22

Irreversibility

It would be much more impressive if it flowed the other way.

Oscar Wilde, on seeing the Niagara Falls

22.1 What Needs to be Explained?

The first thing we must establish is the meaning of the term 'irreversibility'. This is not quite as trivial as it might seem.[1] The irreversible behavior I will try to explain is that which is observed. Every day we see that time runs in only one direction in the real world. If I drop my keys, they fall to the floor and stay there; keys lying on the floor do not suddenly jump into my hand. This asymmetry of time, or 'arrow of time', has seemed to many as being incompatible with the time-reversal invariance of the fundamental equations of both classical and quantum physics.[2] This is the issue addressed in this chapter.

As with the development of statistical mechanics in general, the explanation of irreversibility given here is based on a large number of particles in a macroscopic system for which we have very limited information about the microscopic state. As in the rest of the book, this leads us to describe a macroscopic system on the basis of probability theory.

We will present an explanation of irreversibility using the example of the free expansion of a classical ideal gas. The microscopic equations of motion for this problem are time-reversal invariant, but the macroscopic behavior will nevertheless turn out to be irreversible. Because of the simplicity of the example we will be able to carry out every mathematical step exactly, so that the argument can be analyzed completely.[3]

---

[1] This chapter is based on an earlier paper by the author: R. H. Swendsen, 'Explaining irreversibility', *Am. J. Phys.*, 76, 643–48 (2008).

[2] The history of the debate on the origins of irreversibility is fascinating (and continuing), but we will not have space to delve into it. There are many books on this topic that the interested student might consult.

[3] The mathematical analysis we use was first derived by Harry Lloyd Frisch, an American chemist, who was born in Vienna, Austria (1928–2007). H. L. Frisch, 'An approach to equilibrium,' *Phys. Rev.*, 109, 22–9 (1958).

22.2   Trivial Form of Irreversibility

The irreversibility observed in daily life must be distinguished from a trivial form of irreversibility that appears only in infinite systems.[4]

Consider a particle moving in empty space. At some initial time it is observed to be located in some finite region of space. Since the particle is moving, it will eventually leave this region. If space is infinite, it will not return. This is, technically, irreversible behavior, but of a trivial kind.

This trivial form of irreversibility is real, and occurs in the radiation of light from a star. It is quite general in infinite systems, whether classical or quantum. An open system also displays this trivial form of irreversibility, since it is really just a piece of an infinite system.

However, we would like to separate this trivial irreversibility from the non-trivial irreversibility that we experience every day. Therefore, we will restrict the discussion to an isolated, finite system.

22.3   Boltzmann's H-Theorem

The history of the debate on irreversibility has been closely associated with responses to Boltzmann's famous H-theorem, in which he claimed to explain irreversibility. Although objections were aimed at Boltzmann's equations rather than the apparent paradox itself, discussing them can help clarify the problem.

Boltzmann derived an equation for the time-derivative of the distribution of atoms in a six-dimensional space of positions and momenta by approximating the number of collisions between atoms (*Stosszahlansatz*). This approximation had the effect of replacing the true macroscopic dynamics by a process that included random perturbations of the particle trajectories. Within the limits of his approximation, Boltzmann showed that a particular quantity that he called 'H' could not increase with time.

The fact that Boltzmann used an approximation—and an essential one—in deriving his result meant that his derivation could not be regarded as a proof. The arguments about its validity began soon after his publication appeared, and have continued to the present day.

22.4   Loschmidt's *Umkehreinwand*

The first objection came from Boltzmann's friend, Johann Josef Loschmidt (Austrian physicist, 1821–1895), who noted that if all momenta in an isolated system were reversed, the system would retrace its trajectory. If Boltzmann's H-function had been decreasing at the moment of reversal, its value after time-reversal must increase. Loschmidt argued on this basis that Boltzmann's conclusions could not be correct, and his argument is usually referred to by the German term *Umkehreinwand*.

---

[4] R. H. Swendsen, 'Irreversibility and the thermodynamic limit,' *J. of Stat. Phys.*, 10: 175–7 (1974).

Loschmidt's argument is very close to the central paradox. If every microscopic state that approaches equilibrium corresponds to a time-reversed state that moves away from equilibrium, should they not be equally probable?

22.5 Zermelo's *Wiederkehreinwand*

Ernst Zermelo (German mathematician, 1871–1953) was a prominent mathematician, who raised a different objection. He cited a recently derived theorem by Jules Henri Poincaré (French mathematician and physicist, 1854–1912) that proved that any isolated classical system must exhibit quasi-periodic behavior; that is, the system must return repeatedly to points in phase space that are arbitrarily close to its starting point. Zermelo claimed that Poincaré's theorem is incompatible with Boltzmann's H-theorem, which predicts that the system will never leave equilibrium. Zermelo's argument is referred to by the German term *Wiederkehreinwand*.

22.6 Free Expansion of a Classical Ideal Gas

In this section we will present an exact analysis of a simple model that exhibits irreversible behavior. Since the model is governed by time-reversal-invariant equations of motion, this will serve as a demonstration of the compatibility of irreversible macroscopic behavior and time-reversal invariance.

Consider an ideal gas of N particles in a volume V. Isolate the gas from the rest of the universe and assume that the walls of the box are perfectly reflecting. Initially, the gas is confined to a smaller volume V_o by an inner wall. At time $t = 0$, the inner wall is removed.

For convenience, let the box be rectangular and align the coordinate axes with its sides. The inner wall that initially confines the gas to a subvolume V_o is assumed to be perpendicular to the x-direction and located at $x = L_o$. The length of the box in the x-direction is L. The dependence of the probability distribution on the y- and z-coordinates does not change with time, so that we can treat it as a one-dimensional problem.

At time $t = 0$, the confining wall at L_o is removed and the particles are free to move throughout the box. Following Frisch's 1958 paper, we will eliminate the difficulties involved in describing collisions with the walls at $x = 0$ and $x = L$ by mapping the problem onto a box of length $2L$ with periodic boundary conditions. When a particle bounces off a wall in the original system, it corresponds to a particle passing between the positive and negative sides of the box without change of momentum. The mapping from the periodic system to the original system is then simply $x_j \rightarrow |x_j|$.

The key assumption is that the initial microscopic state must be described by a probability distribution. Assume that before the inner wall is removed, the initial positions of the particles are uniformly distributed in V_o, and the momenta are independent of each other and have a probability distribution $h(p_j)$. This probability distribution is assumed to be time-reversal invariant, $h(-p_j) = h(p_j)$, so that there is no time-asymmetry in the

initial conditions. Since we are assuming that the positions and momenta of different particles are initially independent, we can write the total probability distribution as

$$f_N\left(\{x_j,p_j|j=1\ldots N\},t=0\right)=\prod_{j=1}^{N}f\left(x_j,p_j,t=0\right)=\prod_{j=1}^{N}g\left(x_j\right)h\left(p_j\right) \tag{22.1}$$

where

$$g\left(x\right)=\begin{cases}1/L_o & -L_0<x<L_0.\\0 & |x|\geq L_0.\end{cases} \tag{22.2}$$

This initial probability distribution for the periodic system is shown schematically in Fig. 22.1.

Since the particle probabilities are independent, we can restrict our attention to the distribution function of a single particle. The periodic boundary conditions then allow us to Fourier transform the initial conditions to obtain

$$g(x)=g_o+\sum_{n=1}^{\infty}g_n\cos\left(\frac{\pi n}{L}x\right) \tag{22.3}$$

where no sine terms enter because of the symmetry of the initial conditions in the expanded representation with period $2L$.

Using the standard procedure of multiplying by $\cos\left(\pi n'x/L\right)$ and integrating to obtain the coefficients in eq.(22.3), we find

$$\int_{-L}^{L}g(x)\cos\left(\frac{\pi n'}{L}x\right)dx$$
$$=\int_{-L}^{L}g_o\cos\left(\frac{\pi n'}{L}x\right)dx+\sum_{n=1}^{\infty}g_n\int_{-L}^{L}\cos\left(\frac{\pi n}{L}x\right)\cos\left(\frac{\pi n'}{L}x\right)dx. \tag{22.4}$$
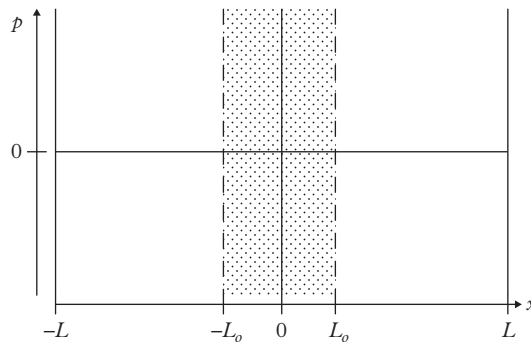

Fig. 22.1 *The initial probability distribution at t = 0.*

For $n' = 0$, this gives us

$$g_o = \frac{1}{L},\tag{22.5}$$

and for $n' \geq 1$

$$g_{n'} = \frac{2}{n'\pi L_o}\sin\left(\frac{n'\pi}{L}L_o\right).\tag{22.6}$$

The time development for a given momentum is now simple:

$$\begin{aligned}f(x,p,t) &= f\left(x - \left(\frac{p}{m}\right)t, p, 0\right)\\ &= h(p)\left[g_o + \sum_{n=1}^{\infty}g_n\cos\left(\frac{n\pi}{L}\left(x - \left(\frac{p}{m}\right)t\right)\right)\right].\end{aligned}\tag{22.7}$$

Fig. 22.2 shows the probability distribution at a time $t > 0$. The shaded areas, representing non-zero probability density, tilt to the right. As time goes on, the shaded areas become increasingly flatter and closer together.

To find an explicit solution for a special case, assume that the initial probability distribution for the momentum is given by the Maxwell–Boltzmann distribution,

$$h(p) = \left(\frac{\beta}{2\pi m}\right)^{1/2}\exp\left[-\beta\left(\frac{p^2}{2m}\right)\right]\tag{22.8}$$

where T is the temperature, and $\beta = 1/k_B T$. The assumption of a Maxwell–Boltzmann distribution is not essential, but it provides an explicit example of how the approach to equilibrium comes about.
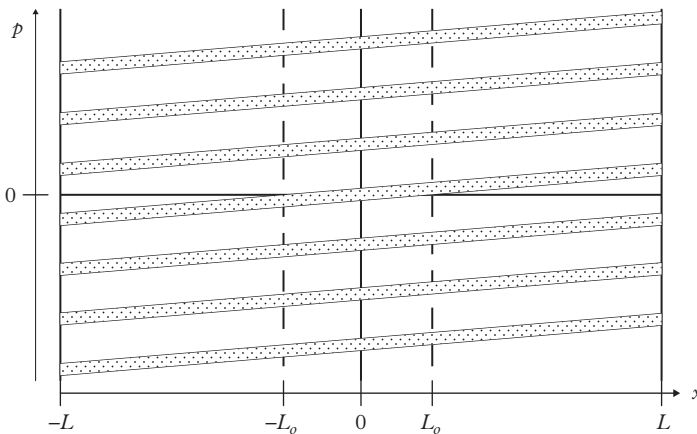

Fig. 22.2 *The probability distribution for t > 0.*

Inserting eq. (22.8) in eq. (22.7), we can integrate over the momenta to calculate the local probability density as a function of time,

$$
\begin{aligned}
g(x,t) &= \int_{-\infty}^{\infty} f(x,t)\, dp \\
&= \int_{-\infty}^{\infty} h(p)\, L^{-1}\, dp \\
&+ \sum_{n=1}^{\infty} g_n \int_{-\infty}^{\infty} h(p) \cos\left(\frac{n\pi}{L}\left(x - \left(\frac{p}{m}\right)t\right)\right) dp.
\end{aligned}
\tag{22.9}
$$

The first integral is trivial, since $h(p)$ is normalized,

$$
\int_{-\infty}^{\infty} h(p)\, dp = 1.
\tag{22.10}
$$

The integrals in the sum are rather tedious, but not particularly difficult when you use the identity $\cos\theta = \frac{1}{2}\left(e^{i\theta} + e^{-i\theta}\right)$ to write them in the form

$$
\int_{-\infty}^{\infty} \exp\left[-\beta\frac{p^2}{2m} \pm \frac{n\pi}{L} i \frac{p}{m} t\right] dp,
\tag{22.11}
$$

and complete the square in the exponent. The result is

$$
g(x,t) = L^{-1} + \frac{2}{\pi L_o} \sum_{n=1}^{\infty} \frac{1}{n} \sin\left(\frac{n\pi}{L}L_o\right) \cos\left(\frac{n\pi}{L}x\right) \exp\left[-\lambda_n^2 t^2\right],
\tag{22.12}
$$

where

$$
\lambda_n^2 = \frac{n^2\pi^2}{2mL^2\beta} = \frac{n^2\pi^2}{2mL^2}k_B T.
\tag{22.13}
$$

Note that since the temperature is related to the one-dimensional root-mean-square velocity by

$$
\frac{1}{2}k_B T = \frac{1}{2m}\left\langle p^2\right\rangle = \frac{m}{2}\left\langle v^2\right\rangle = \frac{m}{2}v_{rms}^2.
\tag{22.14}
$$

The coefficients λ_n can be written as

$$
\lambda_n^2 = \frac{n^2\pi^2}{2L^2}v_{rms}^2.
\tag{22.15}
$$

Eq. (22.12) can also be expressed in terms of the characteristic time for a particle traveling with the speed v_{rms} to cross the box,

$$\tau = \frac{L}{v_{rms}}. \tag{22.16}$$

In terms of τ, the time-dependent local probability density becomes

$$g(x,t) = L^{-1} \left[1 + \frac{L}{L_o} \sum_{n=1}^{\infty} \left(\frac{2}{n\pi} \right) \sin\left(\frac{n\pi}{L} L_o \right) \cos\left(\frac{n\pi}{L} x \right) \exp\left[-\left(\frac{\pi^2 n^2}{2} \right) \left(\frac{t^2}{\tau^2} \right) \right] \right]. \tag{22.17}$$

Since

$$\left| \left(\frac{2}{n\pi} \right) \sin\left(\frac{n\pi}{L} L_o \right) \cos\left(\frac{n\pi}{L} x \right) \right| \leq \frac{2}{n\pi} \tag{22.18}$$

for all x, the sum converges rapidly for $t > 0$. At $t = \tau$, the *n-th* term in the sum is less than $(2/n\pi) \exp\left[-\left(\frac{\pi^2}{2} \right) n^2 \right]$, and even the leading non-constant term has a relative size of less than $0.005 \, (L/L_o)$. The rapidity of convergence is striking. The inclusion of interactions in the model would actually slow down the approach to equilibrium.

It is also possible to calculate the energy density as a function of position and time by including a factor of $p^2/2m$ in the integral over the momentum,

$$\begin{aligned} U(x,t) &= \int_{-\infty}^{\infty} f(x,p,t) \left(\frac{p^2}{2m} \right) dp \\ &= \int_{-\infty}^{\infty} h(p) \left(\frac{p^2}{2m} \right) L^{-1} dp \\ &\quad + \sum_{n=1}^{\infty} g_n \int_{-\infty}^{\infty} h(p) \left(\frac{p^2}{2m} \right) \cos\left(\frac{n\pi}{L} \left(x - \left(\frac{p}{m} \right) t \right) \right) dp. \end{aligned} \tag{22.19}$$

The integrals are even more tedious than those in eq. (22.9), but again not particularly difficult:

$$\begin{aligned} U(x,t) &= \frac{1}{2L\beta} - \sum_{n=1}^{\infty} \left(\frac{n^2\pi^2 t^2 - L^2 m\beta}{L^2 L_o mn\pi \beta^2} \right) \\ &\quad \sin\left(\frac{n\pi}{L} L_o \right) \cos\left(\frac{n\pi}{L} x \right) \exp\left[-\left(\frac{\pi^2 n^2 t^2}{2\tau^2} \right) \right]. \end{aligned} \tag{22.20}$$

After the internal wall is removed, the energy distribution is not uniform and not proportional to the particle density, since the faster particles move more rapidly into

the region that was originally vacuum. However, the energy density converges rapidly to the expected constant, $1/2L\beta = \frac{1}{2}k_BT/L$, as $t \to \infty$.

22.7   Zermelo's *Wiederkehreinwand* Revisited

We have now established that that the particle density and energy density both go to equilibrium. This leaves the question of reconciling the approach to equilibrium with Poincaré recurrences. Fortunately, it is easy to see how quasi-periodic behavior can arise in the ideal gas model, without contradicting the approach to equilibrium. To do this, it helps to have an intuitive picture of how Poincaré cycles occur.

First consider two particles with speeds v_1 and v_2. They will each return to their original states with periods $\tau_1 = 2L/v_1$ and $\tau_2 = 2L/v_2$, respectively. In general, the ratio τ_1/τ_2 will be irrational, but it can be approximated to arbitrary accuracy by a rational number, $\tau_1/\tau_2 \approx n_1/n_2$, where n_1 and n_2 are sufficiently large integers. Therefore, after a time $\tau_{1,2} = n_2\tau_1 \approx n_1\tau_2$, both particles will return to positions and velocities arbitrarily close their initial states .

Now add a third particle with speed v_3 and period $\tau_3 = 2L/v_3$. A rational approximation $\tau_{1,2}/\tau_3 \approx n_{1,2}/n_3$, will give us an approximate recurrence after a time $\tau_{1,2,3} = n_{1,2}\tau_3 \approx n_3\tau_{1,2} \approx n_3n_2\tau_1$. Since n_2 and n_3 will usually be large numbers, $\tau_{1,2,3}$ will usually be a long time.

If we repeat this procedure for 10^{23} particles, we will arrive (with probability one) at a recurrence time that would be enormous even in comparison to the age of the universe. It might naturally be said that we are not interested in such extremely long times, but it is interesting to note that eqs. (22.17) and (22.20) do not exhibit the quasi-periodic behavior that Poincaré recurrence might seem to require.

The resolution of the apparent contradiction lies in our lack of knowledge of the exact initial velocities and the extreme sensitivity of the Poincaré recurrence time to tiny changes in initial velocities. Even with far more detailed information than we are ever going to obtain in a real experiment, we would not be able predict a Poincaré recurrence with an uncertainty of less than many ages of the universe.

Poincaré recurrences are included in the exact solution for the free expansion of an ideal gas. However, since we cannot predict them, they appear as extremely rare large fluctuations. Since such large fluctuations are always possible in equilibrium, there is no contradiction between observed irreversible behavior and Poincaré recurrences.

22.8   Loschmidt's *Umkehreinwand* Revisited

Loschmidt's *Umkehreinwand* was directed against Boltzmann's equation, which was not time-reversal invariant. However, the exact solution to the free expansion of an ideal gas retains the time-reversal properties of the microscopic equations. If we reverse all velocities in the model some time after the inner wall is removed, the particles return to their original positions.

Under normal conditions, the reversal of molecular velocities is experimentally difficult, to say the least. However, the reversal of spin precession can be accomplished and is fundamental to magnetic resonance imaging. It is known as the spin-echo effect. After an initial magnetic pulse aligns the spins in a sample, interactions between the spins and inhomogeneities in the sample lead to decoherence and a decay of this magnetization. If no further action is taken, the signal will not return. However, if a second magnetic pulse is used to rotate the spins by 180° at a time t after the first pulse, it effectively reverses the direction of precession of the spins. The spins realign (or 'refocus') after a total time $2t$ after the initial pulse, and the magnetization appears again.

The reversal of the time-development of an 'irreversible' process is therefore not a flaw in the free-expansion model, but a reflection of the reality that such experiments can be carried out.

22.9   What is 'Equilibrium'?

An interesting feature of the exact solution to the free-expansion model can be seen in Fig. 22.2. Even though the shaded regions, indicating non-zero probabilities, become progressively thinner and closer together, the local probability density at any point along the trajectory of the system remains constant. This is consistent with Liouville's theorem that the total time derivative of the probability distribution vanishes.[5]

This property of all isolated Hamiltonian systems has caused difficulties for those who would like to define equilibrium in terms of a smooth distribution in phase space. Although we assumed that our model experiment started with a canonical probability distribution, it will certainly never evolve to one.

I believe that the difficulty is a confusion about the direction of inference. It has been amply demonstrated that a canonical distribution accurately describes an equilibrium state. However, that does not imply that a system in equilibrium can only be described by a canonical distribution. The probability distribution for our model at long times will give the same macroscopic predictions as the canonical distribution, even though the two distributions are different.

22.10   Entropy

Liouville's theorem has also caused difficulties for the traditional textbook definition of the entropy as the logarithm of a volume in phase space. The theorem requires that this volume remain constant, so the traditional expression for the entropy cannot increase in an isolated system.

This seems to violate the Second Law, but it is correct—in a certain sense. The traditional definition of entropy is related to the total information we have about the

---

[5] See Section 19.5 for a derivation of the Liouville theorem.

system—not the thermodynamic information about the current and future behavior of the system. The information that the system was initially confined to a smaller volume is contained in the layered structure of the probability density shown in Fig. 22.2. Since that information does not change, the traditional entropy does not change. The apparent violation of the Second Law arises because the traditional entropy does not correspond to the thermodynamic entropy.

If we use Boltzmann's definition of the thermodynamic entropy in terms of the probability of a macroscopic state, we obtain an expression that increases with time as expected for the thermodynamic entropy.

The specific form of the entropy depends on what we are measuring in the experiment. Since we are interested in the time development of the system as it approaches a uniform state, it would be reasonable to observe its properties as a function of position. Let us divide the system into M subsystems, each with length $\Delta L = L/M$, and measure the energy and density in each subsystem. For convenience, assume that the first m subsystems were inside the inner wall before the free expansion began, so that $L_o = mL/M$.

If M is large, we can assume that the energy and density are uniform across a subsystem. The number of particles in the j-th subsystem, $N_j(t)$, is given by ΔL times the expression in eq. (22.12), and the energy in the j-th subsystem, $E_j(t)$, is given by ΔL times the expression in eq. (22.20).

If the subsystems are large enough that $1/\sqrt{N_j(t)}$ is much smaller than the relative error in experimental measurements, they can be regarded as macroscopic systems. Their individual entropies might then be well approximated by the equilibrium entropy of a one-dimensional, classical ideal gas,

$$S(E_j(t), L/M, N_j(t)) = N_j(t)k_B T \left[\ln\left(\frac{L/M}{N_j(t)}\right) + \frac{1}{2}\ln\left(\frac{E_j(t)}{N_j(t)}\right) + X \right], \qquad (22.21)$$

where X is the usual constant. The total time-dependent entropy of the whole system is then given by

$$S(\{E_j(t)\}, L, N_j(t), /M) = \sum_{j=1}^{M} S\left(E_j(t), L/M, N_j(t)\right). \qquad (22.22)$$

This expression has the properties that at $t = 0$, it takes on the value of the initial entropy before the inner wall was removed,

$$S(\{E_j(0)\}, L, N_j(0), /M) = S(E, L_o, N), \qquad (22.23)$$

and as $t \to \infty$, it goes rapidly to the equilibrium entropy of the full system,

$$\lim_{t\to\infty} S(\{E_j(t)\}, L, N_j(t), /M) = S(E, L, N). \qquad (22.24)$$

These two properties are independent of the number M of subsystems that are observed.

22.11   Interacting Particles

The explanation of irreversible phenomena given here is, of course, not complete. The apparent conflict between microscopic and macroscopic laws has been resolved, but we have ignored the effects of interactions. However, now that it is clear that there is no real conflict between time-reversal-invariant microscopic laws and macroscopic irreversibility, it should be sufficient to demonstrate the equilibration of the momentum distribution by molecular dynamics computer simulations that are generalizations of those carried out in the assignments in Chapter 19.