

CHAPTER 1

INTRODUCTION

1.1 What is thermodynamics?

Thermodynamics is the science which has evolved from the original investigations in the 19th century into the nature of “heat.” At the time, the leading theory of heat was that it was a type of fluid, which could flow from a hot body to a colder one when they were brought into contact. We now know that what was then called “heat” is not a fluid, but is actually a form of energy – it is the energy associated with the continual, random motion of the atoms which compose macroscopic matter, which we can’t see directly.

This type of energy, which we will call *thermal energy*, can be converted (at least in part) to other forms which we *can* perceive directly (for example, kinetic, gravitational, or electrical energy), and which can be used to do useful things such as propel an automobile or a 747. The principles of thermodynamics govern the conversion of thermal energy to other, more useful forms.

For example, an automobile engine can be thought of as a device which first converts chemical energy stored in fuel and oxygen molecules into thermal energy by combustion, and then extracts part of that thermal energy to perform the work necessary to propel the car forward, overcoming friction. Thermodynamics is critical to all steps in this process (including determining the level of pollutants emitted), and a careful thermodynamic analysis is required for the design of fuel-efficient, low-polluting automobile engines. In general, thermodynamics plays a vital role in the design of any engine or power-generating plant, and therefore a good grounding in thermodynamics is required for much work in engineering.

If thermodynamics only governed the behavior of engines, it would probably be the most economically important of all sciences, but it is much more than that. Since the chemical and physical state of matter depends strongly on how much thermal energy it contains, thermodynamic principles play a central role in any description of the properties of matter. For example, thermodynamics allows us to understand why matter appears in different phases (solid, liquid, or gaseous), and under what conditions one phase will transform to another.

The composition of a chemically-reacting mixture which is given enough time to come to “equilibrium” is also fully determined by thermodynamic principles (even though thermodynamics alone can’t tell us how fast it will get there). For these reasons, thermodynamics lies at the heart of materials science, chemistry, and biology.

Thermodynamics in its original form (now known as *classical* thermodynamics) is a theory which is based on a set of postulates about how macroscopic matter behaves. This theory was developed in the 19th century, before the atomic nature of matter was accepted, and it makes no reference to atoms. The postulates (the most important of which are energy conservation and the impossibility of complete conversion of heat to useful work) can’t be derived within the context of classical, macroscopic physics, but if one accepts them, a very powerful theory results, with predictions fully in agreement with experiment.

When at the end of the 19th century it finally became clear that matter was composed of atoms, the physicist Ludwig Boltzmann showed that the postulates of classical thermodynamics emerged naturally from consideration of the microscopic atomic motion. The key was to give up trying to track the atoms individually and instead take a statistical, probabilistic approach, averaging over the behavior of a large number of atoms. Thus, the very successful postulates of classical thermodynamics were given a firm physical foundation. The science of *statistical mechanics* begun by Boltzmann encompasses everything in classical thermodynamics, but can do more also. When combined with quantum mechanics in the 20th century, it became possible to explain essentially all observed properties of macroscopic matter in terms of atomic-level physics, including esoteric states of matter found in neutron stars, superfluids, superconductors, etc. Statistical physics is also currently making important contributions in biology, for example helping to unravel some of the complexities of how proteins fold.

Even though statistical mechanics (or statistical thermodynamics) is in a sense “more fundamental” than classical thermodynamics, to analyze practical problems we usually take the macroscopic approach. For example, to carry out a thermodynamic analysis of an aircraft engine, it’s more convenient to think of the gas passing through the engine as a continuum fluid with some specified properties rather than to consider it to be a collection of molecules. But we *do* use statistical thermodynamics even here to calculate what the appropriate property values (such as the heat capacity) of the gas should be.

1.2 Energy and Entropy

The two central concepts of thermodynamics are *energy* and *entropy*. Most other concepts we use in thermodynamics, for example temperature and pressure, may actually be defined in terms of energy and entropy. Both energy and entropy are properties of physical systems, but they have very different characteristics. Energy is conserved: it can neither be produced nor destroyed, although it is possible to change its form or move it around. Entropy has a different character: it can't be destroyed, but it's easy to produce more entropy (and almost everything that happens actually does). Like energy, entropy too can appear in different forms and be moved around.

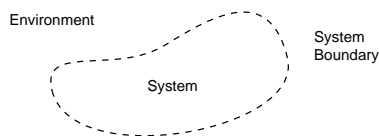
A clear understanding of these two properties and the transformations they undergo in physical processes is the key to mastering thermodynamics and learning to use it confidently to solve practical problems. Much of this book is focused on developing a clear picture of energy and entropy, explaining their origins in the microscopic behavior of matter, and developing effective methods to analyze complicated practical processes¹ by carefully tracking what happens to energy and entropy.

1.3 Some Terminology

Most fields have their own specialized terminology, and thermodynamics is certainly no exception. A few important terms are introduced here, so we can begin using them in the next chapter.

1.3.1 System and Environment

In thermodynamics, like in most other areas of physics, we focus attention on only a small part of the world at a time. We call whatever object(s) or region(s) of space we are studying the *system*. Everything else surrounding the system (in principle including the entire universe) is the *environment*. The boundary between the system and the environment is, logically, the *system boundary*. The starting point of any thermodynamic analysis is a careful definition of the system.



¹Rocket motors, chemical plants, heat pumps, power plants, fuel cells, aircraft engines, . . .

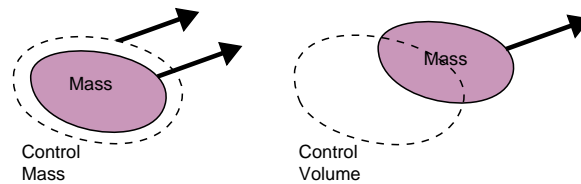


Figure 1.1: Control masses and control volumes.

1.3.2 Open, closed, and isolated systems

Any system can be classified as one of three types: open, closed, or isolated. They are defined as follows:

open system: Both energy and matter can be exchanged with the environment. Example: an open cup of coffee.

closed system: energy, but not matter, can be exchanged with the environment. Examples: a tightly capped cup of coffee.

isolated system: Neither energy nor matter can be exchanged with the environment – in fact, no interactions with the environment are possible at all. Example (approximate): coffee in a closed, well-insulated thermos bottle.

Note that no system can truly be isolated from the environment, since no thermal insulation is perfect and there are always physical phenomena which can't be perfectly excluded (gravitational fields, cosmic rays, neutrinos, etc.). But good approximations of isolated systems can be constructed. In any case, isolated systems are a useful conceptual device, since the energy and mass contained inside them stay constant.

1.3.3 Control masses and control volumes

Another way to classify systems is as either a *control mass* or a *control volume*. This terminology is particularly common in engineering thermodynamics.

A control mass is a system which is defined to consist of a specified piece or pieces of matter. By definition, no matter can enter or leave a control mass. If the matter of the control mass is moving, then the system boundary moves with it to keep it inside (and matter in the environment outside).

A control volume is a system which is defined to be a particular region of space. Matter and energy may freely enter or leave a control volume, and thus it is an open system.

1.4 A Note on Units

In this book, the SI system of units will be used exclusively. If you grew up anywhere but the United States, you are undoubtedly very familiar with this system. Even if you grew up in the US, you have undoubtedly used the SI system in your courses in physics and chemistry, and probably in many of your courses in engineering.

One reason the SI system is convenient is its simplicity. Energy, no matter what its form, is measured in Joules ($1 \text{ J} = 1 \text{ kg}\cdot\text{m}^2/\text{s}^2$). In some other systems, different units are used for thermal and mechanical energy: in the English system a BTU (“British Thermal Unit”) is the unit of thermal energy and a ft-lbf is the unit of mechanical energy. In the cgs system, thermal energy is measured in calories, all other energy in ergs. The reason for this is that these units were chosen before it was understood that thermal energy was like mechanical energy, only on a much smaller scale.²

Another advantage of SI is that the unit of force is identical to the unit of (mass \times acceleration). This is only an obvious choice if one knows about Newton’s second law, and allows it to be written as

$$\mathbf{F} = m\mathbf{a}. \quad (1.1)$$

In the SI system, force is measured in $\text{kg}\cdot\text{m}/\text{s}^2$, a unit derived from the 3 primary SI quantities for mass, length, and time (kg, m, s), but given the shorthand name of a “Newton.” The name itself reveals the basis for this choice of force units.

The units of the English system were fixed long before Newton appeared on the scene (and indeed were the units Newton himself would have used). The unit of force is the “pound force” (lbf), the unit of mass is the “pound mass” (lbm) and of course acceleration is measured in ft/s^2 . So Newton’s second law must include a dimensional constant which converts from Ma units ($\text{lbm ft}/\text{s}^2$) to force units (lbf). It is usually written

$$\mathbf{F} = \frac{1}{g_c}m\mathbf{a}, \quad (1.2)$$

where

$$g_c = 32.1739 \text{ ft}\cdot\text{lbm}/\text{lbf}\cdot\text{s}^2. \quad (1.3)$$

Of course, in SI $g_c = 1$.

²Mixed unit systems are sometimes used too. American power plant engineers speak of the “heat rate” of a power plant, which is defined as the thermal energy which must be absorbed from the furnace to produce a unit of electrical energy. The heat rate is usually expressed in BTU/kw-hr.

In practice, the units in the English system are now *defined* in terms of their SI equivalents (e.g. one foot is defined as a certain fraction of a meter, and one lbf is defined in terms of a Newton.) If given data in Engineering units, it is often easiest to simply convert to SI, solve the problem, and then if necessary convert the answer back at the end. For this reason, we will implicitly assume SI units in this book, and will not include the g_c factor in Newton's 2nd law.

CHAPTER 2

ENERGY, WORK, AND HEAT

2.1 Introduction

Energy is a familiar concept, but most people would have a hard time defining just what it is. You may hear people talk about “an energy-burning workout,” “an energetic personality,” or “renewable energy sources.” A few years ago people were very concerned about an “energy crisis.” None of these uses of the word “energy” corresponds to its scientific definition, which is the subject of this chapter.

The most important characteristic of energy is that it is *conserved*: you can move it around or change its form, but you can’t destroy it, and you can’t make more of it.¹ Surprisingly, the principle of conservation of energy was not fully formulated until the middle of the 19th century. This idea certainly does seem nonsensical to anyone who has seen a ball roll across a table and stop, since the kinetic energy of the ball seems to disappear. The concept only makes sense if you know that the ball is made of atoms, and that the macroscopic kinetic energy of motion is simply converted to microscopic kinetic energy of the random atomic motion.

2.2 Work and Kinetic Energy

Historically, the concept of energy was first introduced in mechanics, and therefore this is an appropriate starting point for our discussion. The basic equation of motion of classical mechanics is due to Newton, and is known as Newton’s second law.² Newton’s second law states that if a net force \mathbf{F} is applied to a body, its center-of-mass will experience an acceleration \mathbf{a} proportional to \mathbf{F} :

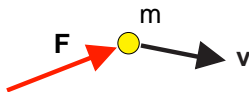
$$\mathbf{F} = m\mathbf{a}. \tag{2.1}$$

The proportionality constant m is the *inertial mass* of the body.

¹Thus, energy can’t be burned (fuel is burned), it is a property matter has (not personalities), there are no sources of it, whether renewable or not, and there is no energy crisis (but there may be a usable energy, or availability, crisis).

²For now we consider only classical, nonrelativistic mechanics.

Suppose a single external force \mathbf{F} is applied to point particle moving with velocity \mathbf{v} . The force is applied for an infinitesimal time dt , during which the velocity changes by $d\mathbf{v} = \mathbf{a} dt$, and the position changes by $d\mathbf{x} = \mathbf{v} dt$.



Taking the scalar product³ (or dot product) of Eq. (2.1) with $d\mathbf{x}$ gives

$$\begin{aligned} \mathbf{F} \cdot d\mathbf{x} &= m\mathbf{a} \cdot d\mathbf{x} \\ &= \left[m \left(\frac{d\mathbf{v}}{dt} \right) \right] \cdot [\mathbf{v} dt] \\ &= m\mathbf{v} \cdot d\mathbf{v} \\ &= d(mv^2/2). \end{aligned} \quad (2.2)$$

Here $v = |\mathbf{v}|$ is the particle speed. Note that only the component of \mathbf{F} along the direction the particle moves is needed to determine whether v increases or decreases. If this component is parallel to $d\mathbf{x}$, the speed increases; if it is antiparallel to $d\mathbf{x}$ the speed decreases. If \mathbf{F} is perpendicular to $d\mathbf{x}$, then the speed doesn't change, although the direction of \mathbf{v} may.

Since we'll have many uses for $\mathbf{F} \cdot d\mathbf{x}$ and $mv^2/2$, we give them symbols and names. We call $\mathbf{F} \cdot d\mathbf{x}$ the infinitesimal *work* done by force \mathbf{F} , and give it the symbol dW :

$$dW = \mathbf{F} \cdot d\mathbf{x} \quad (2.3)$$

(We'll see below why we put a bar through the d in dW .)

The quantity $mv^2/2$ is the *kinetic energy* E_k of the particle:

$$E_k = \frac{mv^2}{2} \quad (2.4)$$

With these symbols, Eq. (2.2) becomes

$$dW = d(E_k). \quad (2.5)$$

Equation (2.5) may be interpreted in thermodynamic language as shown in Fig. 2.1. A system is defined which consists only of the particle; the energy

³Recall that the scalar product of two vectors $\mathbf{A} = \mathbf{i} A_i + \mathbf{j} A_j + \mathbf{k} A_k$ and $\mathbf{B} = \mathbf{i} B_i + \mathbf{j} B_j + \mathbf{k} B_k$ is defined as $\mathbf{A} \cdot \mathbf{B} = A_i B_i + A_j B_j + A_k B_k$. Here \mathbf{i} , \mathbf{j} , and \mathbf{k} are unit vectors in the x , y , and z directions, respectively.

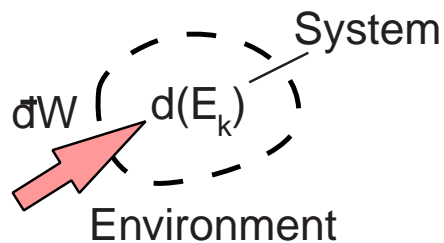
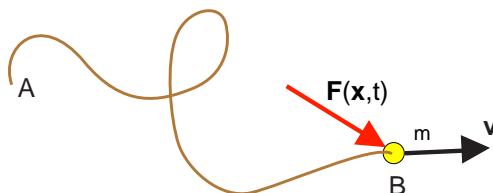


Figure 2.1: Energy accounting for a system consisting of a single point particle acted on by a single force for time dt .

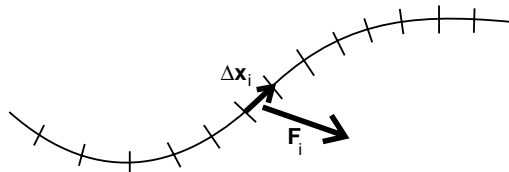
“stored” within the system (here just the particle kinetic energy) increases by $d(E_k)$ due to the work dW done by external force \mathbf{F} . Since force \mathbf{F} is produced by something outside the system (in the environment), we may regard dW as an *energy transfer* from the environment to the system. Thus, *work is a type of energy transfer*. Of course, dW might be negative, in which case $d(E_k) < 0$. In this case, the direction of energy transfer is actually from the system to the environment.

The process of equating energy transfers to or from a system to the change in energy stored in a system we will call *energy accounting*. The equations which result from energy accounting we call *energy balances*. Equation (2.5) is the first and simplest example of an energy balance – we will encounter many more.

If the force \mathbf{F} is applied for a finite time t , the particle will move along some trajectory $\mathbf{x}(t)$.



The change in the particle kinetic energy $\Delta E_k = E_k(B) - E_k(A)$ can be determined by dividing the path into many very small segments, and summing Eq. (2.2) for each segment.



In the limit where each segment is described by an infinitesimal vector $d\mathbf{x}$,

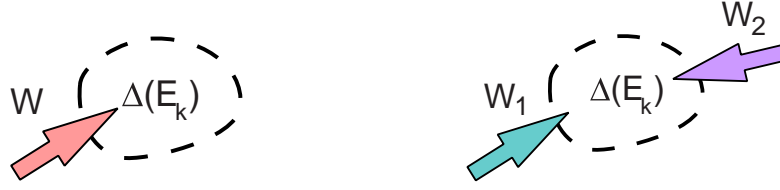


Figure 2.2: Energy accounting for a single particle acted on by (a) a single force (b) multiple forces for finite time.

the sum becomes an integral:

$$\int_{path} dW = \int_{path} d(E_k) \quad (2.6)$$

The right-hand side of this can be integrated immediately:

$$\int_{path} d(E_k) = \Delta E_k. \quad (2.7)$$

The integral on the left-hand side defines the total work done by \mathbf{F} :

$$W = \int_{path} dW = \int_{path} \mathbf{F} \cdot d\mathbf{x}. \quad (2.8)$$

Note that the integral is along the particular path taken. Eq. (2.6) becomes

$$W = \Delta E_k. \quad (2.9)$$

The thermodynamic interpretation of this equation is shown in Fig. 2.2 and is similar to that of Eq. (2.5): work is regarded as a transfer of energy to the system (the particle), and the energy stored in the system increases by the amount transferred in. (Again, if $W < 0$, then the direction of energy transfer is really from the system to the environment, and in this case $\Delta E_k < 0$.)

If two forces act simultaneously on the particle, then the total applied force is the vector sum: $\mathbf{F} = \mathbf{F}_1 + \mathbf{F}_2$. In this case, Eq. (2.9) becomes

$$W_1 + W_2 = \Delta E_k, \quad (2.10)$$

where $W_1 = \int_{path} \mathbf{F}_1 \cdot d\mathbf{x}$ and $W_2 = \int_{path} \mathbf{F}_2 \cdot d\mathbf{x}$.⁴ The generalization to N forces is obvious: the work done by all N forces must be considered to compute ΔE_k .

⁴For now we're considering a point particle, so the path followed is the same for both forces; this won't be true for extended objects, which will be considered in section 2.4.

2.3 Evaluation of Work

Since in general a force may depend on factors such as the instantaneous particle position \mathbf{x} , the instantaneous velocity \mathbf{v} , or may depend explicitly on time, the work done by the force will clearly depend on *the path* the particle takes from A to B, *how fast* it travels, and the particular *time* it passes each point. Since there are infinitely many possible trajectories $\mathbf{x}(t)$ which start at point A at some time and pass through point B at some later time, there are infinitely many possible values for $W = \int_{path} dW$; we need additional information [i.e., $\mathbf{x}(t)$] to evaluate W .

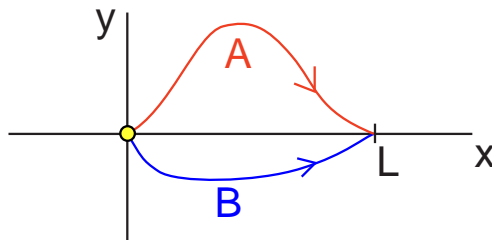
This is the reason we put the bar through dW but not through $d(E_k)$. It's always true that $\int_{path} d(Q)$ may be formally evaluated to yield $Q_B - Q_A$, where Q is some function of the state (position, velocity, etc.) of the particle and of time, and Q_A and Q_B denote the values of Q when the particle is at endpoints of the path.

But dW is not like this: it's only the symbol we use to denote "a little bit of work." It really equals $\mathbf{F} \cdot d\mathbf{x}$, which is not of the form $d(Q)$, so can't be integrated without more information. Quantities like dW are known as "inexact differentials." We put the bar in dW just to remind ourselves that it is an inexact differential, and so its integral depends on the particular path taken, not only on the state of the particle at the beginning and end of the path.

Example 2.1 The position-dependent force

$$\mathbf{F}(x, y, z) = \begin{cases} +iC & \text{if } y > 0 \\ -i2C & \text{if } y \leq 0 \end{cases}$$

is applied to a bead on a frictionless wire. The bead sits initially at the origin, and the wire connects the origin with $(L, 0, 0)$. How much work does \mathbf{F} do to move the bead along wire A? How much along wire B? Does the contact force of the bead against the wire do any work?



Solution:

$$W = \int_{path} \mathbf{F}(x, y, z) \cdot d\mathbf{x}.$$

Since \mathbf{F} always points in the x direction,

$$\mathbf{F}(x, y, z) \cdot d\mathbf{x} = F_x(x, y, z)dx$$

Therefore, along path A, $W = CL$, and along path B, $W = -2CL$.

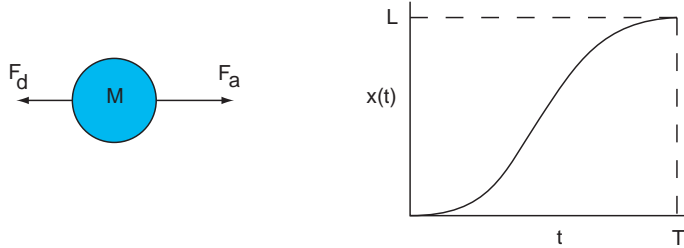
Along path A, the force does work on the particle, while along path B the particle does work on whatever is producing the force. Of course, for motion along path B to be possible at all, the particle would have to have an initial kinetic energy greater than $2CL$. The contact force does no work, since it is always perpendicular to the wire (and therefore to $d\mathbf{x}$), so $\mathbf{F}_{\text{contact}} \cdot d\mathbf{x} = 0$.

If we do know $\mathbf{x}(t)$, we can convert the path integral definition of work [Eq. (2.8)] into a time integral, using $d\mathbf{x} = \mathbf{v}(t)dt$:

$$W = \int_{t_A}^{t_B} \mathbf{F}(\mathbf{x}(t), \mathbf{v}(t), t) \cdot \mathbf{v}(t) dt \quad (2.11)$$

This is often the easiest way to evaluate work. Note that the integrand is $\mathbf{F} \cdot \mathbf{v}$. Therefore, $\mathbf{F} \cdot \mathbf{v}$ is the *rate* at which force \mathbf{F} does work, or in other words the instantaneous *power* being delivered by \mathbf{F} . We denote the power by \dot{W} :

$$\dot{W} = \mathbf{F} \cdot \mathbf{v} \quad (2.12)$$



Example 2.2

A ball initially at rest at $x = 0$ in a viscous fluid is pulled in a straight line by a string. A time-dependent force $F_a(t)$ is applied to the string, which causes the ball to move according to

$$x(t) = \frac{L}{2} \left[1 - \cos\left(\frac{\pi t}{T}\right) \right].$$

At time $t = T$, the ball comes to rest at $x = L$ and the force is removed. As the ball moves through the fluid, it experiences a drag force proportional to its

speed: $F_d = -C\dot{x}(t)$. How much work is done by the applied force to move the ball from $x = 0$ to $x = L$?

Solution: Newton's second law requires

$$F_a + F_d = m\ddot{x}(t), \quad (2.13)$$

so

$$F_a(t) = m\ddot{x}(t) + C\dot{x}(t). \quad (2.14)$$

Since we know $x(t)$, we can differentiate to find

$$\dot{x}(t) = \frac{L}{2} \left(\frac{\pi}{T} \right) \sin \tau \quad (2.15)$$

and

$$\ddot{x}(t) = \frac{L}{2} \left(\frac{\pi}{T} \right)^2 \cos \tau \quad (2.16)$$

where $\tau = \pi t/T$. Substituting these expressions into Eq. (2.14) results in

$$F_a(t) = \frac{CL}{2} \left(\frac{\pi}{T} \right) \sin \tau + \frac{mL}{2} \left(\frac{\pi}{T} \right)^2 \cos \tau.$$

To calculate the work done by $F_a(t)$, we need to evaluate

$$W_a = \int_{path} \mathbf{F}_a \cdot d\mathbf{x} = \int_0^L F_a dx.$$

Since we know both $F_a(t)$ and $x(t)$, it is easiest to convert this path integral to a time integral using $dx = \dot{x}(t)dt$:

$$W_a = \int_0^T F_a(t)\dot{x}(t) dt.$$

Changing the integration variable to τ ($d\tau = (\pi/T)dt$),

$$W_a = \left(\frac{L}{2} \right)^2 \frac{\pi}{T} \int_0^\pi \left[C \sin^2 \tau + \left(\frac{\pi}{T} \right) \sin \tau \cos \tau \right] d\tau.$$

Since $\int_0^\pi \sin^2 \tau d\tau = \pi/2$ and $\int_0^\pi \sin \tau \cos \tau d\tau = 0$,

$$W_a = \frac{\pi^2 CL^2}{8T}.$$

If there were no drag ($C = 0$), then the work would be zero, since the work done to accelerate the ball for $t < T/2$ would be fully recovered in decelerating the ball for $t > T/2$. But in the presence of a drag force, a finite amount of work must be done to overcome drag, even though the ball ends as it began with no kinetic energy.

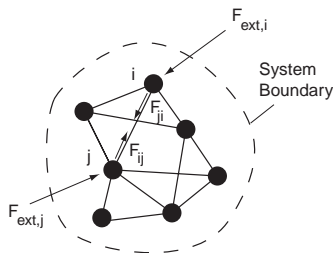


Figure 2.3: External and internal forces acting on two masses of a rigid body.

Note that the work is inversely proportional to the total time T . It takes more work to push the ball rapidly through the fluid (short T) than slowly. By carrying out the process very slowly, it is possible to make W_a as small as desired, and in the limit of $T \rightarrow \infty$ the process requires no work. This behavior is characteristic of systems which exhibit viscous drag.

2.4 Energy Accounting for Rigid Bodies

Up until now we have only considered how to do energy accounting for point masses. To develop energy accounting methods for macroscopic matter, we can use the fact that macroscopic objects are composed of a very large number of what we may regard as point masses (atomic nuclei), connected by chemical bonds. In this section, we consider how to do energy accounting on a macroscopic object if we make the simplifying assumption that the bonds are completely rigid. We'll relax this assumption and complete the development of energy accounting for macroscopic matter in section 2.8.

Consider a body consisting of N point masses connected by rigid, massless rods, and define the system to consist of the body (Fig. 2.3). The rods will transmit forces between the masses. We will call these forces *internal forces*, since they act between members of the system. We will assume the internal forces are directed along the rods. The force exerted on (say) mass j by mass i will be exactly equal in magnitude and opposite in direction to that exerted on mass i by mass j ($\mathbf{F}_{ij} = -\mathbf{F}_{ji}$), since otherwise there would be a force imbalance on the rod connecting i and j . No force imbalance can occur, since the rod is massless and therefore would experience infinite acceleration if the forces were unbalanced. (Note this is Newton's third law.)

Let the masses composing the body also be acted on by arbitrary *external* forces from the environment. The external force on mass i will be denoted

$\mathbf{F}_{ext,i}$.

The energy balance in differential form for one mass, say mass i , is

$$dW_{ext,i} + \left(\sum_j \mathbf{F}_{ji} \right) \cdot d\mathbf{x}_i = d(E_{k,i}), \quad (2.17)$$

where $dW_{ext,i} = \mathbf{F}_{ext,i} \cdot d\mathbf{x}_i$ and of course $\mathbf{F}_{ii} = 0$. Summing the energy balances for all masses results in an energy balance for the entire system:

$$\sum_i dW_{ext,i} + \sum_i \sum_j \mathbf{F}_{ji} \cdot d\mathbf{x}_i = d(E_k), \quad (2.18)$$

where

$$d(E_k) = \sum_i d(E_{k,i}) = \sum_i d(m_i v_i^2 / 2) \quad (2.19)$$

is the change in the total kinetic energy of the body.

Equation (2.18) can be simplified considerably, since the second term on the left is *exactly zero*. To see this, recall that the rods are rigid, so

$$d(|\mathbf{x}_i - \mathbf{x}_j|) = 0 \quad (2.20)$$

for all i and j . Equation (2.20) can be written as

$$(\mathbf{x}_i - \mathbf{x}_j) \cdot d(\mathbf{x}_i - \mathbf{x}_j) = 0. \quad (2.21)$$

Now \mathbf{F}_{ij} is parallel to $(\mathbf{x}_i - \mathbf{x}_j)$, so multiplying Eq. (2.21) by $|\mathbf{F}_{ij}|/|\mathbf{x}_i - \mathbf{x}_j|$ results in

$$\mathbf{F}_{ij} \cdot d(\mathbf{x}_i - \mathbf{x}_j) = 0. \quad (2.22)$$

Since $\mathbf{F}_{ji} = -\mathbf{F}_{ij}$, we can re-write this as

$$\mathbf{F}_{ji} \cdot d\mathbf{x}_i = -\mathbf{F}_{ij} \cdot d\mathbf{x}_j. \quad (2.23)$$

Therefore, because the body is rigid, the work done by \mathbf{F}_{ji} on mass i is precisely equal to the negative of the work done by \mathbf{F}_{ij} on mass j . Thus, the internal forces \mathbf{F}_{ij} cause a transfer of kinetic energy from one mass within the body to another, but considering the body as a whole, do no *net* work on the body.

Mathematically, the second term on the left of Eq. (2.18) is a sum over all pairs of mass indices (i, j) . Because of Eq. (2.23), for every i and j , the (i, j) term in this sum will exactly cancel the (j, i) term, with the result that the double sum is zero.

With this simplification (for rigid bodies), Eq. (2.18) reduces to

$$\sum_i dW_{ext,i} = d(E_k). \quad (2.24)$$

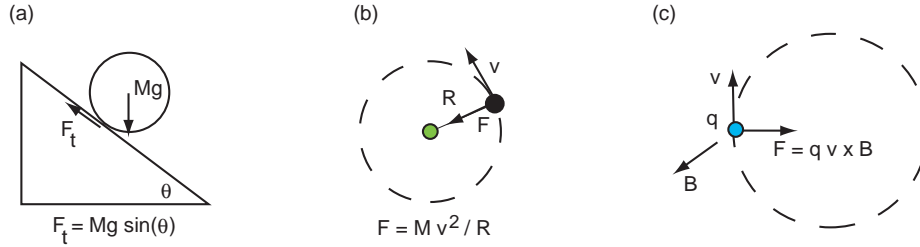


Figure 2.4: Some forces which do no work: (a) traction force on a rolling wheel; (b) centrifugal force; (c) Lorentz force on a charged particle in a magnetic field

We see that to carry out an energy balance on a rigid body, *we only need consider work done by external forces*, not by internal ones. We can always tell which forces are external ones – they are the ones which cross the system boundary on a sketch.

A macroscopic solid object is composed of a huge number of essentially point masses (the atomic nuclei) connected by chemical bonds (actually rapidly moving, quantum-mechanically smeared out electrons). If we ignore for the moment the fact that bonds are not really rigid, a solid object can be approximated as a rigid body. If this approximation holds, then the appropriate energy balance equation will be Eq. (2.24).

For simplicity, assume that the external forces act only at L discrete locations on the surface of the object, where it contacts the environment.⁵ In this case, the external work term in Eq. (2.24) becomes $\sum_{\ell=1}^L \mathbf{F}_{\ell} \cdot d\mathbf{x}_{\ell}$, where $d\mathbf{x}_{\ell}$ is the displacement of the surface of the object *at the point where the force \mathbf{F}_{ℓ} is applied*. The energy balance Eq. (2.24) becomes

$$\sum_{\ell=1}^L \mathbf{F}_{\ell} \cdot d\mathbf{x}_{\ell} = d(E_k). \quad (2.25)$$

It is very important to remember that the displacements to use in this equation are those *where the forces are applied*, and may differ for each force. Do not make the mistake of using the displacement of some other point (e.g. the center of mass).

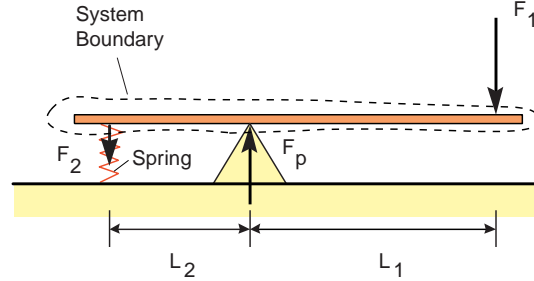
If a force is applied to a macroscopic object at a point where it is stationary, the force does no work no matter how large the force is. (If you push against a stationary wall, you may exert yourself, but you do no work on it.) Also, a force

⁵If the macroscopic force is exerted over some small but finite contact area, the macroscopic force \mathbf{F}_{ℓ} in Eq. (2.25) is simply the sum over the atomic-level forces $\mathbf{F}_{ext,i}$ in Eq. (2.24) for all atoms i in the contact area.

applied *perpendicular* to the instantaneous direction of motion of the contact area can do no work.

Some common forces which do no work are shown in Fig. 2.4. A traction force $|\mathbf{F}_t| = mg \sin \theta$ in the plane of the surface keeps a rolling wheel from sliding down a hill; but since the wheel is instantaneously stationary where it contacts the ground, $\mathbf{F}_t \cdot d\mathbf{x} = 0$ and therefore the traction force does no work. A centrifugal force and the Lorentz force a charged particle experiences in a magnetic field are both perpendicular to the direction of motion, and thus can do no work.

Example 2.3 A downward force \mathbf{F}_1 is applied to a rigid, horizontal lever a distance L_1 to the right of the pivot point. A spring connects the lever to the ground at a distance L_2 to the left of the pivot, and exerts a downward force \mathbf{F}_2 . An upward force \mathbf{F}_p is exerted on the lever at the pivot. Evaluate the work done by each force if end 2 is raised by dy_2 , and determine the value of \mathbf{F}_1 which achieves this motion without changing the kinetic energy of the lever.



Solution: Define the system to consist of the lever only (a rigid body). The body is acted on by three external forces, and so we must evaluate the work input to the system from each force. Since the lever is rigid, if the height of end 2 changes by dy_2 while the height at the pivot point is unchanged, then the height of end 1 must change by $dy_1 = -(L_1/L_2)dy_2$. So the three work inputs are:

$$dW_1 = (-\mathbf{j}F_1) \cdot (-\mathbf{j}L_1 dy_2/L_2) = (F_1 L_1/L_2) dy_2 > 0 \quad (2.26)$$

$$dW_2 = (-\mathbf{j}F_2) \cdot (+\mathbf{j}dy_2) = -F_2 dy_2 < 0 \quad (2.27)$$

$$dW_p = (+\mathbf{j}F_p) \cdot (0) = 0. \quad (2.28)$$

Note that the work due to the pivot force is zero, since the lever does not move at the pivot. Force \mathbf{F}_1 does positive work on the lever, since the force and displacement are in the same direction. The spring which produces force \mathbf{F}_2 does *negative* work on the lever, since the force and displacement are in opposite directions. In this case, we say that the lever does positive work on the spring, since the force exerted *by* the lever *on* the spring is oppositely directed

to \mathbf{F}_2 (Newton's third law).

The energy balance on the lever is then

$$\begin{aligned} dW_1 + dW_2 + dW_p &= d(E_k) \\ (F_1 L_1 / L_2 - F_2) dy_2 &= d(E_k). \end{aligned} \quad (2.29)$$

If we wish to move the lever without increasing its kinetic energy, then we must choose

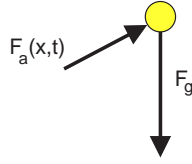
$$F_1 L_1 = F_2 L_2. \quad (2.30)$$

This is the familiar law of the lever, but note that we obtained it from an energy balance, not by balancing torques as would be done in mechanics.

2.5 Conservative Forces and Potential Energy

2.5.1 A Uniform Gravitational Field

Suppose a point particle near the surface of the earth is acted on by gravity, which exerts a constant downward force $\mathbf{F}_g = -\mathbf{j}mg$. It is also acted on by an arbitrary external applied force $\mathbf{F}_a(\mathbf{x}, t)$.



In this case, Eq. (2.10) becomes

$$W_a + W_g = \Delta E_k \quad (2.31)$$

where

$$W_a = \int_{path} \mathbf{F}_a \cdot d\mathbf{x} \quad (2.32)$$

is the work done by the applied force, and

$$W_g = \int_{path} \mathbf{F}_g \cdot d\mathbf{x} \quad (2.33)$$

is the work done by the gravitational force. Due to the special character of \mathbf{F}_g (a constant force), W_g can be evaluated for an arbitrary path from A to B:

$$W_g = - \int_{path} \mathbf{j}mg \cdot d\mathbf{x}$$

$$\begin{aligned}
&= - \int_{y_A}^{y_B} mg dy \\
&= -mg(y_B - y_A) \tag{2.34}
\end{aligned}$$

$$= -mg\Delta y. \tag{2.35}$$

If $\Delta y < 0$, gravity does work *on* the particle, and its kinetic energy increases. If $\Delta y > 0$, $W_g < 0$, which means that the particle must do work against gravity. In this case the kinetic energy decreases.

Note that W_g can be expressed solely in terms of the difference in a property (the height) of the particle at the beginning and end of its trajectory: *any* path connecting A and B would result in the same value for W_g . This is due to the special nature of the force \mathbf{F}_g , which is just a constant. Of course, for an arbitrary force such as $\mathbf{F}_a(\mathbf{x}, t)$, this would not be possible. The force \mathbf{F}_g is the first example of a *conservative force*.

Since W_g is independent of the particular path taken, we can bring it to the other side of Eq. (2.31):

$$\begin{aligned}
W_a &= (-W_g) + \Delta E_k \\
&= mg\Delta y + \Delta E_k \\
&= \Delta(E_k + mgy) \tag{2.36}
\end{aligned}$$

We define mgy to be the *gravitational potential energy* E_g of the particle in this uniform gravitational field:

$$E_g = mgy. \tag{2.37}$$

With this definition, Eq. (2.31) becomes

$$W_a = \Delta(E_k + E_g). \tag{2.38}$$

Equations (2.31) and (2.38) are mathematically equivalent, but have different interpretations, as shown in Fig. 2.5. In Eq. (2.31), the gravitational force is considered to be an *external* force acting on the system; the work W_g it does on the system is included in the energy balance but not any potential energy associated with it. In (b), the source of the gravitational force (the gravitational field) is in effect considered to be part of the system. Since it is now internal to the system, we don't include a work term for it, but *do* include the gravitational potential energy (which we may imagine to be stored in the field) in the system energy. It doesn't matter which point of view we take – the resulting energy balance is the same because ΔE_g is defined to be identical to

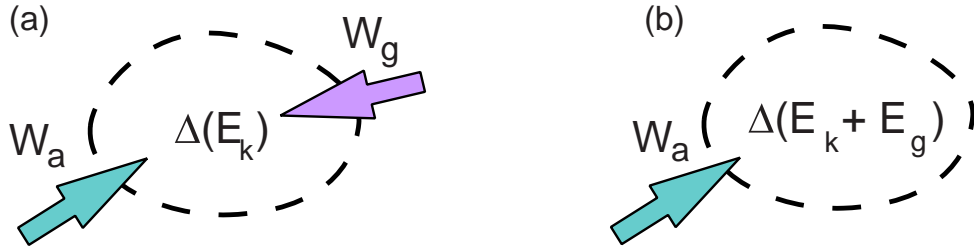


Figure 2.5: Two energy accounting schemes to handle the effects of a constant gravitational force. In (a), the gravitational field is considered to be external to the system, while in (b) the field is part of the system.

$-W_g$. But remember not to *mix* these points of view: don't include *both* W_g and ΔE_g in an energy balance!

We may generalize this analysis to a macroscopic body. In this case, the gravitational potential energy becomes

$$E_g = \int_{body} \rho(\mathbf{x}) g y dV, \quad (2.39)$$

where $\rho(\mathbf{x})$ is the local *mass density* (kg/m^3) at point \mathbf{x} within the body. This can be re-written as

$$E_g = M g y_{cm}, \quad (2.40)$$

where

$$M = \int_{body} \rho(\mathbf{x}) dV \quad (2.41)$$

is the total mass of the body and y_{cm} is the *y*-component of the *center of mass*, defined by

$$\mathbf{x}_{cm} = \frac{1}{M} \int \rho(\mathbf{x}) \mathbf{x} dV. \quad (2.42)$$

2.5.2 General Conservative Forces

A constant force, such as discussed above, is the simplest example of a conservative force. The general definition is as follows:

a force is conservative if and only if the work done by it in going from an initial position \mathbf{x}_A to a final position \mathbf{x}_B depends only on the initial and final positions, and is *independent of the path taken*.

Mathematically, this definition may be stated as follows:

$$W_c = \int_{path} \mathbf{F}_c \cdot d\mathbf{x} = f(\mathbf{x}_B) - f(\mathbf{x}_A), \quad (2.43)$$

where f is some single-valued scalar function of position in space.

For the special case of a closed path ($\mathbf{x}_B = \mathbf{x}_A$), Eq. (2.43) reduces to

$$\oint \mathbf{F}_c \cdot d\mathbf{x} = 0, \quad (2.44)$$

where \oint denotes integrating all the way around the path. Therefore, the work done by a conservative force on a particle traversing any arbitrary closed loop is exactly zero. Either Eq. (2.43) or Eq. (2.44) may be taken as the definition of a conservative force.

Only very special functions $\mathbf{F}(\mathbf{x}, \mathbf{v}, t)$ can satisfy the conditions for a conservative force. First of all, consider the dependence on velocity. The only way Eq. (2.44) can be satisfied by a velocity-dependent force for all possible loops, traversing the loop in either direction at arbitrary speed, is if the velocity-dependent force does no work. This is possible if $\mathbf{F}(\mathbf{x}, \mathbf{v}, t)$ is always perpendicular to \mathbf{v} . Thus, any conservative force can have an arbitrary velocity-dependent force \mathbf{F}_v added to it and still be conservative *as long as* $\mathbf{F}_v \cdot \mathbf{v} = 0$ at all times.

It seems that in nature there is only one velocity-dependent conservative force, which is the Lorentz force felt by a charged particle moving through a magnetic field \mathbf{B} . This Lorentz force is given by

$$\mathbf{F}_L = q\mathbf{v} \times \mathbf{B}, \quad (2.45)$$

which is always perpendicular to both \mathbf{v} and \mathbf{B} . Unless stated otherwise, we will assume from here on that conservative forces do not have a velocity-dependent part, keeping in mind that the Lorentz force is the one exception.

Having dealt with the allowed type of velocity dependence, consider now the time dependence. It is clear that \mathbf{F}_c can have no *explicit* time dependence (i.e., $\mathbf{F}(\mathbf{x}(t))$ is OK but $\mathbf{F}(\mathbf{x}(t), t)$ is not). If F_c depended explicitly on time, then the result for W_c would too, rather than on just the endpoint positions in space. So we conclude that a conservative force (or at least the part which can do work) can depend explicitly only on position: $\mathbf{F}_c(\mathbf{x})$.

2.5.3 How to Tell if a Force is Conservative

If we are given a force function $\mathbf{F}(\mathbf{x})$, how can we tell if it is conservative? First consider the inverse problem: If we know the function $f(\mathbf{x})$, can we derive what \mathbf{F}_c must be? Consider a straight-line path which has infinitesimal length: $\mathbf{x}_B = \mathbf{x}_A + d\mathbf{x}$. Then equation 2.43 reduces to

$$\mathbf{F}_c(\mathbf{x}_A) \cdot d\mathbf{x} = f(\mathbf{x}_A + d\mathbf{x}) - f(\mathbf{x}_A). \quad (2.46)$$

Since $d\mathbf{x}$ is infinitesimal, we may expand $f(\mathbf{x}_A + d\mathbf{x})$ in a Taylor series:⁶

$$f(\mathbf{x}_A + d\mathbf{x}) = f(\mathbf{x}_A) + \nabla f(\mathbf{x}_A) \cdot d\mathbf{x} + O(|d\mathbf{x}|^2), \quad (2.47)$$

where the gradient of f is defined by

$$\nabla f = \mathbf{i} \frac{\partial f}{\partial x} + \mathbf{j} \frac{\partial f}{\partial y} + \mathbf{k} \frac{\partial f}{\partial z}. \quad (2.48)$$

As we let $|d\mathbf{x}|$ go to zero, the higher-order terms go to zero rapidly, so Eq. (2.46) becomes

$$\mathbf{F}_c(\mathbf{x}) \cdot d\mathbf{x} = \nabla f(\mathbf{x}_A) \cdot d\mathbf{x} \quad (2.49)$$

The only way this equation can hold for arbitrary \mathbf{x}_A and $d\mathbf{x}$ is if

$$\boxed{\mathbf{F}_c(\mathbf{x}) = \nabla f(\mathbf{x})}. \quad (2.50)$$

Therefore, a conservative force which depends only on position must be the gradient of some scalar function of position in space $f(\mathbf{x})$.

How can we tell if a given vector function $\mathbf{F}(\mathbf{x})$ is the gradient of some unknown scalar function $f(\mathbf{x})$? The easiest way is to write them both out explicitly:

$$\mathbf{F}(x, y, z) = \mathbf{i}F_i(x, y, z) + \mathbf{j}F_j(x, y, z) + \mathbf{k}F_k(x, y, z) \quad (2.51)$$

$$\nabla f(x, y, z) = \mathbf{i} \frac{\partial f}{\partial x} + \mathbf{j} \frac{\partial f}{\partial y} + \mathbf{k} \frac{\partial f}{\partial z}. \quad (2.52)$$

If these are equal, then each component must be equal, so

$$F_i(x, y, z) = \partial f(x, y, z) / \partial x \quad (2.53)$$

$$F_j(x, y, z) = \partial f(x, y, z) / \partial y \quad (2.54)$$

$$F_k(x, y, z) = \partial f(x, y, z) / \partial z. \quad (2.55)$$

Consider now the mixed second derivatives of $f(x, y, z)$. It doesn't matter which order we do the differentiation:

$$\frac{\partial}{\partial x} \left(\frac{\partial f}{\partial y} \right) = \frac{\partial}{\partial y} \left(\frac{\partial f}{\partial x} \right) = \frac{\partial^2 f}{\partial x \partial y}, \quad (2.56)$$

with similar results for the partial derivatives involving z . Therefore, if $\mathbf{F} = \nabla f$, we may substitute eqs. (2.53) and (2.54) into Eq. (2.56) and obtain

$$\frac{\partial F_j}{\partial x} = \frac{\partial F_i}{\partial y}. \quad (2.57)$$

⁶If this is not clear to you in vector form, write it out component by component.

Similarly,

$$\frac{\partial F_i}{\partial z} = \frac{\partial F_k}{\partial x}, \quad (2.58)$$

and

$$\frac{\partial F_j}{\partial z} = \frac{\partial F_k}{\partial y}, \quad (2.59)$$

Equations (2.57)–(2.59) provide a simple test to determine if $\mathbf{F}(\mathbf{x})$ is conservative. If \mathbf{F} passes this test, it should be possible to integrate equations (2.53)–(2.55) and find a function $f(\mathbf{x})$ such that $\mathbf{F} = \nabla f$. If \mathbf{F} fails the test, then no such $f(\mathbf{x})$ exists.

2.5.4 Energy Accounting with Conservative Forces

We can easily generalize the analysis of the mass in a constant gravitational field to handle an arbitrary conservative force acting on a particle. The energy balance is

$$W_a + W_c = \Delta E_k. \quad (2.60)$$

Since the force is conservative, $W_c = f(\mathbf{x}_B) - f(\mathbf{x}_A) = \Delta f$. Therefore, we may write the energy balance as

$$W_a = \Delta E_k - \Delta f = \Delta(E_k - f). \quad (2.61)$$

Now define the potential energy associated with this conservative force as follows:

$$E_p(\mathbf{x}) = -f(\mathbf{x}) + C. \quad (2.62)$$

Since only differences in potential energy have any physical significance, we can set the additive constant C to any convenient value. The energy balance now becomes

$$W_a = \Delta(E_k + E_p). \quad (2.63)$$

As with the gravitation example, the energy balances (2.60) and (2.63) are completely equivalent mathematically, and we can use whichever one we prefer. They differ only in interpretation. Using Eq. (2.60), we regard whatever produces the conservative force (e.g. a gravitational, electric, or magnetic field, a frictionless spring, etc.) as part of the *environment* – external to the system. Therefore, we include the work W_c done by this force on our system when we do energy accounting. If we write the energy balance as in Eq. (2.63), we are regarding the source of the conservative force as part of the system. Since in this case the force becomes an internal one, we don't include the work W_c in the energy balance, but we must account for the potential energy stored in the field or spring as part of the system energy.

2.6 Elementary Forces and Conservation of Energy

Elementary forces are those forces which are part of the basic structure of physics, such as the gravitational force, electromagnetic forces, nuclear forces, etc. These forces are responsible for all atomic-level or subatomic behavior, including chemical and nuclear bonding and the forces atoms feel when they collide with one another. (But quantum mechanics, rather than classical mechanics, must be used to correctly predict these features).

As far as we know now, *every* elementary force of nature is conservative – that is, it may be derived from some potential energy function. Considering how special conservative forces are (there are infinitely more functions $\mathbf{F}(\mathbf{x})$ which are *not* the gradient of some $f(\mathbf{x})$ than there are functions which are), this can be no accident – it must be a deep principle of physics.

The universe can be thought of as a very large number of elementary particles interacting through conservative, elementary forces. If we do an energy accounting for the entire universe, treating the conservative interactions between particles by adding appropriate potential energy terms to the system energy as discussed in section 2.5.4, we find⁷

$$\Delta(E_k + E_p) = 0, \quad (2.64)$$

where E_k and E_p represent the kinetic and potential energies, respectively, of the entire universe. Of course there can be no external work term, since the entire universe is inside our system!

Therefore, the total energy of the universe (kinetic + all forms of potential) is constant. Everything that has happened since the birth of the universe — its expansion, the condensation of protons and electrons to form hydrogen gas, the formation of stars and heavy nuclei within them, the formation of planets, the evolution of life on earth, you reading this book — all of these processes simply shift some energy from one type to another, never changing the total.

The constancy of the energy of the universe is the *principle of conservation of energy*. Of course, any small part of the universe which is isolated from the rest in the sense that no energy enters or leaves it will also have constant total energy. Another way of stating the principle of conservation of energy is that there are no sinks or sources for energy — you can move it around or change its form, but you can't create it, and you can't destroy it.

⁷Of course, to calculate E_k and E_p correctly we would have to consider not only quantum mechanics but general relativity. These change the details in important ways, but not the basic result that the energy of the universe is constant.

Why is the energy of the universe constant? This is equivalent to asking *why* all elementary forces are conservative. Quantum mechanics provides some insight into this question. In quantum mechanics, a system has a well-defined constant total energy if two conditions are met: a) there are no interactions with external forces, and b) the laws governing the elementary forces are constant in time. If this is applied to the whole universe condition a) is automatically satisfied, and b) says simply that the basic laws of physics have always been the same as they are now. As far as we know, this is true – the laws of physics don't depend on time.

2.7 Non-Conservative Forces

Since all elementary forces are conservative, it might be thought that any macroscopic forces between macroscopic objects (which, after all, are composed of elementary particles interacting through elementary forces) should be conservative. This is actually *not true*, as a simple thought experiment demonstrates.

Imagine sliding an object around in a circle on a table, returning to the starting point. If the table were perfectly frictionless, it would take no net work to do this, since any work you do to accelerate the object would be recovered when you decelerate it. But in reality, you have to apply a force just to overcome friction, and you have to do net work to slide the object in a circle back to its original position. Clearly, friction is not a conservative force.

If we were to look on an atomic scale at the interface between the object and the table as it slides, we don't see a "friction force" acting at all. Instead, we would notice the roughness of both the table and the object – sometimes an atomic-scale bump sticking out of the object would get caught behind an atomic-scale ridge on the table. As the object continued to move, the bonds to the hung-up atoms stretch or bend, increasing their potential energy (like springs or rubber bands); finally, the stuck atoms break free and vibrate violently, as the energy due to bond stretching is released. The increased vibrational kinetic energy of these few atoms is rapidly transferred through the bonds to all of the other atoms in the object, resulting in a small increase in the random, thermal energy of the object.⁸

If we reverse the direction we slide the object, the apparent friction force

⁸Essentially the same process happens in earthquakes as one plate of the earth's crust attempts to slide past another one along faults (such as the San Andreas fault or the many other faults below the LA basin). The sliding slabs of rock get hung up, and as the plates keep moving, huge strain energy is built up. Eventually, the plates break free, converting the pent-up strain energy (potential energy) into the kinetic energy of ground motion, which we experience as an earthquake. Sliding friction is a microscopic version of an earthquake!

reverses direction too, always opposing the direction of motion. This means that the friction force depends on the velocity of the object. For sliding friction, the dependence is usually only on the direction of the velocity vector (not its magnitude). But viscous drag in a fluid (also a type of friction) depends on the magnitude also, increasing with speed. This behavior is in sharp contrast to conservative forces, which only depend on position. For example, the gravitational force on an object of mass m is always mg directed in the same direction (toward the center of the earth) no matter what the velocity of the object is.

We see then that macroscopic forces which are non-conservative (friction) are actually “effective” forces which result from very complex atomic-level motion. Frictional forces always result in an irreversible conversion of macroscopic kinetic energy (the motion of the object) to disorganized, random thermal energy, and always oppose the direction of motion, so $\mathbf{F}_{nc} \cdot d\mathbf{x}$ is always negative.

2.8 The First Law of Thermodynamics

We now wish to do energy accounting for arbitrary macroscopic material systems. We’re already part way there – in Section 2.4 we developed an energy balance equation for macroscopic matter valid *if* the bonds between atoms were rigid. Unfortunately, this is not really the case. Bonds in solids can stretch and bend like springs, so the atoms are continually vibrating. This means that a solid will have kinetic energy associated with this motion, and potential energy due to stretching bonds. In liquids and gases, molecules can move and rotate, as well as vibrate.

In this section, we extend our previous analysis to account for these effects, and develop a purely macroscopic statement of energy accounting, which is the celebrated First Law of Thermodynamics.

2.8.1 The Internal Energy

Consider a macroscopic sample of matter (solid, liquid, or gaseous) at rest. Although no motion is apparent, on a microscopic level the atoms composing the sample are in continual, random motion. The reason we don’t perceive this motion, of course, is that all macroscopic measurements we can do average over a *huge* number of atoms. Since the atomic motion is essentially random, there are just as many atoms travelling to the right with a given speed as to the left. Even though individual atomic speeds may be hundreds of meters per second, the atomic velocities tend to cancel one another when we sum over a large number of atoms.

But the kinetic energies due to the atomic motion don’t cancel, since the

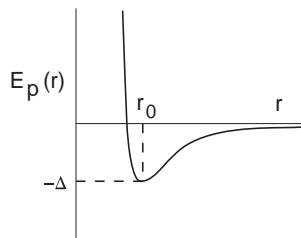


Figure 2.6: Potential energy of a chemical bond as a function of bond length r . The unstretched length is r_0 .

kinetic energies are all positive, scalar numbers. Even a sample of matter at rest (no center-of-mass motion) has microscopic kinetic energy, which we will call the internal kinetic energy:

$$E_{k,int} = \sum_j m_j v_j^2 / 2, \quad (2.65)$$

where the sum is over all atoms in the sample.

The sample has microscopic potential energy too. As the atoms move, they stretch or compress the bonds holding them together. The bonds may be modeled as springs, although ones with a spring constant which depends on bond length. The potential energy of these “springs” as a function of length typically looks something like the curve in Fig. 2.6. If the bond is compressed so that it is shorter than r_0 , the potential energy rises rapidly. If it is stretched, the potential energy rises, too. The bond can be broken ($r \rightarrow \infty$) if work Δ is done to pull the atoms apart.

Other types of interactions between atoms can be modeled in a similar way. Molecules in a gas hardly feel any force from other molecules far away, but when two molecules approach closely (collide) the potential energy rises rapidly, causing them to repel one another and move apart again. Similarly, the interaction of two atoms which are charged may be described by a repulsive or attractive electrostatic potential energy which depends on their separation. If the atoms or molecules have a magnetic moment (e.g. iron or nickel atoms, oxygen molecules), then their interaction is described by a potential energy function which depends on their separation and the relative alignment of their magnetic moment vectors. In fact, *every* atomic-level interaction between atoms can be described in terms of some potential energy function. We know this is possible, since we know atomic-level forces are conservative.

At any instant in time, the sample has a microscopic, internal potential energy, which is the sum of all of the potential energy contributions describing

the interactions between the atoms or molecules:

$$E_{p,int} = \sum_k E_{p,k}. \quad (2.66)$$

The index k runs over all pairs⁹ of atoms which exert forces on one another in any way.

Of course, since the particles are constantly moving and interacting, both the atomic positions and atomic velocities keep changing. Hence, both $E_{k,int}$ and $E_{p,int}$ are constantly changing, as energy shifts from kinetic to potential and back again (for example, as two atoms in a gas collide and rebound).

In doing energy accounting for this sample, we will first of all choose the system to consist of the sample itself (a closed system). We will furthermore choose to treat the conservative interactions between atoms within it by including $E_{p,int}$ in the system energy, rather than accounting explicitly for the work done by these forces. Therefore, the only work terms which will appear are those due to external forces.

If no external forces act on the atoms of the sample (if it is completely isolated from the environment), then energy accounting leads to the conclusion that the sum of $E_{k,int}$ and $E_{p,int}$ must be constant:

$$\Delta(E_{k,int} + E_{p,int}) = 0. \quad (2.67)$$

We define the *internal energy* U by

$$U = E_{k,int} + E_{p,int}, \quad (2.68)$$

For a stationary sample which is isolated from the environment $\Delta U = 0$.

The internal energy includes all of the kinetic energy associated with the atomic-level, random motion of the atoms of the system, and all of the potential energy associated with all possible interactions between the atoms. Since the potential energy associated with chemical bonds is included in $E_{p,int}$, chemical energy is part of the internal energy. Chemical energy is essentially the energy required to break chemical bonds (Δ in Fig. 2.6). Since Δ differs for every different type of bond, if a chemical reaction occurs which breaks bonds of one type and forms bonds of another type, $E_{p,int}$ may go up or down. If the system is isolated, U must be constant, and therefore $E_{k,int}$ must change oppositely to

⁹The potential energy of some interactions – for example, bending of chemical bonds – may depend on the positions of three or more atoms. This doesn't change anything – we simply add these terms too to $E_{p,int}$.

the change in $E_{p,int}$. The change in $E_{k,int}$ would be experienced as a change in temperature.¹⁰

Example 2.4 At sufficiently low density and a temperature of 300 K, the internal energy of gaseous H_2 is -2441 kJ/kmol ¹¹ and the internal energy of gaseous I_2 is $59,993 \text{ kJ/kmol}$. (We will show later that in the limit of low density the internal energy per mole of a gas is a function only of temperature – assume this limit applies here.) The internal energy of gaseous hydrogen iodide HI is given by the formula

$$U_{HI} = 17,655 + 21.22T \text{ kJ/kmol} \quad (2.69)$$

which is valid for $300 < T < 600 \text{ K}$.

If one kmol of H_2 is reacted with one kmol of I_2 to form two kmol of HI in a closed, constant-volume container with no energy transfer to the environment, what is the final temperature if the initial temperature is 300 K?

Solution: The internal energy of the initial mixture of H_2 and I_2 at 300 K is

$$U = (1 \text{ kmol})(-2441 \text{ kJ/kmol}) + (1 \text{ kmol})(59,993 \text{ kJ/kmol}) = 57,552 \text{ kJ}. \quad (2.70)$$

Since the system is isolated (no energy transfer to the environment), U does not change during the reaction. The final state consists of 2 kmol of HI, so the final internal energy per kmol of HI is $28,776 \text{ kJ/kmol}$. From Eq. (2.69), the final temperature is 524 K.

Note that the internal energy of H_2 is negative at 300 K. This is not a problem, since only differences in internal energy matter. It simply reflects a particular choice for the arbitrary constant C in the internal potential energy for H_2 .

Nuclear or even relativistic mass energy ($E = mc^2$) could be included in U if we like. A nuclear physicist would certainly want to do this. But since only changes in energy have physical significance, we can disregard these forms of energy if we don't plan to consider processes in which they change.

The internal energy is defined in a reference frame in which the sample is at rest. If this frame is moving and/or rotating with respect to the lab frame, then the macroscopic kinetic energy associated with these motions must be added to

¹⁰Temperature will be formally introduced in the next chapter. For now, think of it as a measure of the internal kinetic energy per atom. This would be exactly true if atomic motions were really described by classical mechanics, but when quantum effects are important (and the usually are) it is only approximately true.

¹¹One kmol is 6.023×10^{26} molecules. The mass of 1 kmol in kg equals the molecular weight of the molecule.

U to determine the total sample energy E in the lab frame:

$$E = U + \frac{1}{2}Mv_{cm}^2 + \frac{1}{2}I\omega^2, \quad (2.71)$$

where v_{cm} is the center-of-mass speed and ω is the rotation rate (assumed to be about a principal axis which has moment of inertia I).¹²

It is important to note that $E_{p,int}$ does not include any potential energy arising from interactions of atoms in the sample with gravitational, electric, or magnetic fields produced by *external* sources. If we choose to include these macroscopic potential energy terms in the sample energy, we have to add them explicitly. If the sample is near the surface of the earth and has charge q , the total energy including potential energy terms would be

$$E = U + \frac{1}{2}Mv_{cm}^2 + \frac{1}{2}I\omega^2 + Mgy + q\mathcal{E}, \quad (2.72)$$

where \mathcal{E} is the value of the electrostatic potential (volts).

With macroscopic kinetic and potential energy modes, the energy balance for an isolated sample is

$$\Delta E = 0, \quad (2.73)$$

not $\Delta U = 0$. For example, a rubber ball dropped onto a rigid table will eventually come to rest, even if there is no energy loss to the environment. The gravitational potential energy Mgy is converted into an increase in U , which would be experienced as an increase in temperature. E , however, remains constant.

2.8.2 Atomic Level Energy Transfer: Microscopic Work

No sample of matter can really be completely isolated from the environment. Usually, it is in contact with other matter (e.g., a container for a gas or liquid; a table a solid rests on). Even if it were floating in interstellar space it would still exchange energy with the environment through radiation.

We now need to consider how to do energy accounting for a macroscopic sample allowing for external work done by forces from the environment. Consider a sample of gas in a container such as shown in Fig. 2.7 which has one movable wall (a piston). We will take the system to be the gas, and container to be part of the environment.

On an atomic level, both the gas and container consist of atoms which are in constant motion. The atoms of the gas are moving randomly in all directions,

¹²We assume here that all parts of the sample are moving or rotating macroscopically together. If not, then the macroscopic kinetic energy must be determined as $(1/2) \int \rho(\mathbf{x})v(\mathbf{x})^2 dV$, where the integration is over the sample and ρ is the mass density.

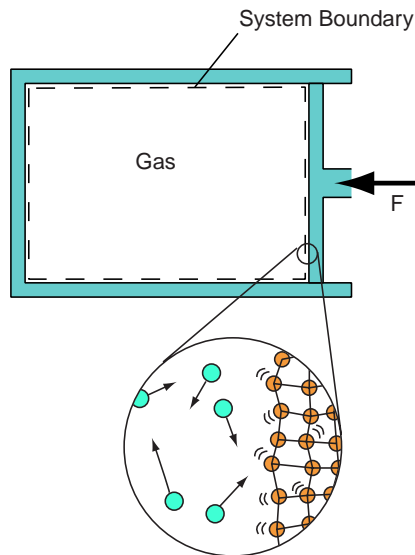


Figure 2.7: A gas in a container, as seen from the macroscopic and microscopic points of view.

colliding with one another and occasionally with the container walls. The atoms in the container are vibrating chaotically about their equilibrium positions as they are buffeted by the neighboring atoms they are bonded to, or (at the surface) by gas atoms.

When a gas atom collides with a wall atom, the gas atom may rebound with either more or less kinetic energy than it had before the collision. If the wall atom happens to be moving rapidly toward it (due to vibration) when they hit, the gas atom may receive a large impulse and rebound with more kinetic energy. In this case, the wall atom does microscopic work on the gas atom: positive microscopic work is done by the environment on the system.

On the other hand, the wall atom may happen to be moving away when the gas atom hits it, or it may rebound significantly due to the impact. In this case, the gas atom will rebound with *less* kinetic energy than it had before — therefore, the gas atom does microscopic work on the wall atom: negative microscopic work is done by the environment on the system.

We see that collisions between the gas atoms and the walls can do microscopic work even if macroscopically the walls appear stationary. If we let time dt elapse,

then the energy balance on the gas is

$$dW_{micro} = dU, \quad (2.74)$$

where dW_{micro} is the total work done on the gas by wall collisions during time dt .

2.8.3 Energy Transfer as Heat

Suppose the piston is held fixed, but the container starts out “hotter” than the gas, meaning that the container atoms have more kinetic energy per atom than do the gas atoms.¹³ Then over time the gas atoms *on average* will pick up kinetic energy from collisions with wall, and wall atoms will lose kinetic energy: dW_{micro} will be positive, U_{gas} will tend to go up, and $U_{container}$ will tend to go down. Of course, if the gas started out hotter, then dW_{micro} would be negative, and the changes in internal energy would be reversed.

Eventually, when their kinetic energies per atom are comparable,¹⁴ the number of collisions per unit time which impart extra energy to the gas atoms will just balance the number per unit time which remove energy from the gas atoms, and U_{gas} and U_{wall} will stop changing *on average*. There would still be very rapid statistical fluctuations about these average values, but for a reasonable sized sample these fluctuations are not observable, since it can be shown from statistics that random fluctuations like this have a relative magnitude proportional to $1/\sqrt{N}$. For example, if $N = 10^{20}$, then $\delta U/U \sim 10^{-10}$: the internal energy is constant to one part in 10^{10} in this case.

The process we have just described is energy transfer between the wall (part of the environment) and the gas (the system) due to microscopic work. However, *macroscopically* it doesn’t appear that any work is being done, since the piston isn’t moving, and we can’t see the microscopic deflections due to atomic motion. Therefore, there is no *observable, macroscopic* $\mathbf{F} \cdot d\mathbf{x}$, and no macroscopic work.

We call this process of energy transfer by microscopic work without observable macroscopic work *energy transfer as heat*, or *heat transfer* for short. The amount of energy transferred in this way is denoted by the symbol Q . For an infinitesimal amount, we use the symbol dQ . As for work, the bar in dQ reminds us that it is not the differential of any function, it only means “a little bit of heat.” (Or the other way to say it is that dQ , like dW , is an inexact differential.)

¹³Of course, “hotness” is really related to temperature, which we’ll introduce in the next chapter.

¹⁴More precisely, when their temperatures are equal.

The energy balance for this process is then

$$dQ = dU. \quad (2.75)$$

2.8.4 Energy Transfer as Macroscopic Work

Each collision of a gas atom with a wall delivers an impulse to the wall. At typical gas densities, the number of collisions per unit area of wall per unit time is very large. For example, objects sitting in room temperature ambient air experience roughly 10^{24} collisions per cm^2 per second. Macroscopically, it is not possible to detect the individual impulses from so many frequent collisions. Instead, a macroscopic force on the wall is felt, which is proportional to wall area:

$$F_{wall} = PA. \quad (2.76)$$

The proportionality constant P is the gas *pressure*.

Suppose the piston is now moved slowly toward the gas a distance dx . The macroscopic work required to do this is

$$dW_{macro} = \mathbf{F} \cdot d\mathbf{x} = (PA)dx. \quad (2.77)$$

The gas atoms which collide with the moving piston have their kinetic energy increased on average slightly more than if the piston had been stationary; therefore, U_{gas} increases. If $dQ = 0$, then the energy balance is

$$dU = dW_{macro} = PA dx. \quad (2.78)$$

Of course, there may also be microscopic work occurring which is not visible macroscopically (heat transfer). To account for this, we must write the energy balance as

$$dU = dQ + dW_{macro}. \quad (2.79)$$

For a more general system, macroscopic kinetic energy and potential energy may also be part of the system energy. If energy is transferred to such a system by macroscopic work and by heat transfer, the most general energy balance for a closed system is

$$\boxed{dE = dQ + dW.} \quad (2.80)$$

We have to stipulate that the system is closed, since if matter were to enter or leave the system, it would carry energy with it which is not accounted for

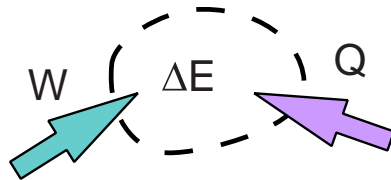


Figure 2.8: The First Law for a Closed System.

in Eq. (2.80). Note that we have removed the subscript “macro” on the work term. In thermodynamics generally, and from here on in this book, the term *work* means *macroscopic work*, unless otherwise stated.

Equation (2.80) is known as the *First Law of Thermodynamics*. The First Law simply states that the change in the total energy of a system equals the energy transfer to it as heat, plus the energy transfer to it as work. It is simply a statement of conservation of energy for a macroscopic system.

Note that there is no formula for dQ like $dW = \mathbf{F} \cdot d\mathbf{x}$. In practice, dQ is determined from equation 2.80 once dW and dE have been evaluated.

We can integrate Eq. (2.80) for some finite change from an initial state to a final one, yielding

$$\Delta E = Q + W \quad (2.81)$$

where

$$W = \int_i^f dW = \int_i^f \mathbf{F}_{macro} \cdot d\mathbf{x}_{macro} \quad (2.82)$$

and

$$Q = \int_i^f dQ. \quad (2.83)$$

The interpretation of Eq. (2.81) is as shown in Fig. 2.8. Both work and heat represent energy transfers across the system boundary; the energy E stored within the system (in whatever form) changes by the amount of energy transferred in.

Alternatively, we may divide Eq. (2.80) by the elapsed time dt to obtain

$$\frac{dE}{dt} = \dot{Q} + \dot{W} \quad (2.84)$$

where the ratio dQ/dt is the *heat transfer rate* \dot{Q} and the ratio dW/dt is the power input or *work rate* we've defined previously.

All three equations (2.80), (2.81), and (2.84) are different forms of First Law of Thermodynamics for a closed system. In solving problems involving the First Law, you should carefully consider which form is most appropriate to use. If the process occurs during an infinitesimal time dt , use Eq. (2.80). If you are given initial and final states of the system, often Eq. (2.81) is the best choice. If you are given a heat transfer or work *rate*, then probably Eq. (2.84) would be easiest to use.

For many processes, both Q and W will be significant, and must be included to correctly calculate the change in the system energy E from the First Law. But in some cases, either Q or W may be very much smaller than the other. In analyzing such processes, it is often acceptable to only include the dominant energy transfer mechanism, although this all depends on how accurate an answer is required for ΔE .

For example, if a solid is heated, it usually expands a little bit. But in many cases the work done in the expansion against atmospheric pressure is so small that $W \ll Q$. In this case, it might be OK to neglect W in calculating ΔE due to heating.

The opposite case would occur, for example, if a rubber band were rapidly stretched. Since heat transfer takes some time to occur, if the stretching is rapid enough it might be OK to neglect Q in calculating the increase in internal energy of the rubber band due to the work done to stretch it.¹⁵ Processes for which $Q = 0$ are called *adiabatic*.

2.9 Reversible and Irreversible Work

An important concept in thermodynamics is the idea of *reversible work*. Work $dW = \mathbf{F}(\mathbf{x}, \mathbf{v}) \cdot d\mathbf{x}$ is reversible if and only if the work done in moving $d\mathbf{x}$ is exactly recovered if the motion is reversed. That is,

$$\text{Forward: } dW_{\text{forward}} = \mathbf{F}(\mathbf{x}, \mathbf{v}) \cdot d\mathbf{x} \quad (2.85)$$

$$\text{Reverse: } dW_{\text{reverse}} = -dW_{\text{forward}} = \mathbf{F}(\mathbf{x}, -\mathbf{v}) \cdot (-d\mathbf{x}). \quad (2.86)$$

Note that the velocity changes sign when the direction of motion is reversed. This condition is satisfied if

$$\mathbf{F}(\mathbf{x}, \mathbf{v}) \cdot \mathbf{v} = \mathbf{F}(\mathbf{x}, -\mathbf{v}) \cdot \mathbf{v}. \quad (2.87)$$

¹⁵You can verify for yourself that rubber bands heat up when stretched by rapidly stretching one and holding it to your lip.

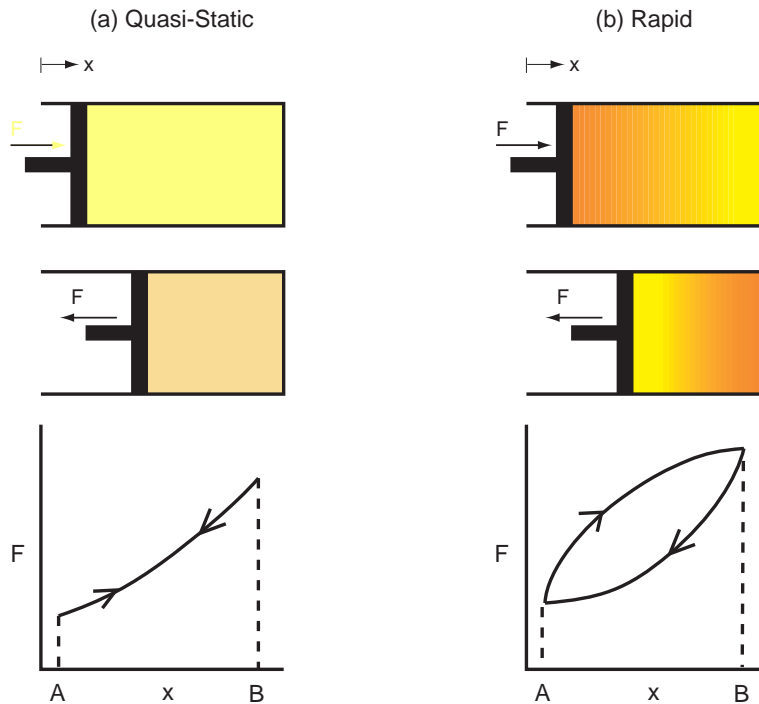


Figure 2.9: (a) Quasi-static and (b) rapid compression and expansion of a gas.

Therefore, the condition is that the force component in the direction of \mathbf{v} be the same for forward and reverse motion. A force which depends only on position \mathbf{x} will satisfy this, and therefore work done by any $\mathbf{F}(\mathbf{x})$ is reversible.

Friction and drag forces always depend on velocity, and act opposite to the direction of motion, changing sign when the direction of motion is reversed. Therefore, work done on a system by a friction or drag force is *always negative* (i.e., the system must do work against the friction force). Work done by or against such forces is never reversible – work must always be done to overcome friction, and you can never recover it.

Consider compressing a gas in a cylinder by pushing in a piston, as shown in Fig. 2.9. As discussed above, the gas exerts a force on the piston due to collisions of gas atoms with the piston surface. To hold the piston stationary, a force $F = PA$ must be applied. We will assume the piston is lubricated, and is well-insulated so the compression process is adiabatic.

The piston can be moved very slowly by applying a force just slightly greater than PA . If the piston velocity is sufficiently small, then the work required to overcome viscous drag in the lubricant will be negligible (example 2.2). Also, if the piston speed is slow enough, the gas molecules which collide with the piston have plenty of time to move away from it and distribute their excess energy with other molecules through collisions before the piston has moved any significant distance. In this case, the state of the gas is the same as would occur if the piston were stationary at the instantaneous value of x : The gas molecules are uniformly distributed in the cylinder, and the force on the piston is the same as if the piston were not moving – it is PA .

In this limit of zero piston speed, the force on the piston approaches PA , no matter whether the piston is moving in or out. In this limit, the compression or expansion process is called *quasi-static*, since the force on the piston is the same as if the piston were static. Therefore, the work done during quasi-static compression of a gas is reversible work.

If the piston velocity is high, two things happen which make the process irreversible. First, the work to overcome viscous drag in the lubricant may no longer be negligible. Second, if the piston velocity is comparable to the average molecular speed in the gas, then the piston will tend to sweep up molecules near it, forming a high-density region just in front of it (similar to a snowplow). Since the rate at which molecules collide with the piston is proportional to their number per unit volume, the piston will experience more collisions at a given x location than if it were moving slowly. Therefore, the applied force F to move the piston must be *greater* than the quasi-static force, and thus the work to compress the gas is greater than in the quasi-static limit. A typical plot of $F(x)$ for rapid compression is shown in Fig. 2.9(b).

If this process is now reversed and the gas is rapidly expanded, we still have to do work to overcome viscous drag in the lubricant (not only do we not get back the work done to overcome drag during compression, we have to do *still more work* to overcome it during expansion). Also, there is now a *low density* gas region near the piston, since the piston is moving away so fast the molecules lag behind. So the gas pushes on the piston with *less* force than if the expansion were done very slowly. Therefore, *the work we get back in the expansion is less than we put in during compression*.

Since $W = \int \mathbf{F} \cdot d\mathbf{x}$, the work input W_{AB} to move the piston from A to B is $\int_A^B F_{comp}(x) dx$, where $F_{comp}(x)$ is the force applied along the compression part of the curve. This of course is simply the area under the $F_{comp}(x)$ curve. The work input to expand the gas from B to A is $W_{BA} = \int_B^A F_{exp}(x) dx =$

$$- \int_A^B F_{exp}(x) dx.$$

If we consider the entire process $A \rightarrow B \rightarrow A$, then the total work input $W = W_{AB} + W_{BA}$. In the quasi-static case $F_{comp}(x) = F_{exp}(x) = PA$, so $W = W_{AB} + (-W_{AB}) = 0$. No net work is required to return the piston to its starting point. From the first law for this process (remember $Q = 0$)

$$W = \Delta U = 0. \quad (2.88)$$

Therefore, the gas internal energy returns to its starting value after quasi-static compression followed by quasi-static expansion.

In the non-quasi-static case, $F_{comp}(x) > F_{exp}(x)$. Therefore, $W > 0$: net work input must be done if the piston is rapidly moved from A to B and then back to A. From the First Law then, $\Delta U > 0$. The gas ends up with more internal energy (hotter) at the end of the process than at the beginning.

2.10 Some Reversible Work Modes

There are several different ways of doing reversible (quasi-static) work on matter. A few of these are described here.

2.10.1 Compression

We saw in the last section that if a gas is slowly compressed, the work required to move the piston dx is $dW = (PA)dx$. The same analysis would apply if the gas in the cylinder were replaced by any compressible substance, so this result is quite general. The volume change of the substance is $dV = -Adx$, so we may write this as

$$\boxed{dW_{qs} = -PdV.} \quad (2.89)$$

We add the subscript “qs” since this only applies if the compression is done quasi-statically. Note this expression is for the work done on the substance (input to the system); For compression, $dV < 0$ and $dW_{qs} > 0$, for expansion $dV > 0$ and $dW_{qs} < 0$.

2.10.2 Stretching a Liquid Surface

If a liquid film is suspended on a wire frame, as shown in Fig. 2.10, a force is exerted on the wire that is proportional to its wetted length L that results from a tensile force¹⁶ per unit length in the surface of the liquid. This is known as the

¹⁶A tensile force is the opposite of a compression force – it pulls, rather than pushes.

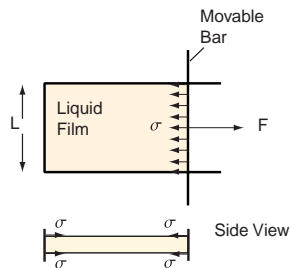


Figure 2.10: Surface tension in a liquid.

surface tension σ , and has units of N/m. For example, for a water/air interface at 25 °C, $\sigma = 0.072$ N/m.

The physical origin of the surface tension is that molecules in a liquid exert attractive forces on one another, which hold the liquid together. These forces are much weaker than covalent chemical bonds, but nevertheless have a dependence on distance similar to that shown in Fig. 2.6. A molecule will have lower potential energy in the bulk, where it is surrounded by molecules on all sides, than at the surface, where it feels the attractive force only on one side. Therefore, surface molecules will try to move into the bulk, until as many have crowded into the bulk as possible and there is a shortage of surface molecules left to cover the area. The remaining surface molecules will be spaced slightly further apart than ideal ($r > r_0$), and therefore they will pull on their neighboring surface molecules, resulting in the surface tension.

Since the film has two surfaces, the force required to hold the movable wire stationary is

$$F = 2\sigma L. \quad (2.90)$$

If the wire is now quasi-statically pulled, so that F is only infinitesimally greater than $2\sigma L$, the work done to move dx is

$$dW_{qs} = 2\sigma L dx. \quad (2.91)$$

During this process the total surface area of the film has increased by $2L dx$. Therefore, we may write

$$\boxed{dW_{qs} = \sigma dA.} \quad (2.92)$$

This expression for the work required to quasi-statically increase the surface area of a liquid is valid for arbitrary geometry.

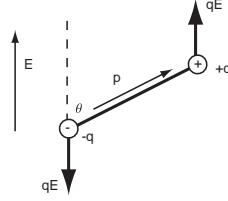


Figure 2.11: Forces exerted by an electric field on a polar diatomic molecule.

2.10.3 Electric Polarization

Many materials are *polar*, which means that although they are electrically neutral, they are composed of positively and negatively charged atoms. Any ionic crystal (NaCl) is polar, as is water (the hydrogen atoms have positive charge, and the oxygen atom has negative charge). If an electric field is applied to a polar material, it is possible to do work on it.

Consider the situation shown in Fig. 2.11. A polar diatomic gas molecule is oriented at a particular instant in time at an angle θ with respect to an applied electric field. (Due to collisions between the gas molecules, at any instant in time there is a distribution of orientations – they are not all lined up with the field, except at absolute zero.)

The force on a charge q in an electric field \mathbf{E} is given by

$$\mathbf{F} = q\mathbf{E}. \quad (2.93)$$

Therefore, the positive end of the molecule at position \mathbf{x}_+ feels a force $q\mathbf{E}$, and the negative end at \mathbf{x}_- feels a force $-q\mathbf{E}$. The molecule will turn and may be stretched by the forces due to the electric field acting on each end. (The center of mass motion is unaffected, since there is no net force.) If, due to the field, the atoms move by $d\mathbf{x}_+$ and $d\mathbf{x}_-$, respectively, then the work done on this one molecule is

$$dW_1 = (q\mathbf{E} \cdot d\mathbf{x}_+) + (-q\mathbf{E} \cdot d\mathbf{x}_-) \quad (2.94)$$

$$= q\mathbf{E} \cdot d(\mathbf{x}_+ - \mathbf{x}_-). \quad (2.95)$$

The *electric dipole moment* \mathbf{p} of the molecule is defined by

$$\mathbf{p} = q(\mathbf{x}_+ - \mathbf{x}_-). \quad (2.96)$$

The dipole moment is a vector which points along the direction from the negative charge to the positive charge. In terms of \mathbf{p} , the work done is

$$dW_1 = \mathbf{E} \cdot d\mathbf{p}. \quad (2.97)$$

This is the work done on one polar molecule; to determine the work done on the entire polar gas, we must sum over all N molecules in the gas:

$$dW = \sum_{n=1}^N \mathbf{E} \cdot d\mathbf{p}_n = \mathbf{E} \cdot d \left(\sum_{n=1}^N \mathbf{p}_n \right). \quad (2.98)$$

The quantity in parenthesis is the total dipole moment of the gas. This may be rewritten in terms of purely macroscopic quantities by defining the *polarization* \mathbf{P} to be the dipole moment per unit volume:

$$\mathbf{P} = \frac{1}{V} \left(\sum_{n=1}^N \mathbf{p}_n \right). \quad (2.99)$$

Then Eq. (2.98) becomes

$$\boxed{dW = \mathbf{E} \cdot d(V\mathbf{P})}. \quad (2.100)$$

Although we derived this equation for a polar gas, it is in fact completely general and applies to polar solids and liquids also.

For static polarization, P is some function of \mathbf{E} and temperature T . Except in extremely strong fields, the dependence on \mathbf{E} is linear, so that \mathbf{P} may be expressed as $\mathbf{P} = \epsilon_0 \chi_e(T) \mathbf{E}$. In Rationalized MKSA units (the most common electromagnetic unit system for engineering work), $\epsilon_0 = 8.90 \times 10^{-12} \text{ C}^2 \cdot \text{s}^2 / \text{kg} \cdot \text{m}^3$. The dimensionless material property $\chi_e(T)$ is the *electric susceptibility*.

The susceptibility for many materials may be expressed as

$$\chi_e(T) = A + B/T, \quad (2.101)$$

where A and B are constants. The A term describes polarization due to stretching of polar chemical bonds, a process which is not strongly temperature dependent. The B/T term describes orientation of polar molecules in a liquid or a gas, as we considered above.

It is not hard to see why this orientation process should be temperature-dependent. If $\mathbf{E} = 0$, then the dipoles are oriented randomly, so just as many point up as down. In this case, $\mathbf{P} = 0$. For non-zero \mathbf{E} , the field will try to align

\mathbf{p} for each molecule with \mathbf{E} , but collisions between molecules will upset this alignment, tending to randomize \mathbf{p} . The net effect of the competition between alignment by \mathbf{E} and randomization by collisions is that molecules point in all directions, but $\mathbf{p} \cdot \mathbf{E}$ is somewhat more likely to be positive than negative. This means that there is non-zero polarization, with \mathbf{P} directed along \mathbf{E} . Since higher \mathbf{E} increases the tendency to align, and higher T increases the tendency to randomize direction, \mathbf{P} typically increases with increasing \mathbf{E} , and decreases with increasing T .

Example 2.5

A *dielectric* material is one which may be polarized, but has no mobile free charges, so no electrical currents can flow through it. Dielectrics are often used to fill the space between the plates in capacitors. A particular dielectric liquid, which obeys Eq. (2.101) with $A = 0$, is quasi-statically polarized at constant temperature starting at $\mathbf{E} = 0$ and ending at $\mathbf{E} = \mathbf{E}_1$. For this material, the internal energy depends only on T . Determine the work and heat transfer during this process.

Solution:

$$W = \int dW = \int \mathbf{E} \cdot d(V\mathbf{P}). \quad (2.102)$$

For quasi-static polarization, the static relationship between \mathbf{P} , \mathbf{E} , and T holds, so

$$W_{qs} = \int_0^{E_1} Ed(V\epsilon_0 BE/T) = \frac{\epsilon_0 V B E_1^2}{2T}. \quad (2.103)$$

Since this process is carried out isothermally, and for this particular material $U = U(T)$, $\Delta U = 0$ for this process. The first law applied to this system is

$$\Delta U = 0 = Q + W_{qs}, \quad (2.104)$$

from which we conclude that $Q = -\epsilon_0 V B E_1^2 / 2T$. Therefore, heat must be removed ($Q < 0$) to polarize this material at constant temperature; if no heat were removed (adiabatic polarization), U would increase by W_{qs} , and the temperature would increase.

2.10.4 Magnetization

Some materials are magnetic – that is, they contain atoms which have magnetic dipole moments and behave just like atomic-scale magnets. Magnetic atoms are usually ones with unpaired electrons, such as iron, nickel, or rare-earth elements. Some molecules can have unpaired electrons also, for example O_2 .

An applied magnetic field can do work on magnetic materials. The analysis is very similar to that for electric polarization. If a single magnetic dipole with

dipole moment \mathbf{m} is placed in a uniform magnetic field \mathbf{H} , the field exerts a torque on the dipole given by

$$\tau = \mu_0 \mathbf{m} \times \mathbf{H}. \quad (2.105)$$

where μ_0 is a constant which depends on the unit system we use to measure \mathbf{m} and \mathbf{H} known as the *permeability of free space*. In Rationalized MKSA units $\mu_0 = 4\pi \times 10^{-7} \text{ kg} \cdot \text{m}/\text{C}^2$.

The work done by this torque is

$$dW_1 = \mu_0 \mathbf{H} \cdot d\mathbf{m}. \quad (2.106)$$

Summing over all magnetic dipoles in the material and defining the *magnetization* \mathbf{M} by

$$\mathbf{M} = \frac{1}{V} \sum_{n=1}^N \mathbf{m}_n, \quad (2.107)$$

we obtain an expression for the work required to change the magnetization of a magnetic material:

$$\boxed{dW = \mu_0 \mathbf{H} \cdot d(V\mathbf{M})} \quad (2.108)$$

Analogous to the discussion above for dielectrics, if a magnetic material is placed in a static magnetic field, it will develop some static magnetization $\mathbf{M}(\mathbf{H}, T)$, which results from the balance between the field trying to align the dipoles and random thermal motion (collisions, vibrations) upsetting perfect alignment. If the field is increased very slowly (quasi-statically), then this relation between \mathbf{M} , \mathbf{H} , and T will still hold, and we may write

$$dW_{qs} = \mu_0 \mathbf{H} \cdot d[V\mathbf{M}(\mathbf{H}, T)] \quad (2.109)$$

Example 2.6 A *Curie substance* is one for which the static magnetization is

$$\mathbf{M} = C \frac{\mathbf{H}}{T}, \quad (2.110)$$

where C is a material-specific constant. Most magnetic materials behave as Curie substances in the limit of high temperature and low field strength.

A Curie substance in a uniform magnetic field \mathbf{H}_0 is quasi-statically, isothermally magnetized by slowly increasing the field to $\mathbf{H}_1 > \mathbf{H}_0$. Calculate the work done on the substance, and the change in its internal energy. It may be shown

that for a Curie substance $U = U(T)$. Calculate the heat transfer which must occur in this process.

Solution: Since the process is quasi-static, the static relationship $\mathbf{M}(\mathbf{H}, T)$ holds at every step in the process. Therefore,

$$\begin{aligned} W &= \int_{H_0}^{H_1} \frac{\mu_0 C}{T} \mathbf{H} \cdot d\mathbf{H} \\ &= \frac{\mu_0 C}{2T} (H_1^2 - H_0^2). \end{aligned} \quad (2.111)$$

$$(2.112)$$

The First Law for this process is

$$\Delta U = Q + W. \quad (2.113)$$

Since the process is isothermal and we are given $U = U(T)$, $\Delta U = 0$. Therefore,

$$Q = -W = -(\mu_0 C / 2T)(H_1^2 - H_0^2). \quad (2.114)$$

Example 2.6 shows that heat must be given off to the environment in order to quasi-statically (reversibly) magnetize a Curie substance at constant temperature. If the process had been done adiabatically instead ($Q = 0$), the internal energy and temperature of the substance would have increased. The reason for this is that the microscopic torque exerted on the individual dipoles by the field imparts to them some rotational kinetic energy, which is then transferred to the rest of the substance by collisions. The reverse process of quasi-static isothermal demagnetization require heat *input* to maintain the sample temperature; if no heat is supplied (adiabatic demagnetization), the sample temperature drops.

These processes may be combined to produce useful devices, such as magnetic engines or magnetic refrigerators. Magnetic refrigerators are used in practice to achieve very low temperatures ($T < 1$ K), where conventional refrigerators cannot function.

2.10.5 Generalized Forces and Displacements

The expressions for quasi-static work are always of the general form $\mathcal{F}dX$ or $\mathcal{F} \cdot d\mathbf{X}$: $-PdV$, σdA , $\mathbf{E} \cdot d\mathbf{P}$, $\mu_0 \mathbf{H} \cdot d\mathbf{M}$, etc. We can think of these as generalizations of $\mathbf{F} \cdot d\mathbf{x}$. We call the \mathcal{F} terms ($-P$, σ , \mathbf{E} , \mathbf{H}) *generalized forces* and the X terms (V , A , \mathbf{P} , \mathbf{M}) *generalized displacements*.

The work done by any of these for a finite change is

$$W = \int_{path} \mathcal{F} \cdot d\mathbf{X}. \quad (2.115)$$

This is still a path integral, but note it is not an integral in physical coordinate space, but in the space defined by \mathbf{X} . For example, polarization work would involve integrating $\mathbf{E} \cdot d\mathbf{P}$ along some $\mathbf{P}(t)$ trajectory.

2.11 Some Irreversible Work Modes

If any of the processes discussed in the last section are done too rapidly, the work done will not be reversible. For example, if the magnetic field is increased too rapidly, the induced magnetization will lag behind the static $\mathbf{M}(\mathbf{H}, T)$. This will result in $\mu_0 \mathbf{H} \cdot d\mathbf{M}$ being *greater* than the quasi-static value for a given $d\mathbf{M}$. Therefore, more magnetization work must be done to effect a given change in magnetization; less work is recovered during demagnetization.

Some other ways of doing work are *inherently* irreversible – if the direction of the motion is reversed, the force changes sign, so you can't recover *any* of the work put in. As we've already discussed, work done to overcome friction or viscosity is like this.

2.11.1 Stirring a Viscous Fluid

An example of purely irreversible work is stirring a viscous fluid.¹⁷ Work must be done to turn the stirrer, no matter which direction it is turned. The fluid will have some macroscopic kinetic energy for a while due to stirring, but eventually it will come to rest, with the energy transfer as work to the system due to stirring appearing finally as an increase in internal energy U .

In fact, the state of the fluid after it is stirred and has come to rest again is no different than if the same amount of energy had been added to it as heat. *Fully irreversible work is equivalent to heat addition.*

2.11.2 Electrical Current Flow Through A Resistor

Another common type of fully irreversible work is electrical current flow through a resistor. As electrons move through a resistor with an electrical potential $\Delta\mathcal{E} > \nu$ across it, they lose electrostatic potential energy in the amount $|e\Delta\mathcal{E}|$ by doing this amount of irreversible work. The work is done by colliding with atomic scattering centers within the resistor, which transfers energy to them, increasing the internal energy of the resistor.

If the number of electrons flowing through the resistor per second is \dot{N}_e , then the rate at which irreversible work is done is

$$\dot{W}_e = \dot{N}_e |e\Delta\mathcal{E}| = \mathcal{I} \cdot \mathcal{E}. \quad (2.116)$$

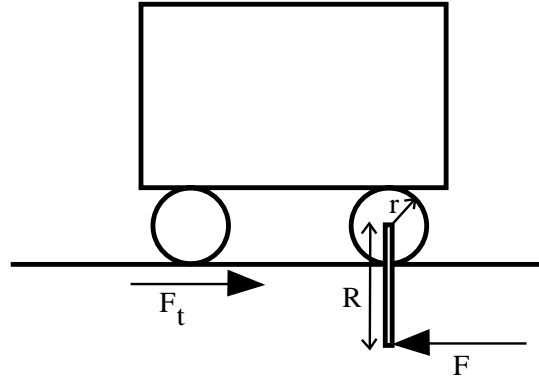
¹⁷All fluids, even water and gases, have some viscosity.

where $I = e\dot{N}_e$ is the electrical current. The current through the resistor is proportional to the voltage across it: $I = \Delta\mathcal{E}/R$, where R is the resistance. Therefore, Eq. (2.116) can be also written $\dot{W}_e = I^2 R$, or $\dot{W}_e = (\Delta\mathcal{E})^2/R$.

As was true for stirring a fluid, passing a current through a resistor is fully irreversible, since changing the sign of the voltage also reverses the direction of current flow. The irreversible electrical work done on the resistor is equivalent thermodynamically to heat addition to the resistor.

Problems

- 2.1** A cart is sitting near the edge of a flat, horizontal table. The wheels of the cart have radius r , and a lever of length $R > r$ is attached to one wheel. When the lever is straight down, it is pushed by a force \vec{F} to the left, causing the wheels to turn a small amount $d\theta$ and the cart to move to the right. A traction force is also present on the wheels, which keeps the cart from sliding when the force \vec{F} is applied. Determine the work done by force \vec{F} and by the traction force. (Hint: calculate displacements with respect to the table carefully, remembering the cart moves. Make sure your answer is sensible when $R = r$.)



- 2.2** Which of the following forces are conservative? The forces are all two-dimensional – the component in the z direction is zero. Here \hat{i} is a unit vector in the x direction, and \hat{j} is a unit vector in the y direction.

$$\vec{F}_1 = \hat{i}y - \hat{j}x \quad (2.117)$$

$$\vec{F}_2 = \frac{\hat{i}x + \hat{j}y}{x^2 + y^2} \quad (2.118)$$

$$\vec{F}_3 = \hat{i}2xy + \hat{j}x^2 \quad (2.119)$$

$$\vec{F}_4 = \hat{i}2x + \hat{j}x^2 \quad (2.120)$$

$$\vec{F}_5 = \mathbf{i}a \sin(ax) \sin(by) - \mathbf{j}b \cos(ax) \cos(by) \quad (2.121)$$

$$\vec{F}_6 = \mathbf{i} \sin(ax) \cos(by) + \mathbf{j} \cos(ax) \sin(by) \quad (2.122)$$

2.3 Show that any force with the properties listed below which acts between two particles must be conservative.

1. The force is directed along the line connecting the particles.
2. The force depends only on the distance r between the particles.

2.4 A ball of mass m is dropped at height $y = H$ onto a plate which at the moment the ball hits is at $y = 0$ and is travelling upward with constant speed v_p . Assuming the ball rebounds elastically and assuming air resistance is negligible, how high does the ball bounce? How much work does the plate do on the ball?

2.5 You are a scientist on your way to the new research station on Mars in the year 2020. Your spacecraft has just undocked from the space station in earth orbit, and is beginning to accelerate to cruising speed for the trip to Mars. The ion engines are turned on at $t = 0$, and the engine thrust begins to increase, producing for a while a spacecraft acceleration linear in time: $a = Ct$. Within the spacecraft, the effects of the acceleration are identical to those of a linearly-increasing gravitational field. Free-floating objects at $t = 0$ begin “falling” toward the rear wall. You notice a ball which at $t = 0$ is a distance H from the rear wall. In terms of the mass of the ball, C , and H , what is the kinetic energy (in the local reference frame within the spacecraft) of the ball when it strikes the wall? Is the concept of potential energy useful to solve this problem? Why or why not?

2.6 An ideal gas is defined to be one which satisfies $PV = N\hat{R}T$, where N is the number of kg-moles (kmol) of gas, and $\hat{R} = 8.3143$ kJ/kmol/K. Consider an ideal gas contained in a vertical cylinder, with a piston of mass M at the top.

1. Initially, no external force is applied to the piston, and it comes to some equilibrium height. If the atmospheric pressure is P_0 and the cylinder area is A , what is the gas pressure in the cylinder?
2. The gas is now heated quasi-statically at constant pressure until its volume is tripled. How much work does the gas do against the environment (which includes the massive piston) during this process?

3. Finally, a force is applied to the top of the piston, and the gas is quasi-statically, isothermally compressed back to its initial volume. How much work must be done?

2.7 A Curie substance undergoes a three-step quasi-static process:

1. The substance is isothermally magnetized from (H_0, M_0) to (H_1, M_1) , where $H_1 = 2H_0$
2. The field is reduced back to H_0 , holding M fixed
3. The magnetization is reduced to M_0 , holding H fixed

At the end of this process, the substance is back in its initial state.

Sketch this process on a plot of H vs. M . For each step, determine

1. Whether work is done on or by the substance, and how much
2. Whether heat is transferred to or from the substance, and how much
3. The temperature at the end of the step.

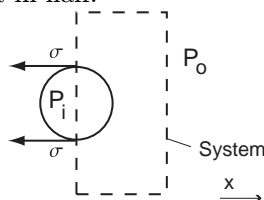
Explain how this process might be used as the basis for a magnetic refrigerator.

2.8 For a particular dielectric material, the susceptibility is given by $\chi_e = A + B/T$. With \vec{E} held constant, the material is cooled to half its initial temperature, causing the polarization to increase. How much work is done on the material by the electric field in this process?

2.9 Due to surface tension, the pressure inside a small water droplet will be greater than the air pressure outside. Show that

$$\Delta P = P_i - P_o = \frac{2\sigma}{r},$$

where σ is the surface tension and r is the droplet radius. Hint: consider a *force* balance in the x direction on the system defined by the dotted line, which cuts the droplet in half.



Small water droplets in a fog may be only a few microns in size. Evaluate numerically the pressure inside a water droplet for diameter $D = 1 \mu\text{m}$ and for $D = 10 \mu\text{m}$ if the air pressure is 1 atm. ($1 \text{ atm} = 1.01325 \times 10^5 \text{ N/m}^2$). Express your answer in atmospheres.

- 2.10** An ideal gas has the properties described in problem 2.6. In addition, for *monatomic* ideal gases (He, Ar, etc.) the internal energy is related to N and T by

$$U = \frac{3}{2}N\hat{R}T.$$

Consider a process in which a monatomic ideal gas is quasi-statically *adiabatically* compressed from a state with V_0, P_0, T_0 to a final state with properties V, P, T .

1. Derive an expression for how the temperature varies with volume during this process – that is, find the function $T(V)$. Do your analysis *systematically*: draw a sketch showing the system and any energy transfers from the environment, state any assumptions, invoke the first law, use necessary property data, etc.
2. Using this formula, evaluate the final temperature (K), final pressure (MPa), and work input (J) if argon at 300 K and 0.1 MPa is quasi-statically, adiabatically compressed to one-fourth of its initial volume. (Recall $1 \text{ Pa} = 1 \text{ N/m}^2$.)

CHAPTER 3

EQUILIBRIUM

3.1 Introduction

In this chapter, we introduce the important idea of *equilibrium*, and some new properties which tell us about equilibrium. We discuss how to measure two of these properties (pressure and temperature), and conclude with some observations about thermodynamic equilibrium states of matter.

3.2 Equilibrium

It's a universal observation that any real system, if left alone for enough time, reaches a state where the macroscopic properties stop changing. If a ball is dropped onto the floor, it will bounce for a while but eventually it will sit motionless. If a drop of dye is placed in a glass of water, the dye will slowly diffuse away until the dye concentration is the same everywhere in the glass, after which no further change is observed. If a fluid is stirred and then left alone, the motion will eventually cease, due to viscosity. We call the motionless, unchanging state a physical system approaches at long times the *equilibrium state*.

The word equilibrium is defined as “a state of balance between opposing forces or actions.”¹ When a physical system is in an equilibrium state, every part of it is in balance (equilibrium) with every other part. There is no net transfer of matter or energy between any two parts of the system. The balance is dynamic, since molecules may leave one region and enter another (in a liquid or gas), but the same number must be going the opposite direction, since otherwise there would be a net transfer of molecules.

Some idealized systems studied in introductory physics courses do not approach equilibrium states. Perfectly elastic balls keep bouncing forever, objects sliding across frictionless tables don't slow down, and a mass hanging from a frictionless spring will oscillate forever. But the motion of every *real* object is subject to non-zero friction or drag forces. No matter how small these forces are, they will eventually cause the object to stop moving and come into equilibrium.

¹Webster's New Collegiate Dictionary

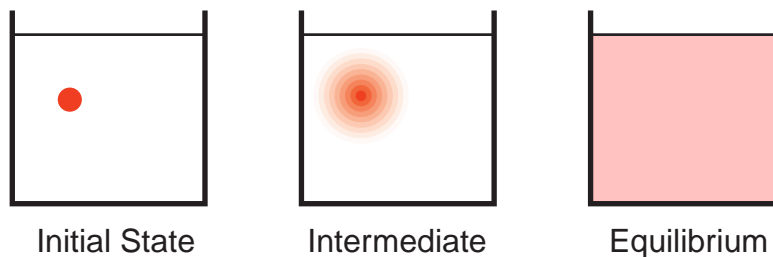


Figure 3.1: Diffusion of a drop of dye in water.

A general characteristic of equilibrium states is that they are much simpler to describe than non-equilibrium states. For example, to describe the state shortly after a drop of dye has been introduced into a glass of water requires specifying the concentration $C(\mathbf{x})$ throughout the glass. Once equilibrium has been achieved, it is only necessary to know how much dye was added to fully specify the state.

The dye/water example illustrates another characteristic of equilibrium states: they depend only on “intrinsic” factors (the amount of dye in this case), not one the time history of the non-equilibrium states which preceded it. In general, the characteristics of systems in equilibrium states depend on intrinsic factors such as the volume available to the system, the energy and mass it contains, and its chemical composition. They do not depend on the time-history of how the state was prepared. Any substance which has properties which depend on past processing history is not in an equilibrium state. For example, the properties of glass depend on the *rate* it is cooled from the molten state. Therefore, glass is not an equilibrium state.

Thermodynamics deals primarily with the properties of matter in the simple-to-describe equilibrium states. To describe matter in non-equilibrium states, we would have to supplement thermodynamics with results from the theories of fluid mechanics, diffusion, heat conduction, electromagnetics, or other areas of physics beyond the scope of thermodynamics. The nature of these states and how they evolve with time is the subject of courses in these other fields.

Although the restriction to equilibrium states seems drastic, it’s really not so bad. In many processes, a system begins and ends in an equilibrium state, even if it passes through non-equilibrium states during the process (for example, rapid expansion of a gas). Since the First Law only requires evaluation of $\Delta U = U_f - U_i$, in many cases we only need to evaluate properties (e.g. U) in equilibrium states.

We can define several different types of equilibrium, depending on what type

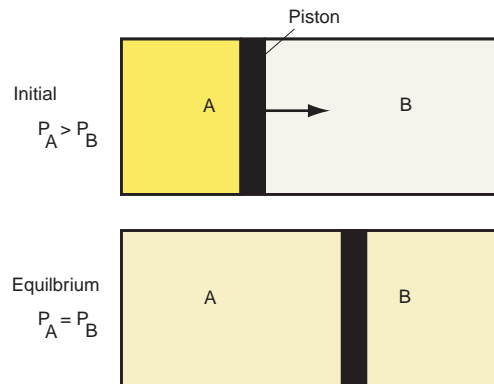


Figure 3.2: Mechanical equilibrium: the piston moves until $P_A = P_B$.

of “force or action” is in balance within a system. For every type of equilibrium, we would like to define some property which can be used to determine if two parts of a system are in equilibrium. In the following sections, various types of equilibrium, and the properties associated with each type, will be discussed.

3.2.1 Mechanical Equilibrium

Suppose a cylinder is divided in two parts (A and B) by a piston, which can move back and forth (Fig. 3.2). The piston is initially locked rigidly in place and some gas is added to each side of the cylinder. The piston is now unlocked, and oscillates back and forth a while. Eventually, friction or viscous forces damp out the piston motion, and it will stop, though perhaps not at its initial position. Once it has stopped, we say A and B are in *mechanical equilibrium* with each other.

If the piston is stationary, the force must be the same on both sides. Therefore, $P_A = P_B$ if A and B are in mechanical equilibrium. Pressure is seen to be the property which tells us if two adjacent parts of a system are in mechanical equilibrium with one another.

3.2.2 Thermal Equilibrium

Consider now two systems connected by a rigid wall which can conduct heat (Fig. 3.3). (We say they are in “thermal contact.”) When they are placed in thermal contact, energy may transfer through the wall as heat, but eventually no *net* energy transfer will occur. When this condition is reached, the two systems are in *thermal equilibrium*.

What property tells us if two systems are in thermal equilibrium? It’s clearly

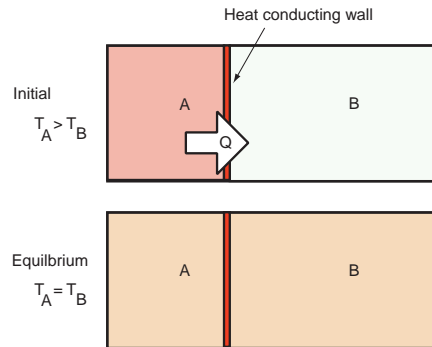


Figure 3.3: Thermal equilibrium: heat flows until $T_A = T_B$.

not the energy, since the two systems may have very different compositions and sizes, and so will have different energies even at thermal equilibrium.

We introduce a new property, specifically for the purpose of determining if systems are in thermal equilibrium. We make the following postulate.

Postulate: There exists a scalar property called *temperature* with the following characteristic: two systems placed in thermal contact will be in thermal equilibrium if and only if they have the same value of temperature.

The so-called zeroth law of thermodynamics is the postulate that if A is in thermal equilibrium with B, and B is with C, then A is with C. The zeroth law is equivalent to the postulate that the temperature property exists.

3.2.3 Diffusive Equilibrium

If a dye drop is added to a glass of water, we know that at equilibrium the dye will diffuse throughout the glass. This type of equilibrium is known as *diffusive equilibrium*. We would like to define some property of the water/dye system we could calculate for various regions in the water, to test if they are in diffusive equilibrium.

We might consider using the dye concentration as this property. But this is not a good choice in general, since there are some situations where the concentration is not equal everywhere, even though the system is in diffusive equilibrium. These situations arise if more than one phase is present or if external conservative forces act on the system. For example, in a gaseous mixture of heavy

and light species in a very tall column, the heavy ones will be preferentially concentrated near the bottom of the column, and the light ones near the top.

We would like a property that can handle even these cases. Not knowing what else to do, let us postulate that such a property exists:

Postulate: There exists a scalar property called *chemical potential* with the following characteristic: two parts of a system will be in diffusive equilibrium if and only if they have the same value of chemical potential.

The chemical potential is given the symbol μ . In situations where multiple chemical species are present, we need to test for diffusive equilibrium for each one, so each will require its own chemical potential. We will show later that μ for a particular species depends on its concentration² and on temperature and pressure. If the system has macroscopic potential energy (e.g. due to gravity), the chemical potential will also depend on the local value of the potential energy. We'll see in Chapter 6 how to calculate the chemical potential.

3.2.4 Phase Equilibrium

If a substance is placed in a sealed container, under some conditions it segregates into separate regions with distinct properties (for example, solid and liquid, or liquid and vapor). We call each region with homogeneous properties a *phase*. When the amount of each phase is constant, the substance is in *phase equilibrium*.

What property tells us when systems are in phase equilibrium? Again it's the chemical potential. Phase equilibrium is similar to diffusive equilibrium, since molecules are moving between the phases, but at equilibrium the rate they move is the same for both directions. For example, in a mixture of ice and liquid water, the rate at which H₂O molecules from the liquid stick onto the solid ice equals the rate molecules detach from the ice and enter the liquid.

The chemical potential of any given phase of a pure substance is some function of T and P . For example, for water there exist two functions $\mu_{ice}(T, P)$ and $\mu_{liquid}(T, P)$. The condition $\mu_{ice}(T_m, P) = \mu_{liquid}(T_m, P)$ determines the temperature where the two phases can co-exist (the melting temperature T_m) for a given P .

²In some cases, it depends on the concentration of other species too.

3.2.5 Chemical Equilibrium

If multiple chemical species are mixed together, they might chemically react. If they do, the concentration of each reactant will decrease, and the concentrations of the products of the reaction will build up. The products may react with one another, too, “reversing” the reaction and re-forming the reactants. Eventually, the concentrations of all species will settle at values such that the forward reaction rate just balances the reverse one. When this happens, the concentrations will no longer change, and we say the system is in *chemical equilibrium*. As we’ll see later, the chemical potential is again the relevant property to determine when a system is in chemical equilibrium, although the analysis is more complicated than for diffusive or phase equilibrium.

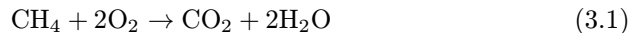
3.2.6 Thermodynamic Equilibrium

When all parts of a system are equilibrated in all ways (mechanical, thermal, diffusive, phase, chemical), we say the system is in *thermodynamic equilibrium*. The temperature, pressure, and chemical potential will be the same in all parts of a system in thermodynamic equilibrium, and the concentrations of every chemical species will be constant in time.

3.2.7 Restricted Equilibrium

Sometimes a system is kept from attaining one type of equilibrium, although it is equilibrated in other ways. For example, a rigid, heat-conducting partition dividing a system in two parts will allow thermal, but not mechanical, equilibrium to be attained. In other cases, there is no partition but one type of equilibrium is attained very slowly compared to other types. When a system is equilibrated in some ways but not others, we say it is in *restricted equilibrium*.

For example, a mixture of methane and air at room temperature is not in chemical equilibrium, although it may be in mechanical, thermal, and diffusive equilibrium. The reaction



is running *very slowly* in the forward direction, while the reverse reaction is hardly occurring due to the very low CO_2 and H_2O concentrations. Therefore, some of the methane is continually being oxidized, but the rate is so small at room temperature to be virtually impossible to detect. If the temperature is increased by adiabatically compressing the mixture or a spark is provided, the mixture may rapidly approach chemical equilibrium (it may burn or explode). Almost always the slow component is some type of chemical, phase, or nuclear

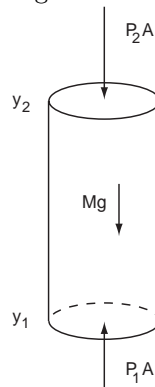
equilibrium. Examples of substances approaching true equilibrium very slowly include radioactive elements with very long half-lives, any type of glass, or cold-worked steel. In many cases, the slow process is so slow that it may be neglected entirely – the properties of cold-worked steel do not change *measurably* over any practical timescale.

3.3 Pressure and Its Measurement

3.3.1 Hydrostatics

The pressure at any point in a fluid at rest is the same in all directions. If a fluid is contained in a cylinder and you push on a piston to apply a force in one direction, the fluid transmits the force in all directions, so the pressure at the piston equals the pressure on the cylinder walls.

If no external body forces act on a fluid at rest, then the pressure is the same at every point. But in a fluid sitting in a gravitational field, the pressure will depend on height y , due to the weight of the fluid above y .



A force balance on a vertical column of fluid yields

$$P_1A = P_2A + Mg. \quad (3.2)$$

Since $M = \rho A(y_2 - y_1)$ (assuming constant density), we find

$$P_1 - P_2 = \rho g(y_2 - y_1). \quad (3.3)$$

Differentiating with respect to y_1 yields

$$\boxed{\frac{dP}{dy} = -\rho g} \quad (3.4)$$

which is the basic equation of *hydrostatics*, and describes how pressure varies with height in a static fluid. Since Eq. (3.4) is a differential equation, it applies even if the density ρ depends on height y , as would be the case if the fluid is a gas. In this case, Eq. (3.3) would be replaced by

$$P_2 - P_1 = - \int \rho(y)g \, dy. \quad (3.5)$$

Example 3.1 A dam of length L holds water to depth H . What is the net horizontal force on the dam if $L = 200$ m and $H = 30$ m?

Solution: Taking the density of water to be constant at $\rho = 10^3$ kg/m³, the pressure distribution on the dam is

$$P(y) = P_0 + \rho gy, \quad (3.6)$$

where P_0 is atmospheric pressure and y is the distance below the surface. The force due to the water is

$$F_w = L \int_0^H P(y)dy = L(P_0H + \rho gH^2/2). \quad (3.7)$$

Since the force due to atmospheric pressure on the other side is P_0LH , the net force is

$$F_{net} = \rho gLH^2/2 = 8.8 \times 10^8 \text{ N}. \quad (3.8)$$

We have neglected the variation in air pressure with height. Since the density of air is about 1.2 kg/m³, for $H = 30$ m, $\Delta P_{air} = 345$ N/m², which is much less than P_0 , which is about 10^5 N/m². If H were several km, then ΔP_{air} would have to be considered.

3.3.2 Manometers

A simple arrangement to measure the pressure of a gas is shown in Figure 3.4. Mercury fills a U-shaped tube of constant cross-sectional area, which is closed on the left side and can be connected to a system of unknown pressure on the right. The region on the left above the mercury column is evacuated.³ A device like this is known as a *manometer*.

At the height of the gas/mercury interface in the right-hand column, the pressure is the gas pressure P_{gas} . The pressure in the left-hand column at the same height must therefore also equal P_{gas} . From Eq. (3.3), the pressure at

³Actually it is filled with mercury vapor. But mercury has a very low vapor pressure, and exerts negligible force on the liquid.

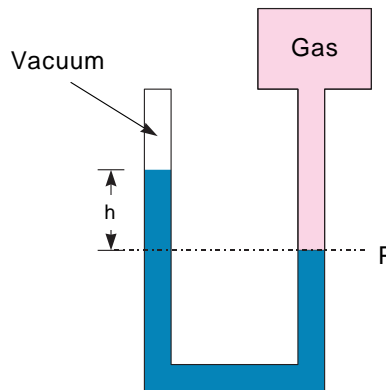


Figure 3.4: A manometer.

this location is also ρgh . Therefore, the unknown gas pressure is related to the measurable height h by

$$P_{gas} = \rho gh. \quad (3.9)$$

A different type of manometer can be used to measure the pressure difference between two systems. If instead of evacuating the region above the column on the left we attach it to a system with some pressure P_A , and attach the other side to a system with P_B , then

$$P_B - P_A = \rho gh. \quad (3.10)$$

The SI unit for pressure is the *Pascal*, which is simply the name given to 1 N/m^2 . A standard atmosphere (1 atm) is defined to be $1.01325 \times 10^5 \text{ Pa}$, or 0.101325 MPa . Other common units for pressure are the Torr ($1/760 \text{ atm}$ or 133.32 Pa), the bar (10^5 Pa), and lbf/in^2 or psi ($1 \text{ atm} = 14.7 \text{ psi}$).

Usually, pressure is measured with other, more convenient instruments which may be calibrated against mercury manometers. In some pressure gauges, gas at the unknown pressure enters a curved, flexible tube. The tube tends to straighten as the gas pressure increases, which causes a needle on a dial to move. Another type of gauge, known as a *capacitance manometer*, consists of an evacuated disk-shaped cavity, one side of which consists of a thin, flexible membrane. As the membrane deflects due to external pressure, the capacitance of the disk changes. Since capacitance can be determined very accurately by electrical measurements, capacitance manometers are designed to produce an electrical signal (usually a voltage) proportional to pressure.

Sometimes you will hear a pressure described as a “gauge pressure.” This is defined as the pressure difference between the actual pressure and the local

atmospheric pressure, and is what is directly measured by some types of pressure gauges.

3.4 Temperature and Its Measurement

3.4.1 Thermometers

To measure temperature, we need some system with an easily-measurable property which changes as its temperature changes (a “thermometer”). A simple thermometer can be constructed by filling a glass tube with some fluid which expands when heated. Mercury is again a convenient choice. The tube is marked at the mercury level corresponding to some reference temperature (maybe the freezing point of water), and marked again at another reference temperature (maybe the boiling point).

The interval between the two marks is then divided into an arbitrary number of equal segments, and the mark the mercury rises is used as the measure of temperature. The marks could be assigned numbers, but letters, names, or any other symbols would work too. The Fahrenheit temperature scale assigns the values 32 and 212 to the freezing and boiling points, respectively, of water at a pressure of 1 atm. The Celsius scale (formerly called the centigrade scale) assigns the values 0 and 100.

This procedure defines a practically-useful temperature scale, but has a few shortcomings for accurate work. For example, the freezing and boiling points of water depend on pressure; if you wanted to check the calibration of the thermometer you’d have to know and reproduce the atmospheric pressure when the marks were made. The thermal expansion properties of mercury also depend slightly on pressure. For a temperature reading to be meaningful, you would need to report the pressure at which the measurement was made.

If a thermometer were constructed using a different fluid (an alcohol, for example), the freezing and boiling points measured by the two thermometers would agree, but other temperatures would not, since the second fluid expands with temperature in a somewhat different way than mercury does. So when the mercury thermometer reads, say, 50 marks above the freezing-point mark, an alcohol thermometer might read 52 marks. To record a meaningful temperature, you’d have to always record what sort of thermometer was used.

3.4.2 Ideal Gas Temperature Scales

It would be preferable to measure temperature in a way which is completely *independent* of the particular thermometer used. One interesting possibility is to use the fact that *all* gases obey the same equation – the ideal gas law – in

the limit of zero density. (No real gas *exactly* satisfies the ideal gas law at any finite density – see Chapter 4).

The procedure to construct a substance-independent ideal-gas thermometer is as follows. A low density gas is put in a small, fixed-volume closed container. This serves as the thermometer. A pressure gauge of some sort (maybe a capacitance manometer) is connected to the container to measure its pressure.

Now a *single* reproducible reference state is chosen, and the ideal-gas thermometer is put in thermal contact with it and allowed to come to thermal equilibrium.⁴ The pressure reading P_{ref} is recorded.

To measure an unknown temperature, the ideal-gas thermometer is placed in thermal contact with the system of unknown temperature and allowed to come to thermal equilibrium. The pressure in the gas in the container will change to a new value P (remember the thermometer volume and amount of gas inside are fixed). We will *define* the temperature on the ideal-gas scale by

$$T = T_{ref} \left(\frac{P}{P_{ref}} \right). \quad (3.11)$$

Here T_{ref} is an arbitrary constant, which sets the size of a degree.

An interesting feature about this temperature scale is that only *one* reference state is needed (not two, as with the Celsius or Fahrenheit scales). It is best to pick a reference state which can be precisely reproduced. It will be shown in Chapter 4 that for any pure substance there is a *single* temperature (and pressure) at which solid, liquid, and vapor all co-exist. This point is called the *triple point*. The most common procedure is to take the triple point of water as the reference point.

Then to specify a particular ideal-gas scale, all that is left is to decide what number we should assign to T_{ref} . It might seem logical to pick a round number (say, 1, 100, 1000, etc.). On the other hand, it might be convenient if the size of a degree was the same as we are accustomed to on the Celsius or Fahrenheit scale.

To find the necessary T_{ref} value to reproduce the degree size of the Celsius scale, the following experimental procedure can be used. Go to the reference temperatures used to establish the Celsius scale (the freezing and boiling points of water at 1 atm), and measure the pressure ratios (P_{fp}/P_{ref}) and (P_{bp}/P_{ref}) with the ideal-gas thermometer. If we want $T_{bp} - T_{fp}$ to be 100 on the ideal-gas

⁴We assume that the ideal gas only exchanges a negligibly small amount of energy with the system in coming to thermal equilibrium, so that it doesn't affect the temperature being measured. We can make this assumption as good as we like by simply decreasing the total amount of gas in the thermometer.

scale, then T_{ref} must satisfy

$$T_{ref} \left[\left(\frac{P_{bp}}{P_{ref}} \right) - \left(\frac{P_{fp}}{P_{ref}} \right) \right] = 100. \quad (3.12)$$

With the measured pressure ratios, the required value for T_{ref} is found to be

$$T_{ref} = 273.16. \quad (3.13)$$

The temperature scale defined in this way is known as the Kelvin scale. Since the triple point of water occurs at 0.01 °C,

$$T(\text{K}) = T(^{\circ}\text{C}) + 273.15. \quad (3.14)$$

Note that the Kelvin temperature scale has an *absolute* zero, where the pressure in the gas in principle becomes zero. The only freedom we have is in setting the size of a degree, not where the zero point is.

To construct an ideal-gas scale with the degree size of the Fahrenheit scale, choose $T_{ref} = 1.8 \times 273.16 = 491.69$. This scale is known as the Rankine temperature scale, and of course has its zero at the same temperature as the Kelvin scale. Since on the Fahrenheit scale the triple point of water is at 32.02 °F,

$$T(\text{R}) = T(^{\circ}\text{F}) + 459.67. \quad (3.15)$$

In practice, ideal gas temperature scales have some limitations. First of all, the pressure readings not entirely independent of the gas chosen, since P may be small but it is not zero. To remove this dependence, multiple thermometers would have to be used, each with a different amount of gas inside, and their readings extrapolated down to zero pressure. Also, gases condense at low temperature. Even using helium it is not possible to measure temperatures below 1 K with an ideal gas thermometer. At the other extreme, at temperatures above a few thousand Kelvin, gases dissociate to form plasmas. This puts an upper limit on the useful temperature range. Outside this range, temperature must be defined and measured some other way.

3.4.3 Thermodynamic Temperature

We'll show in Chapter 6 that there is a more fundamental thermodynamic way to define temperature, which is not dependent on the behavior of any particular type of thermometer, not even an ideal gas one. This *thermodynamic temperature* may be defined over the range zero to infinity, and can be measured with good accuracy in various temperature ranges using different physical phenomena. Temperatures as low as 10^{-8} K and higher than 10^6 K have been

measured using techniques which rely on the definition of the thermodynamic temperature.

In the temperature range where ideal-gas thermometers can be used, the thermodynamic temperature is identical to the temperature defined by the ideal gas scale. Both the Kelvin and Rankine scales may be considered to be thermodynamic temperature scales. The Celsius and Fahrenheit scales are not, however, since they have their zero points in the wrong place. They are acceptable to use for temperature differences, but not for absolute (i.e. thermodynamic) temperatures.

3.5 Intensive and Extensive Properties

In Chapter 2, we introduced several thermodynamic properties of macroscopic matter, such as the volume V , the mass M , the total energy E , the internal energy U , the surface area A , the total electric dipole moment $V\mathbf{P}$, and the total magnetic moment $V\mathbf{M}$. These properties are defined both for equilibrium and non-equilibrium states of matter. They are also additive: if we have a system composed of two parts A and B, $M_{A+B} = M_A + M_B$, $E_{A+B} = E_A + E_B$, etc. Thermodynamic properties with these characteristics are called *extensive*.

In this chapter, we have introduced three new properties (T , P , μ) which serve to determine whether or not two systems or parts of a single system are in equilibrium with one another. These properties have a very different character. For a system composed of parts A and B, T_{A+B} is meaningful only if $T_A = T_B$ (thermal equilibrium). Otherwise, there is no *single* temperature of the whole system. If A and B are in thermal equilibrium, then $T_{A+B} = T_A$ (not $2T_A$). Pressure and chemical potential have similar characteristics: they are only defined for systems in equilibrium, and are not additive for systems composed of multiple parts. Properties like this are called *intensive*. All properties in thermodynamics are either extensive or intensive.

The extensive properties depend on how much matter the system contains. If a substance in an equilibrium state with mass M , volume V and internal energy U is divided into n equal parts, each part will have mass M/n , volume V/n , and internal energy U/n . Because the extensive properties are proportional to how much matter is in the system, it is usually more convenient to work with quantities normalized to a unit amount.

One choice is to normalize to a unit mass. The *specific* internal energy is defined to be the internal energy per unit mass:

$$u = \frac{U}{M}. \quad (3.16)$$

We always use the term “specific” to mean “per unit mass,” and use a lower-case letter to denote a specific property. We can also define the specific volume

$$v = \frac{V}{M}, \quad (3.17)$$

which is simply the reciprocal of the mass density ρ .

Alternatively, we could choose to normalize to one mole. The *molar* internal energy is defined to be

$$\hat{u} = \frac{U}{N}, \quad (3.18)$$

where N is the number of moles, and the molar volume is

$$\hat{v} = \frac{V}{N}. \quad (3.19)$$

We use the hat notation to denote molar properties. Note that both specific and molar properties are intensive (independent of the size of the system), since they are the ratio of 2 extensive properties.

In engineering thermodynamics, it is most common to work with specific properties (per unit mass), since mass is conserved in any process but moles may not be if chemical reactions occur. Also, important engineering quantities like the momentum of a flowing fluid depend on the rate of *mass* flow, not the molar flow rate.

3.6 The Thermodynamic State

We’ve now introduced a rather long list of material properties ($U, V, M, P, T, \sigma, A, \mathbf{E}, \mathbf{P}, \mathbf{H}, \mathbf{M}$) and we will soon introduce more properties. Fortunately, we usually only need to work with a subset of these which relate to the ways energy can be transferred to or from the substance as work or heat.⁵ For equilibrium states, even the properties in this subset are not all independent. Specifying a small number of them is sufficient to determine all the rest. For example, if the mass, volume, and temperature of an equilibrium sample of water are given, then its pressure, internal energy, and phase are fully determined.

When enough property values for an equilibrium state are specified so that all the other relevant properties of that state can be determined, we say that the *thermodynamic state* is specified. The number of properties which must be given to specify the thermodynamic state is called the number of *degrees of freedom* f of the substance.

⁵For example, we don’t usually care about the magnetization of water, since for achievable magnetic fields it is very, very, small, and the possible magnetic work is therefore negligible.

The number of degrees of freedom is also equal to the number of independent ways the equilibrium state can be altered. If there are two independent ways to change the state, two independent properties will be needed to specify the state, etc.

Consider a fixed mass M of some substance for which the only reversible work mode is compression, and which is in an equilibrium state with no macroscopic kinetic or potential energy. The substance is in a cylinder with a piston which can be used to change its volume. The substance may be solid, liquid, or gaseous, or may consist of an equilibrium mixture of phases.

If we add heat $\bar{d}Q$ and do quasi-static work $\bar{d}W_{qs} = -PdV$, the First Law for this process is

$$dU = \bar{d}Q - PdV. \quad (3.20)$$

From this, we see that the two extensive properties U and V may be varied independently. To change U by dU holding V constant, just add heat in the amount $\bar{d}Q = dU$. To change V by dV holding U constant, add heat in the amount $\bar{d}Q = +PdV$ while changing the volume. This is true no matter what the details are of the substance in the container (a single phase, a mixture of phases, etc.)

Suppose U and V are held fixed and we try to vary something else (maybe T or P). Since V is fixed, no compression work can be done. But since U is fixed, no heat can be added either. So there's really no way to change any other properties (T , P , etc.) without changing U or V or both. We conclude that specifying U and V (for given M) fully specifies the thermodynamic state of this substance.

Now consider a substance with two reversible work modes – maybe in addition to being compressible, it has a surface area, and non-negligible surface tension. Then the First Law would become

$$dU = \bar{d}Q - PdV + \sigma dA. \quad (3.21)$$

Now it *is* possible to hold U and V fixed and vary another property: all we have to do is change the shape at constant volume, changing the surface area. This does some work, but as before we just remove some heat to get U back to the original value ($\bar{d}Q = -\sigma dA$). So in this case, A can be varied independently of U and V . But there would be no way to hold U , V , and A fixed for this system and vary something else. We conclude that specifying U , V , and A in this case fully specifies the thermodynamic state of the system.

We begin to see a pattern here: The number of degrees of freedom equals the number of reversible work modes, plus one more (to account for heat addition).

If the number of reversible work modes is r , then we have

$$f = r + 1 \quad (3.22)$$

The independently-variable properties which serve to specify the thermodynamic state are U plus the generalized displacements corresponding to each reversible work mode.

It might be objected that in this analysis we have only considered reversible work, and have ignored irreversible work modes. But as we discussed in Chapter 2, the change of state produced by irreversible work can be reproduced by reversible work plus some amount of heat input. For example, electrical work done on a system by current flowing through a resistor is fully equivalent to heat addition, while rapid compression of a gas results in a state which can be reproduced by quasi-static, reversible compression plus heat input. Irreversible work does not represent an *independent* way to alter the state of the system, and therefore does not affect the calculation of the number of degrees of freedom.

3.7 Equations of State

From Eq. (3.22), we see that a substance with only one reversible work mode has 2 degrees of freedom. Such substances are known as *simple substances*. If only compression work can be done, then the substance is a *simple compressible substance* (SCS); if only magnetic work can be done, it is a *simple magnetic substance* (SMS), and so on. Often, a substance may really have more than one reversible work mode, but only one is important. In this case, the substance is often approximated as a simple one, considering only the dominant work mode. For example, we usually may neglect compression work in comparison to magnetic work when dealing with magnetic solids.

For a simple compressible substance, specifying U and V fixes the thermodynamic state for given M . Since U and V are extensive and simply scale with the total mass, the more important parameters are actually $u = U/M$ and $v = V/M$. Therefore, properties such as temperature and pressure are fixed once u and v are specified: there must exist some single-valued function $T(u, v)$ and some single-valued function $P(u, v)$. We call functions like this which express relationships among properties of substances in equilibrium states *equations of state*.

The two independent variables in the equation of state don't have to be (u, v) . For example, if heat dQ is added to an SCS holding v constant, u increases

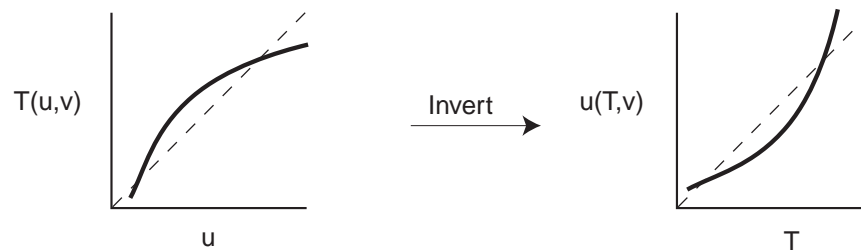


Figure 3.5: Inverting $T(u, v)$ to form $u(T, v)$ by reflecting across the dashed $u = T$ line for specified v .

by dQ (the First Law); it is also always observed that the temperature change dT is positive.⁶ That is, $(\partial T(u, v)/\partial u)_v > 0$, which means that $T(u, v)$ is a monotonically increasing function of u for fixed v . Since monotonic function can be inverted, it is possible to form the function $u(T, v)$ (Fig. 3.5).

Therefore, the pair (T, v) are always independently variable. Now substitute $u(T, v)$ into $P(u, v) = P(u(T, v), v)$ resulting in a new function $P(T, v)$. This new function doesn't involve energy, and is called the *mechanical equation of state*. It is also relatively easy to measure, which is why most experimental investigations of equations of state of simple compressible substances focus on measuring $P(T, v)$. We will examine the nature of this function for an SCS in the next chapter.

Note that $P(T, v)$ is a very different function than $P(u, v)$, and the partial derivative $(\partial P/\partial v)$ depends very much on whether it is u or T we're holding constant. For this reason, we always use a subscript in thermodynamics to state what variable is being held constant in a partial derivative: $(\partial P/\partial v)_T$ is a partial derivative of $P(T, v)$, while $(\partial P/\partial v)_u$ is a partial derivative of $P(u, v)$.⁷

We have to be a little careful in choosing which 2 variables to take as the independent ones. Both (u, v) and (T, v) are always independent, but not all pairs of variables are. In particular, P and T are not always independent. If, for example, the substance consists of an equilibrium solid/liquid mixture, then P and T become coupled – T must equal the melting temperature at the pressure P .

Equations of state also may be defined for substances other than simple compressible substances. For a simple magnetic substance, an argument analogous

⁶This observation will be proven to be true in Chapter 6.

⁷This is no different than with any other functions. For example, consider $f(x, y) = x^2y$. Differentiating with respect to x , $(\partial f/\partial x)_y = 2xy$. Now change variables to (x, z) , where $z = xy$. Then $f(x, z) = xz$ and $(\partial f/\partial x)_z = z = xy$ not $2xy$.

to the one above leads to the conclusion that u and \mathbf{M} are always independently variable, and can be used to define the thermodynamic state. Therefore, there must exist equations of state $T(u, \mathbf{M})$ and $\mathbf{H}(u, \mathbf{M})$. Also, $T(u, \mathbf{M})$ can be inverted to form $u(T, \mathbf{M})$, which can then be used to form the equation of state $\mathbf{H}(T, \mathbf{M})$. Like $P(T, v)$, this equation of state can be easily measured. For other simple substances, the procedure is analogous to that outlined here.

CHAPTER 4

THE SIMPLE COMPRESSIBLE SUBSTANCE

4.1 Introduction

Since all matter can be compressed if a large enough pressure is applied, a study of the thermodynamic properties of the simple compressible substance is a good starting point for any description of the macroscopic properties of matter in equilibrium. Simple compressible substances are also by far the most important ones for engineering thermodynamics, since most (but not all) power plants and engines employ compression, heating, and expansion of a fluid to produce power.

A substance may be approximated as a simple compressible substance if effects due to other reversible work modes are negligible. For example, if the surface-to-volume ratio of a large body of water is small enough, then surface tension will not measurably affect the properties of the water except very near the surface. On the other hand, surface tension will have a dramatic influence on the properties of a very small water droplet, and will, for example, cause the pressure inside the droplet to be elevated above the value predicted if surface tension were neglected. Clearly, a very small water droplet can't be treated accurately as a simple compressible substance, while a large body of water is approximated very well in this way.

In this chapter, we examine the properties of simple compressible substances. We will restrict attention to *pure* substances, which contain only one type of molecule. Mixtures will be considered in a later chapter.

4.2 Phases of a Simple Compressible Substance

A simple compressible substance may exist in different phases: solid, liquid, or gas. Some substances have multiple solid phases, some even have multiple liquid phases (helium), but all have only one gas phase.

An experimental apparatus is shown in Fig. 4.1 which can be used to measure the properties and phases of a simple compressible substance as a function of temperature and pressure. A cylindrical solid sample is placed in a vertical cylinder of the same diameter, which is fitted with a piston. The ambient pressure is P_0 , and the piston weight provides a constant downward force $F =$

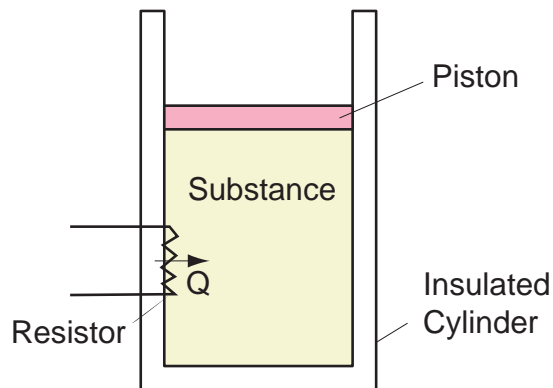


Figure 4.1: Constant-pressure heating experiment.

$M_p g$. The pressure in the cylinder is

$$P = F/A = M_p g/A + P_0. \quad (4.1)$$

The sample height is small enough that the pressure may be taken to be uniform within the sample. The cylinder and piston are well insulated, so there is no heat loss to the environment.

A small amount of heat Q is added by briefly passing current through a resistor mounted in the cylinder wall, after which the system is allowed to re-establish equilibrium. Once the system has come back to equilibrium, both the temperature and the volume may have changed. The new temperature is measured with a thermometer, and the new volume by the piston height.

If Q is sufficiently small, the expansion will occur slowly enough that friction between the piston and cylinder is negligible. In this case, even if the piston oscillates for a while due to the perturbation, once the oscillations have died out and the piston has settled down at a new height, the work done by the substance on the piston will be equal to the work done against atmospheric pressure, plus the change in the gravitational potential energy of the piston:¹

$$W = (M_p g + P_0 A) \Delta y = (M_p g + P_0) (\Delta V / A) = (M_p g / A + P_0) \Delta V = P \Delta V. \quad (4.2)$$

An energy balance on the substance yields the change in its internal energy:

$$\Delta U = Q - P \Delta V. \quad (4.3)$$

¹If friction were not negligible, some kinetic energy of the piston would be converted to internal energy in the piston or cylinder due to friction, and therefore the work would be $> P \Delta V$.

The quantities Q , P , and ΔV are all measured, so we can calculate ΔU . Since P is a constant in this experiment, this equation may be rearranged in the form

$$\Delta(U + PV) = Q. \quad (4.4)$$

The combination $U + PV$ occurs often in analysis of problems at constant pressure. Since U , P , and V are material properties, so is $U + PV$. Rather than always write $U + PV$, we give this property its own name and symbol: the *enthalpy!definition* H is defined by

$$H = U + PV. \quad (4.5)$$

Like U and V , H is an extensive property.

In terms of the enthalpy, Eq. (4.4) becomes

$$\Delta H = Q. \quad (4.6)$$

For heat addition *at constant pressure*, the heat added equals the change in enthalpy of the substance. In contrast, recall from the First Law that if heat is added *at constant volume* ($W = 0$), then $\Delta U = Q$.

Returning to our experiment, the process is now repeated many times, and the resulting property values are recorded at every step: heat Q is added, time is allowed to elapse to re-establish equilibrium, the new T and V are measured, H is incremented by Q .

After n heat addition steps, the volume and temperature have values V_n and T_n , and enthalpy H_n of the substance relative to its starting value H_0 is

$$H_n - H_0 = nQ. \quad (4.7)$$

Since the extensive properties (V and H) depend on how much of the substance was placed in the container, it is preferable to convert them to specific quantities ($v = V/M$, $h = H/M$). In Fig. 4.2, the measured temperature and change in specific enthalpy are shown plotted vs. the measured specific volume (connecting the individual measurements with solid lines).

When heat is first added to the solid, its temperature increases and it expands slightly (region a-b in Fig. 4.2). At point b, the temperature stops increasing, although the volume still increases. A look inside the cylinder reveals the presence of some liquid – the solid is melting.

At point c, all of the solid has melted, and precisely at this point the temperature begins to rise again. But when point d is reached, it stops again and the volume begins to increase significantly. Bubbles are observed to begin forming

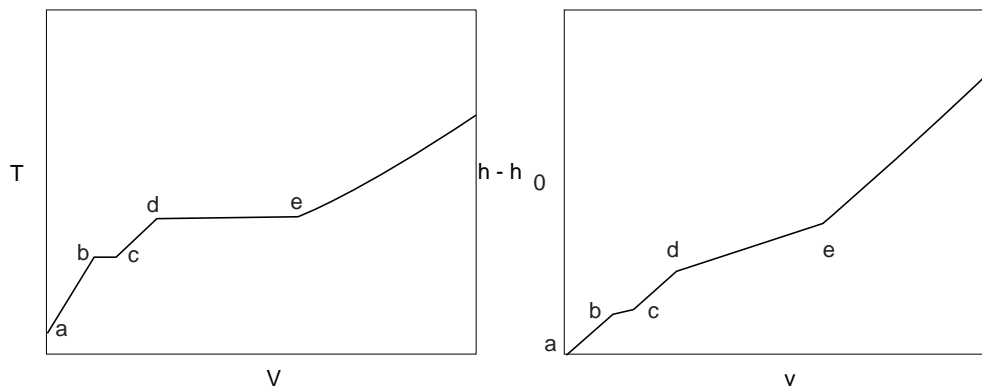


Figure 4.2: Measured temperature (a) and specific enthalpy change (b) vs. measured specific volume for constant-pressure heating.

in the liquid at point d – the liquid is boiling. Moving across from point d to e, the amount of vapor in the cylinder increases, and the amount of liquid decreases. At point e, no liquid remains, and both temperature and volume increase upon further heat addition.

Figure 4.2 shows that when two phases are present (solid/liquid or liquid/vapor), $h - h_0$ continues to increase with v even though T is constant, since energy input is required to convert solid to liquid, or liquid to vapor. Note that the *only* way to measure h is by means of Eq. (4.7), which actually only allows the *change* in h from the initial state to be determined. There is no experiment we could do to measure the value in the initial state h_0 .² Since $h = u + Pv$, if h_0 can't be determined, then u_0 can't be either. This isn't a problem, however, since only *differences* of energy (or enthalpy) have any physical significance. We can start the experiment in some convenient, reproducible state, and simply assign any value we like to h_0 (for example, $h_0 = 0$). We call the initial state with arbitrarily-chosen h_0 the *reference state* or *datum state*.

The results shown in Fig. 4.2 are for a single pressure, $P = M_p g/A + P_0$. By changing the mass of the piston, we can repeat the experiment for different pressures, determining $T(v)$ and $h(v) - h_0$ curves for a range of pressures.

Typical $T(v)$ curves are shown in Fig. 4.3(a) with the points on different curves where the slope is discontinuous connected by dotted lines. The curve corresponding to Fig. 4.2(a) is labeled P_1 on this plot. In Fig. 4.3(b), the dashes

²We could measure it by starting in some other state, but it wouldn't be the initial state; the measured value would be relative to the enthalpy in this new initial state, which would still be undetermined.

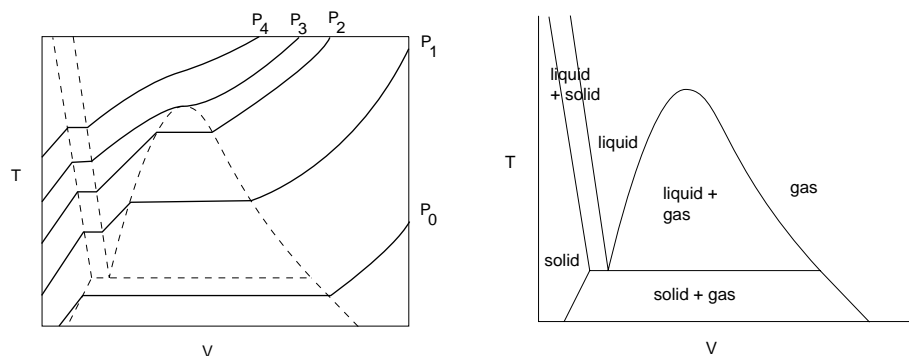


Figure 4.3: (a) $T - v$ curves for constant-pressure heating; (b) phase diagram.

lines are drawn solid, and each region is labeled by which phase or phases are present in the cylinder. Diagrams like this are called *phase diagrams*.

The experimental observations are as follows. As the pressure is increased, the temperature at which liquid appears (the melting point) and the temperature at which vapor appears (the boiling point) both increase. Also, the specific volume of the liquid increases to a larger value before boiling begins, and the specific volume of the vapor once the last liquid has evaporated decreases. Therefore, the change in specific volume upon boiling decreases as the pressure increases.

Beyond a particular pressure P_3 , the $T(v)$ curve changes character. As P_3 is approached, the change in specific volume upon boiling goes to zero – the liquid and vapor approach the same density, and in fact become identical in all respects at P_3 . For $P > P_3$, there is no longer a meaningful distinction between liquid and vapor, and there is no longer any conventional boiling behavior observed. In this pressure regime, as heat is added the high-density fluid simply expands continuously and homogeneously to a low-density fluid, without ever breaking up into separate liquid and vapor regions within the cylinder.

Pressure P_3 is known as the *critical pressure* P_c . Below P_c , the transformation from liquid to vapor upon heating occurs by means of the fluid in the cylinder splitting into two separate regions (high-density liquid and low-density vapor); as more heat is added, the liquid portion shrinks, and the vapor portion grows. Above P_c , the transformation from liquid to vapor occurs continuously, with the fluid remaining uniform throughout the cylinder at all times.

As P approaches P_c from below, the limiting value approached by the boiling

Table 4.1: Critical temperature, pressure, and density for a few substances.

	T_c (K)	P_c (MPa)	ρ_c (kg/m ³)
Helium-4	5.20	0.2275	69.64
Hydrogen	32.94	1.28	31.36
Nitrogen	126.2	3.4	314.03
Oxygen	154.6	5.04	436.15
Methane	190.6	4.60	160.43
Carbon Dioxide	304.2	7.38	464.00
Water	647.3	22.1	317.0

temperature is known as the *critical temperature* T_c . The $T(v)$ curve for $P = P_c$ has an inflection point at $T = T_c$:

$$\left(\frac{\partial T}{\partial v}\right)_P = 0 \quad \text{and} \quad \left(\frac{\partial^2 T}{\partial v^2}\right)_P = 0 \quad \text{at } T = T_c, P = P_c. \quad (4.8)$$

The specific volume of the fluid at the point where $P = P_c$ and $T = T_c$ is known as the *critical specific volume* v_c ; the reciprocal of v_c is the *critical density* ρ_c . The quantities T_c , P_c , and v_c (or ρ_c) define the *critical point*. The critical point quantities for a few substances are listed in Table 4.2.

If the pressure is now lowered below P_1 , another change in the character of the $T(v)$ curve is observed. On the P_0 curve, there is only one segment where T is constant, not two. On the flat segment, a solid/vapor mixture is found in the cylinder, rather than a solid/liquid or liquid/vapor mixture. Evidently, at sufficiently low pressure no liquid phase forms – instead, the solid transforms directly to vapor. This process is known as *sublimation*.

4.3 $P - v - T$ Surfaces

Since any two independent properties serve to define the thermodynamic state for a simple compressible substance, we may regard any other property as a function of these two. For example, at every (T, v) point in Fig. 4.3 we know the pressure, so we can construct $P(T, v)$.

The function $P(T, v)$ defines a surface over the $T - v$ plane. A typical $P - v - T$ surface is shown in Fig. 4.4. Every equilibrium state of the substance corresponds to some point on the $P - v - T$ surface. In the pure solid region and in the pure liquid region below T_c , the slope of the surface is very steep, since compressing a solid or liquid even a little requires a huge increase in pressure. In the gas or vapor region, the surface is gently sloped, since gases are easily

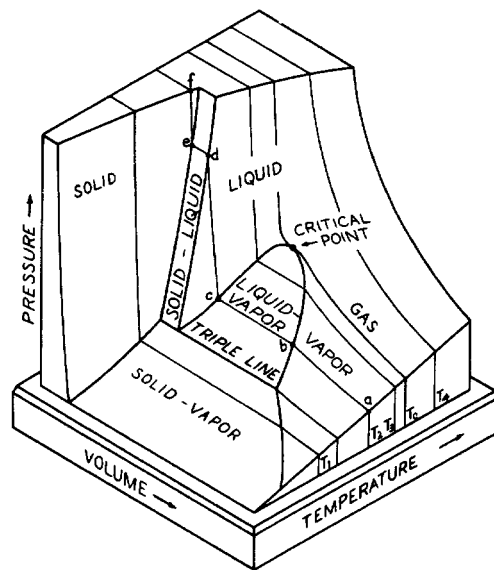


Figure 4.4: A $P - v - T$ surface for a substance which expands upon melting.

compressed. Of course, above the critical point the liquid and gas regions merge smoothly.

In Fig. 4.4 the term “gas” is used above the critical point, and “vapor” below it. This convention dates back to the early 19th century, when it was thought that “gases” like oxygen were different than “vapors” like steam. Vapors could be condensed to liquid, but gases (it was believed) could not be. When it was demonstrated in 1877 that oxygen and nitrogen could be liquified at sufficiently low temperature, it became clear that there was no fundamental distinction between gases and vapors; the only difference is that substances such as oxygen have critical temperatures well below room temperature, while the critical temperature of water is above room temperature. Thus, liquid water is commonplace, but liquid oxygen, nitrogen, hydrogen, or helium are not. However, processes to produce these as liquids are now straightforward, and liquid oxygen, nitrogen, and helium have very significant technological applications.³

The slope of the $P - v - T$ surface is discontinuous on the boundary between the single-phase and the two-phase regions. Note that since T remains constant

³For example, liquid hydrogen and oxygen as used as propellants in the Space Shuttle Main Engine; liquid nitrogen is widely used to cool electronic equipment and photodetectors; liquid helium is used to cool large superconducting magnets to temperatures a few degrees above absolute zero.

Table 4.2: Triple-point temperatures and pressures.

	T_t (K)	P_t (kPa)
Helium-4	2.18	5.1
Hydrogen	13.8	7.0
Nitrogen	63.15	12.5
Oxygen	54.34	0.14
Methane	90.68	11.7
Carbon Dioxide	216.54	517.3
Water	273.16	0.61

in the two-phase regions during constant-pressure heating, any horizontal slice through the surface in a two-phase region must produce a line perpendicular to the T axis; in other words, the slope of the surface in the T direction is zero in the two-phase regions.

The solid-vapor two-phase region intersects the solid-liquid and liquid-vapor regions in a single line parallel to the volume axis. This line is known as the *triple line*, since along this line all three phases may coexist in equilibrium. The pressure and temperature are the same everywhere along the triple line. Therefore, for a given substance, there is only *one* pressure P_t and *one* temperature T_t at which solid, liquid, and vapor may coexist in equilibrium. The combination (T_t, P_t) is known as the *triple point*. Of course it is only a *point* in the $P - T$ plane; when the volume axis is considered, it is a line. This is in contrast to the critical point, which is really a point in (P, v, T) space.

The triple points for several substances are listed in Table 4.3. Note that of the ones listed, only carbon dioxide has $P_t > 1$ atm ($1 \text{ atm} = 101.325 \text{ kPa}$). Therefore, at 1 atm pressure, solid carbon dioxide (“dry ice”) sublimates, while solid water melts.

The $P - v - T$ surface shown in Fig. 4.4 is appropriate for a substance which expands upon melting, which we assumed implicitly in our discussion above. However, a few substances – including water – contract when they melt. For these substances, the $P - v - T$ surface looks like that shown in Fig. 4.5.

If a substance has multiple solid phases, the $P - v - T$ surface can become very complex. A portion of the actual surface for water is shown in Fig. 4.6, showing the different phases of ice.

Two-dimensional phase diagrams may be obtained by projecting the $P - v - T$ surface onto the $P - T$, $T - v$, or $P - v$ planes. We have already looked at the $T - v$ projection in constructing the $P - v - T$ surface [Fig. 4.3(b)]. The $P - T$

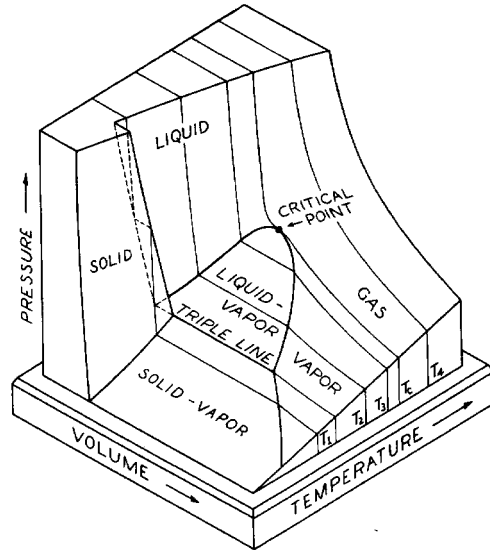


Figure 4.5: A $P - v - T$ surface for a substance which contracts upon melting.

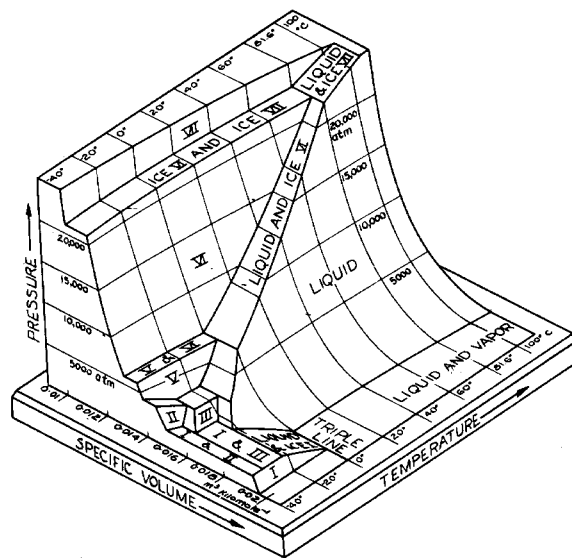


Figure 4.6: A portion of the $P - v - T$ surface for water.

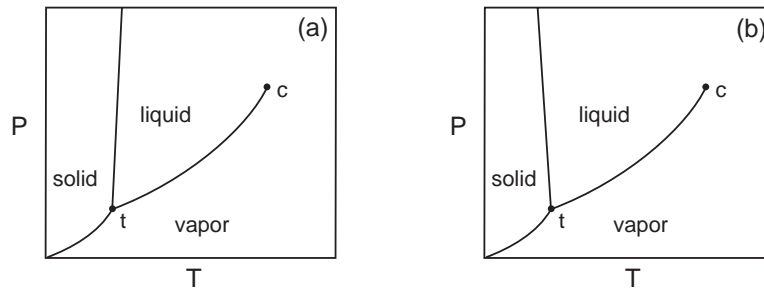


Figure 4.7: $P - T$ phase diagrams for (a) a substance which expands on melting; (b) a substance which contracts on melting.

projection is shown in Fig. 4.7 for a substance which expands on melting (a) and for one which contracts on melting (b). The critical point (c) and the triple point (t) are both shown.

The lines separating the single-phase regions on a $P - T$ plot are known as *coexistence lines*, since on these lines two phases may coexist. The liquid-vapor coexistence line terminates at the critical point. In contrast, the solid-liquid coexistence line never terminates, no matter how high the pressure. This is because solid and liquid are fundamentally different – in a solid, atoms are arranged in a highly regular, periodic way, while in a liquid they are arranged randomly. There is no way for these states with very different symmetry to transform into one another continuously, and so it is not possible for a critical point to exist on the solid-liquid coexistence line.

Each of the coexistence lines in a $P - T$ phase diagram can be described by some function $P(T)$, so clearly P and T are not independent when two phases are simultaneously present. On the liquid-vapor and solid-vapor coexistence lines, the term *vapor pressure* is used to denote $P(T)$, since this is the pressure of the vapor in equilibrium with the solid or liquid. An equivalent term is *saturation pressure* $P_{sat}(T)$. If pressure is specified, the *saturation temperature* $T_{sat}(P)$ is defined to be the temperature on the coexistence curve where the pressure is P . The saturation temperature is just another name for the boiling temperature. For example, for water, $T_{sat}(1 \text{ atm})$ is 373.15 K (100 °C), and $P_{sat}(373.15 \text{ K}) = 1 \text{ atm}$.

4.4 Determining Properties in the Mixed-Phase Regions

Under conditions where two phases coexist in equilibrium, some care must be taken to correctly determine the properties and amount of each phase from a

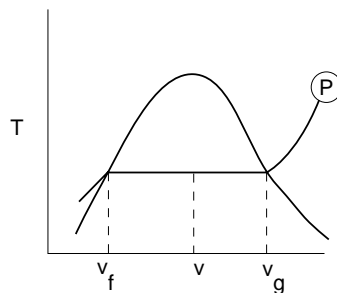


Figure 4.8: In the liquid/vapor two-phase region, the liquid has specific volume v_f and the vapor has specific volume v_g .

phase diagram. The two phases have very different properties. For example, the specific volume of a liquid is much less than that of a gas. How can we determine *both* values from a phase diagram?

The key is to realize that in a two-phase region, the properties of each phase present are those at the “edges” of the region. For example, consider boiling a liquid at constant pressure. Just before the temperature where gas bubbles first appear, the cylinder is still filled with liquid. Call the specific volume at this point v_f .⁴ As more heat is added at constant pressure, the *only* thing that happens is that some liquid becomes vapor – *the properties (per unit mass) of the remaining liquid don’t change*. Although there is less liquid, the liquid remaining still has specific volume v_f .

What is the specific volume of the vapor which has been created? It too is constant during the constant-pressure boiling process, and thus must equal the value of v obtained once the cylinder contains only vapor. Call this value v_g .

Suppose now that the system is somewhere in the two-phase region on the isobar⁵ labeled P in Fig. 4.8 and the measured total volume V results in a value for $v = V/M$ as shown in this figure. This v does not actually correspond to the specific volume of either the liquid or the vapor in the container. These are v_f and v_g , respectively. Instead, v is an *average* of v_f and v_g , weighted by the mass of each in the container.

We use the term “saturated” to denote the states on either side of the vapor dome (Fig. 4.9). Thus, “saturated liquid” has specific volume v_f , and “saturated

⁴Although it is not entirely logical, it is conventional to use the subscript “f” to denote properties of the liquid and “g” to denote properties of the vapor in an equilibrium liquid/vapor mixture.

⁵An *isobar* is a line of constant pressure. The P in a circle simply labels the pressure of this isobar.

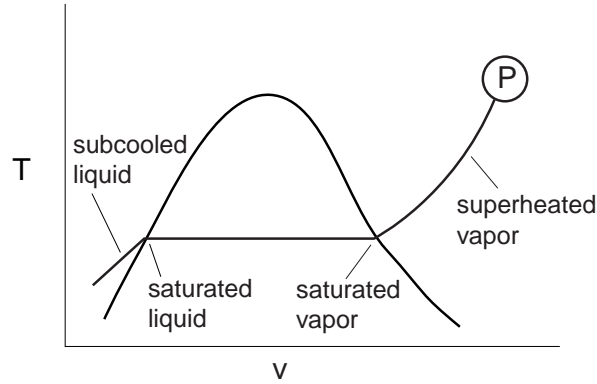


Figure 4.9: Definition of saturated, superheated, and subcooled states.

vapor” has specific volume v_g . When liquid and vapor are present together in equilibrium, the liquid is always saturated liquid, and the vapor is saturated vapor. Vapor at a temperature above $T_{sat}(P)$ is called *superheated vapor*, and liquid at a temperature below $T_{sat}(P)$ is called *subcooled liquid*. Subcooled liquid is also called *compressed liquid*, since it is at a higher pressure than $P_{sat}(T)$.

Let the mass of the *vapor* in the container be Mx , where $0 \leq x \leq 1$. Then the mass of the liquid must be $M(1 - x)$, since the total mass is M . The total volume is then

$$V = M(1 - x)v_f + Mxv_g, \quad (4.9)$$

or

$$v = \frac{V}{M} = (1 - x)v_f + xv_g. \quad (4.10)$$

If we know v , we can then solve for x :

$$x = \frac{v - v_f}{v_g - v_f}. \quad (4.11)$$

The vapor mass fraction x is an intensive thermodynamic property of a liquid/vapor mixture. The common name in engineering thermodynamics for x is the *quality*. This name was given to x by engineers developing steam engines and power plants: the presence of liquid droplets in the steam damages engine parts such as turbine blades, hence from the engineer’s perspective higher “quality” steam had less liquid content. Of course, for other applications the relative merits of liquid and vapor might be reversed. In this book, we will usually refrain from making a value judgement about liquid vs. vapor, and simply call x the vapor mass fraction.

Equation (4.11) is known as the *lever rule*. It may be interpreted in terms Fig. 4.8 as follows. To determine the mass fraction of a phase in a saturated vapor/liquid mixture, locate the system point in the two-phase region on a $T-v$ or $P-v$ plot. Now take the length of the horizontal line segment from v to the saturation line corresponding to the *other* phase, and divide this length by the total width of the 2-phase region ($v_g - v_f$). Note that this rule works for any 2-phase region, for example for liquid/solid mixtures.

Example 4.1 A bottle contains 10 kg of carbon dioxide at 260 K. If the volume of the bottle is 100 liters, does the bottle contain liquid, solid, gas, or a mixture? How much of each? What is the pressure?

Solution: Since 260 K is greater than the triple-point temperature for CO_2 , no solid will be present. Calculate $v = V/M = (100 \text{ liters})/(10 \text{ kg})$: $v = 0.01 \text{ m}^3/\text{kg}$. From a phase diagram for CO_2 at 260 K, we find that $v_f = 0.001001 \text{ m}^3/\text{kg}$, and $v_g = 0.01552 \text{ m}^3/\text{kg}$. Since v is between these two values, the bottle contains a mixture of liquid and gaseous CO_2 . The vapor mass fraction is

$$x = \frac{0.01 - 0.001001}{0.01552 - 0.001001} = 0.62.$$

Since it is in the mixed phase region, the pressure is the saturation pressure at 260 K, which is 2.421 MPa.

4.5 Software for Property Evaluation

To solve problems involving real substances (such as the last example), some source of property data is required. Traditionally in thermodynamics courses, properties were looked up in tables, or estimated from detailed phase diagrams. A typical diagram for oxygen is shown in Fig. 4.10. Here the pressure is plotted against the enthalpy, and many curves representing particular values of other properties are shown. Since P and h are two valid, independent properties, the thermodynamic state is represented by a point on this plot. The point may be fixed by interpolation if any two properties are known for which curves are plotted on the chart. Once the state point is found, any other properties can be read off (with some care and practice).

Most thermodynamics textbooks now come with at least some rudimentary software to evaluate properties. In many cases, the programs are simply electronic versions of tables, which print to the screen the property values. In other cases, more elaborate software is provided which allows you to do a complete thermodynamic analysis. But even these packages are specialized applications, which you use for thermodynamics but nothing else.

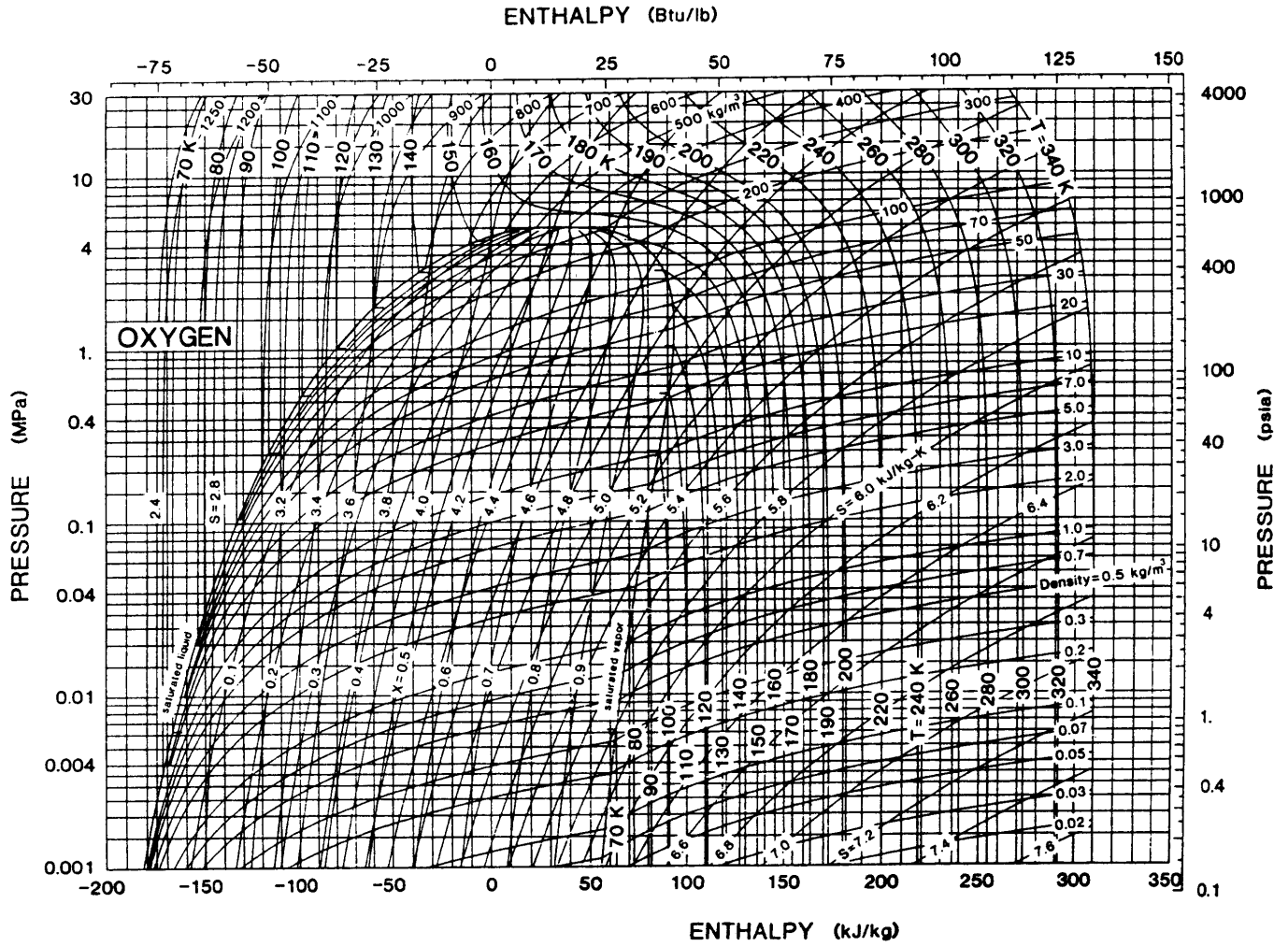


Figure 4.10: A pressure-enthalpy plot for oxygen.

A new software package is provided as a supplement to this book which implements thermodynamic property functions in Microsoft Excel (a spreadsheet program). The thermodynamic property functions are provided by an “add-in” module called TPX (“**T**hermodynamic **P**roperties for **E**Xcel”). Details of how to load it into Excel and use it are given in Appendix A.

Calculating properties is easy with TPX. For example, if you want to know the specific enthalpy of oxygen at 1 MPa and 500 K, you simply type into a cell:

```
=h("o2" , "PT" , 1 , 500)
```

The parameters are: the substance name (case is unimportant); a string stating the properties which will be used to fix the state (here P and T); the value of the first parameter (P); the value of the second parameter (T). The value of the function returned in the cell is 655.83 kJ/kg. You can select any system of units you like; all inputs and outputs will then be in those units. (This example assumes the user selected units of Kelvin for temperature, MPa for pressure, kJ for energy, and kg for mass.)

The real power of using a spreadsheet becomes apparent in more complex analyses. For example, the temperature and pressure may not be specified inputs, but are themselves the result of calculations in other cells. In this case, simply replace the numerical value in the function parameters by the appropriate cell address (e.g. B4).

The same functions implemented by TPX are also available in a WWW property calculator. The calculator is convenient for simple calculations, if you have access to the Web but not to Excel.

Example 4.2 One kg of water is placed in a closed container and heated at constant volume. The initial temperature is 300 K. If the desired final state is the critical point, determine the necessary container volume, initial pressure, initial vapor mass fraction, and energy transfer as heat.

Solution: Two properties are needed to specify the initial state. The temperature is given, so one more is required. Since the process occurs at constant volume, the initial volume V_1 must equal the final volume V_2 . The final state is the critical point, so

$$V_2 = (1 \text{ kg}) \times v_c.$$

The critical specific volume of water may be calculated using TPX:

$$v_c = \text{vcrit}(\text{"h2o"}) = 0.003155 \text{ m}^3/\text{kg}.$$

Therefore, state 1 is fixed by $T_1 = 300 \text{ K}$ and $v_1 = v_c$. Using TPX, any other

desired property for state 1 may be computed:

$$\begin{aligned} P_1 &= 3528.2 \text{ Pa} & P(\text{"H2O"}, \text{"TV"}, 300, \text{vcrit}(\text{"H2O"})) \\ X_1 &= 5.49 \times 10^{-5} & x(\text{"H2O"}, \text{"TV"}, 300, \text{vcrit}(\text{"H2O"})) \\ u_1 &= 112.727 \text{ kJ/kg} & u(\text{"H2O"}, \text{"TV"}, 300, \text{vcrit}(\text{"H2O"})) \\ h_1 &= 112.738 \text{ kJ/kg} & h(\text{"H2O"}, \text{"TV"}, 300, \text{vcrit}(\text{"H2O"})) \end{aligned}$$

Note that the initial state is a mixed liquid/vapor state, and so $P_1 = P_{sat}(300 \text{ K})$. The vapor fraction is very small, since v_c is only slightly greater than the specific volume of the saturated liquid.

The required heat transfer Q is determined from the First Law:

$$\Delta U = Q + W.$$

Since the volume is constant, $W = 0$, so

$$Q = \Delta U = M(u_2 - u_1).$$

Using TPX,

$$u_2 = u(\text{"h2o"}, \text{"tv"}, \text{tcrit}(\text{"h2o"}), \text{vcrit}(\text{"h2o"})) = 2029.6 \text{ kJ/kg.}$$

$$\text{so } Q = (1 \text{ kg})(2029.6 \text{ kJ/kg} - 112.7 \text{ kJ/kg}) = 1916.9 \text{ kJ.}$$

4.6 More Properties: Partial Derivatives of Equations of State

Consider an equation of state like $P(T, v)$. Clearly, a partial derivative of this function, for example $(\partial P / \partial T)_v$, is some new function of (T, v) and may rightly be regarded as a thermodynamic property of the system. Some useful derivative properties are defined here.

4.6.1 Thermal Expansion Coefficient

Most substances expand when heated. The property which tells us how much a substance expands when heated at constant pressure is the *thermal expansion coefficient* β , defined by

$$\beta = \frac{1}{v} \left(\frac{\partial v}{\partial T} \right)_P \quad (4.12)$$

To calculate β , the equation of state $v(T, P)$ would be differentiated with respect to T .

Note the $1/v$ in the definition – β is defined as the *fractional* change in volume per degree of temperature increase. Also, note that this definition is only really meaningful if a single phase is present, since otherwise T can't be increased holding P constant.

Example 4.3 What is the thermal expansion coefficient of liquid water at 300 K and 1 atm?

Solution: β may be calculated approximately using TPX, evaluating the partial derivative by a finite-difference approximation. Taking a small increment of, say, 0.1 K, use TPX (with units set to K and atm) to evaluate $(\partial v/\partial T)_P$:

$$\begin{aligned} \left(\frac{\partial v}{\partial T}\right)_P &\approx (v(\text{"h2o"}, \text{"tp"}, 300.1, 1) - v(\text{"h2o"}, \text{"tp"}, 300, 1))/0.1 \\ &= 2.75 \times 10^{-7} \text{ m}^3/\text{kg-K}. \end{aligned} \tag{4.13}$$

Since

$$\begin{aligned} v &= v(\text{"h2o"}, \text{"tp"}, 300, 1) = 0.001003378 \text{ m}^3/\text{kg}, \\ \beta &\approx 2.74 \times 10^{-4} \text{ K}^{-1}. \end{aligned}$$

4.6.2 Isothermal Compressibility

All matter decreases slightly in volume if the pressure is increased at constant temperature. The property how the volume varies with pressure at constant temperature is the *isothermal compressibility* κ , defined by

$$\kappa = -\frac{1}{v} \left(\frac{\partial v}{\partial P}\right)_T \tag{4.14}$$

As with β , if the equation of state $v(T, P)$ is known, it can be differentiated to find $\kappa(T, P)$. And as with the thermal expansion coefficient, κ is only meaningful if the substance is in a single phase.

4.6.3 Specific Heats

Suppose a unit mass of a substance absorbs an amount of heat $\bar{d}Q$; how much does the temperature increase? It depends in part on how the heating is done. From the First Law,

$$du = \bar{d}Q + \bar{d}W = \bar{d}Q - Pdv. \tag{4.15}$$

Constant Volume

If the heating is done at constant volume, the work term is zero. Therefore

$$du = \bar{d}Q. \quad (4.16)$$

Differentiating the equation of state $u(T, v)$ at constant v , we may also write for du

$$du = \left(\frac{\partial u(T, v)}{\partial T} \right)_v dT. \quad (4.17)$$

Equating these two expressions for du produces the desired relationship between $\bar{d}Q$ and dT :

$$\bar{d}Q = \left(\frac{\partial u}{\partial T} \right)_v dT. \quad (4.18)$$

The *specific heat at constant volume* c_v is defined by

$$c_v = \left(\frac{\partial u}{\partial T} \right)_v \quad (4.19)$$

Then

$$du = c_v dT \quad (4.20)$$

for constant-volume heating. The SI units of c_v are J/kg/K.

Constant Pressure

At constant pressure, we have to be careful to remember to account for the work done against the environment as the sample expands. Rearranging Eq. (4.15), we find $\bar{d}Q = du + Pdv$. But since P is constant in the process, $du + Pdv = d(u + Pv)$. We recognize $u + Pv$ to be the specific enthalpy h .

Therefore, for constant pressure heating, $\bar{d}Q = dh$. We can also write

$$dh = \left(\frac{\partial h(T, P)}{\partial T} \right)_P dT, \quad (4.21)$$

since P is constant. Therefore, for *constant pressure* heating, $\bar{d}Q$ and dT are related by

$$\bar{d}Q = \left(\frac{\partial h}{\partial T} \right)_P dT. \quad (4.22)$$

The *specific heat at constant pressure* c_p is defined by

$$c_p = \left(\frac{\partial h}{\partial T} \right)_P \quad (4.23)$$

The units are the same as those for c_v .

Another common term for c_v and c_p is *heat capacity*. Heat capacities may also be defined per mole of substance (\hat{c}_v and \hat{c}_p) or for a total amount of substance of mass M (C_v and C_p).

Finally, it is worth noting that the names “heat capacity” and “specific heat” both derive from the old idea that a body can store “heat.” The names shouldn’t be taken too literally – c_v and c_p are simply defined in terms of the derivatives given above.

4.7 Model Equations of State

A software package like TPX evaluates properties using some mathematical functions which have been fit to a large number of experimental measurements. The measured data might be $P(T, v)$, $c_p(T, v)$, or more indirect quantities like the speed of sound. The functions are empirical, and contain many adjustable constants which may be set to fit the measurements as accurately as possible. For example, the functions used by TPX to represent the properties of water contain more than 60 constants, with values chosen to provide the best fit to measurements.⁶

Sometimes it is desirable to work with much simpler approximate equations of state. The price for simplicity is lack of accuracy, but in some cases we are interested in examining qualitative behavior rather than calculating precise numbers. Here we examine some common approximate or *model* equations of state, starting with the simplest and working up to more complex but more accurate ones.

4.7.1 The Ideal Gas

The quantity $P\hat{v}/T$ (where \hat{v} is the *molar* volume) is found to approach the same value for all fluids in limit of low density when the molecules are far from one another (Fig. 4.11). The limiting value is found to be

$$\lim_{P \rightarrow 0} \frac{P\hat{v}}{T} = \hat{R} = 8.314 \text{ kJ/mol-K} \quad (4.24)$$

The constant \hat{R} is known as the *universal gas constant*. If the pressure is low enough that $P\hat{v}/T$ has reached this limiting value, then the equation of state is approximated by

$$P\hat{v} = \hat{R}T. \quad (4.25)$$

⁶J. H. Keenan et al., *Steam Tables*, John Wiley and Sons, New York, 1969.

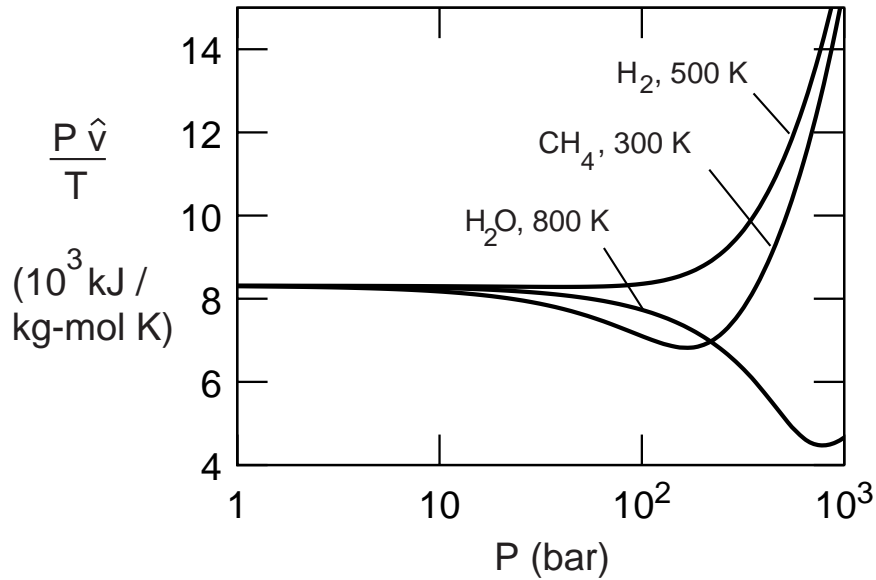


Figure 4.11: Demonstration that $P\hat{v}/T$ approaches the same value for all fluids as $P \rightarrow 0$.

This equation of state is known as the *ideal gas law*, and represents the low-density limit of the equilibrium equation of state of all real simple compressible substances. An *ideal gas* (or *perfect gas*) is defined to be any gas which obeys Eq. (4.25).

Since $\hat{v} = V/N$, the ideal gas law can be written in terms of the total volume V and total number of moles N :

$$PV = N\hat{R}T. \quad (4.26)$$

The value of \hat{R} depends on how N is expressed. If N is expressed as a number of gram-moles (mol), then $\hat{R} = 8.3143 \text{ kJ/mol-K}$; if N is expressed in kg-moles (1 kmol = 1000 mol), then $\hat{R} = 8314.3 \text{ kJ/kmol-K}$. Alternatively, we could dispense with using mole numbers, and express N as the actual number of molecules. In this case, \hat{R} needs to be converted from mol or kmol units to molecule units (1 mol = 6.023×10^{23} molecules; 1 kmol = 6.023×10^{26} molecules). In this case,

$$\hat{R} = (8.3143 \text{ kJ/mol-K})(1000 \text{ J/kJ})(1 \text{ mol}/6.023 \times 10^{23}) = 1.38 \times 10^{-26} \text{ J/K}. \quad (4.27)$$

In these units \hat{R} is usually called *Boltzmann's constant* k_B :

$$k_B = 1.38 \times 10^{-26} \text{ J/K}. \quad (4.28)$$

In this case, the ideal gas law becomes

$$PV = Nk_B T. \quad (4.29)$$

It's important to remember that k_B and \hat{R} are really the same thing – just different units. We'll see Boltzmann's constant again in Chapter 6 when we discuss entropy.

We can also divide Eq. (4.26) by the total mass to obtain

$$Pv = \left(\frac{N\hat{R}}{M} \right) T = \left(\frac{\hat{R}}{\hat{M}} \right) = RT \quad (4.30)$$

where $\hat{M} = M/N$ is the molecular weight, and $R = \hat{R}/\hat{M}$. Unlike \hat{R} , R is different for every gas. For helium, $R = (8314.3 \text{ J/kmol-K})(1 \text{ kmol}/4.0026 \text{ kg}) = 2077.2 \text{ J/kg-K}$, while for water $R = (8314.3 \text{ J/kmol-K})(1 \text{ kmol}/18.016 \text{ kg}) = 461.5 \text{ J/kg-K}$. It is often more convenient to work per unit mass rather than per mole, and so the form of the ideal gas law we will most often use is Eq. (4.30), $Pv = RT$.

It is important to bear in mind that the ideal gas law is not rigorously true, but becomes a good approximation at “sufficiently low” density. Some students reflexively invoke $Pv = RT$ if a problem involves a substance they normally think of as a gas (e.g. helium, oxygen), forgetting that these substances not only can behave as non-ideal gases, but also can be liquid or solid – it all depends on the conditions.

It is simple to check whether the ideal gas law is a suitable approximation under specified conditions. Simply compute the *compressibility factor* Z , defined by

$$Z = \frac{Pv}{RT}. \quad (4.31)$$

If the ideal gas law is valid, then Z should equal 1. The deviation from $Z = 1$ is a measure of the error made in assuming ideal-gas behavior under the specified conditions.

Example 4.4 How much error is made in using the ideal gas law to calculate the specific volume of hydrogen at 20 MPa and 90 K?

Solution: For hydrogen, $R = 8314.3/2.016 = 4124 \text{ J/kg-K}$. Using TPX, $v = v(\text{"h2"}, \text{"tp"}, 90, 20) = 0.023 \text{ m}^3/\text{kg}$ under these conditions, so $Z = (2 \times 10^7)(0.023)/4124 \times 90 = 1.24$. Therefore, the actual volume is 24% larger than would be predicted by the ideal gas law.

Internal Energy, Enthalpy, and Specific Heats

The specific internal energy of any simple compressible substance may be expressed as a function of T and v : $u(T, v)$. Alternatively, we could take the independent variables to be (T, ρ) instead of (T, v) , in which case we would write the specific internal energy as $u(T, \rho)$. Since every simple compressible substance approaches ideal gas behavior as $\rho \rightarrow 0$, the specific internal energy in the ideal-gas limit is $u(T, \rho = 0)$, which is a function of temperature alone: $u(T, 0) = u^0(T)$. We shall use a superscript “0” to denote properties in the ideal-gas, $\rho \rightarrow 0$ limit.

The specific enthalpy of an ideal gas is

$$h^0 = u^0(T) + Pv = u^0(T) + RT. \quad (4.32)$$

Therefore, the specific enthalpy of an ideal gas is also a function only of temperature (independent of pressure).

From the definitions of c_v and c_p , in the ideal gas limit

$$c_v^0(T) = \frac{du^0(T)}{dT}, \quad (4.33)$$

and

$$c_p^0(T) = \frac{dh^0(T)}{dT}. \quad (4.34)$$

Since $h^0(T) = u^0(T) + RT$,

$$\boxed{c_p^0(T) = c_v^0(T) + R.} \quad (4.35)$$

The analogous equation on a molar basis would be $\hat{c}_p^0(T) = \hat{c}_v^0(T) + \hat{R}$. Of course, Eq. (4.35) holds *only in the ideal gas limit*, where $Pv = RT$ applies.

The function $c_p^0(T)/R$ is shown in Fig. 4.12 for several gases. Since c_p^0 and c_v^0 have the same units as R , the ratio $c_p^0/R (= \hat{c}_p^0/\hat{R})$ is dimensionless. By multiplying the non-dimensional c_p^0/R by the appropriate appropriate R (or \hat{R}) value, c_p^0 or \hat{c}_p^0 may be determined in any desired unit system.

If $c_p^0(T)$ is known, Eq. (4.34) may be integrated between any two temperatures to find the change in enthalpy:

$$\boxed{h^0(T_1) = h^0(T_0) + \int_{T_0}^{T_1} c_p^0(T) dT.} \quad (4.36)$$

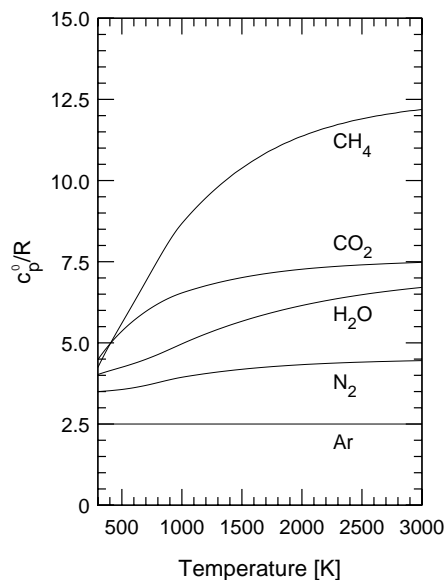


Figure 4.12: The ideal-gas specific heat $c_p^0(T)$ for several gases.

If $u^0(T_1)$ is desired, it could be computed by integrating Eq. (4.33), or from $u^0(T_1) = h^0(T_1) - RT$.

For the special case of c_p^0 independent of T , Eq. (4.36) becomes

$$h^0(T_1) = h^0(T_0) + c_p^0(T_1 - T_0), \quad (4.37)$$

and the analogous expression for u^0 is

$$u^0(T_1) = u^0(T_0) + c_v^0(T_1 - T_0). \quad (4.38)$$

For some gases, c_p^0 and c_v^0 are truly independent of temperature. This is the case for any monatomic gas, such as argon, helium, neon, etc. In other cases, it may be approximately true for a limited temperature range (e.g., for N₂ from 300-500 K or from 2000-3000 K).

Physics of the Temperature Dependence of $c_p(T)$

From Figure 4.12, it is apparent that the behavior of the function $c_p^0(T)/R$ depends on the molecular structure of the gas. Using the methods of statistical physics, it is possible to calculate this dependence exactly. We will do this later, but for now we will discuss qualitatively the physics governing the ideal-gas specific heat, and learn how to compute at least the high-temperature limit of $c_p^0(T)$ from molecular structure.

A principle of classical statistical physics known as the *principle of equipartition of energy* says that every “square term” in the classical expression for the energy of a molecule will contribute a factor of $(1/2)R$ to c_v^0 . For example, the kinetic energy of a single atom contains three square terms:

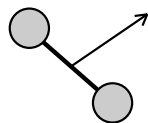
$$\epsilon = \frac{1}{2}m(v_x^2 + v_y^2 + v_z^2). \quad (4.39)$$

So $c_v^0 = (3/2)R$ and $c_p^0 = c_v^0 + R = (5/2)R$.

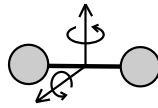
For a diatomic molecule (assuming it can be approximated as two point masses joined by a linear spring), the energy is

$$\epsilon = \underbrace{\frac{1}{2}(m_1 + m_2)(v_x^2 + v_y^2 + v_z^2)}_{\text{translation}} + \underbrace{\frac{1}{2}m_{12}r^2(\omega_y^2 + \omega_z^2)}_{\text{rotation}} + \underbrace{\frac{1}{2}m_{12}v_r^2 + \frac{1}{2}k(r - r_0)^2}_{\text{vibration}} \quad (4.40)$$

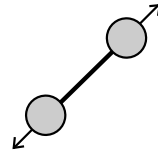
where $m_{12} = m_1 m_2 / (m_1 + m_2)$ is the *reduced mass* and $v_r = dr/dt$.



Translation



Rotation



Vibration

Center-of-mass motion (translation) contributes 3 square terms, rotation about 2 mutually orthogonal axes contributes 2 square terms, and vibration contributes 2 square terms (one for kinetic energy of vibration, and one for potential energy due to stretching the bond). The principle of equipartition of energy would predict $c_v^0/R = 1/2(3 + 2 + 2) = 7/2$, and thus $c_p^0/R = 9/2$.

This is in fact observed *at high temperatures* for diatomic gases (see the N_2 curve in Fig. 4.12). But at lower temperatures, c_p^0 is found to be less than the value predicted by the equipartition principle. The reason for this is that equipartition of energy is a *classical* principle, which is valid only if classical physics provides an acceptable description of the molecular motion.

If we had calculated the energy of the molecule using quantum mechanics, we would have found that the possible energies for translation, rotation, and vibration are quantized. According to the correspondence principle, whenever the spacing between quantum levels is very small compared to the average energy a molecule possesses, the predictions of quantum mechanics approach those of classical mechanics. An appropriate comparison would be the average translational energy per molecule, which is just $(3/2)k_B T$. Actually, we’re interested only in orders of magnitude so we can drop the $3/2$ and just compare to $k_B T$.

Mode	ϵ/k_B (K)
Translation	10^{-15}
Rotation	2.9
Vibration	3390

Table 4.3: Characteristic quantum level spacings divided by k_B , in Kelvin. The value for translation assumes a 1 cm^3 box.

Typical numbers for the energy level spacings of N_2 are shown in Table 4.3. These energies are shown divided by k_B , giving them the units of Kelvin and making the comparison to $k_B T$ easy.

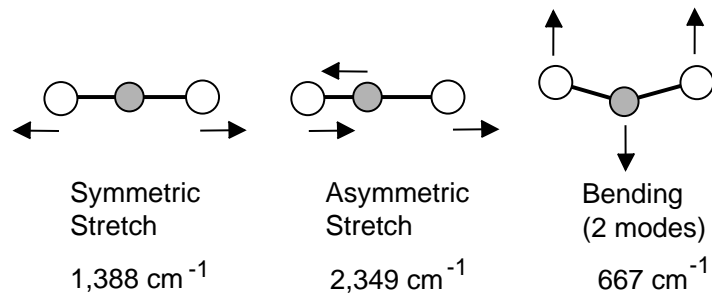
We see that a classical description of translation is essentially always valid, and for rotation it is valid except at very low temperatures, when in any case N_2 wouldn't be a gas. But vibration is another matter entirely. The spacing between vibrational levels for N_2 predicted by quantum mechanics is large compared to the average energy per molecule at room temperature. For this reason, collisions of a room-temperature N_2 molecule with others do not have enough energy to “excite” vibration, and so the molecule can't acquire vibrational energy. Higher temperatures are needed for vibration to become “fully excited.”

This explains the $c_p^0(T)$ behavior for N_2 . At room temperature, $c_p^0 \approx 7/2$. This suggests that the contribution from vibration is missing. The molecule is translating and rotating, but essentially not vibrating at all, since the lowest vibrational level is too high to be reached by collisions with other room-temperature molecules. As the temperature increases, c_p^0 approaches $9/2$ as expected, as vibration becomes excited.

High Temperature Limit for Polyatomic Molecules

Molecules containing more than 2 atoms have several different vibration modes, with different frequencies and therefore different spacings between vibrational levels. The vibrational modes become excited at different temperatures in general. As each mode becomes active, the value of c_p^0 increases by R .

The vibrational modes of CO_2 are shown below, along with the vibrational frequencies of each. (The units used are the standard spectroscopic units of wavenumbers. To convert to Hz, multiply by the speed of light.) There are actually two bending modes with the same frequency, since the molecule may bend in the plane of the paper (as shown) or out of the plane, so CO_2 has a total of four vibrational modes.



We can determine the high-temperature limit for c_p^0 for any polyatomic molecule using the equipartition principle. Suppose the molecule contains N atoms, has N_R rotational modes, and N_V vibrational modes.

If the molecule is linear (such as CO_2), it has only 2 possible rotations, just like a diatomic, so $N_R = 2$. (It is impossible to rotate about the molecular axis, since the nuclei are effectively point masses.) But a nonlinear, bent molecule (such as H_2O) has 3 possible independent rotations ($N_R = 3$).

The number of vibrational modes is given by ⁷

$$N_V = 3N - 3 - N_R. \quad (4.41)$$

From the discussion of the equipartition principle above, we see that each rotational mode contributes $(1/2)R$ to c_p^0 , while each vibrational mode contributes a full R . (Vibration has 2 square terms – one for kinetic energy, one for potential energy.) So the general formula for a polyatomic molecule is

$$\frac{c_p(T \rightarrow \infty)}{R} = \frac{5}{2} + \frac{1}{2}N_R + N_V. \quad (4.42)$$

Therefore,

$$\frac{c_p(T \rightarrow \infty)}{R} = \begin{cases} 3N - 3/2 & \text{linear} \\ 3N - 2 & \text{nonlinear} \end{cases} \quad (4.43)$$

4.7.2 The van der Waals Equation of State

In 1873, The Dutch physicist van der Waals proposed (as part of his doctoral thesis) two simple, empirical modifications to the ideal gas law, in an attempt to find a gas law valid over a wider range of conditions. The first modification was to replace v in the ideal gas law by $(v - b)$, where b is a small positive constant. This accounts for the fact that real gases are not infinitely compressible, since the

⁷It takes $3N$ numbers to specify the instantaneous configuration of the molecule (x , y , and z for each atom). Therefore, the molecule has $3N$ “degrees of freedom.” We can alternatively describe the molecular configuration by specifying its center-of-mass position (3 numbers), angular orientation (1 angle for each rotational mode), and amplitude of each vibrational mode (1 number for each mode). The sum of these must equal $3N$.

molecules have finite volume (v does not approach zero as P increases holding T constant). The constant b corresponds roughly to the specific volume of the fluid when the molecules are packed together — essentially the liquid or solid specific volume.

The second modification accounts approximately for the attractive forces between molecules. Molecules about to strike the wall feel a net restraining force due to their attraction to the other molecules in the gas behind them; this lowers the speed with which they strike the wall and the impulse delivered to the wall in the collision. Since any pressure measurement can be thought of as measuring the force on a wall due to the impulse of many collisions per second, the measured pressure P will be somewhat lower than the value which would be measured if the attractive forces were absent. The magnitude of the lowering, van der Waals argued, should be proportional to $1/v^2$, since the wall collision frequency per unit area is proportional to $1/v$, and to first approximation the net restraining force should be proportional to the number of “nearby” molecules, which also scales with $1/v$.

Based on these considerations, van der Waals proposed replacing P in the ideal gas law by $P + a/v^2$, since the measured P is lower than it “should be” in the ideal gas law by a factor a/v^2 , where a is a positive constant.

With these modifications to the ideal gas law, the van der Waals equation

$$(P + a/v^2)(v - b) = RT \quad (4.44)$$

is obtained. For large v (low density), this equation reduces to the ideal gas law, as it should.

Unlike the ideal gas equation of state, the van der Waals equation has a critical point, where $(\partial T/\partial v)_P = 0$ and $(\partial^2 T/\partial v^2)_P = 0$. It is left as an exercise to show that setting these two partial derivatives to zero results in the solution for T_c and P_c

$$RT_c = \frac{8a}{27b} \quad (4.45)$$

$$P_c = \frac{a}{27b^2}. \quad (4.46)$$

Solving for a and b in terms of T_c and P_c ,

$$a = \frac{27}{64} \frac{(RT_c)^2}{P_c} \quad (4.47)$$

$$b = \frac{RT_c}{8P_c}. \quad (4.48)$$

If we pick a and b using these expressions, the van der Waals equation is guaranteed to reproduce the correct critical temperature and pressure, although this certainly does not mean it must produce any other accurate property values.

In Fig. 4.13, the predictions of the van der Waals equation, the ideal gas equation, and TPX are compared near the critical point of methane. Three isotherms at $0.94T_c$, T_c , and $1.06T_c$ are shown, along with the actual vapor dome of methane. All three agree fairly well at large v , where the ideal gas law holds approximately. For smaller v , the ideal gas law does not adequately approximate the actual $P(v, T)$ at all.

The van der Waals equation does much better. For $T > T_c$, the shape of the isotherm is qualitatively right, although quantitatively it overpredicts $P(v, T)$ at small v . It also overpredicts the critical volume v_c slightly. An interesting thing happens below T_c : the van der Waals $P(v, T)$ is no longer monotonic in v – instead, the isotherm exhibits a local maximum and a local minimum.

Stability and the Liquid-Vapor Phase Transition

Suppose we prepare a van der Waals gas in a cylinder at $T_0 < T_c$, with v_0 chosen such that the state lies on the portion of the T_0 isotherm where $(\partial P/\partial v)_T > 0$ (Fig. 4.14). The piston weight is chosen to balance $P_0 = P(v_0, T_0)$ exerted by the gas. But due to very small random fluctuations in the number of molecules striking the piston from both sides, the piston position (and therefore v) will fluctuate very slightly.

Suppose due to a small fluctuation v increases very, very slightly. Then since $(\partial P/\partial v)_T > 0$ on this portion of the isotherm, P will *increase*. This will create a net upward force on the piston, so v will increase even more. The process will stop only when point b is reached, where the pressure again balances the piston weight.

Now consider the other case: suppose v decreases very slightly. Now the pressure in the gas *drops*, so there is a net downward force on the piston, and it falls until point a is reached.

We conclude from this thought experiment that the entire portion of the isotherm with $(\partial P/\partial v)_T > 0$ is *unstable*. A small fluctuation in v , *no matter how small*, results in the system changing state: it will either go to the low-volume state a , or the high-volume state b (both of which are stable to small fluctuations, since $(\partial P/\partial v)_T < 0$ for these states).

Thus, the van der Waals equation predicts that for a given $T < T_c$, there is a range of pressures for which there are two possible states which are stable to small fluctuations. One state has $v < v_c$, and may be regarded as the liquid

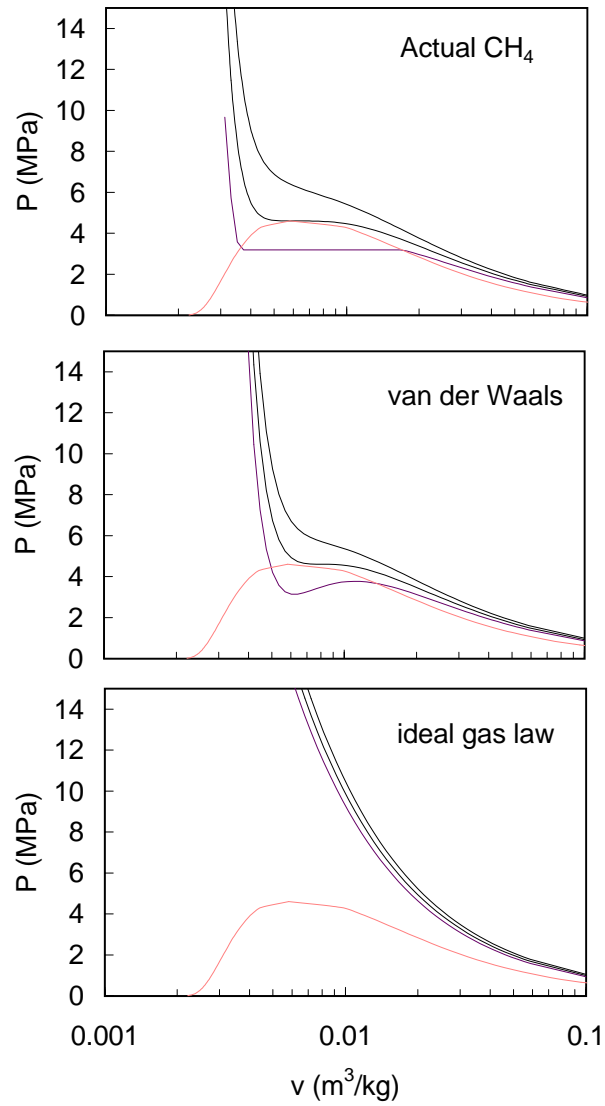


Figure 4.13: Isotherms for methane predicted by TPX, the van der Waals equation, and the ideal gas equation. The temperatures are 179.1, 190.5, and 202.0 K.

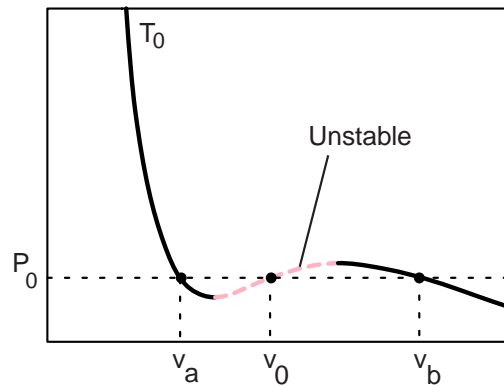
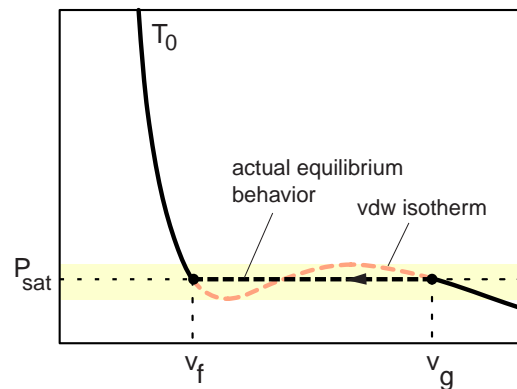
Figure 4.14: An isotherm of the van der Waals equation for $T < T_c$.

Figure 4.15: Comparison of actual behavior during isothermal compression to prediction of van der Waals equation.

state, and the other has $v > v_c$, and may be regarded as the vapor state.

In reality, of course, we don't observe isotherms which have the shape in Fig. 4.14 for $T < T_c$. Instead, at some pressure $P_{sat}(T)$ the fluid abruptly switches from the low-density (vapor) state to the high-density (liquid) state or visa versa, as shown in Fig. 4.15. The reason this happens is that although both the liquid and vapor states for pressures within the shaded region are stable to *small* perturbations, only one is stable to "big" perturbations. For $P > P_{sat}(T)$, it is the liquid state which is most stable, and for $P < P_{sat}(T)$ it is the vapor state. We'll see how to predict $P_{sat}(T)$ soon.

This qualitative behavior applies to real fluids, not only to idealized van der Waals fluids: the liquid-vapor phase transition results from an instabil-

ity of an “underlying” smooth $P(v, T)$ function. By carefully avoiding “big” perturbations (like dirt particles or scratches on the container wall where vapor bubbles can get a start) it is actually possible experimentally to prepare liquid at $P < P_{sat}(T)$. Similarly, by carefully compressing a vapor in very clean conditions pressures greater than $P_{sat}(T)$ can be reached without liquid droplets forming. Such states are called *metastable states*. Metastable states are not thermodynamic equilibrium states, since if the system is given a big enough perturbation (a dust particle, a scratch, gently shaking the container), it switches to the more stable state.

To experimentally prepare metastable states, it is actually more common to vary temperature at a given P . In this case, liquid is stable below $T_{sat}(P)$ (the boiling point), and metastable above it. Liquid heated above the boiling point without boiling is called *superheated liquid* and vapor cooled below the boiling point without condensing is called *supercooled vapor*.

A Generalized Equation of State

The van der Waals equation of state may be put in non-dimensional form by defining the nondimensional *reduced pressure* $P^* = P/P_c$ and *reduced temperature* $T^* = T/T_c$. If we write Eq. (4.44) in terms of P^* and T^* , and substitute for a and b from Eq. (4.47) and Eq. (4.48), it reduces to

$$Z^3 - \left(\frac{P^*}{8T^*} + 1 \right) Z^2 + \left(\frac{27P^*}{64T^{*2}} \right) Z - \frac{27P^{*2}}{512T^{*3}} = 0, \quad (4.49)$$

where Z is the compressibility factor Pv/RT . The roots of this cubic equation determine $Z(T^*, P^*)$. Depending on T^* and P^* , this equation will have either one root, or three. (If three, then as discussed above the middle one is unstable.)

Note that a and b do not appear in Eq. (4.49). Thus, the van der Waals equation of state reduces to a single equation for all fluids, independent of a and b , as long as T and P are expressed in reduced form. Equations of state which depend only on T^* and P^* are called *generalized equations of state*. The van der Waals equation is the simplest example of a generalized equation of state.

At the critical point ($P^* = 1, T^* = 1$) the solution to Eq. (4.49) is $Z = 3/8$. Thus, the van der Waals equation predicts that all fluids should have $Z_c = P_c v_c / RT_c = 0.375$.

Unfortunately, Z_c for real fluids is not 0.375, and differs from one fluid to the next. Most fluids have Z_c in the range from about 0.23 to about 0.33 – less than the van der Waals equation of state would predict. So clearly the van der Waals equation of state is not very accurate for real fluids. This is also clear from Fig. 4.13: the critical specific volume is overpredicted, as is the liquid specific

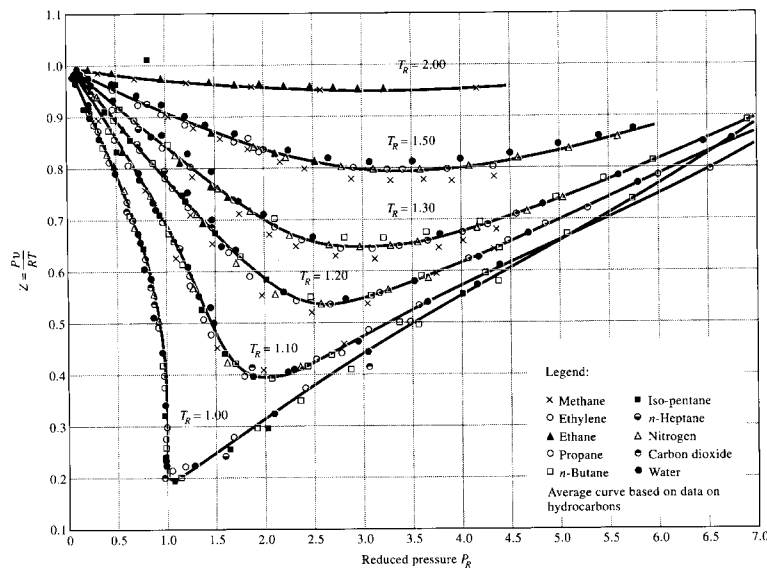


Figure 4.16: Compressibility factor Z vs. P^* (P_r) for several values of T^* (T_r).

volume at high pressure.

4.7.3 The Principle of Corresponding States

Even though the van der Waals equation is not particularly accurate, the ability to write it as a universal function $Z(T^*, P^*)$ is intriguing — maybe Z really *can* be expressed as a function of only T^* and P^* , but the function resulting from the van der Waals equation simply isn't the right one.

To test this hypothesis, we can take measured $P - v - T$ data for different fluids, calculate $Z = Pv/RT$, and the plot Z as a function of $P^* = P/P_c$ and $T^* = T/T_c$, where of course T_c and P_c differ for each fluid.

If $Z = Z(T^*, P^*)$, plotting Z in this way should collapse the experimental $P - v - T$ data onto the same set of curves for all fluids. A plot of this type is shown in Fig. 4.16, in which Z is plotted vs. P^* for T^* values from 1 to 2. It is observed that the experimental values for different fluids do tend to fall onto the same curves for a given T^* , independent of the particular fluid.

However, it is also clear that this result is only approximate. For example, on the $T^* = 1.5$ curve in Fig. 4.16, the propane data points are systematically above the mean, and the methane points systematically below. Also, we have already stated that Z_c varies slightly from one fluid to another, which could not be true if $Z = Z(T^*, P^*)$ were exactly true, since $Z_c = Z(1, 1)$.

The statement that $Z = Z(T^*, P^*)$ is known as the *Principle of Corresponding States*. As we have already seen, it an *approximate* principle, not an exact one. Several empirical fits for $Z(T^*, P^*)$ have been proposed which seek to reproduce the “average” behavior shown in Fig. 4.16. The earliest was a set of charts due to Nelson and Obert in 1954.

A more recent, popular generalized equation of state is one proposed by Lee and Kesler in 1975. The function they propose for $Z(T^*, P^*)$ for “simple” fluids is shown graphically in Fig. 4.17. It can be seen that this function corresponds fairly closely to the experimental data shown in Fig. 4.16. Simple fluids are basically ones composed of small, simple molecules, but the precise definition of a simple fluid is circular: it is one for which $Z(T^*, P^*)$ is as shown in Fig. 4.17.

Lee and Kesler (following earlier work by Pitzer) also introduced a third parameter, the so-called acentric factor ω , to allow better correlation of experimental data than can be achieved using only the parameters T^* and P^* . They observe that simple fluids have $P_{sat}(T)/P_c = 0.1$ when $T/T_c = 0.7$. To parameterize deviations from the behavior of Fig. 4.17, they define ω as

$$\omega = -\log_{10} \left(\frac{P_{sat}(0.7T_c)}{P_c} \right) - 1.0. \quad (4.50)$$

The acentric parameter is zero for a simple fluid, and usually positive for other fluids.

Lee and Kesler assume that the effects of non-zero ω can be accounted for by adding a correction term linear in ω :

$$Z = Z^{(0)}(T^*, P^*) + \omega Z^{(1)}(T^*, P^*). \quad (4.51)$$

Here $Z^{(0)}$ is the function shown in Fig. 4.17, and $Z^{(1)}$ is a correction factor to account for non-simple-fluid effects. Specifically, $Z^{(1)}$ is computed so that an accurate $Z(T^*, P^*)$ function is obtained for octane (C_8H_{18}), which is not a “simple” fluid and has $\omega = 0.3978$. Of course, Eq. (4.51) is still approximate, but it is found to be accurate to within 2% or 3% for most non-polar or slightly polar fluids. For highly-polar fluids (e.g. water) or very light ones (e.g. hydrogen, helium, or neon) for which quantum effects are important, it is less accurate.

Equation 4.51 is implemented in TPX as the function `ZLK(T*, P*, Omega)`. The `Omega` parameter is optional – if omitted, the result will be calculated for $\omega = 0$ (a simple fluid).

4.7.4 The Incompressible Substance

Finally, now that we have considered some rather complex model equations of state, we turn to the simplest possible equation of state. The compressibility

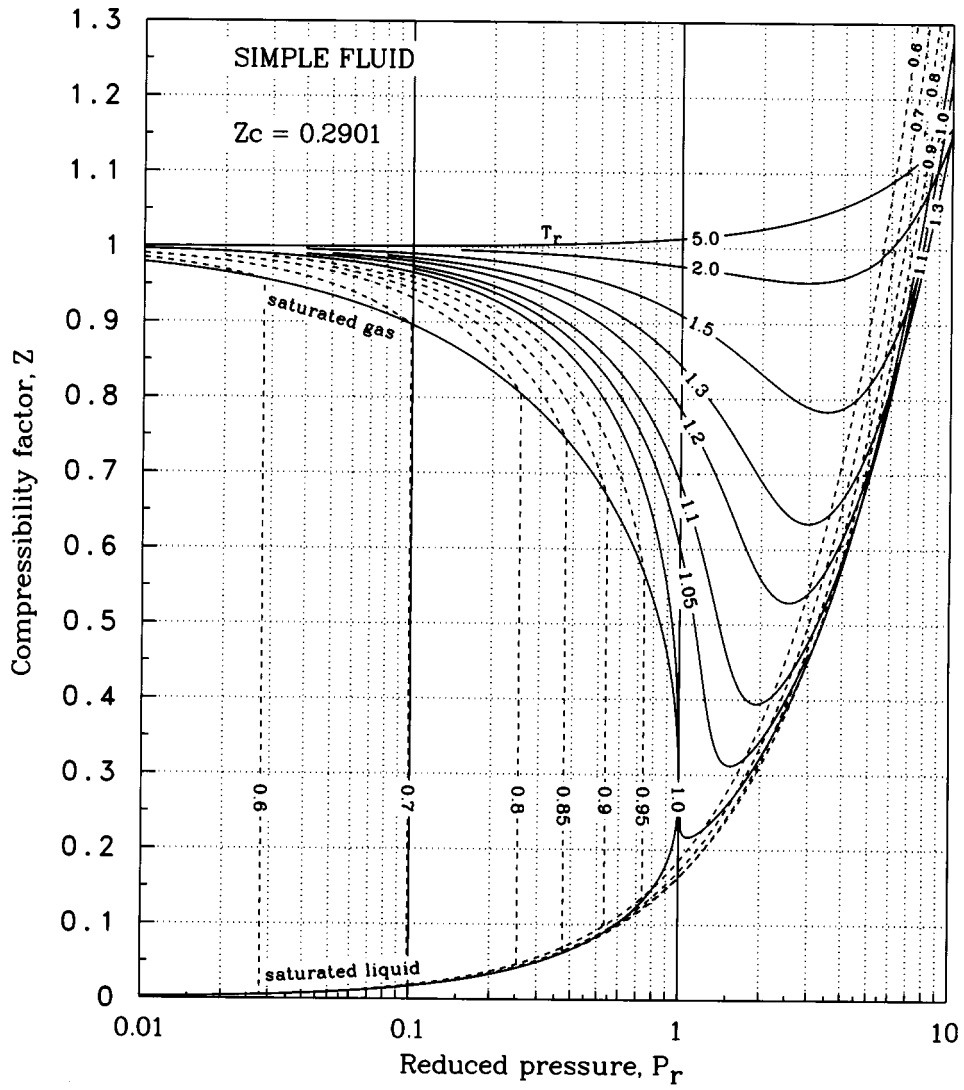


Figure 4.17: The Lee-Kesler generalized compressibility function $Z^{(0)}$ for a simple fluid.

of most liquids and solids is quite low. Unless very large pressures are applied, the specific volume hardly changes at all.

If we wish to ignore entirely the compressibility of a liquid or a solid, we can replace the real $P(v, T)$ behavior by a simple model:

$$v = v_0, \quad (4.52)$$

where v_0 is a constant. A idealized substance which obeys Eq. (4.52) is called an *incompressible substance*. It is not possible to do compression work on an incompressible substance, since by definition $dv = 0$. Therefore, the First Law for an incompressible substance is

$$du = \bar{d}Q. \quad (4.53)$$

Therefore, the only way to change u is by heat addition. This implies that an incompressible substance has only *one* degree of freedom, so $u = u(T)$.

If we choose to approximate a real liquid as incompressible in solving a particular problem, we simply neglect the small pressure dependence of v and u . A common choice is to evaluate v and u at the saturation pressure at the local temperature, and use these values no matter what the real liquid pressure is.

Note that by definition $h = u + Pv$, so h will still depend on P for an incompressible substance, even though u does not. Therefore, for a process in which T and P change, $\Delta h = \Delta u + v\Delta P$. For most liquids and solids, v is small enough that the $v\Delta P$ term is small compared to Δu .

Note also that $(\partial h/\partial T)_P = du/dT$ in this case, so

$$c_p(T) = c_v(T) \quad (4.54)$$

for an incompressible substance. This relationship is very different than the one for ideal gases [Eq. (4.35)].

Problems

4.1 Some properties of ice, liquid water, and water vapor at the triple point $T = 273.16$ K are given below.

Phase	Density (kg/m ³)	Specific enthalpy (kJ/kg)
Ice	917.0	-333.5
Liquid	999.8	0.0
Vapor	4.84×10^{-3}	2501.4

The enthalpy values are relative to the liquid enthalpy.

A closed, *constant volume* 1 liter container initially is at a temperature infinitesimally below 273.16 K. It holds solid and vapor in equilibrium, with 80% solid by mass. Heat is now added until equal masses of liquid and solid are present. Write down equations expressing energy, mass, and volume balance. Determine numerically how much heat must be added.

- 4.2** The air pressure in aircraft cabins is kept lower than sea level atmospheric pressure, since otherwise the pressure force on the airframe would be too great at high altitude. A typical value for a transatlantic flight would be a cabin pressure equivalent to atmospheric pressure at an elevation of 8,000 ft. On such a flight, to what temperature should the flight attendants heat water for it to boil? How much heat must be added to boil a unit mass of water, and how does it compare to the heat needed at sea level? Make any reasonable assumptions you need for the analysis, but state what you are assuming.
- 4.3** Five kg of methane are contained in a closed, 150 liter container. If the pressure is 1 MPa, determine:
1. Whether any liquid is present, and if so how many kg
 2. The temperature
 3. The internal energy U of the methane (J)
- 4.4** A particular substance has an isothermal compressibility $\kappa = aT^4/P^3$ and a thermal expansion coefficient $\beta = bT^3/P^2$. Determine the equation of state $v(T, P)$ to within an additive constant and the ratio a/b .
- 4.5** Ten kg of saturated nitrogen vapor at 90 K is heated at constant pressure until its volume is 3 m³. Determine
1. The pressure
 2. The final temperature
 3. The heat added
 4. The work done by the nitrogen on the environment.
- 4.6** Using TPX, plot the compressibility factor Z for hydrogen and for oxygen at 300 K vs. pressure. For hydrogen, plot Z over the pressure range of 1 atm to 100 MPa, and for oxygen over the pressure range of 0.02 MPa to 20 MPa. Use a logarithmic scale for pressure.

For a particular gas storage tank design, it is necessary to estimate to within 1% accuracy the tank volume required for a given mass of gas and pressure at $T = 300$ K. For both hydrogen and oxygen, determine the pressure below which use of the ideal gas equation provides acceptable accuracy.

- 4.7** Estimate the temperature water at the bottom of a 500 ft deep lake would have to be heated before it begins to boil.
- 4.8** The specific heat at constant volume c_v for insulating solids at low temperature is given by the equation

$$\frac{c_v}{R} = \left(\frac{12\pi^4}{5} \right) \left(\frac{T}{\theta} \right)^3,$$

which is known as the Debye T^3 law. The constant θ is known as the *Debye temperature*, and is a characteristic of the material in question.

For diamond, $\theta = 2200$ K. Treating diamond as incompressible, how much heat in Joules must be added to raise the temperature of 1 kg of diamond from 1 K to 50 K?

- 4.9** Show that for any simple compressible substance

$$\left(\frac{\partial P}{\partial T} \right)_v = \frac{\beta}{\kappa}$$

and

$$\left(\frac{\partial \beta}{\partial P} \right)_T = - \left(\frac{\partial \kappa}{\partial T} \right)_P$$

- 4.10** Derive the results for a van der Waals gas

$$RT_c = \frac{8a}{27b}$$

and

$$P_c = \frac{a}{27b^2}.$$

- 4.11** A particular substance is found to obey the Dieterici equation of state

$$P(v - b) \exp(a/vRT) = RT.$$

1. Derive expressions for the properties at the critical point P_c , v_c , and T_c in terms of the constants a , b , and R .

2. Compare the values for $Z_c = P_c v_c / RT_c$ predicted by this equation with the values from Table 4.2 or calculated using TPX for hydrogen, carbon dioxide, and water.
- 4.12** A particular fluid has critical state parameters $T_c = 300$ K and $P_c = 6$ MPa, and a molecular weight of 30. Estimate its specific volume at $T = 330$ K and $P = 12$ MPa using
1. the van der Waals equation of state
 2. the Lee-Kesler generalized compressibility function, assuming it is a simple fluid.
- 4.13** Determine the high-temperature limit for c_p for Ar, N₂, H₂O, CO₂, and CH₄, and compare to the results in Figure 4.12. What is the value of c_p (kJ/kg/K) in the high-temperature limit for C₆₀?
- 4.14** Calculate the acentric parameter ω for N₂, CH₄, and CF₄H₂ (“HFC134a”) using TPX. Compare the predictions for Z at $T^* = 1.2$, $P^* = 1.5$ using
1. TPX
 2. Lee-Kesler with $\omega = 0$
 3. Lee-Kesler with the actual ω .

Index

- adiabatic, 35
- Boltzmann's constant, 87
- center of mass, 20
- chemical potential, 54
- coexistence lines, 77
- compressibility factor, 88
- conservative force, 19
 - definition, 20
- control mass, 4
- control volume, 4
- Corresponding States
 - Principle of, 100
- critical point, 73
- Curie substance, 43
- datum state, 71
- degrees of freedom, 63
- dielectric, 42
- electric dipole moment, 40
- electric susceptibility, 41
- energy, 3
 - accounting, 9
 - balance, 9
 - conservation of, 24
- enthalpy
 - definition, 70
- entropy, 3
- environment, 3
- equations of state, 65
- equilibrium, 50
 - chemical, 55
 - diffusive, 53
 - mechanical, 52
 - phase, 54
 - restricted, 55
 - thermal, 52
- equipartition of energy, 91
- extensive, 62
- First Law of Thermodynamics, 34
- force
 - centrifugal, 17
 - Lorentz, 17, 21
 - traction, 17
- gas constant, 86
- generalized displacements, 44
- generalized equations of state, 98
- generalized forces, 44
- heat capacity, 86
- heat transfer, 32
 - rate, 35
- hydrostatics, 57
- ideal gas, 87
- incompressible substance, 102
- inertial mass, 7
- intensive, 62
- internal energy, 28
- isobar, 78
- isothermal compressibility, 84
- kinetic energy, 8
- Lee-Kesler generalized compressibility, 100
- lever rule, 80

- liquid
 - compressed, 79
 - subcooled, 79
- magnetization, 43
- manometer, 57
- mass density, 20
- phase, 54
- phase diagrams, 72
- polar, 40
- polarization, 41
- potential energy
 - definition, 23
 - gravitational, 19
- power, 12
- pressure, 33
 - critical, 72
 - reduced, 98
- quality, 79
- quasi-static, 37
- reference state, 71
- saturation, 78
- saturation pressure, 77
- saturation temperature, 77
- specific, 62
- specific heat
 - c_p , 85
 - c_v , 85
- sublimation, 73
- supercooled vapor, 98
- superheated liquid, 98
- surface tension, 39
- system, 3
 - classification, 4
- temperature, 53
 - critical, 73
 - reduced, 98
 - thermodynamic, 61
- thermal energy, 1
- thermal expansion coefficient, 83
- thermodynamic state, 63
- triple line, 75
- triple point, 75
- van der Waals equation of state, 94
- vapor pressure, 77
- work, 8, 10
 - definition of, 10

CHAPTER 5

THE FIRST LAW FOR OPEN SYSTEMS

5.1 Introduction

An open system is one which matter may enter or leave. An uncovered cup of coffee is an open system, since water evaporates from it. Open systems are very common in engineering thermodynamics. A simple example is a jet engine, which takes in air and fuel, mixes and burns them, and then exhausts the hot combustion products at high speed to produce thrust. Design of an efficient jet engine requires a careful thermodynamic analysis, which we will discuss in more detail soon.

In this chapter, we develop one of the major tools needed to carry out such an analysis — an expression of the First Law of Thermodynamics suitable for open systems. Following this development, we will apply it to examine the characteristics of some very useful devices for changing the state of a flowing fluid, and take a first look at how such simple devices may be combined to do useful tasks, such as generate electrical power.

5.2 Some Approximations Involving Flowing Fluids

In most cases of interest to us, the matter which enters or leaves an open system is a fluid. Problems involving flowing fluids are more complicated than the problems we have considered up to now, since strictly speaking a flowing fluid is not in equilibrium — unbalanced forces are causing it to move. Depending on the particular features of the flow, it may be turbulent, and properties such as temperature, pressure, velocity, and density may vary from point to point within the fluid. To make the analysis of problems with flowing fluids tractable, we need to make some approximations.

5.2.1 Equilibrium Properties

What property relationships should we use to determine the properties of a flowing fluid? We now know a great deal about the properties of fluids in *equilibrium* (at rest). But if the velocity varies from point-to-point in a fluid, as it does in all real viscous fluids, then there is no reference frame in which it

appears stationary.

Fortunately, for most situations of engineering interest the relationships among properties in a flowing fluid are indistinguishable from the equilibrium equations of state we discussed in the last chapter, as long as the thermodynamic properties are interpreted as “local” values at a point in a fluid. When this is true, we say the flow is in *local thermodynamic equilibrium*, sometimes abbreviated LTE.

For example, a very low density flowing gas will usually still obey $Pv = RT$ to a very good approximation at every point in the gas, even if P , v , and T differ from point to point, as long as the P , v , and T values are all measured at the *same* point. Similarly, when LTE holds, the local specific internal energy u will have the same dependence on the local temperature as if the gas were truly in equilibrium.

There *can* be some extreme situations where LTE does not hold and equilibrium equations of state break down entirely. For example, in a shock wave the properties of a gas change dramatically over a distance comparable to a molecular mean free path, which in air at 1 atm is about $0.15 \mu\text{m}$. When properties change significantly over a distance of a mean free path, then the equilibrium property relationships no longer hold, and properties such as pressure and temperature may not even be meaningful. However, the properties of the gas just a few mean free paths on either side of the shock wave would be found to be well-described by equilibrium equations of state. Fortunately, we will not need to evaluate any gas properties right at a shockwave.

Another type of system where non-LTE behavior is often found is a plasma, which is simply a gaseous mixture of ions and electrons (and often neutral atoms or molecules too). Plasmas occur naturally in the upper atmosphere and in outer space, and also can be created in the laboratory. Plasmas have been investigated actively since the 1950's as a means to achieve the conditions necessary for nuclear fusion. Smaller scale plasmas are also widely used in industry, for example to etch patterns in semiconductors, to clean surfaces, and to deposit thin coatings. Plasmas can have strange behavior – often the electrons have much more kinetic energy on average than do the heavy ions or atoms. In plasmas used in the semiconductor industry, the heavy particles are at room temperature, while the electrons behave as if they had a much higher temperature, often greater than 10,000 K. Clearly a gas which behaves as if different species have different temperatures is not close to equilibrium, and equilibrium gas properties could not predict the correct properties of such a plasma.

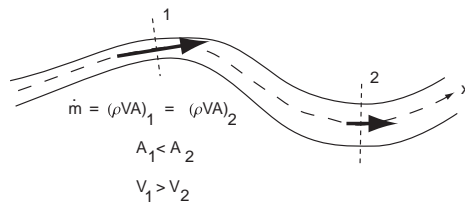


Figure 5.1: Approximate one-dimensional flow of a river.

Another situation where deviations from equilibrium equations of state can be important is if the fluid properties are changing too rapidly for the fluid to adjust. For example, if the pressure in a liquid is rapidly reduced, it may drop below $P_{sat}(T)$ without vapor forming, since formation of vapor bubbles is not instantaneous. When $P_{sat} - P$ reaches some critical value, vapor bubbles will form explosively, which is known as *cavitation*. Some “fluids” like glasses and polymer melts can take a very long time to re-establish equilibrium when perturbed, and thus often exhibit non-equilibrium properties.

Situations like these in which the fluid properties differ significantly from the equilibrium properties are relatively rare. In this book, we will usually assume that equilibrium equations of state may also be used to compute the properties of fluids slightly out of equilibrium (due, for example, to flow). We will call this the assumption of *equilibrium properties*.

5.2.2 One-Dimensional Flow

Another often-useful assumption is the assumption of *one-dimensional flow*. Contrary to what its name seems to imply, it does not mean the molecules follow one after another along a mathematical line. Instead, it only means that the flow velocity is assumed to depend on only a single co-ordinate (not two or three, which would define two-dimensional or three-dimensional flow, respectively).

The one-dimensional flow assumption is easily understood by considering the flow of a river (Fig. 5.1). A detailed description of a river flow would be quite complicated, since it would have to describe in detail the turbulent flow around over and around rocks, etc. But on a more global scale, a river

flow is simple: the water flows basically downstream. When we make the one-dimensional flow assumption, we ignore the details of the actual fluid flow and focus on the principal direction the fluid is moving. We approximate the flow velocity as being the same everywhere on a cross-section normal to the local downstream direction. Essentially, we are assuming that all fluid “packets” at a given downstream location are moving with the *average* velocity of the real three-dimensional flow at that downstream location.

One-dimensional flow does not necessarily mean the velocity is constant. In a river, the flow is rapid where the river is narrow, and slow where the river is wide. What is constant is the water *mass flow rate* \dot{m} (kg/s). The mass flow rate of water is

$$\dot{m}_w = \rho_w V(x)A(x), \quad (5.1)$$

where x is the distance downstream, ρ_w is the water density (constant), $V(x)$ is the one-dimensional flow speed at x , and $A(x)$ is the cross sectional area of the river at x .

Therefore, for a river with constant \dot{m} , the product $V(x)A(x)$ is constant. In practice, Eq. (5.1) would be the equation used to calculate $V(x)$, and so serves to define the one-dimensional flow speed.

We will be interested in many situations in which a gas or vapor flows through a tube of variable cross-sectional area $A(x)$. As long as the changes in area are gradual, the flow down the tube is approximately one-dimensional. The gas mass flow rate through the tube is given by the same equation as for the river flow, except that we should allow for the possibility that the gas density may also depend on x :

$$\dot{m} = \rho(x)V(x)A(x). \quad (5.2)$$

The density would depend on x if, for example, the gas heats up or cools down as it flows down the tube (perhaps the tube walls are heated or cooled). As we’ll see later, the density can also vary greatly if the flow speed is comparable to or greater than the speed of sound and the tube area changes. Of course, Eq. (5.2) also applies to liquids, for which $\rho(x) \approx \text{constant}$.

5.2.3 Steady Flow

Equation 5.2 always holds in one-dimensional flow, even if \dot{m} is time-dependent. In this case, the flow is *unsteady*. If all quantities appearing in Eq. (5.2) do *not* depend on time, then the flow in the tube is by definition *steady*.

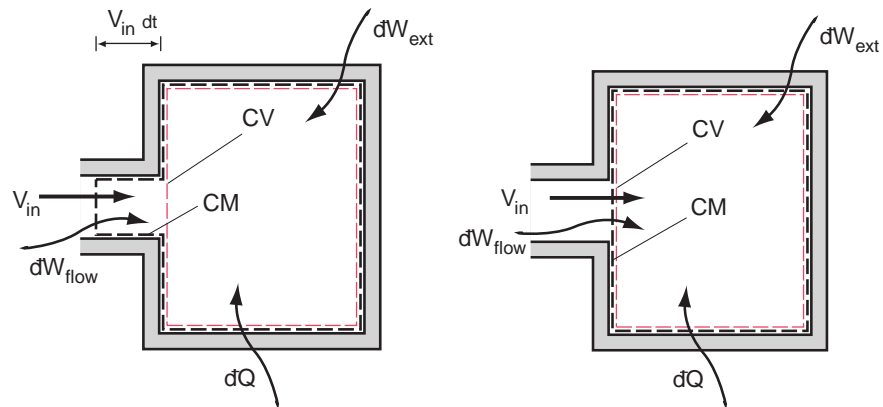


Figure 5.2: Control mass and control volumes for analysis of a system with fluid entering through an inlet pipe. (a) time t ; (b) time $t + dt$.

5.3 Energy Accounting for an Open System

In order to develop the First Law for an open system, consider the simple situation shown in Fig. 5.2. A container is connected to an inlet tube, and fluid (either liquid or gas, it doesn't matter) is flowing in. There is no outlet, so fluid builds up in the container.

The flow into the container is assumed to be one-dimensional, with speed V_{in} .

Since we know how to write the First Law for closed systems, let's begin by analyzing the control mass system labeled CM. Suppose we start the analysis at time t and allow an infinitesimal time dt to elapse, during which energy flows from the environment to the system in the form of heat dQ and work dW .

The control mass boundary at time t in the inlet pipe is chosen to include the fluid which flows into the container in time dt . At time $t + dt$, this fluid has moved into the container, so the control mass boundary has moved flush with the end of the inlet pipe.

The work done on the system in time dt contains two contributions, which we will call dW_{ext} and dW_{flow} . The first is the work done by external forces (such as a stirrer, or electrical work from a battery, or any other type of work). The second contribution is the force on the moving system boundary in the inlet pipe. Remember any time a system boundary moves by $d\mathbf{x}$ under action of an external force \mathbf{F} , work in the amount $\mathbf{F} \cdot d\mathbf{x}$ is done by the external force on the system.

In the present case, the external force is provided by the fluid in the inlet

pipe just outside the control mass, pushing on the fluid just inside the control mass. The flow work done in time dt is

$$\bar{d}W_{flow} = F dx = (P_{in} A_{in}) dx = P_{in} A_{in} V_{in} dt. \quad (5.3)$$

This expression may be re-written in terms of the mass flow rate \dot{m}_{in} in the inlet pipe. The mass flow rate (kg/s) is

$$\dot{m}_{in} = \rho_{in} V_{in} A_{in}. \quad (5.4)$$

Therefore, in terms of \dot{m}_{in} ,

$$\bar{d}W_{flow} = \dot{m}_{in} (P/\rho)_{in} dt = (\dot{m} P v)_{in} dt. \quad (5.5)$$

The first law for system CM is:

$$\begin{aligned} dE_{CM} &= \bar{d}Q + \bar{d}W \\ &= \bar{d}Q + \bar{d}W_{ext} + (\dot{m} P v)_{in} dt. \end{aligned} \quad (5.6)$$

The term E_{CM} includes all energy in the control mass, whether internal, kinetic or potential:

$$E_{CM} = (U + E_k + E_p)_{CM}. \quad (5.7)$$

Now consider the control volume CV, which consists only of the container. At time t , CV does not contain the fluid in CM which is still in the inlet tube. Therefore,

$$E_{CM}(t) = E_{CV}(t) + (\dot{m}_{in} dt)(e_{in}) \quad (5.8)$$

where e_{in} is the total specific energy of the fluid in the inlet:

$$e_{in} = u_{in} + e_{k,in} + e_{p,in}. \quad (5.9)$$

If the inlet flow is one-dimensional and if the potential energy is due to gravitation then

$$e_{in} = u_{in} + \frac{V_{in}^2}{2} + g y_{in}, \quad (5.10)$$

where y_{in} is the elevation of the inlet tube.

At time $t + dt$, CM and CV coincide, so

$$E_{CM}(t + dt) = E_{CV}(t + dt). \quad (5.11)$$

Subtracting Eq. (5.8) from Eq. (5.11), we find

$$dE_{CM} = dE_{CV} - (\dot{m}_{in} dt)(e_{in}). \quad (5.12)$$

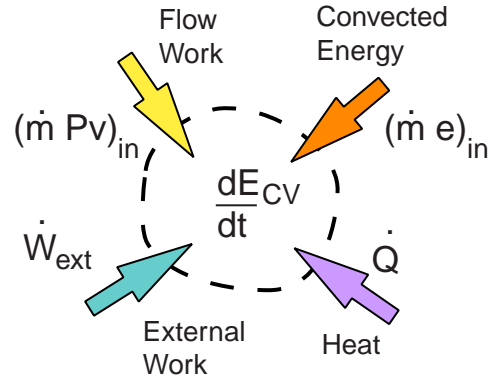


Figure 5.3: Energy accounting for a control volume with fluid entering through one inlet.

Putting this into Eq. (5.6) yields

$$dE_{CV} - (\dot{m}_{in} dt)(e_{in})E = dQ + dW_{ext} + (\dot{m}Pv)_{in} dt. \quad (5.13)$$

Rearranging, we have

$$dE_{CV} = dQ + dW_{ext} + [\dot{m}(e + Pv)]_{in} dt. \quad (5.14)$$

We may divide this equation by the elapsed time dt to express this equation on a rate basis (per unit time):

$$\frac{d}{dt}E_{CV} = \dot{Q} + \dot{W}_{ext} + [\dot{m}(e + Pv)]_{in}, \quad (5.15)$$

where

$$\dot{Q} = \frac{dQ}{dt} \quad (5.16)$$

is the rate at which heat is added to CV, and

$$\dot{W}_{ext} = \frac{dW_{ext}}{dt} \quad (5.17)$$

is the rate at which external work is done on CV. Note that dQ/dt should be interpreted as a fraction (dQ divided by dt), and not as a derivative; the same holds for dW/dt .

Both Eq. (5.14) and Eq. (5.15) are statements of the First Law for an open system with one fluid inlet. They may be interpreted as shown in Fig. 5.3: the rate of increase of energy stored in the control volume (dE_{CV}/dt) is equal to the sum of the rates of energy transfer into the system from the environment. The energy transfer terms are due to

1. external work: \dot{W}_{ext}
2. heat transfer: \dot{Q}
3. flow work required to push the fluid into the control volume: $(\dot{m}Pv)_{in}$
4. energy which is carried in (or *convected in*) by the entering fluid: $(\dot{m}e)_{in}$.

Equation (5.15) may be easily generalized to handle an arbitrary control volume with multiple fluid inlets and multiple fluid outlets. For the i^{th} inlet, there is a term $[\dot{m}(e + Pv)]_i$ in the energy balance to account for the energy convected in with the fluid, and the flow work done by the environment to push it in. For outlets, the analysis is the same as for inlets except $V < 0$. This simply changes the sign: for the o^{th} outlet, a term $-[\dot{m}(e + Pv)]_o$ must be added. Therefore, the general statement of the First Law for an open system is

$$\frac{d}{dt}E_{CV} = \dot{Q} + \dot{W}_{ext} + \sum_{inlets} [\dot{m}(e + Pv)]_i - \sum_{outlets} [\dot{m}(e + Pv)]_o \quad (5.18)$$

Note that \dot{m} , e , P , and v may differ for each inlet and outlet, and may differ from the state of the fluid inside the control volume. Also, all quantities in this equation may be time-dependent.

We have not said anything about the state of the matter inside the control volume. It might be near equilibrium, or might be very far from it, with shock waves, plasmas, chemical reactions, or other complex phenomena occurring. Of course, if we wish to relate E_{CV} to other properties (e.g. pressure or temperature), we would need to know more about just what is happening inside CV.

We do need to evaluate $(e + Pv)$ for the fluid in each inlet and in each outlet. Typically, we will make the equilibrium properties assumption for the inlet and outlet streams to allow evaluating $e + Pv$.

Since $e = u + e_k + e_p$,

$$\begin{aligned} e + Pv &= (u + Pv) + e_k + e_p \\ &= h + e_k + e_p. \end{aligned} \quad (5.19)$$

Therefore, $e + Pv$ is simply the specific enthalpy h plus any macroscopic kinetic and potential energy per unit mass. For the common case of a one-dimensional

flow where the only potential energy is gravitational,

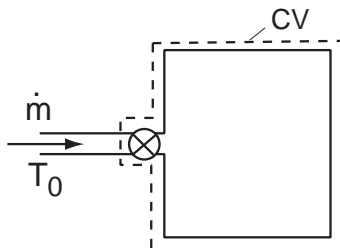
$$e + Pv = h + \frac{|\mathbf{V}|^2}{2} + gy \quad (5.20)$$

Another way to write Eq. (5.18) for this case is

$$\frac{d}{dt}E_{CV} = \dot{Q} + \dot{W}_{ext} + \sum_{inlets} \left[\dot{m} \left(h + \frac{|\mathbf{V}|^2}{2} + gy \right) \right]_i - \sum_{outlets} \left[\dot{m} \left(h + \frac{|\mathbf{V}|^2}{2} + gy \right) \right]_o \quad (5.21)$$

Example 5.1

An insulated tank is to be pressurized with air. A high-pressure air line is connected to the tank inlet through a flow-regulating valve which produces a constant mass flow of air \dot{m} . The temperature of the air in the high-pressure line is the ambient temperature T_0 . Assuming that air may be idealized as an ideal gas with constant specific heats, determine the temperature in the tank as a function of time.



Solution: Since mass is entering the tank, this problem is most easily solved by defining a control volume and using the First Law in the form of Eq. (5.18). The surface of the control volume should always be placed where we know some information. Since we don't know the state of the air at the outlet of the valve, but we do know the state at the valve inlet, let the control surface cut through the inlet before the valve.

We will assume the following:

1. One-dimensional, steady flow in air line
2. Air inside tank is uniform, with negligible kinetic energy
3. Potential energy may be neglected
4. Kinetic energy in air line may be neglected

5. Equilibrium properties in air line and in tank
6. Adiabatic process ($\dot{Q} = 0$)
7. Ideal gas with constant c_v

For this problem, Eq. (5.18) becomes

$$\frac{d}{dt}E_{CV} = [\dot{m}(e + Pv)]_{in}. \quad (5.22)$$

Note that $\dot{W}_{ext} = 0$, since there is no external work in this problem.

Since we are assuming kinetic and potential energies are negligible, $E_{CV} = U_{CV}$ and $(e + Pv)_{in} = h_{in}$. Therefore,

$$\frac{d}{dt}U_{CV} = [\dot{m}h]_{in}. \quad (5.23)$$

Since the flow is steady, the right-hand side of this equation is a constant, and therefore it is easily integrated:

$$U_{CV}(t) = U_{CV}(0) + \dot{m}h_{in}t. \quad (5.24)$$

We can also write a mass balance on the control volume:

$$M_{CV}(t) = M_{CV}(0) + \dot{m}t \quad (5.25)$$

The rate of these two equations gives the specific internal energy of the air in the tank as a function of time:

$$u_{CV}(t) = \frac{U_{CV}(t)}{M_{CV}(t)} = \frac{U_{CV}(0) + \dot{m}h_{in}t}{M_{CV}(0) + \dot{m}t}. \quad (5.26)$$

At long times, the terms linear in t will dominate over the constant terms, so

$$\lim_{t \rightarrow \infty} u_{CV}(t) = h_{in}. \quad (5.27)$$

For an ideal gas with constant c_v , we may write $u = c_v T$ and $h = c_p T$, taking the reference state at $T = 0$. Then

$$\lim_{t \rightarrow \infty} T_{CV}(t) = \frac{c_p}{c_v} T_{in} \quad (5.28)$$

Thus, the air in the tank ends up *hotter* than the air in the high-pressure line. The physical reason for this is that to put more air in the tank, the air already there has to be compressed to make room. We've seen before that adiabatic compression of a gas causes it to heat up, due to the compression work done on it. This is one more example of this effect.

An interesting feature of this problem is that the temperature in the tank depends only on T_{in} and c_p/c_v , and not on the final pressure. For air, $c_p/c_v \approx 1.4$, so if $T_{in} = 300$ K, the final temperature in the tank is 420 K.

5.4 Steady State

If the energy and mass contained within the control volume are not changing with time, we say the control volume is in *steady state*. In this case, $dE_{CV}/dt = 0$, so Eq. (5.18) becomes

$$\dot{Q} + \dot{W}_{ext} + \sum_{inlets} [\dot{m}(e + Pv)]_i = \sum_{outlets} [\dot{m}(e + Pv)]_o. \quad (5.29)$$

This equation simply states that at steady state the energy inflow rate to the control volume (the left-hand side) must equal the energy outflow rate (the right-hand side).

Since no mass can be accumulating inside the control volume at steady state, we also require

$$\sum_{inlets} \dot{m}_i = \sum_{outlets} \dot{m}_o. \quad (5.30)$$

It is important to remember the difference between the assumption of steady flow and the assumption of steady state. Steady flow applies to an individual inlet or outlet stream, and means that \dot{m} and the fluid properties of the stream are constant. Steady state applies to the entire control volume, and means that energy and mass inflows and outflows balance, so that there is no net change in the energy or mass contained in the control volume. In the last example, the inlet flow was a steady flow, but the system was not in steady state, since there was no outlet and both mass and energy continued to accumulate in the tank.

The steady state assumption is very useful to analyze many real engineering systems, including most power plants, chemical plants, jet engines, rocket motors, refrigeration systems, etc. Although they are not always *truly* in steady-state (for example, the thrust of an aircraft engine changes from take-off to cruise), the transient terms in the energy and mass balances (dE_{CV}/dt and dM_{CV}/dt , respectively) are usually small enough compared to the other terms that they may be neglected, and the problem treated as a steady state one.

Most of these complex engineering systems are constructed from a set of simple steady-flow, steady-state devices which alter the state of a flowing fluid in some way, and may exchange energy as heat or work with the environment. We will introduce some of the major ones in the next section. Analyzing these simple systems shows how the First Law for open systems is used, and it allows us to begin assembling a “toolkit” of devices which we will use to build more complex systems in later chapters.

5.5 Some Steady-Flow, Steady-State Devices

5.5.1 Nozzles

A *nozzle* is a device which accelerates a fluid flowing steadily through it. For liquids or subsonic gases, a nozzle simply consists of a converging tube (Fig. 5.4).¹

To first approximation, nozzles are often idealized as adiabatic. In a real nozzle, there may be some heat transfer to or from the environment if the fluid is hotter or colder than the surroundings. But often the fluid flows through the nozzle so rapidly that there is little time for a significant amount of heat transfer to occur.

We will make the following assumptions to analyze flow through an adiabatic nozzle:

1. One-dimensional, steady flow at inlet and outlet (states 1 and 2)
2. Equilibrium properties at states 1 and 2
3. $\dot{Q} = 0$
4. Negligible change in potential energy
5. Steady state

With these assumptions, the energy balance is

$$\left[\dot{m} \left(h + \frac{V^2}{2} \right) \right]_1 = \left[\dot{m} \left(h + \frac{V^2}{2} \right) \right]_2. \quad (5.31)$$

Because of the steady-state assumption, $\dot{m}_1 = \dot{m}_2$, so

$$h_1 + \frac{V_1^2}{2} = h_2 + \frac{V_2^2}{2}. \quad (5.32)$$

Since the fluid accelerates through a nozzle, the pressure must decrease in the direction of flow. Thus, $P_2 < P_1$. The action of an adiabatic, steady nozzle on the thermodynamic state of the fluid is as shown in the $P-h$ plot in Fig. 5.4: both the pressure and the enthalpy are lower at the outlet than at the inlet. The horizontal distance $h_1 - h_2$ between the inlet and outlet state points equals the increase in kinetic energy.

The line connecting states 1 and 2 is shown as dashed in Fig. 5.4. A dashed, straight line is used simply to indicate that state 1 is transformed to state 2 by

¹A nozzle for a *supersonic* gas consists of a diverging tube, as we'll discuss in more detail later. To accelerate a gas from subsonic to supersonic speed, a converging-diverging tube is required (an hourglass shape).

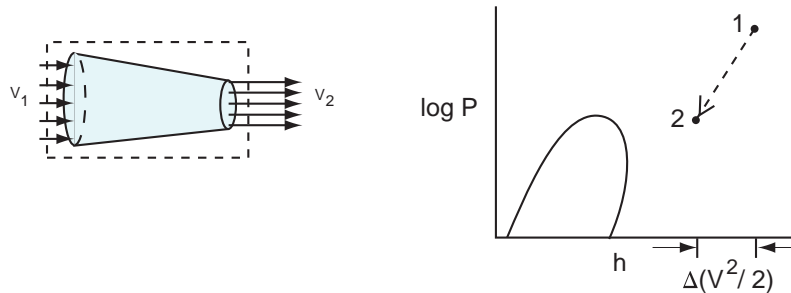


Figure 5.4: An adiabatic nozzle.

the nozzle. It does not imply that the fluid state really followed a straight-line trajectory in (P, h) in going from state 1 to state 2. In fact, we have only made the equilibrium properties assumption for states 1 and 2. Inside the nozzle, the fluid might not be describable by *any* equilibrium state – we don't know.

Of course, the positions of state 1 and state 2 do not need to be as shown in this diagram. States 1 and 2 could be located anywhere, including in the liquid or two-phase regions. The only requirement is that $P_2 < P_1$ and $h_2 < h_1$.

Example 5.2 Nitrogen at 300 K and 2 atm flows at 1 m/s into an adiabatic nozzle which exhausts into the ambient air at 1 atm. If the exit speed is 300 m/s, what is the temperature at the exit?

Solution: Using TPX,

$$h_1 = h(\text{"N2"}, \text{"tp"}, 300, 2) = 4.613 \times 10^5 \text{ J/kg.}$$

The initial specific kinetic energy $V_1^2/2 = 0.5 \text{ J/kg}$, and the final specific kinetic energy $V_2^2/2 = 4.5 \times 10^4 \text{ J/kg}$. Therefore,

$$h_2 = 4.613 \times 10^5 + (0.5 - 4.5 \times 10^4) = 4.163 \times 10^5 \text{ J/kg.}$$

Since P_2 is given, P_2 and h_2 fix the exit state. From TPX,

$$T_2 = \text{Temp}(\text{"N2"}, \text{"PH"}, 1, 4.163\text{E}5) = 256.57 \text{ K.}$$

Note that in the last example the initial kinetic energy of the gas is very small compared to the final kinetic energy, and thus $\Delta(V^2/2) \approx V_2^2/2$. When this is the case, Eq. (5.32) reduces to $V_2 = \sqrt{2(h_1 - h_2)}$. We will often assume that the initial kinetic energy is negligible when analyzing flow through nozzles.

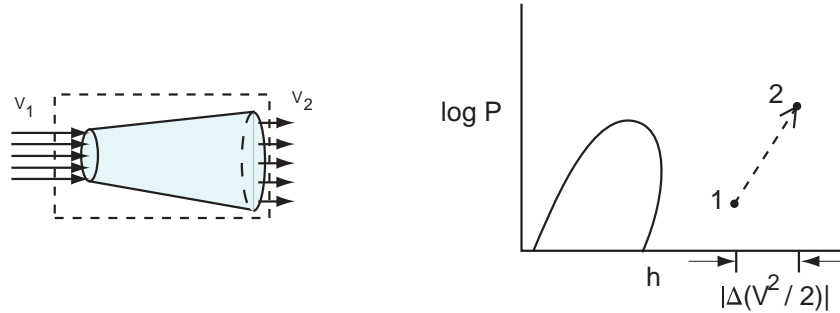


Figure 5.5: An adiabatic diffuser.

5.5.2 Diffusers

A *diffuser* is in a sense the opposite of a nozzle. In a diffuser, a flow decelerates, with a corresponding pressure rise. (The pressure *must* rise, since fluid elements will only decelerate if they experience a net force directed opposite to the flow direction.) For liquids and subsonic gases, a diffuser simply consists of a diverging tube (Fig. 5.5). The energy balance for a steady-state adiabatic diffuser is the same as that for an adiabatic nozzle (with the same assumptions):

$$h_1 + \left(\frac{V^2}{2}\right)_1 = h_2 + \left(\frac{V^2}{2}\right)_2. \quad (5.33)$$

Now, however, $V_2 < V_1$, so $h_2 > h_1$. The state of the fluid is transformed by the diffuser as shown in Fig. 5.5.

If the exit area of the diffuser is much larger than the inlet area and the flow is one-dimensional, then $V_2 \ll V_1$ and $\Delta(V^2/2) \approx -V_1^2/2$. In this limit, $h_2 = h_1 + V_1^2/2$. The quantity

$$h^* = h + \frac{V^2}{2} \quad (5.34)$$

is known as the *stagnation enthalpy* of a fluid with enthalpy h and speed V , since it is the enthalpy the fluid would have if it were brought to rest (stagnation) in an adiabatic diffuser. In terms of the stagnation enthalpy, the energy balance for an adiabatic nozzle or diffuser may be written

$$h_1^* = h_2^*. \quad (5.35)$$

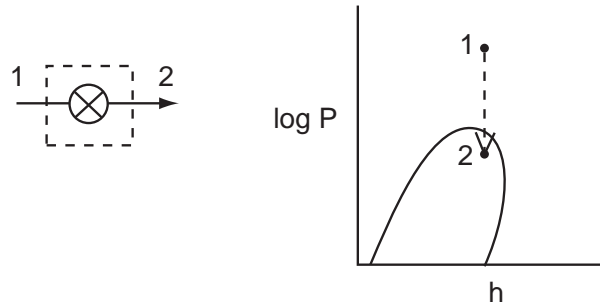


Figure 5.6: An adiabatic valve.

5.5.3 Valves

A *valve* is a device which drops the pressure in a flowing fluid without significantly increasing its kinetic energy. The interior of a valve looks something like a nozzle: there is a constriction in the flow passage, requiring the fluid to accelerate to pass through. But in a valve, the flow passage opens up abruptly after the constriction, causing one-dimensional flow to break down. A highly three-dimensional turbulent flow develops, in which most of the kinetic energy of the flow is converted into internal energy through viscous (friction) forces.

Valves are highly non-equilibrium, irreversible devices. But all we require to carry out a steady-state analysis of a valve is the ability to evaluate properties just upstream and just downstream of the valve. As long as the flow is reasonably one-dimensional and describable with equilibrium properties at these points, what is going on within the valve is irrelevant for the analysis.

Like nozzles and diffusers, valves are often approximated as adiabatic, since usually the fluid does not spend enough time in the valve for appreciable heat transfer to occur. Therefore we will make the following assumptions for the analysis of a valve: steady, one-dimensional flow at the inlet and outlet (states 1 and 2); equilibrium properties at 1 and 2; adiabatic; steady state; negligible change in kinetic and potential energy.

With these assumptions, the energy balance is simply

$$h_1 = h_2. \quad (5.36)$$

Since $P_2 < P_1$, the action of a valve on the state of the fluid is as shown in Fig. 5.6. Of course, state 1 can be anywhere in the (P, h) plane, and state 2

can be any point directly below state 1. Depending on conditions, the fluid emerging from the valve could be liquid, vapor, or mixed.

Example 5.3 An adiabatic valve in a liquid oxygen line causes a pressure drop of 0.5 MPa. If saturated liquid at 1 MPa enters the valve, what is the state of the oxygen emerging from the valve?

Solution: Assume:

1. steady, 1D flow, equilibrium properties at inlet and outlet
2. steady state
3. adiabatic
4. negligible change in P.E. or K.E.

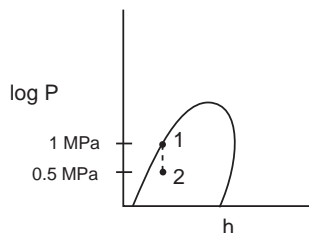
Then the energy balance is

$$h_1 = h_2.$$

From TPX, $h_1 = h_f(P_1) = \text{h}(\text{"o2"}, \text{"px"}, 1, 0) = 114.16 \text{ kJ/kg}$. The inlet temperature $T_1 = T_{sat}(P_1) = \text{temp}(\text{"o2"}, \text{"px"}, 1, 0) = 119.68 \text{ K}$. State 2 is fixed by $P_2 = 0.5 \text{ MPa}$ and $h_2 = h_1$. Therefore,

$$T_2 = \text{temp}(\text{"o2"}, \text{"ph"}, 0.5, 114.16) = 108.86 \text{ K}$$

$$X_2 = \text{x}(\text{"o2"}, \text{"ph"}, 0.5, 114.16) = 0.106.$$



5.5.4 Compressors and Pumps

Compressors and pumps have the same function – they increase the pressure of the fluid flowing through them. We usually use the term *pump* if the fluid is a liquid, and *compressor* if the fluid is a gas. Work input is required to push the fluid from the low-pressure inlet to the high-pressure outlet. (Unlike a diffuser, the flow entering a compressor or pump does not have significant kinetic energy at the inlet.) Internally, a compressor or pump has a set of blades or vanes, which are mounted on a central rotating shaft. The blades are designed so that

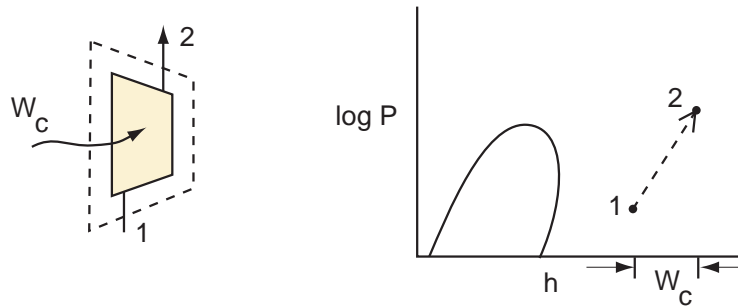


Figure 5.7: An adiabatic compressor.

as the shaft rotates the blades impart momentum to the fluid, pushing it toward the outlet.

Again, compressors and pumps are usually idealized as adiabatic devices. Making appropriate assumptions about the system, the energy balance for a compressor is

$$\dot{m}h_1 + \dot{W}_c = \dot{m}h_2. \quad (5.37)$$

See if you can list the assumptions implied by this equation.²

Since \dot{m} is a constant, it is convenient to divide through by it to obtain

$$h_1 + W_c = h_2. \quad (5.38)$$

Here $W_c = \dot{W}_c/\dot{m}$ is the input compressor work per unit mass of fluid flowing through the compressor (J/kg).

Of course, for a pump the analogous energy balance is

$$h_1 + W_p = h_2, \quad (5.39)$$

where $W_p = \dot{W}_p/\dot{m}$.

5.5.5 Turbines

A turbine is a device which produces continuous power by expanding a fluid flowing through it. The most familiar example is a windmill, or *wind turbine*. A wind turbine has blades mounted on a rotating shaft. As air flows over the blades, the shaft turns. If the shaft is connected to a load, such as an electrical generator, power is delivered to the load.

²At inlet and outlet: 1D, steady flow, equilibrium properties, negligible K.E. and P.E.; steady state; adiabatic.

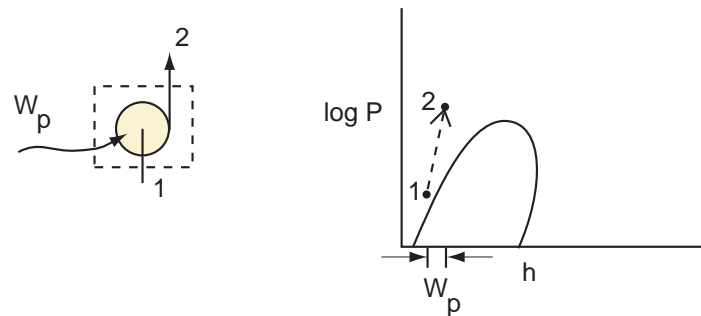


Figure 5.8: An adiabatic pump.

Although a wind turbine extracts the power from the kinetic energy of the air stream, an industrial turbine incorporates internal nozzles to accelerate a high-pressure input stream to high velocity before it impacts the turbine blades. Many turbines incorporate several rows of blades, one behind the other, to convert as much of the kinetic energy at the outlet of the nozzle to useful shaft power as possible. The flow emerges from the turbine at lower pressure than it went in, with negligible kinetic energy.

Turbines too are usually idealized as adiabatic. Therefore, to analyze a turbine we will assume: steady, 1D flow at inlet and outlet; negligible K.E. and P.E. at inlet and outlet; adiabatic; steady state.

Up until now, when writing down energy balance equations, we have always regarded heat and work as positive if the energy transfer is from the environment to the system. Energy transfer in the other direction can be treated as *negative* energy transfer to the system. Since a turbine delivers power to an external load, if we do the energy balance this way we would have to regard the turbine power as negative.

However, it is usually more convenient to regard heat and work transfers as positive in doing an analysis. It is simple to modify our procedure to do this. We only need to remember whether we are defining \dot{W} and \dot{Q} as an energy inflow or outflow, and put it in the appropriate place in the energy balance equation (i.e., on the same side of the equation as the other energy inflows or outflows). To avoid confusion, it is best to *always draw a sketch of the process, with the direction of energy flows clearly labeled*.

With the assumptions listed above, and the turbine power regarded as an

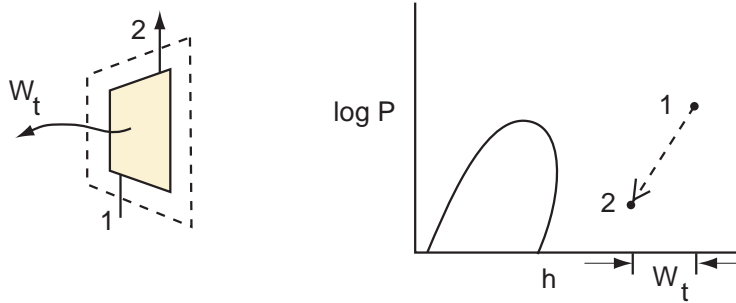


Figure 5.9: An adiabatic turbine.

energy outflow (as shown in the sketch in Fig. 5.9), the energy balance is

$$\dot{m}h_1 = \dot{m}h_2 + \dot{W}_t. \quad (5.40)$$

Dividing by \dot{m} ,

$$h_1 = h_2 + W_t, \quad (5.41)$$

where $W_t = \dot{W}_t/\dot{m}$. An adiabatic turbine changes the state of the fluid as shown in Fig. 5.9.

5.5.6 Heat Exchangers

The last devices we will introduce are *heat exchangers*. As the name implies, these are devices which transfer heat from one fluid stream to another. An example of a heat exchanger is a car radiator, which transfers heat from the hot engine coolant to air which is forced to flow over the radiator coils by the fan.

The simplest heat exchanger is a tube which one fluid (“A”) flows through, with the other one (“B”) flowing over the tube. If A and B are at different temperatures and the tube wall is a good conductor of heat (usually a metal) then heat will flow between the two fluids. If fluid B is the atmosphere (as in a car radiator) or a body of water such as a river, this open design is fine. Otherwise, fluid B should be enclosed in a tube also. One common design is to use concentric tubes, with A flowing one way through the center tube and B flowing the other way through the annulus surrounding the center tube.

The change of state for each fluid can be analyzed separately by defining a control volume which encompasses only that fluid. For example, the control volume shown in Fig. 5.10 is appropriate to determine the change in state of fluid A.

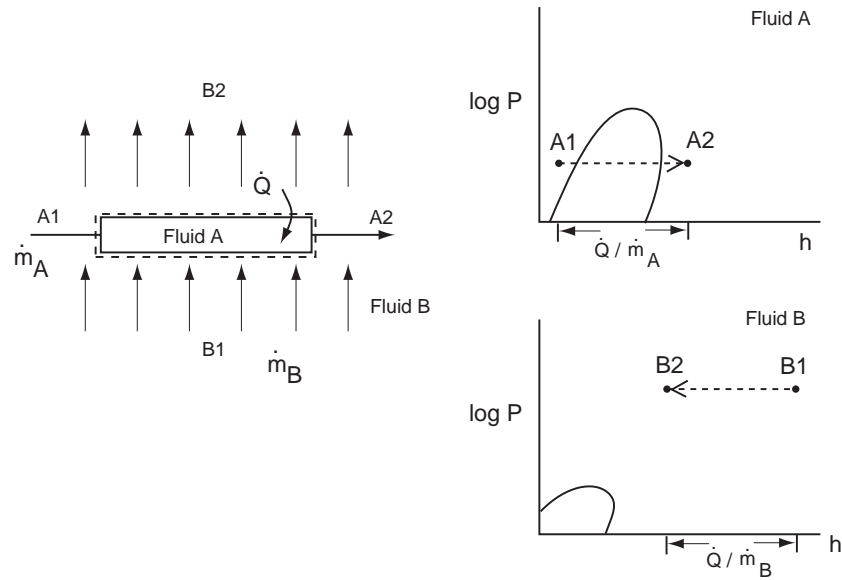


Figure 5.10: A simple constant-pressure heat exchanger. Note that A and B may be different fluids, with different mass flow rates.

Heat exchangers are usually constructed of such that the pressure drop due from the inlet to the outlet for either fluid due to viscous drag is small. For this reason, the flow through a heat exchanger is usually modeled (at least to first approximation) as occurring at constant pressure.

An appropriate set of assumptions to analyze flow through a heat exchanger is:

1. Steady, 1D flow at inlet and outlet; equilibrium properties at these points
2. Steady state
3. Constant pressure

For stream A in Fig. 5.10, with the direction of heat transfer as shown, the energy balance is

$$\dot{m}_A h_{A1} + \dot{Q} = \dot{m}_A h_{A2}. \quad (5.42)$$

Dividing by \dot{m}_A ,

$$h_{A1} + Q_A = h_{A2}, \quad (5.43)$$

where $Q_A = \dot{Q}/\dot{m}_A$.

Of course, the heat added to stream A came from stream B, so the energy balance on B is

$$\dot{m}_B h_{B1} = \dot{Q} + \dot{m}_B h_{B2}, \quad (5.44)$$

or

$$h_{B1} = Q_B + h_{B2}, \quad (5.45)$$

where $Q_B = \dot{Q}/\dot{m}_B$. Note that in general $\dot{m}_A \neq \dot{m}_B$, so $Q_A \neq Q_B$, even though \dot{Q} is the same for both.

The inlet state for either fluid may be anywhere on the $P-h$ plot, and the outlet state may be anywhere at the same pressure. Some particular types of heat exchangers are given names, depending on their primary function.

An *evaporator* or *boiler* is a heat exchanger which takes in liquid and produces vapor (as shown for stream A in Fig. 5.10). The minimum heat input required to do this is for the case where the input is saturated liquid, and the output saturated vapor. In this case,

$$Q_{in} = h_g - h_f. \quad (5.46)$$

The symbol h_{fg} is used to denote $h_g - h_f$, and is called the *enthalpy of vaporization* or *heat of vaporization*. If the output of the evaporator is superheated vapor, the last segment of the heat exchanger where liquid is no longer present is often called the *superheater*.

A *condenser* is just the opposite: it takes in vapor, removes heat causing the vapor to condense, and outputs liquid. A condenser must transfer at least h_{fg} of heat to the other fluid stream in order to condense the vapor.

CHAPTER 6

ENTROPY AND THE SECOND LAW

6.1 Introduction

We have now developed the First Law of Thermodynamics for both closed and open systems, and shown how it may be used to solve practical problems. But the First Law is not the end of the story. There are many imaginable processes which satisfy the First Law, but are nevertheless don't seem to happen. We need another principle, or "law," to explain why not. In this chapter, we introduce this principle – the *second* law of thermodynamics and the new property associated with it, the entropy.

6.2 Possible and Impossible Processes

Of all processes which satisfy the First Law, some actually happen and some never do. If a few ice cubes are added to a thermos bottle containing boiling water, the ice spontaneously melts and the initially-hot water cools. When equilibrium is reached, only warm water is left. The *reverse*, however, never happens: warm water has never been observed to spontaneously transform into hot water + ice cubes, even though it is possible to do so in a way which conserves energy.

If a rubber ball is held above a table and then dropped, it bounces a few times and comes to rest. The initial gravitational potential energy of the ball is converted to internal energy of the ball (and possibly of the table): the ball ends up slightly warmer than it began. The reverse process — a ball cooling off slightly and jumping off a table – has never been observed, although it could conserve energy.

In fact, for every process that really happens, the time-reversed version — in which the initial and final states of the system + environment are switched — never seems to. Eggs break when dropped on the floor; broken eggs have never been seen to "unbreak" and rise off the floor. Heat flows spontaneously from a high temperature body to a low temperature one it contacts; it has never been observed to flow spontaneously from low temperature to high. A helium balloon slowly deflates due to diffusion of helium atoms through the balloon

skin; a deflated balloon has never been observed to spontaneously inflate due to helium atoms diffusing from the surroundings to the interior.

Despite the observed one-way nature of real processes, the First Law makes no distinction between possible process and their impossible time-reversed versions. Consider, for example, a closed system which undergoes a process during which heat Q and work W are transferred to it from the environment. The First Law requires

$$E_{final} - E_{initial} = Q + W. \quad (6.1)$$

For the time-reversed version, the energy transfers occur in the other direction, so

$$E_{final}^{(rev)} - E_{initial}^{(rev)} = (-Q) + (-W). \quad (6.2)$$

But since $E_{initial}^{(rev)} = E_{final}$ and $E_{final}^{(rev)} = E_{initial}$, this is equivalent to

$$-(E_{final} - E_{initial}) = (-Q) + (-W), \quad (6.3)$$

which is of course equivalent to Eq. (6.1). We see that the First Law is satisfied equally by the forward and reverse processes. This must be true, since the First Law takes the form of an *equality*: if $A = B$, then $-A = -B$. What we would need to distinguish forward from backward would be an *inequality*: if A and B satisfy $A > B$, then $(-A)$ and $(-B)$ do not. Evidently, the reason some energy-conserving processes occur spontaneously and others don't has nothing to do with the First Law.

6.3 The Microscopic View of Spontaneous Processes

When we consider what happens on an atomic level, it is clear why some processes happen spontaneously and others don't. We'll consider two specific examples — dropping a ball, and letting a gas expand into vacuum. The general principles apply to any spontaneous process.

6.3.1 Dropping A Ball

Consider what happens on an atomic level when a rubber ball is dropped on a table (Fig. 6.1). To simplify matters, let's neglect the thermal energy the ball has before being dropped, ignore any energy transfer to the table, and assume the ball is dropped in vacuum, so there is no air resistance. Let's also imagine we have some way of taking an atomic-level “snapshot” of the positions and velocities of all atoms in the ball at any desired time — we can't do this in a real experiment, but it's quite easy in a thought experiment.

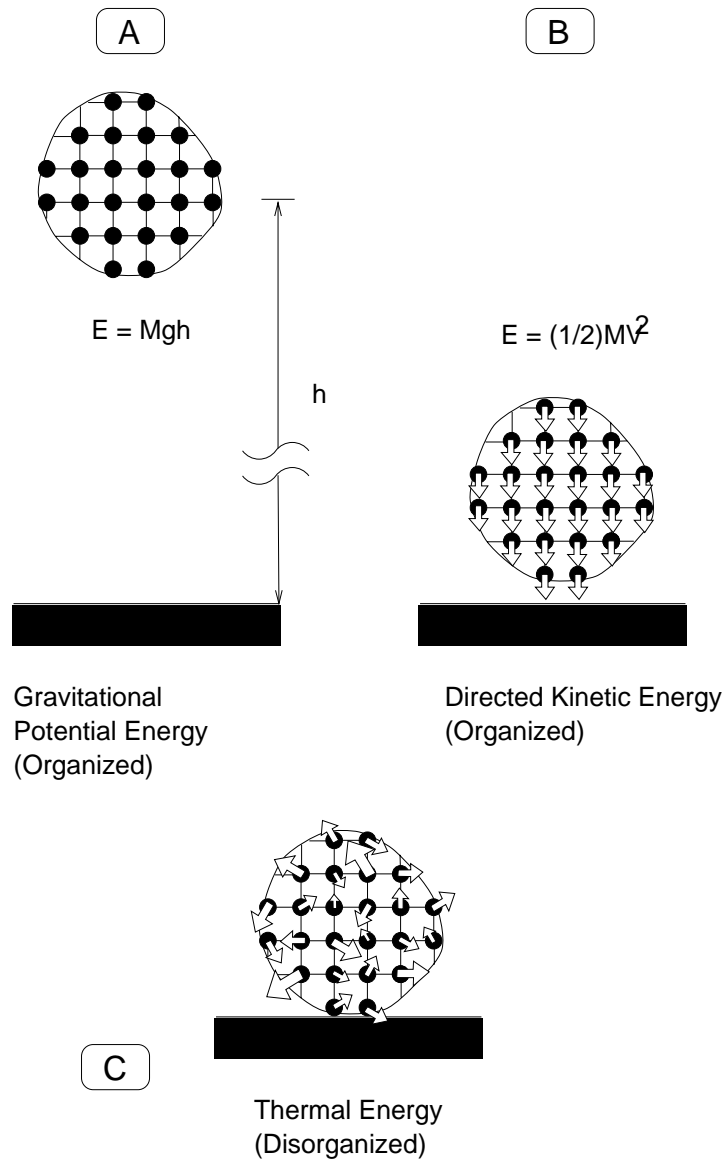


Figure 6.1: A rubber ball dropped on a table.

The ball has mass M and is dropped from a distance h above the table. Three different macroscopic states of the ball are shown in Fig. 6.1. In state A, the ball hasn't yet been dropped, so is stationary at a height h above the table. In state B, the ball is just about to strike the table the first time. In state C, the ball has stopped bouncing and reached the equilibrium state, resting on the table.

A set of atomic-level snapshots are taken in each state A, B, and C. In state A, all atoms are initially stationary (no thermal energy), so repeated snapshots always find the same atomic-level configuration. In state B, all atoms moving downward in unison at speed $V = \sqrt{2gh}$. Still, every snapshot is the same.

But in state C, every snapshot is different. The atoms are vibrating chaotically, and the velocity of each atom changes in an apparently random way from one snapshot to the next as the atom is pulled or pushed by the stretched or compressed bonds to neighboring atoms. Of course, energy is still conserved. If we computed for any snapshot the sum of the kinetic energies of all atoms plus the potential energies of all stretched or compressed bonds, the result would always be Mgh .

Each snapshot fully determines the instantaneous microscopic state of the ball, which we will call the *microstate*. In contrast, the *macroscopic* state of the ball is specified by only a few numbers, such as the height, center-of-mass velocity, temperature, etc. A macroscopic description of a ball (or any other system) is certainly simpler than a detailed, atomic-level specification of the microstate. But it is also much less precise, since a system in a given macroscopic state (sometimes called a *macrostate*) might be found in any one of a very large number of microstates.

If we tried to compile a list of all observed microstates for the ball in macrostate C, it would be a *very* long list. In fact, if position and velocity are measured with arbitrary precision (and therefore specified by real numbers), we would never find *exactly* the same value for any atomic position or velocity; there would be an uncountably infinite number of microstates.

But usually there is some uncertainty in measurements. For example, if the measurements are acquired by a computer or stored in one, they need to be represented by a finite number of bits, leading to round-off error. In this case, positions which differ by less than the round-off error Δx would be recorded as the same position, and velocities which differ by less than the velocity round-off error Δv would be recorded as the same velocity. In this case, the number of microstates observed in state C would be finite, but extremely large.

If we were extraordinarily patient, we could take enough snapshots to build

up statistics on the number of times each microstate is found. If after taking N_s snapshots, the ball was found in the j^{th} microstate in $N_{j,s}$ snapshots, then as $N_s \rightarrow \infty$ the probability p_j of finding the ball in microstate j is

$$p_j = \lim_{N_s \rightarrow \infty} \frac{N_{j,s}}{N_s}. \quad (6.4)$$

There is no particular reason to think that any one microstate of the ball in macrostate C should be found more often than any other one. Remember, we said the ball is in equilibrium in state C, which means it has been sitting on the table a very long time (at least compared to atomic vibration time scales). After enough time has elapsed, the atoms of the ball have “forgotten” about any coordinated motion (bouncing) they were executing previously, and are all moving more or less randomly.

Of course, we’re only considering microstates which are “accessible,” given the initial conditions and the conserved quantities (total energy, number of atoms, etc.). There are microstates of the ball which have total energy different than Mgh , or with a few atoms missing or bonded differently, but with the given initial conditions the ball will never be found in one of these inaccessible states.

With the assumptions we’ve made, the ball does not interact in any way with the environment, which is why its total energy must be constant. This means we are treating the ball as an isolated system.

The assumption that all microstates of the ball are equally probable at equilibrium can be generalized to any isolated system. In fact, this is the basic *postulate* of statistical mechanics:

Postulate: At equilibrium, all accessible microstates of an isolated system are equally probable.

We can’t prove this postulate, but we can ask what consequences follow from it, and whether they agree with experiment.

Let’s now return to the question of why some processes occur, and some don’t. On a microscopic level, all accessible microstates are equally likely at equilibrium. A particular microstate of the ball with atoms moving in various directions such that there is no center of mass velocity is just as likely as another one which has all atoms moving upward with the same speed, or one with the ball hovering completely stationary a distance h above the table. At first glance, this might lead you to conclude that a ball should be just as likely to jump off a table as to remain sitting there.

But of course, once the atoms are moving randomly in state C, it is highly unlikely that they would just happen to all move in such a way to push off from the table and rise up in unison, although there is no physical law preventing it.

Another way of saying this is that the number of microstates in which the ball is sitting on the table with all atoms moving more or less randomly (and thus with negligible net center-of-mass velocity) is *huge* compared to those few special microstates with the atoms all moving in one direction, or all displaced above the table. All accessible microstates — even ones corresponding to bizarre macroscopic behavior — are equally probable, but there are *far, far more* microstates which correspond to the ball simply sitting on the table.

Therefore, we would expect to hardly ever observe one of the microstates corresponding to the ball doing anything but sitting on the table. If we worked out the numbers, we would find that we would have to wait *much* longer than the age of the universe to see the ball spontaneously move — during which time, of course, other processes would occur which would make the whole experiment moot (disintegration of the ball and table, the end of life on earth, etc.)

So the basic idea is as follows. Denote by Ω the number of microstates which a system might be in when it is in some specified macroscopic state. Suppose an isolated system starts out in some equilibrium macrostate 1 which has a certain number of microstates Ω_1 . Now some process occurs within the system (e.g. the ball is dropped) which changes the number of available microstates to $\Omega_2 \gg \Omega_1$. Once enough time has elapsed, the system is again in equilibrium, so it could be in any one of its Ω_2 microstates with equal probability. It is still possible that it could be found in one of the original microstates, which would mean that macroscopically it would appear to be in macrostate 1. But the probability of this happening is Ω_1/Ω_2 , which is very small if $\Omega_2 \gg \Omega_1$.

6.3.2 Irreversible Expansion of an Ideal Gas

To see how the numbers work out, we need a system so simple that we can actually compute the number of microstates. (This would be hard to do in practice for a rubber ball.) Consider a gas of N identical point atoms in a container of volume V .

For the purposes of this discussion, we'll only specify the microstate by the spatial location of the atoms, and not consider how many ways there are to assign velocities to the atoms consistent with a specified energy. We'll come back to the question of how to partition the energy in Section 6.7.

If we specify the microstate by the *exact* position of each atom ($\{\mathbf{x}_1, \dots, \mathbf{x}_N\}$), then there are an uncountably infinite number of microstates, since each coordi-

nate position is specified by a real number. To make the number of microstates countable, let's assume each position measurement has some finite round-off error Δx . This is equivalent to dividing the volume into small cubes of volume $(\Delta x)^3$, and only recording *which* small cube an atom is in, rather than the exact co-ordinates within the cube. This makes the number of microstates finite and countable, although the answer will depend on Δx .

The number of small cubes n_c is

$$n_c = \frac{V}{(\Delta x)^3}. \quad (6.5)$$

We will assume that $n_c \gg N$, so that if we place the N atoms in cubes randomly, the odds of finding two or more in the same cube are very low. The vast majority of states do not have any cubes multiply-occupied. We'll allow multiple-occupancy in counting states, but it won't affect the result much one way or the other as long as $n_c \gg N$.

The number of spatial arrangements Ω_x is the number of ways of placing N identical atoms in n_c cubes. Since the atoms are identical,¹ we won't count arrangements which differ only by a permutation of atom labels as being different. That is, for $N = 2$, the arrangement with atom 1 in cube 47 and atom 2 in cube 129 is no different than arrangement with atom 2 in cube 47 and atom 1 in cube 129. We should count this arrangement once, but not twice. Therefore, for $N = 2$, $\Omega_x(N = 2) = n_c^2/2$. For $N = 3$, we have to divide the number of arrangements of 3 labeled atoms (n_c^3) by the number of ways to permute labels among 3 atoms ($3! = 6$). So $\Omega_x(N = 3) = n_c^3/6$. The general result is

$$\Omega_x(V, N) = \frac{1}{N!} n_c^N = \frac{1}{N!} \left[\frac{V}{(\Delta x)^3} \right]^N. \quad (6.6)$$

Let's use this result to analyze the process shown in Fig. 6.2.

A container is divided into two compartments, of volume V_A and V_B , respectively. Initially, compartment A contains an equilibrium gas of N identical point atoms, and compartment B is empty. A shutter is now opened in the partition, allowing atoms to move freely between A and B. Enough time is allowed to elapse so that the gas comes to equilibrium in the new volume $V_A + V_B$.

Now a sensitive detector is turned on, which can detect the presence of even one atom in compartment B. The detector looks for atoms in B every Δt

¹Quantum mechanics requires that identical atoms (the same isotope of the same element) are indistinguishable *even in principle*. So just exchanging two identical atoms, keeping everything else the same, can't yield a new microstate.

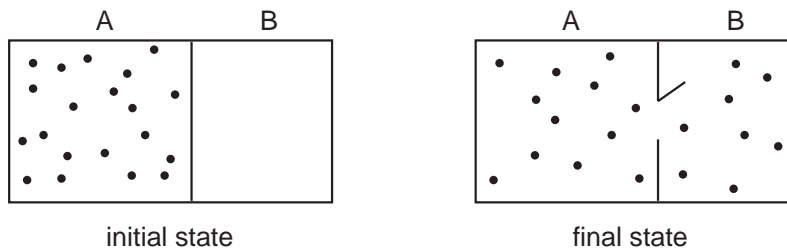


Figure 6.2: Irreversible expansion of a gas.

seconds. If the detector ever detects that B is empty, it sends a signal which closes the shutter, which returns the gas to its original state of being confined in compartment A. How much time elapses on average before B is found to be empty and the shutter closes?

Once the shutter has opened and the atoms have had ample time to move back and forth between A and B, the number of available microstates has increased by the factor

$$\frac{\Omega_{final}}{\Omega_{initial}} = \left(\frac{V_{A+B}}{V_A} \right)^N. \quad (6.7)$$

Since all microstates are equally probable at equilibrium, the probability p_{empty} that the detector will find none of the N atoms in B in any one measurement is simply the ratio of the number of microstates which have all of the atoms in A to the total number of microstates:

$$p_{empty} = \frac{\Omega_{initial}}{\Omega_{final}} = \left(\frac{V_A}{V_{A+B}} \right)^N. \quad (6.8)$$

If N is small, then p_{empty} is fairly large, and only a few measurements would be needed, on average, before the detector finds B empty. For example, if $V_A = V_B$ and $N = 2$, then $p_{empty} = 1/4$, so the detector would be expected to find B empty after only about 4 measurements. If $N = 20$, $p_{empty} = 9.5 \times 10^{-7}$. Now it is unlikely to find B empty on any given measurement, but statistically once every million or so times it should happen, and the shutter would then close, trapping the gas in A. If the detector took a measurement, say, every second, then it would take about 12 days (a million seconds) before we would expect the shutter to close and effectively “reverse” the effect of opening the shutter. We conclude that if N is small (say, 20 or so) then the process of opening the shutter can be reversed, if we are willing to wait a little while.

Suppose, however, that we have a macroscopic gas, with $N \sim 10^{23}$. Then

$$p_{empty} = \left(\frac{V_A}{V_{A+B}} \right)^{10^{23}}. \quad (6.9)$$

If $V_A = V_B$,

$$\begin{aligned} p_{10^{23}} &= 2^{-10^{23}} \\ &= 10^{-10^{23} \log_{10} 2} \\ &\approx 10^{-3 \times 10^{22}}. \end{aligned} \quad (6.10)$$

Now *that* is a small probability — its reciprocal would be written as a “1” followed by about 3×10^{22} zeros.

The universe is of order 10^{10} years old, which is 3×10^{17} seconds. Therefore, the chance of detecting no atoms in B during the age of the universe, checking every second, is

$$(3 \times 10^{17})(10^{-3 \times 10^{22}}) = 3 \times 10^{-(3 \times 10^{22} - 17)}.$$

Thus, it is *overwhelmingly* unlikely to occur in the age of the universe. Even if the detector looked every microsecond, this would only change the probability to $3 \times 10^{-(3 \times 10^{22} - 23)}$ — still far too small. Even if we could wait 10^{10} universe lifetimes it would still be overwhelmingly unlikely. We conclude therefore that if $N \sim 10^{23}$, the process of opening the shutter is truly *irreversible*: it won't spontaneously reverse, even over times much longer than the age of the universe. For this reason, the process of a macroscopic gas expanding into vacuum is called *irreversible expansion*. It is also called *unrestrained expansion*, since there is no piston restraining the gas expansion that the gas must do work against.

This example illustrates a common feature of macroscopic systems. Since the number of atoms is very large (say, $> 10^{20}$), and the formula for Ω typically has N in an exponent, making a minor change like opening a shutter to increase the volume expands the number of available microstates by an unimaginably huge degree. Once the system can be found in any of these with equal probability, the probability of finding it in one of its original allowed microstates is effectively zero, *even if we watch it for the age of the universe*. When this is the case, we say that the “minor change” which was made is *irreversible*.

Of course, it is possible to restore a system to its initial state, but only at the cost of altering the environment irreversibly. In the previous example, a piston could be used to compress the gas back into compartment A. However, this would require work input, which would increase the energy of the gas.

This extra energy would have to be removed to the environment as heat, which would irreversibly alter the microscopic state of the environment. In particular, the energy added to the environment would increase its number of possible microstates by a huge factor.

6.4 The Second Law of Thermodynamics

The Second Law of Thermodynamics simply expresses as a general “law” the characteristics of spontaneous, irreversible processes discussed in the last section. Consider first an isolated system (constant energy), since in this case all accessible microstates have the same energy, and we don’t have to be concerned about what is happening in the environment. The Second Law states that

No process will occur within an isolated, macroscopic system that decreases the number of accessible microstates.

All this says is that isolated, macroscopic systems will proceed in the direction of increasing Ω if allowed to (say, by letting go of a ball, or opening an internal shutter in a gas container, etc.), but will never proceed in the direction of decreasing Ω . Of course, we mean “never” in the sense of the irreversible expansion example of the last section.

The second law says only that the number of possible microstates of an isolated system cannot decrease; it does not require the number to increase by any specific amount. In particular, processes which do not change the number of microstates are allowed. Such processes are called *reversible*, since they are the only ones which may be reversed without violating the Second Law.

For non-isolated systems which interact with the environment, the system + environment together constitute an isolated system, so the Second Law may be state more generally:

No process that decreases the number of accessible microstates of the system + environment will occur.

6.5 The Entropy

The number of microstates Ω associated with some macroscopic, equilibrium state of matter is a useful quantity to know. As we saw in the irreversible expansion example, knowing Ω and its dependence on macroscopic parameters such as the volume allows us to determine what sort of processes can occur (expansion of gas through an opening in a partition), and what sort can’t (spontaneous collection of gas atoms in one part of a container).

However, there are two reasons why working with Ω directly is inconvenient. First, it is difficult to work with numbers as large as, say, $10^{10^{23}}$ – we don't have much intuitive feel for them.

The second problem becomes clear if we consider a system which is composed of two non-interacting parts, A and B . If A and B are independent, and A has Ω_A microstates, and B has Ω_B microstates, then

$$\Omega_{A+B} = \Omega_A \Omega_B. \quad (6.11)$$

Therefore, the property Ω is multiplicative for systems composed of separate parts. Since other properties we've worked with like mass, volume, and energy are additive (extensive), it is inconvenient to have to work with a multiplicative property.

Both problems can be solved by working with $\ln \Omega$, rather than Ω . The logarithm of Ω typically is proportional to N , which is still a big number but is quite manageable compared to 10^N . Also, $\ln \Omega$ is additive:

$$\ln \Omega_{A+B} = \ln \Omega_A + \ln \Omega_B. \quad (6.12)$$

We will define the *entropy* S as

$$S = k_B \ln \Omega. \quad (6.13)$$

The constant k_B is “Boltzmann's constant,” named after Ludwig Boltzmann, who first proposed this equation. It is introduced only for convenience, and can be set to any desired value. Given S , it is always possible to recover $\Omega = \exp(S/k_B)$. Sometimes it is most convenient to use $k_B = 1$, in which case the entropy is dimensionless. But we'll soon see that the choice of k_B affects the units of other properties, in particular temperature. If we have already decided what units we want to use for temperature (e.g. Kelvin), then this will determine the value for k_B . We'll discuss this point more in Section 6.9.

It is important to remember that we have been considering equilibrium states of isolated systems (constant energy U , volume V , number of atoms N). For isolated systems, the only accessible microstates are the ones with the right values of U and N , and for which all atoms are contained within V . The number of these will depend on U , V , and N : $\Omega(U, V, N)$. Therefore, the entropy has the same functional dependence: $S(U, V, N)$.

Since the system is in equilibrium, all accessible microstates are assumed to be equally likely, and therefore all that really matters is how many of them there are. If it were not in equilibrium, then some microstates would be more

probable than others. For example, in the irreversible expansion example, just after the shutter is opened microstates with most atoms in A are more probable than ones with most atoms in B. It can be shown that in this case the entropy should be defined as

$$S = k_B \sum_j p_j \ln p_j, \quad (6.14)$$

where the sum is over all accessible microstates. If all microstates are equally probable, ($p_j = 1/\Omega$) Eq. (6.14) reduces to Eq. (6.13).

For our purposes, we won't need to evaluate S for non-equilibrium states, so we won't need Eq. (6.14). The only sort of non-equilibrium states we'll need to consider explicitly are ones composed of separate parts, each of which is in local thermodynamic equilibrium. Since S is extensive, in this case S can be written as

$$S = \sum_{k=1}^K S_k(U_k, V_k, N_k). \quad (6.15)$$

Here the system is divided into K parts, each of which is internally in equilibrium (all of its microstates are equally probable), but which may not be in equilibrium with one another. Each term S_k is evaluated as the entropy of the equilibrium state with energy U_k , volume V_k , and N_k atoms.

Once we have $S(U, V, N)$, we could change variables from number of atoms to the total mass, since $M = mN$, where m is the mass of one atom. This would produce the function $S(U, V, M)$, which depends only on macroscopically-measurable quantities.

Now we also know that S is extensive, like U , V , N , and M . The extensive properties have the characteristic that they all scale with the amount of the substance present. Therefore, if we scale up the size of the system by some factor λ , all extensive variables must be scaled by λ . The entropy function evaluated for the scaled-up system $S(\lambda U, \lambda V, \lambda M)$ must be λ times larger than the entropy function evaluated for the original system $S(U, V, M)$:

$$\lambda S(U, V, M) = S(\lambda U, \lambda V, \lambda M). \quad (6.16)$$

Taking $\lambda = 1/M$,

$$\frac{S}{M} = S\left(\frac{U}{M}, \frac{V}{M}, 1\right). \quad (6.17)$$

Therefore, as for the other extensive properties, we define the *specific entropy* s by

$$s = \frac{S}{M} = s(u, v), \quad (6.18)$$

which from Eq. (6.17) depends only on (u, v) .

Therefore, for a simple compressible substance in equilibrium, there exists some entropy equation of state $s(u, v)$. Of course, as we discussed in Chapter 3, any two independent properties can be used to specify the thermodynamic state. To construct, for example, $s(h, P)$ we only need to know the equations of state $u(h, P)$ and $v(h, P)$: $s(h, P) = s(u(h, P), v(h, P))$. The specific entropy can also be used as one of the properties to fix the state. For example, $s(h, P)$ could be inverted to form $h(s, P)$.

For most substances, figuring out all of the microstates consistent with specified (U, V, N) is too difficult to do in practice, although in *principle* it can always be done. But the function $s(u, v)$ still exists, even if we have difficulty calculating it directly from $S = k_B \ln \Omega$. Fortunately, as we'll discuss below, there are ways to determine $s(u, v)$ purely from macroscopic measurements except for an arbitrary constant of integration.

6.6 Entropy Production

If some process occurs within an isolated system, the Second Law requires $\Omega_{final} \geq \Omega_{initial}$. Therefore,

$$\Delta S = S_{final} - S_{initial} = k_B \ln \left(\frac{\Omega_{final}}{\Omega_{initial}} \right) \geq 0. \quad (6.19)$$

Since this extra entropy wasn't transferred into the system from the environment (the system is isolated), *we must regard the extra entropy to have been produced inside the system during this process.*

Thus, irreversible processes produce entropy – unlike energy and mass, entropy is not conserved. We'll call the amount of entropy produced by a process the *entropy production* \mathcal{P}_s . Every irreversible process produces entropy ($\mathcal{P}_s > 0$). Besides unrestrained expansion, some other processes which produce entropy include motion against frictional forces, combustion, electrical current flow through a resistor, heat transfer through a finite temperature difference, mixing of two different fluids, and even clearing data in the memory of a digital computer.

For a process occurring in a non-isolated system, the entropy produced by the process equals the increase in the entropy of the system + environment. A very compact, general way of writing the Second Law is

$$\boxed{\mathcal{P}_s \geq 0} \quad (6.20)$$

That is, entropy can be produced, but it can never be destroyed. Since this form of the Second Law applies to all types of systems, it is the form we will use most often.

The special class of processes for which $\mathcal{P}_s = 0$ are reversible, since reversing them would not result in destruction of entropy, as it would for a process with $\mathcal{P}_s > 0$. Reversible processes can't be actually achieved in practice, since friction or other irreversibilities are always present, even if only in very small amounts. However, real processes can approach reversible ones as a limiting case, as we make irreversibilities (electrical resistance, friction, etc.) smaller and smaller.

6.7 The Entropy of a Monatomic Ideal Gas

One of the few systems which is simple enough to allow computing $s(u, v)$ from first principles is the monatomic ideal gas. We've introduced the ideal gas previously as the low-density limit of a simple compressible substance, and stated that it satisfies $Pv = RT$. From a microscopic viewpoint, an ideal gas is a gas of particles of negligible size, which have kinetic energy but have negligible interaction potential energy. A real gas approximates this when the container volume per atom is much larger than the atomic size, and the average distance between atoms is much greater than the distance over which they exert appreciable attractive or repulsive forces on one another. A *monatomic* ideal gas (e.g. He, Ar, Ne) is particularly simple, since in this case the energy of each particle is simply $(1/2)m|\mathbf{v}|^2$ – there is no rotational or vibrational energy, as there would be in a molecular ideal gas.

Consider an isolated system consisting of a monatomic ideal gas of N atoms contained in volume V . For simplicity, assume the volume is a cube with side length L . Since it is isolated, the total energy is fixed at some value; call it U . Then every possible microstate $\{\mathbf{x}_1, \dots, \mathbf{x}_N; \mathbf{v}_1, \dots, \mathbf{v}_N\}$ of the system must satisfy

$$\sum_{n=1}^N \frac{mv_n^2}{2} = U. \quad (6.21)$$

It will turn out to be more convenient to work with the momentum of an atom $\mathbf{p}_n = m\mathbf{v}_n$, rather than velocity. In terms of momentum, this equation becomes

$$\sum_{n=1}^N \frac{p_n^2}{2m} = U. \quad (6.22)$$

Also, all N atoms must be in the container, so the microstate must satisfy

$$0 \leq x_n \leq L, \quad n = 1, \dots, N$$

$$\begin{aligned} 0 \leq y_n \leq L, \quad n = 1, \dots, N \\ 0 \leq z_n \leq L, \quad n = 1, \dots, N. \end{aligned} \quad (6.23)$$

How many microstates are there which satisfy these constraints? Clearly, if we can really measure position and momentum with arbitrary accuracy, there is an uncountably-infinite number. So let's introduce some small round-off error to make the states countable. Say the position round-off error is Δx , and the momentum round-off error is Δp .

We already solved part of this problem when we calculated how many ways there are to arrange N atoms in volume V [Eq. (6.6)]. Since for an ideal gas the energy depends only on the atomic momenta, not on their positions, the calculation of the number of ways to arrange the atoms in space (Ω_x) and the number of ways to distribute the energy as kinetic energy among the atoms (Ω_p) can be done independently. The total number of microstates is simply the product:

$$\Omega = \Omega_x \Omega_p. \quad (6.24)$$

The details of the calculation of Ω_p are relegated to Appendix A. The result is

$$\Omega_p = \frac{(2\pi mU)^{3N/2}}{(\Delta p)^{3N} \left(\frac{3N}{2}\right)!}. \quad (6.25)$$

Therefore, using Eq. (6.6), the total number of microstates is

$$\Omega = \Omega_x \Omega_p = \frac{1}{N!} \frac{V^N (2\pi mU)^{3N/2}}{(\Delta x \Delta p)^{3N} \left(\frac{3N}{2}\right)!}. \quad (6.26)$$

Classically, the choice of $\Delta x \Delta p$ is arbitrary. Nevertheless, when $S = k_B \ln \Omega$ is evaluated, the term involving $\Delta x \Delta p$ just becomes an additive constant equal to $-3Nk_B \ln(\Delta x \Delta p)$, so this classical treatment determines the entropy of an ideal gas to within an arbitrary additive constant. Since we only need to know *differences* in entropy to determine if a process satisfies the Second Law, any additive constant will cancel when the difference is taken.

We can do a little better if we supplement this classical analysis with the uncertainty principle of quantum mechanics. The uncertainty principle states that a particle (which in quantum mechanics is a wave packet) cannot simultaneously have precise values of position and momentum — \mathbf{x} or \mathbf{p} or both must have some “fuzziness” or uncertainty. The momentum uncertainty in any direction δp and the position uncertainty in that direction δx are related by

$$(\delta x)(\delta p) \geq h, \quad (6.27)$$

where

$$h = 6.626 \times 10^{-34} \text{ J-s} \quad (6.28)$$

is *Planck's constant*.

It seems natural to choose $\Delta x \Delta p = h$ in Eq. (6.26). In fact, it can be shown that this choice reduces Eq. (6.26) to what would result if we had done the analysis quantum-mechanically from the start. Making this substitution,

$$\Omega = \frac{1}{N!} \frac{V^N (2\pi m U)^{3N/2}}{h^{3N} \left(\frac{3N}{2}\right)!}. \quad (6.29)$$

The entropy is $k_B \ln \Omega$. If N is larger than about 100, $\ln N!$ is well-approximated by $\ln N! \approx N \ln N - N$, which is known as *Stirling's approximation*. It is left as an exercise to show that in this case the entropy is

$$S = Nk_B \left[\frac{3}{2} \ln \left(\frac{U}{N} \right) + \ln \left(\frac{V}{N} \right) + \frac{3}{2} \ln \left(\frac{4\pi m}{3h^2} \right) + \frac{5}{2} \right]. \quad (6.30)$$

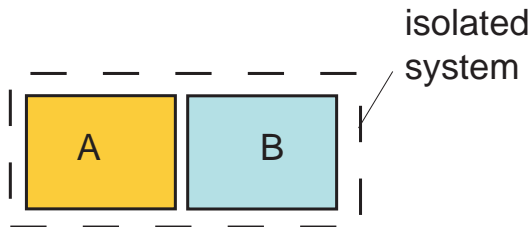
This expression for the absolute entropy of a monatomic ideal gas.

6.8 The Second Law and Equilibrium

The Second Law has significant implications regarding the nature of equilibrium. In this section, we'll take a look at equilibrium from the point of view of the Second Law. This will extend our analysis of Chapter 3, and result in relationships between the properties defined there (temperature, pressure, and chemical potential) and the entropy.

6.8.1 The Entropy Maximum Principle

Consider the process shown below. Two solid blocks, A and B, are brought into contact briefly and then separated. Heat may flow from one to the other during the contact time, but there is no energy or mass transfer to the environment. Therefore, the two blocks together constitute an isolated system.

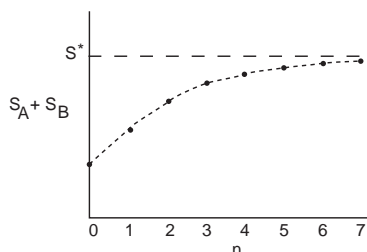


The Second Law ($\mathcal{P}_s \geq 0$) prohibits any process occurring within an isolated system which would decrease its entropy. We will also exclude the possibility of

a reversible process occurring ($\mathcal{P}_s = 0$), since every *real* process which occurs in finite time produces some entropy, although \mathcal{P}_s can be made arbitrarily small.

Therefore, if energy transfers as heat between the blocks, it must do so in a way which increases the entropy $S = S_A(U_A) + S_B(U_B)$ of the whole system, measured after the blocks have been separated and each has again come into internal equilibrium (uniform temperature).

If this process is repeated many times, the entropy of the whole system must increase each time. But the entropy can't go on increasing *without bound*, since there is a finite number of ways to distribute the total energy $U_A + U_B$ among all the atoms in block A and block B. Since Ω is bounded, so is S . Therefore, the entropy must approach some limiting value as the number of contact periods goes to infinity.



In the limit of infinite contact periods (infinite contact time) the state is no longer changing, and by definition this is the condition of thermal equilibrium between blocks A and B. The same analysis would apply to any other process occurring in an isolated system, not only heat flow (chemical reaction, mixing of two different species, fluid motion, etc.). The process would occur for a while, continually increasing the entropy of the system, until it has maximized the entropy as much as it can, then no further change will be observed. The system will be in equilibrium.

So we have a general criterion for the attainment of equilibrium in an isolated system:

The equilibrium state of an isolated macroscopic system has the maximum entropy, compared to all other possible states of the system.

This is known as the *Entropy Maximum Principle*. By “state” we mean of course the macroscopic state, and by “possible” states, we mean ones with the same total energy, volume, mass, and elemental composition as the system began with, since these quantities are fixed for an isolated system.

A few examples will make clear how the Entropy Maximum Principle is used.

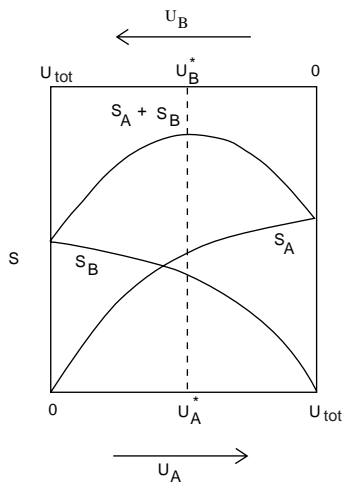


Figure 6.3: The equilibrium state of an isolated system maximizes the total entropy $S_A + S_B$ with respect to the quantity being varied (here U_A).

6.8.2 Thermal Equilibrium

Consider again the example of the last section. Suppose block A initially has internal energy $U_{A,0}$, and block B has $U_{B,0}$. After they have come to thermal equilibrium, energy may have transferred as heat from one to the other, changing U_A and U_B . What are the equilibrium values U_A^* and U_B^* ?

To answer this question, we can use the Entropy Maximum Principle. First note that the total energy $U_{tot} = U_A + U_B$ is fixed, since the system is isolated. So at any time, $U_B = U_{tot} - U_A$.

The entropy of each block depends on its energy, volume, and number of atoms. So the total system entropy is

$$S = S_A(U_A, V_A, N_A) + S_B(U_B, V_B, N_B), \quad (6.31)$$

where

$$U_B = U_{tot} - U_A. \quad (6.32)$$

Note that in this problem the only independent variable we can change is U_A ; everything else is fixed.

The Entropy Maximum Principle says that at equilibrium, U_A will have the value which maximizes S . Therefore, by plotting S vs. U_A , it is possible to determine the equilibrium value U_A^* (Fig. 6.3).

There is another way to find the maximum of S . Instead of plotting it, we

could calculate dS/dU_A , and set this derivative to zero to find U_A^* :

$$\frac{dS}{dU_A} = \left(\frac{\partial S_A}{\partial U_A} \right)_{V_A, N_A} + \left(\frac{dU_B}{dU_A} \right) \left(\frac{\partial S_B}{\partial U_B} \right)_{V_B, N_B} = 0 \quad (6.33)$$

at equilibrium. From Eq. (6.32),

$$\frac{dU_B}{dU_A} = -1. \quad (6.34)$$

Putting this into Eq. (6.33), we find

$$\frac{dS}{dU_A} = \left(\frac{\partial S_A}{\partial U_A} \right)_{V_A, N_A} - \left(\frac{\partial S_B}{\partial U_B} \right)_{V_B, N_B}. \quad (6.35)$$

At the maximum (thermal equilibrium), $dS/dU_A = 0$, so we conclude that at thermal equilibrium

$$\left(\frac{\partial S_A}{\partial U_A} \right)_{V_A, N_A} = \left(\frac{\partial S_B}{\partial U_B} \right)_{V_B, N_B} \quad (6.36)$$

or changing the notation slightly

$$\left[\left(\frac{\partial S}{\partial U} \right)_{V, N} \right]_A = \left[\left(\frac{\partial S}{\partial U} \right)_{V, N} \right]_B. \quad (6.37)$$

This equation states that at thermal equilibrium, A and B will have the same value of the property $(\partial S/\partial U)_{V, N}$. The functions S_A and S_B may differ, since A and B may be different substances. But when equilibrium is attained, the partial derivatives of these two different functions with respect to internal energy will be equal. Clearly, this is an important and useful partial derivative property.

Note that $(\partial S/\partial U)_{V, N} = (\partial s/\partial u)_v$, so we can state the thermal equilibrium condition more simply as

$$\left[\left(\frac{\partial s}{\partial u} \right)_v \right]_A = \left[\left(\frac{\partial s}{\partial u} \right)_v \right]_B. \quad (6.38)$$

In chapter 3, we already introduced the property *temperature* to tell us if two systems are in thermal equilibrium (they are if $T_A = T_B$; otherwise they are not). Evidently, there must be a close relationship between temperature and $(\partial s/\partial u)_v$. We could in fact use any monotonic function of this derivative $f((\partial s/\partial u)_v)$ to *define* a thermodynamic temperature scale.

What function should we pick? Let's first look at how heat flows, from Fig. 6.3. Suppose we start with $U_A^{(0)} < U_A^*$ (to the left of the maximum in

Fig. 6.3. Then $dS/U_A > 0$, which from Eq. (6.35) means that $[(\partial s/\partial u)_v]_A > [(\partial s/\partial u)_v]_B$. As the system approaches equilibrium, U_A increases to U_A^* . Thus, heat must flow *into* A *from* B.

Now consider the other possibility: suppose we start with $U_A^{(0)} > U_A^*$ (to the right of the maximum in Fig. 6.3). Then $dS/U_A < 0$, which from Eq. (6.35) means that $[(\partial s/\partial u)_v]_A < [(\partial s/\partial u)_v]_B$. As the system approaches equilibrium, U_A decreases to U_A^* . Thus, heat must flow *from* A *into* B.

So it seems that heat flows *from* the system with lower $(\partial s/\partial u)_v$ to the system with higher $(\partial s/\partial u)_v$. By convention, we choose to define temperature so that heat flows from high to low temperature. So $(\partial s/\partial u)_v$ is not a suitable definition of thermodynamic temperature, if we want heat to flow from high to low temperature. But we could fix this up by taking the reciprocal: if $[(\partial s/\partial u)_v]_A > [(\partial s/\partial u)_v]_B$, then $[(\partial s/\partial u)_v]_A^{-1} < [(\partial s/\partial u)_v]_B^{-1}$. Therefore, we define the *thermodynamic temperature* as

$$\boxed{\frac{1}{T} = \left(\frac{\partial s}{\partial u} \right)_v} \quad (6.39)$$

Of course, Eq. (6.39) is not the only possibility. We could have put $1/T^2$ on the left instead of $1/T$, for example. What we need to show is that Eq. (6.39) is equivalent to temperature as measured on the ideal-gas temperature scale. It is, but we'll save the proof for the exercises.

6.8.3 Mechanical Equilibrium

Consider again the problem of mechanical equilibrium, first discussed in Chapter 3. Suppose a cylinder is divided into two parts A and B by a piston, which may freely move back and forth (Fig. 3.2). The cylinder is isolated from the environment, and has fixed total volume

$$V_{tot} = V_A + V_B \quad (6.40)$$

and total energy

$$U_{tot} = U_A + U_B. \quad (6.41)$$

The piston will be assumed to not only move, but also conduct heat. Therefore, at equilibrium, A and B will be in both thermal and mechanical equilibrium.

As before, according to the Entropy Maximum Principle, the equilibrium state will maximize S . But now we have two independent parameters to vary:

U_A and V_A . We seek to maximize the function

$$S = S_A(U_A, V_A, N_A) + S_B(U_{tot} - U_A, V_{tot} - V_A, N_B), \quad (6.42)$$

$$0 \leq U_A \leq U_{tot}; 0 \leq V_A \leq V_{tot}. \quad (6.43)$$

At the maximum (the equilibrium state),

$$\left(\frac{\partial S}{\partial U_A} \right) = 0 \text{ and } \left(\frac{\partial S}{\partial V_A} \right) = 0. \quad (6.44)$$

The derivative with respect to U_A is the same condition as we discussed above; it leads to the conclusion that $T_A = T_B$ at equilibrium.

But the derivative with respect to V_A gives us some new information:

$$\left(\frac{\partial S}{\partial V_A} \right) = \left(\frac{\partial S_A}{\partial V_A} \right)_{U_A, N_A} + \left(\frac{dV_B}{dV_A} \right) \left(\frac{\partial S_B}{\partial V_B} \right)_{U_B, N_B} \quad (6.45)$$

$$= \left(\frac{\partial S_A}{\partial V_A} \right)_{U_A, N_A} + (-1) \left(\frac{\partial S_B}{\partial V_B} \right)_{U_B, N_B}. \quad (6.46)$$

Here we used Eq. (6.40) to set $dV_B/dV_A = -1$.

Setting $(\partial S/\partial V_A) = 0$ at equilibrium, we find

$$\left[\left(\frac{\partial S}{\partial V} \right)_{U, N} \right]_A = \left[\left(\frac{\partial S}{\partial V} \right)_{U, N} \right]_B. \quad (6.47)$$

or

$$\left[\left(\frac{\partial s}{\partial v} \right)_u \right]_A = \left[\left(\frac{\partial s}{\partial v} \right)_u \right]_B. \quad (6.48)$$

Thus, when the piston stops moving and A and B are in both thermal and mechanical equilibrium, A and B will have the same value for $(\partial s/\partial v)_u$.

We define the *thermodynamic pressure* P as follows:

$$\boxed{\frac{P}{T} = \left(\frac{\partial s}{\partial v} \right)_u} \quad (6.49)$$

Then at equilibrium,

$$\left(\frac{P}{T} \right)_A = \left(\frac{P}{T} \right)_B. \quad (6.50)$$

Since the wall is heat-conducting, $T_A = T_B$. Therefore, if the piston can move, then we must have $P_A = P_B$ at equilibrium. The proof that the thermodynamic pressure equals the mechanical pressure is left as an exercise.

6.8.4 Diffusive Equilibrium

Finally, let's consider diffusive equilibrium. Imagine an isolated system divided into two parts A and B by a partition which is rigid but allows molecules to pass through it. It also conducts heat.

For now, we will say that substances A and B are the same chemically, but may be in different states (e.g. different pressures, temperatures, and/or phases). For example, we might place in A a cold, low-pressure mixture of liquid water and ice, and in B hot, high pressure water vapor.

The analysis to find the equilibrium state is very similar to what we did in the previous section, except that the two variable quantities are U_A and N_A . The total energy and number of molecules are fixed, so

$$U_B = U_{tot} - U_A \quad (6.51)$$

and

$$N_B = N_{tot} - N_A. \quad (6.52)$$

Following the same procedures as above, we find that in the equilibrium state,

$$\left[\left(\frac{\partial S}{\partial N} \right)_{U,V} \right]_A = \left[\left(\frac{\partial S}{\partial N} \right)_{U,V} \right]_B \quad (6.53)$$

We define the *chemical potential* μ by

$$\boxed{-\frac{\mu}{T} = \left(\frac{\partial S}{\partial N} \right)_{U,V}} \quad (6.54)$$

and therefore at equilibrium

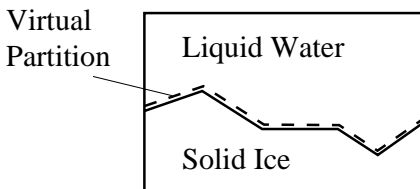
$$\left[\frac{\mu}{T} \right]_A = \left[\frac{\mu}{T} \right]_B. \quad (6.55)$$

Since the partition is heat-conducting, $T_A = T_B$, so the condition for equilibrium with respect to mass transfer is $\mu_A = \mu_B$. The negative sign is included in the definition so that mass will flow from the system with higher μ to the one with lower μ . The proof is left as an exercise.

Note that in this definition, the *total* values of U and V are being held constant (not \hat{u} and \hat{v} , or u and v). So as more matter is added (N increases) at constant U and V , $\hat{u} = U/N$ and $\hat{v} = V/N$ must decrease.

It is not necessarily the case that $P_A = P_B$ for diffusive equilibrium, even though $T_A = T_B$ and $\mu_A = \mu_B$. For example, if region A is a very small water

droplet and region B is the space around it containing water vapor, the liquid and vapor may be in equilibrium even though there is a pressure difference due to surface tension.



But in most cases other than small droplets or bubbles, the pressure is the same in all phases present in equilibrium. In this case, the phase-equilibrium conditions between, say, co-existing liquid and solid phases would be

$$T_\ell = T_s \quad (6.56)$$

$$P_\ell = P_s \quad (6.57)$$

$$\mu_\ell = \mu_s. \quad (6.58)$$

Clearly, these conditions would hold for any other type of phase equilibrium also (liquid/vapor, solid/vapor).

6.9 The Value of Boltzmann's Constant

We have not yet said what value should be assigned to k_B in $S = k_B \ln \Omega$. We have seen now that $1/T = (\partial S / \partial U)_{V,N}$, so the units of entropy must be [energy]/[temperature]. If we wish to use SI units, then, entropy must have units of J/K, and specific entropy J/kg-K. Since $\ln \Omega$ is dimensionless, we must then give k_B the units of J/K.

Using the expression we have derived for the entropy of a monatomic, ideal gas along with the definitions of thermodynamic temperature and pressure, it is possible to prove that for an ideal gas

$$PV = Nk_B T. \quad (6.59)$$

Converting from units of molecules to units of moles,

$$PV = (N/N_A)(N_A k_B)T. \quad (6.60)$$

where $N_A = 6.023 \times 10^{26} \text{ kmol}^{-1}$ is *Avagadro's Number*. This may be written in terms of molar volume as

$$\frac{P\hat{v}}{T} = N_A k_B \quad (6.61)$$

Experimentally, we know that this quantity equals 8314.4 J/kmol-K in the low-density, ideal gas limit (Chapter 4).

The experimental result uses temperature measured on the Kelvin scale (an ideal gas scale). If we wish the thermodynamic temperature to coincide with the Kelvin scale, then we must choose

$$k_B = (8314.3 \text{ J/kmol-K})(1 \text{ kmol}/6.023 \times 10^{26}). \quad (6.62)$$

Therefore,

$$k_B = 1.38 \times 10^{-23} \text{ J/K} \quad (6.63)$$

6.10 The Gibbs Equation

Often, we want to know how the entropy of a substance in equilibrium changes if its internal energy, volume, and number of molecules change by a small amount. Differentiating the function $S(U, V, N)$ yields

$$dS = \left(\frac{\partial S}{\partial U} \right)_{V,N} dU + \left(\frac{\partial S}{\partial V} \right)_{U,N} dV + \left(\frac{\partial S}{\partial N} \right)_{U,V} dN. \quad (6.64)$$

Using the definitions for thermodynamic temperature, pressure, and chemical potential, this equation becomes

$$dS = \left(\frac{1}{T} \right) dU + \left(\frac{P}{T} \right) dV - \left(\frac{\mu}{T} \right) dN. \quad (6.65)$$

This equation is known as the *Gibbs equation*. It expresses a relationship among equilibrium *properties* of a substance, and doesn't depend on any particular type of process for its validity (e.g. reversible, etc.).

We can write the Gibbs equation on a per-mole or per-kg basis also. In this case, $dN = 0$, so

$$d\hat{s} = \frac{1}{T} d\hat{u} + \frac{P}{T} d\hat{v}, \quad (6.66)$$

and

$$ds = \frac{1}{T} du + \frac{P}{T} dv. \quad (6.67)$$

We can also solve Eq. (6.65) for dU :

$$\boxed{dU = TdS - PdV + \mu dN.} \quad (6.68)$$

This equation is completely equivalent to Eq. (6.65).

Now the property U exists, so it must be that Eq. (6.68) is what we would get if we differentiated the function $U(S, V, N)$. Therefore, we can identify

$$T = \left(\frac{\partial U}{\partial S} \right)_{V,N}; \quad P = - \left(\frac{\partial U}{\partial V} \right)_{S,N}; \quad \mu = \left(\frac{\partial U}{\partial N} \right)_{S,V}. \quad (6.69)$$

These equations are alternative ways to define the thermodynamic temperature, pressure, and chemical potential.

Applying Eq. (6.68) to a unit mass

$$du = Tds - Pdv, \quad (6.70)$$

and so

$$T = \left(\frac{\partial u}{\partial s} \right)_v; \quad P = - \left(\frac{\partial u}{\partial v} \right)_s. \quad (6.71)$$

Adding a “hat” would give the corresponding equations on a molar basis.

6.11 Macroscopic Determination of the Entropy

For most substances, it is virtually impossible to figure out $\Omega(U, V, N)$. Fortunately, the Gibbs equation gives us another way to determine S . If the equations of state $T(u, v)$ and $P(u, v)$ are known, then these may be inserted into Gibbs equation. For a unit mass,

$$ds = \frac{1}{T(u, v)} du + \frac{P(u, v)}{T(u, v)} dv. \quad (6.72)$$

This can be integrated along some path in the (u, v) plane, starting from some point (u_0, v_0) :

$$s(u, v) - s(u_0, v_0) = \int_{path} \left[\frac{1}{T(u, v)} du + \frac{P(u, v)}{T(u, v)} dv \right]. \quad (6.73)$$

Since we know the function $s(u, v)$ exists, the value of the integral won't depend on the particular path chosen. We can choose any convenient path.

This procedure, in slightly modified form, is how the entropy is usually determined for real substances. We will discuss this procedure more in Chapter 8.

We cannot determine the entropy of the initial state $s_0 = s(u_0, v_0)$. We have two choices. We can simply choose a convenient “reference” state (u_0, v_0) and assign the entropy in this state an arbitrary value (for example, zero). Then all entropies $s(u, v)$ are really relative to s_0 , the entropy in the reference state.

Or we can choose the reference state so that the ideal gas approximation holds for the reference state. In this case, we can use expressions from statistical mechanics for the absolute entropy of an ideal gas to get s_0 . If the substance is monatomic, then Eq. (6.30) would be the right expression to evaluate s_0 . If it is a molecular gas, the expression is slightly more complicated, but s_0 still can be computed.

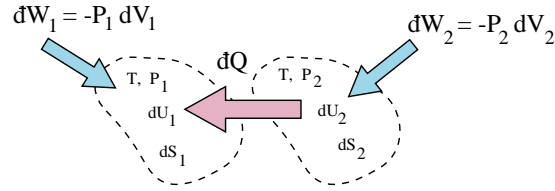
The choice of how to get s_0 differs from one source of thermodynamic data to another. Care must be taken when comparing s (or h or u) from different sources. In TPX, for most substances the reference state is taken to be saturated liquid at the lowest temperature for which the data are valid. Both the entropy and the enthalpy are assigned the value zero in the reference state. Of course, the value of s_0 is irrelevant in computing entropy differences, as long as both entropy values are taken from the same source.

6.12 Entropy Flow with Heat

Consider two systems, each composed of some simple compressible substance which is internally in equilibrium. System 1 has temperature T_1 , and system 2 has temperature $T_2 = T_1 + \Delta T$. We bring them into contact, and heat Q flows from 2 to 1.

Let’s take the limit $\Delta T \rightarrow 0$. In this limit, the temperature difference becomes infinitesimal (dT) and the process approaches heat transfer at constant T . Of course, as $\Delta T \rightarrow 0$, the time t required to transfer a finite amount of heat Q increases proportional to $Q/\Delta T$. We’ll transfer only an infinitesimal amount of heat dQ (which only takes finite time).

The two systems start out internally in equilibrium. Since the heat transferred is infinitesimal, we can assume that the systems stay in internal equilibrium during this process, with the result that their intensive properties (T and P) don’t change, and their extensive properties (U , V , and S) change only infinitesimally. This also means that there are no internal irreversibilities (viscous forces, electrical power dissipation, etc.) since these only exist when a system is not in internal equilibrium. Therefore, no entropy is being produced in either system, and this process is reversible.



Writing the first law for the process,

$$dU_1 = dQ - P_1 dV_1, \quad (6.74)$$

and

$$dU_2 = -dQ - P_2 dV_2. \quad (6.75)$$

Since 1 and 2 are not necessarily in mechanical equilibrium with one another, we allow P_1 and P_2 to differ. (For example, they may be sitting in environments at different pressures, and are connected by a copper wire allowing heat to flow.) Note that we have used the assumption of reversibility in writing the work terms as $-PdV$.

We can also write the Gibbs equation for each substance:

$$dS_1 = \frac{1}{T} dU_1 + \frac{P_1}{T} dV_1 \quad (6.76)$$

and

$$dS_2 = \frac{1}{T} dU_2 + \frac{P_2}{T} dV_2. \quad (6.77)$$

Since $dN_1 = dN_2 = 0$, we have left off the $-(\mu/T)dN$ term. Substituting for dU from the first-law expressions, we find

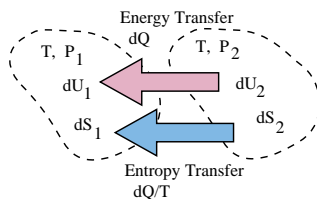
$$dS_1 = \frac{dQ}{T} \quad (6.78)$$

and

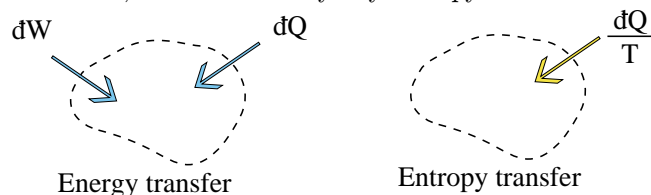
$$dS_2 = -\frac{dQ}{T}. \quad (6.79)$$

Although no entropy was produced by this process of heat transfer at constant T , the entropy of system 1 increased, and that of system 2 decreased by exactly the same amount. It must be, then, that entropy was *transferred* from 2 to 1 along with the energy transfer dQ . Therefore,

When heat dQ enters or leaves a system at temperature T , it carries with it entropy in the amount dQ/T .



Note that the work terms canceled out of the above equations. Evidently, entropy transfer is associated only with energy transfer as heat – when energy is transferred as work, it does not carry any entropy with it.



Problems

6.1 A rectangular box contains N molecules. Suppose that each molecule can be on either side of the box with equal probability.

1. What is the probability that all N molecules are on one side? Compute this number for $N = 1, 2, 20,$ and 10^{20} .
2. Astronomers believe the universe is between 8 and 13 billion years old. Suppose you examine this system once every second, and the molecules are distributed completely randomly each time. Estimate the largest value N can have so that on average you will find all molecules on one side at least once in the age of the universe.

6.2 Using Eq. (6.30) and the definitions of thermodynamic temperature and pressure, show that for a monatomic gas of N atoms

$$U = \frac{3}{2}Nk_B T$$

and

$$PV = Nk_B T.$$

6.3 Prove that the thermodynamic pressure P equals the mechanical pressure $P_{mech} = F/A$. To do so, consider a quasi-static, reversible, adiabatic compression in which the volume changes by dV . The work done on the system is $dW_{qs} = -P_{mech}dV$. Use the first law, and the fact that $dS = 0$ for this process. (Hint: write $ds(u, v)$ in terms of du and dv .)

6.4 Show that if $\mu_A \neq \mu_B$, mass flows from the one with higher μ to the one with lower μ .

6.5 Using

$$S(U, V, N) = N\hat{s}(\hat{u}, \hat{v}) = N\hat{s}(U/N, V/N)$$

show that

$$\mu = \hat{u} + P\hat{v} - T\hat{s}.$$

The right-hand side of this equation is the definition of the molar Gibbs free energy, and thus $\mu = \hat{g}$.

(N may be interpreted as either the number of molecules or the number of moles, and all “molar” quantities as per molecule or per mole, whichever is more convenient.)

6.6 Consider a crystal composed of \mathcal{N} atoms, each of which can occupy one of *two* quantum states. The first state (the ground state) has energy zero, and the other one (the upper state) has energy $\epsilon > 0$.

1. Suppose the total internal energy of the crystal is U , where U is an integral multiple of ϵ . Show that the number of different microscopic quantum states of the crystal which have this energy U is given by

$$\Omega = \frac{N!}{\left(\frac{U}{\epsilon}\right)! \left(N - \frac{U}{\epsilon}\right)!} \quad (6.80)$$

Neglect any other forms of energy, such as vibration.

2. In the limit $N \rightarrow \infty$, determine the entropy of the system at equilibrium. (Note that for large n , $\ln(n!) \approx n \ln n - n$.)
3. From the entropy and the definition of thermodynamic temperature, determine the thermal equation of state $U(T, N)$.

6.7 We showed that when two phases 1 and 2 of a substance are in equilibrium with one another, $P_1 = P_2$, $T_1 = T_2$, and $\mu_1 = \mu_2$. Using TPX, verify that $\mu_1 = \mu_2$ for saturated liquid water and water vapor at 300 K.

6.8 A mixture of saturated liquid nitrogen and saturated vapor at 1 MPa is contained on one side of an insulated container which is divided in half by a partition. The initial vapor mass fraction is 0.7. The other side is evacuated. Now the partition is broken, and the nitrogen expands unrestrained to fill the container, doubling its volume. Using TPX:

1. Find the new pressure and temperature of the system once equilibrium has been re-established.
2. Determine how much entropy was produced by this process, per unit mass of nitrogen.

CHAPTER 7

ENTROPY ACCOUNTING AND APPLICATIONS

7.1 Introduction

In the last chapter, we found two important properties of the entropy:

1. Entropy can be *produced* by irreversible processes, but can never be destroyed: $\mathcal{P}_s \geq 0$ for every possible process.
2. When heat Q enters or leaves a system at temperature T , it carries entropy in the amount Q/T with it.

With these two results, it is possible to write down entropy accounting expressions for closed or open systems. In this chapter, we will develop these expressions, and look at several types of problems which can be solved by combining energy and entropy accounting.

7.2 Some General Accounting Principles

We must take some care in developing entropy accounting expressions, since, unlike energy, entropy is not conserved. First let's consider in general how to do accounting for things which can be produced. Consider the monthly statement you might receive from a bank if you have an interest-bearing bank account. The account balance at the end of a month is related to the previous month's balance by

$$(\text{new balance}) = (\text{old balance}) + (\text{deposits}) - (\text{withdrawals}) + (\text{interest}). \quad (7.1)$$

The account balance is the amount of money “stored” in the account; the deposits represent an “inflow” of money, the withdrawals an “outflow” of money, and we can think of the interest as money “produced” within the account during the month, since it appears in the account even though you didn't deposit it. Therefore, the accounting principle for a bank account is

$$\Delta(\text{stored money}) = (\text{money inflow}) - (\text{money outflow}) + (\text{money production}). \quad (7.2)$$

Of course, this basic accounting principle applies to *anything* which can be stored, transported, and produced. Suppose a factory manufactures a product. The change in the inventory stored at the factory over some period of time equals the number of products produced in that period, plus the number returned from stores (inflow to the factory), minus the number shipped to stores (outflow from the factory). So Eq. (7.2) can be written more generally:

$$\Delta(\text{storage}) = (\text{inflow}) - (\text{outflow}) + (\text{production}). \quad (7.3)$$

This general accounting principle can be applied to the storage, transportation, and production of physical quantities like mass, momentum, energy, and entropy. For mass and energy, the production term is always zero, since these are conserved. Therefore, the energy balances we developed in Chapters 2 and 4 were always of the form

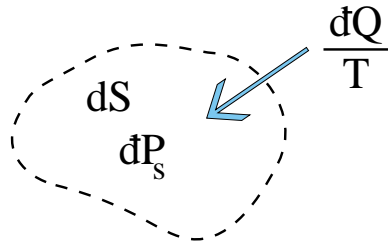
$$\Delta(\text{energy stored in system}) = (\text{energy inflow}) - (\text{energy outflow}). \quad (7.4)$$

For entropy, the production term is non-zero, so we have

$$\begin{aligned} \Delta(\text{entropy stored in system}) &= (\text{entropy inflow}) - (\text{entropy outflow}) \\ &+ (\text{entropy production}). \end{aligned} \quad (7.5)$$

7.3 Entropy Accounting for a Closed System

For a closed system, the only way entropy can enter or leave the system is with heat which enters or leaves. The simplest case occurs when heat enters at a single temperature, as shown below.



During elapsed time dt , heat dQ crosses the system boundary where the temperature is T . We saw in the last chapter that when heat dQ enters a system at temperature T , it brings with it entropy in the amount dQ/T . Therefore, the entropy inflow is dQ/T , and there is no entropy outflow.

During the elapsed time dt , entropy in the amount $d\mathcal{P}_s$ may have been also *produced* within the system due to internal irreversible processes (friction, viscous fluid flow, electrical current flow, unrestrained expansion, etc.). Therefore, according to Eq. (7.5), the increase in the amount of entropy stored within the system is

$$dS = \frac{dQ}{T} + d\mathcal{P}_s. \tag{7.6}$$

Note that work dW might have also been done on or by the system during dt , but since energy transfer as work does not carry entropy, it does not enter into the entropy accounting expression.

Since the second law requires $d\mathcal{P}_s \geq 0$,

$$dS \geq \frac{dQ}{T}. \tag{7.7}$$

Thus, the system entropy increases by *more* than the amount which enters with heat; the difference is what is produced inside. Equation (7.7) also holds if $dQ < 0$, in which case it says the the entropy decreases by less than the amount of entropy which leaves with heat. For a reversible process, Eq. (7.7) reduces to

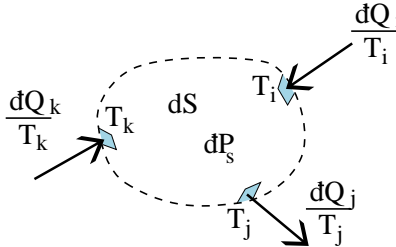
$$dQ_{rev} = TdS_{rev}. \tag{7.8}$$

From Eq. (7.6), a process which is both adiabatic ($dQ = 0$) and reversible ($d\mathcal{P}_s = 0$) would not change the entropy ($dS = 0$). Therefore,

the entropy is constant during any reversible, adiabatic process in a closed system.

We call processes which occur at constant entropy *isentropic*.

Equation (7.6) can be easily generalized to the case where heat flows across the boundary at various spots, each of which may have a different temperature (which might even be time-dependent).



In time dt , the total entropy S of the system changes due to the entropy entering or leaving with heat, and the entropy produced by irreversible processes

inside. If heat enters through I places, and leaves through O places, then

$$dS = \sum_{\text{in}} \left(\frac{dQ_{\text{in}}}{T_{\text{in}}} \right) - \sum_{\text{out}} \left(\frac{dQ_{\text{out}}}{T_{\text{out}}} \right) + d\mathcal{P}_s. \quad (7.9)$$

Note that the dQ/T terms are evaluated *at the temperature on the system boundary where the heat flow is occurring*.

Equation (7.9) can also be written on a per-unit-time basis (a rate basis). Dividing by dt ,

$$\frac{dS}{dt} = \sum_{\text{in}} \left(\frac{\dot{Q}_{\text{in}}}{T_{\text{in}}} \right) - \sum_{\text{out}} \left(\frac{\dot{Q}_{\text{out}}}{T_{\text{out}}} \right) + \dot{\mathcal{P}}_s. \quad (7.10)$$

If the temperatures on the boundaries where heat is entering or leaving are constant in time, this equation can be integrated for a finite time interval $(0, t)$:

$$\Delta S = S(t) - S(0) = \sum_{\text{in}} \left(\frac{Q_{\text{in}}}{T_{\text{in}}} \right) - \sum_{\text{out}} \left(\frac{Q_{\text{out}}}{T_{\text{out}}} \right) + \mathcal{P}_s. \quad (7.11)$$

Here each Q is just the total heat which enters: $Q = \int_0^t dQ$.

Equations (7.9) – (7.10) all are simply statements of entropy accounting. They equate the net change in the total entropy of a system to the *net* entropy inflow (inflow - outflow) plus the amount of entropy produced inside the system. They apply to any closed (control mass) system.

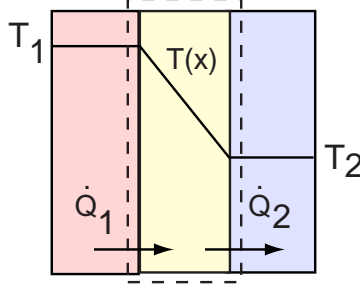
In many (but not all) processes we will consider, the system is in steady state. In this case, the entropy contained within the system is not changing, so on a rate basis $dS/dt = 0$, or on a total time basis $\Delta S = 0$. For a steady-state problem, Eq. (7.10) reduces to

$$\sum_{\text{in}} \left(\frac{\dot{Q}_{\text{in}}}{T_{\text{in}}} \right) - \sum_{\text{out}} \left(\frac{\dot{Q}_{\text{out}}}{T_{\text{out}}} \right) + \dot{\mathcal{P}}_s = 0. \quad (7.12)$$

A great deal can be learned about real processes by carrying out energy and entropy accounting. Let's consider a few examples of processes occurring in closed systems.

7.4 Heat Transfer Through Finite ΔT

Consider the situation shown below. Heat is flowing from a high-temperature region at T_1 to low-temperature region at T_2 through a thin slab of insulation material, which has thickness L and area A . The temperature is everywhere constant in time, and the insulation is in steady state.



From the theory of heat conduction in solids, the heat flow through the insulation is related to the temperature difference across it by

$$\dot{Q} = \frac{\kappa A}{L} (T_1 - T_2), \quad (7.13)$$

where κ is a material property called the *thermal conductivity* (SI units: W/m-K). This equation is analogous to $I = V/R$ for current flow through a resistor, with \dot{Q} playing the role of current, ΔT the potential difference, and $L/\kappa A$ the resistance. For this reason, $L/\kappa A$ is called the *thermal resistance* R_{th} of the insulation.

We will define a system to consist of the insulation, with the left system boundary at T_1 and the right one at T_2 . We will assume the heat flow is one dimensional, so heat only crosses the system boundaries perpendicular to x . Since the insulation is in steady state, no energy or entropy are building up in the system: $dU/dt = 0$ and $dS/dt = 0$. Therefore, the energy balance reduces to

$$\dot{Q}_h = \dot{Q}_c, \quad (7.14)$$

and so we may drop the subscripts and call the heat transfer rate \dot{Q} .

Applying Eq. (7.10) with $dS/dt = 0$ and one heat inflow and one heat outflow, the entropy accounting expression becomes

$$0 = \frac{\dot{Q}}{T_1} - \frac{\dot{Q}}{T_2} + \dot{\mathcal{P}}_s. \quad (7.15)$$

Solving for $\dot{\mathcal{P}}_s$,

$$\dot{\mathcal{P}}_s = \dot{Q} \left(\frac{1}{T_2} - \frac{1}{T_1} \right). \quad (7.16)$$

Since $T_1 > T_2$, $\dot{\mathcal{P}}_s > 0$, in accord with the Second Law. Note that as $T_2 \rightarrow T_1$, the ratio $\dot{\mathcal{P}}_s/\dot{Q} \rightarrow 0$. Therefore, the amount of entropy produced per unit of heat transferred can be made arbitrarily small by making $\Delta T = T_1 - T_2$ sufficiently small. We conclude that heat transfer through a finite ΔT is *irreversible* (produces entropy), but in the limit as T_2 approaches T_1 , the process approaches reversibility.

Of course, from Eq. (7.13), decreasing ΔT decreases the heat transfer rate, and as $T_2 \rightarrow T_1$ $\dot{Q} \rightarrow 0$. Therefore, in the reversible limit transferring a finite amount of heat would take infinite time.

7.5 Heat Engines

The original engineering problem which led to the development of thermodynamics was how to efficiently use the thermal energy of a fire (produced by burning wood or coal) to do useful work, for example to turn the wheels of a locomotive.

It's easy to design a process which can take in heat from a high-temperature source and produce work for a little while, but much harder to design one which can go on doing it indefinitely, which is what is really desired. For example, gas in a cylinder fitted with a piston will expand when heated, pushing the piston, doing work against an external load. This is fine until the piston has reached the top of the cylinder, at which time the process must stop. Figuring out how to get the piston back to its initial position to start the "cycle" again without having to put more work in than you get out is the challenge.

A *heat engine* is a device which operates either steadily or in a cycle, which takes in energy as heat and converts at least part of it into useful work. Heat engines are a vital part of modern technology. Most current and proposed electric power plants are heat engines, which take in heat at high temperature given off by a chemical or nuclear reaction, and convert part of this heat into electrical work delivered to the grid. Even very sophisticated concepts such as nuclear fusion are simply different sources of heat — a heat engine of some sort is still needed to convert this to useful work.¹ Automobile, aircraft, and rocket engines are other examples of practical heat engines.

Since a heat engine operates either continuously or in a cycle, in which it returns to its initial state after some time period T , we can use a steady-

¹One promising power plant concept which is not a heat engine is a fuel cell, which converts the chemical energy in fuel directly into electrical energy, without first converting it into thermal energy in a flame. Thermodynamics still governs fuel cell performance, however, as we'll discuss later.

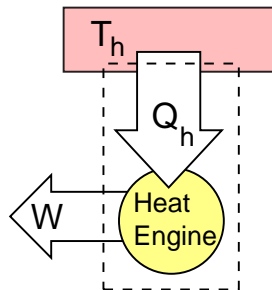


Figure 7.1: A hypothetical one-temperature heat engine.

state analysis to study heat engines. For continuous operation, $dE/dt = 0$ and $dS/dt = 0$, while for cyclic operation we will choose the time period T for the analysis, in which case $\Delta E = 0$ and $\Delta S = 0$. In the following, we will assume continuous operation, but the results apply equally for cyclic operation.

To analyze heat engines, it is useful to introduce the conceptual device of a *thermal reservoir*. A thermal reservoir is defined as a substance with a very large heat capacity, so that heat can be taken from it or added to it without significantly perturbing its temperature. Furthermore, it has a very large thermal conductivity, so that transferring heat to or from it causes only negligible temperature differences within it [Eq. (7.13)]. A large block of copper or a large lake would *approximate* a thermal reservoir.

In the limit where the heat capacity and thermal conductivity become infinite, the temperature of a thermal reservoir remains perfectly uniform, and constant in time, no matter how much heat is added or removed. In this limit, heat flow *within* the reservoir is perfectly reversible. (The electrical analog of a thermal reservoir would be an infinitely-large perfect conductor, which would have constant electrical potential even if current were being drawn from it.)

7.5.1 A (Hypothetical) One-Temperature Heat Engine

Consider the heat engine shown in Fig. 7.1. The engine takes in heat \dot{Q}_h from a thermal reservoir at temperature T_h , and outputs power \dot{W} . Since the engine is in steady state, energy accounting yields

$$\dot{Q}_h = \dot{W} \quad (7.17)$$

and entropy accounting [Eq. (7.12)] yields

$$\frac{\dot{Q}_h}{T_h} + \dot{\mathcal{P}}_s = 0. \quad (7.18)$$

Since the Second Law requires $\dot{\mathcal{P}}_s \geq 0$ and also $T_h > 0$, Eq. (7.18) requires $\dot{Q}_h < 0$, and therefore $\dot{W} < 0$.

Thus, it is impossible to produce positive power *by any process* which only takes in heat at a single temperature T_h .

It is not hard to see why this doesn't work. Entropy is entering the engine at a rate \dot{Q}_h/T_h , but no entropy is leaving. (Remember work does not carry entropy with it.) Thus, operation of this engine would require entropy to be destroyed inside the engine, but that violates the Second Law. Therefore, we conclude

a one-temperature heat engine cannot be built.

This statement is sometimes called the *Kelvin-Planck statement of the Second Law*. In some classical treatments of thermodynamics which make no reference to the existence of atoms or microstates, this is introduced as a *postulate*. With this postulate, it is possible to work “backwards” (from our point of view) using some very clever thought experiments to prove that there must exist a property, the entropy, which has all of the macroscopic characteristics we've discussed already.

7.5.2 A Two-Temperature Heat Engine

To allow production of positive power, some means of continually removing entropy from the engine must be provided. The only way to do this in a closed system is to remove some heat from the engine, which will carry entropy with it. Since we want to convert at least part of \dot{Q}_h to \dot{W} , we can only afford to remove a portion of \dot{Q}_h . But we need to remove *at least as much* entropy as entered (since some might have also been produced inside), so we need to remove heat in such a way that the ratio (entropy flow)/(heat flow) is larger going out than coming in. Since the amount of entropy carried with a unit amount of heat flow is $1/T$, this suggests we should try rejecting some heat to a reservoir at a cold temperature T_c (large $1/T_c$).

Doing energy and entropy accounting for the two-temperature heat engine in Fig. 7.2, we have

$$\dot{Q}_h = \dot{W} + \dot{Q}_c, \quad (7.19)$$

and

$$\frac{\dot{Q}_h}{T_h} + \dot{\mathcal{P}}_s = \frac{\dot{Q}_c}{T_c}. \quad (7.20)$$

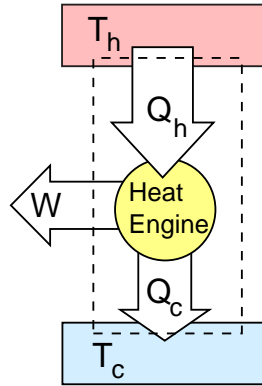


Figure 7.2: A two-temperature heat engine.

Eliminating \dot{Q}_c between these equations,

$$\dot{W} = \dot{Q}_h \left(1 - \frac{T_c}{T_h} \right) - T_c \dot{\mathcal{P}}_s. \quad (7.21)$$

We see that positive power *can* be produced by a two-temperature heat engine. Since the second law requires $\dot{\mathcal{P}}_s \geq 0$, the *maximum* power for a given \dot{Q}_h is produced for $\dot{\mathcal{P}}_s = 0$. Therefore a reversible heat engine produces the most power possible for a given heat transfer rate \dot{Q}_h , and given T_h and T_c . For a real engine, the term $-T_c \dot{\mathcal{P}}_s$ represents power “lost” due to irreversibilities.

The *thermal efficiency* η of a heat engine is defined as

$$\eta = \frac{\dot{W}}{\dot{Q}_h}. \quad (7.22)$$

This definition makes sense, since the desired output is \dot{W} , and typically it is \dot{Q}_h which you must “pay for” in the cost of fuel burned to produce it.

A reversible heat engine has the highest thermal efficiency of any heat engine operating between T_h and T_c . From Eq. (7.21), this limiting value of the efficiency is

$$\eta^{max} = \left(1 - \frac{T_c}{T_h} \right), \quad (7.23)$$

This upper-limit efficiency is often called the *Carnot efficiency*, after Sadi Carnot, a French engineer who was the first person to systematically analyze heat engines in the 19th century. Note that it depends only on T_h and T_c , and is the same for all possible heat engines operating between these temperatures.

Most practical engines and generators, including automotive engines, aircraft engines, and electric power generation facilities, can be approximated as

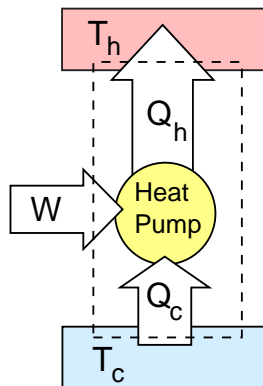


Figure 7.3: A heat pump or refrigerator “pumps” heat from low temperature to high temperature.

two-temperature heat engines. In order to produce a fuel-efficient engine, η should be as large as practically possible. Of course, the first thing to do is to try to eliminate as much unnecessary entropy production within the engine as possible, to minimize the term $-T_c \dot{\mathcal{P}}_s$. But once this has been done to the extent practical, further increases in η require either increasing T_h , or decreasing T_c .

For most engines, T_c is the temperature of the ambient surroundings and can't be easily varied. This leaves increasing T_h . A flame can easily be used to produce a temperature of, say, 2000 K. If we use $T_h = 2000$ K and $T_c = 300$ K, we find $\eta^{max} = 1 - 300/2000 = 0.85$. This is a very high efficiency, but unfortunately not many materials can withstand this temperature for long. Metals (which turbines are usually made of) fail at much lower temperatures, so the peak temperature in a practical engine is usually limited by materials constraints. Some of the most significant developments in recent years have been ceramic components in engines which can take much higher temperatures and allow designing engines with higher T_h .

7.6 Heat Pumps and Refrigerators

Although heat will not spontaneously flow from low temperature to high, a process can be designed which takes in heat at low temperature and expels heat at high temperature. However, such a process will not run by itself — it requires work input to “pump” the heat from low to high temperature.

A *heat pump* is a device which operates either continuously or in a cycle,

which takes in heat at low temperature and expels it at higher temperature (Fig. 7.3). A refrigerator is a common example of a heat pump. Heat continually leaks into the refrigerated compartment from the surroundings through the insulation, or when the door is opened. This heat must be removed and expelled to the surroundings at higher temperature to keep the refrigerated space cold.

We may analyze heat pumps assuming steady-state, as we did to analyze heat engines, since they operate continuously ($dS/dt = 0$) or in a cycle ($S(T) - S(0) = 0$). For the heat pump shown in Fig. 7.3, energy accounting yields

$$\dot{W} + \dot{Q}_c = \dot{Q}_h, \quad (7.24)$$

and entropy accounting yields

$$\frac{\dot{Q}_c}{T_c} - \frac{\dot{Q}_h}{T_h} + \dot{\mathcal{P}}_s = 0. \quad (7.25)$$

If the heat pump is used as a refrigerator, what we really care about is how much power \dot{W} we have to supply to run the heat pump for a given rate of heat removal \dot{Q}_c from the cold space. In this case, eliminating \dot{Q}_h between equations (7.24) and (7.25),

$$\dot{W} = \dot{Q}_c \left(\frac{T_h}{T_c} - 1 \right) + T_h \dot{\mathcal{P}}_s. \quad (7.26)$$

Since the Second Law requires $\dot{\mathcal{P}}_s \geq 0$, clearly a reversible heat pump ($\dot{\mathcal{P}}_s = 0$) requires the least power input for given \dot{Q}_c , T_c , and T_h . Note that if \dot{W} were set to zero in Eq. (7.26), there would be no solution with $\dot{Q}_c > 0$, even if the heat pump were reversible. Therefore,

a zero-work heat pump is impossible to build.

This statement is sometimes called the *Clausius statement of the Second Law*. Like the Kelvin-Planck statement, it is sometimes introduced as a postulate in purely macroscopic treatments of thermodynamics, from which the existence of entropy is inferred.

The second law requires $\dot{\mathcal{P}}_s \geq 0$, so for a given \dot{Q}_c , the *minimum* power required is obtained when $\dot{\mathcal{P}}_s = 0$. Therefore, a hypothetical heat pump which produces no entropy (a reversible heat pump) would require the least work:

$$\dot{W}_{min} = \dot{Q}_c \left(\frac{T_h}{T_c} - 1 \right). \quad (7.27)$$

Since no real refrigerator is truly reversible, the actual power required to run it will be greater than this value. The term $T_h \dot{\mathcal{P}}_s$ in Eq. (7.26) represents the

“extra” work a real refrigerator does, due to irreversibilities such as friction and finite-temperature-difference heat transfer. Design of an energy-efficient refrigerator would require a careful thermodynamic analysis to find where entropy is being produced, and how entropy production could be minimized.

The performance of a refrigerator is characterized by the *coefficient of performance*, defined as the watts of heat removed from the cold space per watt of power input: $\text{COP}_{ref} = \dot{Q}_c/\dot{W}$. From Eq. (7.27), the maximum value is

$$\text{COP}_{ref}^{(max)} = \frac{1}{T_h/T_c - 1}. \quad (7.28)$$

Note that this can be greater than one: more energy can be removed as heat than is required as input to run the process.

An interesting thing about this equation is that it only depends on T_h and T_c , not on the details of how we implement the heat pump – *every* possible refrigerator design has the same limiting COP as it approaches reversibility, no matter whether the working substance is a fluid, or a magnetic crystal, or anything else.

If a heat pump is used to heat a house by pumping heat \dot{Q}_h into the house at T_h from the outside surroundings at T_c , then eqs. (7.24) and (7.25) still apply, but we are most concerned with the ratio \dot{Q}_h/\dot{W} . If we eliminate \dot{Q}_c between equations (7.24) and (7.25), we find

$$\dot{W} = \dot{Q}_h \left(1 - \frac{T_c}{T_h} \right) + T_c \dot{\mathcal{P}}_s. \quad (7.29)$$

The performance parameter which is used to characterize heat pumps used for heating is also called the coefficient of performance, but it is defined slightly differently:

$$(\text{COP})_{hp} = \frac{\dot{Q}_h}{\dot{W}}. \quad (7.30)$$

For a reversible heat pump the coefficient of performance is maximized:

$$(\text{COP})_{hp}^{(max)} = \frac{1}{1 - T_c/T_h}. \quad (7.31)$$

7.6.1 The Carnot Cycle

We’ve shown based on very general considerations that no heat engine can have an efficiency greater than $1 - T_c/T_h$. In deriving this result, we didn’t have to say anything about *how* the engine operates – it is a general result, which applies to all conceivable heat engines.

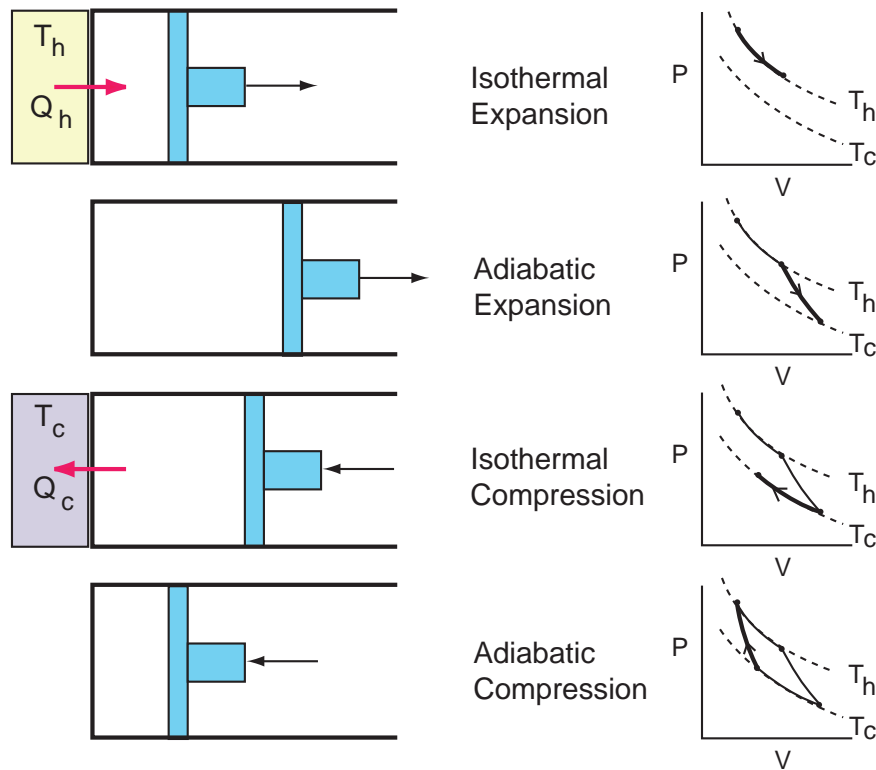


Figure 7.4: The Carnot Cycle two-temperature heat engine. Each step is performed quasi-statically, and therefore the process is reversible. By reversing the cycle, it also can function as a heat pump.

It's instructive to consider a simple idealized, reversible cycle which achieves this limiting efficiency. While not a very practical design for an engine, this cycle, known as the *Carnot cycle*, is conceptually the simplest possible reversible two-temperature heat engine. By running it in reverse, it functions as a heat pump.

In a Carnot-cycle engine, a fluid (not necessarily an ideal gas) is placed in a piston/cylinder system. The following 4 steps are executed in a reversible manner in sequence:

1. The cylinder is placed in contact with a thermal reservoir at T_h , and the fluid is expanded isothermally, drawing in heat Q_h .
2. Now the thermal reservoir is removed, and the expansion continues adiabatically. Since no heat is being added in this step, but the fluid is doing

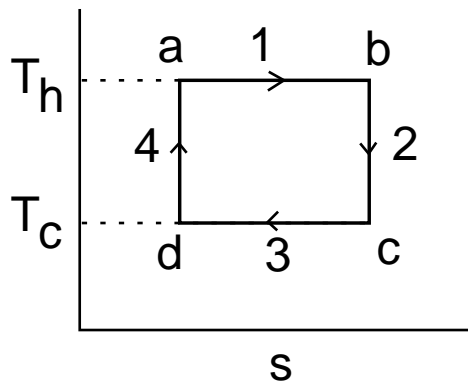


Figure 7.5: The Carnot cycle as represented on a $T - S$ plot.

work, its temperature drops. This step continues until the temperature drops to T_c .

3. Now a thermal reservoir at T_c is placed in contact with the cylinder, and the the fluid is isothermally compressed, expelling heat Q_c .
4. Finally, the thermal reservoir is removed, and the compression continues adiabatically, ending with the piston reaching its initial position and the fluid temperature back at T_h .

In Fig. 7.4, this process is represented on a $P - V$ plot, assuming the fluid is a gas. Since the process is carried out quasi-statically, the fluid remains in equilibrium at all times during the cycle. Therefore, its state may be represented at every time by a point in the $P - V$ plane, and the entire cycle is a closed curve in the $P - V$ plane.

A simpler representation of this cycle is a temperature-entropy ($T - S$) plot, as shown in Fig. 7.5.

The Carnot cycle traces out a rectangle on a $T - S$ plot, and this is true no matter what the fluid is. Consider first steps 2 and 4, which are both reversible and adiabatic. We have shown that reversible, adiabatic processes are isentropic, so the entropy of the fluid does not change during steps 2 and 4. They must appear as vertical lines on a $T - S$ plot.

Steps 1 and 3 are isothermal, so they must appear as horizontal lines on a $T - S$ plot. Step 1 draws in heat Q_h , which brings in entropy Q_h/T_h with it. Since the process is reversible, $\mathcal{P}_s = 0$ and therefore

$$\Delta S_1 = S_b - S_a = \frac{Q_h}{T_h}. \quad (7.32)$$

For step 3, heat Q_c is rejected and the entropy of the system decreases by Q_c/T_c :

$$\Delta S_3 = S_a - S_b = -\frac{Q_c}{T_c}. \quad (7.33)$$

Since step 3 brings the system entropy back to its initial value (it is unchanged by step 4), we must have

$$\frac{Q_c}{T_c} = \frac{Q_h}{T_h} = \Delta S_1. \quad (7.34)$$

The net heat taken in by one complete cycle is then

$$Q_h - Q_c = T_h \Delta S - T_c \Delta S = (T_h - T_c) \Delta S, \quad (7.35)$$

which is simply the area enclosed by the cycle on the $T - S$ plot.

An energy balance for one complete cycle is

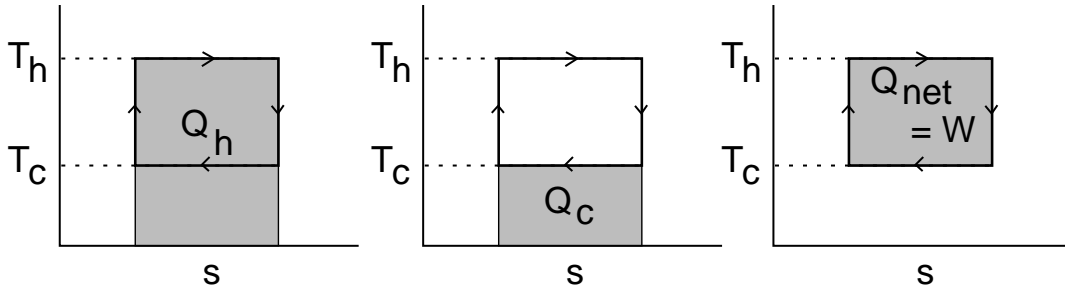
$$Q_h = W_{net} + Q_c, \quad (7.36)$$

where W_{net} is the *net* work output for the entire cycle. (If we denote the work output done in step n by W_n , then $W_{net} = W_1 + W_2 + W_3 + W_4$, where W_1 and W_2 are positive, and W_3 and W_4 are negative.) Therefore,

$$W_{net} = Q_h - Q_c = (T_h - T_c) \Delta S \quad (7.37)$$

is also given by the area enclosed on a $T - S$ plot.

We can also relate Q_h , Q_c , and W to areas on the $T - S$ plot:



Then the efficiency $\eta = W/Q_h$ is simply the ratio of the areas of two rectangles, and is easily seen to be $(1 - T_c/T_h)$.

Temperature-entropy plots are very useful for analyzing any reversible cycle. For a reversible process, $dQ = TdS$. Integrating this around a complete cycle, returning to the initial state,

$$Q_{net} = \oint dQ = \oint TdS. \quad (7.38)$$

Thus, the net heat taken in during one complete cycle equals the area enclosed by the cycle on a $T - S$ plot, no matter what the shape of the curve is (not only for a rectangle). From the First Law for a complete cycle,

$$\oint dU = 0 = \oint \bar{d}Q - \oint \bar{d}W \tag{7.39}$$

where we are taking work to be positive which is done *by* the system. Therefore, the net work done during one cycle is also given by the enclosed area:

$$W_{net} = Q_{net} = \oint TdS. \tag{7.40}$$

7.6.2 Maximum Power Output of a Heat Engine

Anyone who has compared the performance and fuel efficiency of a Geo Metro and a Corvette will not be surprised to learn that there is a *trade-off* between the efficiency of an engine and its power output. Let’s look at a simple example which shows this clearly.

Suppose we have a two-temperature heat engine operating between specified T_h and T_c , and we want to maximize the power output

$$\dot{W} = \eta \dot{Q}_h. \tag{7.41}$$

The highest η will be achieved by minimizing $\dot{\mathcal{P}}_s$, in which case η will approach the Carnot efficiency.

To minimize $\dot{\mathcal{P}}_s$, the engine must be run very slowly — in the limit, quasi-statically. This minimizes the work lost to friction, and also insures that heat transfer from the hot reservoir to the engine and from the engine to the cold reservoir is slow enough that large temperature differences (needed to “drive” the heat transfer) don’t develop. But if we run the engine that slowly, \dot{Q}_h will be very small!

The engine will be very *efficient* at converting Q_h into W , but the *rate* of heat extraction \dot{Q}_h will be very low, as will $\dot{W} = \eta \dot{Q}_h$, and in the limit of a reversible process (infinitely slow) both will go to zero. So it seems that building a reversible engine is not what we want to do if we want to maximize power output. The conditions which maximize η don’t maximize the product $\eta \dot{Q}_h$.

To maximize *power* instead of *efficiency*, we must allow some irreversibility. Let’s consider a simple case which only has one type of irreversibility – finite ΔT heat transfer. In reality, this source of irreversibility is inescapable. For example, suppose heat is taken from a reservoir at T_h by passing a fluid through tubes immersed in it. In order for heat to flow from T_h into the fluid in the tubes,

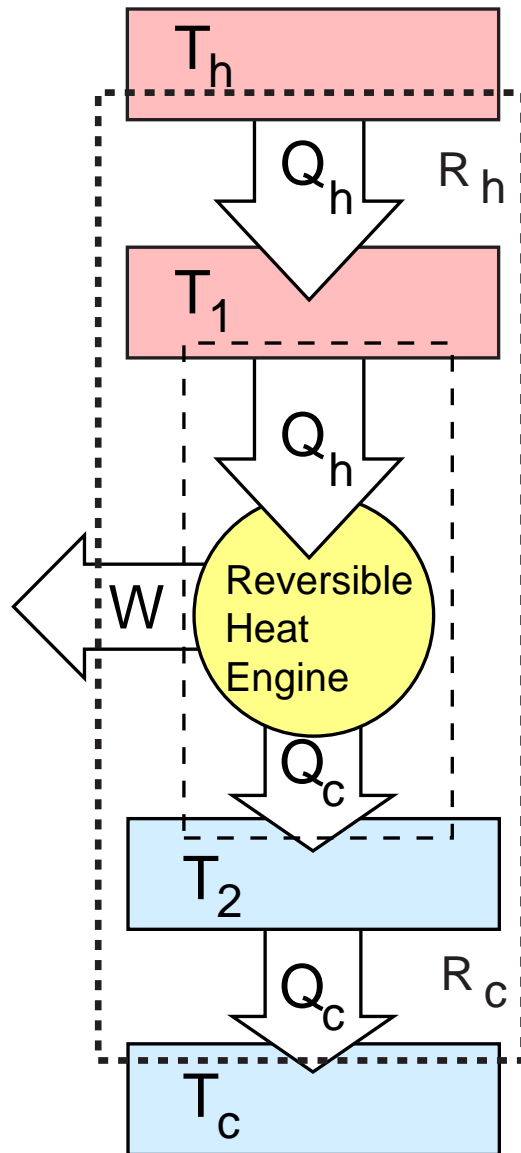


Figure 7.6: An “endoreversible” engine, in which the only irreversibility is due to the thermal resistances for heat transfer.

the fluid temperature must be less than T_h ; call it T_1 . The heat transfer rate is generally proportional to the temperature difference:

$$\dot{Q}_h = \frac{T_h - T_1}{R_h} \quad (7.42)$$

where R_h is the overall thermal resistance, which depends on factors such as the total tube area, the fluid flow rate, and the thermal conductivities of the materials involved. But if these are constant, then $Q_h \propto (T_h - T_1)$.

Also, to reject heat to a reservoir at T_c , the fluid (again flowing through tubes) must be hotter than T_c ; let's say it has temperature $T_2 > T_c$. Then the heat rejection rate will be

$$\dot{Q}_c = \frac{T_2 - T_c}{R_c}, \quad (7.43)$$

where R_c is the thermal resistance for heat transfer to the cold reservoir.

Assume for simplicity that there are no other irreversibilities – the engine is a reversible one, operating between its own maximum and minimum temperatures T_1 and T_2 . An hypothetical engine which is internally reversible, but must transfer heat in or out through thermal resistances is called *endoreversible*, meaning that the interior (*endo*) is reversible.

Since the engine is internally reversible, its efficiency is

$$\eta = \frac{\dot{W}}{\dot{Q}_h} = \left(1 - \frac{T_2}{T_1}\right). \quad (7.44)$$

But since $T_1 < T_h$ for $Q_h > 0$, and $T_2 > T_c$ for $Q_c > 0$, this efficiency is less than the value it would have if it were operating reversibly between T_h and T_c .

Temperatures T_1 and T_2 are not fixed, but change depending on the heat extraction rate. We can use Eq. (7.42) and Eq. (7.43) to substitute for T_1 and T_2 in terms of the constant reservoir temperatures T_h and T_c . If we also invoke the first law ($\dot{Q}_h = \dot{W} + \dot{Q}_c$), we find (after a bit of algebra):

$$\eta = 1 - \frac{T_c}{T_h - \dot{Q}_h(R_h + R_c)}. \quad (7.45)$$

As we increase the rate at which we take heat from the hot reservoir, the engine efficiency goes down, since T_1 decreases and T_2 increases. If we keep increasing \dot{Q}_h , eventually T_1 and T_2 approach each other, and $\eta \rightarrow 0$. From Eq. (7.45), this occurs when

$$\dot{Q}_h^{(max)} = \frac{T_h - T_c}{R_h + R_c}. \quad (7.46)$$

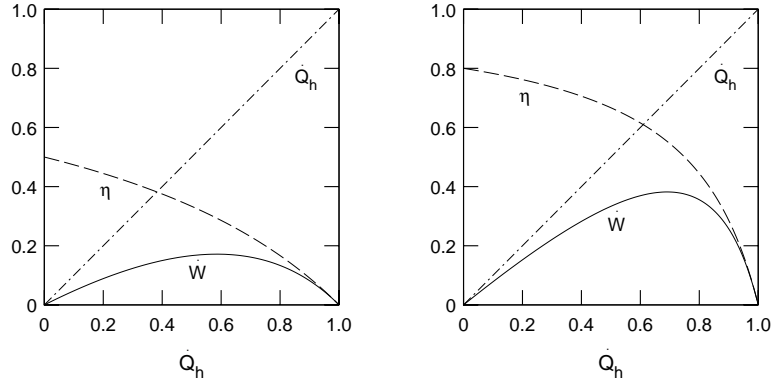


Figure 7.7: Efficiency, heat input, and power output for an endoreversible engine. Both \dot{Q}_h and \dot{W} are shown normalized by $\dot{Q}_h^{(max)}$ [Eq. (7.46)]. Left: $T_c/T_h = 0.5$; right: $T_c/T_h = 0.2$.

Equation (7.45) may be written as

$$\eta = 1 - \frac{(T_c/T_h)}{1 - (\dot{Q}_h/\dot{Q}_h^{(max)})(1 - T_c/T_h)}. \quad (7.47)$$

The power output is given by

$$\dot{W} = \eta \dot{Q}_h. \quad (7.48)$$

The quantities η , \dot{Q}_h , and \dot{W} shown below in Fig. 7.7 as a function of \dot{Q}_h for two values of T_c/T_h (0.2 and 0.5). The maximum *efficiency* is obtained for $\dot{Q}_h \rightarrow 0$, but the maximum power output occurs at a value of $\dot{Q}_h > 0$, where the engine efficiency is less than its maximum value.

To find the maximum power point, we can differentiate \dot{W} with respect to \dot{Q}_h and set the resulting expression to zero. Doing this, the maximum power condition is found to occur for

$$\dot{Q}_h^{(maxpower)} = \dot{Q}_h^{(max)} \frac{1}{1 + \sqrt{T_c/T_h}}. \quad (7.49)$$

Substituting this expression into Eq. (7.47), the engine efficiency at the maximum power condition is

$$\eta^{(maxpower)} = 1 - \sqrt{\frac{T_c}{T_h}}. \quad (7.50)$$

Note that this (unlike the Carnot efficiency) is not an upper limit on the efficiency of an engine. It's simply the efficiency an internally reversible engine will

have in the presence of thermal resistance for heat transfer, when it is operating at the maximum-power-output condition.

The surprising thing about this result is that $\eta^{(maxpower)}$ depends only on T_c and T_h , not on R_c and R_h . This analysis was first done by Curzon and Ahlborn at the University of British Columbia in 1975 and for this reason $1 - \sqrt{T_c/T_h}$ is sometimes called the Curzon and Ahlborn efficiency.

Curzon and Ahlborn pointed out that this efficiency is in practice quite close to that of many large power plants. This suggests that finite- ΔT heat transfer is one of the largest irreversibilities encountered in real power plants, and that they are in effect designed for maximum power output achievable for given T_h , T_c , R_h , and R_c . This in turn suggests that capital cost, not fuel cost, is the dominant factor in determining how many power plants are designed. (If fuel cost were dominant, it would make sense to operate below maximum power, where the efficiency is higher. To produce the required power, a larger (and more expensive) power plants could be built.)

7.7 Entropy Accounting for Open Systems

To extend our analysis to open systems, we have to modify the entropy accounting expressions of Section 7.3 to account for entropy carried by matter entering or leaving the system. We've already seen that when matter enters a system, it carries with it *energy* of e units per kg. Recall e is the total specific energy: $e = u + |\vec{V}|^2/2 + gz$.

It should come as no surprise that matter also carries entropy into a system when it enters. After all, entropy is an extensive property like energy or mass. The amount brought in is s units of entropy per kg, where s is the specific entropy (J/K-kg) of the matter entering. So if a fluid is flowing into a control volume at rate \dot{m} , and at the inlet the fluid has specific entropy s_{in} , the rate at which entropy is carried into the system by the fluid will be $\dot{m}s_{in}$.

We can easily add terms to our entropy accounting equations to account for this *convected* entropy. For example, Eq. (7.10) can be generalized as:

$$\frac{dS}{dt} = \sum_{\text{heat in}} \left(\frac{\dot{Q}_{in}}{T_{in}} \right) - \sum_{\text{heat out}} \left(\frac{\dot{Q}_{out}}{T_{out}} \right) \quad (7.51)$$

$$+ \sum_{\text{inlets}} (\dot{m}s)_{in} - \sum_{\text{outlets}} (\dot{m}s)_{out} + \dot{P}_s. \quad (7.52)$$

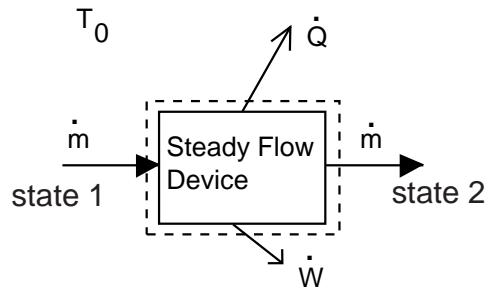


Figure 7.8: A steady-flow device.

7.8 Steady-Flow Devices

In Chapter 5, a set of useful steady-flow devices was introduced: turbines, compressors, pumps, nozzles, diffusers, and heat exchangers. These devices change the state of a fluid flowing steadily through them. In Chapter 5, we did an energy accounting on each of them. Let's now combine this with an entropy accounting.

We'll make the assumptions of steady state (no change in the amount of energy or entropy contained within the device), steady flow (constant \dot{m} in and out), and equilibrium properties at the inlet and outlet, where we need to evaluate them. We will assume that heat transfer occurs at a steady rate \dot{Q} to the surroundings at T_0 , and power \dot{W} is being produced at a steady rate. With these assumptions, the system is as shown in Fig. 7.8.

With the system as shown in the sketch, the entropy accounting is

$$\dot{m}s_{in} + \dot{\mathcal{P}}_s = \dot{m}s_{out} + \frac{\dot{Q}}{T_0} \quad (7.53)$$

Note that we are taking the system boundary just *outside* the device, where the temperature is T_0 . In that way, the irreversibility due to any finite- ΔT heat transfer occurs within the system, and is contained in the $\dot{\mathcal{P}}_s$ term. If we put the system boundary inside the device, then entropy would be being produced in the environment – we could do the analysis that way, too, but it would be more complicated.

Let's apply this equation together with an energy balance to some specific devices.

7.8.1 Adiabatic steady-flow devices

In Chapter 5, we introduced five steady-flow devices which are often modeled as adiabatic. These were (1) compressors, (2) turbines, (3) nozzles, (4) diffusers, and (5) valves.

For any *adiabatic* steady-flow device, Eq. (7.53) reduces to

$$\dot{m}s_{in} + \dot{\mathcal{P}}_s = \dot{m}s_{out}. \quad (7.54)$$

Since the second law requires $\dot{\mathcal{P}}_s \geq 0$, $s_{out} \geq s_{in}$:

the specific entropy of the fluid leaving an adiabatic steady-flow device is greater than or equal to the specific entropy of the fluid entering.

Equation (7.54), together with the steady-flow energy equation introduced in Chapter 4, is what we need to analyze any of these devices. Recall the forms of the steady-flow energy equation appropriate for adiabatic compressors, turbines, nozzles, and diffusers:

$$W_c = \dot{W}_c/\dot{m} = h_2 - h_1 \quad (\text{compressor}) \quad (7.55)$$

$$W_t = \dot{W}_t/\dot{m} = h_1 - h_2 \quad (\text{turbine}) \quad (7.56)$$

$$\Delta [V^2/2] = h_1 - h_2 \quad (\text{nozzle or diffuser}) \quad (7.57)$$

$$(7.58)$$

Since h and s appear in the energy and entropy accounting expressions, an $h - s$ plot is a convenient way to represent the change in the state of the fluid as it passes through any of these devices, as shown in Fig. 7.9.

In Fig. 7.9, the inlet state 1 and the outlet state 2 are shown connected by a dashed line. This is because the process may be irreversible, and so the fluid *inside* the device may not be in an equilibrium state, and therefore would not correspond to any point on the plot. Since the second law requires $s_2 \geq s_1$, state 2 can be *anywhere to the right of point 1, but not to the left*.

In these plots, another state “2s” is also shown. This is the outlet state which *would* result for an “ideal” reversible adiabatic device operating between P_1 and P_2 . The fluid coming out of the idealized device ends up in a different state than the outlet state of the real, irreversible device. State 2s has the pressure of state 2, but the entropy of state 1.

The turbine work output, compressor work input, or nozzle/diffuser change in kinetic energy is given by $|h_2 - h_1|$. This is shown on the plot. Since constant-

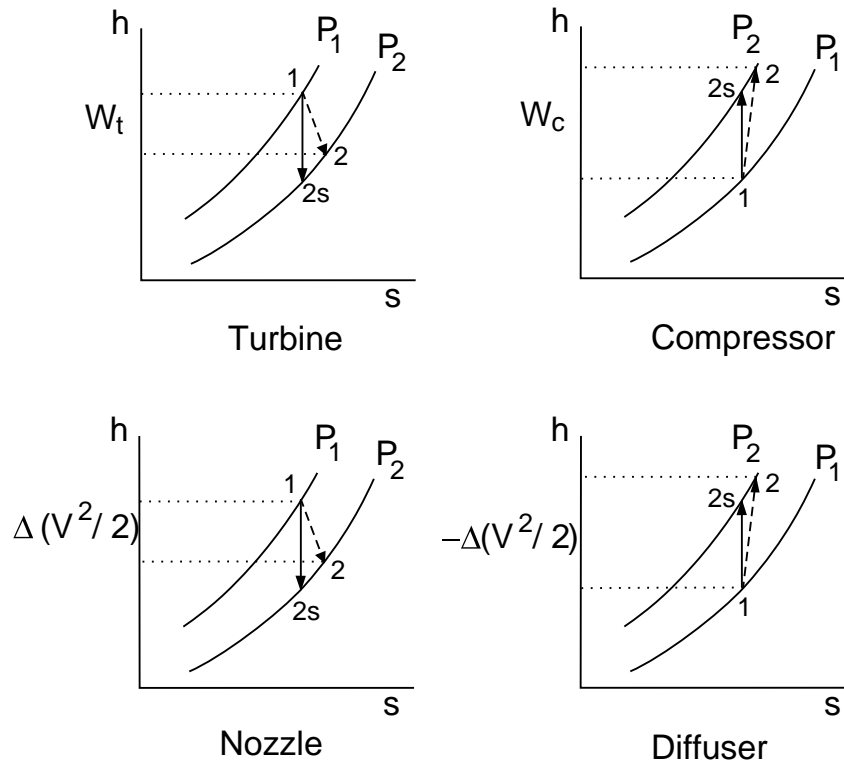


Figure 7.9: Enthalpy-entropy process representations for some adiabatic steady-flow devices. Isobars at P_1 and P_2 are also shown.

pressure lines have positive slope on an $h - s$ plot [i.e., $(\partial h/\partial s)_P > 0$],² we see that the reversible device always has the best performance. Specifically,

1. A reversible turbine gives the *maximum* work output W_t
2. A reversible compressor requires the *least* work input W_c
3. A reversible nozzle gives the *maximum* change in kinetic energy.

Example 7.1 Steam enters an adiabatic turbine at 900 K and 10 MPa and leaves at 1 MPa. What is the maximum work output per unit mass of steam, and what is the exit temperature under maximum work conditions?

Solution: at the inlet,

$$\begin{aligned} h_1 &= \text{h}(\text{"h2o"}, \text{"tp"}, 900, 10) = 3691.5 \text{ kJ/kg} \\ s_1 &= \text{s}(\text{"h2o"}, \text{"tp"}, 900, 10) = 6.977 \text{ kJ/kg-K}. \end{aligned}$$

Since the Second Law requires $s_2 \geq s_1$, $s_2 \geq 6.977 \text{ kJ/kg-K}$, and we're given $P_2 = 1 \text{ MPa}$. From TPX, the state with $s_2 = s_1$ at 1 MPa has temperature 535.87 K and enthalpy 2970.6 kJ/kg. Any state at this pressure with *greater* entropy has a higher temperature, and a higher enthalpy (try it and see).

The work output per kg of steam is $W_t = h_1 - h_2$. Thus, the most work is produced when h_2 is as small as possible; the smallest h_2 is achieved when $s_2 = s_1$. Therefore, the greatest work output possible is $W_{max} = 3691.5 - 2970.6 \text{ kJ/kg} = 720.9 \text{ kJ/kg}$, and the exit temperature is 535.87 K.

The performance of an adiabatic steady-flow device is often characterized by its *isentropic efficiency* η_s . The definition of η_s is different for each type of device, but in each case is a comparison of the actual performance (work input or output, or change in kinetic energy) with the performance of an ideal isentropic device operating between the same pressures. The isentropic efficiency is always defined so that it is ≤ 1 .

For a turbine,

$$\eta_{s,t} = \frac{W_t}{W_{t,s}} = \frac{h_1 - h_2}{h_1 - h_{2s}}. \quad (7.59)$$

Here W_t is the actual work output, and $W_{t,s}$ is the work output of the ideal isentropic turbine. Values of η_s for real turbines depend on size, varying from perhaps 60% for a very small turbine to 95% for a very large, well-designed turbine. A typical value would be around 90%.

²Since $dh = d(u + Pv)$, $dh = (Tds - Pdv) + vdP + Pdv = Tds + vdP$. Therefore, $(\partial h/\partial s)_P = T$, which is positive.

For a compressor,

$$\eta_{s,c} = \frac{W_{c,s}}{W_c} = \frac{h_{2s} - h_1}{h_2 - h_1}. \quad (7.60)$$

Compressor efficiencies are typically lower than turbine efficiencies, with a typical value of about 85%. A very large compressor might approach 90% efficiency.

7.9 Steady-Flow Availability

A question which is often of interest is how much power can be produced by *any conceivable* steady-flow process which takes a steadily-flowing fluid from a specified inlet state to a specified outlet state. To answer this question, consider again Fig. 7.8. From Eq. (7.53), the entropy accounting is

$$\dot{m}s_1 + \dot{\mathcal{P}}_s = \dot{m}s_2 + \frac{\dot{Q}}{T_0}. \quad (7.61)$$

The energy accounting for this steady flow process is

$$\dot{m}(e + Pv)_1 = \dot{W} + \dot{Q} + \dot{m}(e + Pv)_2. \quad (7.62)$$

(We are allowing the fluid to have kinetic and/or potential energy in addition to internal energy in states 1 and 2.)

Eliminating \dot{Q} between these equations and solving for \dot{W} ,

$$\dot{W} = \dot{m} [(e + Pv)_1 - (e + Pv)_2] - [T_0\dot{m}(s_1 - s_2) + T_0\dot{\mathcal{P}}_s], \quad (7.63)$$

or

$$\dot{W} = \dot{m}(b_1 - b_2) - T_0\dot{\mathcal{P}}_s, \quad (7.64)$$

where

$$b = e + Pv - T_0s \quad (7.65)$$

is the *steady flow availability function*. (Note that it is the environment temperature T_0 , not the fluid temperature T , which appears in this expression.)

From Eq. (7.64), we see that a real, irreversible process always produces less power than a reversible process acting between the same inlet and outlet states, for the same mass flow rate. The term $T_0\dot{\mathcal{P}}_s$ represents power which is “lost” due to the irreversibilities in the real process. For this reason, it is called the *irreversibility rate* \dot{I} :

$$\dot{I} = T_0\dot{\mathcal{P}}_s. \quad (7.66)$$

The maximum power produced occurs for $\dot{I} = 0$, and is

$$\dot{W}^{(max)} = \dot{m}(b_1 - b_2). \quad (7.67)$$

In some cases, the outlet state 2 is not fixed, and therefore we may ask what state 2 conditions maximize $\dot{W}^{(max)}$. Clearly, this will be the state which minimizes b_2 .

To minimize

$$b_2 = u_2 + e_{k,2} + e_{p,2} + P_2 v_2 - T_0 s_2,$$

the first thing to do is pick state 2 to have zero kinetic energy, and the lowest achievable potential energy. For example, if the potential energy is due to gravitation, pick state 2 at the lowest achievable elevation. (Elevations below the surface of the earth result in lower potential energy, but are not usually achievable.)

Once this is done, we are left with the task of minimizing $h_2 - T_0 s_2$. It can be shown (using methods to be discussed in the next chapter) that this is minimized for $T_2 = T_0$ and P_2 as small as possible. The lowest possible P_2 is P_0 , since eventually the fluid must flow to the environment. Thus, the optimal state 2 is the one which is in thermal and mechanical equilibrium with the environment.

Example 7.2

A certain industrial process exhausts 10 kg/s of steam at 2 atm and 400 K to the atmosphere through an exhaust valve. It has been suggested that the valve should be replaced with some process which can generate electrical power. Before deciding whether to pursue this, it is desired to know how much power could be produced in principle. The environment conditions are $P_0 = 1$ atm and $T_0 = 300$ K.

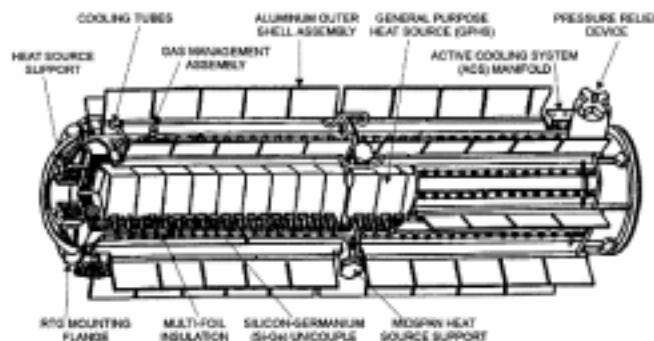
Solution: The maximum power output is $\dot{m}(b_1 - b_2)$. We will take state 2 to be liquid water at 1 atm and 300 K. Using TPX, the properties are

State	T (K)	P (atm)	h (kJ/kg)	s (kJ/kg-K)	b (kJ/kg)
1	400	2	2720.4	7.156	573.6
2	300	1	112.7	0.393	-5.3

Therefore, $\dot{W}^{(max)} = (10 \text{ kg/s})(578.9 \text{ kJ/kg}) = 5.8 \text{ MW}$. Of course, any real process will produce less power, but a well-designed process should be able to produce at least a couple of megawatts. Whether it makes sense to pursue this would depend on how much the process costs to build and run, and on the cost of electricity saved.

Problems

- 7.1** Gaseous water at 5 MPa in a piston/cylinder system is adiabatically expanded until its temperature is 400 K. If it is desired that there be no liquid droplets in the final state, what is the minimum possible initial temperature?
- 7.2** An insulated container contains saturated liquid nitrogen at 1 atm. You wish to design a process to produce work from this liquid nitrogen.
1. If the surroundings are at 300 K and 1 atm, what is the maximum work which could be produced in principle?
 2. Suppose you use a simple process in which you transfer the liquid to a piston-cylinder system, and let it warm up to room temperature at constant pressure. Sketch this process on an appropriate process diagram, and calculate the work produced.
 3. If the work produced in (b) is less than the maximum calculated in (a), discuss why.
 4. Can you suggest ways to improve this process?
- 7.3** The Cassini mission to Saturn was recently launched by NASA, and should arrive at Saturn in 2004. Like all NASA deep space missions, on-board electrical power on Cassini is generated by RTGs (Radioisotope Thermoelectric Generators). These consist of pellets of radioactive plutonium fuel which produce heat, bundles of silicon-germanium thermocouples, and radiator fins which radiate waste heat into space. The thermocouples produce electrical power directly when connected between the high-temperature “General Purpose Heat Source” containing the plutonium and the cold radiator fins.



The heat rejected from the radiator fins follows the Stefan-Boltzmann law:

$$\dot{Q}_c = \epsilon A \sigma T_c^4.$$

Here T_c is the temperature of the radiator fins, A is the total fin area, and ϵ is a material property known as the *emissivity* ($0 < \epsilon < 1$).

The array of thermocouples constitutes a heat engine operating between the plutonium fuel temperature (T_h) and the radiator temperature T_c .

To minimize the weight and spacecraft size, it is desired to keep the radiator fins as small as possible. For a given power generation requirement \dot{W} , ϵ , and T_h , determine

1. the *minimum radiator fin area* required, if the thermocouples act as “ideal” heat engines.
2. the thermodynamic efficiency $\eta = \dot{W}/\dot{Q}_h$ and fin temperature T_c under the conditions of part (a).

Assume that there is no heat transfer limitation on the rate at which heat may be taken from the radioactive source.

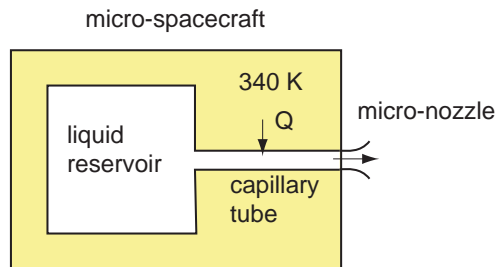
7.4 You have 2 identical copper blocks of mass M . One is initially at temperature $T_1(0)$, and the other is initially at temperature $T_2(0) < T_1(0)$. Both may be assumed to be incompressible substances with constant c . You now let them interact through some process which extracts heat from block 1, generates work, and dumps some heat into block 2. Since the blocks have finite mass, $T_1(t)$ and $T_2(t)$ change with time. Eventually, they come to the same temperature $T_1(\infty) = T_2(\infty) = T_f$ and no more work can be done.

1. Taking both blocks together as the system and assuming no heat is transferred to or from the environment, write down expressions for $\Delta U = U(\infty) - U(0)$ and $\Delta S = S(\infty) - S(0)$. Assume that the total entropy produced during the process is \mathcal{P}_s .
2. Combine these equations to obtain expressions for W and T_f . Your expression should contain only M , c , $T_1(0)$, $T_2(0)$, and \mathcal{P}_s .
3. Show that the *maximum possible* work is

$$W_{max} = Mc \left(\sqrt{T_1} - \sqrt{T_2} \right)^2$$

and determine T_f in this case.

- 7.5** Nitrogen is held in a container at its critical pressure and temperature. If the environment is at 300 K and 1 atm, what is the maximum work which could be done per kg of N_2 ?
- 7.6** A concept for a positioning thruster for a very small (1 kg) spacecraft is shown below.



Liquid HFC-134a at $T = 250$ K and 0.8 MPa flows from a small on-board reservoir into a capillary tube 0.5 mm in diameter and 1 cm long, where it absorbs waste heat given off by the on-board electronic equipment, causing it to evaporate. It emerges from the tube as vapor at 320 K at a pressure of 0.65 MPa. (Due to wall friction, there is a substantial pressure drop down the tube.) The vapor emerging from the tube enters an adiabatic micro-nozzle and is accelerated to produce thrust. The micro-nozzle has an isentropic efficiency of 0.9, and the pressure at its exit is 0.05 MPa.

Determine:

1. The irreversibility rate \dot{I} in the capillary tube, if the environment surrounding the tube is at 340 K.
 2. The velocity and temperature at the exit of the micro-nozzle
 3. The mass flow rate in mg/s required to produce a thrust of 1 mN. (Thrust is mass flow rate multiplied by exit velocity.)
- 7.7** What is the minimum power required by an adiabatic compressor which compresses 10 kg/min of saturated water vapor at 0.2 MPa to 0.4 MPa?
- 7.8** What is the minimum power required by *any* process (not necessarily adiabatic) which compresses 10 kg/min of saturated water vapor at 0.2 MPa to 0.4 MPa? The surroundings are at $P_0 = 1$ atm and $T_0 = 300$ K. Compare this process to that in problem 1. Are the final states the same? Does this process involve heat transfer? If so, in which direction?

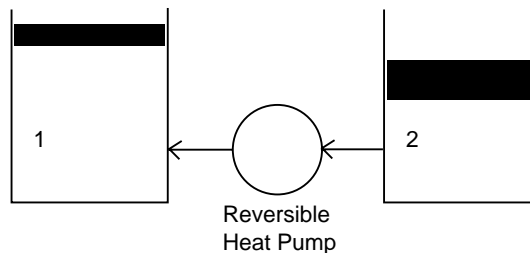
7.9 An industrial process to produce liquid nitrogen starts with gaseous nitrogen at 300 K and 1 atm, which enters at a steady rate of 1 kg/s. The product is saturated liquid nitrogen at 1 atm.

1. If the environment is at 300 K and 1 atm, does this process require work input, or can it produce work?
2. Determine the minimum work input, or maximum work output (depending on your answer to (a)).
3. Determine the heat transfer rate for the conditions of part (b).

7.10 (NOTE: Requires Helium data; can't be done with current version of TPX.)

Two cylinders with pistons each contain 10 kg of helium. The weight of piston 1 keeps P_1 constant at 0.05 MPa, and the weight of piston 2 keeps P_2 constant at 0.2 MPa. Initially, both cylinders are at a temperature of 5.2 K. They are now thermally connected to a reversible heat pump, which heats cylinder 1 and cools cylinder 2 with no heat transfer to the surroundings. When cylinder 1 reaches 8 K, the process stops.

1. Write equations expressing energy and entropy accounting for this process.
2. Find the final state of the helium in cylinder 2.
3. Find the work input to the heat pump.



7.11 A heat-powered portable air compressor consists of three components: (a) an adiabatic compressor; (b) a constant pressure heater (heat supplied from an outside source); and (c) an adiabatic turbine. The compressor and turbine both may be idealized as reversible, and air may be assumed to be an ideal gas with $k = c_p/c_v = 1.4$.

Ambient air enters the compressor at 0.1 MPa and 300 K, and is compressed to 0.6 MPa. All of the power from the turbine goes to run the

compressor, and the turbine exhaust is the supply of compressed air. If this pressure is required to be 0.2 MPa, what must the temperature be at the exit of the heater?

- 7.12** For a particular industrial process, a stream of gaseous oxygen at 2 MPa and 1000 K is required. You have available a stream of saturated liquid oxygen at 100 K, and you wish to design a process to continuously produce the desired gaseous oxygen from this liquid oxygen. The surroundings are at 300 K. Determine whether such a process requires work input or can produce work output, and the minimum work input or maximum work output per kg of O₂.
- 7.13** An adiabatic pump pumps 200 kg/hr of saturated liquid HFC-134a at 320 K to 2.5 MPa. The temperature of the emerging liquid is 321.26 K. Determine:
1. The power required to run the pump
 2. The isentropic efficiency of the pump
 3. The entropy production rate
 4. The irreversibility rate if the environment is at 300 K.

CHAPTER 9

POWER GENERATION AND PROPULSION

9.1 Introduction

One of the most significant engineering applications of thermodynamics is the design and optimization of energy conversion devices which take in heat and produce useful work (“heat engines”). In Chapter 7, we discussed in very general terms the principles governing the performance of heat engines. In this chapter, we look at some specific processes which are actually used for power generation and propulsion.

9.2 Cycles

Any thermodynamic process which takes a substance through a sequence of states eventually returning it to the original state is known as a *cycle*. A familiar example is the Carnot cycle, discussed in Chapter 7.

Virtually all heat engines are implemented as cycles. Usually, the substance the cycle operates on is a fluid, although in principle any substance could be used, as long as it has some reversible work mode. In this chapter, we’ll only consider cycles which operate on fluids, since all important engineering cycles are of this type.

The fluid used in the cycle is known as the *working fluid*. During the cycle, the working fluid may or may not change phase. In *vapor power cycles*, the working fluid changes phase from liquid to vapor and back to liquid over the course of the cycle; in *gas power cycles*, the working fluid remains gaseous during the entire cycle. As we will see, they each have certain advantages and disadvantages.

In many types of cycles, the working fluid circulates continuously in a loop. For example, the major cycles used for electric power generation are of this type. In these cycles, the thermodynamic state of the working fluid at any one *location* in the loop is constant in time. But if we focus our attention on a single “packet” of fluid (say, a unit mass) moving around the loop, its state changes as it passes through various components such as pumps, turbines, heat exchangers, etc. To analyze such cycles, we follow a fluid packet around the loop, applying

the steady-flow equations expressing energy and entropy accounting to each device the fluid packet encounters to determine how its state changes.

In another type of cycle, a fixed mass of fluid is made to undergo a sequential set of operations. The Carnot cycle is an idealized example of such a cycle. Practical examples include spark-ignition and Diesel internal combustion engines. To analyze these, we apply the control-mass energy and entropy accounting equations to each step in the cycle.

To analyze real cycles, we need to account for the fact that they are irreversible (they produce entropy). As we saw in Chapter 7, some irreversibility (for example due to finite- ΔT heat transfer) is inevitable if we want to produce *power*, since reversible processes operate infinitely slowly. In addition to finite- ΔT heat transfer, some other common irreversibilities are those due to friction, fluid viscosity, shock waves, and combustion.

Nevertheless, we can learn a lot about real cycles by first considering ones which are fully reversible, and then considering how the results for reversible cycles need to be modified to account for irreversibilities. We'll do this in the next two sections, and then apply these ideas to look at the characteristics of several different cycles.

9.2.1 Process Representations

To carry out thermodynamic analysis of cycles, we need to keep track of how the thermodynamic state of the fluid changes as the cycle proceeds. Since for a fluid in equilibrium only two properties are required to specify its state, a fluid in thermodynamic equilibrium may be represented as a point on a plot of one thermodynamic property vs. another one (e.g. T vs. s , P vs. v , etc.) Any such plot of the working fluid state at various points in the cycle is known as a *process representation*.

If the fluid remains in equilibrium at every point during the cycle, then the fluid state always corresponds to some point on the process representation and as the cycle proceeds, a closed curve is traced out. Since the fluid is always in equilibrium, no internal irreversible processes (which act only when a system is perturbed from equilibrium) can be occurring. Therefore, such a cycle is reversible. We call reversible cycles *ideal cycles*. Of course, real cycles can approach ideal cycles, but there will always be some entropy production in reality and the fluid will always be at least a little out of equilibrium.

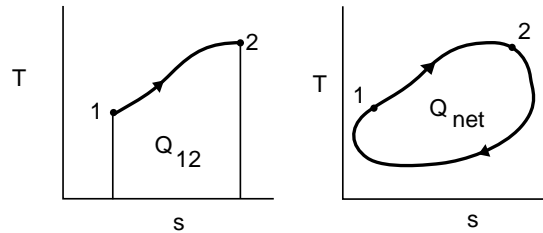
To analyze real, irreversible cycles, we usually assume the fluid is approximately in equilibrium at the inlet and outlet of devices like valves, turbines, pumps, compressors, etc., but not necessarily within them. Points in the pro-

process representation plane corresponding to the inlet and outlet states are simply connected with a dashed line to represent the action of the device on the fluid.

9.2.2 Analysis of Ideal Cycles

The $T - s$ process representation is particularly useful to represent ideal cycles. Since for any reversible process $dQ = Tds$, the area under a curve traced out by a reversible process which takes the working fluid from some state 1 to some state 2 is the total heat added:

$$Q_{12} = \int_1^2 T ds. \quad (9.1)$$

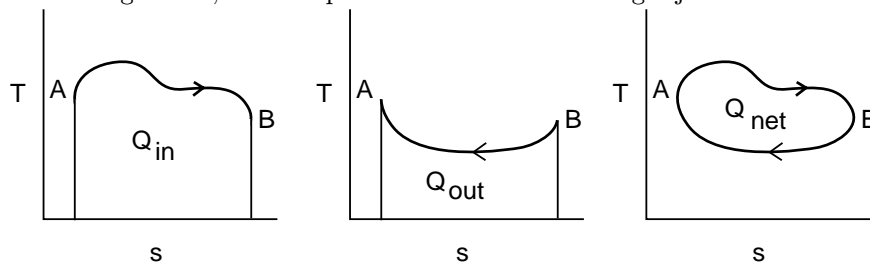


If we consider a complete cycle, then the *net* heat taken in by the cycle is the area enclosed by the curve in the $T - s$ plane:

$$Q_{net} = \oint T ds. \quad (9.2)$$

Of course, on the portion of the cycle where $ds < 0$, $dQ < 0$ so heat is actually being transferred out of the working fluid to the environment.

It is convenient to break up the path of a closed cycle into the portion where heat is being added, and the portion where heat is being rejected:



Then the heat absorbed per unit mass by one complete cycle is the area under the upper portion of the curve (left), and the heat rejected to the environment per unit mass is the area under the lower portion (center). The net heat taken in is the difference, which is the enclosed area (right). From the first law, the net heat taken in must equal the net work output, since for a complete cycle $\Delta u = 0$.

Therefore, *the area enclosed by an ideal cycle in the $T - s$ plane is the net work produced per unit mass of working fluid for one complete cycle.*

The net work per unit mass is an important quantity, since the total power produced by the cycle is the product of W_{net} and the mass flow rate \dot{m} :

$$\dot{W} = \dot{m}W_{net}. \quad (9.3)$$

The mass flow rate needed to produce a given desired power scales inversely with W_{net} . Since the overall size of a powerplant scales roughly with the working fluid mass flow rate, higher W_{net} results in smaller, more-compact (and usually less expensive) power plants. This is especially important in applications such as aircraft propulsion, where the weight of the engine must be minimized.

The thermodynamic efficiency η of any power cycle is defined as the net work output divided by the heat *input*:

$$\eta = \frac{W_{net}}{Q_{in}}. \quad (9.4)$$

Thus, the efficiency of an *ideal* cycle is the ratio of the area enclosed on a T-s plot to the total area under the curve.

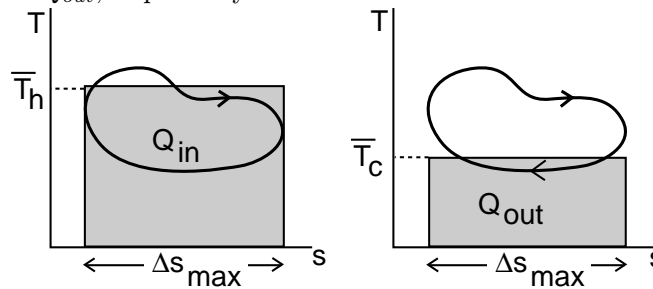
Two important quantities for ideal cycles are the *average* temperature at which heat is added (\bar{T}_h), and the average temperature at which heat is rejected (\bar{T}_c). These are defined by

$$\bar{T}_h = \frac{Q_{in}}{\Delta S_{max}}, \quad (9.5)$$

and

$$\bar{T}_c = \frac{Q_{out}}{\Delta S_{max}}, \quad (9.6)$$

where $\Delta S_{max} = s_B - s_A$ is the difference between the maximum entropy and minimum entropy points in the cycle. Then the rectangles shown below have area Q_{in} and Q_{out} , respectively.



The thermodynamic efficiency may be expressed in terms of \bar{T}_h and \bar{T}_c :

$$\eta = \frac{W_{net}}{Q_{in}} = \frac{Q_{in} - Q_{out}}{Q_{in}} \quad (9.7)$$

$$= 1 - \frac{Q_{out}}{Q_{in}} \quad (9.8)$$

$$= 1 - \frac{\bar{T}_c \Delta s_{max}}{\bar{T}_h \Delta s_{max}}. \quad (9.9)$$

Therefore,

$$\eta = 1 - \frac{\bar{T}_c}{\bar{T}_h}. \quad (9.10)$$

This expression holds for any ideal cycle. For the special case when all heat is taken in at a single temperature T_h and all heat is rejected at a single temperature T_c (as in the Carnot cycle), $\bar{T}_h = T_h$ and $\bar{T}_c = T_c$ and this expression reduces to the Carnot efficiency. We may regard the cycle as being equivalent to a Carnot cycle with temperatures \bar{T}_h and \bar{T}_c .

Equation (9.10) shows that the thermodynamic efficiency of an ideal cycle can be increased either by increasing the average temperature at which heat is added or by lowering the average temperature at which heat is rejected. This insight will prove to be vital for improving the efficiency of real cycles, although of course with real cycles we also have to be concerned with minimizing internal irreversibilities.

9.2.3 Component Models

To analyze real cycles, we need models for the action of various components the working fluid encounters (pumps, compressors, heaters, turbines, nozzles, etc.) which include the irreversible character of these components. Several such models of common steady-flow devices were discussed in Section 7.8. We will use these in the analysis below, and introduce some new component models also.

9.3 Vapor Power Cycles

The first power cycle to be developed was the *Rankine cycle*, and in modified form the Rankine cycle is still the most common vapor power cycle. It is in widespread use for electric power generation in large central power plants, and will be for the foreseeable future. Whatever the type of fuel burned in the power plant (coal, natural gas, oil, or nuclear fuel), a modified Rankine cycle using water as the working fluid is invariably used to convert the heat produced into useful power (sometimes in conjunction with other cycles we'll discuss later).

In the sections below, we work through some specific examples of the Rankine cycle, progressively adding modifications which improve performance and lead

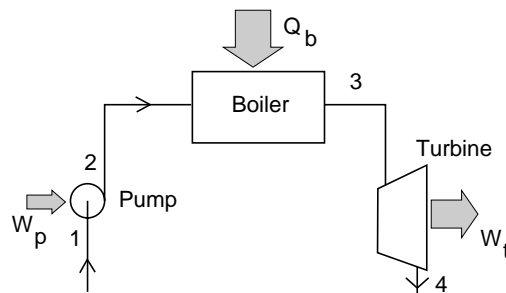


Figure 9.1: The Open Rankine Cycle

to the types of cycles actually used in modern power plants.

9.3.1 The Open Rankine Cycle

The simplest version of the Rankine cycle is the open Rankine cycle, shown in Figure 9.1. In the open Rankine cycle, a fluid (usually water) is pumped to high pressure, and then heat from a flame is added in a boiler to boil the water at constant pressure. Saturated steam is extracted from the boiler, and expanded back to atmospheric pressure, producing work. In Figure 9.1, the expansion is shown occurring in a turbine, which is how modern implementations of the Rankine cycle operate.

The open Rankine cycle was originally developed in the 19th century to power steam locomotives and other steam-operated heavy equipment. The development of steam engines made possible the industrial revolution and everything which followed from it. In a steam locomotive, the high pressure steam entered a cylinder and pushed a piston, which propelled the train forward. At the end of the stroke, an exhaust valve opened and on the reverse stroke the low-pressure steam was expelled to the atmosphere.

Let's analyze this cycle using some typical numbers. We'll take state 1 to be water at 30°C and 1 atm, $P_2 = 0.8\text{ MPa}$, and assume that state 3 is saturated vapor. The pump and turbine are both adiabatic, with isentropic efficiencies of 0.6 and 0.8 respectively. To analyze the cycle, we simply work our way around it, putting a control volume around each component and writing down expressions for energy and entropy accounting.

The Initial State

We start the analysis in state 1, since enough information is given to fully determine the state here. With $T = 303.15\text{ K}$ and $P = 1\text{ atm}$, we using TPX

that

$$\begin{aligned}v_1 &= 0.00100 \text{ m}^3/\text{kg} \\h_1 &= 125.2 \text{ kJ/kg} \\s_1 &= 0.4348 \text{ kJ/kg-K}\end{aligned}$$

The Pump

To analyze the pump, it is simplest to assume that the liquid water is incompressible. In this case, the work of an ideal (isentropic) pump is

$$W_{p,s} = v_1(P_2 - P_1). \quad (9.11)$$

Since these quantities are known, we find

$$W_{p,s} = 0.7 \text{ kJ/kg}.$$

By definition of the isentropic efficiency for a pump, the actual work W_p is

$$W_p = \frac{W_{p,s}}{\eta_s} = 1.2 \text{ kJ/kg}.$$

Knowing the pump work, we can calculate the enthalpy of state 2:

$$h_2 = h_1 + W_p = 126.4 \text{ kJ/kg}.$$

Knowing h_2 and P_2 is sufficient to fix all other properties in state 2, and so we find

$$T_2 = 303.26 \text{ K}$$

$$s_2 = 0.4366 \text{ kJ/kg-K}.$$

As required by the second law, $s_2 > s_1$. We see that pumping the liquid produces only a very small temperature rise (0.1 K).

The Boiler

Let us assume the heat addition in the boiler occurs at constant pressure. Then

$$P_3 = P_2 = 0.8 \text{ MPa}.$$

We are not given how much heat is added in the boiler, but we know that state 3 is saturated vapor. Calculating the properties for saturated vapor at 0.8 MPa, we find

$$T_3 = 443.58 \text{ K}$$

$$h_3 = 2768.7 \text{ kJ/kg}$$

$$s_3 = 6.662 \text{ kJ/kg-K.}$$

From an energy balance on the boiler, $h_2 + Q_b = h_3$, so the heat added in the boiler is

$$Q_b = 2642.3 \text{ kJ/kg.}$$

The Turbine

To analyze the turbine, we first must find the outlet state which would result from an isentropic turbine. Calling this state $4s$, we know

$$P_{4s} = 1 \text{ atm}$$

$$s_{4s} = s_3 = 6.662 \text{ kJ/kg-K.}$$

This is enough to fully specify state $4s$, so we find

$$T_{4s} = 373.144 \text{ K}$$

$$h_{4s} = 2417.3 \text{ kJ/kg}$$

$$s_{4s} = 6.662 \text{ kJ/kg-K}$$

$$x_{4s} = 0.885$$

Note that state $4s$ lies within the vapor dome, with nearly 12% liquid content by mass.

Having found $4s$, we now can get state 4 from the given isentropic efficiency of the turbine:

$$\eta_s = \frac{h_3 - h_4}{h_3 - h_{4s}}.$$

Putting in $\eta_s = 0.8$ and solving for h_4 ,

$$h_4 = 2487.6 \text{ kJ/kg.}$$

With h_4 and P_4 known, we can find the remaining properties:

$$s_4 = 6.850 \text{ kJ/kg-K}$$

$$x_4 = 0.917$$

We see $s_4 > s_3$, consistent with the second law, and the actual turbine outlet state also contains some liquid.

The vapor mass fraction x at the turbine outlet is an important parameter, since if it is too low the water droplets which form within the turbine impact on the turbine blades and cause serious erosion. Typically, x of about 0.9 is the

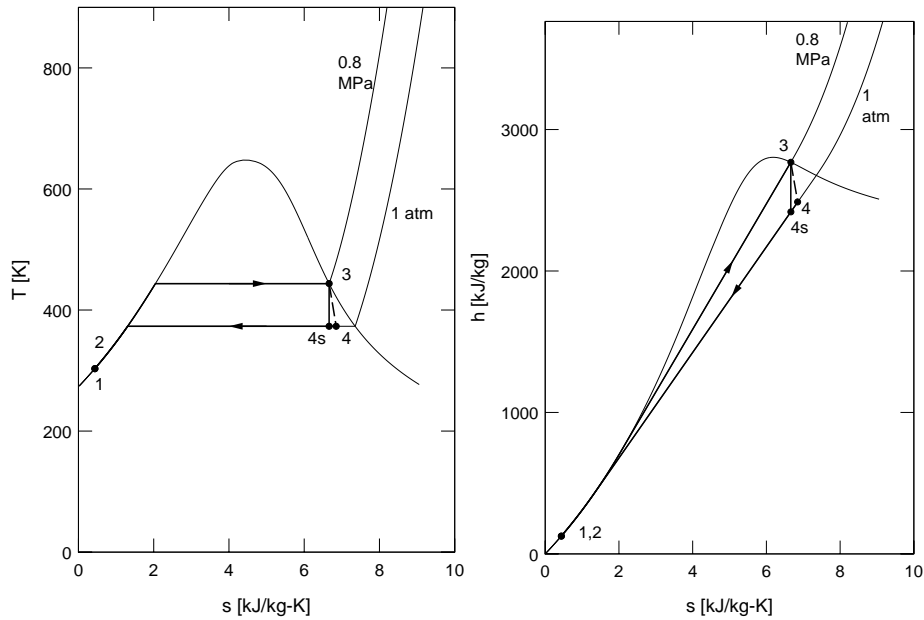


Figure 9.2: An open Rankin cycle (shown as closed by the environment)

lowest tolerable value. Thus, the conditions at state 4 are probably acceptable, but just barely. Since the vapor mass fraction is so important, it is often called the *steam quality*.

Process representations for this cycle in the $T-s$ and $h-s$ planes are shown in Figure 9.2. Note that states 1 and 2 are different, but overlap on this scale. Also, the cycle is shown in these process representations as being closed by a constant-pressure line connecting state 4 to state 1. This allows us to regard an open Rankin cycle as being closed by the environment, and the step from 4 to 1 is the heat rejection to the environment required to condense the steam and cool the liquid down to T_1 .

Having now found all of the states of the cycle, we can calculate its performance. The work produced by the turbine per unit mass of steam flowing through it is

$$W_t = h_3 - h_4 = 281.1 \text{ kJ/kg.}$$

Some of this work (1.2 kJ/kg) must be used to run the pump, so the net work output is

$$W_{net} = W_t - W_p = 279.9 \text{ kJ/kg.}$$

To produce this work, $Q_b = 2642.3$ kJ/kg of heat had to be added to the steam

in the boiler. Therefore, the efficiency of this process at converting heat into work is

$$\eta = \frac{W_{net}}{Q_b} = 10.6\%.$$

We see in this example that the work output from the turbine is much greater than the work required to run the pump. Their ratio is known as the *back work ratio* BWR:

$$\text{BWR} = \frac{W_p}{W_t}. \quad (9.12)$$

For this example, $\text{BWR} = 0.004$. This is one of the major advantages of vapor power cycles: the work required to compress the liquid is negligible compared to the work produced by the vapor expanding through the same pressure difference.

The Carnot efficiency for an ideal heat engine operating between the same maximum and minimum temperatures encountered in this cycle is

$$\eta_{Carnot} = 1 - \frac{303.15}{443.58} = 31.6\% \quad (9.13)$$

Therefore, the cycle in this example has an efficiency only 1/3 of the theoretical maximum for these temperatures. One reason the efficiency is low is the fact that the turbine and the pump are not ideal. But if we recalculated the performance of an ideal cycle with $\eta_{s,p} = \eta_{s,t} = 1$, we would find an efficiency of 13.2%. So even the ideal cycle has an efficiency far below the Carnot efficiency.

By examining the $T - s$ plot in Fig. 9.2, we can see some of the reasons for the low efficiency. The turbine exhaust is steam at 373.15 K. Much of the total heat rejection to the atmosphere occurs as this steam condenses to liquid at this temperature. Therefore, \bar{T}_c is much greater than 30 °C. Thus, this cycle needlessly throws away hot steam, from which more work could still be extracted.

Another factor limiting the efficiency is the heat addition from state 2 up to the saturation point. This heat addition in the boiler occurs at less than T_{sat} , so $\bar{T}_h < 443.58$ K.

In the next example, we consider a modification to allow rejecting heat at lower temperature. We'll look at how to raise \bar{T}_h in the following example.

9.3.2 The Closed Rankine Cycle

In the last example, state 4 had a temperature of 373.15 K, which is the result of exhausting at 1 atm. But suppose we didn't simply exhaust to the atmosphere, but instead provided a low-pressure region for the turbine to exhaust into. For $P_4 < 1$ atm, $T_4 = T_{sat}(P_4)$ will be less than 373.15 K, and we will be able to lower the temperature at which heat is rejected.

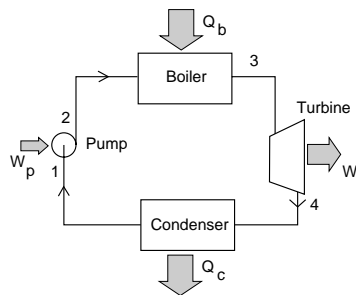


Figure 9.3: The Closed Rankine Cycle.

To do this, we must close the cycle by adding a condenser, in which heat will be removed from the turbine exhaust, leaving saturated liquid water at state 1. This is shown in Fig. 9.3.

We can choose the temperature at which we will condense the steam based on the temperature of the reservoir into which we will be dumping the heat. In many cases, the heat will be removed by cooling water supplied from a nearby lake, river, or ocean. If a source of cooling water is available at, say, 20 °C (293 K), then we could choose a condensing temperature of, say, 300 K. The 7 degree temperature difference should be enough to drive heat transfer in the condenser. The condenser tube area is determined by the total amount of heat which must be rejected, and by the temperature difference between the steam and the cooling water. There is always a trade-off between ΔT and total condenser area. For given Q_c , smaller ΔT means larger A . The optimum values would be determined in reality by balancing the efficiency gain by lowering the condensing temperature with the cost required to build a bigger condenser. A ΔT value of a few degrees is would be typical.

Let us specify $T_4 = 300$ K. Then since state 4 will be in the vapor dome,

$$P_4 = P_{sat}(300) = 0.003536 \text{ MPa.}$$

This is about 1/29 atm. To start the cycle running (for example, after shutting down for maintainance), a vacuum pump is needed to pump out the air. Also, since the pressure is less than 1 atm, any small leaks will result in air entering the condenser. If enough air enters, the pressure will begin to rise, causing T_{sat} to rise, and lowering the cycle efficiency. So the vacuum pump must be used during operation also to remove any air entering through leaks. Thus, the system will be somewhat more complex than the open Rankine cycle. But the gain in efficiency will more than offset the increased complexity.

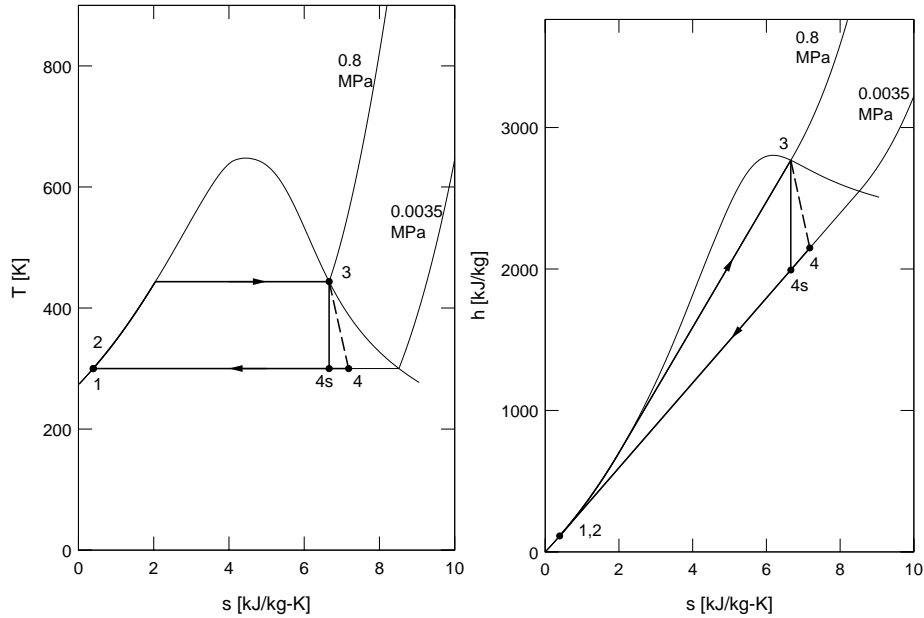


Figure 9.4: A closed Rankin cycle

With T_4 set to 300 K, and state 1 set to saturated liquid at 300 K, we can re-do the analysis of Example 1. The numbers are summarized in the table below. (You should verify these.)

State	P (MPa)	T (K)	h (kJ/kg)	s (kJ/kg-K)
1	0.003536	300	111.7	0.3900
2	0.8	300.08	113.6	0.3924
3	0.8	443.58	2768.7	6.662
4s	.003536	300	1993.3	6.662
4	.003536	300	2148.4	7.179

These states are shown in $T-s$ and $h-s$ plots in Fig. 9.4.

The pump work is now

$$W_p = h_2 - h_1 = 1.33 \text{ kJ/kg.}$$

This is slightly greater than in Example 1, which is not surprising since now state 1 has a lower pressure, so $P_2 - P_1$ is greater than before.

The turbine work output is

$$W_t = h_3 - h_4 = 620.3 \text{ kJ/kg}$$

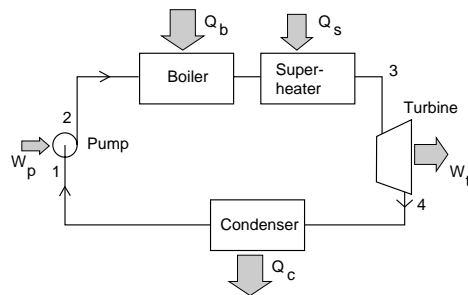


Figure 9.5: A Rankine Cycle with Superheat.

This is a factor of 2.2 greater than the turbine work in Example 1. Lowering the outlet pressure allowed the steam to expand much more, producing more work. This is clear in comparing the $h - s$ plots in Figures 9.2 and 9.4.

The heat added in the boiler Q_b is unchanged, so the cycle efficiency is

$$\eta = 23.3\%$$

This is quite an improvement over the open Rankine cycle efficiency.

9.3.3 Addition of Superheat

Having added the condenser, we've lowered the temperature at which heat is rejected about as much as possible. We now need to consider how to raise \bar{T}_h if we wish to improve the efficiency more.

The temperature of the steam in the boiler is set by the pressure, since $T = T_{sat}(P)$. So the next improvement to make is to increase the boiler pressure. The limits on boiler pressure are determined by how much pressure the material the tubes carrying the high-pressure steam can withstand at high temperature for long durations. Originally, boilers were constructed from cast iron, which limited the safe operating pressures. Now with high-strength steel boilers, much higher pressures can be used. Let's take a typical value of 10 MPa.

If this is all we do, we have a problem. If we expand saturated vapor at 10 MPa to 0.0035 MPa in a turbine, a very large liquid fraction would result. But there is no reason to restrict ourselves to saturated vapor. We can extract saturated steam from the boiler, but add more heat to it before sending it through the turbine. Doing this is known as *superheating* the steam, and all modern power plants employ superheat.

The upper limit on how hot the steam may be heated is determined by the materials constraints of the turbine. Hot steam is corrosive, and if it is too hot

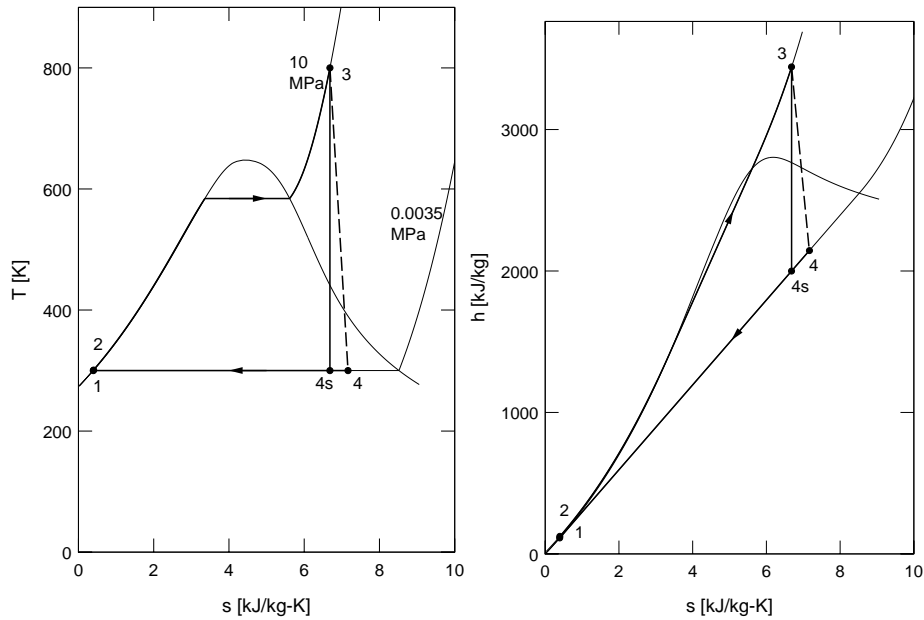


Figure 9.6: A Rankin cycle with superheat.

when it enters the turbine it will rapidly corrode the turbine blades. Advances in high temperature materials for turbine blades and elaborate cooling methods have resulted in allowable turbine inlet temperatures of more than 800 K. We'll take this value for this example.

Finally, we can work on improving the isentropic efficiencies of the pump and turbine, to more closely approach the ideal cycle. Large-scale turbines can now be manufactured with isentropic efficiency above 90%; let's take a value of $\eta_{s,t} = 0.9$. The isentropic efficiency of the pump hardly matters, due to the low BWR, but these have gotten better too. We'll take $\eta_{s,p} = 0.8$.

The process representation is shown in Fig. 9.6.

The analysis is very similar to what we've done in the last two examples, except that state 3 is now fixed by the specified T_3 and P_3 . The numbers work out as shown below.

State	P (MPa)	T (K)	h (kJ/kg)	s (kJ/kg-K)	x
1	0.003536	300	111.7	0.3900	0
2	10	300.72	124.3	0.3993	0
3	10	800	3442.0	6.683	1
4s	.003536	300	1999.7	6.683	0.774
4	.003536	300	2143.9	7.164	0.833

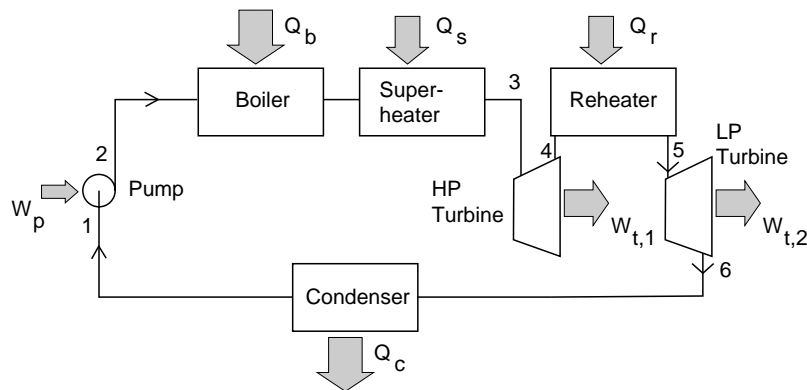


Figure 9.7: A Rankine cycle with superheat and one stage of reheat.

Heat added:	3317.7 kJ/kg
Turbine work:	1298.1 kJ/kg
Pump work:	12.5 kJ/kg
Efficiency:	38.7%
BWR:	0.0096

These modifications have produced a marked increase in efficiency, and this value is in the ballpark of real steam power plant efficiencies. The work output per unit mass of steam has also significantly increased. The steam quality leaving the turbine is too low, however, so we're not done yet.

9.3.4 Addition of Reheat

To improve the steam quality at the turbine exit, we need to move to the right on the $T - s$ plot. We can do this and also increase the average temperature of heat addition by employing *reheat*, as shown in Fig. 9.7.

Due to the large pressure range we're expanding the steam through, the turbine must be constructed of multiple stages or else multiple separate turbines must be used (small ones for high pressure and progressively larger ones as the pressure becomes lower). In the reheat cycle, we expand the steam first through a high pressure turbine to some intermediate pressure, extract it and add more heat, and then expand it through a second low-pressure turbine. The effect of this is seen in the process representations (Fig. 9.8). The work output (the sum of both turbines) per unit mass of steam is increased, as is the efficiency.

For this example, we will take the intermediate reheat pressure to be 1 MPa. Then carrying out an analysis similar to what we've done before, we find

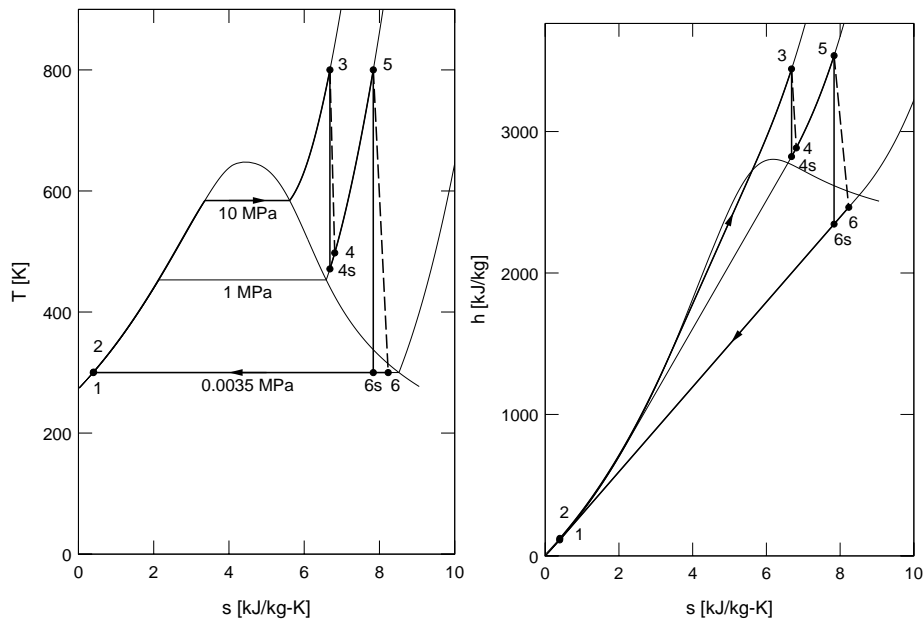


Figure 9.8: A Rankin cycle with superheat and one stage of reheat.

State	P (MPa)	T (K)	h (kJ/kg)	s (kJ/kg-K)	x
1	0.003536	300	111.7	0.3900	0
2	10	300.72	124.3	0.3993	0
3	10	800	3442.0	6.683	1
4s	1	471.2	2822.7	6.683	1
4	1	497.6	2884.7	6.811	1
5	1	800	3536.4	7.835	1
6s	.003536	300	2345.4	7.835	0.916
6	.003536	300	2464.5	8.232	0.965

Heat added in boiler / superheater:	3317.7 kJ/kg
Heat added in reheater:	651.7 kJ/kg
HP Turbine work:	557.3 kJ/kg
LP Turbine work:	1071.9 kJ/kg
Pump work:	12.5 kJ/kg
Efficiency:	40.7%
BWR:	0.0076

9.3.5 Regeneration

One remaining source of inefficiency is the heating of the water when it first enters the boiler. Since the water is initially cold (300 K), the heat addition required to bring the water to the boiling point constitutes low-temperature

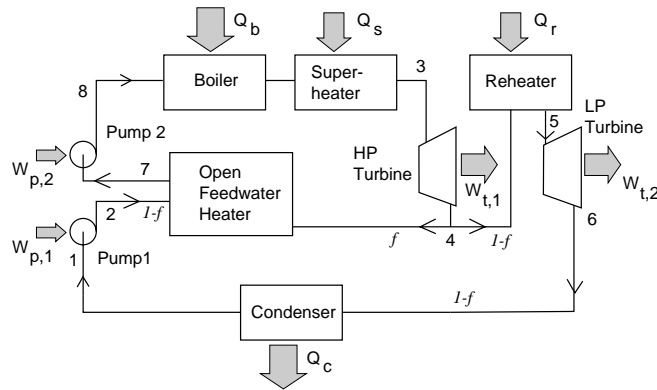


Figure 9.9: A Rankine cycle with superheat, one stage of reheat, and one open feedwater heater.

heat addition. If we can eliminate some of the low-temperature heat addition, we can raise the *average* temperature at which heat is added, and thus the efficiency of the cycle.

A technique known as *regeneration* uses heat from the hot steam to preheat the water going into the boiler (the “feedwater”). There are several ways to implement regeneration. A particularly simple way (which however only partially eliminates low-temperature external heat addition) is shown in Fig. 9.9. A portion f of the steam leaving the high-pressure turbine is extracted and simply mixed with the cold water before it enters the boiler. The remaining fraction $1 - f$ goes through the reheater, low-pressure turbine, and condenser as usual.

The chamber where the mixing occurs is known as an *open feedwater heater*. Since the mixing must occur at the pressure at the outlet of the high-pressure turbine, this scheme requires two pumps, as shown in Fig. 9.9. The water which leaves the feedwater heater is pumped up to the boiler pressure in the second pump. Since it is now hotter than it would have been without the feedwater heater, less heat must be added in the boiler to reach the boiling point.

Let’s consider the effect of adding an open feedwater heater to the previous example. We will specify that just enough steam is extracted at state 4 to produce saturated liquid at the outlet of the feedwater heater (state 7). Most of the states remain the same as in example 4. The only changes are state 2 ($P_2 = 1$ MPa now, not 10 MPa) and the two new states 7 and 8.

State 7 is determined by our specification that it is saturated liquid at 1 MPa, and state 8 is determined by analyzing pump 2, in the same way we found state 2 earlier. The properties in each state are listed below, and the $T - s$ plot

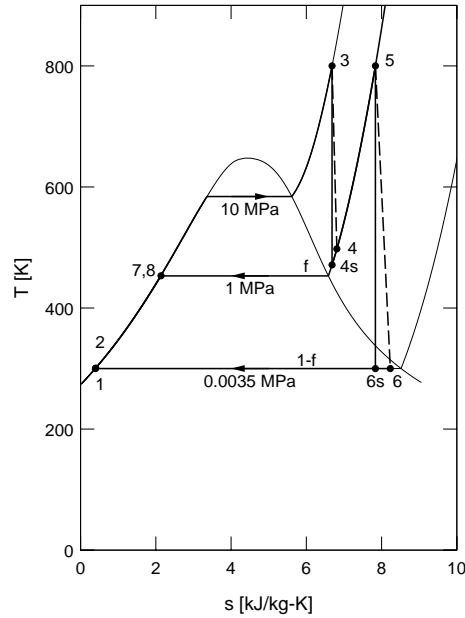


Figure 9.10: A Rankin cycle with superheat, one stage of reheat, and one open feedwater heater. Note that only a fraction f of the total mass flow rate flows from 4 to 7, and a fraction $1 - f$ flows in the loop 4-5-6-1.

for this cycle is shown in Fig. 9.10.

State	P (MPa)	T (K)	h (kJ/kg)	s (kJ/kg-K)	x
1	0.003536	300	111.7	0.3900	0
2	1	300.01	113.0	0.3915	0
3	10	800	3442.0	6.683	1
4s	1	471.2	2822.7	6.683	1
4	1	497.6	2884.7	6.811	1
5	1	800	3536.4	7.835	1
6s	.003536	300	2345.4	7.835	0.916
6	.003536	300	2464.5	8.232	0.965
7	1	453.06	762.5	2.138	0
8	10	454.92	775.2	2.144	0

Knowing all the states, we now need to calculate the extraction fraction f which is required to produce the desired state 7 (saturated liquid at 1 MPa). To find the extraction fraction, consider an energy balance on the feedwater heater:

$$\dot{m}_4 h_4 + \dot{m}_2 h_2 = \dot{m}_7 h_7. \quad (9.14)$$

Note that we have to keep track of the mass flow rate in each stream, since they are not equal. Dividing by \dot{m}_7 ,

$$fh_4 + (1 - f)h_2 = h_7. \quad (9.15)$$

Solving for f ,

$$f = \frac{h_7 - h_2}{h_4 - h_2} \quad (9.16)$$

which works out to $f = 0.234$.

Now we can calculate the energy transfer for each component. For the boiler, superheater, and pump 2, the full \dot{m} flows through them, and we calculate the work per unit mass in the usual way. But note that only $(1 - f)\dot{m}$ flows through pump 1, the reheater, and turbine 2. It is most convenient to express all energy transfers as the power divided by the total \dot{m} (not the actual mass flow rate through a given component). Therefore, we must multiply the enthalpy difference by the fraction of the mass flowing through a given component:

$$Q_b = h_3 - h_8 \quad (9.17)$$

$$W_{p,2} = h_8 - h_7 \quad (9.18)$$

$$W_{t,1} = h_3 - h_4 \quad (9.19)$$

$$Q_r = (1 - f)(h_5 - h_4) \quad (9.20)$$

$$W_{p,1} = (1 - f)(h_2 - h_1) \quad (9.21)$$

$$W_{t,2} = (1 - f)(h_5 - h_6) \quad (9.22)$$

Evaluating these expressions, we find

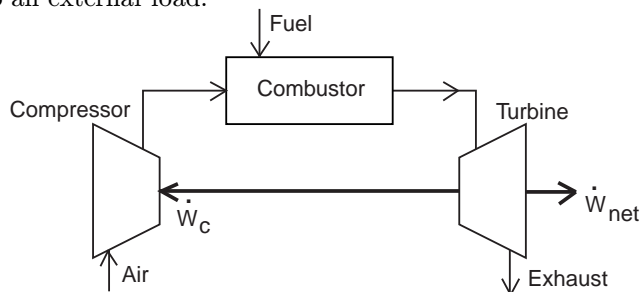
Heat added in boiler / superheater:	2666.8 kJ/kg
Heat added in reheater:	499.0 kJ/kg
Turbine work 1:	557.3 kJ/kg
Turbine work 2:	820.7 kJ/kg
Pump work 1:	.96 kJ/kg
Pump work 2:	12.68 kJ/kg
Extraction fraction:	23.4%
Efficiency:	43.1%
BWR:	0.0099

Using one open feedwater heater is seen to increase the efficiency of the cycle from 40.7% to 43.1%. While this may seem not much considering the added complexity, in fact a gain of 2.4 percentage points in efficiency is quite significant, and well-worth any added costs due to complexity in even a moderately-sized power plant.

A typical real power plant would use multiple feedwater heaters, with steam extracted from the turbine at various pressures, in order to achieve even more regeneration. For example, the largest unit at the Pasadena power plant makes use of 5 feedwater heaters.

9.4 Gas Power Cycles

A simple gas power cycle is shown below. A steady flow of air is compressed and enters a combustor, where fuel is injected into the high-pressure air stream and the fuel/air mixture burns. The hot products of combustion¹ expand through a turbine, producing power \dot{W}_t . Some fraction of the turbine power is used to drive the compressor, and the remaining power $\dot{W}_{net} = \dot{W}_t - \dot{W}_c$ may be delivered to an external load.



Variations on this basic *gas turbine cycle* are widely used both for power generation and propulsion. Aircraft jet engines use a modified version of this cycle, which we will discuss in more detail shortly. Gas turbines have become increasingly popular for ground-based power generation also, as the efficiency of these cycles has improved. They are particularly attractive as topping cycles in conjunction with vapor power cycles, or when relatively small, modular power generation units are needed.

The two most important performance parameters for a gas turbine cycle are the thermal efficiency η and the net work developed per unit mass of air flow, $W_{net} = \dot{W}_{net}/\dot{m}_{air}$. The efficiency determines the fuel economy; a more efficient gas turbine cycle for electric power generation consumes less fuel per kW-hr of electricity generated. This lowers both the cost of electricity and the pollutant emissions. For an aircraft engine, higher efficiency translates into lower cost, longer range, and lower pollution.

W_{net} is especially important for transportation applications, since the size

¹for lean combustion, a mixture composed primarily of nitrogen, oxygen, water vapor, carbon dioxide, and smaller amounts of other species, including pollutants such as oxides of nitrogen (NO_x).

and weight of the engine scale approximately with the air flow rate. For a fixed power output requirement, a cycle with higher W_{net} will be lighter and smaller.

Gas turbine engines are almost always run with significantly more air than required to burn all of the fuel injected (so-called “lean combustion”). A major reason lean combustion is used is to limit pollutant emissions. By burning lean, combustion is complete, so there is no unburned fuel vapor in the exhaust, nor is there much CO. Also, lean combustion limits the temperature of the gas leaving the combustor. This is important both to minimize the production of smog-producing oxides of nitrogen (NO_x) and to protect the turbine blades, which cannot survive excessive temperatures.

A complete analysis of a gas turbine cycle would require considering the combustion process, including calculating the amount of heat released, and the composition of the combustion products. A design of a real gas turbine cycle would consider this in some detail, with particular focus on pollutant formation, since the emissions from both ground-based gas turbines and aircraft engines are strictly regulated by government agencies.

9.4.1 The air-standard Brayton cycle

We can learn a lot about gas turbine cycles by considering a simple model which captures their essential features, without requiring full consideration of the combustion process. Since the gas mixture is mostly air, even downstream of the combustor, as a first approximation we may take the gas to be air throughout the cycle. In this approximation, the combustor is modeled as an air heater, with the heat of combustion now an external heat input. Since in this *air-standard cycle* the fluid is the same at the turbine exhaust as at the compressor inlet, we may close the cycle by adding a cooler, which removes heat from the air to the environment. The cycle which results from these approximations is called the air-standard *Brayton cycle*, shown in Fig. 9.11. It is also called the *simple* Brayton cycle, to distinguish it from modified versions we will discuss below.

To analyze the air-standard Brayton cycle, we will assume the compressor and turbine are both adiabatic, with isentropic efficiencies $\eta_{s,c}$ and $\eta_{s,t}$, respectively. We will neglect any pressure drop in the heater and cooler, so these are modeled as constant-pressure heat addition or removal processes. With these assumptions, the T - s process representation is as shown in Fig. 9.11.

Let us put in some “typical” numbers. Suppose the inlet to the compressor is air at 1 atm and 300 K, the pressure ratio $P^* = P_2/P_1$ across the compressor is 4, and the temperature at the inlet to the turbine is fixed at 800 K by materials

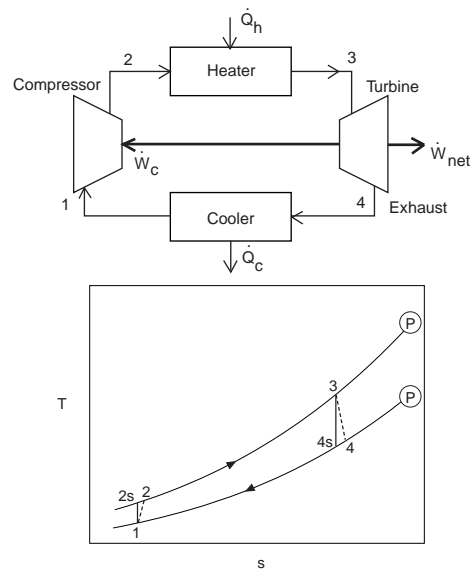


Figure 9.11: The air-standard Brayton cycle and T - s process representation.

constraints. The compressor has isentropic efficiency $\eta_{s,c} = 0.7$, and the turbine has isentropic efficiency $\eta_{s,t} = 0.8$.

To determine the performance of this cycle, we first must find all states, and their properties. The procedure is very much like what we did to analyze vapor power cycles: start with a known state (state 1), and simply work around the loop. Since this procedure is now familiar to you, we will simply summarize the set of relations needed in Table 9.1.

The equations in Table 9.1 may be easily implemented in Excel, using TPX to calculate the properties. Since TPX does not calculate properties for air, we will calculate properties for nitrogen, which of course is the major constituent (78%) of air, with most of the remainder being oxygen. For the parameter values specified above, the states are as follows:

state	T (K)	P (atm)	h (kJ/kg)	s (kJ/kg-K)
1	300	1	461.5	4.41
2s	445.5	4	613.0	4.41
2	507.1	4	677.9	4.55
3	800	4	996.4	5.05
4s	548.2	1	721.6	5.05
4	599.6	1	776.6	5.14

Table 9.1: Relations needed to solve for the states in a simple air-standard Brayton cycle.

state	T	P	h	s
1	given	given	$h(T, P)$	$s(T, P)$
2s	$T(s, P)$	P_2	$h(s, P)$	s_1
2	$T(h, P)$	given	$h_1 + (h_{2s} - h_1)/\eta_{s,c}$	$s(h, P)$
3	given	P_2	$h(T, P)$	$s(T, P)$
4s	$T(s, P)$	P_1	$h(s, P)$	s_3
4	$T(h, P)$	P_1	$h_3 - \eta_{s,t}(h_3 - h_{4s})$	$s(h, P)$

The turbine work output per kg of air is given by

$$W_t = h_3 - h_4 = 219.8 \text{ kJ/kg.}$$

The compressor work input is

$$W_c = h_2 - h_1 = 216.4 \text{ kJ/kg.}$$

We see that almost *all* of the turbine work output must be used simply to run the compressor! Only 3.4 kJ/kg is left over to deliver to an external load. In terms of the back work ratio, $\text{BWR} = 0.98$. This is in sharp contrast to the results we obtained for vapor power cycles, which had $\text{BWR} < 0.01$. This is a major difference between these two classes of cycles.

The heat input for this example is

$$Q_h = h_3 - h_2 = 318.5 \text{ kJ/kg,}$$

so the overall thermal efficiency is

$$\eta = (W_t - W_c)/Q_h = 0.011.$$

Therefore, this cycle only converts 1.1 % of the heat input into useful work — terrible performance, considering that the Carnot efficiency evaluated between the minimum and maximum temperatures in this cycle is $1 - 300/800 = 0.625$, or 62.5%.

Actually, the numbers used in this example are characteristic of the state of the art several decades ago, when gas turbine engines were first being developed. Due to intensive research, it is now possible to build compressors with isentropic efficiency in the range 0.8–0.9, and turbines with isentropic efficiency as high as 0.95. If we re-work the above example with, say, $\eta_{s,c} = 0.85$ and $\eta_{s,t} = 0.9$, we find $W_{net} = 69.1 \text{ kJ/kg}$ and $\eta = 0.19$ — quite a substantial improvement.

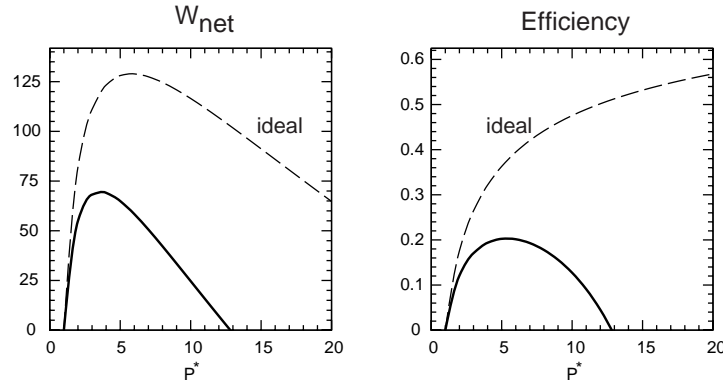


Figure 9.12: Net work per kg of air and efficiency vs. P^* . $T_1 = 300$ K, $P_1 = 1$ atm, $T_3 = 800$ K. The dashed curves are for an ideal cycle, and the solid curves for $\eta_{s,c} = 0.85$, $\eta_{s,t} = 0.9$.

Effect of Pressure Ratio on Performance

The effect of pressure ratio $P^* = P_2/P_1$ on W_{net} and on η for these conditions is shown in Fig. 9.12. For an ideal cycle with $\eta_{s,c} = \eta_{s,t} = 1$ (dashed curves), the net work per kg of air is maximized for P^* near 6, while the efficiency continues to increase with P^* . Taking the more realistic values $\eta_{s,c} = 0.85$, $\eta_{s,t} = 0.9$, the solid curves are obtained. In this case, the net work is maximized near $P^* = 4$. The efficiency reaches a maximum of about 20% near $P^* = 5$, and falls off at higher P^* , unlike the ideal case. A T - s plot for this cycle for $P^* = 5$ is shown in Fig. 9.13.

The dependence of W_{net} and η on P^* shown in Fig. 9.12 for the *ideal* case can be explained as in Fig. 9.14. For the ideal cycle, the enclosed area is W_{net} . Since $T_{max} = T_3$ is fixed, the area is maximized for an intermediate pressure ratio. The *efficiency* of an ideal cycle, on the other hand, is given by $\eta = 1 - \overline{T_c}/\overline{T_h}$. This is an increasing function of P^* , and approaches the Carnot efficiency $1 - T_1/T_{max}$ as P^* approaches P_{max}^* , which is the pressure ratio at which the air emerges from the compressor at T_{max} .

The Effect of Turbine Inlet Temperature on Performance

Another important development in the last couple of decades has been in advanced high-temperature materials for turbine blades, and development of blade cooling techniques. In the most advanced current gas turbines, turbine inlet temperatures up to 1700 K can be tolerated. If we recompute the above cycle performance using, say, $T_3 = 1400$ K, we find BWR = 0.41, $W_{net} = 261.4$ kJ/kg,

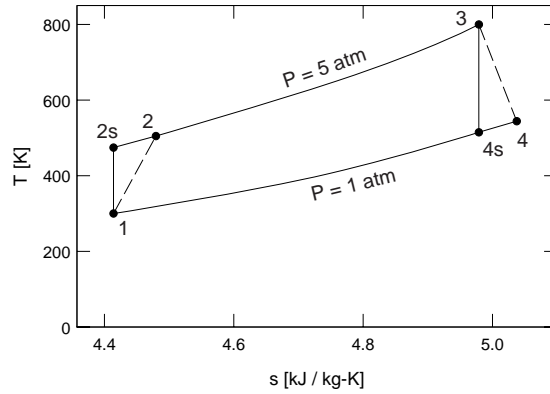


Figure 9.13: Process representation for simple Brayton cycle with $T_1 = 300$ K, $P_1 = 1$ atm, $P^* = 5$, $T_3 = 800$ K, $\eta_{s,c} = 0.85$, $\eta_{s,t} = 0.9$.

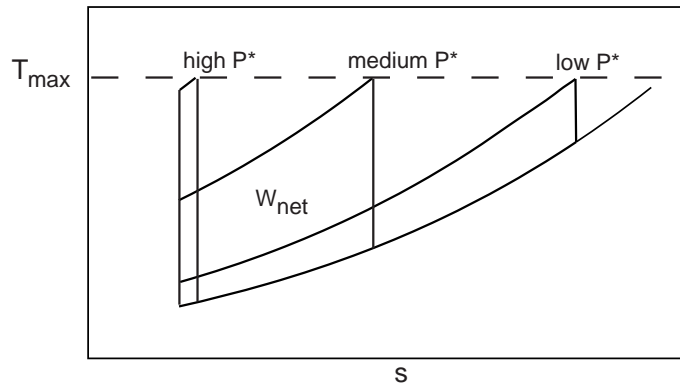


Figure 9.14: An intermediate P^* maximizes the enclosed area W_{net} for fixed T_{max} .

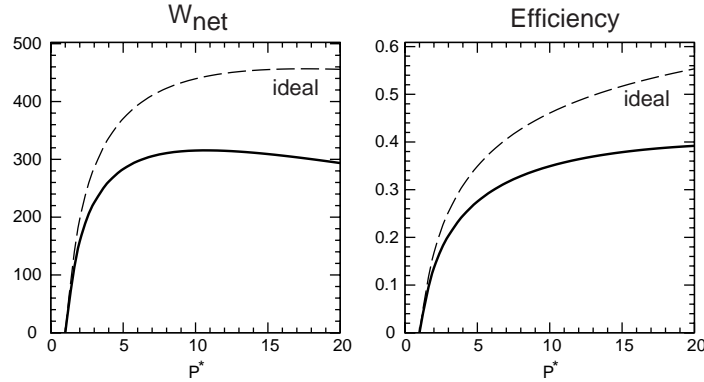


Figure 9.15: Net work per kg of air and efficiency vs. P^* . $T_1 = 300$ K, $P_1 = 1$ atm, $T_3 = 1400$ K. The dashed curves are for the ideal cycle, and the solid curves for $\eta_{s,c} = 0.85$, $\eta_{s,t} = 0.9$.

and $\eta = 0.25$. For this higher value of T_3 , the effect of pressure ratio is as shown in Fig. 9.15. The optimal pressure ratio is higher — a value of $P^* = 10$ would be a reasonable choice. This choice would result in an efficiency of 34.9%, and $W_{net} = 315.3$ kJ/kg. The T - s plot for this case is shown in Fig. 9.16.

Analysis assuming ideal gas with constant c_p

Since the peak pressure in a gas-turbine cycle is typically not too high (say 10 atm), air-standard cycles are often analyzed by treating air as an ideal gas. For even greater simplicity, the specific heat is often assumed constant. The ideal gas approximation should be fairly good at these pressures (cf. Fig. 4.11), but the assumption of constant c_p is more questionable, due to the large temperature variation around the cycle. Let's examine how much error is made for the conditions of Fig. 9.15.

If we assume an ideal gas with constant c_p , then

$$h(T) - h(T_0) = c_p(T - T_0) \quad (9.23)$$

and

$$s(T, P) - s(T_0, P_0) = c_p \ln \left(\frac{T}{T_0} \right) - R \ln \left(\frac{P}{P_0} \right). \quad (9.24)$$

Therefore, for the isentropic steps such as $1 \rightarrow 2s$,

$$c_p \ln \left(\frac{T_{2s}}{T_1} \right) = R \ln \left(\frac{P_{2s}}{P_1} \right). \quad (9.25)$$

Since $R/c_p = (c_p - c_v)/c_p = 1 - c_v/c_p$ and $P_{2s}/P_1 = P^*$, this equation may be

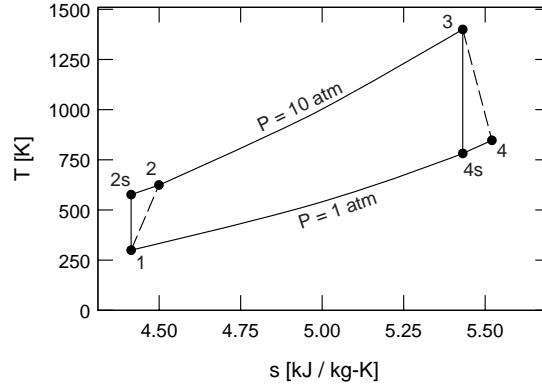


Figure 9.16: Process representation for a simple Brayton cycle with $T_1 = 300$ K, $P_1 = 1$ atm, $P^* = 10$, $T_3 = 1400$ K, $\eta_{s,c} = 0.85$, $\eta_{s,t} = 0.9$. Performance: $W_{net} = 315.3$ kJ/kg, $\eta = 0.349$.

written

$$T_{2s} = T_1(P^*)^{1-1/k}, \quad (9.26)$$

where the specific heat ratio k is defined as

$$k = \frac{c_p}{c_v}. \quad (9.27)$$

Similarly,

$$T_{4s} = T_3(1/P^*)^{1-1/k}. \quad (9.28)$$

These expressions may be used instead of TPX to solve for the temperatures and enthalpies in Table 9.1. For nitrogen at room temperature, $c_p = 1.04$ kJ/kg, and $k = (7/2)/(5/2) = 1.4$ (for air, $c_p = 1.004$ kJ/kg, and k is also 1.4). If we re-do the cycle analysis as a function of P^* with the ideal gas, constant c_p approximation, we come up with the dashed curves in Fig. 9.17, which are compared to the results using TPX (solid curves). The approximation is seen to be not too bad; certainly the trends with P^* are well-predicted. The major effect is an underprediction of W_{net} by roughly 10%, and a slight over-prediction of η .

9.4.2 The Regenerative Brayton Cycle

Although the cycle of Fig. 9.16 is far more efficient than the one we first considered (34.9% vs. 1.1%), more improvements are still possible. Note that due to the high turbine inlet temperature of 1400 K, the turbine exhaust (state 4) is still quite hot — 847 K. Since this temperature is higher than the air temperature at the exit of the compressor (state 2; 624 K), it should be possible to

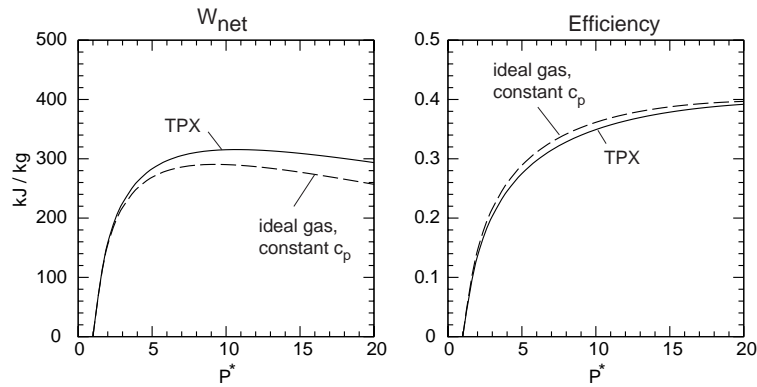


Figure 9.17: Comparison of results calculated assuming ideal gas with constant c_p to results calculated using TPX, for cycle conditions of Fig. 9.15.

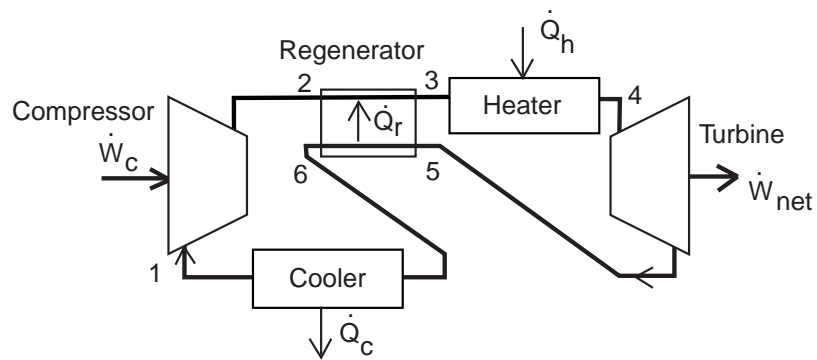


Figure 9.18: The regenerative air-standard Brayton cycle.

Table 9.2: Relations needed to solve for the states in a regenerative air-standard Brayton cycle.

state	T	P	h	s
1	given	given	$h(T, P)$	$s(T, P)$
2s	$T(s, P)$	P_2	$h(s, P)$	s_1
2	$T(h, P)$	given	$h_1 + (h_{2s} - h_1)/\eta_{s,c}$	$s(h, P)$
3	$T(h, P)$	given	$h_2 + \epsilon_r(h_5 - h_2)$	$s(h, P)$
4	given	P_2	$h(T, P)$	$s(T, P)$
5s	$T(s, P)$	P_1	$h(s, P)$	s_4
5	$T(h, P)$	P_1	$h_4 - \eta_{s,t}(h_4 - h_{5s})$	$s(h, P)$
6	$T(h, P)$	P_1	$h_5 - (h_3 - h_2)$	$s(h, P)$

use the turbine exhaust to pre-heat the air entering the heater, thus requiring less external heat input Q_h to reach 1400 K. This can be done using a heat exchanger, as shown in Fig. 9.18. Since the work output is unchanged by this modification, the efficiency W_{net}/Q_h will increase.

The heat exchanger, called the *regenerator*, transfers heat from the hot turbine exhaust to the air emerging from the compressor. The heat transfer in the regenerator per kg of air is

$$Q_r = h_3 - h_2 \quad (9.29)$$

$$= h_5 - h_6. \quad (9.30)$$

If the regenerator were infinitely long, the air leaving it at state 3 would have the same temperature as the hot air entering at state 5. Thus, the *most* heat transfer which could occur in the regenerator is given by

$$Q_{r,max} = h_5 - h_2. \quad (9.31)$$

It is impractical to build a regenerator so large that T_3 closely approaches T_5 — in practice, T_3 will always be less than T_5 , and therefore $Q_r < Q_{r,max}$.

The *regenerator effectiveness* is defined by

$$\epsilon_r = \frac{Q_r}{Q_{r,max}} = \frac{h_3 - h_2}{h_5 - h_2}. \quad (9.32)$$

A typical realistic value for ϵ_r would be about 0.8. With ϵ_r specified, this equation may be used to solve for h_3 . The complete set of relations needed to solve the regenerative air-standard Brayton cycle is given in Table 9.2.

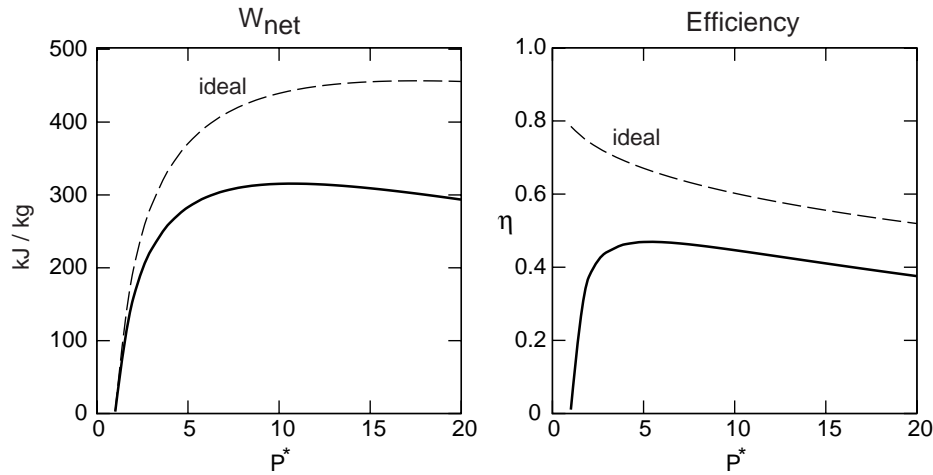


Figure 9.19: Net work per kg of air and efficiency vs. P^* . $T_1 = 300$ K, $P_1 = 1$ atm, $T_3 = 1400$ K, $\epsilon_r = 0.8$. The dashed curves are for the ideal cycle, and the solid curves for $\eta_{s,c} = 0.85$, $\eta_{s,t} = 0.9$.

If we add a regenerator with $\epsilon_r = 0.8$ to the simple Brayton cycle of Fig. 9.15, and use TPX to calculate the state properties, we find the performance characteristics shown in Fig. 9.19. The net work is unchanged from Fig. 9.15, but the efficiency is much different. For the ideal case ($\epsilon_r = \eta_{s,c} = \eta_{s,t} = 1$), the efficiency is maximized at $P^* = 1$ (although of course W_{net} is zero there).

For the non-ideal case, the efficiency is maximized near $P^* = 5$, at a value of 46.9%. This is much better than the simple Brayton cycle, which from Fig. 9.15 has an efficiency of 27.5% at $P^* = 5$. Adding a regenerator is seen to produce a big increase in efficiency, particularly at lower P^* values, where the difference ($T_5 - T_2$) is largest.

Based upon the performance results shown in Fig. 9.19, the best choice for P^* would likely be between 5 and 10, depending on whether maximizing W_{net} or η was more important. The T - s plot for $P^* = 10$ is shown in Fig. 9.20. Note that this is the same as Fig. 9.16, except for the location of the state points.

9.4.3 Intercooling and Reheat

Two more modifications are often used. In order to reduce the compressor work requirement, the compression process may be broken up into two or more stages, with heat removal between stages. This process is known as *intercooling*.

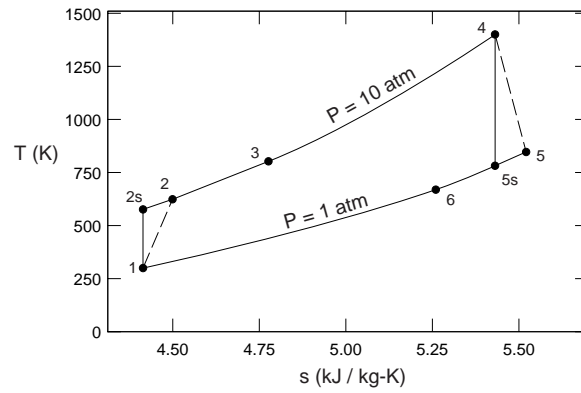
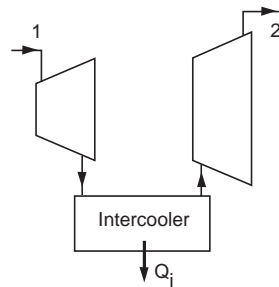
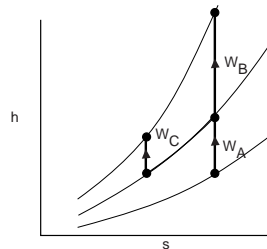


Figure 9.20: Process representation for a regenerative Brayton cycle with $T_1 = 300$ K, $P_1 = 1$ atm, $P^* = 10$, $T_3 = 1400$ K, $\eta_{s,c} = 0.85$, $\eta_{s,t} = 0.9$, $\epsilon_r = 0.8$. Performance: $W_{net} = 315.3$ kJ/kg, $\eta = 0.447$.



That this lowers the work requirement is easiest to see by considering an h - s process representation of an ideal, 2-stage compressor (below). With no heat removal between stages, the total compressor work to compress from P_1 to P_2 is $W_A + W_B$. With intercooling the work is $W_A + W_C$. Since constant-pressure lines on an h - s plot diverge (slope = T), $W_C < W_B$ and thus the compressor work is lower with intercooling.



Reheat is the same process as for vapor power cycles — heat is added between turbine stages. Both intercooling and reheat increase W_{net} , but may or may not

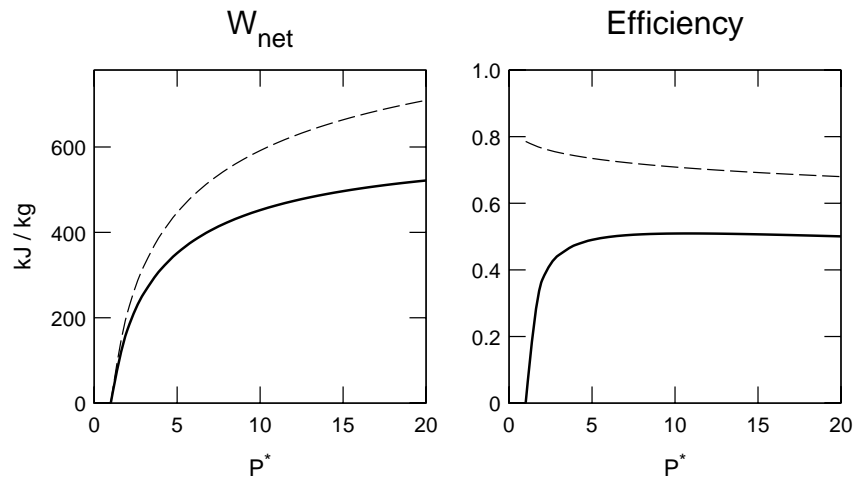


Figure 9.21: Net work per kg of air and efficiency vs. P^* for regenerative Brayton cycle with intercooling and reheat.

improve the efficiency. If they are used together with regeneration, significant gains in efficiency are possible, since they increase the potential for regeneration.

In Fig. 9.21, the performance of a regenerative cycle with one stage of intercooling and one stage of reheat is presented. Most of the parameters are the same as in Fig. 9.19. The intercooler effectiveness is 0.8, and the reheat temperature is 1400 K. The intercooling and reheat pressures are both taken to be $(P^*)^{1/2}$. (For an ideal cycle and assuming constant c_p , it can be shown this is the optimal value.)

It can be seen that the performance is quite good. The net work is above 400 kJ/kg, and above $P^* = 5$ the efficiency is slightly greater than 50%.

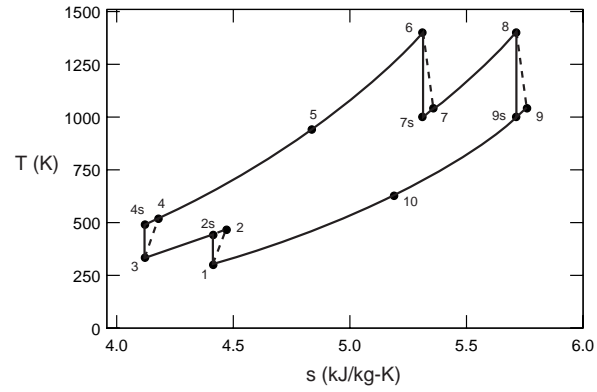


Figure 9.22: Process representation for a regenerative Brayton cycle with intercooling and reheat. $T_1 = 300$ K, $P_1 = 1$ atm, $P^* = 10$, $P_i = P_r = 3.16$ atm, $T_3 = 1400$ K, $\eta_{s,c} = 0.85$, $\eta_{s,t} = 0.9$, $\epsilon_r = \epsilon_i = 0.8$.

CHAPTER 10

CHEMICAL REACTIONS AND CHEMICAL EQUILIBRIUM

10.1 Introduction

(NOTE: This chapter is incomplete; consult the hand-written notes I passed out in class.)

Chemically-reacting flows are vital to many engineering applications, including almost all aerospace and terrestrial propulsion systems, and all systems which produce power by burning fossil fuels. Solving current technological problems, such as how to reduce undesirable pollutant emission from these devices, requires an understanding of some of the chemistry occurring in these systems.

10.2 The Reaction $A + B \rightleftharpoons AB$

Consider a mixture of 2 chemically-reactive elemental species, A and B. Suppose that these can react with one another to form a molecule AB:



Once an AB molecule is formed, it may break up again via



We can write a short-hand notation for these two reactions as follows:



Suppose we introduce $N_{A,0}$ moles of A and $N_{B,0}$ moles of B into a container with volume V at time $t = 0$. We allow the system to evolve with no interaction with the surroundings (isolated). As time proceeds, we find that the mole numbers change, due to formation of AB molecules. If we wait long enough, the mole numbers come to steady-state values, which we will denote N_A^* , N_B^* , and N_{AB}^* . Evidently, once steady state is reached, the rate at which AB is formed by reaction (10.1) just balances the rate at which it is destroyed by reaction (10.2).

When this condition is reached, we say the system is in *chemical equilibrium*. Depending on the nature of the chemical reaction, and also on the temperature

and pressure, chemical equilibrium may be reached in microseconds (explosions) or may take thousands of years. Determining the rate at which the reaction proceeds is the subject of chemical kinetics. However, the final state achieved at long times does not depend on kinetics, but only on the thermodynamic properties of the atoms or molecules involved.

To determine the condition for chemical equilibrium, we apply the same basic principle we have used to determine other equilibrium conditions: the system will attain the state for which the entropy is maximal, compared to all other possible states with the same total U , V , $N_{A,0}$, and $N_{B,0}$.

In our previous discussions of equilibrium, we have imagined a system divided by a partition, which can pass energy, volume, or molecules. For the present problem, the “constraint” we will relax is somewhat different. Let us imagine that we can turn a knob on the system which can (somehow) stop the reaction from occurring, or can run it in the forward direction [reaction (10.1)] or in the reverse direction [reaction (10.2)]. We will turn the knob back and forth, calculating the entropy of the system for each set of conditions, and find the state which maximizes S . This then must be the equilibrium state the system will come to when “unconstrained.”

Suppose we initially set the knob so that the reaction is stopped, and the mole numbers are N_A , N_B , and N_{AB} . We now turn the knob to allow the reaction to proceed infinitesimally in the forward direction, and then reset it to stop the reaction at this new condition. We find that N_{AB} has increased by an amount dN_{AB} . Since the only way to increase N_{AB} is by consuming an equivalent amount of A and B atoms, we must have that

$$dN_A = -dN_{AB} \quad (10.4)$$

$$dN_B = -dN_{AB}. \quad (10.5)$$

How much has $S(U, V, N_A, N_B, N_{AB})$ changed? Using the chain rule,

$$\begin{aligned} dS &= \left(\frac{\partial S}{\partial N_A} \right)_{U, V, N_B, N_{AB}} dN_A + \left(\frac{\partial S}{\partial N_B} \right)_{U, V, N_A, N_{AB}} dN_B \\ &\quad + \left(\frac{\partial S}{\partial N_{AB}} \right)_{U, V, N_A, N_B} dN_{AB}. \end{aligned} \quad (10.6)$$

We know that the chemical potential is defined by

$$\left(\frac{\partial S}{\partial N_i} \right)_{U, V, N_j} = -\frac{\mu_i}{T}. \quad (10.7)$$

Therefore, if we also use equations (10.4) and (10.5) to substitute for dN_A and dN_B , Eq. (10.6) becomes

$$dS = \frac{1}{T} [\mu_A + \mu_B - \mu_{AB}] dN_{AB}. \quad (10.8)$$

At the equilibrium state, S will be maximal with respect to N_{AB} , and therefore $dS = 0$. The chemical equilibrium condition for this reaction is then

$$\mu_{AB} - \mu_A - \mu_B = 0. \quad (10.9)$$

This equilibrium condition is very general, and applies whatever type of solution the species A , B , and AB form. For example, it could be a liquid or even a solid solution. By using the form for the chemical potential appropriate for the phase of the mixture, we can develop an explicit condition for equilibrium in terms of experimental parameters, such as the mole numbers, temperature, and pressure.

It is worth noting that, although we imagined an isolated system to derive Eq. (10.9), now that we have this equation we no longer need to restrict our attention only to isolated systems. The condition expressed by Eq. (10.9) must be satisfied at the equilibrium state, *no matter how this state was reached*. It could have been reached in an isolated system (constant U and V) or, for example, by allowing the reaction to proceed at constant T and P , adding or removing heat as necessary to keep T constant, and allowing the volume to change to keep P constant.

10.3 Reactions in Ideal Gases

If the mixture of A , B , and AB is an ideal gas mixture, then we have already derived the form of the equation for the chemical potential. In this case, recall that

$$\mu_i = \hat{g}(T, P_i) = \hat{g}^0(T) - \hat{R}T \ln \left(\frac{P_i}{P_0} \right) \quad (10.10)$$

That is, the chemical potential of species i in the mixture is the molar Gibbs free energy (of the pure species i) evaluated at the partial pressure P_i .

Substituting this expression into Eq. (10.9), we have

$$\ln \left[\frac{(P_A/P_0)(P_B/P_0)}{(P_{AB}/P_0)} \right] = - \frac{\hat{g}_A^0(T) + \hat{g}_B^0(T) - \hat{g}_{AB}^0(T)}{\hat{R}T} \quad (10.11)$$

or

$$\frac{P_A P_B}{P_{AB}} = P_0 e^{-\Delta G^0(T)/\hat{R}T}, \quad (10.12)$$

where

$$\Delta G^0(T) = \hat{g}_{AB}^0(T) - \hat{g}_A^0(T) - \hat{g}_B^0(T) \quad (10.13)$$

is the *standard free energy of reaction* for this reaction. We can re-write this equation in terms of mole numbers as

$$\frac{N_A^* N_B^*}{N_{AB}^*} = \left(\frac{P_0 V}{\hat{R} T} \right) e^{-\Delta G^0(T)/RT}. \quad (10.14)$$

This equation provides us with one equation for the 3 unknowns N_A^* , N_B^* , and N_{AB}^* . The remaining 2 equations come from the *atom conservation* conditions:

$$N_A^* + N_{AB}^* = N_{A,0} \quad (10.15)$$

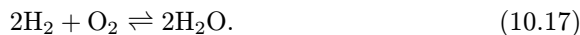
and

$$N_B^* + N_{AB}^* = N_{B,0}. \quad (10.16)$$

These equations simply state that the number of moles of A atoms (independent of whether they appear as free A atoms or in AB molecules) is the same as initially placed in the container, and similarly for B. These three equations together allow solving uniquely for the values of N_A^* , N_B^* , and N_{AB}^* at equilibrium.

10.4 A General Reaction

The above results are easily generalized to any chemical reaction. Let's consider a reaction such as



We may write this reaction (or any other one) symbolically as

$$\sum_j \nu_j A_j = 0, \quad (10.18)$$

where A_j stands for the chemical symbol of the j th species (e.g., “H₂”, “O₂”, or “H₂O”) and ν_j is the *stoichiometric coefficient* for species j . The stoichiometric coefficient is defined to be positive for products (H₂O) and negative for reactants (H₂ and O₂). Therefore, for this reaction

$$\nu(\text{H}_2) = -2 \quad (10.19)$$

$$\nu(\text{O}_2) = -1 \quad (10.20)$$

$$\nu(\text{H}_2\text{O}) = +2 \quad (10.21)$$

To find the equilibrium condition for a general reaction such as reaction (10.18), we again maximize the entropy at fixed U and V as we did above. For

the general case, the stoichiometric relation among the mole numbers as the reaction proceeds forward or backward is:

$$\frac{dN_j}{\nu_j} = d\tilde{N}. \quad (10.22)$$

That is, each mole number changes by an amount equal to ν_j multiplied by the change in a species-independent “extent of reaction” variable \tilde{N} . Note that for the simple $A + B \rightleftharpoons AB$ reaction considered above, Eq. (10.22) reduces to equations (10.4) and (10.5).

The condition that S is maximal with respect to \tilde{N} then leads to

$$\sum_{j=1}^r \mu_j \nu_j = 0. \quad (10.23)$$

This is the general form of the condition for chemical equilibrium for a single reaction. (Note that ν_j is non-zero only for those species in the mixture which participate in the reaction.)

For ideal gases, we can again substitute in the known form of the chemical potential. This leads us to the generalization of Eq. (10.12):

$$\prod_{j=1}^r \left(\frac{P_j}{P_0} \right)^{\nu_j} = e^{-\Delta G^0(T)/\hat{R}T}, \quad (10.24)$$

where

$$\Delta G^0(T) = \sum_{j=1}^r \hat{g}_j^0(T) \nu_j \quad (10.25)$$

is again the standard free energy of reaction. Equation (10.24) is known as the *law of mass action*.

Since the right-hand side of Eq. (10.24) is a function only of temperature, it is convenient to give it a symbol. We define the *equilibrium constant* $K_p(T)$ as

$$K_p(T) = e^{-\Delta G^0(T)/\hat{R}T}. \quad (10.26)$$

We can write the law of mass action in terms of mole fractions, since $P_j = X_j P$. Making this substitution into Eq. (10.24), we have

$$\prod_{j=1}^r X_j^{\nu_j} = \left(\frac{P_0}{P} \right)^{\sum_{j=1}^r \nu_j} K_p(T). \quad (10.27)$$

We can also write the law of mass action in terms of mole numbers or concentrations $C_j = N_j/V$ as follows:

$$\prod_{j=1}^r C_j^{\nu_j} = \left(\frac{P_0}{\hat{R}T} \right)^{\sum_{j=1}^r \nu_j} K_p(T), \quad (10.28)$$

and

$$\prod_{j=1}^r N_j^{\nu_j} = \left(\frac{P_0 V}{\hat{R}T} \right)^{\sum_{j=1}^r \nu_j} K_p(T). \quad (10.29)$$

All of the forms (10.27), (10.28), and (10.29) of the law of mass action have an additional factor in front of $K_p(T)$, which arises due to the conversion from pressure to mole fraction, concentration, or mole number units. The exponent this factor is raised to is the *net* change in mole numbers as the reaction proceeds from pure reactants to pure products. For example, for the reaction $A+B \rightleftharpoons AB$, $\sum_{j=1}^r \nu_j = -1$. On the other hand, for a reaction of the type $A + B \rightleftharpoons C + D$, this sum is zero, and the factor in front of $K_p(T)$ does not appear.

Remember that P_0 is the *reference* pressure at which the thermodynamic properties are computed, not the actual pressure of the problem. It might seem odd that P_0 appears in the law of mass action, since the equilibrium composition can't depend on our choice of the reference pressure. In fact, it does not, since $\Delta G^0(T)$ depends on P_0 in just such a way that it cancels the P_0 factor multiplying $K_p(T)$. (Recall that $\hat{g}(T, P)$ has an $\hat{R}T \ln P$ term in it.)

If the reference pressure $P_0 = 1$ atm, then

$$\frac{P_0}{\hat{R}T} = 40.62 \left(\frac{300}{T} \right) \text{ moles/m}^3. \quad (10.30)$$

Let us see how this works in practice by looking at a simple example.

Example 10.1 Suppose oxygen gas (O_2) is heated to 2500 K. We wish to know the mole fraction of atomic oxygen O which forms at chemical equilibrium.

The reaction forming O is



Therefore,

$$\Delta G^0 = 2\hat{g}(O) - \hat{g}(O_2). \quad (10.32)$$

We can look up the molar Gibbs functions in a table such as the JANAF tables, or calculate them from polynomial fits (using $\hat{g} = \hat{h} - T\hat{s}$). Doing this, we find that for $T = 2500$ K,

$$\hat{g}(O) = -219.13 \text{ kJ/mole} \quad (10.33)$$

$$\hat{g}(O_2) = -614.67 \text{ kJ/mole.} \quad (10.34)$$

Therefore,

$$\Delta G^0(2500) = 176.41 \text{ kJ/mole} \quad (10.35)$$

and

$$K_p(2500) = 2.06 \times 10^{-4}. \quad (10.36)$$

The law of mass action then becomes

$$\frac{X_{\text{O}}^2}{X_{\text{O}_2}} = \left(\frac{P_0}{P}\right)^{+1} K_p(T). \quad (10.37)$$

The atom conservation condition, expressed in mole fraction units, is simply

$$X_{\text{O}} + X_{\text{O}_2} = 1. \quad (10.38)$$

Therefore,

$$\frac{X_{\text{O}}^2}{1 - X_{\text{O}}} = 2.06 \times 10^{-4} \left(\frac{P_0}{P}\right). \quad (10.39)$$

We can solve this simple quadratic equation for the atomic oxygen mole fraction if the pressure is specified. As P increases, the atomic oxygen mole fraction decreases. For example, at $P = 1$ atm at 2500 K, we have that $X_{\text{O}} = 0.0142$. But at $P = 0.01$ atm, $X_{\text{O}} = 0.133$.

The pressure dependence seen in Example 10.1 occurs because the mole numbers change by +1 in this reaction. A reaction of the type $A + B \rightleftharpoons C + D$, in which the mole numbers do not change, would have no pressure dependence to the equilibrium mole fractions, since the term (P_0/P) in the law of mass action would not appear (the exponent would be zero).

This is one example of *Le Chatelier's Principle*, which states that when a change is imposed on a system (such as increasing the pressure), the system will respond in such a way to try to counteract the change. In this case, as the pressure is increased, some O atoms recombine to O_2 in order to reduce the total number of moles, and therefore reduce the system pressure compared to what it would be if the composition remained constant.

CHAPTER 11

ONE-DIMENSIONAL COMPRESSIBLE FLOW

11.1 Introduction

Thermodynamics plays an important role in fluid mechanics, particularly when the flow speed is high (comparable to the speed of sound). High-speed flow in gases is called *compressible flow*, since it turns out that only in this case does the density vary much from point to point in the flow; if the velocity is much less than the speed of sound in a gas, then the density is nearly constant everywhere. In this chapter, we introduce some of the basic ideas of compressible flow.

11.2 The Momentum Principle

To analyze fluid flow, we need a principle from mechanics to supplement the laws of thermodynamics. This is the *momentum principle*. When fluid enters or leaves a control volume, it carries momentum with it. Since a packet of fluid of mass δm moving with velocity \vec{V} has momentum $(\delta m)\vec{V}$, the specific momentum (momentum per unit mass) of the fluid packet is simply equal to the velocity \vec{V} . Therefore, if fluid is flowing into a control volume with a mass flow rate of \dot{m} kg/s and each kg brings momentum \vec{V} with it, the momentum inflow rate is $\dot{m}\vec{V}$.

The momentum principle is simply Newton's second law applied to a control volume. It states that the net force on a control volume is equal to the *net* rate of momentum outflow from the control volume, plus the rate of change of momentum stored inside. The momentum stored inside the control volume only

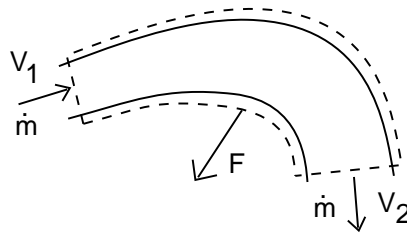


Figure 11.1: Flow in a curved tube.

changes in unsteady flows, which we will not consider here.

Consider the steady flow in a curved tube shown in Fig. 11.1. The flow is assumed to be one-dimensional at points 1 and 2, which means that the velocity is constant over any cross-section and is oriented perpendicular to the cross-section. In this flow, the momentum inflow rate is $\dot{m}\vec{V}_1$, and the momentum outflow rate is $\dot{m}\vec{V}_2$. The *net* momentum outflow rate is $\dot{m}(\vec{V}_2 - \vec{V}_1)$. Therefore, the momentum principle states that the net force \vec{F} on the control volume is

$$\vec{F} = \dot{m} (\vec{V}_2 - \vec{V}_1). \quad (11.1)$$

Note that \vec{F} is the *resultant* of all forces applied to the control volume. The momentum principle doesn't tell us anything about *where* these forces appear, just their vector sum. In general, the forces on the control volume result both from pressure in the fluid where the control surface intersects the fluid, and from stresses in the tube walls where the control surface intersects the walls.

11.3 The Speed of Sound

Sound waves are small pressure disturbances, which travel through a fluid with a characteristic speed. As long as the wavelength of the sound wave λ is long compared to the mean free path in the fluid,¹ the speed of propagation does not depend on the wavelength. (If you listen to a concert from far away the high notes and the low notes both reach you at the same time.)

To calculate the speed of sound, consider the situation shown in Fig. 11.2. A piston at the end of a long tube is suddenly displaced a small distance to the left, locally compressing the gas next to it. This launches a wave down the tube, which travels at the speed of sound. If the piston displacement is very sudden, then the density and pressure will change abruptly across the wavefront, from the undisturbed values ρ and P to the perturbed values $\rho + \Delta\rho$ and $P + \Delta P$ behind the wave. (If the piston were replaced by a loudspeaker, then both the diaphragm displacement and the disturbances would instead be periodic.)

Since the piston displacement is rapid, there is little time for heat transfer, and consequently the compression process may be taken as adiabatic. Therefore, the temperature rises to $T + \Delta T$ behind the wave. Since the gas behind the wave is expanding, it is in motion; let its velocity be ΔV . (Note that $\Delta V \ll c$; the wave propagates into the undisturbed gas, much like a water wave moves without transporting water with it.)

We want to determine the speed c the disturbance propagates with. To do

¹The mean free path in air is about 10^{-7} m, so this condition is easily satisfied in air.

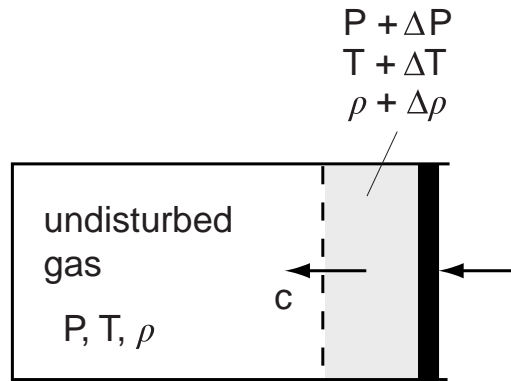


Figure 11.2: The piston is rapidly pushed in a small distance, launching a sound wave travelling to the left.

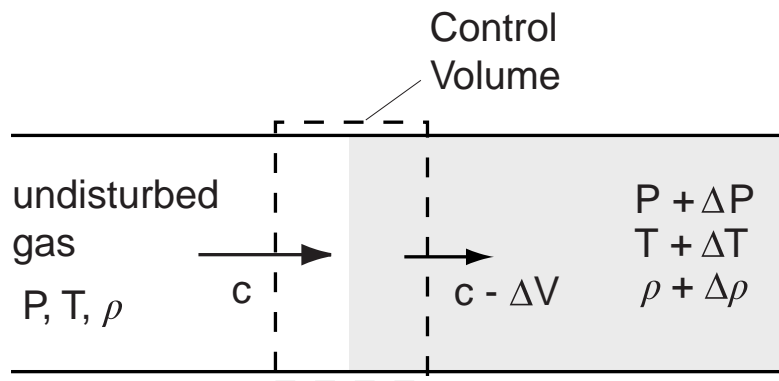


Figure 11.3: The situation as seen by an observer moving with the wave at speed c .

this, it is convenient to switch to a reference frame travelling with the wave (Fig. 11.3). Defining a control volume surrounding the wave, undisturbed fluid enters the control volume from the left at speed c , passes through the wave, and leaves the control volume at speed $c - \Delta V$.

If this wave propagates steadily, then there can be no mass buildup within the control volume. The mass flow rate in the left is

$$\dot{m}_{in} = \rho c A, \quad (11.2)$$

where A is the tube cross-sectional area. The mass flow rate out the right side is

$$\dot{m}_{out} = (\rho + \Delta\rho)(c - \Delta V)A. \quad (11.3)$$

Equating \dot{m}_{in} and \dot{m}_{out} yields

$$c\Delta\rho - \rho\Delta V - \Delta\rho\Delta V = 0. \quad (11.4)$$

The magnitudes of the pressure, density, and velocity disturbances in real sound waves are very small. In fact, the sound speed is formally defined as the speed *infinitesimal* disturbances propagate with. We will soon let all disturbance quantities become infinitesimal, which means we can ignore the very small $\Delta\rho\Delta V$ term in this equation. With this approximation, we have

$$\Delta V = c \frac{\Delta\rho}{\rho}. \quad (11.5)$$

Since $\Delta\rho/\rho \ll 1$, this justifies our claim above that $\Delta V \ll c$.

Equation (11.5) resulted from doing a mass balance on the control volume; let's now see what results from applying the momentum principle. The x momentum leaving the control volume on the right is $\dot{m}(c - \Delta V)$; the x momentum entering on the left is $\dot{m}c$. Therefore, the *net* x momentum leaving the control volume is $-\dot{m}\Delta V$.

The momentum principle says that this equals the net force in the x direction on the control volume. The net force is due to pressure:

$$F_{net} = AP - A(P + \Delta P) = -A\Delta P. \quad (11.6)$$

Therefore, the momentum principle requires that

$$A\Delta P = \dot{m}\Delta V. \quad (11.7)$$

Using $\dot{m} = \rho Ac$, this becomes

$$\Delta P = \rho c \Delta V. \quad (11.8)$$

Substituting for ΔV from Eq. (11.5), we find

$$\Delta P = c^2 \Delta \rho. \quad (11.9)$$

Finally, an energy balance on the control volume (including the kinetic energy in this reference frame) yields

$$h + \frac{c^2}{2} = (h + \Delta h) + \frac{(c - \Delta V)^2}{2}. \quad (11.10)$$

Since the propagation is assumed to be adiabatic, there is no heat transfer term Q in this equation.

To first order in ΔV , the energy balance becomes

$$\Delta h = c \Delta V. \quad (11.11)$$

Using Eq. (11.5), this can be written

$$\Delta h = c^2 \frac{\Delta \rho}{\rho}. \quad (11.12)$$

Now let us take the limit as the disturbances become infinitesimal. Equations (11.9) and (11.12) become

$$dP = c^2 d\rho, \quad (11.13)$$

$$dh = c^2 \frac{d\rho}{\rho}. \quad (11.14)$$

$$(11.15)$$

At this point, it is interesting to ask what happens to the entropy of the gas as it passes through the wave. We know

$$dh = T ds + v dP = T ds + dP/\rho, \quad (11.16)$$

so solving for ds yields

$$ds = \frac{1}{T} \left(dh - \frac{dP}{\rho} \right). \quad (11.17)$$

Substituting from above for dh and dP ,

$$ds = \frac{1}{T} \left((c^2 d\rho) - \frac{(c^2 d\rho/\rho)}{\rho} \right) = 0. \quad (11.18)$$

Therefore,

the propagation of an infinitesimal sound wave is isentropic.

(This would not be true for finite-sized disturbances, which are shock waves.)

With this result, we can use Eq. (11.13) to solve for c . Since s is constant, $dP/d\rho$ for this process is equivalent to $(\partial P/\partial\rho)_s$. Therefore,

$$c = \sqrt{\left(\frac{\partial P}{\partial\rho}\right)_s} \quad (11.19)$$

Example 11.1 Estimate the speed of sound in liquid water at 400 K and 1 MPa.

Solution: using the WWW thermodynamic property calculator or TPX, for these conditions $\rho = 937.88 \text{ kg/m}^3$. Now let's perturb ρ by a small amount, say 1.0 kg/m^3 . Holding s constant and setting $\rho = 938.88 \text{ kg/m}^3$, we find $P = 3.2676 \text{ MPa}$. Thus,

$$c \approx \left(\frac{2.2676 \times 10^6}{1.0}\right)^{1/2} = 1506 \text{ m/s}.$$

11.3.1 Speed of sound in an ideal gas

In an ideal gas with constant c_p , Pv^k is constant for any isentropic process. Therefore, since $\rho = 1/v$,

$$P = C\rho^k \quad (11.20)$$

for an isentropic process. Differentiating this expression,

$$\left(\frac{\partial P}{\partial\rho}\right)_s = kC\rho^{k-1} = \frac{kP}{\rho}. \quad (11.21)$$

Using $P = \rho RT$,

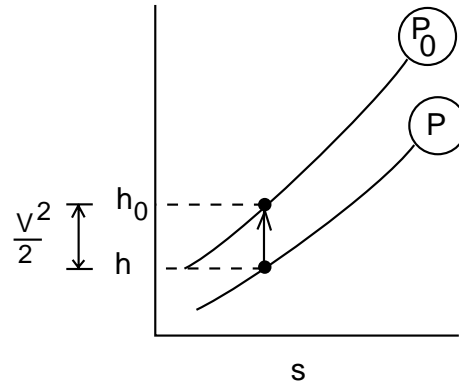
$$\left(\frac{\partial P}{\partial\rho}\right)_s = kRT \quad (11.22)$$

and therefore the speed of sound in an ideal gas is

$$c = \sqrt{kRT}. \quad (11.23)$$

Recall that R is the gas constant per unit mass: $R = \hat{R}/\hat{M}$.

Example 11.2 Calculate the speed of sound in air at 300 K.

Figure 11.4: The stagnation state on an $h - s$ plot

Solution: For air, $k = 1.4$ and $\hat{M} = 28.97$. Therefore,

$$c = \sqrt{(1.4)(8314/28.97)(300)} = 347 \text{ m/s.} \quad (11.24)$$

Note that the speed of sound in air is much less than in water.

The *Mach Number* M is defined to be the ratio of the fluid velocity to the local speed of sound:

$$M = \frac{V}{c}. \quad (11.25)$$

If $M < 1$, the flow is *subsonic*, and if $M > 1$ the flow is *supersonic*.

11.4 Stagnation Properties

When a high-speed flowing gas is decelerated to zero velocity (for example in a diffuser, or when the gas must deflect around an object inserted into it), its temperature, pressure, and density increase, since as a fluid packet decelerates the kinetic energy in the flow is expended by doing compression work on the fluid packet. The state reached by decelerating a fluid moving at speed V reversibly and adiabatically to zero velocity is called the *stagnation state*. The stagnation process is shown in Fig. 11.4.

The stagnation state is defined by the two equations

$$h_0 = h + \frac{V^2}{2} \quad (11.26)$$

and

$$s_0 = s. \quad (11.27)$$

Here the subscript 0 denotes the stagnation state, and h_0 is known as the stagnation enthalpy. The pressure attained in the stagnation state is the stagnation pressure $P_0 = P(h_0, s_0)$, and the temperature $T_0 = T(h_0, s_0)$ is the stagnation temperature. Note that given h , s , and V it is always possible to calculate h_0 , s_0 , P_0 , and T_0 , whether or not the problem at hand actually involves decelerating the fluid to a stationary state; the stagnation state is the state which would result *if* we were to do it. We may regard the stagnation properties as properties of the flowing fluid.

Example 11.3 Determine the stagnation temperature and pressure for nitrogen at 10 MPa and 300 K, with speed $V = 200$ m/s.

Solution: Using the WWW calculator, we find $h = 442.14$ kJ/kg (4.4214×10^5 J/kg) and $s = 2.988$ kJ/kg-K. Therefore, $h_0 = 4.4214 \times 10^5 + (300)^2/2 = 4.8714 \times 10^5$ J/kg. Using $h_0 = 487.14$ kJ/kg, and $s_0 = s = 2.988$ kJ/kg-K, the WWW calculator finds $T_0 = 343.44$ K, and $P_0 = 15.83$ MPa.

For the special case of an ideal gas with constant c_p , Eq. (11.26) becomes

$$c_p T_0 = c_p T + \frac{V^2}{2}. \quad (11.28)$$

We can re-write this equation in terms of k and Mach number, using $c_p = kR/(k-1)$, $c = \sqrt{kRT}$, and $M = V/c$:

$$\frac{T_0}{T} = 1 + \frac{k-1}{2} M^2 \quad (11.29)$$

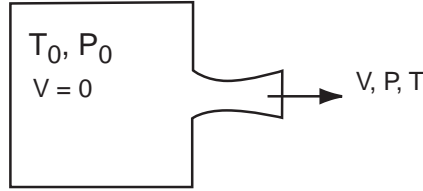
For any isentropic process in an ideal gas with constant c_p ,

$$\frac{P_2}{P_1} = \left(\frac{T_2}{T_1} \right)^{k/k-1}. \quad (11.30)$$

Taking state 1 to be the state with velocity V and state 2 to be the stagnation state,

$$\frac{P_0}{P} = \left(\frac{T_0}{T} \right)^{k/k-1} = \left(1 + \frac{k-1}{2} M^2 \right)^{k/k-1} \quad (11.31)$$

If we consider the reverse situation, in which a stationary gas at T_0 and P_0 is accelerated in a nozzle to speed V , Eq. (11.28) still applies, and sets an upper limit on the attainable speed, since T cannot be negative.



Setting $T = 0$,

$$V_{max} = \sqrt{2c_p T_0}. \quad (11.32)$$

This can be rewritten as

$$V_{max} = \sqrt{\frac{2}{k-1}} c_0, \quad (11.33)$$

where $c_0 = \sqrt{kRT_0}$ is the sound speed upstream of the nozzle. For $k = 1.4$, this becomes $V_{max} = 2.2c_0$. Therefore, the maximum velocity air can be accelerated to using an adiabatic nozzle is a little more than twice the upstream sound speed. Of course, since T is dropping as the gas accelerates, the *local* sound speed c is going down, and the *local* Mach number $M = V/c$ diverges to infinity as $V \rightarrow V_{max}$.

11.5 Isentropic Flow with Area Change

Consider one-dimensional, steady, flow down a tube with an area $A(x)$ which changes with distance. We will assume the irreversible processes of viscous friction and heat conduction are negligible, in which case the flow is isentropic.

First of all, conservation of mass requires that

$$\rho(x)V(x)A(x) = \text{constant}. \quad (11.34)$$

Differentiating this and dividing by ρVA yields

$$\frac{d\rho}{\rho} + \frac{dV}{V} + \frac{dA}{A} = 0. \quad (11.35)$$

Conservation of energy requires that

$$h(x) + \frac{V^2(x)}{2} = \text{constant}. \quad (11.36)$$

Differentiating yields

$$dh = -VdV. \quad (11.37)$$

We also know from basic thermodynamic principles that

$$dh = Tds + vdP = Tds + (1/\rho)dP. \quad (11.38)$$

Since the flow is isentropic, $ds = 0$, and therefore

$$dh = \frac{dP}{\rho} \quad (11.39)$$

Equating Eq. (11.37) and Eq. (11.39),

$$dP = -\rho V dV. \quad (11.40)$$

This relationship makes sense qualitatively, since if the pressure decreases down the tube ($dP/dx < 0$), we expect the fluid to accelerate ($dV/dx > 0$), and visa versa.

Let's now find how the density changes with distance. Since the flow is isentropic,

$$dP = \left(\frac{\partial P}{\partial \rho} \right)_s d\rho. \quad (11.41)$$

But $(\partial P/\partial \rho)_s = c^2$, so

$$dP = c^2 d\rho. \quad (11.42)$$

Putting this into Eq. (11.39), we find

$$\frac{d\rho}{\rho} = -\frac{V dV}{c^2} = -M \frac{dV}{c}. \quad (11.43)$$

Therefore, the density decreases if the flow accelerates and visa versa ($d\rho < 0 \iff dV > 0$). Note that the relative change in density for a given dV is greater at high Mach number than at low.

We may eliminate density from Eq. (11.35) by substituting from Eq. (11.43):

$$-\frac{V dV}{c^2} + \frac{dV}{V} + \frac{dA}{A} = 0 \quad (11.44)$$

or

$$\frac{dA}{A} = -\frac{dV}{V} \left[1 - \left(\frac{V}{c} \right)^2 \right]. \quad (11.45)$$

In terms of the Mach number,

$$\boxed{\frac{dA}{A} = -\frac{dV}{V} (1 - M^2)} \quad (11.46)$$

Note that we have not made any ideal gas assumption, so this equation applies for one-dimensional, isentropic flow of any fluid.

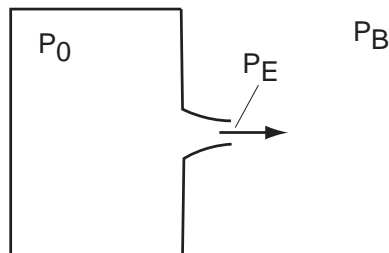


Figure 11.5: Flow in a converging nozzle.

Equation (11.46) tells us how the velocity of the fluid changes as it passes down the tube. An interesting result of this equation is that whether the fluid accelerates ($dV > 0$) or decelerates ($dV < 0$) as the tube area changes depends on whether the Mach number is greater than or less than 1.0 (supersonic or subsonic). We may identify 3 cases:

Subsonic Flow: $M < 1$. In this case, a converging tube ($dA < 0$) causes the gas to accelerate ($dV > 0$), and a diverging tube causes the gas to decelerate. Therefore, a subsonic nozzle is a converging tube, and a subsonic diffuser is a diverging tube.

Supersonic Flow: $M > 1$. Now the sign of the right-hand side of Eq. (11.46) is positive, so a converging tube causes the gas to *decelerate*, and a diverging tube causes the gas to accelerate. Therefore, a nozzle to accelerate a supersonic flow must be constructed as a diverging tube, and a supersonic diffuser must be converging.

Sonic Flow: $M = 1$. If the Mach number equals one, then the right-hand side becomes zero for any dV . Therefore, sonic flow *requires* $dA = 0$: **it is not possible to have $M = 1$ in a portion of the tube where the area is changing with x .**

11.6 Flow in a Converging Nozzle

Suppose gas is escaping from a gas tank at pressure P_0 through a small converging nozzle, as shown in Fig. 11.5. The pressure of the surroundings is $P_B < P_0$ (the “back pressure”). The pressure at the exit plane of the nozzle is P_E .

Since the nozzle is converging, $dA/dx < 0$ and therefore the accelerating flow in the nozzle must be subsonic – it is impossible for supersonic flow to exist in this nozzle. At the end, the flow area expands abruptly. We will assume

that $dA/dx = 0$ at some point very close to the nozzle exit, since the lip will inevitably be at least a little rounded.

Let us consider the flow in the nozzle as a function of P_B for given P_0 . If $P_B = P_0$, then the system is in mechanical equilibrium, and there is no flow, so $\dot{m} = 0$. For P_B slightly less than P_0 , gas will flow, and the pressure through the nozzle will fall continuously from P_0 to $P_E = P_B$. As P_B is lowered further, \dot{m} increases, and still $P_E = P_B$.

But at some P_B the flow becomes sonic at the exit; call this value P^* . What happens if P_B is now reduced below P^* ? The way the flow “communicates” the change in P_B to the flow upstream is through acoustic waves, which travel at the sound speed. As long as the flow is everywhere subsonic, it is possible for sound waves to propagate upstream, and the flow everywhere in the nozzle adjusts to the change in P_B .

But at $P_B = P^*$, the exit flow is coming out at the local speed of sound. Sound waves propagate upstream with this speed *relative to the moving gas*. In the lab frame, the sound waves get nowhere – it is analogous to a fish trying to swim upstream in a river when the river is flowing downstream as fast as the fish can swim upstream.

Therefore, the information that P_B is now below P^* can’t be conveyed to the flow in the nozzle; it is unaware of the change. Therefore, everything about the flow in the nozzle – the total mass flow rate, $P(x)$, $V(x)$, etc. – is identical to the results for $P_B = P^*$, even though now $P_B < P^*$.

In particular, $P_E = P^*$, not P_B . Once the gas emerges from the nozzle, it finds itself suddenly at higher pressure than the surroundings. It rapidly adjusts to P_B through a set of complex, three-dimensional expansion waves. The pressure distribution would look something like that shown in Fig. 11.6.

11.7 Choked Flow

Once the flow at the exit of a converging nozzle becomes sonic, the downstream conditions no longer affect the flow upstream of the shock wave. Therefore, the mass flow rate through the nozzle is *unaffected* by downstream pressure once $M = 1$ at the nozzle exit. When this occurs, we say the flow is *choked*.

We can calculate the pressure ration P_0/P_B above which the flow becomes choked, and the mass flow rate under choked conditions for the case of an ideal gas with constant c_p . Let the upstream conditions be T_0 and P_0 , and denote the conditions where $M = 1$ at the exit by T^* and P^* . Then from Eqs. 11.29

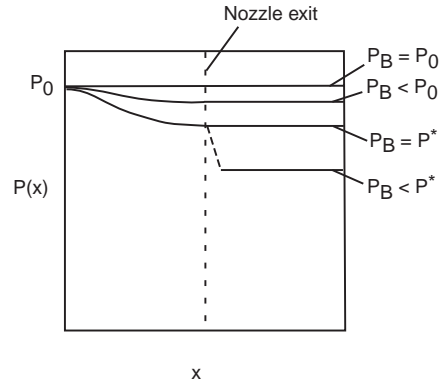


Figure 11.6: Pressure vs. distance for flow through a converging nozzle.

and 11.31,

$$\frac{T_0}{T^*} = 1 + \frac{k-1}{2} \quad (11.47)$$

and

$$\frac{P_0}{P^*} = \left(1 + \frac{k-1}{2}\right)^{k/k-1}. \quad (11.48)$$

For $k = 1.4$, these expressions become $T_0/T^* = 1.2$ and $P_0/P^* = 1.89$. Thus, a pressure ratio from inlet to outlet of the tube of only 1.9 is sufficient to choke the flow for air.

To solve for the choked mass flow rate, we may use the result that \dot{m} is equal to $\rho V A$ at any cross-section of the nozzle. For simplicity let's use the throat:

$$\dot{m} = \rho^* A^* c. \quad (11.49)$$

Here A^* is the area of the throat, and we have used the fact that the flow is sonic at the throat. Using the expression for c for an ideal gas, we have

$$\dot{m} = A^* \frac{P^*}{RT^*} \sqrt{kRT^*} \quad (11.50)$$

$$= \left[\frac{P^*}{P_0} \sqrt{\frac{k}{R} \frac{T_0}{T^*}} \right] \frac{P_0 A^*}{\sqrt{T_0}} \quad (11.51)$$

The terms in brackets can be evaluated for a given k and \hat{M} (needed to calculate R). For air, this expression reduces to

$$\dot{m} = 0.0404 \frac{A^* P_0}{\sqrt{T_0}} \quad (11.52)$$

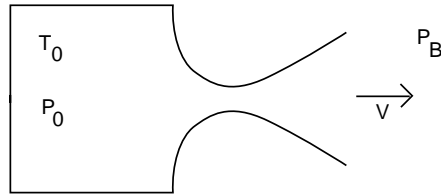


Figure 11.7: A converging-diverging nozzle.

in SI units.

Therefore, for a given gas, the maximum (choked) mass flow rate through a nozzle depends only on three quantities: the area of the nozzle where the flow is sonic, and the stagnation temperature and pressure upstream.

Example 11.4 Air enters a converging nozzle with an entrance diameter of 15 cm and an exit diameter of 2 cm. The entrance conditions are 6 MPa and 300 K. Treating air as an ideal gas with constant c_p , determine the maximum mass flow rate possible through the nozzle.

Solution: The maximum flow rate is the choked flow rate. From Eq. (11.52),

$$\dot{m}_{max} = 0.0404 \frac{(\pi)(0.01 \text{ m})^2(6 \times 10^6 \text{ Pa})}{\sqrt{300 \text{ K}}} = 4.39 \text{ kg/s.}$$

11.8 Flow in a Converging-Diverging Nozzle

Consider now attaching a diverging section onto the converging nozzle of the previous section, producing the situation shown in Fig. 11.7.

As before, upstream of the nozzle the flow velocity is negligible, and the pressure and temperature are the stagnation values P_0 and T_0 , respectively. Outside the nozzle, the pressure is P_B .

As P_B is lowered below P_0 , gas begins to flow through the nozzle. In the converging section, the flow accelerates and the pressure and temperature drop. If the flow is still subsonic when it reaches the minimum-area position in the nozzle (the “throat”), then as it passes into the diverging portion it decelerates, with a corresponding rise in P and T . It emerges with $P_e = P_B$, and $T_e = T_0 - V_e^2/2c_p$. Note that this behavior is completely consistent with Eq. (11.46). Since $M \neq 1$ at the throat where $dA/dx = 0$, it must be that $dV/dx = 0$ here, as it is.

If P_B is reduced further to P^* , the flow at the throat will become sonic. From Eq. (11.46), in this case it is no longer necessary for $dV/dx = 0$ at the

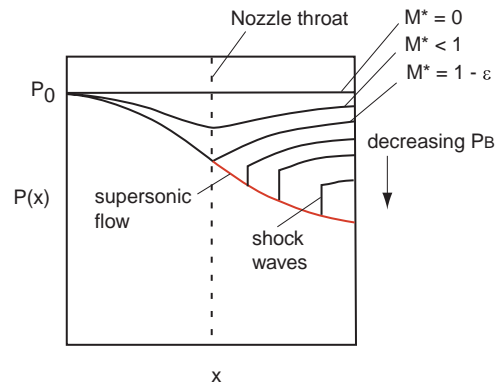


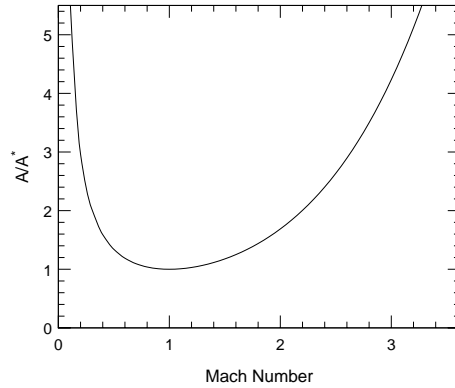
Figure 11.8: Flow in a converging-diverging nozzle. If P_B is low enough that the flow at the throat is sonic ($M^* = 1$), then the flow continues to expand supersonically in the diverging portion (red line). In order to match P_B at the nozzle exit, at some point the flow suddenly transitions from supersonic to subsonic in a shock wave, with a sudden rise in pressure. After the shock, the diverging nozzle acts as a diffuser, and the pressure rises to P_B . The location of the shock depends on P_B , beginning at the throat when P_B is just low enough to produce supersonic flow, and moving out toward the nozzle exit as P_B decreases.

throat, so the flow can continue to accelerate into the diverging portion, where (now that it is supersonic) it continues to be accelerated.

As the supersonic flow accelerates in the diverging portion, the pressure continues to drop. The value of P_B has no effect on the flow in this portion of the nozzle, since any information about P_B would propagate at the speed of sound, and so would be swept downstream and would not reach the upstream flow. Therefore, if the gas continues to accelerate isentropically to the exit of the nozzle the pressure of the gas at the nozzle exit would not generally match P_B .

But somehow the pressure must come to P_B at or after the nozzle exit. What really happens is that if the pressure in the nozzle drops below P_B , then at some point in the diverging portion the gas undergoes a sudden and irreversible transition to subsonic flow. This transition is known as a *shock wave*, and will be discussed in the next section. Typically a shock wave normal to the flow will stand somewhere in the diverging portion of the nozzle, at a position such that $P_e = P_B$. The pressure vs. distance plot for various P_B values now looks something like the plot shown in Fig. 11.8.

In the isentropic portion of the flow (before the shock wave), we can use the condition $\rho VA = \text{constant}$ to solve for the velocity at any position in the nozzle.

Figure 11.9: Area ratio A/A^* for $k = 1.4$.

We may write

$$\rho V A = \rho^* c A^* \quad (11.53)$$

which using the ideal gas equation of state and $c = \sqrt{kRT}$ becomes²

$$\frac{A}{A^*} = \frac{1}{M} \frac{P^*}{P} \sqrt{\frac{T}{T^*}}. \quad (11.54)$$

If we now write this as

$$\frac{A}{A^*} = \frac{1}{M} \frac{P^*/P_0}{P/P_0} \sqrt{\frac{T/T_0}{T^*/T_0}} \quad (11.55)$$

and use eqs. (11.29) and (11.29), we find

$$\frac{A}{A^*} = \frac{1}{M} \left[\left(\frac{2}{k+1} \right) \left(1 + \frac{k-1}{2} M^2 \right) \right]^{(k+1)/2(k-1)}. \quad (11.56)$$

This equation is plotted in Fig. 11.9 for $k = 1.4$. Note that for a given area ratio, there are two possible Mach numbers — one subsonic, and the other supersonic.

11.9 Normal Shock Waves

Shock waves which stand perpendicular to the flow direction are called *normal shock waves*, in contrast to *oblique* shock waves which can occur when two- and

²You should verify this.

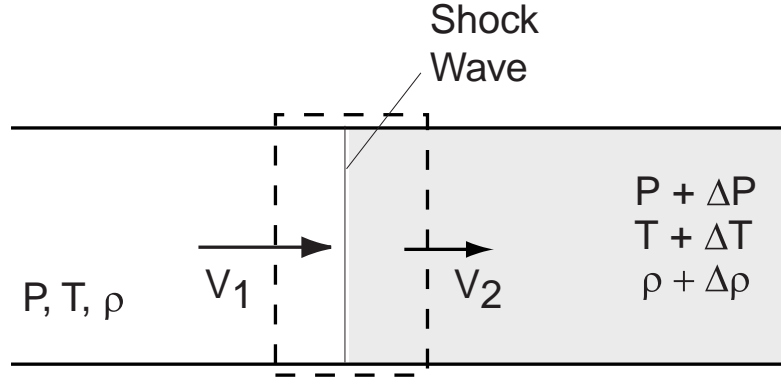


Figure 11.10: A steady shock wave. Note that ΔP , $\Delta \rho$, and ΔT are *not* assumed to be infinitesimal.

three-dimensional flow is considered. A shock wave is a discontinuity in the pressure, velocity, density, and temperature of the gas. It may seem surprising to you that these properties can be discontinuous, but in fact the equations describing conservation of mass, momentum, and energy, along with the ideal gas law, do admit discontinuous solutions.³

By carrying out balances of mass, momentum, and energy on a shock wave, we may derive relationships between the gas properties on either side of the shock. The approach is very much like that we did to derive the speed of sound, except that now the perturbations are not infinitesimal.

Consider a situation in which gas with velocity V_1 in the x direction passes through a stationary normal shock and leaves with velocity V_2 . Conservation of mass requires

$$\rho_1 V_1 = \rho_2 V_2. \quad (11.57)$$

We have divided through by the tube area, since it is the same on each side of the shock. The momentum principle states

$$(P_1 - P_2) = (\dot{m}/A)(V_2 - V_1), \quad (11.58)$$

which can also be written

$$P_1 + \rho_1 V_1^2 = P_2 + \rho_2 V_2^2, \quad (11.59)$$

since $\dot{m}/A = \rho_1 V_1 = \rho_2 V_2$. The energy balance is

$$h_1 + \frac{V_1^2}{2} = h_2 + \frac{V_2^2}{2}, \quad (11.60)$$

³A more detailed treatment considering viscosity and heat conduction would show that a shock wave has a finite thickness, which is of order the mean free path.

which is equivalent to

$$h_{0,1} = h_{0,2}. \quad (11.61)$$

Therefore, the stagnation enthalpy is constant across a shock.

Also, we can write equations of state that have to be satisfied in both states. If the fluid may be treated as an ideal gas in both states, then

$$P_1 = \rho_1 R T_1, \quad (11.62)$$

$$P_2 = \rho_2 R T_2 \quad (11.63)$$

If state 1 is fully specified, the four equations (11.57), (11.59), (11.60), and (11.63) may be used to solve for the four unknowns P_2 , T_2 , ρ_2 , and V_2 .

For the case of an ideal gas with constant c_p , this set of equations may be solved analytically. Equation (11.60) becomes

$$T_{0,1} = T_{0,2}, \quad (11.64)$$

which, using Eq. (11.29), may be written

$$\frac{T_2}{T_1} = \frac{1 + \frac{k-1}{2} M_1^2}{1 + \frac{k-1}{2} M_2^2} \quad (11.65)$$

Using $\rho = P/RT$, $c = \sqrt{kRT}$, and $M = V/c$, Eq. (11.59) reduces to

$$\frac{P_2}{P_1} = \frac{1 + k M_1^2}{1 + k M_2^2} \quad (11.66)$$

Equation (11.57) becomes

$$\frac{P_2}{P_1} = \sqrt{\frac{T_2}{T_1}} \frac{M_2}{M_1}. \quad (11.67)$$

Substituting eqs. (11.65) and (11.66) into Eq. (11.67), we can solve for M_2 :

$$M_2^2 = \frac{M_1^2 + \frac{2}{k-1}}{\frac{2k}{k-1} M_1^2 - 1} \quad (11.68)$$

Since there is no heat transfer in the shock, the entropy production rate as the gas flows through the shock is simply

$$\dot{\mathcal{P}}_s = \dot{m}(s_2 - s_1) = \dot{m}(c_p \ln(T_2/T_1) - R \ln(P_2/P_1)). \quad (11.69)$$

It may be shown that this is positive only if $M_1 > 1$. Therefore, the flow entering the shock must be supersonic. Or, in the frame of reference of the gas upstream of the shock, this is equivalent to saying the shock wave must propagate at a supersonic (not subsonic) speed.

CHAPTER 12

THERMODYNAMICS OF MAGNETIC SYSTEMS

12.1 Introduction

Thermodynamics applies to much more than just simple compressible substances. We've dealt briefly with magnetic work. In this chapter, we'll discuss in more depth the thermodynamics of materials on which magnetic work can be done.

12.2 The Simple Magnetic Substance

Magnetic fields can interact with some materials, causing work to be done on or by them. Here we consider the simplest case, when magnetic work is the only possible reversible work mode.

If a substance contains molecules or ions which have permanent magnetic dipole moments (for example, O_2 or rare-earth ions), then an external magnetic field can do work on the sample by turning these dipoles to align with the field. Such materials are known as *paramagnetic*.

A magnetic field can also do work on a substance by inducing electrical currents to flow. These currents may be macroscopic ones, like those induced in the surface layer of a perfect conductor when a magnetic field is applied, or atomic-scale ones due to slight changes in the motion of the electrons orbiting about the nuclei due to the magnetic field. In either case, materials in which currents are induced when a magnetic field is applied are called *diamagnetic*.

The net magnetic dipole moment of a sample per unit volume defines the *magnetization* \vec{M} . Note that \vec{M} is a vector. For the paramagnetic case, it is the vector sum of all individual magnetic dipole moments. If they point in random directions (as they do with no applied field), then their vector sum is zero and $\vec{M} = 0$. But if a field is applied so they are at least partially aligned, then there is a net magnetization pointing in the direction of the applied field. (They are not perfectly aligned since collisions with other molecules or with the lattice in a solid tend to upset the alignment.)

For diamagnetic materials, there is also a magnetization, since current loops have a magnetic moment associated with them. It may be shown that the

magnetization of a diamagnetic material points *opposite* to the direction of the magnetic field.

All materials have some diamagnetic response, but it is usually weak. Except in extremely strong fields, most diamagnetic materials can be considered nonmagnetic. If the material contains permanent magnetic dipoles, then the paramagnetic response is usually much larger and the diamagnetism can be neglected.

Finally, the most familiar form of magnetism is *ferromagnetism*. Permanent magnets, which by definition are magnetized even in zero applied field, are made of ferromagnetic materials. Only the elements Fe, Co, Ni, and Gd exhibit ferromagnetism, and all permanent magnets contain at least one of these elements. The ferromagnetic elements have permanent magnetic dipoles, like paramagnetic materials. The difference is that the dipoles exert forces on one another which tend to lock them into alignment even without an external magnetic field. If a ferromagnetic material is heated, beyond a certain temperature the forces between dipoles are overcome, and the material becomes paramagnetic. This temperature is known as the *Curie Point*, and for iron it is 770 °C.

In general, a paramagnetic, diamagnetic, or ferromagnetic substance may have other work modes too. For example, oxygen gas is paramagnetic but also is compressible. A liquid oxygen droplet will have a surface tension, so deforming the droplet is another work mode. But in most cases, magnetic materials are solid, and reasonably incompressible. In this case, the only important work mode is magnetic work, and we may then treat the substance as a *simple magnetic substance*. This is the only case we will consider here.

When an external magnetic field is applied to a sample, the response of the sample (turning dipoles or induced currents) modifies the local value of the field within the sample. We define \vec{H} to represent the applied field which would exist without the sample, and \vec{B} to represent the true field value in the sample. The quantities \vec{B} , \vec{H} , and \vec{M} are related by ¹

$$\vec{B} = \mu_0(\vec{H} + \vec{M}). \quad (12.1)$$

The value of the field without the sample is really $\mu_0\vec{H}$, but since μ_0 is just a constant ($4\pi \times 10^{-7}$ tesla-m/amp) we can treat \vec{H} as representing the applied magnetic field.

¹This equation depends on the unit system. The form here is for the Rationalized MKS system, which is the one used most commonly in applied science and engineering. In another unit system widely used in physics (cgs), \vec{B} and \vec{H} have the same units and equation 12.1 is replaced by $\vec{B} = \vec{H} + 4\pi\vec{M}$

We've shown previously that the work required to reversibly magnetize a magnetic substance is

$$dW = \mu_0 \vec{H} \cdot d(\vec{M}V). \quad (12.2)$$

The Gibbs equation for a simple magnetic substance may be derived by considering a reversible process in which magnetic work dW is done and heat $dQ = TdS$ is added. Since $dU = dQ + dW$,

$$dU = TdS + \mu_0 \vec{H} \cdot d(\vec{M}V). \quad (12.3)$$

On a unit mass basis,

$$du = Tds + \mu_0 \vec{H} \cdot d(\vec{M}v). \quad (12.4)$$

We can simplify the notation by defining

$$\vec{m} = \mu_0 v \vec{M}. \quad (12.5)$$

Also, the magnetization is generally either parallel to \vec{H} (paramagnetic substances) or anti-parallel (diamagnetic substances). Therefore, $\vec{H} \cdot d\vec{m} = Hdm$, where $m < 0$ in the diamagnetic case.

The magnetic Gibbs equation then becomes

$$du = Tds + Hdm. \quad (12.6)$$

Starting from this equation, we can carry out an analysis very similar to the one in chapter 6 for a simple compressible substance. The only difference is that we replace $-v$ by m , and P by H . For example, we can define a magnetic enthalpy, Helmholtz free energy, and Gibbs free energy by

$$h = u - Hm \quad (12.7)$$

$$f = u - Ts \quad (12.8)$$

$$g = h - Ts = u - Ts - Hm. \quad (12.9)$$

Although we use the same symbols for these as we did for a simple compressible substance, it's important to keep in mind the difference in definition (e.g. $h = u - Hm$, not $h = u + Pv$) since the reversible work mode is now magnetization, not compression.

The differential equations of state for these properties are:

$$dh = Tds - mdH \quad (12.10)$$

$$df = -sdT + Hdm \quad (12.11)$$

$$dg = -sdT - mdH \quad (12.12)$$

Integrating these equations would yield $u(s, m)$, $h(s, H)$, $f(T, m)$, and $g(T, H)$, all of which are fundamental relations. Experimentally, it is usually easiest to work with temperature and magnetic field strength as independent variables, so $g(T, H)$ is in many cases the most useful of these fundamental relations.

Of course, analogs of the Maxwell relations may be easily written down.

The above differential equations of state are general and apply to any simple magnetic substance. To determine the specific properties of a particular substance, we need to know the functional form of one of the fundamental relations [say, $g(T, H)$] or else a set of experimental equations of state like $m(T, H)$ and $c_H(T, H)$. Let's look at a couple of important magnetic substances.

12.2.1 The Curie Substance

For paramagnetic materials, the tendency for the dipoles to be aligned with the field is balanced by the effects of random collisions (in a gas) or vibrations of the lattice (in a solid), which tend to randomize the direction the dipole points. Since the energy associated with the random, thermal motion scales with temperature, we expect that the magnetization of a paramagnetic material will increase with H at constant T , and decrease with T at constant H . In fact it is found that $M = f(H/T)$. The function $f(H/T)$ is linear for small H/T , and saturates at an asymptotic value for large H/T , since in this limit the dipoles are perfectly aligned and therefore M can't increase further.

A *Curie substance* is defined as any substance which obeys

$$\vec{M} = C \frac{\vec{H}}{T}. \quad (12.13)$$

All paramagnetic substances have this behavior for small H/T . In terms of m , or

$$m = C' \frac{\vec{H}}{T}, \quad (12.14)$$

where $C' = \mu_0 v C$.

Like an ideal gas, a Curie substance has an internal energy which depends only on temperature. The proof follows exactly the procedure we used to prove $u = u(T)$ for an ideal gas. First solve Eq. (12.6) for ds :

$$ds = \frac{1}{T} du - \frac{H}{T} dm. \quad (12.15)$$

Substituting for H/M ,

$$ds = \frac{1}{T} du - \frac{1}{C'} m dm. \quad (12.16)$$

Equating mixed partial derivatives,

$$\frac{\partial}{\partial m} \left[\frac{1}{T} \right]_u = -\frac{1}{C'} \left(\frac{\partial m}{\partial u} \right)_m \quad (12.17)$$

But $(\partial m / \partial u)_m = 0$, so $1/T$ depends only on u , not on m . Therefore, $T = T(u)$, and inverting this we have $u = u(T)$.

In fact, it can be seen from this analysis that any material which obeys

$$m = f(H/T) \quad (12.18)$$

will have an internal energy which is a function of temperature alone.

Knowing that $u = u(T)$, we may now integrate the Gibbs equation and find the entropy of a Curie substance.

$$ds = \left(\frac{1}{T} \right) \left(\frac{du}{dT} \right) dT - \frac{H}{T} dm. \quad (12.19)$$

The coefficient multiplying dT is $(\partial s / \partial T)_m$, which by definition is c_m/T , where c_m is the specific heat at constant magnetization. Therefore, we see

$$c_m(T) = \frac{du(T)}{dT} \quad (12.20)$$

which is a function only of temperature for a Curie substance. The Gibbs equation then is

$$ds = \frac{c_m(T)}{T} dT - \frac{H}{T} dm. \quad (12.21)$$

Since $m = C'H/T$,

$$ds = \frac{c_m(T)}{T} dT - \frac{m}{C'} dm. \quad (12.22)$$

Integrating this,

$$s(T_1, m_1) - s(T_0, m_0) = \int_{T_0}^{T_1} \frac{c_m(T)}{T} dT - \frac{1}{2C'} (m_1^2 - m_0^2), \quad (12.23)$$

or in terms of M ,

$$s(T_1, M_1) - s(T_0, M_0) = \int_{T_0}^{T_1} \frac{c_m(T)}{T} dT - \frac{\mu_0 v}{2C} (M_1^2 - M_0^2), \quad (12.24)$$

Note that the entropy decreases if a Curie substance is magnetized isothermally. This makes sense physically: as individual atomic magnetic dipoles become more aligned with the field (increasing M), the sample becomes more ordered. Therefore, we would expect s to decrease, as it does.

This also means that heat must be removed from the sample during isothermal magnetization. If a sample is magnetized reversibly at T from zero magnetization to some value M ,

$$Q = \frac{\mu_0 v M^2}{2C} \quad (12.25)$$

must be rejected to the environment. If instead the sample is magnetized adiabatically, then it will increase in temperature. Similarly, if a Curie substance is demagnetized (M decreased) at constant temperature, heat must be supplied. If demagnetization is carried out adiabatically, the temperature will decrease.

Thus, magnetizing a Curie substance is similar to compressing a gas, and demagnetizing a Curie substance is similar to expanding a gas. This analogy suggests it should be possible to construct a heat pump or a heat engine by taking a Curie substance through a sequence of states in the (T, m) plane (or equivalently, the (T, H) plane) returning to the initial state. While no known practical heat engine operates on this cycle, it is in fact used as a heat pump (refrigerator) to reach very low temperatures near absolute zero. Temperatures as low as 10^{-8} K have been reached using magnetic refrigerators. To reach these very low temperatures, nuclear magnetic moments (rather than atomic ones) are employed.

12.3 Superconductors

Many metals exhibit superconductivity when cooled to very low temperatures. This effect was first discovered by Onnes in 1911, shortly after he succeeded in liquifying helium. Until 1986, all known superconductors required cooling below 20 K before they would become superconductors. Beginning in 1986, some ceramic materials which become superconductors at abnormally high temperatures (above 100 K) were discovered. One of the most useful ones is yttrium barium copper oxide: $\text{YBa}_2\text{Cu}_3\text{O}_{7-x}$. The discovery of high temperature superconductors has enormous practical consequences, since their transition temperatures are above the boiling point of N_2 at 1 atm. Therefore, inexpensive liquid nitrogen can be used to keep these materials superconducting, rather than expensive liquid helium which is required for most lower-temperature superconductors.

Superconductors exhibit several strange properties. As they are cooled below

the transition temperature T_c , the electrical resistance drops suddenly to exactly zero. If a current is set up in a superconducting ring (for example by cooling it below T_c in a magnetic field and then removing the field), it persists indefinitely. Measurements have been done which show currents remaining unchanged for months. This can only be possible if the resistance is really zero, not just very small.

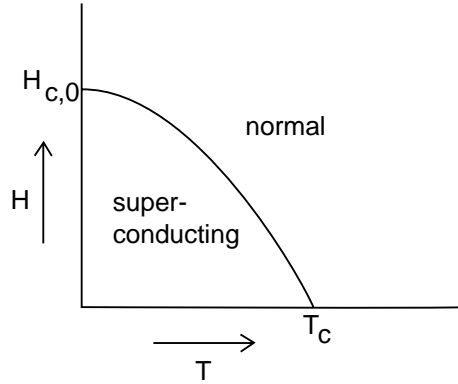
If a superconductor is brought into a magnetic field, currents will flow in the surface layer of the superconductor, which perfectly shield the the magnetic field from the interior, so that the magnetic field inside is zero. This is the behavior predicted from EM theory for any perfect conductor. In contrast, for a real conductor with finite resistance, the induced currents would eventually die out and the field would penetrate.

But superconductors display an additional property which is not simply a consequence of zero resistance. If a superconducting material is cooled from above T_c to below T_c in a magnetic field, *the magnetic field is completely expelled from the interior at T_c* . Since the magnetic field penetrates above T_c , there would be no reason for currents to start flowing in the surface to shield the interior from the magnetic field, if a superconductor were just like a metal with zero resistance.

But the currents *do* start flowing in the surface layer at T_c , and the field *is* expelled. Evidently, there is something intrinsic about the superconducting state which is incompatible with the presence of a magnetic field inside. This effect, first discovered in 1933 by Meissner and Ochsenfeld, is called the *Meissner effect*. This is what causes a permanent magnet to levitate above a superconducting dish, and is employed in several schemes for magnetic levitation trains.

Thermodynamics can be used to understand some of the properties of superconductors. When a superconductor expels a magnetic field it does magnetic work on the external currents producing the field, so magnetic work is a relevant work mode for a superconductor. In principle, we could include $-Pdv$ work too, but it is usually acceptable to assume the superconductor is incompressible. In this case, a superconductor is a simple magnetic substance.

It is found that applying a sufficiently strong magnetic field to a superconductor can destroy the superconductivity, causing it to revert to the “normal” state. We may regard the superconducting state exactly like a thermodynamic phase. The phase diagram for a superconductor looks something like shown below.



What is the magnetic equation of state of a superconductor? Since we know $\vec{B} = 0$ in a superconductor, and also

$$\vec{B} = \mu_0(\vec{H} + \vec{M}) \tag{12.26}$$

the magnetic equation of state in the superconducting state is

$$\vec{M} = -\vec{H}. \tag{12.27}$$

In contrast, in the normal state, $\vec{M} = 0$ since typically metals which become superconducting do not contain paramagnetic ions.

Since the Gibbs free energy is particularly useful if T and H are held constant, let's calculate g for the normal and superconducting phases. We can do this by integrating the differential equation of state

$$dg = -sdT - mdH, \tag{12.28}$$

using the appropriate magnetic equation of state for the normal and superconducting phases ($m = 0$ or $m = -\mu_0 vH$).

For the normal phase, $m = 0$, and therefore $Hdm = 0$ so it is impossible to do magnetic work on the substance when it is in the normal phase. As in the case of an incompressible fluid, this means that the properties of the substance in the normal state are fixed by only one quantity, the temperature. For the free energy we have

$$dg_n = -s_n(T)dT. \tag{12.29}$$

Therefore,

$$g_n(T_1) - g_n(T_0) = - \int_{T_0}^{T_1} s_n(T) dT. \tag{12.30}$$

For the superconducting phase,

$$dg_s = -s_s dT - m_s dH = -s_s dT + (\mu_0 v H) dH. \quad (12.31)$$

If we equate the mixed partial derivatives of the coefficients in this equation,

$$-\left(\frac{\partial s_s}{\partial H}\right)_T = \mu_0 v \left(\frac{\partial H}{\partial T}\right)_H = 0, \quad (12.32)$$

we see that the entropy in the superconducting phase has no dependence on H , and therefore depends only on T . Thus, s is a function just of T for both the superconducting and normal phases. (Note that this was not true for the Curie substance.)

With this information, we can proceed to integrate Eq. (12.31) to find the Gibbs free energy of a superconductor.

$$g_s(T_1, H_1) - g_s(T_0, H_0) = -\int_{T_0}^{T_1} s_s(T) dT + \frac{\mu_0 v}{2} (H_1^2 - H_0^2). \quad (12.33)$$

We showed in Chapter six that the stable phase of a simple compressible substance at specified (T, P) is the one with the lowest Gibbs free energy $g = u + Pv - Ts$. The same analysis can be repeated for the magnetic case, simply replacing P with H and v with $-m$ to show that the stable phase of a simple magnetic substance at specified (T, H) is the one with lowest magnetic Gibbs free energy $g = u - Hm - Ts$.

Therefore, in the region of the (H, T) plane where the superconducting phase is thermodynamically stable, $g_s < g_n$; outside this region $g_n < g_s$. On the coexistence line where they can exist in equilibrium with one another $g_s = g_n$.

Let's consider a temperature $T < T_c$, where the superconducting phase is stable for $H = 0$, and so $g_s(T, 0) < g_n(T)$. Consider what happens as the magnetic field is increased from zero. The Gibbs free energy of the normal phase (metastable at $H = 0$) is not affected by H . But the Gibbs free energy of the superconducting phase increases quadratically with H . As shown in Figure 12.1, $g_s(T, H)$ becomes equal to $g_n(T)$ at some value of H ; this defines the value of the critical magnetic field strength $H_c(T)$, above which the superconductor reverts to the normal phase:

$$g_s(T, H_c) = g_n(T)$$

$$g_s(T, 0) + \frac{\mu_0 v H_c^2}{2} = g_n(T). \quad (12.34)$$

$$(12.35)$$

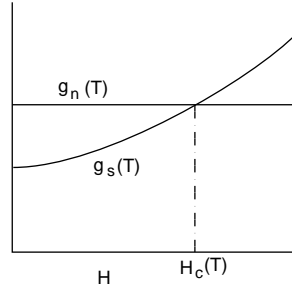


Figure 12.1: Gibbs free energy for the superconducting and normal phases.

Therefore,

$$H_c^2(T) = \frac{2}{\mu_0 v} (g_n(T) - g_s(T, 0)). \quad (12.36)$$

Note that the work done *by* the superconductor on the surroundings to keep the magnetic field out as the external field is raised from zero to H is

$$W = - \int_0^H H dm = +\mu_0 v \int_0^H H dH = \frac{\mu_0 v}{2} H^2. \quad (12.37)$$

So the increase in Gibbs free energy of the superconductor is due to the work it has to do to keep the field out; eventually, (at H_c) it is no longer worth it energetically and the substance reverts to the normal phase and lets the field in.

Now consider the latent heat associated with the normal to superconducting transition. Doing entropy accounting for a substance which undergoes a phase change from the normal to the superconducting state,

$$s_s(T) = s_n(T) + \frac{Q}{T}, \quad (12.38)$$

where Q is the heat added to bring about the normal \rightarrow superconducting transition.

Since $s = -(\partial g / \partial T)_H$,

$$\begin{aligned} Q &= T (s_s(T) - s_n(T)) \\ &= -T \left(\frac{\partial (g_s(T, H) - g_n(T))}{\partial T} \right) \\ &= -T \left(\frac{\partial (g_s(T, 0) + \mu_0 v H^2 - g_n(T))}{\partial T} \right) \\ &= -T \left(\frac{d(g_s(T, 0) - g_n(T))}{dT} \right) \end{aligned}$$

$$= T \left(\frac{d(\mu_0 v H_c^2(T)/2)}{dT} \right) \quad (12.39)$$

$$(12.40)$$

Therefore,

$$Q = \mu_0 v T H_c(T) \frac{dH_c(T)}{dT} \quad (12.41)$$

From the phase diagram, we see that $H_c = 0$ at T_c , and the transition occurs at $T < T_c$ if a field is applied. Therefore, from Eq. (12.41), at zero applied field the normal \rightarrow superconducting transition has a latent heat of zero. Also, as $T \rightarrow 0$, $Q \rightarrow 0$, so $|Q|$ must have a maximum at a temperature between 0 and T_c . Since $dH_c(T)/dT < 0$, $Q < 0$ for non-zero H . Thus, when a substance is cooled and becomes superconducting, heat is given off, similar to when a gas condenses to form a liquid.

Finally, we can derive an expression for the difference between the specific heats of the normal and superconducting phases. Since for both phases the entropy is a function only of temperature, the specific heat is simply $c = T(ds/dT)$, and is the same whether H or m is held constant.

Since $s_s(T) - s_n(T) = Q/T$, we have from Eq. (12.41)

$$s_s(T) - s_n(T) = \mu_0 v H_c(T) \frac{dH_c(T)}{dT} \quad (12.42)$$

Differentiating with respect to T and multiplying by T ,

$$c_s(T) - c_n(T) = \mu_0 v T \left[\left(\frac{dH_c(T)}{dT} \right)^2 + H_c(T) \frac{d^2 H_c}{dT^2} \right]. \quad (12.43)$$

At zero field, the transition occurs at $T = T_c$. Since $H_c = 0$ in this case,

$$c_s(T_c) - c_n(T_c) = \mu_0 v T_c \left(\frac{dH_c(T)}{dT} \right)_{T=T_c}^2. \quad (12.44)$$

This important formula, known as *Rutger's formula*, shows that there will be a jump in specific heat as the substance is cooled below T_c and becomes superconducting. The magnitude of the jump depends only on v , T_c , and the slope of the coexistence curve at $T = T_c$. The specific heat of tin is shown in Figure 12.2, which illustrates the jump in specific heat when the tin becomes superconducting.

This relationship provides a good way to determine how much of a sample is superconducting. In many cases, a sample may have impurities or other phases

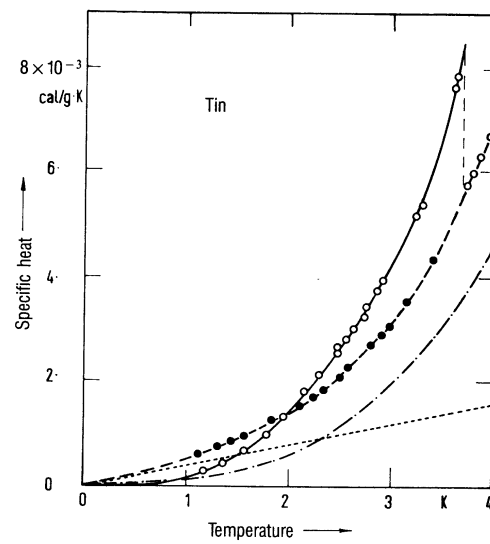


Figure 12.2: The specific heat of tin as a function of temperature. Open circles: no magnetic field; solid circles: $H > H_c$ (normal phase); dotted line: electron contribution to normal specific heat; dashed line: lattice contribution to normal specific heat. From W. Buckel, *Superconductivity: Fundamentals and Applications*, VCH, 1991.

present which do not become superconducting. As long as the superconducting portion forms a continuous network, the electrical resistance will be zero, even if the superconduction portion is very small. But the jump in specific heat is dependent on the amount of the sample which undergoes the phase transition (through v) and so can be used to determine the superconducting fraction.

12.3.1 Microscopic Interpretation

The results derived above were based on a few experimental observations, such as the Meissner effect ($M = -H$). They did not require us to know anything about what causes superconductivity, or why $M = -H$. In fact, these thermodynamic results were known long before a satisfactory microscopic theory of superconductivity had been developed.

The theory which finally gave a clear microscopic picture of superconductivity was developed in the 1950's by Bardeen, Cooper, and Schrieffer, and their theory is known as the BCS theory. The BCS theory shows that the reason some materials become superconducting at low temperature is that the electrons pair up in quantum-mechanical bound states called "Cooper pairs." The pairing is brought about by an interaction between the electrons and the positively-charged lattice. Since an electron has negative charge, the lattice atoms near an electron are attracted to it, leading to excess positive charge around an electron. This positive charge in turn attracts another electron, which creates an "effective" attractive interaction between the two electrons. Of course, the electrons also repel one another, but in some cases the attractive interaction can be stronger than the repulsion and the two electrons become bound together.

The situation is actually very similar to how two protons and two electrons bind together to form an H_2 molecule. Even though the electrons repel one another and so do the protons, by keeping the electrons most of the time between the protons, the electron-proton attraction is greater than the proton-proton and electron-electron repulsion and a stable molecule forms. The electrons and positively-charged lattice distortions ("phonons") arrange themselves similarly to form a stable Cooper pair. Just like a molecule, a Cooper pair requires energy to break it up.

It turns out that the number of Cooper pairs $n_c(T)$ is a function of temperature, but hardly depends on H . The number of pairs goes continuously to zero at $T = T_c$. This is why the specific heat of a superconductor is greater than in the normal phase – it requires energy to break up some Cooper pairs if the temperature is raised.

A Cooper pair has zero total angular momentum, and so are bosons obeying

Bose-Einstein statistics (the individual electrons, in contrast are fermions and obey Fermi-Dirac statistics). There is no restriction on the number of bosons which can occupy a single quantum level. In fact, the BCS theory shows the most-stable state of the system (the ground state) is one in which *all* Cooper pairs occupy the *same* quantum state. Thus, a single quantum-mechanical wavefunction which extends throughout the superconductor describes all of the Cooper pairs, which are rigidly locked together in phase (much like photons in laser light). It is this macroscopic quantum-mechanical aspect which is responsible for all of the strange properties such as zero resistance.

APPENDIX A

STATISTICAL MECHANICS OF THE MONATOMIC IDEAL GAS

NOTE: this appendix is taken unmodified from some older lecture notes I wrote for a different class. The notation may be a little different.

Consider a container of volume V which contains N atoms, which we will assume are "structureless" – that is, they contain no internal energy modes (vibration, rotation, or electronic excitation). Let us also assume that the potential energy due to interatomic forces is negligible compared to the kinetic energy of the atoms. These assumptions define the "monatomic ideal gas."

For simplicity, let the container be a cube of side length L , such that $V = L^3$. (The results we obtain will be independent of the assumed container geometry.)

The microstate of the system is specified by the positions and momenta of all atoms. If there are N atoms, then $3N$ coordinates (3 for each atom) must be specified, and $3N$ momenta, for a total of $6N$ degrees of freedom. The microstate of the gas is then fully specified by the $6N$ numbers $\{x_1, \dots, x_{3N}, p_1, \dots, p_{3N}\}$. We may represent the microstate geometrically as a point in a $6N$ -dimensional space, where the space is spanned by $3N$ orthogonal spatial coordinate axes x_i and $3N$ orthogonal momentum axes p_i . We call this $6N$ -dimensional space "phase space."

It is also useful to consider 2 subspaces of phase space. We will define "coordinate space" to be that space spanned by the $3N$ coordinate axes, and "momentum space" to be that space spanned by the $3N$ momentum axes.

In order keep the number of microstates finite, we divide up each coordinate axis into segments of length Δx , and each momentum axis into segments of length Δp . We can then construct a grid, or lattice, which divides phase space into small hypercubes of volume $\Delta W = (\Delta x \Delta p)^{3N}$. We will take all points within a given cube to correspond to one microstate. The number of microstates in a region of phase space with volume W is then simply

$$\Omega = W/\Delta W. \tag{A.1}$$

Therefore, rather than trying to count microstates directly (counting becomes

difficult when the numbers get very large!), we can find the volume of the region of phase space which satisfies some criteria, and then get Ω from Eq. (A.1).

Of course, the number of microstates will depend upon our choices of Δx and Δp . We'll see later that the only implication of this is that the entropy can only be determined to within some additive constant. But this was also true in thermodynamics – for the ideal gas, we integrated the Gibbs equation to get the entropy, which introduced an arbitrary integration constant s_0 . Recall that we introduced a postulate (the 3rd law of thermodynamics) which fixed the entropy at one temperature ($T = 0$) and removed the ambiguity. In statistical mechanics, no such postulate is needed – the problem is taken care of by a proper quantum-mechanical treatment of the problem.

Our objective is to count how many microstates (i.e., cubes in phase space) there are which are consistent with our macroscopic knowledge of the state of the system, for example the volume V and the internal energy U . Some of the microstates are consistent with our macroscopic specifications, but many are not (for example, those for which one or more atom lies outside the container).

There are two constraints on the system, which determine the region of phase space we are interested in. The first is that all atoms are located within the container:

$$0 < x_i < L, \quad i = 1, \dots, 3N. \quad (\text{A.2})$$

The second constraint is the specification of the total internal energy of the system. The total kinetic energy E of the gas is given by

$$E = \sum_{i=1}^{3N} \frac{1}{2} m v_i^2 = \sum_{i=1}^{3N} \frac{p_i^2}{2m}, \quad (\text{A.3})$$

where m is the mass of one atom. This equation shows that microstates with a given kinetic energy all lie on a sphere of radius $R = \sqrt{2mE}$ in momentum space.

Since we are assuming that translational kinetic energy is the only contribution to the total internal energy U of the gas (no potential energy or internal energy modes), the specification of the energy of the system is that $E = U$. Actually, we will assume that there is a very small, but non-zero, uncertainty in the energy. That is, the constraint on energy is

$$U - \delta U < E < U, \quad (\text{A.4})$$

where $\delta U \ll U$. We'll assume that Δx and Δp are chosen to be small enough that there are very many microstates with kinetic energy in this range.

Let us now find the volume W of the region of phase space satisfying eqs. (A.2) and (A.4). This volume may be expressed as a multidimensional integral:

$$W = \int_0^L dx_1 \int_0^L dx_2 \dots \int_0^L dx_{3N} W_{shell} \quad (\text{A.5})$$

where W_{shell} is the volume of the spherical shell in momentum space defined by Eq. (A.4). Since the kinetic energy does not depend on the spatial coordinates (it only depends on the momentum coordinates), the integral in momentum space defining W_{shell} is independent of the coordinate values, and we may reverse the order of integration and do the coordinate integrals first. Each of these contributes a factor L , and therefore

$$W = L^{3N} W_{shell} = V^N W_{shell} \quad (\text{A.6})$$

where $V = L^3$ is the container volume (in real 3D space).

To determine W_{shell} we may use a result from mathematics for the volume $W_n(R)$ of a sphere of radius R in n -dimensional space. It may be shown that

$$W_n(R) = \frac{(\pi R^2)^{n/2}}{(\frac{n}{2})!}. \quad (\text{A.7})$$

Note that if n is odd we can evaluate $(n/2)!$ in terms of the Gamma function: $z! = \Gamma(z+1)$ for any z . The Gamma function is defined by

$$\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt \quad (\text{A.8})$$

and satisfies the recurrence formula $\Gamma(z+1) = z\Gamma(z)$. Therefore, $\Gamma(n/2)$ may be determined from this recurrence formula, given that $\Gamma(3/2) = \sqrt{\pi}/2$.

For the familiar cases of $n = 2$ and $n = 3$, Eq. (A.7) reduces to $W_2 = \pi R^2$ and $W_3 = (4/3)\pi R^3$, respectively, as expected.

W_{shell} may be determined by the difference of two sphere volumes:

$$W_{shell} = W_{3N}(\sqrt{2mU}) - W_{3N}(\sqrt{2m(U - \delta U)}) \quad (\text{A.9})$$

Or, using Eq. (A.7),

$$W_{shell} = W_{3N}(\sqrt{2mU}) \left[1 + \left(1 - \frac{\delta U}{U} \right)^{3N/2} \right]. \quad (\text{A.10})$$

Note that the term $(1 - \delta U/U)$ is very slightly less than one. Therefore, for any finite (but very small) δU , we have that

$$\lim_{N \rightarrow \infty} \left(1 - \frac{\delta U}{U} \right)^{3N/2} = 0 \quad (\text{A.11})$$

and therefore

$$\lim_{N \rightarrow \infty} W_{shell} = W_{3N}(\sqrt{2mU}). \quad (\text{A.12})$$

This seemingly-paradoxical result states that in the limit of large N , the shell volume equals the volume of the entire sphere! This is in fact the case: as the dimensionality of the space increases, the distribution of the volume of the sphere with radius becomes highly peaked near the surface of the sphere. For any small but finite shell width, if n is large enough then virtually all of the sphere volume is located within the shell.

This means that for large N , the number of microstates available to a gas which have kinetic energy very close to U is *overwhelmingly* larger than the number of microstates which have kinetic energy less than U . In the large N limit (which is what we are interested in), we have then that

$$\Omega = W_{3N}(\sqrt{2mU})/\Delta W \quad (\text{A.13})$$

or

$$\Omega = \frac{(2\pi mU)^{3N/2}}{\Delta W \left(\frac{3N}{2}\right)!}. \quad (\text{A.14})$$

Actually, there is one correction we need to make to Eq. (A.14). In counting microstates by integrating over phase space, we have implicitly assumed that the atoms are *distinguishable*. That is, suppose we interchange the position and momentum coordinates of particles 1 and 2. Then if the original microstate was

$$\{\underbrace{x_1, x_2, x_3}_{\text{atom 1}}, \underbrace{x_4, x_5, x_6}_{\text{atom 2}}, x_7, \dots, x_{3N}, \underbrace{p_1, p_2, p_3}_{\text{atom 1}}, \underbrace{p_4, p_5, p_6}_{\text{atom 2}}, p_7, \dots, p_{3N}\},$$

the microstate with atoms 1 and 2 interchanged will be

$$\{\underbrace{x_4, x_5, x_6}_{\text{atom 2}}, \underbrace{x_1, x_2, x_3}_{\text{atoms 1}}, x_7, \dots, x_{3N}, \underbrace{p_4, p_5, p_6}_{\text{atom 2}}, \underbrace{p_1, p_2, p_3}_{\text{atom 1}}, p_7, \dots, p_{3N}\}.$$

The second microstate is different from the first unless atoms 1 and 2 happen to have the *same* position and momentum vectors, such that $x_1 = x_4$, etc. For a low-density gas, this is extremely unlikely, so in general exchanging atoms 1 and 2 produces a different microstate of the gas, which corresponds to a different point in phase space. Since this microstate satisfies the macroscopic constraints [eqs. Eq. (A.2) and Eq. (A.4)], both of the above microstates are in the accessible region of phase space, and therefore we have counted both of them in Ω .

This is correct if atoms 1 and 2 are distinguishable, so that exchanging them really does produce a new microstate of the gas. However, it turns out

that atoms which are identical (i.e., the same isotope of the same element) are actually *indistinguishable* in nature. That is, exchanging them *does not* produce a new microstate of the system.

Since we have in effect treated the atoms as distinguishable when counting microstates, we have overcounted the number of microstates, and need to correct our expression for Ω . Let us assume that the gas density is low enough that we can neglect the possibility that any 2 atoms are at the same position (to within Δx) and have the same momentum (to within Δp). Then we simply need to divide Ω by the number of ways to permute the atom labels among the N atoms, which is $N!$. Therefore, the correct expression for Ω is

$$\Omega = \frac{(2\pi mU)^{3N/2}}{\Delta W N! \left(\frac{3N}{2}\right)!}. \quad (\text{A.15})$$

Using Stirling's formula to approximate the factorial terms ($n! \approx n^n e^{-n}$), we have

$$\Omega = \left(\frac{eV}{N}\right)^N \left[\frac{4\pi emU}{3N}\right]^{3N/2} \frac{1}{(\Delta x \Delta p)^{3N}} \quad (\text{A.16})$$

Let us take the product $\Delta x \Delta p$ to be some constant a . Then the entropy is

$$S = k \ln \Omega \quad (\text{A.17})$$

$$= Nk \left[\ln \hat{v} + \frac{3}{2} \ln \left(\frac{4\pi m \hat{u}}{3a^2} \right) + \frac{5}{2} \right] \quad (\text{A.18})$$

where $\hat{v} = V/N$ and $\hat{u} = U/N$. Since this expresses S as a function of U , V , and N , we have succeeded in deriving a fundamental relation for the ideal gas.

The only uncertainty remaining is the additive constant resulting from the unknown value of a . If we had analyzed this problem from the point of view of quantum mechanics (rather than classical mechanics), we would have found the same result for S as in Eq. (A.18), but with the constant a replaced by *Planck's constant* h , which is a fundamental constant, given by

$$h = 6.626 \times 10^{-34} \text{ J-s}. \quad (\text{A.19})$$

With the substitution $a = h$, we then have

$$\hat{s} = k \left[\ln \hat{v} + \frac{3}{2} \ln \left(\frac{4\pi m \hat{u}}{3h^2} \right) + \frac{5}{2} \right]. \quad (\text{A.20})$$

(Note that the units of \hat{s} are J/molecule/K. If we would rather work with moles, we only need replace k by \hat{R} .)

Evaluating the temperature from $1/T = (\partial\hat{s}/\partial\hat{u})_{\hat{v}}$ we find

$$\hat{u} = \frac{3}{2}kT. \quad (\text{A.21})$$

Therefore, Eq. (A.20) may also be expressed as

$$\hat{s} = k \left[\ln \hat{v} + \frac{3}{2} \ln \left(\frac{2\pi mkT}{h^2} \right) + \frac{5}{2} \right]. \quad (\text{A.22})$$

Index

- adiabatic, 35
- air-standard cycle, 250
- Avagadro's Number, 149

- boiler, 126
- Boltzmann's constant, 87
- Brayton cycle, 250

- cavitation, 108
- center of mass, 20
- chemical potential, 54
- choked, 281
- coexistence lines, 77
- compressibility factor, 88
- condenser, 126
- conjugate variables, 218
- conservative force, 19
 - definition, 20
- control mass, 4
- control volume, 4
- COP, 168
- Corresponding States
 - Principle of, 100
- critical point, 73
- Curie substance, 43

- datum state, 71
- degrees of freedom, 63
- dielectric, 42
- differential equations of state, 192
- diffuser, 119

- electric dipole moment, 40
- electric susceptibility, 41
- endoreversible, 174

- endothermic, 200
- energy, 3
 - accounting, 9
 - balance, 9
 - conservation of, 24
- enthalpy
 - definition, 70
 - stagnation, 119
- enthalpy of vaporization, 126
- entropy, 3
 - ideal gas, 142
- environment, 3
- equations of state, 65
- equilibrium, 50
 - chemical, 55
 - diffusive, 53
 - mechanical, 52
 - phase, 54
 - restricted, 55
 - thermal, 52
- equilibrium properties, 108
- equipartition of energy, 91
- evaporator, 126
- exact differential, 189
- exothermic, 200
- extensive, 62

- First Law of Thermodynamics, 34
- force
 - centrifugal, 17
 - Lorentz, 17, 21
 - traction, 17
- fusion, 206

- gas constant, 86

- gas turbine cycle, 249
- generalized displacements, 44
- generalized equations of state, 98
- generalized forces, 44
- Gibbs free energy, 196

- heat capacity, 86
- heat exchangers, 124
- heat of vaporization, 126
- heat transfer, 32
 - rate, 35
- Helmholtz free energy, 195
- hydrostatics, 57

- ideal cycles, 229
- ideal gas, 87
- incompressible substance, 102
- inertial mass, 7
- intensive, 62
- intercooling, 259
- internal energy, 28
- irreversibility rate, 181
- irreversible expansion, 135
- isentropic, 159
- isobar, 78
- isothermal compressibility, 84

- kinetic energy, 8

- latent heat, 200
- Lee-Kesler generalized compressibility, 100
- lever rule, 80
- liquid
 - compressed, 79
 - subcooled, 79
- local thermodynamic equilibrium, 107

- Mach Number, 276
- macrostate, 130

- magnetization, 43
- manometer, 57
- mass density, 20
- microstate, 130

- normal shock waves, 285
- nozzle, 117

- oblique, 285
- one-dimensional flow, 108

- phase, 54
- phase diagrams, 72
- Planck's constant, 142
- polar, 40
- polarization, 41
- potential energy
 - definition, 23
 - gravitational, 19
- power, 12
- pressure, 33
 - critical, 72
 - reduced, 98

- quality, 79
- quasi-static, 37

- reference state, 71
- regenerator effectiveness, 258

- saturation, 78
- saturation pressure, 77
- saturation temperature, 77
- Second Law
 - Clausius statement, 167
 - Kelvin-Planck statement, 164
- specific, 62
- specific heat
 - c_p , 85
 - c_v , 85

- stagnation state, 276
- steady flow, 109
- steady flow availability function, 181
- steady state, 116
- Stirling's approximation, 142
- sublimation, 73
- subsonic, 276
- supercooled vapor, 98
- superheated liquid, 98
- supersonic, 276
- surface tension, 39
- system, 3
 - classification, 4

- temperature, 53
 - critical, 73
 - reduced, 98
 - thermodynamic, 61
- thermal conductivity, 161
- thermal efficiency, 165
- thermal energy, 1
- thermal expansion coefficient, 83
- thermal reservoir, 163
- thermodynamic pressure, 147
- thermodynamic state, 63
- thermodynamic temperature, 146
- triple line, 75
- triple point, 75

- unrestrained expansion, 135

- valve, 120
- van der Waals equation of state, 94
- vapor pressure, 77

- work, 8, 10
 - definition of, 10