhance the student's understanding of the subject matter, and to help develop physical insight. The student should do them all.

I am particularly indebted to the students who have taken my course, and who have enthusiastically hunted down errors in my notes.

This page intentionally left blank

Contents

STATISTICAL MECHANICS OF SOLIDS

This page intentionally left blank

1

The Basics of Thermodynamics

1.1 The existence of equilibrium and state functions

Thermodynamics addresses those relations among the macroscopic properties of macroscopic systems that do not depend on the properties of its atomic constituents. It is a fact of experience that systems exist whose properties do not change with time, within the limits of measurement. It is also a fact that systems exist whose properties do change with time. In making such statements, the time scale must be considered. Thus, a piece of iron may have properties whose values are constant over some time period, say, one week, but over a period of months or years the iron might develop rust on its surface.

To develop thermodynamic theory, we first restrict our analysis to systems in the absence of external fields, such as gravity or magnetism, and that consist of isotropic phases. An *isotropic phase* is a homogeneous substance none of whose properties depend on spatial directions. If the system consists of just one phase, then we will call it a *simple* system. A heterogeneous system is one that consists of several simple systems. The extension of the theory to heterogeneous complex systems, and to systems in external fields, is straightforward once the thermodynamics of simple systems is understood.

A system is said to be in *internal equilibrium* if none of its properties change with time. Such properties might be the pressure exerted by the system, or its volume, chemical composition, magnetic moment, refractive index, or temperature. Experiments have shown that the values of these properties are not all independent. For example, in a one-component system the temperature, volume, and pressure are related through the equation of state. In general, the *state* of a system at equilibrium is defined as the minimum amount of information needed to determine all of the system's properties. One of the tasks of thermodynamics is to find the number and types of system properties that define state.

The measured properties of a system at equilibrium are independent of the system's history. Thus, a system could arrive at a particular composition, volume, and pressure by any of an infinite number of ways, each of which would result in the system having the same value for these properties. Not all system properties are independent of a system's history. The magnetization of a piece of iron depends on how it was magnetized; the electrical resistivity of a metal depends on the defects present, and these depend on how the metal was treated.

It is part of the definition of equilibrium that the system properties not only are constant in time but also are independent of how the system got to equilibrium. This is a definition of what may be called *absolute equilibrium*. Partial

equilibrium also exists, in which the properties are time independent although the values of some of them may be history dependent.

A property of a system at equilibrium that is independent of the systems history is called a *state function.*

1.2 Empirical temperature scales

The idea of temperature requires some special comment. Other quantities can be directly measured and have their origin in mechanics and in the basic measurements of mass, length, and time. Temperature, on the other hand, has its origin in our perception of hot and cold bodies and their interactions. We note that these perceptions are accompanied by properties of the bodies that can be measured quantitatively. Thus, if a thin metal rod is exposed to hot water, we notice that its length increases. We can qualitatively sense that water gets progressively hotter as it sits in a kettle over a fire, and this sense can be given a quantitative form by noting that the length of the metal rod progressively increases as the water approaches boiling.

An empirical temperature scale can be constructed by choosing some physical property that we know is different for hot and for cold bodies, constructing an instrument that measures this property, putting the instrument in contact with a system whose temperature we wish to know, and then measuring that property. Instruments of this sort (thermometers) can be based on such properties as length (mercury or alcohol thermometers), electrical resistance (resistance thermometers), voltage (thermocouples), or volume (gas thermometers). An obvious problem is that the empirical temperature scales depend on the physical property chosen to measure it and even on the material the thermometer is made of. If this could not be overcome, a temperature scale would be arbitrary and not have any fundamental significance.

1.3 The ideal gas temperature

The gas thermometer occupies a special place in defining empirical temperature scales. The reason for this is that, if the gas is sufficiently dilute, all gas thermometers give the same temperature scale. At low pressures and high volumes, the gases approach an ideal behavior, and it is therefore possible to construct a thermometer based on the properties of an ideal gas.

Experiment shows that in the limits of high temperature and/or low pressure, the volume, pressure, and temperature of a gas are related by

$$\frac{PV}{(A+t)} = \frac{P_0V_0}{(A+t_0)} \tag{1.3.1}$$

where (P, V) and (P_0, V_0) denote the pressure and volume at any two different temperatures t and t_0, respectively, the temperature being measured on an empirical temperature scale. This is the ideal gas law. If we adopt the centigrade scale of temperature, then the constant A is found to be 273.16. We are thus led to define a new temperature scale by the relation

$$T = t + 273.16 \tag{1.3.2}$$

T is the temperature on the ideal gas temperature scale. We expect this to have a more fundamental significance than t because the ideal gas law is valid for all gases that are sufficiently dilute and are at sufficiently high temperature.

That is, the ideal gas law holds when the intermolecular interactions are very small. In fact, it will be shown that the temperature defined by ideal gases is the same as the absolute temperature of thermodynamics.

With this definition of temperature, equation (1.3.1) becomes

$$PV = NkT \tag{1.3.3}$$

where k is Boltzmann's constant and N is the number of molecules in the gas.

A gas thermometer is just a bulb or glass tube filled with a dilute gas. The gas can be held at either constant volume or constant pressure. The pressure, or the volume, is then measured and converted to temperature via equation (1.3.3).

The ideal gas is an example of a system whose thermodynamic state is completely specified by two of the three variables (P, T, V). It is an empirical fact that if any two of these are known, all properties of the system are determined. The ideal gas equation is the simplest example of an equation of state that connects the state parameters of a system.

1.4 The mechanical equivalent of heat

A large array of experiments have shown that doing work on a system raises its temperature. Starting with Count Rumford's observation that the mechanical work in boring cannon raises the temperature of the brass, and the extensive experiments of James Prescott Joule that related the performance of many kinds of work to the temperature rise of systems, it is an experimental fact that work has important thermal effects. The results of these experiments can be summarized in the statement that work can be converted into heat and that the ratio of the work expended to the heat generated is always the same constant. This implies, of course, that a concept of heat has been developed and that a method exists for its measurement. Much of the verbal description of thermal phenomena is a holdover from the days of the caloric theory, when it was believed that heat was a substance that flows from hotter to colder bodies. While this is in some respects unfortunate, it still provides a convenient descriptive language. In modern terms, the concept of heat and its measurement can be made precise as follows.

Consider a system, which for the sake of specificity we take to consist of pure water, enclosed in a perfectly insulating envelope (this can be approximated by a Dewar flask or asbestos). Assume that we have an empirical temperature scale so that we know that any temperature change in the environment has no effect on the temperature of the system and any temperature change in the system has no effect on the temperature of the environment. The walls are said to be thermally insulating or *adiabatic*. Assume also that the system can be accessed by thermometers or other instruments in such a way that the thermally insulating property of the walls is not sensibly disturbed. Now do work on the system by turning a paddle wheel that is immersed in the water. The experimental result is that the temperature of the water goes up. We also notice that, if a certain amount of work raises the temperature of the water by one degree, twice as much work will raise the temperature by two degrees.

In general, the temperature rise is proportional to the amount of work done. (This is only approximately true if the temperature range over which measurements are made is large.) Also, the temperature rise is inversely proportional to the mass of the water. It takes twice as much mechanical work to raise the temperature by one degree if there are two kilograms of water rather than one. This law is quite exact. We are thus led to define a unit quantity of heat

in terms of a unit temperature rise for a unit amount of water. In fact, the definition of a calorie is that amount of heat needed to raise the temperature of one gram of water from 14.5 to 15.5°C. It is found that the amount of work required to do this in an adiabatic system is always the same. This experiment is described by saying that the mechanical work generates heat. The ratio of the mechanical work, in mechanical units, to the heat generated, in thermal units, is called the *mechanical equivalent of heat.*

Since work is done on the system, and since we insist on the law of the conservation of energy, we associate the rise in temperature, and therefore the heat generated, with an increase in the internal energy of the system.

Different substances exhibit different temperature rises for the same amount of mechanical work, and the ratio of the two temperature rises is used to define the relative heat capacities of the substances. An enormous number of calorimetric experiments are readily rationalized and understood on the basis of this concept.

The extension of the above experiments to other kinds of systems and to other kinds of work yields consistent results. For work produced by mechanical, gravitational, electrical, or magnetic means acting on gases, liquids, or solids of any composition, the mechanical equivalent of heat is always the same.

1.5 Walls and the zeroth law of thermodynamics

We know that it is possible to effectively isolate a system from its surroundings in the sense that any changes in the environment have no effect on the values of the properties of the system. A system so isolated is said to be surrounded by a completely isolating wall. For purposes of illustration, let the system be a closed bottle of gas surrounded by an adiabatic wall and immersed in a pool of water. The gas has a fixed volume, pressure, composition, and empirical temperature, and let these be constant in time. The gas is then said to be in internal equilibrium because of the constancy of its properties and because, by definition, it is not interacting with its surroundings. If some change is effected in the pool of water, such as a rise in temperature or pressure, or a change in chemical composition, then there is no change in the physical properties of the gas in the sealed bottle. This is the meaning of a completely isolating wall.

It is possible to have a partially isolating wall. Thus, if our bottle is fitted with a piston and cylinder at its neck instead of being sealed, a change in pressure of the pool of water will effect a change in the gas because the change in pressure will cause the piston to move, thereby changing the volume of the gas in the bottle. But a change in temperature of the water, while leaving the pressure constant, will not result in a change in temperature of the gas. Such a wall is said to be *adiabatic,* and the system is said to be *thermally isolated.*

Not all walls are adiabatic. In fact, special care must be taken to thermally isolate a system. Most containers will allow the systems within them to respond to temperature changes in their surroundings sooner or later. For many walls, such as those made of metal, the response is quite rapid. A system is said to be surrounded by a *diathermic wall* if it responds to temperature changes in its environment.

Now consider systems surrounded by diathermic walls, each system being at internal equilibrium. The following statements are found to be true from experiment:

a. If two systems that are each in internal equilibrium but at different temperatures are brought into contact through a diathermic wall, they will each come to a new state of internal equilibrium. Both will then have the same temperature, whose value is between the two initial temperatures. The two systems are then said to be in thermal equilibrium with each other.
b. If a system A is in thermal equilibrium with a system B, and if B is in thermal equilibrium with a system C, then systems A and C are in thermal equilibrium with each other. That is, if two systems are each in equilibrium with a third system, they are in equilibrium with each other. This is often called the zeroth law of thermodynamics.

According to the zeroth law, for a system with diathermic walls to be in internal equilibrium, it must be in thermal equilibrium with its surroundings.

1.6 Spontaneous, reversible, and irreversible processes

For the purposes of thermodynamics it is important to distinguish among the general processes by which a system can change its state. These processes are called spontaneous, reversible, and irreversible.

A *spontaneous* process is one that takes place naturally without any intervention. Examples of spontaneous processes are the flowing of water downhill, the rusting of iron in moist air, and the dissolution of sugar in hot coffee. If a system is isolated and it is found that its properties change with time, it is undergoing a spontaneous internal process. If a system is not isolated, but immersed in an environment that is sensibly constant (such as a large pool of water at constant temperature and pressure) and its properties change with time, it, too, is undergoing a spontaneous internal process but may also be involved in an interaction with its environment. The combined system plus environment is then undergoing a spontaneous process. If the environment does not have constant properties, the details of the system–environment interaction must be analyzed to determine whether or not a spontaneous process is taking place.

Consider a system that is initially at equilibrium with given values for its properties. For specificity let the system be a one component gas with pressure, volume, and temperature given by (P_1, V_1, T_1). Now suddenly change the volume to V_2, by rapidly pushing on a piston, and wait. The system eventually comes to a new state of equilibrium with pressure, volume, and temperature given by (P_2, V_2, T_2), but the process of getting there is rather chaotic. Because the volume is changed suddenly, turbulent flow takes place in the gas as it rushes to adjust to the new volume. The frictional effect of this turbulence generates heat that raises the temperature either of the gas or of its surroundings, or both. Because of this frictional effect, if the volume is restored to its original value of V_1 by pulling up on the piston, and the system is then allowed to come to equilibrium, the original values of the pressure and temperature, (P_1, T_1), are not recovered. The process of going from state 1 to state 2 via a rapid change in volume is said to be *irreversible* because it cannot be undone by just changing the volume back to its original value. It is clear that the states the system goes through between 1 and 2 are not equilibrium states. There are an infinite number of equilibrium states of the system, but none of these occur during the irreversible process because equilibrium does not exist until we wait for the system to calm down.

The irreversible effects can be mitigated by pushing on the piston more slowly to get from V_1 to V_2. The degree of turbulence and the internal frictional

effects are then less. In fact, the degree of irreversibility clearly depends on the rate at which the pressure is applied to effect the volume change. If this rate is very slow, then at any instant of time, the difference between the internal pressure and the applied pressure is very low and the system is near an equilibrium state. In the limit of zero rate of change of the pressure, the system is always at an equilibrium state. This leads us to define a *reversible* process as one in which the rate of change of all relevant parameters is so low that the system is always in an equilibrium state. Clearly, the process is truly reversible only if the process is infinitely slow. Thus, there are no real processes that are reversible, but the slower the process the more nearly it approaches reversibility. In practice, a process can be called reversible if at every stage it is in equilibrium within the accuracy of measurements made on the system. These concepts apply to any sort of system and any changes in state functions, not only to (P, V, T) changes in gases.

The difference between reversible and irreversible processes is of great physical importance. From a mathematical point of view, reversible processes connect equilibrium states. Thus, a system can exist in an infinite number of equilibrium states, each state being characterized by specific values for the state functions of the system so that if F is a state function that depends on parameters x_i,

$$F = F(x_1, x_2, \ldots) \tag{1.6.1}$$

then two equilibrium states that are very close together and differ by an amount dF are related by

$$dF = \sum_i \frac{\partial F}{\partial x_i} dx_i \tag{1.6.2}$$

and since F is a state function, its value for any state is independent of the system's history. In a reversible process, state functions exist throughout the process, and all differentials of state functions have the form of equation (1.6.2).

1.7 Work and the dependence of work on the path

Changes in the equilibrium state of a system can be effected by changing its temperature or by doing work. If the system has diathermic boundaries, its temperature can be changed by changing that of its environment. Work can be done on systems that have either diathermic or adiabatic boundaries. If the temperature of the system is maintained at a constant value while work is being done, the process is said to be isothermal. This can be done by surrounding the system with diathermic walls and placing it in contact with another system at the same temperature that is very large compared to the system of interest.

If the system is surrounded by an adiabatic wall while work is being done on (or by) it, the process is said to be adiabatic. In an *adiabatic process*, the temperature of the system may change, but this does not affect the temperature of any other system or of the environment.

A system at equilibrium has a definite energy that depends only on the equilibrium parameters and not on its history. The energy is therefore a state function.

The energy of a system surrounded by adiabatic walls can be changed by doing work on the system. Thus, in the bottle of gas surrounded by asbestos and fitted with a piston, as described in section 1.5, work can be done on the

system by applying a pressure on the piston that is greater than the gas pressure. The piston then moves, compressing the gas until its pressure rises to equal the external pressure. Since the system is impervious to all other influences, the work done on the system increases its internal energy. Let U_i be the initial energy of the system and let U_f be the energy of the system after a certain amount of work has been done. Also, let us adopt the convention that work done *by* the system is positive and work done *on* the system is negative. Then, from the law of the conservation of energy,

$$U_f - U_i = -W_{i,f} \qquad \text{(adiabatic)} \tag{1.7.1}$$

where $W_{i,f}$ is the work done by the system in going from state i to state f. Equation (1.7.1) merely states that the decrease in energy of a system resulting from an adiabatic process is equal to the work done by the system.

In general, there is a temperature change accompanying an adiabatic process such that the temperature of the final state is not the same as that of the initial state.

Equation (1.7.1) is correct whether or not the process in going from the initial to the final state is reversible because only work can get through to an adiabatic system.

Now let us shift our viewpoint and consider the system doing work rather than having work done on it. That is, the system exerts a pressure on its environment, thereby changing the volume of the system from V_i to V_f. If the process is carried out reversibly so that the difference between the internal pressure and the external pressure is always very small, a certain amount of work will be done given by $\int_{V_i}^{V_f} P dV = W_{i,f}$, the relation between P and V being given by the equation of state at every stage in the process.

What if the process is carried out irreversibly? In this case, frictional processes will be set up that create turbulent internal motion and some of the energy that would have gone into work against the external pressure is dissipated in the internal friction of the system. We thus get the result that the maximum amount of work that can be performed by an adiabatic system in irreversibly going from one state to another is less than that for the reversible process. The irreversible processes do less external work.

The same is true for work done in an isothermal process, because in a reversible process the pressure exerted by the system is infinitesimally close to the external pressure. But if the process were irreversible, the external pressure would be appreciably less than that exerted by the system, and then the work done by the expanding system would be decreased. We then have a general conclusion that work done in going from one state to another is a maximum if the process is reversible.

Since the energy is a state function, equation (1.7.1) shows that the work done in an adiabatic process is independent of the detailed path taken in going from the initial to the final state. That is, the pressure could have been applied rapidly or slowly or in varying increments; the work in going adiabatically from a given equilibrium initial state to a given equilibrium final state is always the same. This is not true for all kinds of processes, and it can be shown that the amount of work done depends on the particular way the process is carried out. It is easiest to demonstrate this for a system consisting of an ideal gas.

Let us assume that the gas is surrounded by diathermic walls and is in equilibrium with a large water bath at a constant temperature. Let the system undergo a change in state from an initial pressure, volume, and temperature of (P_i, V_i, T_i) to a final pressure, volume, and temperature (P_f, V_f, T_f). Note that the temperature is constant because the walls are diathermic.

Let us examine two reversible processes by which we can go from the initial to the final state. For process 1, we first change the volume of the gas from its initial to its final value while keeping the temperature constant. The amount of work done in this step is

$$\int_{V_i}^{V_f} P dV = NkT_i \ln\left(\frac{V_f}{V_i}\right) \qquad (\text{at } T_i) \tag{1.7.2}$$

Next we change the temperature to its final value T_f by changing the temperature of the reservoir while keeping the pressure and volume constant. No work is done in this second step and the total work done is given by (1.7.2).

In comparison, in process 2 let the first step consist of changing the temperature to its final value while keeping the pressure and volume constant. No work is done in this step. For the second step of process 2, change the volume to its final value by increasing the pressure. The amount of work done is now given by

$$\int_{V_i}^{V_f} P dV = NkT_f \ln\left(\frac{V_f}{V_i}\right) \qquad (\text{at } T_f) \tag{1.7.3}$$

The amount of work done by the two processes is clearly different, although both processes started at the same initial state and ended up at the same final state. The work done on or by a system is therefore *not* a state function of the system. This conclusion is not restricted to ideal gas systems. The reasoning is easily extended to any kind of work and to any equation of state.

The above considerations lead to a definition of heat that is based purely on mechanical concepts as follows: in any process, the work done by a system in going from one state to another is equal to the internal energy change only if the process is adiabatic. For all other processes, the work done is *not* equal to the change in internal energy. We define the heat absorbed by the system as the difference between the work done and the internal energy change.

1.8 The first law of thermodynamics

We are now ready to formulate the first law of thermodynamics. In fact, the last sentence in the preceding section is a statement of the first law that we have obtained as a mere definition of the heat change of a process. Written in symbols, the definition is

$$Q_{i,f} = U_{i,f} + W_{i,f} \tag{1.8.1}$$

Again note that work done *by* the system is taken as positive, so (1.8.1) defines heat as the change in internal energy minus the work done on the system.

There are two facts that give the first law a status that goes beyond that of mere definition. The first is that the internal energy is taken to be a state function, and every consequence of its being so, that can be tested by experiment, has been verified. The second fact is the existence of the mechanical equivalent of heat, which is one of the most accurately verified experimental results in all of science. The first law is therefore nothing but a statement of the law of conservation of energy extended to processes that involve thermal changes. To emphasize this, we write (1.8.1) as

$$\delta U = \delta Q - \delta W \tag{1.8.2}$$

where δQ is the heat absorbed by the system and δW is the work done by the system in any process for which the change in internal energy is δU. Equation (1.8.2) is true for spontaneous and irreversible processes as well as for reversible processes; we adopt the convention that δ is used to label changes that can be either reversible or irreversible, while lowercase d is used to label differential changes for reversible processes.

1.9 Heat capacity, energy, and enthalpy

Consider a system that undergoes a reversible process from a state A to a state B. We assume heat can be absorbed or liberated by the system during the process, that only PV work can be done by or on the system, and that the composition of the system is constant. (It will be clear from the following discussion that the results are easily generalized to systems for which other kinds of work can be done.) The change in the state of the system is then defined by the change in the values of pressure, volume, and temperature. The ideal gas temperature scale is used to measure the temperature.

First assume that the volume is kept constant during the entire process in going from A to B, such that no work is done on or by the system. Then, from the first law, the energy change is just equal to the heat liberated or absorbed by the system during the process. If we call this heat δQ_V, then the energy change is

$$\delta U = \delta Q_V \qquad \text{(constant volume process)} \tag{1.9.1}$$

In general, the absorption of heat is accompanied by a temperature change. The heat capacity is defined as the heat absorbed when the temperature of the system is increased by a small amount. That is, if the system absorbs an amount of heat δQ when the temperature is increased by an amount δT, then the heat capacity is defined by

$$C = \operatorname*{Lim}_{\delta T \to 0} \frac{\delta Q}{\delta T} \tag{1.9.2}$$

If the volume is constant during the temperature increase in a reversible process, then (1.9.1) shows that (1.9.2) becomes the derivative of the energy with respect to temperature at constant volume. The constant volume heat capacity is therefore defined by

$$C_V = \left(\frac{\partial U}{\partial T}\right)_V \tag{1.9.3}$$

where the heat is transferred to the system reversibly.

Now consider a constant pressure process in which the volume goes from V_1 to V_2. The work done during this process is

$$W = \int_{V_1}^{V_2} P dV = P(V_2 - V_1) \tag{1.9.4}$$

From the first law, the energy change is the heat absorbed minus the work done, so if U_1 and U_2 are the energies in the initial and final states, respectively, then

$$U_2 - U_1 = \delta Q_P - P(V_2 - V_1) \tag{1.9.5}$$

or

$$\delta Q_P = (U_2 + PV_2) - (U_1 + PV_1) \tag{1.9.6}$$

δQ_P being the heat absorbed in a constant pressure process.

Equation (1.9.6) leads us to define the function

$$H = U + PV \tag{1.9.7}$$

called the *enthalpy*. For a constant pressure process, the heat absorbed is the change in enthalpy.

Clearly, the heat capacity at constant pressure for a reversible increase in temperature is the derivative of the enthalpy at constant pressure:

$$C_P = \left(\frac{\partial H}{\partial T}\right)_P \tag{1.9.8}$$

1.10 The second law of thermodynamics and entropy

We know from experience that many processes cannot take place even if they satisfy the law of conservation of energy. Water left to itself, for example, is never seen to flow uphill, and if two bodies at different temperatures are brought into contact and then left alone, heat never flows from the colder to the hotter body. This experience is embodied in the second law of thermodynamics, which has been expressed in two basic forms. One form states that a quantity of heat extracted from a system cannot be converted entirely into work while leaving everything else unchanged. This is called the *Kelvin statement*. Another version of the second law states that heat cannot be transferred from a colder to a hotter body while leaving everything else unchanged. This is called the *Clausius statement* of the second law.

The two statements of the second law are equivalent, as can be seen by showing that if one is violated then so is the other. For example, if the Kelvin statement is false, then a quantity of heat from a system at temperature T_1 can be completely converted to work. This work could then be completely used to heat a system at a higher temperature T_2, thereby violating the Clausius statement. Thus, if the Kelvin statement is false, so is the Clausius statement.

Conversely, if the Clausius statement is false, then a quantity of heat Q can be extracted from a system at a temperature T_1 and completely transferred to a body at a higher temperature T_2. Now let the second system do an amount of work equivalent to this amount of heat. The result is that an amount of heat Q is converted completely into work, thereby violating the Kelvin statement. Thus, if the Clausius statement is false, so is the Kelvin statement. The Kelvin and Clausius statements of the second law are thus seen to be equivalent.

A principle reason we value energy is because the performance of work is so essential to so much human activity. From this anthropomorphic point of view, therefore, some forms of energy (such as the potential energy at the top of a waterfall, or the chemical energy in fuel) are more desirable than others. The operation of heat engines depends on the transfer of heat between two temperatures, so thermal energy at a uniform temperature is useless for performing work. This is often described by saying that the generation of heat results in the degradation of energy to a less useful form.

Heat engines are devices that operate in cycles to convert heat into work, and their efficiency is defined as the ratio of the work obtained from the engine to the heat absorbed by the engine from its environment during a cycle. A cycle is merely a sequence of events that starts with a system in a particular state and ends up with the system in the same state.

It is clear from the Kelvin statement of the second law that the efficiency of heat engines must be less than unity. But it turns out that there is a maximum efficiency that is the same for all heat engines, regardless of the material nature of the engine. This maximum efficiency can be found by considering a material system that can accept heat from its surroundings and use it to perform work by going through an idealized cycle called the *Carnot cycle.*

For specificity, we consider a system consisting of a gas in a cylinder with a piston that can be placed in contact with heat reservoirs. Thus, heat can flow into and out of the system and the system can do work. Our conclusions, however, are valid for any kind of system that can be surrounded by heat sources or sinks and for any kind of work, not just PV work. That this is true will be obvious from an inspection of the steps in the cycle. There are four steps:

1. Start with the system that is in thermal equilibrium with a heat bath at a temperature T_2. Label this initial state A. (For a gas, the state is determined by the pressure, volume, and temperature, P_2, V_2, T_2.) Let the system do work reversibly to bring it to a new state labeled B. The temperature is still T_2, but work has been done by the system, so an amount of heat Q_2 had to be absorbed from the reservoir. (For a gas, this is a reversible isothermal expansion.) The energy change for this process is $\delta E_1 = |Q_2| - |W_1|$. To emphasize that work done on the system is negative and heat absorbed by the system is positive, this equation was written in terms of absolute values.
2. Now take the system in the state B, surround it with insulating walls, and let it do an amount of work W_2 on its surroundings reversibly. This brings the system to a new state C. (For a gas, this amounts to a reversible adiabatic expansion.) No heat is transferred, so the energy change is just the work done by the system: $\delta E_2 = -|W_2|$. Since no heat is transferred, the temperature must fall to some new value T_1.
3. Put the system in state C in contact with a heat bath at temperature T_1 and remove the insulating wall. Now do an amount of work $-W_3$ on the system reversibly and isothermally. (For a gas, this is a reversible isothermal compression.) An amount of heat Q_1 is thereby transferred to the heat bath and the first law requires that $\delta E_3 = -|Q_1| + |W_3|$.
4. To complete the cycle, the system must be brought back to its original state. This can be done by surrounding it with an insulating wall and doing an amount of reversible work that brings it back to its original temperature T_2. (For a gas, this is a reversible adiabatic compression such that the final pressure and volume are equal to the original pressure and volume so that, because of the equation of state, the final temperature is T_2.) The energy change for this step is $\delta E_4 = |W_4|$.

The sum of the energies for each of the four steps described above step must add up to zero since the energy is a state function. Thus, $|Q_2| - |W_1| - |W_2| - |Q_1| + |W_3| + |W_4| = 0$. The total amount of work done by the system during this cycle is $|W| = |W_1| + |W_2| - |W_3| - |W_4|$, so

$$W = |Q_2| - |Q_1| \tag{1.10.1}$$

That is, the amount of work done is equal to the amount of heat extracted from the reservoir at T_2 minus the amount given up to the reservoir at T_1.

The efficiency η of the cycle is the ratio of work done to heat taken from the reservoir. That is,

$$\eta = \frac{W}{|Q_2|} = \frac{|Q_2| - |Q_1|}{|Q_2|} = 1 - \frac{|Q_1|}{|Q_2|} \tag{1.10.2}$$

The temperatures T_2 and T_1 of the two heat reservoirs were written as if they were from the ideal gas temperature scale. In fact, they do not appear in the final result of the Carnot cycle. But a temperature scale that is independent of any material can be defined in purely thermodynamic terms by recognizing that the ratio of the heats in (1.10.2) is a function of the temperatures of the two reservoirs. By considering two Carnot cycles, each with the same lower temperature, it can be shown that

$$\frac{|Q_1|}{|Q_2|} = \frac{f(T_1)}{f(T_2)} \tag{1.10.3}$$

where $f(T)$ is some function of the temperature.

The absolute temperature scale is defined by requiring that the function $f(T)$ is just the thermodynamic temperature such that

$$\frac{|Q_1|}{|Q_2|} = \frac{T_2}{T_1} \tag{1.10.4}$$

This defines a thermodynamic temperature scale that is independent of any particular material. The same symbol T is used here as in the definition of the ideal gas temperature scale. In fact, if the Carnot cycle is carried out using an ideal gas as the working substance, then the ratio of the heats is indeed equal to the ratio of the ideal gas temperatures. The thermodynamic absolute temperature is therefore equal to the ideal gas temperature. The reason for this equality is that in an ideal gas there are no intermolecular interactions and therefore nothing that identifies the differences among different substances. Note that using (1.10.4) in (1.10.2) gives the maximum thermodynamic efficiency as it is usually written: $\eta = 1 - (T_1/T_2)$. Let us rewrite (1.10.4) as

$$\frac{|Q_2|}{|T_2|} - \frac{|Q_1|}{|T_1|} = 0 \tag{1.10.5}$$

or

$$\frac{Q_2}{T_2} + \frac{Q_1}{T_1} = 0 \tag{1.10.6}$$

where the removal of the absolute value signs defines positive heat as being heat absorbed by the system.

Since the heat absorbed in going from one state to another depends on the path, it is not a state function. But (1.10.6) can be used to define a state function that includes the concept of heat. To do this, consider any cycle that brings a system through a succession of states that returns the system to its original state. This is not necessarily a Carnot cycle, but simple construction shows that it can be approximated by a succession of many small Carnot cycles, for each of which (1.10.6) is true. By taking the limit of an infinite number of Carnot cycles, it is easy to show that the integral of dQ/T over the cycle is zero, where dQ is the infinitesimal amount of heat reversibly transferred to the system at temperature T. Thus, if the cycle is separated into two parts such that the system first goes from state A to state B and then from state B to state A, we have

$$\int_A^B \frac{dQ}{T} + \int_B^A \frac{dQ}{T} = 0 \tag{1.10.7}$$

This means that no matter how we get from A to B, the result is always the same integral. That is, integrating over any path from A to B always gives the negative of the integral from B to A. Since both paths are arbitrary, the integrals must both be path independent. The integral is therefore a state function. This is the definition of the entropy:

$$S_B - S_A = \int_A^B \frac{dQ}{T} \tag{1.10.8}$$

The importance of the entropy arises from two basic (and related) attributes: the first is that the total entropy change in physical processes is always positive, and the second is that this fact can be used to determine the conditions required for a system to be in thermodynamic equilibrium. To show this, we start with the fact that the efficiency of the transformation of a given amount of heat into work is a maximum for a reversible Carnot cycle. The reason for this is that in a Carnot cycle the work done by the system in steps 1 and 2 is a maximum and the work done on the system in steps 3 and 4 is a minimum because the work is done reversibly. Thus, the work done by the system in a Carnot cycle is the maximum that can be done in bringing the system from its initial state to its final state.

Since the energy change of the system is zero for the cycle, the heat absorbed by the system is a minimum. Remember that this minimum is relative to all the irreversible paths in going through a cycle in which heat is absorbed and work is done in a system such that the system ends up in its initial state. Thus, the maximum thermodynamic efficiency is

$$\frac{T_2 - T_1}{T_2} = \left.\frac{Q_2 + Q_1}{Q_2}\right|_{\text{rev}} > \left.\frac{Q_2 + Q_1}{Q_2}\right|_{\text{irrev}} \tag{1.10.9}$$

where the first equality comes from equation (1.10.4). From (1.10.9), it follows that

$$\left|\frac{Q_1}{T_1} + \frac{Q_2}{T_1}\right| \leq 0 \tag{1.10.10}$$

where the equality holds for a reversible process and the inequality holds for an irreversible process. Since a cycle is irreversible, if any part of it is irreversible, it follows from (1.10.10) that

$$\oint \frac{dQ}{T} \leq 0 \tag{1.10.11}$$

where the integral is taken over the entire cycle. Now consider the cycle as the sum of two processes, the first of which takes the system from its starting state A to some state B, and the second of which takes the system from state B back to state A. Then, (1.10.11) is

$$\int_A^B \frac{dQ}{T} + \int_B^A \frac{dQ}{T} \leq 0 \tag{1.10.12}$$

Now assume that the return part of the cycle, from B to A, is done reversibly. Then the second integral in (1.10.12) is just the entropy change of the system in going from state B to state A and (1.10.12) becomes

$$S_B - S_A \geq \int_A^B \frac{dQ}{T} \tag{1.10.13}$$

Equation (1.10.13) is equivalent to

$$\delta S \geq \frac{dQ}{T} \tag{1.10.14}$$

That is, the entropy difference connecting two infinitesimally close states of a system is equal to the heat absorbed by the system divided by the temperature if the heat is transferred reversibly, and is greater than this if the heat is transferred irreversibly.

Note that the first law requires that $\delta Q = \delta U + \delta W$, so because of (1.10.14) the entropy change at constant energy and volume is

$$\delta S \geq \frac{\delta U + \delta W}{T} \tag{1.10.15}$$

This equation is the basis of the equilibrium conditions of thermodynamics. If, for example, the only work involved is pressure–volume work, then it follows from (1.10.15) that

$$\delta S_{U,V} \geq 0 \tag{1.10.16}$$

where the equality holds for a reversible process and the inequality holds for an irreversible process.

Equation (1.10.16) states that the entropy for a system with constant energy and constant volume is a maximum. Of course, if other forms of work are involved, then (1.10.16) becomes

$$\delta S_{U,\Phi} \geq 0 \tag{1.10.17}$$

where Φ represents all the extensive parameters for all the types of work done by the system.

1.11 Free energies and equilibrium conditions

From (1.10.14), the heat change in a reversible process is related to the entropy by $dQ = TdS$, so the first law can be written as

$$dU = TdS - dW_r \tag{1.11.1}$$

where dW_r is the work done *reversibly* in increasing the energy of the system by dU. For an isothermal process that brings the system from state A to state B, integration of equation (1.11.1) gives $-W_r\,(A \rightarrow B) = (U_B - U_A) - (TS_B - TS_A)$ (isothermal process). This leads us to define a function A by

$$A = U - TS \tag{1.11.2}$$

A is called the Helmholtz free energy, and ΔA is the maximum available work when a system goes from one state to another via an isothermal process.

The Gibbs free energy function is defined by

$$G = A + PV = H - TS \tag{1.11.3}$$

It is related to the maximum work that can be done by the system in a constant temperature, constant volume process. To see this, take the differential of (1.11.3) at constant pressure to get

$$dG_P = dA_P + PdV \tag{1.11.4}$$

If we also require that the change in state is isothermal, then since the Helmholtz free energy is the negative of the maximum (reversible) work done by the system in an isothermal process, (1.11.4) becomes

$$-dG_{T,P} = dW_r - PdV \tag{1.11.5}$$

so the change in Gibbs free energy at constant temperature and volume is equal to the maximum available work done by the system, exclusive of pressure–volume work.

The Helmholtz and Gibbs free energies both provide very useful criteria for a system to be at equilibrium. The fundamental equilibrium condition is given by equation (1.10.17) [or when only P–V work is present by equation (1.10.16)]. This states that the entropy is a maximum for any process in which the energy and the extensive work parameters (such as volume) are constant. In practice, most thermodynamic processes occur at constant temperature and pressure, or at constant temperature and volume. It is convenient, therefore, to have equilibrium criteria for such processes. The free energies provide such criteria.

If we consider only reversible processes so that differentials of thermodynamic functions connect two equilibrium states, then in going from one state to the other, any heat involved is transferred reversibly. That is, $dQ = TdS$. The equilibrium condition (1.10.16) then becomes

$$dS_{U,V} = 0 \tag{1.11.6}$$

For convenience, we restrict ourselves to processes involving PV work only; the generalization to include other types of work will generally be obvious. Then, since we consider only reversible processes, the first law becomes

$$dU = TdS - PdV \tag{1.11.7}$$

from which it follows that

$$dU_{S,V} = 0 \tag{1.11.8}$$

From the definition of the Helmholtz free energy given by (1.11.2), we get $dA = dU - TdS - SdT$, or, using (1.11.7),

$$dA = -PdV - SdT \tag{1.11.9}$$

so

$$dA_{V,T} = 0 \tag{1.11.10}$$

Now start with the definition of the Gibbs free energy given by equation (1.11.3) to get $dG = dU - TdS - SdT + PdV + VdP$, and again use the first law to reduce this to

$$dG = -SdT + VdP \tag{1.11.11}$$

At constant temperature and pressure this gives

$$dG_{T,P} = 0 \tag{1.11.12}$$

Equations (1.11.6), (1.11.8), (1.11.10), and (1.11.12) are all conditions for equilibrium. For completeness, the enthalpy can also yield an equilibrium condition, and in fact, it is easy to show that

$$dH_{S,P} = 0 \tag{1.11.13}$$

1.12 Thermodynamic potentials and Legendre transformations

Since equation (1.11.7) arises directly from the combination of the first and second laws, the energy is said to be a natural function of entropy and volume, so the thermodynamic state of the system is defined as a relation among the energy, volume, and entropy:

$$U = U(S, V) \tag{1.12.1}$$

Taking the differential of (1.12.1) gives

$$dU = \left(\frac{\partial U}{\partial S}\right)_V dS + \left(\frac{\partial U}{\partial V}\right)_S dV \tag{1.12.2}$$

Comparing this with (1.11.7) gives

$$\left(\frac{\partial U}{\partial S}\right)_V = T, \quad \left(\frac{\partial U}{\partial V}\right)_S = -P \tag{1.12.3}$$

The enthalpy, Helmholtz free energy, and Gibbs free energy were defined in the preceding section as natural consequences of considering simple cases of heat transfer and work done by a system. These quantities, along with the energy and entropy, are called *thermodynamic potentials*, and their importance lies in the fact that each of them is a natural function of a different pair of independent variables that can be connected to differing experimental conditions. The adjective "natural" means that the relationships between the thermodynamic potential and the independent variables arise directly from the connections of the potential to work or heat processes. It is, of course, possible to express a potential in terms of variables other than the natural ones, but the differentials of the potentials take their simplest form when expressed in terms of the natural independent variables. The notation adopted for the partial derivatives recognizes that the potentials can be expressed as functions of different sets of variables by explicitly specifying them. Thus, in equation (1.12.3), the partial derivative that defines the temperature clearly refers to the energy as a function of entropy and volume as the independent variables.

But entropy and volume are not convenient variables for describing experiments, so we are led to look for other descriptions. The Helmholtz free energy, for example, is most useful for describing processes at constant temperature and volume, while the Gibbs free energy is naturally connected to constant pressure, constant temperature conditions as shown by (1.11.9) and (1.11.11).

The natural variables for the enthalpy are readily obtained from the energy equation and equation (1.9.7) defining the enthalpy. Taking the differential of (1.9.7) and combining the result with (1.11.7) for the differential of the energy gives

$$dH = TdS + VdP \tag{1.12.4}$$

so the natural variables for the enthalpy are entropy and pressure.

The expression of the potentials in terms of their natural variables immediately yields thermodynamic relations analogous to (1.12.3). From (1.12.4)

$$\left(\frac{\partial H}{\partial S}\right)_P = T, \quad \left(\frac{\partial H}{\partial P}\right)_S = V \tag{1.12.5}$$

while (1.11.9) and (1.11.11) yield

$$\left(\frac{\partial A}{\partial V}\right)_T = -P, \quad \left(\frac{\partial A}{\partial T}\right)_V = -S \tag{1.12.6}$$

$$\left(\frac{\partial G}{\partial T}\right)_P = -S, \quad \left(\frac{\partial G}{\partial P}\right)_T = V \tag{1.12.7}$$

These equations lead to expressions for the free energies that are extremely useful. If the second equation in (1.12.6) is substituted for the entropy in the definition of the Helmholtz free energy given by (1.11.2), the result is

$$A = U + T\left(\frac{\partial A}{\partial T}\right)_V \tag{1.12.8}$$

Similarly, if the first of equations (1.12.7) is used to replace the entropy in the definition of the Gibbs free energy, (1.11.3), then

$$G = H + T\left(\frac{\partial G}{\partial T}\right)_P \tag{1.12.9}$$

Equations (1.12.8) and (1.12.9) are the Gibbs–Helmholtz equations for the Helmholtz and Gibbs free energies, respectively.

In statistical mechanical applications it is sometimes more convenient to express heat capacities in terms of derivatives of free energies rather than energy or enthalpy. This is easily done by using the Gibbs–Helmholtz equations. Differentiating (1.12.8) with respect to temperature at constant volume gives

$$C_V = -T\left(\frac{\partial^2 A}{\partial T^2}\right)_V \tag{1.12.10}$$

Similarly, differentiating (1.12.9) with respect to temperature at constant pressure gives

$$C_P = -T\left(\frac{\partial^2 G}{\partial T^2}\right)_P \tag{1.12.11}$$

While the definitions of the potentials are directly connected to work and heat changes, they are also simply described as mathematical transformations of each other that all describe the same thermodynamic information. To see this, start with the energy equation (1.11.7) for a system in which only PV work is done when a system undergoes an infinitesimal reversible change, and consider a system with a given, fixed volume whose energy can vary through heat transfer. The possible states of the system are then determined by the relation between energy and entropy and can be represented by a two-dimensional plot of energy versus entropy, as shown in figure 1.1. The curve represents the energy–entropy relation for the system, while the straight line is a tangent to the line at a given point. Clearly, if every tangent to the curve is specified, then the energy–entropy curve is determined. That is, giving the slope and intercept of the tangent to every point of the curve is fully equivalent to specifying the curve itself.

For the given point with energy U, the slope $(\partial U/\partial S)_V = T$ is given by $(\partial U/\partial S)_V = (U - b)/S$, where b is the intercept. Solving for the intercept gives

$$b = U - \left(\frac{\partial U}{\partial S}\right)_V S \tag{1.12.12}$$

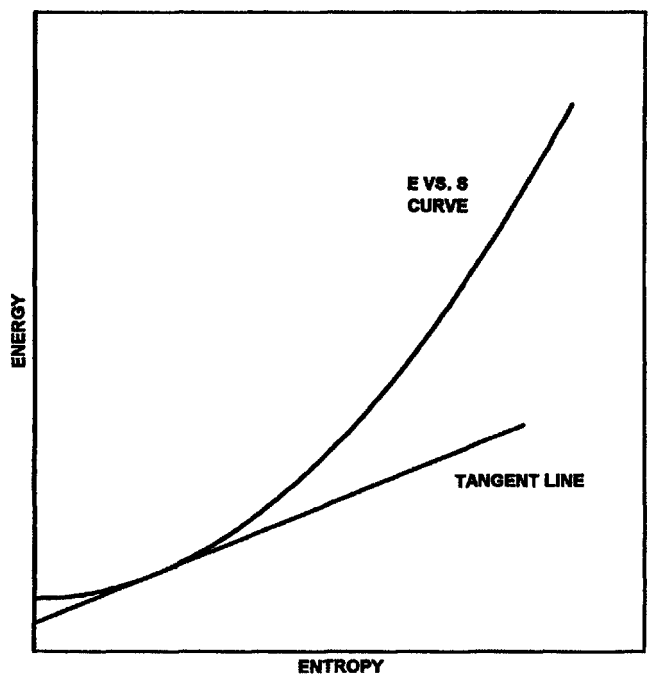

Figure 1.1. Illustration of Legendre transformation.

Equation (1.12.12) is an nple of the Legendre transformation, which expresses a function in terms of its derivative. Since the derivative in (1.12.12) is the temperature, the intercept is just the Helmholtz free energy, which we see is a Legendre transformation of the energy.

The generalization to more than one variable is straightforward. In general, if $F = F(X_1, X_2, \ldots, X_t)$ is a function of t variables with derivatives $p_j = \partial F/X_j$, the Legendre transformation is defined by

$$b_F(X_1, X_2, \ldots) = F - \sum_j p_j X_j \tag{1.12.13}$$

The transform $b_F(X_1, X_2, \ldots)$ is now the intercept of lines that are parallel to the surface defined by the function F in a t-dimensional space. It is the transform of the function F with respect to the variables $(X_1, X_2, \ldots)$. Equation (1.12.13) may be called the *complete* Legendre transformation in that all the variables were transformed. This is indicated by writing the transform as a function of all the variables. A partial Legendre transformation is one in which only some of the variables are transformed. The Helmholtz free energy, for example, is a partial transform of the energy with respect to the entropy.

The partial Legendre transform of the energy with respect to the volume is

$$b_U(V) = U - \left(\frac{\partial U}{\partial V}\right)_S V \tag{1.12.14}$$

But the derivative in (1.12.14) is just the negative of the pressure [equation (1.12.3)], so

$$b_U(V) = U + PV \tag{1.12.15}$$

which is the definition of the enthalpy. Finally, if a Legendre transform of the energy with respect to entropy and volume is performed, the result is

$$b_U(S, V) = U - \left(\frac{\partial U}{\partial S}\right)_V S - \left(\frac{\partial U}{\partial V}\right)_S V \tag{1.12.16}$$

But from (1.12.3) the first derivative on the right of (1.12.16) is the temperature, and the other derivative is the negative of the pressure, so

$$b_U(S, V) = U - TS + PV \tag{1.12.17}$$

which is just the Gibbs free energy. The four thermodynamic potentials are therefore related by Legendre transformations. Other potentials can be defined by additional Legendre transforms, but these four are the most useful. A simple scheme describing them is given in the following diagram:

$$\begin{array}{ccc} S & + & U & - & V \\ + & & & & - \\ H & & & & A \\ + & & & & - \\ P & + & G & - & T \end{array}$$

where each function has its independent variables on either side of it, with the appropriate plus or minus sign indicated.

1.13 Chemical potentials

Up to this point, we have assumed that our system was closed. That is, there was no transfer of matter during any change from one equilibrium state to another, so the dependence of energy on composition did not have to be made explicit. But there are many cases in which the composition changes during a thermodynamic process; the system is then an open system in that matter can pass in and out of it. The composition dependence of the energy must then be included explicitly and equation (1.12.1) must then be written as

$$U = U(S, V, N_1, N_2, \ldots, N_C) \tag{1.13.1}$$

This equation refers to a system that has c components, N_j being the number of molecules of the jth component. The differential form of (1.13.1) is

$$dU = \left(\frac{\partial U}{\partial S}\right)_{V,N_j} dS + \left(\frac{\partial U}{\partial V}\right)_{S,N_j} dV + \sum_i \left(\frac{\partial U}{\partial N_i}\right)_{S,V,N_{j\neq i}} dN_i \tag{1.13.2}$$

The first two derivatives are at constant composition, as well as constant volume and entropy, respectively, while the derivative in the ith term of the sum is at constant entropy, volume, and number of molecules of every type except for the ith. To be consistent with (1.11.7), the derivatives in the first two terms must still be given by temperature and negative pressure as in equation (1.12.3). Also, the derivatives with respect to the number of molecules are given their own symbol μ_i, so we have

$$\left(\frac{\partial U}{\partial S}\right)_{V,N_j} = T \tag{1.13.3}$$

$$\left(\frac{\partial U}{\partial V}\right)_{S,N_j} = -P \tag{1.13.4}$$

$$\left(\frac{\partial U}{\partial N_i}\right)_{S,V,j(\neq i)} = \mu_i \tag{1.13.5}$$

where μ_i is called the chemical potential of the ith component. We rewrite (1.13.2) as

$$dU = TdS - PdV + \sum_i \mu_i dN_i \tag{1.13.6}$$

The other thermodynamic potentials are easily generalized to include the composition variables and the chemical potentials. Combining (1.13.6) first with (1.12.4), then with (1.11.9), and then with (1.11.11) gives

$$dH = TdS + VdP + \sum_i \mu_i dN_i \tag{1.13.7}$$

$$dA = -SdT - PdV + \sum_i \mu_i dN_i \tag{1.13.8}$$

$$dG = -SdT + VdP + \sum_i \mu_i dN_i \tag{1.13.9}$$

Note that the chemical potential is given by partial derivatives of each of the potentials as

$$\mu_i = \left(\frac{\partial U}{\partial N_i}\right)_{S,V,N_{j\neq i}} = \left(\frac{\partial H}{\partial N_i}\right)_{S,P,N_{j\neq i}} = \left(\frac{\partial A}{\partial N_i}\right)_{T,V,N_{j\neq i}} = \left(\frac{\partial G}{\partial N_i}\right)_{T,P,N_j} \tag{1.13.10}$$

1.14 Conditions of phase equilibria and stability

Equation (1.13.6) is valid for any system; it gives the energy change when a system goes from one state to another by a reversible process, the two states differing infinitesimally in their entropy, volume, and composition. Solving for the entropy gives

$$dS = \frac{dU}{T} + \frac{PdV}{T} - \sum_i \frac{\mu_i}{T} dN_i \tag{1.14.1}$$

Since (1.14.1) is valid for any system, it is valid for a system that consists of two phases. Let one phase be labeled by a superscript A and the other by a superscript B, and consider a process in which the energy, volume, and composition of phases A and B change because of transfers of matter and energy between the two phases. Then, if we assume the total system is isolated so no heat or work is transferred to it, the total entropy vanishes and (1.14.1) becomes

$$\begin{aligned} dS &= dS^A + dS^B \\ &= \frac{dU^A}{T^A} + \frac{P^A dV^A}{T^A} - \sum_i \frac{\mu_i^A}{T^A} dN_i^A + \frac{dU^B}{T^B} + \frac{P^B dV^B}{T^B} - \sum_i \frac{\mu_i^B}{T^B} dN_i^B \\ &= 0 \end{aligned} \tag{1.14.2}$$

But since the system as a whole is isolated, we must also have

$$\begin{aligned} dU^A &= -dU^B \\ dV^A &= -dV^B \\ dN_i^A &= -dN_i^B \end{aligned} \tag{1.14.3}$$

Using (1.14.3) in (1.14.2) gives

$$\left(\frac{1}{T^A} - \frac{1}{T^B}\right) dU^A + \left(\frac{P^A}{T^A} - \frac{P^B}{T^B}\right) dV^A - \sum_i \left(\frac{\mu_i^A}{T^A} - \frac{\mu_i^B}{T^B}\right) dN_i^A = 0 \tag{1.14.4}$$

But the variations in energy, volume, and composition are independent, so each term in (1.14.4) must vanish, and therefore

$$T_A = T_B \tag{1.14.5}$$

$$P_A = P_B \tag{1.14.6}$$

$$\mu_i^A = \mu_i^B \tag{1.14.7}$$

This result can obviously be extended to any number of phases, so for internal equilibrium of a system consisting of c components distributed among p phases, the temperature must be the same in every phase (thermal equilibrium), the pressure must be the same in every phase (hydrostatic equilibrium), and the chemical potential of each component must be the same in every phase (chemical equilibrium).

From the mode of their derivation, the conditions of equality of temperatures, pressures, and chemical potentials obviously apply to a single-phase system as well as multiphase systems. For a single-phase system these conditions state that, for equilibrium, the temperature, pressure, and chemical potentials must each be constant throughout the phase.

The phase rule addresses the following question: for a multiphase system, how many thermodynamic variables are needed to specify the state of the system? Since the most useful thermodynamic independent variables for laboratory work are temperature and pressure, it is most convenient to use the Gibbs free energy to answer this question. The equilibrium condition in terms of the Gibbs free energy is that the differential given by (1.13.9) must vanish for a reversible process. The Gibbs free energy change for the Kth phase is given by an equation just like (1.13.9):

$$dG^K = -S^K dT + V^K dP + \sum_{i=1}^{c} \mu_i^K dN_i^K \tag{1.14.8}$$

The entropy and volume are extensive quantities, and the temperature and pressure are the same for every phase, so summing (1.14.8) over all phases gives the total Gibbs free energy differential for the multiphase system. The total free energy change for the entire system must vanish because it is isolated, so

$$dG = -SdT + VdP + \sum_{K=1}^{p} \sum_{i=1}^{c} \mu_i^K dN_i^K = 0 \tag{1.14.9}$$

This equation defines the equilibrium state of the system in that, if the temperature, pressure, and composition variables are given, the state is, determined. There are c composition variables for each phase, giving a total of $cp + 2$ variables including the temperature and pressure. But these variables are not all independent. In the first place, because the system as a whole is closed, the total number of molecules of each component is fixed, so the total variation over all phases for each component must vanish. That is,

$$\sum_{K=1}^{p} dN_i^K = 0 \tag{1.14.10}$$

There are p such equations, one for each phase so this imposes p conditions on the variables. Furthermore, the equality of chemical potentials for a given component among phases gives

$$\mu_i^1 = \mu_i^2 = \mu_i^3 \ldots = \mu_i^p \tag{1.14.11}$$

This is a set of $(p - 1)$ equations for each component, so (1.14.11) represents an additional $c(p - 1)$ conditions that must be satisfied. If we call the difference between the number of variables and the number of conditions f, then

$$f = (cp+2) - p - c(p-1)$$
$$= c - p + 2 \tag{1.14.12}$$

f is called the number of *degrees of freedom* of the system since it is the number of variables that may be independently varied. This is the famous phase rule of Gibbs.

Equations (1.14.5)–(1.14.7) are necessary for equilibrium, but not sufficient. They arise from the requirement that the first variation of the entropy with respect to any possible variations of the system variables is equal to, or greater than, zero. But in order for this to represent a maximum, the second differential of the entropy must be negative. That is,

$$d^2S < 0 \tag{1.14.13}$$

or, using (1.14.1),

$$d(dS) = \frac{\partial(dS)}{\partial U}dU + \frac{\partial(dS)}{\partial V}dV - \sum_j \frac{\partial(dS)}{\partial N_j}dN_j < 0 \tag{1.14.14}$$

$$\begin{aligned} d^2S = {} & \frac{\partial}{\partial U}\left(\frac{1}{T}\right)d^2U + \frac{\partial}{\partial U}\left(\frac{P}{T}\right)dVdU - \sum_j \frac{\partial}{\partial U}\left(\frac{\mu_j}{T}\right)dN_j dU \\ & + \frac{\partial}{\partial V}\left(\frac{1}{T}\right)dVdU + \frac{\partial}{\partial V}\left(\frac{P}{T}\right)d^2V - \sum_j \frac{\partial}{\partial V}\left(\frac{\mu_j}{T}\right)dN_j dU \\ & + \sum_i \frac{\partial}{\partial N_i}\left(\frac{1}{T}\right)dN_i dU + \sum_i \frac{\partial}{\partial N_i}\left(\frac{P}{T}\right)dN_i dV - \sum_{i,j} \frac{\partial}{\partial N_i}\left(\frac{\mu_j}{T}\right)dN_j dN_i < 0 \end{aligned} \tag{1.14.15}$$

If (1.14.11) is to be true, then if the composition and volume are held constant, the first term in (1.14.15) must be less than zero. That is,

$$\left[\frac{\partial}{\partial U}\left(\frac{1}{T}\right)\right]_{V,N} < 0 \tag{1.14.16}$$

But

$$\left[\frac{\partial}{\partial U}\left(\frac{1}{T}\right)\right]_{V,N} = -T^2\left(\frac{\partial T}{\partial U}\right)_{V,N} = -\frac{T^2}{C_V} < 0 \tag{1.14.17}$$

because the derivative in the middle term in (1.14.17) is the reciprocal of the heat capacity at constant volume. Thus, the condition for stable thermal equilibrium is that the heat capacity is positive:

$$C_V > 0 \tag{1.14.18}$$

By considering the case of constant energy and composition, the condition for mechanical equilibrium is easily shown to be

$$\left[\frac{\partial}{\partial V}\left(\frac{P}{T}\right)\right]_{U,N_j} < 0 \tag{1.14.19}$$

By keeping all compositions constant except one, as well as the energy and volume, the condition for chemical equilibrium is found to be

$$\sum_{i,j}\left[\frac{\partial}{\partial N_j}\left(\frac{\mu_i}{T}\right)\right]_{U,V,N_{i\neq j}} dN_i dN_j > 0 \tag{1.14.20}$$

By starting with the Helmholtz free energy (1.13.8), and requiring the second differential to be positive (because the equilibrium condition requires the free energy to be a minimum), a similar derivation gives somewhat more useful forms for the conditions of mechanical and chemical equilibrium as

$$\left(\frac{\partial P}{\partial V}\right)_{T,N_j} < 0 \tag{1.14.21}$$

$$\sum_{i,j}\frac{\partial \mu_i}{\partial N_j} dN_i dN_j > 0 \tag{1.14.22}$$

Equations (1.14.18) and (1.14.21) are obviously true on physical grounds. If the temperature is raised, the system must gain energy, and if the pressure is increased, the volume of the system decreases (the isothermal compressibility is positive).

If the conditions arising from both the first and second differentials of the entropy (or free energy) are both satisfied, the system is in equilibrium. These are *local* conditions in that they are the result of considering changes in the system close to the maximum of the entropy function, or the minimum of the free energy. They do not guarantee that the system is in the lowest of all free energy states (or maximum of all entropy states). It is entirely possible for a system to have a number of states for which the entropy is a maximum and the free energy a minimum. There is always at least one state for which the entropy is higher, and the free energy smaller, than all the others. These are called states of *absolute*, or *stable, equilibrium*. The states for which all equilibrium conditions are satisfied, but which are not states of the highest entropy or lowest free energy, are called *metastable states*.

1.15 Euler's theorem and the Gibbs–Duhem equation

A major task in thermodynamics is to find the relationships among measurable physical quantities. These are needed to enable the enormous amount of possible experimental results to be described in a rational, ordered theoretical structure. Also, thermodynamics places restrictions on the possible relations among physical quantities; if experimental results violate these restrictions, they must be in error. Furthermore, a knowledge of the thermodynamic relations is necessary if any microscopic theory is to be used to understand macroscopic systems. There are a number of simple mathematical results that allow us to connect the basic laws of thermodynamics to physical properties and to obtain the desired relations among them.

The energy is a natural function of entropy, volume, and composition. These are all extensive variables, so if the number of molecules of each component in a phase increases by some fraction, the entropy volume and energy both increase by the same fraction. That is, if each composition variable is multiplied by the same number λ, then

$$\lambda U(S, V, N_1, N_2, \ldots, N_c) = U(\lambda S, \lambda V, \lambda N_1, \lambda N_2, \ldots, \lambda N_c) \tag{1.15.1}$$

That is, the energy is a homogeneous function of its natural variables of order unity.

The Euler theorem for homogeneous functions can therefore be applied to (1.15.1) to yield very useful results. This theorem states that, given a homogeneous function of order n of variables x_j, which is defined by

$$\lambda^n f(x_1, x_2, x_3, \ldots) = f(\lambda x_1, \lambda x_2, \lambda x_3, \ldots), \tag{1.15.2}$$

then

$$nf(x_1, x_2, x_3, \ldots) = x_1 \frac{\partial f}{\partial x_1} + x_2 \frac{\partial f}{\partial x_2} + x_3 \frac{\partial f}{\partial x_3} \cdots \tag{1.15.3}$$

To prove this, differentiate (1.15.2) with respect to λ to get

$$\begin{aligned} n\lambda^{n-1} f(x_1, x_2, x_3, \ldots) &= \frac{\partial f}{\partial(\lambda x_1)} \frac{\partial(\lambda x_1)}{\partial \lambda} + \frac{\partial f}{\partial(\lambda x_2)} \frac{\partial(\lambda x_2)}{\partial \lambda} \cdots \\ &= x_1 \frac{\partial f}{\partial(\lambda x_1)} + x_2 \frac{\partial f}{\partial(\lambda x_2)} \cdots \end{aligned} \tag{1.15.4}$$

which, upon setting $\lambda = 1$, gives (1.15.3). For $n = 1$, applying this to the energy, we get

$$\begin{aligned} &U(S, V, N_1, N_2, \ldots, N_c) \\ &\quad = S\left(\frac{\partial U}{\partial S}\right)_{V,N} + V\left(\frac{\partial U}{\partial V}\right)_{S,N} + \sum_{j=1}^{c} N_j \left(\frac{\partial U}{\partial N_j}\right)_{V,N(\neq j)} \end{aligned}$$

Using the relation of the derivatives to temperature, pressure, and chemical potential given by equations (1.13.3)–(1.13.5), this becomes

$$U = TS - PV + \sum_{j=1}^{c} \mu_j N_j \tag{1.15.5}$$

Thus, the Euler theorem yields an integration of the energy, equation.

The integrated forms of the enthalpy, Helmholtz free energy, and Gibbs free energy in terms of composition are easily obtained by using (1.15.5) in equations (1.9.7), (1.11.2), and (1.11.3) to get

$$H = TS + \sum_{j=1}^{c} \mu_j N_j \tag{1.15.6}$$

$$A = -PV + \sum_{j=1}^{c} \mu_j N_j \tag{1.15.7}$$

$$G = \sum_{j=1}^{c} \mu_j N_j \tag{1.15.8}$$

These equations are quite general and apply to any homogeneous phase of any number of components. From (1.15.8),

$$dG = \sum_{j=1}^{c} \mu_j dN_j + \sum_{j=1}^{c} N_j d\mu_j \tag{1.15.9}$$

which, when combined with (1.13.9), gives

$$-SdT + VdP + \sum_{j=1}^{c} N_j d\mu_j = 0 \tag{1.15.10}$$

This is the Gibbs–Duhem equation.

If the temperature and pressure are kept constant, then equilibrium requires the Gibbs free energy change to be constant and (1.15.10) reduces to

$$\sum_{j=1}^{c} N_j d\mu_j = 0 \qquad \text{(constant } P, T\text{)} \tag{1.15.11}$$

Equation (1.15.11) is the Gibbs–Duhem equation at constant temperature and pressure. It is valid for any process involving an infinitesimal change in the composition of a phase but carried out at constant temperature and pressure, and is very useful for analysis of the equilibria of solutions and chemical reactions.

1.16 Reciprocity relations of Maxwell

The differentials of the thermodynamic potentials given by equations (1.13.6)–(1.13.9) all have the form

$$dF = MdX + NdY + \sum_{i=1}^{c} \mu_i dN_j \tag{1.16.1}$$

where dF is a perfect differential and (M, N, μ_i) are all functions of (X, Y, N_i). That is, the integral of dF is independent of the path of integration, so F is a state function. The coefficients of the differentials in (1.16.1) are therefore

$$M = \left(\frac{\partial F}{\partial X}\right)_{Y,N}, \quad N = \left(\frac{\partial F}{\partial Y}\right)_{X,N}, \quad \mu_i = \left(\frac{\partial F}{\partial N_i}\right)_{X,Y,N(\neq i)} \tag{1.16.2}$$

For simplicity of notation, let us consider a phase of constant composition, with all $dN_i = 0$, so that (1.16.1) and (1.16.2) reduce to

$$dF = MdX + NdY \tag{1.16.3}$$

$$M = \left(\frac{\partial F}{\partial X}\right)_{Y}, \quad N = \left(\frac{\partial F}{\partial Y}\right)_{X} \tag{1.16.4}$$

Since the order of differentiation is immaterial, differentiating the first equation in (1.16.4) with respect to Y and the second with respect to X gives the same result, so

$$\left(\frac{\partial M}{\partial Y}\right)_{X} = \left(\frac{\partial N}{\partial X}\right)_{Y} \tag{1.16.5}$$

[Note that there is no loss of generality by considering a system of constant composition. If the composition is allowed to vary, we simply specify that the differentiation is carried out at constant composition in (1.16.5).]

If (1.16.5) is applied to the differentials of the potentials in (1.13.6)–(1.13.9), we get that

$$\left(\frac{\partial T}{\partial V}\right)_S = -\left(\frac{\partial P}{\partial S}\right)_V \tag{1.16.6}$$

$$\left(\frac{\partial T}{\partial P}\right)_S = \left(\frac{\partial V}{\partial S}\right)_P \tag{1.16.7}$$

$$\left(\frac{\partial S}{\partial V}\right)_T = \left(\frac{\partial P}{\partial T}\right)_V \tag{1.16.8}$$

$$\left(\frac{\partial S}{\partial P}\right)_T = -\left(\frac{\partial V}{\partial T}\right)_P \tag{1.16.9}$$

These are the Maxwell reciprocity relations. They have many applications in finding relations among thermodynamic quantities, such as those that are important for chemical and phase equilibria and in relating heat capacities to other physical quantities.

Since the Maxwell relations are a simple result of the fact that the order of differentiating makes no difference in the value of a second derivative, similar relations among derivatives containing composition variables also exist. They can be found by a straightforward extension of the above procedure.

1.17 Useful differential relations

The state of a thermodynamic system is completely determined by any one of the thermodynamic potentials. The Gibbs free energy, for example, specifies the state as a function of the temperature, pressure, and composition. For a single-phase system, the phase rule tells us that the number of independent variables is one more than the number of components. For a single-phase, one-component system, there are only two independent variables. That is, specifying any two variables determines all other thermodynamic quantities. Since these quantities are derivatives of the potentials, we need to know the connections among the derivatives. Finding these connections is easily accomplished with the aid of some properties of derivatives of a function of several variables, which we hereby establish.

Let z be a function of two independent variables x and y,

$$z = z(x, y) \tag{1.17.1}$$

so that

$$dz = \left(\frac{\partial z}{\partial x}\right)_y dx + \left(\frac{\partial z}{\partial y}\right)_x dy \tag{1.17.2}$$

Divide (1.17.2) through by dx to get

$$\frac{dz}{dx} = \left(\frac{\partial z}{\partial x}\right)_y + \left(\frac{\partial z}{\partial y}\right)_x \frac{dy}{dx} \tag{1.17.3}$$

This equation is true for any dz, so it is true for $dz = 0$, and therefore

$$\left(\frac{\partial z}{\partial x}\right)_y + \left(\frac{\partial z}{\partial y}\right)_x \left(\frac{\partial y}{\partial x}\right)_z = 0 \tag{1.17.4}$$

Similarly, dividing (1.17.2) through by dy and applying the condition of constant z gives

$$\left(\frac{\partial z}{\partial y}\right)_x + \left(\frac{\partial z}{\partial x}\right)_y \left(\frac{\partial x}{\partial y}\right)_z = 0 \tag{1.17.5}$$

Equations (1.17.4) and (1.17.5) give

$$\left(\frac{\partial y}{\partial x}\right)_z = -\frac{\left(\frac{\partial z}{\partial x}\right)_y}{\left(\frac{\partial z}{\partial y}\right)_x} \tag{1.17.6}$$

$$\left(\frac{\partial x}{\partial y}\right)_z = -\frac{\left(\frac{\partial z}{\partial y}\right)_x}{\left(\frac{\partial z}{\partial x}\right)_y} \tag{1.17.7}$$

and from (1.17.6) and (1.17.7),

$$\left(\frac{\partial x}{\partial y}\right)_z = \frac{1}{\left(\frac{\partial y}{\partial x}\right)_z} \tag{1.17.8}$$

Now multiply (1.17.4) by the derivative of y with respect to z at constant x and use (1.17.6) to get

$$\left(\frac{\partial y}{\partial z}\right)_x \left(\frac{\partial z}{\partial y}\right)_x + \left(\frac{\partial x}{\partial y}\right)_z \left(\frac{\partial y}{\partial z}\right)_x \left(\frac{\partial z}{\partial x}\right)_y = 0 \tag{1.17.9}$$

which, because of (1.17.8), becomes

$$\left(\frac{\partial x}{\partial y}\right)_z \left(\frac{\partial y}{\partial z}\right)_x \left(\frac{\partial z}{\partial x}\right)_y = -1 \tag{1.17.10}$$

Equations (1.17.7)–(1.17.10) are very useful in manipulating derivatives of thermodynamic functions.

1.18 Equations of state and heat capacity relations

Consider a simple system consisting of a single phase of constant composition. The properties of the system are then determined by any two thermodynamic variables, and the state of the system is completely described by any of the thermodynamic potentials as a function of two variables. In this sense, the thermodynamic equations are equations of state, but this name is normally applied to the relation between pressure, temperature, and volume.

A thermodynamic equation of state is easily obtained as follows: start with equation (1.11.7), divide by dV, and then specify that the temperature is constant to get the derivative of the energy with respect to volume at constant temperature:

$$\left(\frac{\partial U}{\partial V}\right)_T = T\left(\frac{\partial S}{\partial V}\right)_T - P \tag{1.18.1}$$

Substitution of the Maxwell relation (1.16.8) into (1.18.1) and solving for the pressure gives

$$P = T\left(\frac{\partial P}{\partial T}\right)_V - \left(\frac{\partial U}{\partial V}\right)_T \tag{1.18.2}$$

which is a thermodynamic equation of state for the pressure as a function of volume and temperature.

Another equation of state can be derived by starting with equation (1.12.4) for the reversible differential of the enthalpy. Dividing (1.12.4) by dP and then requiring the temperature to be constant gives

$$\left(\frac{\partial H}{\partial P}\right)_T = T\left(\frac{\partial S}{\partial P}\right)_T + V \tag{1.18.3}$$

Using the Maxwell relation (1.16.9) in (1.18.3) and solving for the volume gives

$$V = T\left(\frac{\partial V}{\partial T}\right)_P + \left(\frac{\partial H}{\partial P}\right)_T \tag{1.18.4}$$

which is a thermodynamic equation of state for the volume as a function of temperature and pressure.

The heat capacity, thermal expansion, and bulk modulus (or its reciprocal, compressibility) are among the more important thermodynamic derivatives, not only because they can be experimentally measured and are of practical use, but also because they are readily related to atomistic theories of matter. The heat capacities at constant pressure and at constant volume have been defined in section 1.9. The definitions of the isothermal compressibility κ and the isobaric thermal expansion coefficient α are

$$\kappa = -\frac{1}{V}\left(\frac{\partial V}{\partial P}\right)_T \tag{1.18.5}$$

$$\alpha = \frac{1}{V}\left(\frac{\partial V}{\partial T}\right)_P \tag{1.18.6}$$

The bulk modulus is defined by

$$B = -V\left(\frac{\partial P}{\partial V}\right)_T \tag{1.18.7}$$

and because of (1.17.8) this is just the reciprocal of the compressibility. The compressibility is the fractional decrease in volume per unit of applied pressure, and the thermal expansion coefficient is the fractional increase in volume per unit temperature increase.

Measured thermal expansions are usually reported as linear thermal expansion coefficients, which are defined as

$$\alpha_L = \frac{1}{L}\left(\frac{\partial L}{\partial T}\right)_P \tag{1.18.8}$$

where L is a linear dimension of the sample.

The relation between the volume and linear expansion coefficient is readily obtained by writing the volume as a product of the linear dimensions L_1, L_2, L_3 so that $dV = L_1L_2dL_3 + L_1L_3dL_2 + L_2L_3dL_1$ and $dV/V = dL_1/L_1 + dL_2/L_2 + dL_3/L_3$. Putting this in (1.18.6) gives

$$\alpha = \frac{1}{L_1}\left(\frac{\partial L_1}{\partial T}\right)_P + \frac{1}{L_2}\left(\frac{\partial L_2}{\partial T}\right)_P + \frac{1}{L_3}\left(\frac{\partial L_3}{\partial T}\right)_P \tag{1.18.9}$$

and if the system is isotropic, all three terms in (1.18.9) are equal, so comparing it to (1.18.8) gives

$$\alpha = 3\alpha_L \tag{1.18.10}$$

For an anisotropic system, such as an orthorhombic crystal, the linear expansion coefficient varies with direction.

A relation between the heat capacities, the thermal expansion coefficient and the compressibility is found as follows. First, write the energy as a function of volume and temperature, instead of its natural variables volume and entropy, so that its differential is

$$dU = \left(\frac{\partial U}{\partial V}\right)_T dV + \left(\frac{\partial U}{\partial T}\right)_V dT \tag{1.18.11}$$

from which $(\partial U/\partial T)_P = (\partial U/\partial V)_T (\partial V/\partial T)_P + (\partial U/\partial T)_V$. But the last term in this equation is just the constant volume heat capacity, so

$$\left(\frac{\partial U}{\partial T}\right)_P = \left(\frac{\partial U}{\partial V}\right)_V \left(\frac{\partial V}{\partial T}\right)_T + C_V \tag{1.18.12}$$

Now start with the definition of the heat capacity at constant pressure to get

$$C_P = \left(\frac{\partial H}{\partial T}\right)_P = \left[\frac{\partial (U + PV)}{\partial T}\right]_P = \left(\frac{\partial U}{\partial T}\right)_P + P\left(\frac{\partial V}{\partial T}\right)_P \tag{1.18.13}$$

From (1.18.12) and (1.18.13), the difference between the two heat capacities is

$$\begin{aligned} C_P - C_V &= \left(\frac{\partial U}{\partial V}\right)_T \left(\frac{\partial V}{\partial T}\right)_P + P\left(\frac{\partial V}{\partial T}\right)_P \\ &= \left(\frac{\partial V}{\partial T}\right)_P \left[P + \left(\frac{\partial U}{\partial V}\right)_T\right] \end{aligned} \tag{1.18.14}$$

Because of the thermodynamic equation of state (1.18.2) the term in the square bracket is

$$P+\left(\frac{\partial U}{\partial V}\right)_T = T\left(\frac{\partial P}{\partial T}\right)_V \tag{1.18.15}$$

so (1.18.14) becomes

$$C_P - C_V = T\left(\frac{\partial V}{\partial T}\right)_P \left(\frac{\partial P}{\partial T}\right)_V \tag{1.18.16}$$

and using (1.17.7), this becomes

$$C_P - C_V = -T\frac{\left(\frac{\partial V}{\partial T}\right)_P^2}{\left(\frac{\partial V}{\partial P}\right)_T} \tag{1.18.17}$$

From the definitions of the compressibility and thermal expansion coefficient given by (1.18.5) and (1.18.6), this becomes

$$C_P - C_V = \frac{\alpha^2 TV}{\kappa} \tag{1.18.18}$$

Experimental measurements are most easily carried out under conditions of constant temperature and pressure, but the statistical mechanical theories of heat capacity are most easily developed with temperature and volume as independent variables. Equation (1.18.8) provides the needed connection between the two heat capacities.

1.19 Magnetic systems

To this point, pressure was assumed to be the only external parameter acting on the system, so it was subject only to PV work. It is obvious that the procedures used to develop the ideas of thermodynamics are the same for any kind of work. All that is needed is to treat the external force and its corresponding property of the system just as pressure and volume were treated. An important example of this in the study of solids is from magnetic materials, in which an external magnetic field H[1] acts on a system with a net magnetization M. The magnetic work accompanying a change in external field is MdH, so assuming that this is the only work (zero pressure), the thermodynamic equations for a magnetic system are obtained by replacing P with H and V with M in the equations for a PV system. Thus, the heat capacity at constant volume is replaced with the heat capacity at constant magnetization:

$$C_M = \left(\frac{\partial U}{\partial T}\right)_M = -T\left(\frac{\partial^2 A}{\partial T^2}\right)_M \tag{1.19.1}$$

the last equality being the magnetic analog of (1.12.10). The analog of the entropy equations in (1.12.6) and (1.12.7), for a system in a magnetic field, is

$$S = -\left(\frac{\partial G}{\partial T}\right)_H = -\left(\frac{\partial A}{\partial T}\right)_M \tag{1.19.2}$$

Instead of the compressibility, we now have the derivative of the magnetization with respect to external field, so the analog of the compressibility is the isothermal magnetic susceptibility:

$$\chi_T = \left(\frac{\partial M}{\partial H}\right)_T \tag{1.19.3}$$

For our purpose, we can take the external field as given in the same sense that we took the pressure as given. However, there is a difference between magnetic field and pressure in that the magnetic field carries energy.[2]

Exercises

1.1 Prove that

$$\left[\frac{\partial(A/T)}{\partial T}\right]_V = -\frac{U}{T^2} \tag{1.A}$$

and that

$$\left[\frac{\partial(G/T)}{\partial T}\right]_P = -\frac{H}{T^2} \tag{1.B}$$

These are alternate forms of the Gibbs–Helmholtz equations.

1.2 The Stefan–Boltzmann law for black body radiation states that the energy density of radiation in equilibrium within a cavity varies as the fourth power of the absolute temperature. Derive this law from the fact that the radiation pressure P is related to the energy density u by $P = u/3$, where $u = U/V$, U being the total energy and V being the volume of the cavity, and the energy density is a function of temperature only. (These results can be derived from the electromagnetic theory of radiation.)

1.3 The grand potential for a one-component system containing N molecules is defined by $\Psi = U - TS - \mu N$. Show that

$$N = -\left(\frac{\partial \Psi}{\partial \mu}\right)_{T,V}$$

$$P = -\left(\frac{\partial \Psi}{\partial V}\right)_{T,\mu}$$

$$S = -\left(\frac{\partial \Psi}{\partial T}\right)_{V,\mu}$$

Notes

1. The symbol for the magnetic field is the same as for the enthalpy. But we will be considering only magnetic systems at zero pressure, so there will be no need to specify constant pressure properties and the enthalpy will not

appear in the equations. I have therefore chosen to retain the traditional symbols rather than invent a new one.

2. A clear exposition of the relation of the external field to total magnetic energy is given in chapter 1 of Goodstein, David L.; 1975; *States of Matter*; chapter 1; Prentice-Hall, Englewood Cliffs, N.J.; reprinted by Dover Publications, 1985.

2

Principles of Statistical Mechanics

2.1 Definitions for statistical mechanics

The prime article of faith in scientific theory is that experimentally observed phenomena can be derived from a small number of general principles. The acceptance of this idea implies that, despite the constant changes evident in physical systems, they possess some immutable attributes from which all of their interesting properties can be calculated. A central concept in such a calculation is that of the state of an isolated system. If all of the experimentally measurable properties of a system are known at all times, then this would certainly constitute an adequate description of the state of the system. However, such a definition would be useless for theoretical purposes and is contrary to the assumption that all properties can be related to a few general laws. The most fruitful definition of state is this:

> The state of a physical system is the minimum amount of information required for a calculation of its properties.

A *theory* can be regarded as the set of concepts and mathematical apparatus needed to go from the definition of the state of a system to its observable properties.

A *system* is defined as any physical entity that can be separated from the rest of the world, at least to a good approximation. That is, the energy of interaction of the system with the rest of the world is small compared to the system's energy. No system is truly isolated since there are always some interactions with its surroundings. In fact, by definition, no measurements can be made on isolated systems, and they are therefore of limited interest. However, a system that is too strongly coupled to its environment cannot be treated as a separate entity. The systems that are of interest to us are those for which there is a small interaction with the environment so that a meaning can be assigned to its properties as an entity that is almost independent of the rest of the world. Examples of systems are an atom or molecule in a dilute gas, a gas in an insulating container, a liquid in a bottle, a crystal, and the earth with its atmosphere.

The "nature" of a system is defined by specifying its constituent particles (such as electrons, nuclei, atoms, and/or molecules), the number of different constituent particles, and the interaction potentials among these particles. Often, as a matter of convenience, we go further and specify its phases and structure (i.e., gas, liquid, solid; crystals, polymers, amorphous solid). In principle, the phases and structure can be derived from the constitution and inter-

action potentials, but this is a difficult task and we are often interested in systems whose phases and structure are given.

The above discussion applies to conservative systems. For nonconservative systems, the effects of the nonconservative forces must be included. But while there may be nonconservative processes at the macro level, they often reduce to conservative processes at the micro level, for example, the phenomena of friction and turbulence. Statistical mechanics is primarily concerned with *macroscopic systems*, which are defined as systems whose dimensions are large compared to atomic dimensions and whose number of constituent particles is large compared to unity.

The definition of state is different for different branches of physical theory. Given the constituent particles, their masses, and the forces acting among them, the state in classical mechanics is completely specified by the coordinates and momenta of the particles at any given instant. The reason for this is that, in classical mechanics, all properties of systems are functions of coordinates and momenta, and once the positions and momenta are given at one particular time, they can be calculated at any future (or past) time. In quantum mechanics, the state is completely defined by the wave function if the constituents, their masses, and their interaction potentials are given, since all measurable physical properties can be obtained from the wave function.

2.2 Thermodynamic state

The first and second laws of thermodynamics provide a basis for constructing relations among the macroscopic properties of physical systems. These properties include temperature, pressure, volume, energy, specific heat, compressibility, and so forth. In a one-component, one-phase system, two variables (e.g., temperature and pressure) are enough to define the thermodynamic state. All other properties are then determined by the equations of thermodynamics. For systems in equilibrium, the thermodynamic state is defined by a number of macroscopic properties determined by the phase rule given by (1.14.12).

The thermodynamic parameters of a system are of two kinds: mechanical and nonmechanical. *Mechanical variables* are those quantities that can be interpreted in mechanical terms, such as energy and internal pressure. *Nonmechanical variables* are those that have no analog in mechanics and are peculiar to thermodynamics. These include temperature and entropy.

Thermodynamic quantities can also be classified as being intensive or extensive. *Intensive parameters* are those that are independent of the size of the system, such as temperature, pressure, and concentration of components. *Extensive parameters* are those that are directly proportional to the amount of matter in the system, such as volume, energy, and heat capacity.

Also, it is useful to distinguish between the properties of the system and external parameters. The *external parameters* define the conditions under which the system exists and its interaction with the environment. These include such quantities as external fields (gravitational, electric, magnetic), the pressure of a movable piston, and the temperature of a heat bath.

Two important points must be noted. First, thermodynamics deals with equilibrium systems, so time and the concept of temporal causality do not enter into it at all. Equilibrium thermodynamic systems are macroscopically static, and the definition of state given here is incapable of describing systems whose properties are changing with time. The idea of thermodynamic equilibrium needs to be applied with care because, in thermodynamics, equilibrium is defined as being the state with the lowest free energy. But there are states whose properties do not change with time, or change so slowly that no change can be mea-

sured over laboratory time scales, and yet are not states with the lowest free energy compared to all other possible states. This can happen when the transition to a lower free energy state is hindered by extremely slow kinetic rates, as when nonequilibrium structures are quenched into solids, or when chemical reaction rates are slow, for example, the reaction between hydrogen and oxygen at room temperatures. It is often valid to apply equilibrium thermodynamics to such systems because, while the free energy may not be at an absolute minimum, it might be at a relative minimum. It is then in a metastable state.

The second point is that no actual numerical computations of the thermodynamic properties can be made from the specification of the thermodynamic state of the system. For example, if we are given the temperature and pressure of a gas, thermodynamics tells us that the volume is determined, but it cannot tell us the numerical value of this volume. To get the volume, an equation of state is needed, and this can only be obtained from some source outside of thermodynamics such as experimental data or a microscopic theory. In this sense, the specification of state in thermodynamics is incomplete.

Sometimes a macroscopic state cannot be defined. This is the case for systems that are so chaotic and changing so rapidly that their future properties cannot be determined, or even expressed, in macroscopic terms. For some nonequilibrium systems, however, it may still be possible to define a macroscopic state. For example, a system whose temperature is changing with time can be described by the equations of heat conduction if the temperature gradients are not too large. The state is then defined by including in it the thermal conductivity and the temperature as a function of position at a given time. This constitutes a definition of state for the nonequilibrium system because it allows the temperature distribution to be computed for all future times.

2.3 Comparison of microscopic and macroscopic state

It is always possible to define a microscopic state of the system since, in principle, the equations of motion (quantum or classical) of the particle constituents completely define the system. Microscopic states can be defined in classical and quantum mechanics whether or not the system is in equilibrium since the equations of motion are known.

To appreciate the task of statistical mechanics, it is useful to contrast the following characteristics of the microscopic and macroscopic definitions of state:

1. For macroscopic systems the microscopic definition of state is extremely detailed and requires knowledge about every particle in the system. Classically, it is necessary to specify all coordinates at some particular time. In quantum mechanics, all relevant properties are computed from the wave function at a particular time, the wave function being obtained from a second-order partial differential equation involving all particle coordinates. For macroscopic systems, however, there is no way to get such detailed information, and even if it were available, it would be useless.

In thermodynamics, the macroscopic state says nothing about microscopic parameters and does not even recognize their existence. Furthermore, the macroscopic state is specified by a small number of parameters (the phase rule), in sharp contrast to the enormous number of variables that enter into the microscopic state.

2. The microscopic definition of state is totally causal in the sense that the state at any time is determined by the state at some earlier time. This means that all measurable quantities are determined as a function of time. This is a precise statement even in quantum mechanics, despite the uncertainty princi-

ple, because the state is defined by the wave function whose evolution in time is precisely determined by the wave equation. (It is true that, at an atomistic level, the measurable quantities are subject to quantum mechanical probabilities, but these are determined by the wave function.) The macroscopic state, on the other hand, is only partially causal. For equilibrium systems, it is causal in the trivial sense that it does not change in time. For nonequilibrium systems, it may be causal to a good enough approximation for certain phenomenan such as diffusion and heat transfer (when gradients are not too large) in the sense that future temperature or concentration distributions can be computed from initial conditions. But for systems that are far from equilibrium, the future macroscopic state cannot always be predicted from a past macroscopic state. Note that in those cases that may be described causally, laws specific to the phenomenon must be introduced (e.g., Fick's law and Fourier's law).

3. The microscopic state is reversible with respect to time. That is, for every process evolving into the future that brings the system from a state A to a state B, there is a corresponding process evolving into the past that brings the system from state B to state A. What this means is that every temporal process is described by equations that give solutions for the properties as a function of time and valid solutions are obtained if the time is reversed in these equations. For macroscopic states, however, temporal processes are irreversible. That is, for the evolution of a macroscopic system in time, there is no corresponding process that undoes the evolution if the time is reversed.

4. If the microscopic state is known, along with some fundamental constants (such as the mass and charge of the electron, Planck's constant, and the velocity of light), then actual numerical calculations of the system's properties can be made, at least in principle (within the limitations of the uncertainty principle). Numerical values of macroscopic properties cannot be computed even in principle from the definition of macroscopic state. The equations defining macroscopic state are merely general relations among macroscopic parameters.

2.4 The relation between microscopic and macroscopic state

Our belief in the ultimate unity of scientific knowledge leads us to conclude that there must be some connection between the macroscopic (thermodynamic) and microscopic (mechanical) definitions of state. The first task of statistical mechanics is to find that connection. For the many particle systems that concern us, it is impossible to obtain an actual specification of the microscopic state and, even if such a specification were available, the equations of motion would be so complicated that they could not be solved. Also, we are not particularly interested in such detailed information. We are, however, very much interested in the possibility of interpreting thermodynamic data in microscopic terms, and in particular, it would be very desirable to obtain a method of computing macroscopic quantities from atomic properties, thereby making up for the inherent deficiency of thermodynamics.

In view of these considerations, any bridge connecting macroscopic and microscopic descriptions of material systems must have the following characteristics:

1. A great deal of microscopic information must be erased in constructing macro- from microstates so that we are not encumbered with an enormous number of microscopic parameters.
2. Enough microscopic information must be retained to enable the calculation of bulk properties from atomic properties.