1
Thermodynamics

1.1 Introduction

Thermodynamics *is a phenomenological description of properties of macroscopic systems in thermal equilibrium.*

Imagine yourself as a post-Newtonian physicist intent on understanding the behavior of such a simple system as a container of gas. How would you proceed? The prototype of a successful physical theory is classical mechanics, which describes the intricate motions of particles starting from simple basic laws and employing the mathematical machinery of calculus. By analogy, you may proceed as follows:

- Idealize the *system* under study as much as possible (as is the case of a point particle). The concept of mechanical work on the system is certainly familiar, yet there appear to be complications due to exchange of heat. The solution is first to examine *closed systems*, insulated by *adiabatic walls* that don’t allow any exchange of heat with the surroundings. Of course, it is ultimately also necessary to study *open systems*, which may exchange heat with the outside world through *diathermic walls*.
- As the state of a point particle is quantified by its coordinates (and momenta), properties of the macroscopic system can also be described by a number of *thermodynamic coordinates* or *state functions*. The most familiar coordinates are those that relate to mechanical work, such as pressure and volume (for a fluid), surface tension and area (for a film), tension and length (for a wire), electric field and polarization (for a dielectric), etc. As we shall see, there are additional state functions not related to mechanical work. The state functions are well defined only when the system is in *equilibrium*, that is, when its properties do not change appreciably with time over the intervals of interest (observation times). The dependence on the observation time makes the concept of equilibrium subjective. For example, window glass is in equilibrium as a solid over many decades, but flows like a fluid over time scales of millennia. At the other extreme, it is perfectly legitimate to consider the equilibrium between matter and radiation in the early universe during the first minutes of the Big Bang.

- Finally, the relationship between the state functions is described by the laws of thermodynamics. As a *phenomenological* description, these laws are based on a number of empirical observations. A coherent logical and mathematical structure is then constructed on the basis of these observations, which leads to a variety of useful concepts, and to testable relationships among various quantities. The laws of thermodynamics can only be justified by a more fundamental (microscopic) theory of nature. For example, statistical mechanics attempts to obtain these laws starting from classical or quantum mechanical equations for the evolution of collections of particles.

1.2 The zeroth law

The zeroth law of thermodynamics describes the transitive nature of thermal equilibrium. It states:

If two systems, A and B, are separately in equilibrium with a third system, C, then they are also in equilibrium with one another.

Despite its apparent simplicity, the zeroth law has the consequence of implying the existence of an important state function, the *empirical temperature* Θ, such that systems in equilibrium are at the same temperature.

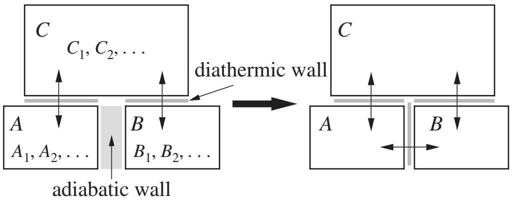


Fig. 1.1 Illustration of the zeroth law: systems A and B, which are initially separately in equilibrium with C, are placed in contact with each other.

Let the equilibrium state of systems A, B, and C be described by the coordinates $\{A_1, A_2, \cdots\}$, $\{B_1, B_2, \cdots\}$, and $\{C_1, C_2, \cdots\}$, respectively. The assumption that A and C are in equilibrium implies a constraint between the coordinates of A and C, that is, a change in A_1 must be accompanied by some changes in $\{A_2, \cdots ; C_1, C_2, \cdots\}$ to maintain equilibrium of A and C. Denote this constraint by

$$f_{AC}(A_1, A_2, \cdots ; C_1, C_2, \cdots) = 0. \tag{1.1}$$

The equilibrium of B and C implies a similar constraint

$$f_{BC}(B_1, B_2, \cdots ; C_1, C_2, \cdots) = 0. \tag{1.2}$$

Note that each system is assumed to be separately in mechanical equilibrium. If they are allowed also to do work on each other, additional conditions (e.g., constant pressure) are required to describe their joint mechanical equilibrium.

Clearly we can state the above constraint in many different ways. For example, we can study the variations of C_1 as all of the other parameters are changed. This is equivalent to solving each of the above equations for C_1 to yield [1]

$$\begin{aligned} C_1 &= F_{AC}(A_1, A_2, \cdots; C_2, \cdots), \\ C_1 &= F_{BC}(B_1, B_2, \cdots; C_2, \cdots). \end{aligned} \tag{1.3}$$

Thus if C is separately in equilibrium with A and B, we must have

$$F_{AC}(A_1, A_2, \cdots; C_2, \cdots) = F_{BC}(B_1, B_2, \cdots; C_2, \cdots). \tag{1.4}$$

However, according to the zeroth law there is also equilibrium between A and B, implying the constraint

$$f_{AB}(A_1, A_2, \cdots; B_1, B_2, \cdots) = 0. \tag{1.5}$$

We can select any set of parameters $\{A, B\}$ that satisfy the above equation, and substitute them in Eq. (1.4). The resulting equality must hold quite independently of any set of variables $\{C\}$ in this equation. We can then change these parameters, moving along the manifold constrained by Eq. (1.5), and Eq. (1.4) will remain valid irrespective of the state of C. Therefore, it must be possible to simplify Eq. (1.4) by canceling the coordinates of C. Alternatively, we can select any fixed set of parameters C, and ignore them henceforth, reducing the condition (1.5) for equilibrium of A and B to

$$\Theta_A(A_1, A_2, \cdots) = \Theta_B(B_1, B_2, \cdots), \tag{1.6}$$

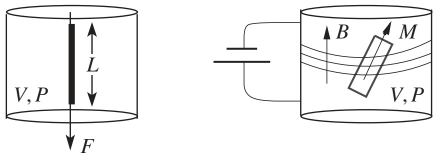
that is, equilibrium is characterized by a function Θ of thermodynamic coordinates. This function specifies the *equation of state*, and *isotherms* of A are described by the condition $\Theta_A(A_1, A_2, \cdots) = \Theta$. While at this point there are many potential choices of Θ, the key point is the existence of a function that constrains the parameters of each system in thermal equilibrium.

There is a similarity between Θ and the force in a mechanical system. Consider two one-dimensional systems that can do work on each other as in the case of two springs connected together. Equilibrium is achieved when the forces exerted by each body on the other are equal. (Of course, unlike the scalar temperature, the vectorial force has a direction; a complication that we have ignored. The pressure of a gas piston provides a more apt analogy.) The mechanical equilibrium between several such bodies is also transitive, and the latter could have been used as a starting point for deducing the existence of a mechanical force.

---

[1] From a purely mathematical perspective, it is not necessarily possible to solve an arbitrary constraint condition for C_1. However, the requirement that the constraint describes real physical parameters clearly implies that we can obtain C_1 as a function of the remaining parameters.

As an example, let us consider the following three systems: (A) a wire of length L with tension F, (B) a paramagnet of magnetization M in a magnetic field B, and (C) a gas of volume V at pressure P.



(A)&(C) (A)&(B)

Fig. 1.2 Equilibria of a gas (A) and a magnet (B), and a gas (A) and a wire (C).

Observations indicate that when these systems are in equilibrium, the following constraints are satisfied between their coordinates:

$$\begin{aligned}\left(P+\frac{a}{V^2}\right)(V-b)(L-L_0)-c[F-K(L-L_0)]=0,\\ \left(P+\frac{a}{V^2}\right)(V-b)M-dB=0.\end{aligned} \tag{1.7}$$

The two conditions can be organized into three empirical temperature functions as

$$\Theta \propto \left(P+\frac{a}{V^2}\right)(V-b)=c\left(\frac{F}{L-L_0}-K\right)=d\frac{B}{M}. \tag{1.8}$$

Note that the zeroth law severely constrains the form of the constraint equation describing the equilibrium between two bodies. Any arbitrary function cannot necessarily be organized into an equality of two empirical temperature functions.

The constraints used in the above example were in fact chosen to reproduce three well-known equations of state that will be encountered and discussed later in this book. In their more familiar forms they are written as

$$\begin{cases}(P+a/V^2)(V-b)=Nk_BT & \text{(van der Waals gas)}\\ M=(N\,\mu_B^2B)/(3k_BT) & \text{(Curie paramagnet)}\\ F=(K+DT)(L-L_0) & \text{(Hooke's law for rubber)}\end{cases}. \tag{1.9}$$

Note that we have employed the symbol for Kelvin temperature T, in place of the more general empirical temperature Θ. This concrete temperature scale can be constructed using the properties of the ideal gas.

The ideal gas temperature scale: while the zeroth law merely states the presence of isotherms, to set up a practical temperature scale at this stage, a reference system is necessary. The *ideal gas* occupies an important place in thermodynamics and provides the necessary reference. Empirical observations indicate that the product of pressure and volume is constant along the isotherms of any gas that is sufficiently dilute. The ideal gas refers to this *dilute* limit of

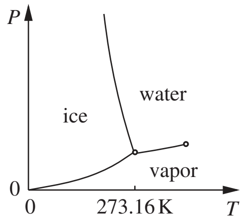


Fig. 1.3 The triple point of ice, water, and steam occurs at a unique point in the (P, T) phase diagram.

real gases, and the ideal gas temperature is proportional to the product. The constant of proportionality is determined by reference to the temperature of the triple point of the ice–water–gas system, which was set to 273.16 (degrees) kelvin (K) by the 10th General Conference on Weights and Measures in 1954. Using a dilute gas (i.e., as $P \to 0$) as thermometer, the temperature of a system can be obtained from

$$T(\mathrm{K}) \equiv 273.16 \times \left(\lim_{P\to 0} (PV)_{\text{system}} / \lim_{P\to 0} (PV)_{\text{ice-water-gas}} \right). \tag{1.10}$$

1.3 The first law

In dealing with simple mechanical systems, conservation of energy is an important principle. For example, the location of a particle can be changed in a potential by external work, which is then related to a change in its potential energy. Observations indicate that a similar principle operates at the level of macroscopic bodies *provided that the system is properly insulated*, that is, when the only sources of energy are of mechanical origin. We shall use the following formulation of these observations:

The amount of work required to change the state of an otherwise adiabatically isolated system depends only on the initial and final states, and not on the means by which the work is performed, or on the intermediate stages through which the system passes.

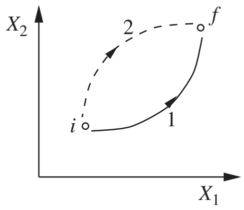


Fig. 1.4 The two adiabatic paths for changing macroscopic coordinates between the initial and final point result in the same change in internal energy.

For a particle moving in a potential, the required work can be used to construct a potential energy landscape. Similarly, for the thermodynamic system we can construct another state function, the internal energy $E(\mathbf{X})$. Up to a constant,

$E(\mathbf{X})$ can be obtained from the amount of work ΔW needed for an *adiabatic* transformation from an initial state $\mathbf{X_i}$ to a final state $\mathbf{X_f}$, using

$$\Delta W = E(\mathbf{X_f}) - E(\mathbf{X_i}). \tag{1.11}$$

Another set of observations indicate that once the adiabatic constraint is removed, the amount of work is no longer equal to the change in the internal energy. The difference $\Delta Q = \Delta E - \Delta W$ is defined as the *heat* intake of the system from its surroundings. Clearly, in such transformations, ΔQ and ΔW are not separately functions of state in that they depend on external factors such as the means of applying work, and not only on the final states. To emphasize this, for a differential transformation we write

$$đQ = \mathrm{d}E - đW, \tag{1.12}$$

where $\mathrm{d}E = \sum_i \partial_i E \mathrm{d}X_i$ can be obtained by differentiation, while $đQ$ and $đW$ generally cannot. Also note that our convention is such that the signs of work and heat indicate the energy *added* to the system, and not vice versa. The first law of thermodynamics thus states that to change the state of a system we need a fixed amount of energy, which can be in the form of mechanical work or heat. This can also be regarded as a way of defining and quantifying the exchanged heat.

A *quasi-static* transformation is one that is performed sufficiently slowly so that the system is always in equilibrium. Thus, at any stage of the process, the thermodynamic coordinates of the system exist and can in principle be computed. For such transformations, the work done on the system (equal in magnitude but opposite in sign to the work done by the system) can be related to changes in these coordinates. As a mechanical example, consider the stretching of a spring or rubber band. To construct the potential energy of the system as a function of its length L, we can pull the spring sufficiently slowly so that at each stage the external force is matched by the internal force F from the spring. For such a quasi-static process, the change in the potential energy of the spring is $\int F \mathrm{d}L$. If the spring is pulled abruptly, some of the external work is converted into kinetic energy and eventually lost as the spring comes to rest.

Generalizing from the above example, one can typically divide the state functions $\{\mathbf{X}\}$ into a set of *generalized displacements* $\{\mathbf{x}\}$, and their conjugate *generalized forces* $\{\mathbf{J}\}$, such that for an infinitesimal quasi-static transformation[2]

$$đW = \sum_i J_i \mathrm{d}x_i. \tag{1.13}$$

---

[2] I denote force by the symbol **J** rather than **F**, to reserve the latter for the free energy. I hope the reader is not confused with currents (sometimes also denoted by **J**), which rarely appear in this book.

Table 1.1 *Generalized forces and displacements*

System	Force		Displacement	
Wire	tension	F	length	L
Film	surface tension	$\mathcal{S}$	area	A
Fluid	pressure	$-P$	volume	V
Magnet	magnetic field	H	magnetization	M
Dielectric	electric field	E	polarization	P
Chemical reaction	chemical potential	μ	particle number	N

Table 1.1 provides some common examples of such conjugate coordinates. Note that pressure P is by convention calculated from the force exerted by the system on the walls, as opposed to the force on a spring, which is exerted in the opposite direction. This is the origin of the negative sign that accompanies hydrostatic work.

The displacements are usually *extensive* quantities, that is, proportional to system size, while the forces are *intensive* and independent of size. The latter are indicators of equilibrium; for example, the pressure is uniform in a gas in equilibrium (in the absence of external potentials) and equal for two equilibrated gases in contact. As discussed in connection with the zeroth law, temperature plays a similar role when heat exchanges are involved. Is there a corresponding displacement, and if so what is it? This question will be answered in the following sections.

The ideal gas: we noted in connection with the zeroth law that the equation of state of the ideal gas takes a particularly simple form, $PV \propto T$. The internal energy of the ideal gas also takes a very simple form, as observed for example by *Joule's free expansion experiment*: measurements indicate that if an ideal gas expands adiabatically (but not necessarily quasi-statically), from a volume V_i to V_f, the initial and final temperatures are the same. As the transformation is adiabatic ($\Delta Q = 0$) and there is no external work done on the system ($\Delta W = 0$), the internal energy of the gas is unchanged. Since the pressure and volume of the gas change in the process, but its temperature does not, we conclude that the internal energy depends only on temperature, that is, $E(V, T) = E(T)$. This property of the ideal gas is in fact a consequence of the form of its equation of state, as will be proved in one of the problems at the end of this chapter.

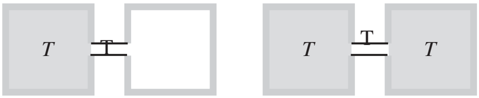


Fig. 1.5 A gas initially confined in the left chamber is allowed to expand rapidly to both chambers.

Response functions are the usual method for characterizing the macroscopic behavior of a system. They are experimentally measured from the changes of thermodynamic coordinates with external probes. Some common response functions are as follows.

Heat capacities are obtained from the change in temperature upon addition of heat to the system. Since heat is not a function of state, the path by which it is supplied must also be specified. For example, for a gas we can calculate the heat capacities at constant volume or pressure, denoted by $C_V = đQ/\mathrm{d}T|_V$ and $C_P = đQ/\mathrm{d}T|_P$, respectively. The latter is larger since some of the heat is used up in the work done in changes of volume:

$$\begin{aligned} C_V &= \left.\frac{đQ}{\mathrm{d}T}\right|_V = \left.\frac{\mathrm{d}E - đW}{\mathrm{d}T}\right|_V = \left.\frac{\mathrm{d}E + P\mathrm{d}V}{\mathrm{d}T}\right|_V = \left.\frac{\partial E}{\partial T}\right|_V, \\ C_P &= \left.\frac{đQ}{\mathrm{d}T}\right|_P = \left.\frac{\mathrm{d}E - đW}{\mathrm{d}T}\right|_P = \left.\frac{\mathrm{d}E + P\mathrm{d}V}{\mathrm{d}T}\right|_P = \left.\frac{\partial E}{\partial T}\right|_P + P\left.\frac{\partial V}{\partial T}\right|_P. \end{aligned} \tag{1.14}$$

Force constants measure the (infinitesimal) ratio of displacement to force and are generalizations of the spring constant. Examples include the isothermal compressibility of a gas $\kappa_T = -\left.\partial V/\partial P\right|_T/V$, and the susceptibility of a magnet $\chi_T = \left.\partial M/\partial B\right|_T/V$. From the equation of state of an ideal gas $PV \propto T$, we obtain $\kappa_T = 1/P$.

Thermal responses probe the change in the thermodynamic coordinates with temperature. For example, the expansivity of a gas is given by $\alpha_P = \left.\partial V/\partial T\right|_P/V$, which equals $1/T$ for the ideal gas.

Since the internal energy of an ideal gas depends only on its temperature, $\left.\partial E/\partial T\right|_V = \left.\partial E/\partial T\right|_P = \mathrm{d}E/\mathrm{d}T$, and Eq. (1.14) simplifies to

$$C_P - C_V = P\left.\frac{\partial V}{\partial T}\right|_P = PV\alpha_P = \frac{PV}{T} \equiv Nk_B. \tag{1.15}$$

The last equality follows from extensivity: for a given amount of ideal gas, the constant PV/T is proportional to N, the number of particles in the gas; the ratio is Boltzmann's constant with a value of $k_B \approx 1.4 \times 10^{-23}\,\mathrm{JK}^{-1}$.

1.4 The second law

The practical impetus for development of the science of thermodynamics in the nineteenth century was the advent of heat engines. The increasing reliance on machines to do work during the industrial revolution required better understanding of the principles underlying conversion of heat to work. It is quite interesting to see how such practical considerations as the efficiency of engines can lead to abstract ideas like entropy.

An idealized *heat engine* works by taking in a certain amount of heat Q_H, from a heat source (for example a coal fire), converting a portion of it to work

W, and dumping the remaining heat Q_C into a heat sink (e.g., atmosphere). The efficiency of the engine is calculated from

$$\eta = \frac{W}{Q_H} = \frac{Q_H - Q_C}{Q_H} \leq 1. \tag{1.16}$$

An idealized *refrigerator* is like an engine running backward, that is, using work W to extract heat Q_C from a cold system, and dumping heat Q_H at a higher temperature. We can similarly define a figure of merit for the performance of a refrigerator as

$$\omega = \frac{Q_C}{W} = \frac{Q_C}{Q_H - Q_C}. \tag{1.17}$$

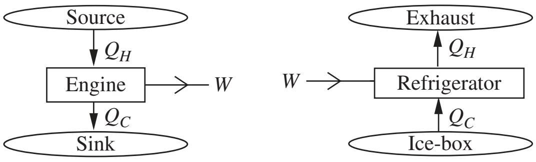


Fig. 1.6 The idealized engine and refrigerator.

The first law rules out so-called "perpetual motion machines of the first kind," that is, engines that produce work without consuming any energy. However, the conservation of energy is not violated by an engine that produces work by converting water to ice. Such a "perpetual motion machine of the second kind" would certainly solve the world's energy problems, but is ruled out by the second law of thermodynamics. The observation that the natural direction for the flow of heat is from hotter to colder bodies is the essence of the second law of thermodynamics. There is a number of different formulations of the second law, such as the following two statements:

Kelvin's statement. *No process is possible whose sole result is the complete conversion of heat into work.*

Clausius's statement. *No process is possible whose sole result is the transfer of heat from a colder to a hotter body.*

A perfect engine is ruled out by the first statement, a perfect refrigerator by the second. Since we shall use both statements, we first show that they are equivalent. Proof of the equivalence of the Kelvin and Clausius statements proceeds by showing that if one is violated, so is the other.

(a) Let us assume that there is a machine that violates Clausius's statement by taking heat Q from a cooler region to a hotter one. Now consider an engine operating between these two regions, taking heat Q_H from the hotter one and dumping Q_C at the colder sink. The combined system takes $Q_H - Q$ from the hot source, produces work equal to $Q_H - Q_C$, and dumps $Q_C - Q$ at the cold sink. If we adjust the engine output such that $Q_C = Q$, the net result is a 100% efficient engine, in violation of Kelvin's statement.

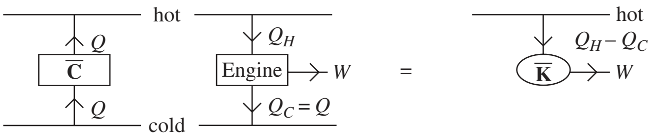


Fig. 1.7 A machine violating Clausius's statement ($\overline{\mathrm{C}}$) can be connected to an engine, resulting in a combined device ($\overline{\mathrm{K}}$) that violates Kelvin's statement.

(b) Alternatively, assume a machine that violates Kelvin's law by taking heat Q and converting it completely to work. The work output of this machine can be used to run a refrigerator, with the net outcome of transferring heat from a colder to a hotter body, in violation of Clausius's statement.

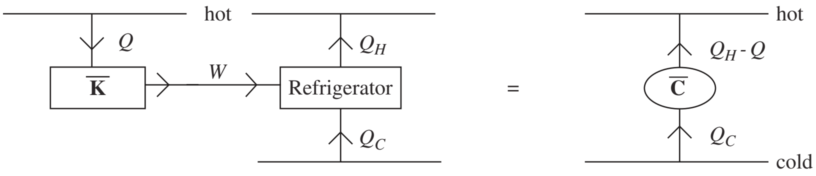


Fig. 1.8 A machine violating Kelvin's statement can be connected to a refrigerator, resulting in violation of Clausius's statement.

1.5 Carnot engines

A Carnot engine is any engine that is reversible, runs in a cycle, with all of its heat exchanges taking place at a source temperature T_H*, and a sink temperature* T_C.

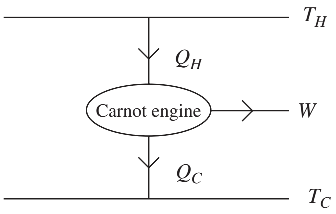


Fig. 1.9 A Carnot engine operates between temperatures T_H and T_C, with no other heat exchanges.

A *reversible* process is one that can be run backward in time by simply reversing its inputs and outputs. It is the thermodynamic equivalent of frictionless motion in mechanics. Since time reversibility implies equilibrium, a reversible transformation must be quasi-static, but the reverse is not necessarily true (e.g., if there is energy dissipation due to friction). An engine that runs in a *cycle*