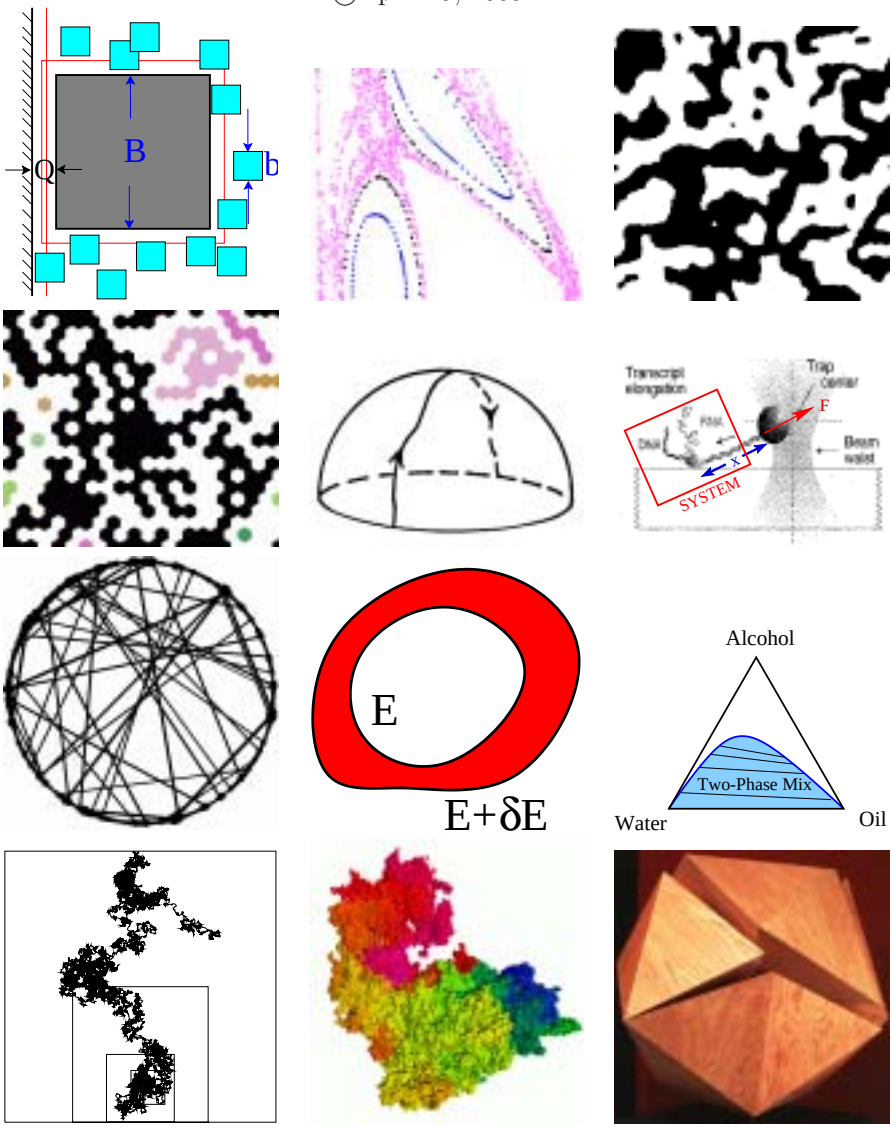
# Statistical Mechanics: Entropy, Order Parameters and Complexity

James P. Sethna, Physics, Cornell University, Ithaca, NY






Electronic version of text available at
http://www.physics.cornell.edu/sethna/StatMech/book.pdf

# Contents

# Why Study Statistical Mechanics?

# 1

Many systems in nature are far too complex to analyze directly. Solving for the motion of all the atoms in a block of ice – or the boulders in an earthquake fault, or the nodes on the Internet – is simply infeasible. Despite this, such systems often show simple, striking behavior. We use statistical mechanics to explain the simple behavior of complex systems.

Statistical mechanics brings together concepts and methods that infiltrate many fields of science, engineering, and mathematics. Ensembles, entropy, phases, Monte Carlo, emergent laws, and criticality – all are concepts and methods rooted in the physics and chemistry of gases and liquids, but have become important in mathematics, biology, and computer science. In turn, these broader applications bring perspective and insight to our fields.

Let's start by briefly introducing these pervasive concepts and methods.

**Ensembles:** The trick of statistical mechanics is not to study a single system, but a large collection or *ensemble* of systems. Where understanding a single system is often impossible, calculating the behavior of an enormous collection of similarly prepared systems often allows one to answer most questions that science can be expected to address.

For example, consider the random walk (figure 1.1). (You might imagine it as the trajectory of a particle in a gas, or the configuration of a polymer in solution.) While the motion of any given walk is irregular (left) and hard to predict, simple laws describe the distribution of motions of an infinite ensemble of random walks starting from the same initial point (right). Introducing and deriving these ensembles are the themes of chapters 3, 4, and 6.

**Entropy:** Entropy is the most influential concept arising from statistical mechanics (chapter 5). Entropy, originally understood as a thermodynamic property of heat engines that could only increase, has become science's fundamental measure of disorder and information. Although it controls the behavior of particular systems, entropy can only be defined within a statistical ensemble: it is the child of statistical mechanics, with no correspondence in the underlying microscopic dynamics. Entropy now underlies our understanding of everything from compression algorithms for pictures on the Web to the heat death expected at the end of the universe.

**Phases.** Statistical mechanics explains the existence and properties of

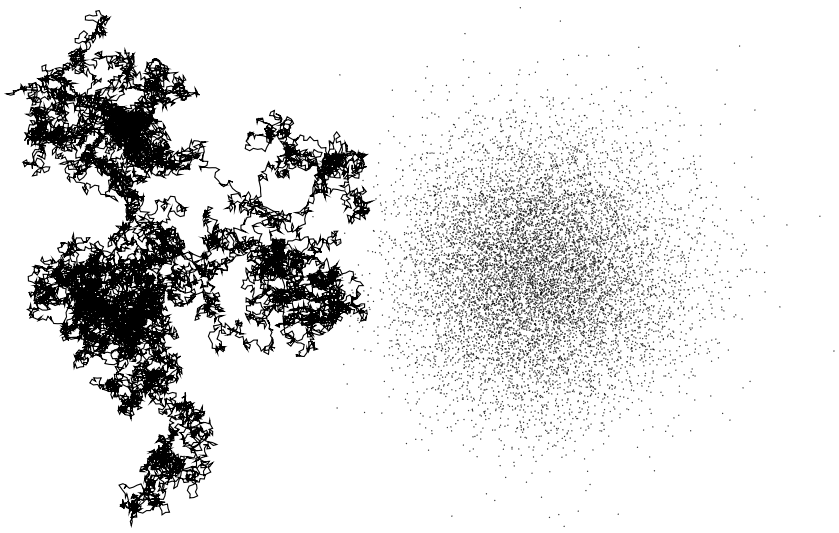


**Fig. 1.1 Random Walks.** The motion of molecules in a gas, or bacteria in a liquid, or photons in the Sun, is described by an irregular trajectory whose velocity rapidly changes in direction at random. Describing the specific trajectory of any given random walk (left) is not feasible or even interesting. Describing the statistical average properties of a large number of random walks is straightforward; at right is shown the endpoints of random walks all starting at the center. The deep principle underlying statistical mechanics is that it is often easier to understand the behavior of *ensembles* of systems.

phases. The three common phases of matter (solids, liquids, and gases) have multiplied into hundreds: from superfluids and liquid crystals, to vacuum states of the universe just after the Big Bang, to the pinned and sliding 'phases' of earthquake faults. Phases have an integrity or stability to small changes in external conditions or composition[1] – with deep connections to *perturbation theory*, section 8.1. Phases often have a rigidity or stiffness, which is usually associated with a *spontaneously broken symmetry*. Understanding what phases are and how to describe their properties, excitations, and topological defects will be the themes of chapters 7,[2] and 9.

**Computational Methods:** Monte–Carlo methods use simple rules to allow the computer to find ensemble averages in systems far too complicated to allow analytical evaluation. These tools, invented and sharpened in statistical mechanics, are used everywhere in science and technology – from simulating the innards of particle accelerators, to studies of traffic flow, to designing computer circuits. In chapter 8, we introduce the Markov–chain mathematics that underlies Monte–Carlo.

**Emergent Laws.** Statistical mechanics allows us to derive the new

[2]Chapter 7 focuses on quantum statistical mechanics: quantum statistics, metals, insulators, superfluids, Bose condensation, ... To keep the presentation accessible to a broad audience, the rest of the text is not dependent upon knowing quantum mechanics.

[1]Water remains a liquid, with only perturbative changes in its properties, as one changes the temperature or adds alcohol. Indeed, it is likely that all liquids are connected to one another, and indeed to the gas phase, through paths in the space of composition and external conditions.

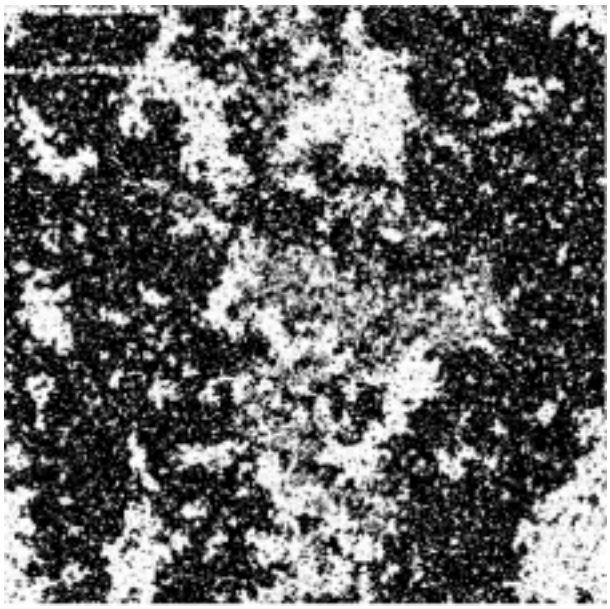


**Fig. 1.2 Temperature: the Ising model** at the critical temperature. Traditional statistical mechanics focuses on understanding phases of matter, and transitions between phases. These phases – solids, liquids, magnets, superfluids – are *emergent* properties of many interacting molecules, spins, or other degrees of freedom. Pictured here is a simple two-dimensional model at its magnetic transition temperature $T_c$. At higher temperatures, the system is non-magnetic: the magnetization is on average zero. At the temperature shown, the system is just deciding whether to magnetize upward (white) or downward (black). While predicting the time dependence of all these degrees of freedom is not practical or possible, calculating the average behavior of many such systems (a *statistical ensemble*) is the job of statistical mechanics.

laws that emerge from the complex microscopic behavior. These laws become exact only in certain limits. Thermodynamics – the study of heat, temperature, and entropy – becomes exact in the limit of large numbers of particles. Scaling behavior and power laws – both at phase transitions and more broadly in complex systems – emerge for large systems tuned (or self–organized) near critical points. The right figure 1.1 illustrates the simple law (the diffusion equation) that describes the evolution of the end-to-end lengths of random walks in the limit where the number of steps becomes large. Developing the machinery to express and derive these new laws are the themes of chapters 9 (phases), and 12 (critical points). Chapter 10 systematically studies the fluctuations about these emergent theories, and how they relate to the response to external forces.

**Phase Transitions.** Beautiful spatial patterns arise in statistical mechanics at the transitions between phases. Most of these are abrupt phase transitions: ice is crystalline and solid until abruptly (at the edge of the ice cube) it becomes unambiguously liquid. We study nucleation and the exotic structures that evolve at abrupt phase transitions in chapter 11.

Other phase transitions are continuous. Figure 1.2 shows a snapshot of the Ising model at its phase transition temperature $T_c$. The Ising model is a lattice of sites that can take one of two states. It is used as a simple model for magnets (spins pointing up or down), two component crystalline alloys (A or B atoms), or transitions between liquids and gases (occupied and unoccupied sites).[3] All of these systems, at their critical

[3]The Ising model has more far-flung applications: the three–dimensional Ising model has been useful in the study of quantum gravity.

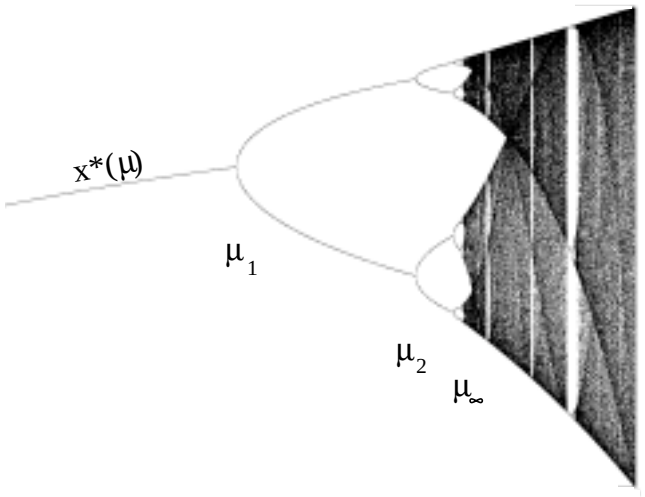



**Fig. 1.3 Dynamical Systems and Chaos.** The ideas and methods of statistical mechanics have close ties to many other fields. Many nonlinear differential equations and mappings, for example, have qualitative changes of behavior (bifurcations) as parameters are tuned, and can exhibit chaotic behavior. Here we see the long–time 'equilibrium' dynamics of a simple mapping of the unit interval into itself as a parameter $\mu$ is tuned. Just as an Ising magnet goes from one unmagnetized state above $T_c$ to two magnetized states below $T_c$, so this system goes from a periodic state below $\mu_1$ to a period–two cycle above $\mu_1$. Above $\mu_\infty$, the behavior is chaotic. The study of chaos has provided us with our fundamental explanation for the increase of entropy in statistical mechanics. Conversely, tools developed in statistical mechanics have been central to the understanding of the onset of chaos.

points, share the self-similar, fractal structures seen in the figure: the system can't decide whether to stay gray or to separate into black and white, so it fluctuates on all scales. Another self–similar, fractal object emerges from random walks (left figure 1.1, also figure 2.2) even without tuning to a critical point: a blowup of a small segment of the walk looks statistically similar to the original path. Chapter 12 develops the scaling and renormalization–group techniques that we use to understand these self–similar, fractal properties.

**Applications.** Science grows through accretion, but becomes potent through distillation. Each generation expands the knowledge base, extending the explanatory power of science to new domains. In these explorations, new unifying principles, perspectives, and insights lead us to a deeper, simpler understanding of our fields.

The period doubling route to chaos (figure 1.3) is an excellent example of how statistical mechanics has grown tentacles into disparate fields, and has been enriched thereby. On the one hand, renormalization–group methods drawn directly from statistical mechanics (chapter 12) were used to explain the striking scaling behavior seen at the onset of chaos (the geometrical branching pattern at the left of the figure). These methods also predicted that this behavior should be *universal*: this same period–doubling cascade, with quantitatively the same scaling behavior, would be seen in vastly more complex systems. This was later verified everywhere from fluid mechanics to models of human walking. Conversely, the study of chaotic dynamics has provided our most convincing microscopic explanation for the increase of entropy in statistical mechanics (chapter 5), and is the fundamental explanation of why ensembles

are useful and statistical mechanics is possible.

We provide here the distilled version of statistical mechanics, invigorated and clarified by the accretion of the last four decades of research. The text in each chapter will address those topics of fundamental importance to all who study our field: the exercises will provide in-depth introductions to the accretion of applications in mesoscopic physics, astrophysics, dynamical systems, information theory, low–temperature physics, statistics, biology, lasers, and complexity theory. The goal is to broaden the presentation to make it useful and comprehensible to sophisticated biologists, mathematicians, computer scientists, or complex–systems sociologists – thereby enriching the subject for the physics and chemistry students, many of whom will likely make excursions in later life into these disparate fields.

## Exercises

Exercises 1.1–1.3 provide a brief review of probability distributions. *Quantum Dice* explores discrete distributions and also acts as a gentle preview into Bose and Fermi statistics. *Probability Distributions* introduces the form and moments for the key distributions for continuous variables and then introduces convolutions and multidimensional distributions. *Waiting Times* shows the paradoxes one can concoct by confusing different ensemble averages. *Stirling* part (a) derives the useful approximation $n! \sim \sqrt{2\pi n}(n/e)^n$; more advanced students can continue in the later parts to explore *asymptotic series*, which arise in typical perturbative statistical mechanics calculations. *Random Matrix Theory* briefly introduces a huge field, with applications in nuclear physics, mesoscopic physics, and number theory; part (a) provides a good exercise in histograms and ensembles, and the remaining more advanced parts illustrate level repulsion, the Wigner surmise, universality, and emergent symmetry.

### (1.1) Quantum Dice. (Quantum) (With Buchan. [15])

You are given several unusual 'three-sided' dice which, when rolled, show either one, two, or three spots. There are three games played with these dice, *Distinguishable*, *Bosons* and *Fermions*. In each turn in these games, the player rolls one die at a time, starting over if required by the rules, until a legal combination occurs. In *Distinguishable*, all rolls are legal. In *Bosons*, a roll is legal only if the new number is larger or equal to the preceding number. In *Fermions*, a roll is legal only if the new number is strictly larger than the preceding number. See figure 1.4 for a table of possibilities after rolling two dice.

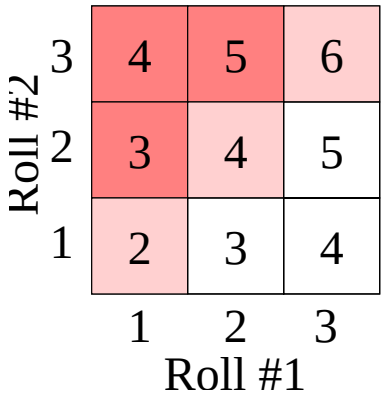



**Fig. 1.4** Rolling two dice. In *Bosons*, one accepts only the rolls in the shaded squares, with equal probability 1/6. In *Fermions*, one accepts only the rolls in the darkly shaded squares (not including the diagonal), with probability 1/3.

*(a) Presume the dice are fair: each of the three numbers of dots shows up* 1/3 *of the time. For a legal turn rolling a die twice in Bosons, what is the probability* $\rho(4)$ *of rolling a 4? Similarly, among the legal Fermion turns rolling two dice, what is the probability* $\rho(4)$*?*

Our dice rules are the same ones that govern the quantum statistics of identical particles.

*(b) For a legal turn rolling three 'three-sided' dice in Fermions, what is the probability $\rho(6)$ of rolling a 6?* (Hint: there's a Fermi exclusion principle: when playing Fermions, no two dice can have the same number of dots showing.) Electrons are fermions; no two electrons can be in exactly the same state.

When rolling two dice in *Bosons*, there are six different legal turns (11), (12), (13), ..., (33): half of them are doubles (both numbers equal), when for plain old *Distinguishable* turns only one third would be doubles[4]: the probability of getting doubles is enhanced by 1.5 times in two-roll *Bosons*. When rolling three dice in *Bosons*, there are ten different legal turns (111), (112), (113), ..., (333). When rolling $M$ dice each with $N$ sides in *Bosons*, one can show that there are $\binom{N+M-1}{M} = \frac{(N+M-1)!}{M!(N-1)!}$ legal turns.

*(c) In a turn of three rolls, what is the enhancement of probability of getting triples in Bosons over that in Distinguishable? In a turn of $M$ rolls, what is the enhancement of probability for generating an $M$-tuple (all rolls having the same number of dots showing)?*

Notice that the states of the dice tend to cluster together in *Bosons*. Examples of real bosons clustering into the same state include Bose condensation (section 7.6.3) and lasers (exercise 7.7).

(1.2) **Probability Distributions.** (Basic)

Most people are more familiar with probabilities for discrete events (like coin flips and card games), than with probability distributions for continuous variables (like human heights and atomic velocities). The three continuous probability distributions most commonly encountered in physics are: (*i*) **Uniform:** $\rho_{\text{uniform}}(x) = 1$ for $0 \le x < 1$, $\rho(x) = 0$ otherwise; produced by random number generators on computers; (*ii*) **Exponential:** $\rho_{\text{exponential}}(t) = e^{-t/\tau}/\tau$ for $t \ge 0$, familiar from radioactive decay and used in the collision theory of gases; and (*iii*) **Gaussian:** $\rho_{\text{gaussian}}(v) = e^{-v^2/2\sigma^2}/(\sqrt{2\pi}\sigma)$, describing the probability distribution of velocities in a gas, the distribution of positions at long times in random walks, the sums of random variables, and the solution to the diffusion equation.

*(a)* **Likelihoods.** *What is the probability that a random number uniform on $[0, 1)$ will happen to lie between $x = 0.7$ and $x = 0.75$? That the waiting time for a radioactive decay of a nucleus will be more than twice the exponential decay time $\tau$? That your score on an exam with Gaussian distribution of scores will be greater than $2\sigma$ above the mean?* (Note: $\int_2^\infty (1/\sqrt{2\pi}) \exp(-v^2/2)\, dv = (1 - \text{erf}(\sqrt{2}))/2 \sim 0.023$.)

*(b)* **Normalization, Mean, and Standard Deviation.** *Show that these probability distributions are normalized: $\int \rho(x)dx = 1$. What is the mean $x_0$ of each distribution? The standard deviation $\sqrt{\int (x - x_0)^2 \rho(x)dx}$?* (You may use the formulas $\int_{-\infty}^{\infty} (1/\sqrt{2\pi}) \exp(-x^2/2)\, dx = 1$ and $\int_{-\infty}^{\infty} x^2 (1/\sqrt{2\pi}) \exp(-x^2/2)\, dx = 1$.)

*(c)* **Sums of variables.** *Draw a graph of the probability distribution of the sum $x + y$ of two random variables drawn from a uniform distribution on $[0, 1)$. Argue in general that the sum $z = x + y$ of random variables with distributions $\rho_1(x)$ and $\rho_2(y)$ will have a distribution given by the convolution $\rho(z) = \int \rho_1(x)\rho_2(z - x)\, dx$.*

**Multidimensional probability distributions.** In statistical mechanics, we often discuss probability distributions for many variables at once (for example, all the components of all the velocities of all the atoms in a box). Let's consider just the probability distribution of one molecule's velocities. If $v_x$, $v_y$, and $v_z$ of a molecule are independent and each distributed with a Gaussian distribution with $\sigma = \sqrt{kT/M}$ (section 3.2.2) then we describe the combined probability distribution as a function of three variables as the product of the three Gaussians:

$$\begin{aligned}\rho(v_x, v_y, v_z) &= 1/(2\pi(kT/M))^{3/2} \exp(-M\mathbf{v}^2/2kT) \\ &= \left(\sqrt{\frac{M}{2\pi kT}} e^{\frac{-Mv_x^2}{2kT}}\right)\left(\sqrt{\frac{M}{2\pi kT}} e^{\frac{-Mv_y^2}{2kT}}\right) \\ &\quad \left(\sqrt{\frac{M}{2\pi kT}} e^{\frac{-Mv_z^2}{2kT}}\right). \end{aligned} \tag{1.1}$$

*(d) Show, using your answer for the standard deviation of the Gaussian in part (b), that the mean kinetic energy is $kT/2$ per dimension. Show that the probability that the speed is $v = |\mathbf{v}|$ is given by a Maxwellian distribution*

$$\rho_{\text{Maxwell}}(v) = \sqrt{2/\pi}(v^2/\sigma^3) \exp(-v^2/2\sigma^2). \tag{1.2}$$

(Hint: What is the shape of the region in 3D velocity space where $|\mathbf{v}|$ is between $v$ and $v + \delta v$? The area of a sphere of radius $R$ is $4\pi R^2$.)

(1.3) **Waiting times.** (Math) (With Brouwer. [14])

On a highway, the average numbers of cars and buses going east are equal: each hour, on average, there are 12 buses and 12 cars passing by. The buses are scheduled: each bus appears exactly 5 minutes after the previous one. On the other hand, the cars appear at random: in a short interval $dt$, the probability that a car comes by is $dt/\tau$,

[4] For *Fermions*, of course, there are no doubles.

with $\tau = 5$ minutes. An observer is counting the cars and buses.

*(a) Verify that each hour the average number of cars passing the observer is 12.*

*(b) What is the probability $P_{\rm bus}(n)$ that $n$ buses pass the observer in a randomly chosen 10 minute interval? And what is the probability $P_{\rm car}(n)$ that $n$ cars pass the observer in the same time interval?* (Hint: For the cars, one way to proceed is to divide the interval into many small slivers of time $dt$: in each sliver the probability is $dt/\tau$ that a car passes, and $1 - dt/\tau \approx e^{-dt/\tau}$ that no car passes. However you do it, you should get a Poisson distribution, $P_{\rm car}(n) = a^n e^{-a}/n!$ See also exercise 3.5.)

*(c) What is the probability distribution $\rho_{\rm bus}$ and $\rho_{\rm car}$ for the time interval $\Delta$ between two successive buses and cars, respectively? What are the means of these distributions?* (Hint: To answer this for the bus, you'll need to use the Dirac $\delta$-function,[5] which is zero except at zero and infinite at zero, with integral equal to one: $\int_a^c f(x)\delta(x-b)\,dx = f(b)$.)

*(d) If another observer arrives at the road at a randomly chosen time, what is the probability distribution for the time $\Delta$ she has to wait for the first bus to arrive? What is the probability distribution for the time she has to wait for the first car to pass by?* (Hint: What would the distribution of waiting times be just after a car passes by? Does the time of the next car depend at all on the previous car?) *What are the means of these distributions?*

The mean time between cars is 5 minutes. The mean time to the next car should be 5 minutes. A little thought should convince you that the mean time since the last car should also be 5 minutes. But $5 + 5 \neq 5$: how can this be?

The same physical quantity can have different means when averaged in different ensembles! The mean time between cars in part (c) was a gap average: it weighted all gaps between cars equally. The mean time to the next car from part (d) was a time average: the second observer arrives with equal probability at every time, so is twice as likely to arrive during a gap between cars that is twice as long.

*(e) In part (c), $\rho_{\rm car}^{\rm gap}(\Delta)$ was the probability that a randomly chosen gap was of length $\Delta$. Write a formula for $\rho_{\rm car}^{\rm time}(\Delta)$, the probability that the second observer, arriving at a randomly chosen time, will be in a gap between cars of length $\Delta$.* (Hint: Make sure it's normalized.) *From $\rho_{\rm car}^{\rm time}(\Delta)$, calculate the average length of the gaps between cars, using the time–weighted average measured by the second observer.*

(1.4) **Stirling's Approximation and Asymptotic Series.** (Mathematics)

One important approximation useful in statistical mechanics is Stirling's approximation [102] for $n!$, valid for large $n$. It's not a traditional Taylor series: rather, it's an *asymptotic series*. Stirling's formula is extremely useful in this course, and asymptotic series are important in many fields of applied mathematics, statistical mechanics [100], and field theory [101], so let's investigate them in detail.

*(a) Show, by converting the sum to an integral, that $\log(n!) \sim (n + \frac{1}{2})\log(n + \frac{1}{2}) - n - \frac{1}{2}\log(\frac{1}{2})$, where (as always in this book) $\log$ represents the natural logarithm, not $\log_{10}$. Show that this is compatible with the more precise and traditional formula $n! \approx (n/e)^n\sqrt{2\pi n}$; in particular, show that the difference of the logs goes to a constant as $n \to \infty$. Show that the latter is compatible with the first term in the series we use below, $n! \sim (2\pi/(n+1))^{\frac{1}{2}} e^{-(n+1)}(n+1)^{n+1}$, in that the difference of the logs goes to zero as $n \to \infty$.* Related formulæ: $\int \log x\,dx = x\log x - x$, and $\log(n+1) - \log(n) = \log(1 + 1/n) \sim 1/n$ up to terms of order $1/n^2$.

We want to expand this function for large $n$: to do this, we need to turn it into a continuous function, interpolating between the integers. This continuous function, with its argument perversely shifted by one, is $\Gamma(z) = (z-1)!$. There are many equivalent formulas for $\Gamma(z)$: indeed, any formula giving an analytic function satisfying the recursion relation $\Gamma(z+1) = z\Gamma(z)$ and the normalization $\Gamma(1) = 1$ is equivalent (by theorems of complex analysis). We won't use it here, but a typical definition is $\Gamma(z) = \int_0^\infty e^{-t}t^{z-1}dt$: one can integrate by parts to show that $\Gamma(z+1) = z\Gamma(z)$.

*(b) Show, using the recursion relation $\Gamma(z+1) = z\Gamma(z)$, that $\Gamma(z)$ is infinite (has a pole) at all the negative integers.*

Stirling's formula is extensible [9, p.218] into a nice expansion of $\Gamma(z)$ in powers of $1/z = z^{-1}$:

$$\begin{aligned}\Gamma[z] &= (z-1)! \qquad (1.3)\\ &\sim (2\pi/z)^{\frac{1}{2}} e^{-z} z^z (1 + (1/12)z^{-1}\\ &\quad + (1/288)z^{-2} - (139/51840)z^{-3}\\ &\quad - (571/2488320)z^{-4} + (163879/209018880)z^{-5}\\ &\quad + (5246819/75246796800)z^{-6}\\ &\quad - (534703531/902961561600)z^{-7}\\ &\quad - (4483131259/86684309913600)z^{-8} + \ldots)\end{aligned}$$

[5] Mathematically, this isn't a function, but rather a distribution or a measure.

This looks like a Taylor series in $1/z$, but is subtly different. For example, we might ask what the radius of convergence [104] of this series is. The radius of convergence is the distance to the nearest singularity in the complex plane.

*(c) Let $g(\zeta) = \Gamma(1/\zeta)$; then Stirling's formula is some stuff times a Taylor series in $\zeta$. Plot the poles of $g(\zeta)$ in the complex $\zeta$ plane. Show, that the radius of convergence of Stirling's formula applied to g must be zero, and hence no matter how large z is, Stirling's formula eventually diverges.*

Indeed, the coefficient of $z^{-j}$ eventually grows rapidly; Bender and Orszag [9, p.218] show that the odd coefficients ($A_1 = 1/12$, $A_3 = -139/51840\ldots$) asymptotically grow as

$$A_{2j+1} \sim (-1)^j 2(2j)!/(2\pi)^{2(j+1)}. \tag{1.4}$$

*(d) Show explicitly, using the ratio test applied to formula 1.4, that the radius of convergence of Stirling's formula is indeed zero.*[6]

This in no way implies that Stirling's formula isn't valuable! An asymptotic series of length $n$ approaches $f(z)$ as $z$ gets big, but for fixed $z$ it can diverge as $n$ gets larger and larger. In fact, asymptotic series are very common, and often are useful for much larger regions than are Taylor series.

*(e) What is 0!? Compute 0! using successive terms in Stirling's formula (summing to $A_N$ for the first few N.) Considering that this formula is expanding about infinity, it does pretty well!*

Quantum electrodynamics these days produces the most precise predictions in science. Physicists sum enormous numbers of Feynman diagrams to produce predictions of fundamental quantum phenomena. Dyson argued that quantum electrodynamics calculations give an asymptotic series [101]; the most precise calculation in science takes the form of a series which cannot converge!

(1.5) **Random Matrix Theory.** (Math, Quantum) (With Brouwer. [14])

One of the most active and unusual applications of ensembles is *random matrix theory*, used to describe phenomena in nuclear physics, mesoscopic quantum mechanics, and wave phenomena. Random matrix theory was invented in a bold attempt to describe the statistics of energy level spectra in nuclei. In many cases, the statistical behavior of systems exhibiting complex wave phenomena – almost any correlations involving eigenvalues and eigenstates – can be quantitatively modeled using ensembles of matrices with completely random, uncorrelated entries!

To do this exercise, you'll need to find a software environment in which it is easy to (i) make histograms and plot functions on the same graph, (ii) find eigenvalues of matrices, sort them, and collect the differences between neighboring ones, and (iii) generate symmetric random matrices with Gaussian and integer entries. Mathematica, Matlab, Octave, and Python are all good choices. For those who are not familiar with one of these packages, I will post hints on how to do these three things on the Random Matrix Theory site in the computer exercise section on the book Web site [108].

The most commonly explored ensemble of matrices is the Gaussian Orthogonal Ensemble. Generating a member $H$ of this ensemble of size $N \times N$ takes two steps:

- Generate a $N \times N$ matrix whose elements are random numbers with Gaussian distributions of mean zero and standard deviation $\sigma = 1$.
- Add each matrix to its transpose to symmetrize it.

As a reminder, the Gaussian or normal probability distribution gives a random number $x$ with probability

$$\rho(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/2\sigma^2}. \tag{1.5}$$

One of the most striking properties that large random matrices share is the distribution of level splittings.

*(a) Generate an ensemble with $M = 1000$ or so GOE matrices of size $N = 2$, 4, and 10. (More is nice.) Find the eigenvalues $\lambda_n$ of each matrix, sorted in increasing order. Find the difference between neighboring eigenvalues $\lambda_{n+1} - \lambda_n$, for n, say, equal to*[7] *$N/2$. Plot a histogram of these eigenvalue splittings divided by the mean splitting, with bin–size small enough to see some of the fluctuations. (Hint: debug your work with $M = 10$, and then change to $M = 1000$.)*

What is this dip in the eigenvalue probability near zero? It's called *level repulsion*.

[6] If you don't remember about radius of convergence, see [104]. Here you'll be using every other term in the series, so the radius of convergence is $\sqrt{|A_{2j-1}/A_{2j+1}|}$.

[7] In the experiments, they typically plot all the eigenvalue splittings. Since the mean splitting between eigenvalues will change slowly, this smears the distributions a bit. So, for example, the splittings between the largest and second–largest eigenvalues will be typically rather larger for the GOE ensemble than for pairs near the middle. If you confine your plots to a small range near the middle, the smearing would be small, but it's so fast to calculate new ones we just keep one pair.

For $N = 2$ the probability distribution for the eigenvalue splitting can be calculated pretty simply. Let our matrix be $M = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$.

*(b) Show that the eigenvalue difference for $M$ is $\lambda = \sqrt{(c-a)^2 + 4b^2} = 2\sqrt{d^2 + b^2}$ where $d = (c-a)/2$.*[8] *If the probability distribution of matrices $\rho_M(d, b)$ is continuous and finite at $d = b = 0$, argue that the probability density $\rho(\lambda)$ of finding an energy level splitting near zero vanishes at $\lambda = 0$, giving us level repulsion. (Both d and b must vanish to make $\lambda = 0$.)* (Hint: go to polar coordinates, with $\lambda$ the radius.)

*(c) Calculate analytically the standard deviation of a diagonal and an off-diagonal element of the GOE ensemble (made by symmetrizing Gaussian random matrices with $\sigma = 1$).* You may want to check your answer by plotting your predicted Gaussians over the histogram of $H_{11}$ and $H_{12}$ from your ensemble in part (a). *Calculate analytically the standard deviation of $d = (c-a)/2$ of the $N = 2$ GOE ensemble of part (b), and show that it equals the standard deviation of b.*

*(d) Calculate a formula for the probability distribution of eigenvalue spacings for the $N = 2$ GOE, by integrating over the probability density $\rho_M(d, b)$.* (Hint: polar coordinates again.)

If you rescale the eigenvalue splitting distribution you found in part (d) to make the mean splitting equal to one, you should find the distribution

$$\rho_{\text{Wigner}}(s) = \frac{\pi s}{2} e^{-\pi s^2/4}. \tag{1.6}$$

This is called the *Wigner surmise*: it is within 2% of the correct answer for larger matrices as well.[9]

*(e) Plot equation 1.6 along with your $N = 2$ results from part (a). Plot the Wigner surmise formula against the plots for $N = 4$ and $N = 10$ as well.*

Let's define a $\pm 1$ ensemble of real symmetric matrices, by generating a $N \times N$ matrix whose elements are independent random variables each $\pm 1$ with equal probability.

*(f) Generate an ensemble with $M = 1000$ $\pm 1$ symmetric matrices with size $N = 2$, 4, and 10. Plot the eigenvalue distributions as in part (a). Are they universal (independent of the ensemble up to the mean spacing) for $N = 2$ and 4? Do they appear to be nearly universal*[10] *(the same as for the GOE in part (a)) for $N = 10$? Plot the Wigner surmise along with your histogram for $N = 10$.*

The GOE ensemble has some nice statistical properties. The ensemble is invariant under orthogonal transformations

$$H \to R^T H R \qquad \textit{with} \quad R^T = R^{-1}. \tag{1.7}$$

*(g) Show that $\mathrm{Tr}[H^T H]$ is the sum of the squares of all elements of $H$. Show that this trace is invariant under orthogonal coordinate transformations (that is, $H \to R^T H R$ with $R^T = R^{-1}$).* (Hint: Remember, or derive, the cyclic invariance of the trace: $\mathrm{Tr}[ABC] = \mathrm{Tr}[CAB]$.)

Note that this trace, for a symmetric matrix, is the sum of the squares of the diagonal elements plus *twice* the squares of the upper triangle of off–diagonal elements. That is convenient, because in our GOE ensemble the variance (squared standard deviation) of the off–diagonal elements is half that of the diagonal elements.

*(h) Write the probability density $\rho(H)$ for finding GOE ensemble member $H$ in terms of the trace formula in part (g). Argue, using your formula and the invariance from part (g), that the GOE ensemble is invariant under orthogonal transformations: $\rho(R^T H R) = \rho(H)$.*

This is our first example of an *emergent symmetry*. Many different ensembles of symmetric matrices, as the size $N$ goes to infinity, have eigenvalue and eigenvector distributions that are invariant under orthogonal transformations *even though the original matrix ensemble did not have this symmetry*. Similarly, rotational symmetry emerges in random walks on the square lattice as the number of steps $N$ goes to infinity, and also emerges on long length scales for Ising models at their critical temperatures.[11]

[8] Note that the eigenvalue difference doesn't depend on the trace of $M$, $a + c$, only on the difference $c - a = 2d$.

[9] The distribution for large matrices is known and universal, but is much more complicated to calculate.

[10] Note the spike at zero. There is a small probability that two rows or columns of our matrix of $\pm 1$ will be the same, but this probability vanishes rapidly for large $N$.

[11] A more exotic emergent symmetry underlies Fermi liquid theory: the effective interactions between electrons disappear near the Fermi energy: the fixed point has an emergent gauge symmetry.

# Random Walks and Emergent Properties

2

What makes physics possible? Why are humans able to find simple mathematical laws that describe the real world? Our physical laws are not direct statements about the underlying reality of the universe. Rather, our laws emerge out of far more complex microscopic behavior.[1] Statistical mechanics provides powerful tools for understanding simple behavior that emerges from underlying complexity.

In this chapter, we will explore the emergent behavior for *random walks*. Random walks are paths that take successive steps in random directions. They arise often in statistical mechanics: as partial sums of fluctuating quantities, as trajectories of particles undergoing repeated collisions, and as the shapes for long, linked systems like polymers. They have two kinds of emergent behavior. First, an individual random walk, after a large number of steps, becomes fractal or *scale invariant* (section 2.1). Secondly, the endpoint of the random walk has a probability distribution that obeys a simple continuum law: the diffusion equation (section 2.2). Both of these behaviors are largely independent of the microscopic details of the walk: they are *universal*. Random walks in an external field (section 2.3) provide our first examples of conserved currents, linear response, and Boltzmann distributions. Finally we use the diffusion equation to introduce Fourier and Greens function solution techniques (section 2.4). Random walks encapsulate many of the themes and methods of statistical mechanics.

[1] You may think that Newton's law of gravitation, or Einstein's refinement to it, is more fundamental than the diffusion equation. You would be correct: gravitation applies to everything. But the simple macroscopic law of gravitation emerges, from a quantum exchange of immense numbers of virtual gravitons, just as the diffusion equation emerges from large numbers of long random walks. The diffusion equation and other continuum statistical mechanics laws are special to particular systems, but they emerge from the microscopic theory in much the same way as gravitation and the other fundamental laws of nature do.

## 2.1 Random Walk Examples: Universality and Scale Invariance

We illustrate random walks with three examples: coin flips, the drunkard's walk, and polymers.

**Coin Flips.** Statistical mechanics often demands sums or averages of a series of fluctuating quantities: $s_N = \sum_{i=1}^{N} \ell_i$. The energy of a material is a sum over the energies of the molecules composing the material; your grade on a statistical mechanics exam is the sum of the scores on many individual questions. Imagine adding up this sum one term at a time: the path $s_1, s_2, \ldots$ forms an example of a one-dimensional random walk.

For example, consider flipping a coin, recording the difference $s_N = h_N - t_N$ between the number of heads and tails found. Each coin flip

contributes $\ell_i = \pm 1$ to the total. How big a sum $s_N = \sum_{i=1}^{N} \ell_i =$ (heads − tails) do you expect after $N$ flips? The average of $s_N$ is of course zero, because positive and negative steps are equally likely. A better measure of the characteristic distance moved is the root–mean–square (RMS) number[2] $\sqrt{\langle s_N^2 \rangle}$. After one coin flip,

$$\langle s_1^2 \rangle = 1 = \tfrac{1}{2}(-1)^2 + \tfrac{1}{2}(1)^2; \tag{2.1}$$

after two and three coin flips

$$\begin{aligned} \langle s_2^2 \rangle &= 2 = \tfrac{1}{4}(-2)^2 + \tfrac{1}{2}(0)^2 + \tfrac{1}{4}(2)^2; \\ \langle s_3^2 \rangle &= 3 = \tfrac{1}{8}(-3)^2 + \tfrac{3}{8}(-1)^2 + \tfrac{3}{8}(1)^2 + \tfrac{1}{8}(3)^2. \end{aligned} \tag{2.2}$$

Does this pattern continue? Because $\ell_N = \pm 1$ with equal probability independent of the history, $\langle \ell_N s_{N-1} \rangle = \frac{1}{2}\langle (+1)s_{N-1} \rangle + \frac{1}{2}\langle (-1)s_{N-1} \rangle = 0$. We know $\langle \ell_N^2 \rangle = 1$; if we assume $\langle s_{N-1}^2 \rangle = N - 1$ we can prove by induction on $N$ that

$$\begin{aligned} \langle s_N^2 \rangle &= \langle (s_{N-1} + \ell_N)^2 \rangle = \langle s_{N-1}^2 \rangle + \cancel{2\langle s_{N-1} \ell_N \rangle} + \langle \ell_N^2 \rangle \\ &= \langle s_{N-1}^2 \rangle + 1 = N. \end{aligned} \tag{2.3}$$

Hence the RMS average of (heads-tails) for $N$ coin flips,

$$\sigma_s = \sqrt{\langle s_N^2 \rangle} = \sqrt{N}. \tag{2.4}$$

Notice that we chose to count the difference between the number of heads and tails. Had we instead just counted the number of heads $h_N$, then $\langle h_N \rangle$ would grow proportionately to $N$: $\langle h_N \rangle = N/2$. We would then be interested in the fluctuations of $h_N$ about $N/2$, measured most easily by squaring the difference between the particular random walks and the average random walk: $\sigma_h^2 = \langle (h_N - \langle h_N \rangle)^2 \rangle = N/4$.[3] The variable $\sigma_h$ is the *standard deviation* of the sum $h_N$: this is an example of the typical behavior that the standard deviation of the sum of $N$ random variables grows proportionally to $\sqrt{N}$.

The sum, of course, grows linearly with $N$, so (if the average isn't zero) the fluctuations become tiny in comparison to the sum. This is why experimentalists often make repeated measurements of the same quantity and take the mean. Suppose we were to measure the mean number of heads per coin toss, $a_N = h_N/N$. We see immediately that the fluctuations in $a_N$ will also be divided by $N$, so

$$\sigma_a = \sigma_h/N = 1/(2\sqrt{N}). \tag{2.5}$$

*The standard deviation of the mean of $N$ measurements is proportional to $1/\sqrt{N}$.*

**Drunkard's Walk.** Random walks in higher dimensions arise as trajectories that undergo successive random collisions or turns: for example, the trajectory of a perfume molecule in a sample of air.[4] Because

[2] We use angle brackets $\langle X \rangle$ to denote averages over various ensembles: we'll add subscripts to the brackets where there may be confusion about which ensemble we are using. Here our ensemble contains all $2^N$ possible sequences of $N$ coin flips.

[3] It's $N/4$ for $h$ instead of $N$ for $s$ because each step changes $s_N$ by $\pm 2$, and $h_N$ only by $\pm 1$: the standard deviation $\sigma$ is in general proportional to the step size.

[4] Real perfume in a real room will primarily be transported by convection; in liquids and gases, diffusion dominates usually only on short length scales. Solids don't convect, so thermal or electrical conductivity would be more accurate – but less vivid – applications for random walks.

the air is dilute and the interactions are short-ranged, the molecule will basically travel in straight lines, with sharp changes in velocity during infrequent collisions. After a few substantial collisions, the molecule's velocity will be uncorrelated with its original velocity. The path taken by the molecule will be a jagged, random walk through three dimensions.

The random walk of a perfume molecule involves random directions, random velocities, and random step sizes. It's more convenient to study steps at regular time intervals, so we'll instead consider the classic problem of a drunkard's walk. The drunkard is presumed to start at a lamppost at $x = y = 0$. He takes steps $\boldsymbol{\ell}_N$ each of length $L$, at regular time intervals. Because he's drunk, the steps are in completely random directions, each uncorrelated with the previous steps. This lack of correlation says that the average dot product between any two steps $\boldsymbol{\ell}_m$ and $\boldsymbol{\ell}_n$ is zero, since all relative angles $\theta$ between the two directions are equally likely: $\langle \boldsymbol{\ell}_m \cdot \boldsymbol{\ell}_n \rangle = L^2 \langle \cos(\theta) \rangle = 0$.[5] This implies that the dot product of $\boldsymbol{\ell}_N$ with $\mathbf{s}_{N-1} = \sum_{m=1}^{N-1} \boldsymbol{\ell}_m$ is zero. Again, we can use this to work by induction:

$$\begin{aligned} \langle \mathbf{s}_N^2 \rangle &= \langle (\mathbf{s}_{N-1} + \boldsymbol{\ell}_N)^2 \rangle = \langle \mathbf{s}_{N-1}^2 \rangle + \langle 2\mathbf{s}_{N-1} \cdot \boldsymbol{\ell}_N \rangle + \langle \boldsymbol{\ell}_N^2 \rangle \\ &= \langle \mathbf{s}_{N-1}^2 \rangle + L^2 = \cdots = NL^2, \end{aligned} \tag{2.6}$$

so the RMS distance moved is $\sqrt{N}L$.

Random walks introduce us to the concepts of *scale invariance* and *universality*.

**Scale Invariance.** What kind of path only goes $\sqrt{N}$ total distance in $N$ steps? Random walks form paths which look jagged and scrambled. Indeed, they are so jagged that if you blow up a small corner of one, the blown up version looks just as jagged (figure 2.2). Clearly each of the blown-up random walks is different, just as any two random walks of the same length are different, but the ensemble of random walks of length $N$ looks much like that of length $N/4$, until $N$ becomes small enough that the individual steps can be distinguished. Random walks are *scale invariant*: they look the same on all scales.[6]

**Universality.** On scales where the individual steps are not distinguishable (and any correlations between steps is likewise too small to see) we find that all random walks look the same. Figure 2.2 depicts a drunkard's walk, but any two–dimensional random walk would give the same behavior (statistically). Coin tosses of two coins (penny sums along $x$, dime sums along $y$) would produce, statistically, the same random walk ensemble on lengths large compared to the step sizes. In three dimensions, photons[7] in the Sun (exercise 2.2) or in a glass of milk undergo a random walk with fixed speed $c$ between collisions. Nonetheless, after a few steps their random walks are statistically indistinguishable from that of our variable–speed perfume molecule. This independence of the behavior on the microscopic details is called *universality*.

Random walks are simple enough that we could probably show that each individual case behaves like the others. In section 2.2 we will generalize our argument that the RMS distance scales as $\sqrt{N}$ to simultaneously cover both coin flips and drunkards; with more work we could

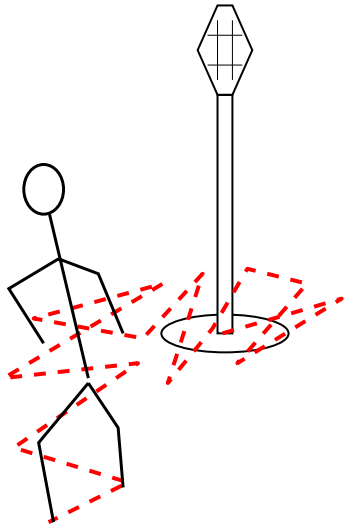


**Fig. 2.1** The drunkard takes a series of steps of length $L$ away from the lamppost, but each with a random angle.

[5] More generally, if two variables are uncorrelated then the average of their product is the product of their averages: in this case this would imply $\langle \boldsymbol{\ell}_m \cdot \boldsymbol{\ell}_n \rangle = \langle \boldsymbol{\ell}_m \rangle \cdot \langle \boldsymbol{\ell}_n \rangle = \mathbf{0} \cdot \mathbf{0} = 0$.

[6] They are also fractal with dimension two, in all spatial dimensions larger than two. This just reflects the fact that a random walk of 'volume' $V = N$ steps roughly fits into a radius $R \sim s_N \sim N^{1/2}$. The fractal dimension $D$ of the set, defined by $R^D = V$, is thus two.

[7] A photon is a quantum of light or other electromagnetic radiation.

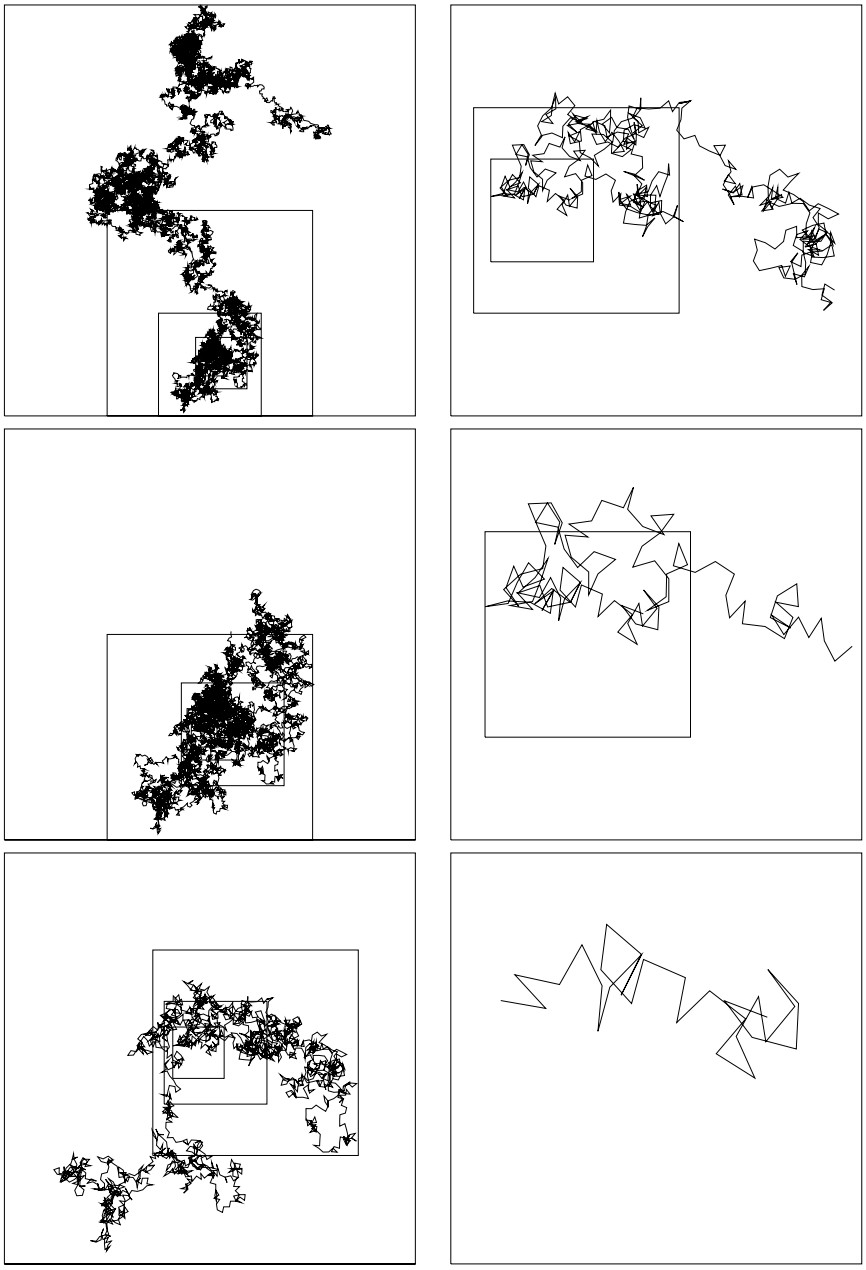


**Fig. 2.2 Random Walk: Scale Invariance** Random walks form a jagged, fractal pattern which looks the same when rescaled. Here each succeeding walk is the first quarter of the previous walk, magnified by a factor of two; the shortest random walk is of length 31, the longest of length 32,000 steps. The left side of figure 1.1 is the further evolution of this walk to 128,000 steps.

include variable times between collisions and local correlations to cover the cases of photons and molecules in a gas. We could probably also calculate properties about the jaggedness of paths in these systems, and show that they too agree with one another after many steps. Instead, we'll wait for chapter 12 (and specifically exercise 12.7), where we will give a deep but intuitive explanation of why each of these problems is scale invariant, and why all of these problems share the same behavior on long length scales. Universality and scale invariance will be explained there using renormalization–group methods, originally developed to study continuous phase transitions.

**Polymers.** Finally, random walks arise as the shapes for *polymers*. Polymers are long molecules (like DNA, RNA, proteins, and many plastics) made up of many small units (called monomers) attached to one another in a long chain. Temperature can introduce fluctuations in the angle between two adjacent monomers; if these fluctuations dominate over the energy,[8] the polymer shape can form a random walk. Here the steps are not increasing with time, but with monomers (or groups of monomers) along the chain.

The random walks formed by polymers are *not* the same as those in our first two examples: they are in a different universality class. This is because the polymer cannot intersect itself: a walk that would cause two monomers to overlap is not allowed. Polymers undergo *self-avoiding* random walks. In two and three dimensions, it turns out that the effects of these self–intersections is not a small, microscopic detail, but changes the properties of the random walk in an essential way.[9] One can show that these intersections will often arise on far–separated regions of the polymer, and that in particular they change the dependence of squared radius $\langle s_N^2 \rangle$ on the number of segments $N$ (exercise 2.8). In particular, they change the power law $\sqrt{\langle s_N^2 \rangle} \sim N^\nu$ from the ordinary random walk value $\nu = 1/2$ to a higher value, $\nu = 3/4$ in two dimensions and $\nu \approx 0.59$ in three dimensions. Power laws are central to the study of scale–invariant systems: $\nu$ is our first example of a *universal critical exponent*.

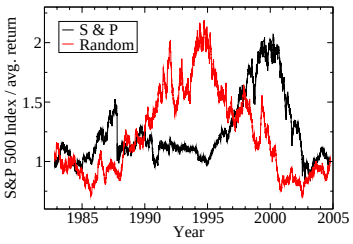



**Fig. 2.3 S&P 500, normalized.** Standard and Poor's 500 stock index daily closing price since its inception, corrected for inflation, divided by the average 6.4% return over this time period. Stock prices are often modeled as a biased random walk. Notice that the fluctuations (risk) in individual stock prices will typically be much higher. By averaging over 500 stocks, the random fluctuations in this index are reduced, while the average return remains the same: see [67] and [68]. For comparison, a one-dimensional multiplicative random walk is also shown.

[8] Plastics at low temperature can be crystals; functional proteins and RNA often packed tightly into well–defined shapes. Molten plastics and denatured proteins form self–avoiding random walks. Double–stranded DNA is rather stiff: the step size for the random walk is many nucleic acids long.

[9] Self–avoidance is said to be a *relevant* perturbation that changes the *universality class*. In (unphysical) spatial dimensions higher than four, self–avoidance is irrelevant: hypothetical hyper–polymers in five dimensions would look like regular random walks on long length scales.

## 2.2 The Diffusion Equation

In the *continuum limit* of long length and time scales, simple behavior emerges from the ensemble of irregular, jagged random walks: their evolution is described by the *diffusion equation*:[10]

$$\frac{\partial \rho}{\partial t} = D \nabla^2 \rho = D \frac{\partial^2 \rho}{\partial x^2}. \tag{2.7}$$

The diffusion equation can describe the evolving density $\rho(x,t)$ of a local cloud of perfume as the molecules random–walk through collisions with the air molecules. Alternatively, it can describe the probability density of an individual particle as it random walks through space: if the particles are non-interacting, the probability distribution of one particle describes the density of all particles.

[10] In the remainder of this chapter we specialize for simplicity to one dimension. We also change variables from the sum $s$ to position $x$.

Consider a general, uncorrelated random walk where at each time step $\Delta t$ the particle's position $x$ changes by a step $\ell$:

$$x(t+\Delta t) = x(t) + \ell(t). \tag{2.8}$$

Let the probability distribution for each step be $\chi(\ell)$.[11] We'll assume that $\chi$ has mean zero and standard deviation $a$, so the first few moments of $\chi$ are

$$\begin{aligned} \int \chi(z)\,dz &= 1, \\ \int z\chi(z)\,dz &= 0, \text{ and} \\ \int z^2\chi(z)\,dz &= a^2. \end{aligned} \tag{2.9}$$

What is the probability distribution for $\rho(x, t+\Delta t)$, given the probability distribution $\rho(x', t)$?

Clearly, for the particle to go from $x'$ at time $t$ to $x$ at time $t+\Delta t$, the step $\ell(t)$ must be $x - x'$. This happens with probability $\chi(x-x')$ times the probability density $\rho(x', t)$ that it started at $x'$. Integrating over original positions $x'$, we have

$$\begin{aligned} \rho(x, t+\Delta t) &= \int_{-\infty}^{\infty} \rho(x', t)\chi(x-x')\,dx' \\ &= \int_{-\infty}^{\infty} \rho(x-z, t)\chi(z)\,dz \end{aligned} \tag{2.10}$$

where we change variables to $z = x - x'$.[12]

Now, suppose $\rho$ is broad: the step size is very small compared to the scales on which $\rho$ varies (figure 2.4). We may then do a Taylor expansion of 2.10 in $z$:

$$\begin{aligned} \rho(x, t+\Delta t) &\approx \int \left[\rho(x,t) - z\frac{\partial \rho}{\partial x} + \frac{z^2}{2}\frac{\partial^2 \rho}{\partial x^2}\right] \chi(z)\,dz \\ &= \rho(x,t)\cancelto{1}{\int \chi(z)\,dz} - \frac{\partial \rho}{\partial x}\cancelto{0}{\int z\chi(z)\,dz} + \tfrac{1}{2}\frac{\partial^2 \rho}{\partial x^2}\int z^2\chi(z)\,dz. \\ &= \rho(x,t) + \tfrac{1}{2}\frac{\partial^2 \rho}{\partial x^2}a^2 \end{aligned} \tag{2.11}$$

using the moments of $\chi$ in 2.9. Now, if we also assume that $\rho$ is slow, so that it changes only slightly during this time step, we can approximate $\rho(x, t+\Delta t) - \rho(x,t) \approx \frac{\partial \rho}{\partial t}\Delta t$, and we find

$$\frac{\partial \rho}{\partial t} = \frac{a^2}{2\Delta t}\frac{\partial^2 \rho}{\partial x^2}. \tag{2.12}$$

This is the diffusion equation[13] (2.7), with

$$D = a^2/2\Delta t. \tag{2.13}$$

The diffusion equation applies to all random walks, so long as the probability distribution is broad and slow compared to the individual steps.

[11] In our two examples the distribution $\chi(\ell)$ was discrete: we can write it using the Dirac $\delta$-function. (The $\delta$ function $\delta(x-x_0)$ is a probability density which has 100% chance of finding the particle in any box containing $x_0$: thus $\delta(x-x_0)$ is zero unless $x = x_0$, and $\int f(x)\delta(x-x_0)dx = f(x_0)$ so long as the domain of integration includes $x_0$.) In the case of coin flips, a 50/50 chance of $\ell = \pm 1$ can be written as $\chi(\ell) = \frac{1}{2}\delta(\ell+1) + \frac{1}{2}\delta(\ell-1)$. In the case of the drunkard, $\chi(\boldsymbol{\ell}) = \delta(|\boldsymbol{\ell}| - L)/(2\pi L)$, evenly spaced around the circle.

[12] Notice that although $dz = -dx'$, the limits of integration $\int_{-\infty}^{\infty} \to \int_{\infty}^{-\infty} = -\int_{-\infty}^{\infty}$, canceling the minus sign. This happens often in calculations: watch out for it.


**Fig. 2.4** We suppose the step sizes $\ell$ are small compared to the broad ranges on which $\rho(x)$ varies, so we may do a Taylor expansion in gradients of $\rho$.

[13] One can understand this intuitively. Random walks and diffusion tend to even out the hills and valleys in the density. Hills have negative second derivatives $\frac{\partial^2 \rho}{\partial x^2} < 0$ and should flatten $\frac{\partial \rho}{\partial t} < 0$, valleys have positive second derivatives and fill up.

## 2.3 Currents and External Forces.

As the particles in our random walks move around, they never are created or destroyed: they are *conserved*.[14] If $\rho(x)$ is the density of a conserved quantity, we may write its evolution law (see figure 2.5) in terms of the current $J(x)$ passing a given point $x$:

$$\frac{\partial \rho}{\partial t} = -\frac{\partial J}{\partial x}. \tag{2.14}$$

Here the current $J$ is the amount of stuff flowing to the right through the point $x$; since the stuff is conserved, the only way the density can change is by flowing from one place to another. From equation 2.7 and equation 2.14, the current for the diffusion equation is

$$J_{\text{diffusion}} = -D\frac{\partial \rho}{\partial x}; \tag{2.15}$$

particles diffuse (random–walk) on average from regions of high density towards regions of low density.

In many applications one has an average drift term along with a random walk. In some cases (like the total grade in a multiple-choice test, exercise 2.1) there is naturally a non-zero mean for each step in the random walk. In other cases, there is an external force $F$ that is biasing the steps to one side: the mean net drift is $F\,\Delta t$ times a mobility $\gamma$:

$$x(t + \Delta t) = x(t) + F\gamma\Delta t + \ell(t). \tag{2.16}$$

We can derive formulas for this mobility given a microscopic model. On the one hand, if our air is dilute and the diffusing molecule is small, we can model the trajectory as free acceleration between collisions separated by $\Delta t$, and we can assume the collisions completely scramble the velocities. In this case, the net motion due to the external force is half the acceleration $F/m$ times the time squared: $\frac{1}{2}(F/m)(\Delta t)^2 = F\,\Delta t\frac{\Delta t}{2m}$ so $\gamma = \frac{\Delta t}{2m}$ Using equation 2.13, we find

$$\gamma = \frac{\Delta t}{2m}\left(D\frac{2\Delta t}{a^2}\right) = \frac{D}{m(a/\Delta t)^2} = \frac{D}{m\bar{v}^2} \tag{2.17}$$

where $\bar{v} = a/\Delta t$ is the velocity of the unbiased random walk step.

On the other hand, if our air is dense and the diffusing molecule is large, we might treat the air as a viscous fluid of kinematic viscosity $\eta$; if we also simply model the molecule as a sphere of radius $r$, a fluid mechanics calculation tells us that the mobility is $\gamma = 1/(6\pi\eta r)$.

Starting from equation 2.16, we can repeat our analysis of the continuum limit (equations 2.10 through 2.12) to derive the diffusion equation

[14] More subtly, the probability density $\rho(x)$ of a single particle undergoing a random walk is also conserved: like particle density, probability density cannot be created or destroyed, it can only slosh around.

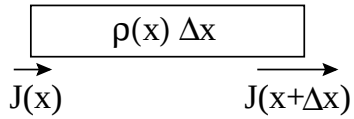



**Fig. 2.5** Let $\rho(x,t)$ be the density of some conserved quantity (# of molecules, mass, energy, probability, etc.) varying in one spatial dimension $x$, and $J(x)$ be the rate at which $\rho$ is passing a point $x$. The the amount of $\rho$ in a small region $(x, x+\Delta x)$ is $n = \rho(x)\,\Delta x$. The flow of particles into this region from the left is $J(x)$ and the flow out is $J(x+\Delta x)$, so $\frac{\partial n}{\partial t} = J(x) - J(x+\Delta x) \approx \frac{\partial \rho}{\partial t}\Delta x$, and we derive the conserved current relation $\frac{\partial \rho}{\partial t} = -\frac{J(x+\Delta x) - J(x)}{\Delta x} = -\frac{\partial J}{\partial x}$.

in an external force,[15]

$$J = \gamma F \rho - D \frac{\partial \rho}{\partial x} \tag{2.18}$$

$$\frac{\partial \rho}{\partial t} = -\gamma F \frac{\partial \rho}{\partial x} + D \frac{\partial^2 \rho}{\partial x^2}. \tag{2.19}$$

The sign of the new term can be explained intuitively: if $\rho$ is increasing in space (positive slope $\frac{\partial \rho}{\partial x}$) and the force is dragging the particles forward, then $\rho$ will decrease with time because the high-density regions ahead of $x$ are receding and the low density regions behind $x$ are moving in.

The diffusion equation describes how systems of random–walking particles approach equilibrium (see chapter 3). The diffusion equation in the absence of external force describes the evolution of perfume density in a room. A time–independent equilibrium state $\rho^*$ obeying the diffusion equation 2.7 must have $\partial^2 \rho^* / \partial x^2 = 0$, so $\rho^*(x) = \rho_0 + Bx$. If the perfume cannot penetrate the walls, $\frac{\partial \rho^*}{\partial x} = 0$ at the boundaries so $B = 0$. Thus, as one might expect, the perfume evolves to a rather featureless equilibrium state $\rho^*(x) = \rho_0$, evenly distributed throughout the room.

In the presence of a constant external force (like gravitation) the equilibrium state is more interesting. Let $x$ be the height above the ground, and $F = -mg$ be the force due to gravity. By equation 2.19, the equilibrium state $\rho^*$ satisfies

$$0 = \frac{\partial \rho^*}{\partial t} = \gamma m g \frac{\partial \rho^*}{\partial x} + D \frac{\partial^2 \rho^*}{\partial x^2} \tag{2.20}$$

which has general solution $\rho^*(x) = A \exp(-\frac{\gamma}{D} m g x) + B$. We assume that the density of perfume $B$ in outer space is zero,[16] so the density of perfume decreases exponentially with height:

$$\rho^*(x) = A \exp(-\frac{\gamma}{D} m g x). \tag{2.21}$$

The perfume molecules are pulled downward by the gravitational force, and remain aloft only because of the random walk. If we generalize from perfume to oxygen molecules (and ignore temperature gradients and weather) this gives the basic explanation for why it becomes harder to breath as one climbs mountains.[17]

[16] Non-zero $B$ would correspond to a constant-density rain of perfume.

[15] Warning: if the force is not constant in space, the evolution also depends on the gradient of the force: $\frac{\partial \rho}{\partial t} = -\frac{\partial J}{\partial x} = -\gamma \frac{\partial F(x)\rho(x)}{\partial x} + D \frac{\partial^2 \rho}{\partial x^2} = -\gamma \rho \frac{\partial F}{\partial x} - \gamma F \frac{\partial \rho}{\partial x} + D \frac{\partial^2 \rho}{\partial x^2}$. Similar problems can arise if the diffusion constant is density dependent. *When working with a conserved property, write your equations first in terms of the current, to guarantee that it is conserved.* $J = -D(\rho, \mathbf{x}) \nabla \rho + \gamma(\mathbf{x}) F(\mathbf{x}) \rho(\mathbf{x})$ The author has observed himself and a variety of graduate students wasting up to a week at a time when this rule is forgotten.

[17] In chapter 6 we shall derive the Boltzmann distribution, implying that the probability of having energy $mgh = E$ in an equilibrium system is proportional to $\exp(-E/k_B T)$, where $T$ is the temperature and $k_B$ is Boltzmann's constant. This has just the same form as our solution (equation 2.21), if

$$D/\gamma = k_B T. \tag{2.22}$$

## 2.4 Solving the Diffusion Equation

We take a brief mathematical interlude, to review two important methods for solving the diffusion equation: Fourier transforms and Greens functions. Both rely upon the fact that the diffusion equation is linear: if a family of solutions $\rho_n(x,t)$ are known, then any linear combination of these solutions $\sum_n a_n \rho_n(x,t)$ is also a solution. If we can then expand the initial density $\rho(x,0) = \sum_n a_n \rho_n(x,0)$, we've formally found the solution.

Fourier methods are wonderfully effective computationally, because of fast Fourier Transform (FFT) algorithms for shifting from the real-space density to the solution space. Greens function methods are more important for analytical calculations and as a source of approximate solutions.[18]

[18] One should note that much of quantum field theory and many-body quantum mechanics is framed in terms of something also called Greens functions. These are distant, fancier cousins of the simple methods used in linear differential equations.

### 2.4.1 Fourier

The Fourier transform method decomposes $\rho$ into a family of *plane wave* solutions $\widetilde{\rho}_k(t)e^{-ikx}$.

The diffusion equation is homogeneous in space: our system is *translationally invariant*. That is, if we have a solution $\rho(x,t)$, another equally valid solution is given by $\rho(x-\Delta,t)$, which describes the evolution of an initial condition translated by $\Delta$ in the positive $x$ direction.[19] Under very general circumstances, a linear equation describing a translation–invariant system will have solutions given by plane waves $\rho(x,t) = \widetilde{\rho}_k(t)e^{-ikx}$.

[19] Make sure you know that $g(x) = f(x-\Delta)$ shifts the function in the positive direction: for example, the new function $g(\Delta)$ is at $\Delta$ what the old one was at the origin, $g(\Delta) = f(0)$.

We argue this important truth in detail in in the appendix (section A.4). Here we just try it. Plugging a plane wave into the diffusion equation 2.7, we find

$$\frac{\partial \rho}{\partial t} = \frac{d\widetilde{\rho}_k}{dt}e^{-ikx} = D\frac{\partial^2 \rho}{\partial x^2} = -Dk^2\widetilde{\rho}_k e^{-ikx} \tag{2.23}$$

$$\frac{d\widetilde{\rho}_k}{dt} = -Dk^2\widetilde{\rho}_k \tag{2.24}$$

$$\widetilde{\rho}_k(t) = \widetilde{\rho}_k(0)e^{-Dk^2 t}. \tag{2.25}$$

Now, these plane wave solutions by themselves are unphysical: we must combine them to get a sensible density. First, they are complex: we must add plane waves at $k$ and $-k$ to form cosine waves, or subtract them and dividing by $2i$ to get sine waves. Cosines and sines are also not by themselves sensible densities (because they go negative), but they in turn can be added to one another (for example, added to a

This is called the *Einstein relation*. Our rough derivation (equation 2.17) suggested that $D/\gamma = m\bar{v}^2$, which suggests that $k_B T$ must equal twice the kinetic energy along $x$ for the Einstein relation to hold: this is also true, and is called the *equipartition theorem* (section 3.2.2). The constants in the (non–equilibrium) diffusion equation are related to one another, because the density must evolve toward the equilibrium distribution dictated by statistical mechanics.

constant background $\rho_0$) to make for sensible densities. Indeed, we can superimpose all different wave-vectors to get the general solution

$$\rho(x,t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \widetilde{\rho}_k(0) e^{-ikx} e^{-Dk^2 t}\, dk. \tag{2.26}$$

Here the coefficients $\rho_k(0)$ we use are just the the Fourier transform of the initial density profile

$$\widetilde{\rho}_k(0) = \int_{-\infty}^{\infty} \rho(x,0) e^{ikx}\, dx \tag{2.27}$$

and we recognize equation 2.26 as the inverse Fourier transform of the solution time–evolved in Fourier space

$$\widetilde{\rho}_k(t) = \widetilde{\rho}_k(0) e^{-Dk^2 t}. \tag{2.28}$$

Thus, by writing $\rho$ as a superposition of plane waves, we find a simple law: the short–wavelength parts of $\rho$ are squelched as time $t$ evolves, with wavevector $k$ being suppressed by a factor $e^{-Dk^2 t}$.

### 2.4.2 Green

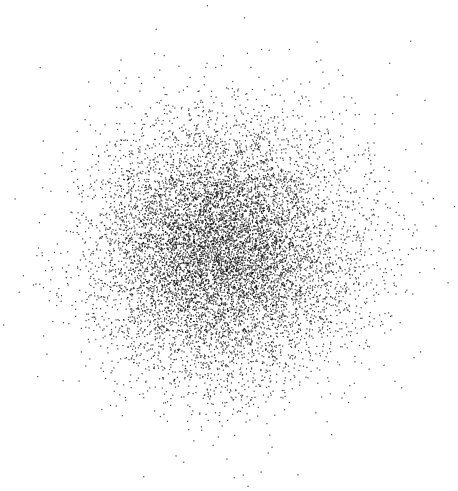


**Fig. 2.6** 10,000 endpoints of random walks, each 1000 steps long. Notice that after 1000 steps, the distribution of endpoints looks quite Gaussian. Indeed after about five steps the distribution is extraordinarily close to Gaussian, except far in the tails.

The Greens function method decomposes $\rho$ into a family of solutions $G(x-y,t)$ where all of the diffusing particles start at a particular point $y$.

Let's first consider the case where all particles start at the origin. Suppose we have one unit of perfume, released at the origin at time $t = 0$. What is the initial condition $\rho(x, t = 0)$? Clearly $\rho(x,0) = 0$ unless $x = 0$, and $\int \rho(x,0)dx = 1$, so $\rho(0,0)$ must be really, really infinite. This is of course the Dirac delta function $\delta(x)$, which mathematically (when integrated) is a linear operator on functions returning the value of the function at zero:

$$\int f(y)\delta(y)\, dy = f(0). \tag{2.29}$$

Let's define the Greens function $G(x,t)$ to be the time evolution of the density $G(x,0) = \delta(x)$ with all the perfume at the origin. Naturally, $G(x,t)$ obeys the diffusion equation $\frac{\partial G}{\partial t} = D\frac{\partial^2 G}{\partial x^2}$. We can use the Fourier transform methods of the previous section to solve for $G(x,t)$. The Fourier transform at $t = 0$ is

$$\widetilde{G}_k(0) = \int G(x,0) e^{ikx}\, dx = \int \delta(x) e^{ikx}\, dx = 1 \tag{2.30}$$

(independent of $k$). Hence the time evolved Fourier transform is $\widetilde{G}_k(t) = e^{-Dk^2 t}$, and the time evolution in real space is

$$G(x,t) = \frac{1}{2\pi}\int e^{-ikx}\widetilde{G}_k(0) e^{-Dk^2 t} dk = \frac{1}{2\pi}\int e^{-ikx} e^{-Dk^2 t} dk. \tag{2.31}$$

This last integral is the Fourier transform of a Gaussian. This transform can be performed[20] giving another Gaussian[21]

$$G(x,t) = \frac{1}{\sqrt{4\pi Dt}}\, e^{-x^2/4Dt}. \tag{2.33}$$

This is the Greens function for the diffusion equation.

The Greens function directly tells us the distribution of the endpoints of random walks centered at the origin (figure 2.6). Does it agree with our formula $\langle x^2 \rangle = Na^2$ for $N$-step random walks of step size $a$ (section 2.1)? At time $t$, the Greens function (equation 2.33) is a Gaussian with standard deviation $\sigma(t) = \sqrt{2Dt}$; plugging in our diffusion constant $D = \frac{a^2}{2\Delta t}$ (equation 2.13), we find an RMS distance of $\sigma(t) = a\sqrt{\frac{t}{\Delta t}} = a\sqrt{N}$, where $N = \frac{t}{\Delta t}$ is the number of steps taken in the random walk: our two methods do agree.

Finally, since the diffusion equation has translational symmetry, we can solve for the evolution of random walks centered at any point $y$: the time evolution of an initial condition $\delta(x-y)$ is $G(x-y,t)$. Since we can write any initial condition $\rho(x,0)$ as a superposition of $\delta$-functions

$$\rho(x,0) = \int \rho(y,0)\delta(x-y)\, dy \tag{2.34}$$

we can write a general solution $\rho(x,t)$ to the diffusion equation

$$\rho(x,0) = \int \rho(y,0)\delta(x-y)\, dy = \int \rho(y,0)G(x-y,0)\, dy \tag{2.35}$$

$$\rho(x,t) = \int \rho(y,0)G(x-y,t)\, dy = \int \rho(y,0)\frac{e^{-(y-x)^2/4Dt}}{\sqrt{4\pi Dt}}\, dy. \tag{2.36}$$

This equation states that the current value of the density is given by the original values of the density in the neighborhood, smeared sideways (convolved) with the function $G$.

Thus by writing $\rho$ as a superposition of point sources, we find that the diffusion equation smears out all the sharp features, averaging $\rho$ over ranges that grow proportionally to the typical random walk distance $\sqrt{2Dt}$.

[21] It's useful to remember that the Fourier transform of a normalized Gaussian $\frac{1}{\sqrt{2\pi}\sigma}\exp(-x^2/2\sigma^2)$ is another Gaussian, $\exp(-\sigma^2 k^2/2)$ of standard deviation $1/\sigma$ and with no prefactor.

[20] If we complete the square in the integrand $e^{-ikx}e^{-Dk^2t} = e^{-Dt(k+\frac{ix}{2Dt})^2}e^{-\frac{x^2}{4Dt}}$, and change variables to $\kappa = k + \frac{ix}{2Dt}$.

$$G(x,t) = \frac{1}{2\pi}e^{-\frac{x^2}{4Dt}} \int_{-\infty+\frac{ix}{2Dt}}^{\infty+\frac{ix}{2Dt}} e^{-Dt\kappa^2} d\kappa. \tag{2.32}$$

If we then shift the limits of integration downward to the real axis, the integral gives $\sqrt{\pi/Dt}$ yielding equation 2.33. This last step (shifting the limits of integration), is not trivial: we must rely on Cauchy's theorem, which allow one to deform the integration contour in the complex plane: see footnote 21 in exercise A.5.

# Exercises

Exercises 2.1, 2.2, and 2.3 give simple examples of random walks in different contexts. Exercises 2.4 and 2.5 illustrate the qualitative behavior of the Fourier and Greens function approaches to solving the diffusion equation. Exercises 2.6 and 2.7 apply the diffusion equation in the familiar context of thermal conductivity.[22] Exercise 2.8 explores self–avoiding random walks: in two dimensions, we find that the constraint that the walk must avoid itself gives new critical exponents and a new universality class (see also chapter 12).

Random walks also arise in nonequilibrium situations.

- They arise in living systems. Bacteria search for food (*chemotaxis*) using a biased random walk, randomly switching from a swimming state (random walk step) to a tumbling state (scrambling the velocity), see [10].
- They arise in economics: Black and Scholes [115] analyze the approximate random walks seen in stock prices (figure 2.3) to estimate the price of options – how much you charge a customer who wants a guarantee that they can by stock $X$ at price $Y$ at time $t$ depends not only on whether the average price will rise past $Y$, but also whether a random fluctuation will push it past $Y$.
- They arise in engineering studies of failure. If a bridge strut has $N$ microcracks each with a failure stress $\sigma_i$, and these stresses have probability density $\rho(\sigma)$, the engineer is not concerned with the average failure stress $\langle\sigma\rangle$, but the minimum. This introduces the study of *extreme value statistics*: in this case, the failure time distribution is very generally described by the *Weibull distribution*.

(2.1) **Random walks in Grade Space.**

Let's make a model of the prelim grade distribution. Let's imagine a multiple-choice test of ten problems of ten points each. Each problem is identically difficult, and the mean is 70. How much of the point spread on the exam is just luck, and how much reflects the differences in skill and knowledge of the people taking the exam? To test this, let's imagine that all students are identical, and that each question is answered at random with a probability 0.7 of getting it right.

*(a) What is the expected mean and standard deviation for the exam? (Work it out for one question, and then use our theorems for a random walk with ten steps.)*

A typical exam with a mean of 70 might have a standard deviation of about 15.

*(b) What physical interpretation do you make of the ratio of the random standard deviation and the observed one?*

(2.2) **Photon diffusion in the Sun.** (Basic)

Most of the fusion energy generated by the Sun is produced near its center. The Sun is $7 \times 10^5$ km in radius. Convection probably dominates heat transport in approximately the outer third of the Sun, but it is believed that energy is transported through the inner portions (say to a radius $R = 5 \times 10^8$ m) through a random walk of X-ray photons. (A photon is a quantized package of energy: you may view it as a particle which always moves at the speed of light $c$. Ignore for this exercise the index of refraction of the Sun.) Assume that the mean free path $\ell$ for the photon is $\ell = 5 \times 10^{-5} m$.

*About how many random steps $N$ will the photon take of length $\ell$ to get to the radius $R$ where convection becomes important? About how many years $\Delta t$ will it take for the photon to get there? (You may assume for this exercise that the photon takes steps in random directions, each of equal length given by the mean-free path.)* Related formulæ: $c = 3 \times 10^8$ m/s; $\langle x^2 \rangle \sim 2Dt$; $\langle s_n^2 \rangle = n\sigma^2 = n\langle s_1^2 \rangle$. There are $31,556,925.9747 \sim \pi \times 10^7 \sim 3 \times 10^7$ seconds in a year.

(2.3) **Ratchet and Molecular Motors.** (Basic, Biology)

Read Feynman's *Ratchet and Pawl* discussion in reference [89, I.46] for this exercise. Feynman's ratchet and pawl discussion obviously isn't so relevant to machines you can make in your basement shop. The thermal fluctuations which turn the wheel to lift the flea are too small to be noticeable on human length and time scales (you need to look in a microscope to see Brownian motion). On the other hand, his discussion turns out to be surprisingly close to how real cells move things around. Physics professor Michelle Wang studies these molecular motors in the basement of Clark Hall.

[22] We haven't derived the law of thermal conductivity from random walks of phonons. We'll give general arguments in chapter 10 that an energy flow linear in the thermal gradient is to be expected on very general grounds.

Inside your cells, there are several different molecular motors, which move and pull and copy (figure 2.7). There are molecular motors which contract your muscles, there are motors which copy your DNA into RNA and copy your RNA into protein, there are motors which transport biomolecules around in the cell. All of these motors share some common features: (1) they move along some linear track (microtubule, DNA, ...), hopping forward in discrete jumps between low-energy positions, (2) they consume energy (burning ATP or NTP) as they move, generating an effective force pushing them forward, and (3) their mechanical properties can be studied by seeing how their motion changes as the external force on them is changed (figure 2.8).

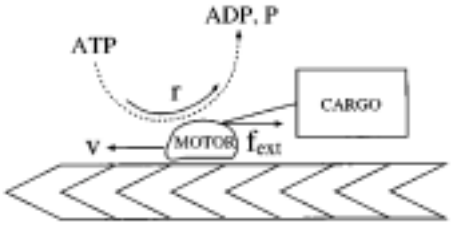



**Fig. 2.7** Cartoon of a motor protein, from reference [50]. As it carries some cargo along the way (or builds an RNA or protein, ...) it moves against an external force $f_{\text{ext}}$ and consumes $r$ ATP molecules, which are hydrolyzed to ADP and phosphate (P).

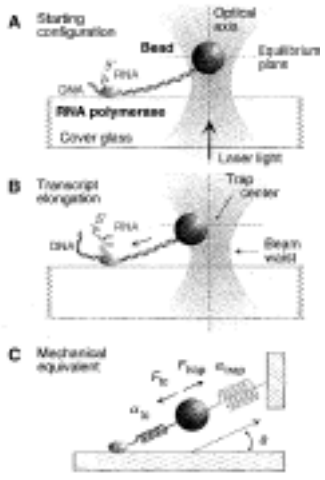



**Fig. 2.8** Cartoon of Cornell professor Michelle Wang's early laser tweezer experiment, (reference [123]). (A) The laser beam is focused at a point (the "laser trap"); the polystyrene bead is pulled (from dielectric effects) into the intense part of the light beam. The "track" is a DNA molecule attached to the bead, the motor is an RNA polymerase molecule, the "cargo" is the glass cover slip to which the motor is attached. (B) As the motor (RNA polymerase) copies DNA onto RNA, it pulls the DNA "track" toward itself, dragging the bead out of the trap, generating a force resisting the motion. (C) A mechanical equivalent, showing the laser trap as a spring and the DNA (which can stretch) as a second spring.

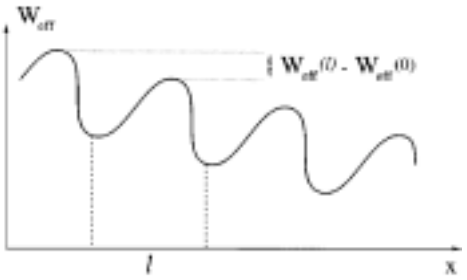



**Fig. 2.9** The effective potential for moving along the DNA (from reference [50]). Ignoring the tilt $W_e$, Feynman's energy barrier $\epsilon$ is the difference between the bottom of the wells and the top of the barriers. The experiment changes the tilt by adding an external force pulling $\ell$ to the left. In the absence of the external force, $W_e$ is the (Gibbs free) energy released when one NTP is burned and one RNA nucleotide is attached.

For transcription of DNA into RNA, the motor moves on average one base pair (A, T, G or C) per step: $\Delta\ell$ is about 0.34nm. We can think of the triangular grooves in the ratchet as being the low-energy states of the motor when it is resting between steps. The barrier between steps has an asymmetric shape (figure 2.9), just like the energy stored in the pawl is ramped going up and steep going down. Professor Wang showed (in a later paper) that the motor stalls at an external force of about 27 pN (pico-Newton).

*(a) At that force, what is the energy difference between neighboring wells due to the external force from the bead? (This corresponds to $L\theta$ in Feynman's ratchet.) Let's assume that this force is what's needed to balance the natural force downhill that the motor develops to propel the transcription process. What does this imply about the ratio of the forward rate to the backward rate, in the absence of the external force from the laser tweezers, at a temperature of $300K$, (from Feynman's discussion preceding equation 46.1)? ($k_B = 1.381 \times 10^{-23} J/K$).*

The natural force downhill is coming from the chemical reactions which accompany the motor moving one base pair: the motor burns up an NTP molecule into a $PP_i$ molecule, and attaches a nucleotide onto the RNA. The

net energy from this reaction depends on details, but varies between about 2 and 5 times $10^{-20}$ Joule. This is actually a Gibbs free energy difference, but for this exercise treat it as just an energy difference.

*(b) The motor isn't perfectly efficient: not all the chemical energy is available as motor force. From your answer to part (a), give the efficiency of the motor as the ratio of force-times-distance produced to energy consumed, for the range of consumed energies given.*

(2.4) **Solving Diffusion: Fourier and Green.** (Basic)

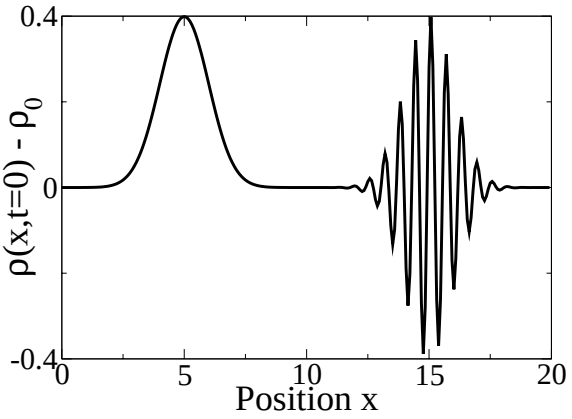



**Fig. 2.10** Initial profile of density deviation from average.

An initial density profile $\rho(x, t = 0)$ is perturbed slightly away from a uniform density $\rho_0$, as shown in figure 2.10. The density obeys the diffusion equation $\partial\rho/\partial t = D\partial^2\rho/\partial x^2$, where $D = 0.001$ m$^2$/s. The lump centered at $x = 5$ is a Gaussian $\exp(-x^2/2)/\sqrt{2\pi}$, and the wiggle centered at $x = 15$ is a smooth envelope function multiplying $\cos(10x)$.

*(a)* **Fourier.** *As a first step in guessing how the pictured density will evolve, let's consider just a cosine wave. If the initial wave were $\rho_{\cos}(x, 0) = \cos(10x)$, what would it be at $t = 10s$?* Related formulæ*: $\tilde{\rho}(k, t) = \tilde{\rho}(k, t')\tilde{G}(k, t-t')$; $\tilde{G}(k, t) = \exp(-Dk^2t)$.*

*(b)* **Green.** *As a second step, let's check how long it would take to spread out as far as the Gaussian on the left. If the wave at some earlier time $-t_0$ were a $\delta$ function at $x = 0$, $\rho(x, -t_0) = \delta(x)$, what choice of the time elapsed $t_0$ would yield a Gaussian $\rho(x, 0) = \exp(-x^2/2)/\sqrt{2\pi}$ for the given diffusion constant $D = 0.001m^2/s$?* Related formulæ*: $\rho(x, t) = \int \rho(y, t')G(y - x, t - t')\, dy$; $G(x, t) = (1/\sqrt{4\pi Dt}) \exp(-x^2/(4Dt))$.*

*(c)* **Pictures.** *Now consider time evolution for the next ten seconds. The initial density profile $\rho(x, t = 0)$ is as shown in figure 2.10. Which of the choices in figure 2.11 represents the density at $t = 10s$?* (Hint: compare $t = 10$s to the time $t_0$ from part (B).) Related formulæ*: $\langle x^2 \rangle \sim 2Dt$;*

(2.5) **Solving the Diffusion Equation.** (Basic) [23]

Consider a one-dimensional diffusion equation $\partial\rho/\partial t = D\partial^2\rho/\partial x^2$, with initial condition periodic in space with period $L$, consisting of a $\delta$ function at every $x_n = nL$: $\rho(x, 0) = \sum_{n=-\infty}^{\infty} \delta(x - nL)$.

*(a) Using the Greens function method, give an approximate expression for the the density, valid at short times and for $-L/2 < x < L/2$, involving only one term (not an infinite sum).* (Hint: how many of the Gaussians are important in this region at early times?)

*(b) Using the Fourier method,*[24] *give an approximate expression for the density, valid at long times, involving only two terms (not an infinite sum).* (Hint: how many of the wavelengths are important at late times?)

*(c) Give a characteristic time $\tau$ in terms of $L$ and $D$, such that your answer in (a) is valid for $t \ll \tau$ and your answer in (b) is valid for $t \gg \tau$.*

(2.6) **Frying Pan** (Basic)

An iron frying pan is quickly heated on a stove top to 400 degrees Celsius. Roughly how long it will be before the handle is too hot to touch (within, say, a factor of two)? (Adapted from reference [93, p. 40].)

Do this three ways.

*(a) Guess the answer from your own experience. If you've always used aluminum pans, consult a friend or parent.*

*(b) Get a rough answer by a dimensional argument. You need to transport heat $c_p\rho V\Delta T$ across an area $A = V/\Delta x$. How much heat will flow across that area per unit time, if the temperature gradient is roughly assumed to be $\Delta T/\Delta x$? How long $\delta t$ will it take to transport the amount needed to heat up the whole handle?*

*(c) Roughly model the problem as the time needed for a pulse of heat at $x = 0$ on an infinite rod to spread*

[23] Math reference: [71, sec. 8.4].

[24] If you use a Fourier transform of $\rho(x, 0)$, you'll need to sum over $n$ to get $\delta$-function contributions at discrete values of $k = 2\pi m/L$. If you use a Fourier series, you'll need to unfold the sum over $n$ of partial Gaussians into a single integral over an unbounded Gaussian.

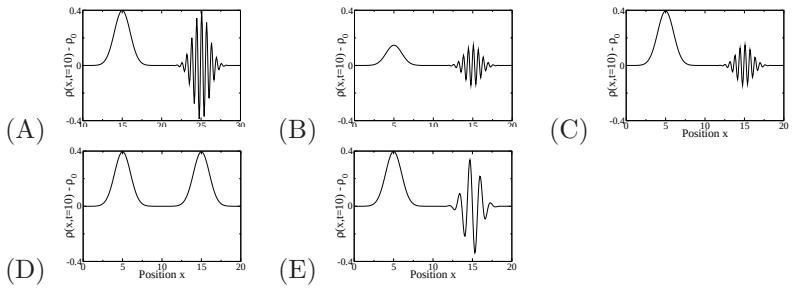



**Fig. 2.11** Final states of diffusion example

*out a distance equal to the length of the handle, and use the Greens function for the heat diffusion equation (exercise 2.7). How long until the pulse spreads out a root-mean square distance $\sigma(t)$ equal to the length of the handle?*

Note: For iron, the specific heat $c_p = 450J/kg \cdot C$, the density $\rho = 7900kg/m^3$, and the thermal conductivity $k_t = 80W/m \cdot C$.

(2.7) **Thermal Diffusion.** (Basic)

The rate of energy flow in a material with thermal conductivity $k_t$ and a temperature field $T(x, y, z, t) = T(\mathbf{r}, t)$ is $\mathbf{J} = -k_t \nabla T$.[25] Energy is locally conserved, so the energy density $E$ satisfies $\partial E/\partial t = -\nabla \cdot \mathbf{J}$.

*(a) If the material has constant specific heat $c_p$ and density $\rho$, so $E = c_p \rho T$, show that the temperature $T$ satisfies the diffusion equation $\partial T/\partial t = \frac{k_t}{c_p \rho} \nabla^2 T$.*

*(b) By putting our material in a cavity with microwave standing waves, we heat it with a periodic modulation $T = \sin(kx)$ at $t = 0$, at which time the microwaves are turned off. Show that amplitude of the temperature modulation decays exponentially in time. How does the amplitude decay rate depend on wavelength $\lambda = 2\pi/k$?*

(2.8) **Polymers and Random Walks.**

Polymers are long molecules, typically made of identical small molecules called *monomers* that are bonded together in a long, one-dimensional chain. When dissolved in a solvent, the polymer chain configuration often forms a good approximation to a random walk. Typically, neighboring monomers will align at relatively small angles: several monomers are needed to lose memory of the original angle. Instead of modeling all these small angles, we can produce an equivalent problem focusing all the bending in a few hinges: we approximate the polymer by an uncorrelated random walk of straight segments several monomers in length. The equivalent segment size is called the *persistence length.*[26]

*(a) If the persistence length to bending of DNA is 50nm, with 3.4Å per nucleotide base pair, what will the root-mean-square distance $\sqrt{\langle R^2 \rangle}$ be between the ends of a gene in solution with 100,000 base pairs, if the DNA is accurately represented as a random walk?*

Polymers are not accurately represented as random walks, however. Random walks, particularly in low dimensions, often intersect themselves. Polymers are best represented as *self-avoiding* random walks: the polymer samples all possible configurations that does not cross itself. (Greg Lawler, in the math department here, is an expert on self-avoiding random walks.)

Let's investigate whether avoiding itself will change the basic nature of the polymer configuration. In particular, does the end-to-end typical distance continue to scale with the square root of the length $L$ of the polymer, $R \sim \sqrt{L}$?

*(b)* **Two dimensional self-avoiding random walk.** *Give a convincing, short argument explaining whether or not a typical, non-self-avoiding random walk in two dimensions will come back after large numbers of monomers and cross itself.* (Hint: how big a radius does it extend to? How many times does it traverse this radius?)

**BU java applet.** Run the Java applet linked to at reference [72]. (You'll need to find a machine with Java enabled.) They model a 2-dimensional random walk as a connected line between nearest-neighbor neighboring lattice points on the square lattice of integers. They start random walks at the origin, grow them without allowing backtracking, and discard them when they hit the same lattice point twice. As long as they survive, they average the squared length as a function of number of steps.

[25] We could have derived this law of thermal conductivity from random walks of phonons, but we haven't yet done so.

[26] Some seem to define the persistence length with a different constant factor.

*(c) Measure for a reasonable length of time, print out the current state, and enclose it. Did the simulation give $R \sim \sqrt{L}$? If not, what's the estimate that your simulation gives for the exponent relating $R$ to $L$? How does it compare with the two-dimensional theoretical exponent given at the Web site?*

# Temperature and Equilibrium

# 3

We now turn to study the equilibrium behavior of matter: the historical origin of statistical mechanics. We will switch in this chapter between discussing the general theory and applying it to a particular system – the ideal gas. The ideal gas provides a tangible example of the formalism, and its solution will provide a preview of material coming in the next few chapters.

A system which is not acted upon by the external world[1] is said to approach *equilibrium* if and when it settles down at long times to a state which is independent of the initial conditions (except for conserved quantities like the total energy). Statistical mechanics describes the equilibrium state as an average over all states consistent with the conservation laws: this *microcanonical ensemble* is introduced in section 3.1. In section 3.2, we shall calculate the properties of the ideal gas using the microcanonical ensemble. In section 3.3 we shall define *entropy* and *temperature* for equilibrium systems, and argue from the microcanonical ensemble that heat flows to maximize the entropy and equalize the temperature. In section 3.4 we will derive the formula for the *pressure* in terms of the entropy, and define the *chemical potential*. Finally, in section 3.5 we calculate the entropy, temperature, and pressure for the ideal gas, and introduce some refinements to our definitions of phase space volume.

[1] If the system is driven (*e.g.*, there are externally imposed forces or currents) we instead call this final condition the *steady state*. If the system is large, the equilibrium state will also usually be time independent and 'calm', hence the name. Small systems will continue to fluctuate substantially even in equilibrium.

## 3.1 The Microcanonical Ensemble

Statistical mechanics allows us to solve *en masse* many problems that are impossible to solve individually. In this chapter we address the general equilibrium behavior of $N$ atoms in a box of volume $V$ – any kinds of atoms, in arbitrary external conditions. Let's presume for simplicity that the walls of the box are smooth and rigid, so that energy is conserved when atoms bounce off the walls. This makes our system *isolated*, independent of the world around it.

How can we solve for the behavior of our atoms? If we ignore quantum mechanics, we can in principle determine the positions[2] $\mathbb{Q} = (x_1, y_1, z_1, x_2, \ldots x_N, y_N, z_N) = (q_1 \ldots q_{3N})$ and momenta $\mathbb{P} = (p_1, \ldots p_{3N})$ of the particles at any future time given their initial positions and mo-

[2] The $3N$ dimensional space of positions $\mathbb{Q}$ is called *configuration space*. The $3N$ dimensional space of momenta $\mathbb{P}$ is called *momentum space*. The $6N$ dimensional space $(\mathbb{P}, \mathbb{Q})$ is called *phase space*.

menta using Newton's laws

$$\begin{aligned} \dot{\mathbb{Q}} &= \mathbf{m}^{-1}\mathbb{P} \\ \dot{\mathbb{P}} &= \mathbb{F}(\mathbb{Q}) \end{aligned} \tag{3.1}$$

(where $\mathbb{F}$ is the $3N$–dimensional force due to the other particles and the walls, and $\mathbf{m}$ is the particle mass).[3]

In general, solving these equations is plainly not feasible.

- Many systems of interest involve far too many particles to allow one to solve for their trajectories.
- Most systems of interest exhibit chaotic motion, where the time evolution depends with ever increasing sensitivity on the initial conditions – you cannot know enough about the current state to predict the future.
- Even if it were possible to evolve our trajectory, knowing the solution would for most purposes be useless: we're far more interested in the typical number of atoms striking a wall of the box, say, than the precise time a particular particle hits.[4]

How can we extract the simple, important predictions out of the complex trajectories of these atoms? The chaotic time evolution will rapidly scramble[5] whatever knowledge we may have about the initial conditions of our system, leaving us effectively knowing only the conserved quantities – for our system, just the total energy $E$.[6] Rather than solving for the behavior of a particular set of initial conditions, let us hypothesize that the energy is all we need to describe the equilibrium state. This leads us to a statistical mechanical description of the equilibrium state of our system as an ensemble of all possible initial conditions with energy $E$ – the *microcanonical ensemble.*

We calculate the properties of our ensemble by averaging over states with energies in a shell $(E, E+\delta E)$ taking the limit[7] $\delta E \to 0$ (figure 3.1). Let's define the function $\Omega(E)$ to be the phase-space volume of this thin shell:

$$\Omega(E)\,\delta E = \int_{E<\mathcal{H}(\mathbb{P},\mathbb{Q})<E+\delta E} d\mathbb{P}\,d\mathbb{Q}. \tag{3.2}$$

Here $\mathcal{H}(\mathbb{P},\mathbb{Q})$ is the Hamiltonian for our system.[8] Finding the average $\langle A \rangle$ of a property $A$ in the microcanonical ensemble is done by averaging $A(\mathbb{P},\mathbb{Q})$ over this same energy shell,[9]

$$\langle A \rangle_E = \frac{1}{\Omega(E)\delta E} \int_{E<\mathcal{H}(\mathbb{P},\mathbb{Q})<E+\delta E} A(\mathbb{P},\mathbb{Q})\,d\mathbb{P}\,d\mathbb{Q}. \tag{3.8}$$

[3] $\mathbf{m}$ is a diagonal matrix if the particles aren't all the same mass.

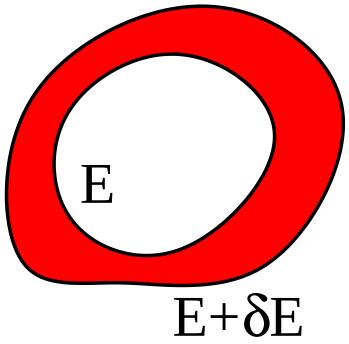



**Fig. 3.1** The shell of energies between $E$ and $E + \delta E$ can have an irregular "thickness". The volume of this shell in $6N$–dimensional phase space, divided by $\delta E$, is the definition of $\Omega(E)$. Notice that the microcanonical average weights the thick regions more heavily. We shall see in section 4.1 that this is the correct way to take the average: just as a water drop in a river spends more time in the deep sections where the water flows slowly, so also a trajectory in phase space spends more time in the thick regions where it moves more slowly.

[5] This scrambling, of course, is precisely the approach to equilibrium.

[6] If our box were spherical, angular momentum would also be conserved.

[7] What about quantum mechanics, where the energy levels in a finite system are discrete? In that case (chapter 7), we will need to keep $\delta E$ large compared to the spacing between energy eigenstates, but small compared to the total energy.

[8] The Hamiltonian $\mathcal{H}$ is the function of $\mathbb{P}$ and $\mathbb{Q}$ that gives the energy. For our purposes, this will always be $\mathbb{P}^2/2m + V(\mathbb{Q}) = \sum_{\alpha=1}^{3N} {p_\alpha}^2/2m + V(q_1, \ldots, q_{3N})$, where the force in Newton's laws 3.1 is $F_\alpha = -\frac{\partial V}{\partial q_\alpha}$.

[4] Of course, there are applications where the precise evolution of a particular system is of interest. It would be nice to predict the time at which a particular earthquake fault will yield, so as to warn everyone to go for a picnic outdoors. Statistical mechanics, broadly speaking, is helpless in computing such particulars. The budget of the weather bureau is a good illustration of how hard such system-specific predictions are.

[9] It is convenient to write the energy shell $E < \mathcal{H}(\mathbb{P},\mathbb{Q}) < E + \delta E$ in terms of the Heaviside step function $\Theta(x)$:

$$\Theta(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases}; \tag{3.3}$$

Notice that, by averaging equally over all states in phase space compatible with our knowledge about the system (that is, the conserved energy), we have made a hidden assumption: all points in phase space (with a given energy) are *a priori* equally likely, so the average should treat them all with equal weight. In section 3.2, we will see that this assumption leads to sensible behavior, by solving the simple case of an ideal gas. We will fully justify this equal-weighting assumption in chapter 4, where we will also discuss the more challenging question of why so many systems actually reach equilibrium.

The fact that the microcanonical distribution describes equilibrium systems should be amazing to you. The long-time equilibrium behavior of a system is precisely the typical behavior of all systems with the same value of the conserved quantities. This fundamental "regression to the mean" is the basis of statistical mechanics.

## 3.2 The Microcanonical Ideal Gas

We can *talk* about a general collection of atoms, and derive general statistical mechanical truths for them, but to *calculate* specific properties we must choose a particular system. The simplest statistical mechanical system is the monatomic[10] ideal gas. You can think of helium atoms at high temperatures and low densities as a good approximation to this ideal gas – the atoms have very weak long-range interactions and rarely collide. The ideal gas will be the limit when the interactions between

[10] Air is a mixture of gases, but most of the molecules are diatomic: $O_2$ and $N_2$, with a small admixture of triatomic $CO_2$ and monatomic Ar. The properties of diatomic ideal gases are almost as simple: but one must keep track of the internal rotational degree of freedom (and, at high temperatures, the vibrational degrees of freedom).

we see that $\Theta(E+\delta E-\mathcal{H})-\Theta(E-\mathcal{H})$ is one precisely inside the energy shell (see figure 3.1). In the limit $\delta E \to 0$, we can write $\Omega(E)$ as a derivative

$$\begin{aligned}\Omega(E)\delta E &= \int_{E<\mathcal{H}(\mathbb{P},\mathbb{Q})<E+\delta E} d\mathbb{P}\, d\mathbb{Q} \\ &= \int d\mathbb{P}\, d\mathbb{Q}\ [\Theta(E+\delta E-\mathcal{H})-\Theta(E-\mathcal{H})] \\ &= \delta E \frac{\partial}{\partial E}\int d\mathbb{P}\, d\mathbb{Q}\, \Theta(E-\mathcal{H}) \end{aligned} \tag{3.4}$$

and the expectation of a general operator $A$ as

$$\begin{aligned}\langle A\rangle &= \frac{1}{\Omega(E)}\int d\mathbb{P}\, d\mathbb{Q}\ [\Theta(E+\delta E-\mathcal{H})-\Theta(E-\mathcal{H})]\, A(\mathbb{P},\mathbb{Q}) \\ &= \frac{1}{\Omega(E)}\frac{\partial}{\partial E}\int d\mathbb{P}\, d\mathbb{Q}\, \Theta(E-\mathcal{H})A(\mathbb{P},\mathbb{Q}). \end{aligned} \tag{3.5}$$

It will be important later to note that the derivatives in equations 3.4 and 3.5 are at constant $N$ and constant $V$: $\frac{\partial}{\partial E}\Big|_{V,N}$. Finally, we know the derivative of the Heaviside function is the the Dirac $\delta$-function. (You may think of $\delta(x)$ as the limit as $\epsilon \to$ zero of a function which is $1/\epsilon$ in the range $(0,\epsilon)$. Mathematicians may think of it as a point mass at the origin.)

$$\Omega(E) = \int d\mathbb{P}\, d\mathbb{Q}\, \delta\left(E-\mathcal{H}(\mathbb{P},\mathbb{Q})\right), \tag{3.6}$$

$$\langle A\rangle = \frac{1}{\Omega(E)}\int d\mathbb{P}\, d\mathbb{Q}\, \delta\left(E-\mathcal{H}(\mathbb{P},\mathbb{Q})\right) A(\mathbb{P},\mathbb{Q}). \tag{3.7}$$

Thus the microcanonical ensemble can be written as a probability density $\delta\left(E-\mathcal{H}(\mathbb{P},\mathbb{Q})\right)/\Omega(E)$ in phase space.

particles vanish.[11]

For the ideal gas, the energy does not depend upon the spatial configuration $\mathbb{Q}$ of the particles. This allows us to study the positions (section 3.2.1 separately from the momenta (section 3.2.2).

### 3.2.1 Configuration Space

Since the energy is independent of the position, our microcanonical ensemble must weight all configurations equally. That is to say, it is precisely as likely that all the particles will be within a distance $\epsilon$ of the middle of the box as it is that they will be within a distance $\epsilon$ of any other particular configuration.

What is the probability density $\rho(\mathbb{Q})$ that the ideal gas particles will be in a particular configuration $\mathbb{Q} \in \mathbb{R}^{3N}$ inside the box of volume V? We know $\rho$ is a constant, independent of the configuration. We know that the gas atoms are in some configuration, so $\int \rho \, d\mathbb{Q} = 1$. The integral over the positions gives a factor of $V$ for each of the $N$ particles, so $\rho(\mathbb{Q}) = 1/V^N$.

It may be counterintuitive that unusual configurations, like all the particles on the right half of the box, have the same probability density as more typical configurations. If there are two non-interacting particles in a $L \times L \times L$ box centered at the origin, what is the probability that both are on the right (have $x > 0$)? The probability that two particles are on the right half is the integral of $\rho = 1/L^6$ over the six dimensional volume where both particles have $x > 0$. The volume of this space is $(L/2) \times L \times L \times (L/2) \times L \times L = L^6/4$, so the probability is $1/4$, just as one would calculate by flipping a coin for each particle. The probability that $N$ such particles are on the right is $2^{-N}$ – just as your intuition would suggest. Don't confuse probability density with probability! The unlikely states for molecules are not those with small probability density. Rather, they are states with small net probability, because their allowed configurations and/or momenta occupy insignificant volumes of the total phase space.

Notice that configuration space typically has dimension equal to several times Avogadro's number.[12] Enormous–dimensional vector spaces have weird properties – which directly lead to to important principles in statistical mechanics. For example, most of configuration space has almost exactly half the $x$-coordinates on the right side of the box.

If there are $2N$ non-interacting particles in the box, what is the probability $P_m$ that $N + m$ of them will be on the right half? There are $2^{2N}$ equally likely ways the distinct particles could sit on the two sides of the box. Of these, $\binom{2N}{N+m} = (2N)!/((N+m)!(N-m)!)$ have $m$ extra particles on the right half.[13] So,

[12] A gram of hydrogen has approximately $N = 6.02 \times 10^{23}$ atoms, known as Avogadro's number. So, a typical $3N$ will be around $10^{24}$.

[13] $\binom{p}{q}$ is the number of ways of choosing an unordered subset of size $q$ from a set of size $p$. There are $p(p-1)...(p-q+1) = p!/(p-q)!$ ways of choosing an ordered subset, since there are $p$ choices for the first member and $p-1$ for the second ... There are $q!$ different ordered sets for each disordered one, so $\binom{p}{q} = p!/(q!(p-q)!)$.

[11] With no interactions, how can the ideal gas reach equilibrium? If the particles never collide, they will forever be going with whatever initial velocity we started them. We imagine delicately taking the long time limit first, before taking the limit of weak interactions, so we can presume an equilibrium distribution has been established.

$$P_m = 2^{-2N}\binom{2N}{N+m} = 2^{-2N}(2N)!/((N+m)!(N-m)!). \tag{3.9}$$

We can calculate the fluctuations in the number on the right using Stirling's formula,[14]

$$n! \sim (n/e)^n\sqrt{2\pi n} \sim (n/e)^n. \tag{3.10}$$

For now, let's use the second, less accurate form: keeping the factor $\sqrt{2\pi n}$ would fix the prefactor in the final formula (exercise 3.5) which we will instead derive by normalizing the total probability to one. Using Stirling's formula, equation 3.9 becomes

$$\begin{aligned} P_m &\approx 2^{-2N}\left(\frac{2N}{e}\right)^{2N} \Big/ \left(\frac{N+m}{e}\right)^{N+m}\left(\frac{N-m}{e}\right)^{N-m} \\ &= N^{2N}(N+m)^{-(N+m)}(N-m)^{-(N-m)} \\ &= (1+m/N)^{-(N+m)}(1-m/N)^{-(N-m)} \\ &= ((1+m/N)(1-m/N))^{-N}\,(1+m/N)^{-m}(1-m/N)^{m} \\ &= \left((1-m^2/N^2)\right)^{-N}\,(1+m/N)^{-m}(1-m/N)^{m} \end{aligned} \tag{3.11}$$

and, since $m \ll N$ we may substitute $1+\epsilon \approx \exp(\epsilon)$, giving us

$$P_m \approx \left(e^{-m^2/N^2}\right)^{-N}\left(e^{m/N}\right)^{-m}\left(e^{-m/N}\right)^{m} \approx P_0\exp(-m^2/N). \tag{3.12}$$

where $P_0$ is the prefactor we missed by not keeping enough terms in Stirling's formula. We know that the probabilities must sum to one, so again for $m \ll N$, $1 = \sum_m P_m \approx \int P_0\exp(-m^2/N)\,dm = P_0\sqrt{\pi N}$. Hence

$$P_m \approx \sqrt{1/\pi N}\exp(-m^2/N). \tag{3.13}$$

This is a nice result: it says that the number fluctuations are distributed according to a Gaussian or normal distribution[15] $(1/\sqrt{2\pi}\sigma)\exp(-x^2/2\sigma^2)$ with a standard deviation $\sigma_m = \sqrt{N/2}$. If we have Avogadro's number of particles $N \sim 10^{24}$, then the fractional fluctuations $\sigma_m/N = \frac{1}{\sqrt{2N}} \sim 10^{-12} = 0.0000000001\%$. In almost all the volume of a box in $\mathbb{R}^{3N}$, almost exactly half of the coordinates are on the right half of their range. In section 3.2.2 we will find another weird property of high–dimensional spaces.

We will find that the relative fluctuations of most quantities of interest in equilibrium statistical mechanics go as $1/\sqrt{N}$. *For many properties of macroscopic systems, statistical mechanical fluctuations about the average value are very small.*

### 3.2.2 Momentum Space

Working with the microcanonical momentum distribution is more challenging, but more illuminating, than working with the ideal gas configuration space of the last section. Here we must study the geometry of spheres in high dimensions.

[14] Stirling's formula tells us that the "average" number in the product $n(n-1)\ldots1$ is roughly $n/e$. See exercise 1.4.

[15] We derived exactly this result in section 2.4.2 using random walks and a continuum approximation, instead of Stirling's formula: this Gaussian is the Green's function for the number of heads in $2N$ coin flips. We'll derive it again in exercise 12.7 by deriving the central limit theorem using renormalization-group methods.

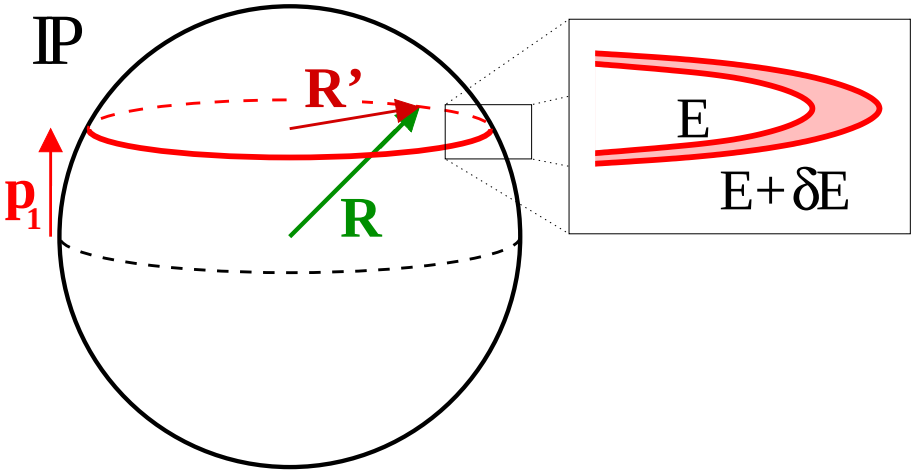



**Fig. 3.2** The energy surface in momentum space is the $3N-1$ – sphere of radius $R=\sqrt{2mE}$. The conditions that the x-component of the momentum of atom #1 is $p_1$ restricts us to a circle (or rather $3N-2$ – sphere) of radius $R'=\sqrt{2mE-{p_1}^2}$. The condition that the energy is in the shell $(E, E+\delta E)$ leaves us with the annular region shown in the inset.

The kinetic energy for interacting particles is $\sum_{\alpha=1}^{3N} \frac{1}{2} m_\alpha {v_\alpha}^2 = \sum_{\alpha=1}^{3N} {p_\alpha}^2/2m_\alpha$. If we assume all of our atoms have the same mass $m$, this simplifies to $\mathbb{P}^2/2m$. Hence the condition that the particles in our system have energy $E$ is that the system lies on a sphere in $3N$–dimensional momentum space of radius $R=\sqrt{2mE}$. Mathematicians[16] call this the $3N-1$ sphere, $\mathbb{S}_R^{3N-1}$. Specifically, if the energy of the system is known to be in a small range between $E$ and $E+\delta E$, what is the corresponding volume of momentum space? The volume $\mu\left(\mathbb{S}_R^{\ell-1}\right)$ of the $\ell-1$ sphere (in $\ell$ dimensions) of radius $R$ is[17]

$$\mu\left(\mathbb{S}_R^{\ell-1}\right) = \pi^{\ell/2} R^\ell / \tfrac{\ell}{2}! \tag{3.14}$$

The volume of the thin shell[18] between $E$ and $E+\delta E$ is given by

$$\begin{aligned}\frac{\text{Momentum Shell Volume}}{\delta E} &= \frac{\mu(\mathbb{S}^{3N-1}_{\sqrt{2M(E+\delta E)}}) - \mu(\mathbb{S}^{3N-1}_{\sqrt{2ME}})}{\delta E} \\ &= d\mu\Big(\mathbb{S}^{3N-1}_{\sqrt{2mE}}\Big)\Big/dE \\ &= \frac{d}{dE}\left(\pi^{\frac{3N}{2}}(2mE)^{\frac{3N}{2}}/\tfrac{3N}{2}!\right) \\ &= \pi^{\frac{3N}{2}}(3Nm)(2mE)^{\frac{3N}{2}-1}/\tfrac{3N}{2}! \\ &= (3N/2E)\pi^{\frac{3N}{2}}(2mE)^{\frac{3N}{2}}/\tfrac{3N}{2}!. \end{aligned} \tag{3.15}$$

Formula 3.15 is the main result of this section. Given our microcanonical ensemble that equally weights all states with energy $E$, the probability density for having any particular set of particle momenta $\mathbb{P}$ is the inverse of this shell volume.

[16] Mathematicians like to name surfaces, or manifolds, for the number of dimensions or local coordinates internal to the manifold, rather than the dimension of the space the manifold lives in. After all, one can draw a circle embedded in any number of dimensions (down to two). Thus a basketball is a two sphere $\mathbb{S}^2$, the circle is the one-sphere $\mathbb{S}^1$, and the zero sphere $\mathbb{S}^0$ consists of the two points $\pm 1$.

[17] Check this in two dimensions. Using $\frac{1}{2}! = \sqrt{\pi}/2$ and $\frac{3}{2}! = 3\sqrt{\pi}/4$, check it in one and three dimensions (see exercise 1.4 for $n!$ for non-integer $n$.) Is $n! = n\,(n-1)!$ valid for $n = 3/2$?

[18] This is not quite the surface area, since we're taking a shell of energy rather than radius. That's why its volume goes as $R^{3N-2}$, rather than $R^{3N-1}$.

Let's do a tangible calculation. Let's calculate the probability density $\rho(p_1)$ that the $x$-component of the momentum of the first atom is $p_1$.[19] The probability density that this momentum is $p_1$ and the energy is in the range $(E, E+\delta E)$ is proportional to the area of the annular region (between two $3N-2$ – spheres) in figure 3.2. The sphere has radius $R = \sqrt{2mE}$, so by the Pythagorean theorem, the circle has radius $R' = \sqrt{2mE - {p_1}^2}$. The volume in momentum space of the $3N-2$–dimensional annulus is given by using equation 3.14 with $\ell = 3N-1$:

[19] It is a sloppy physics convention to use $\rho$ to denote probability densities of all sorts. Earlier, we used it to denote probability density in $3N$–dimensional configuration space; here we use it to denote probability density in one variable. The argument of the function $\rho$ tells us which function we're considering.

$$\begin{aligned}
\text{Annular Area}/\delta E &= d\mu\left(\mathbb{S}^{3N-2}_{\sqrt{2mE-{p_1}^2}}\right) \Big/ dE \\
&= \frac{d}{dE}\left(\pi^{(3N-1)/2}(2mE-{p_1}^2)^{(3N-1)/2}/\tfrac{3N-1}{2}!\right) \\
&= \pi^{(3N-1)/2}((3N-1)m)(2mE-{p_1}^2)^{(3N-3)/2}/\tfrac{3N-1}{2}! \\
&= (3N-1)m\pi^{(3N-1)/2}R'^{3N-3}/\tfrac{3N-1}{2}! \\
&= [Constants]R'^{\,3N-3}, \qquad (3.16)
\end{aligned}$$

where we've dropped multiplicative factors that are independent of $p_1$ and $E$. The probability density of being in the annulus is its area divided by the shell volume in equation 3.15; this shell volume can be simplified as well, dropping terms that do not depend on $E$:

$$\begin{aligned}
\frac{\text{Momentum Shell Volume}}{\delta E} &= \pi^{\frac{3N}{2}}(3Nm)(2mE)^{\frac{3N}{2}-1}/\tfrac{3N}{2}! \\
&= 3Nm\pi^{\frac{3N}{2}}R^{3N-2}/\tfrac{3N}{2}! \\
&= [\text{Constants}]R^{3N-2}. \qquad (3.17)
\end{aligned}$$

Our formula for the probability density $\rho(p_1)$ is thus

$$\begin{aligned}
\rho(p_1) &= \frac{\text{Annular Area}}{\text{Momentum Shell Volume}} \\
&= \frac{(3N-1)m\pi^{(3N-1)/2}R'^{3N-3}/\frac{3N-1}{2}!}{3Nm\pi^{\frac{3N}{2}}R^{3N-2}/\frac{3N}{2}!} \\
&= [\text{Constants}](R^2/R'^3)(R'/R)^{3N} \qquad (3.18) \\
&= [\text{Constants}](R^2/R'^3)(1-{p_1}^2/2mE)^{\frac{3N}{2}}.
\end{aligned}$$

The probability density $\rho(p_1)$ will be essentially zero unless $R'/R = \sqrt{1-{p_1}^2/2mE}$ is nearly equal to one, since this factor is taken to a power $3N/2$ of around Avogadro's number. We can thus simplify $R^2/R'^3 \approx 1/R = 1/\sqrt{2mE}$ and $(1-{p_1}^2/2mE) = (1-\epsilon) \approx \exp(-\epsilon) = \exp(-{p_1}^2/2mE)$, giving us

$$\rho(p_1) \propto 1/\sqrt{2mE}\exp\left(\frac{-{p_1}^2}{2m}\frac{3N}{2E}\right) \qquad (3.19)$$

The probability density $\rho(p_1)$ is a Gaussian distribution of standard deviation $\sqrt{2mE/3N}$; we again can set the constant of proportionality

to normalize the Gaussian, leading to

$$\rho(p_1) = \frac{1}{\sqrt{2\pi m(2E/3N)}} \exp\left(\frac{-{p_1}^2}{2m}\frac{3N}{2E}\right) \tag{3.20}$$

Our ensemble assumption has allowed us to calculate the momentum distribution of our particles explicitly in terms of $E$, $N$, and $m$, without ever considering a particular trajectory: this is what makes statistical mechanics powerful.

Formula 3.20 tells us that most of the surface area of a large–dimensional sphere is very close to the equator. Think of $p_1$ as the latitude on the sphere: The range of latitudes containing most of the area is $\delta p \approx \pm\sqrt{2mE/3N}$, and the total range of latitudes is $\pm\sqrt{2mE}$: the belt divided by the height is the square root of Avogadro's number. This is true *whatever equator you choose*, even intersections of several equators. Geometry is weird in high dimensions.

In the context of statistical mechanics, this seems much less strange: typical configurations of gases have the kinetic energy divided roughly equally among all the components of momentum: configurations where one atom has most of the kinetic energy are vanishingly rare.

This formula foreshadows four key results that will emerge from our systematic study of equilibrium statistical mechanics in the following few chapters.

(1) **Temperature.** In our calculation, a single momentum component competed for the available energy with the rest of the ideal gas. In section 3.3 we will study the competition in general between two large subsystems for energy, and will discover that the balance is determined by the *temperature*. The temperature $T$ for our ideal gas will be given (equation 3.57) by $k_B T = \frac{2E}{3N}$.[20] Equation 3.20 then gives us the important formula

$$\rho(p_1) = 1/\sqrt{2\pi m k_B T}\, \exp(-{p_1}^2/2mk_BT). \tag{3.21}$$

(2) **Boltzmann distribution.** The probability of the x-momentum of the first particle having kinetic energy $K = p_1^2/2m$ is proportional to $\exp(-K/k_BT)$ (equation 3.21). This is our first example of a *Boltzmann distribution*. We shall see in section 6.1 that the probability of a small subsystem being in a particular state[21] of energy $E$ will in completely general contexts have probability proportional to $\exp(-E/k_BT)$.

(3) **Equipartition theorem.** The average kinetic energy $\langle p_1^2/2m\rangle$ from equation 3.21 is $k_BT/2$. This is an example of the *equipartition theorem* (section 6.2): each harmonic degree of freedom in an equilibrium classical system has average energy $k_BT/2$.

(4) **General classical momentum distribution.** Our derivation was in the context of a monatomic ideal gas. But we could have done an analogous calculation for a system with several gases of different masses: our momentum sphere would become an ellipsoid,

[20] We shall see that temperature is naturally measured in units of energy. Historically we measure temperature in degrees and energy in various other units (Joules, ergs, calories, eV, foot-pounds, ...); Boltzmann's constant $k_B$ is the conversion factor between units of temperature and units of energy.

[21] This is different from the probability of the subsystem *having* energy $E$, which is the product of the Boltzmann probability times the number of states with that energy.

but the calculation would still give the same distribution.[22] What is more surprising, we shall see when we introduce the canonical ensemble (section 6.1), that interactions don't matter either, so long as the system is classical:[23] the calculation factors and the probability densities for the momenta are given by equation 3.21, independent of the potential energies.[24] The momentum distribution in the form equation 3.21 is correct for nearly all equilibrium systems of classical particles.

[23] Notice that almost all molecular dynamics simulations are done classically: their momentum distributions are given by equation 3.21.

## 3.3 What is Temperature?

When a hot body is placed beside a cold one, our ordinary experience suggests that heat energy flows from hot to cold until they reach the same temperature. In statistical mechanics, the distribution of heat between the two bodies is determined by the assumption that all possible states of the two bodies at fixed total energy are equally likely. Do these two definitions agree? Can we define the temperature so that two large bodies in equilibrium with one another will have the same temperature?

Consider a general, isolated system of total energy $E$ consisting of two parts, labeled 1 and 2. Each subsystem has fixed volume and number of particles, and is energetically weakly connected to the other subsystem. The connection is weak in that we assume we can neglect the dependence of the energy $E_1$ of the first subsystem on the state $s_2$ of the second one, and vice versa.[25]

[25] A macroscopic system attached to the external world at its boundaries is usually weakly connected, since the interaction energy is only important at the surfaces, which are a negligible fraction of the total. Also, the momenta and positions of classical particles without magnetic fields are weakly connected in this sense: no terms in the Hamiltonian mix them (although the dynamical evolution certainly does).

Our microcanonical ensemble then asserts that the equilibrium ensemble of the total system is an equal weighting of all possible states of the two subsystems having total energy $E$. A particular state of the whole system is given by a pair of states $(s_1, s_2)$ with $E = E_1 + E_2$. This immediately implies that a particular configuration or state $s_1$ of the first subsystem at energy $E_1$ will occur with probability density[26]

$$\rho(s_1) \propto \Omega_2(E - E_1) \tag{3.22}$$

where $\Omega_1(E_1)\,\delta E_1$ and $\Omega_2(E_2)\,\delta E_2$ are the phase-space volumes of the energy shell for the two subsystems. The volume of the energy surface for the total system at energy $E$ will be given by adding up the product of the volumes of the subsystems for pairs of energy summing to $E$. This

[26] That is, if we compare the probabilities of two states of the subsystems with energies $E_a$ and $E_b$, and if $\Omega_2(E - E_a)$ is 50 times larger than $\Omega_2(E - E_b)$, then $\rho(E_a) = 50\,\rho(E_b)$ because the former has 50 times as many partners that it can pair with to get an allotment of probability.

[22] Molecular gases will have internal vibration modes that are often not well described by classical mechanics. At low temperatures, these are often frozen out: including rotations and translations but ignoring vibrations leads to the traditional formulas used, for example, for air (see note 10 on page 31).

[24] Quantum mechanics, however, couples the kinetic and potential terms: see chapter 7. Quantum mechanics is important for atomic motions only at low temperatures, so equation 3.21 will be reasonably accurate for all gases, all liquids but helium, and many solids that are not too cold.

may be intuitively clear, but we can also formally derive it:

$$\begin{aligned}\Omega(E) &= \frac{1}{\delta E}\int_{E<\mathcal{H}_1+\mathcal{H}_2<E+\delta E} d\mathbb{P}_1\, d\mathbb{Q}_1\, d\mathbb{P}_2\, d\mathbb{Q}_2 \\ &= \int d\mathbb{P}_1\, d\mathbb{Q}_1 \left(\frac{1}{\delta E}\int_{E-\mathcal{H}_1<\mathcal{H}_2<E-\mathcal{H}_1} d\mathbb{P}_2\, d\mathbb{Q}_2\right) \\ &= \int d\mathbb{P}_1\, d\mathbb{Q}_1 \Omega_2(E-\mathcal{H}_1(\mathbb{P}_1,\mathbb{Q}_1)) \\ &\approx \int dE_1 \left(\frac{1}{\delta E}\int_{E_1<\mathcal{H}_1<E_1+\delta E} d\mathbb{P}_1\, d\mathbb{Q}_1\right)\Omega_2(E-E_1) \\ &= \int dE_1 \Omega_1(E_1)\Omega_2(E-E_1) \end{aligned} \tag{3.23}$$

Notice that the integrand in equation 3.23, normalized by the total integral, is just the probability density[27] of the subsystem having energy $E_1$:

$$\rho(E_1) = \Omega_1(E_1)\Omega_2(E-E_1)/\Omega(E). \tag{3.24}$$

[27] Warning: again we're being sloppy: we use $\rho(s_1)$ in equation 3.22 for the probability that the subsystem is in a particular state $s_1$ and we use $\rho(E_1)$ in equation 3.24 for the probability that a subsystem is in any of many particular states with energy $E_1$.

We will show in a moment that if the two subsystems have a large number of particles then $\rho(E_1)$ is a very sharply peaked function near its maximum at $E_1^*$. Hence, the energy in subsystem 1 is given (apart from small fluctuations) by the maximum in the integrand $\Omega_1(E_1)\Omega_2(E-E_1)$. The maximum is found when the derivative $\frac{d\Omega_1}{dE_1}\Omega_2 - \Omega_1\frac{d\Omega_2}{dE_2}$ is zero, so

$$\frac{1}{\Omega_1}\frac{d\Omega_1}{dE_1} = \frac{1}{\Omega_2}\frac{d\Omega_2}{dE_2}. \tag{3.25}$$

This is the condition for thermal equilibrium between the two subsystems.

It is more convenient not to work with $\Omega$, but rather to work with its logarithm. We define the *equilibrium entropy*

$$S_{\text{equil}}(E) = k_B \log(\Omega(E)) \tag{3.26}$$

for each of our systems, where $k_B$ is a scale factor representing the unfortunate fact that temperature is not measured in the same units as energy.[28] Then $dS/dE = k_B(1/\Omega)d\Omega/dE$, and the condition 3.25 for thermal equilibrium between two macroscopic bodies is precisely the condition

$$\frac{d}{dE_1}\left(S_1(E_1)+S_2(E-E_1)\right) = \left.\frac{dS_1}{dE_1}\right|_{E_1} - \left.\frac{dS_2}{dE_2}\right|_{E-E_1} = 0 \tag{3.27}$$

[28] See note 20 on page 36.

that entropy is an extremum. Indeed, since the sum of the entropies is the logarithm of the integrand in equation 3.23 which by assumption is expanded about a local maximum, the condition of thermal equilibrium *maximizes the entropy.*[29]

[29] We shall discuss different aspects of entropy and its growth in chapter 5.

We want to define the temperature so that it becomes equal when the two subsystems come to equilibrium. We've seen that

$$dS_1/dE = dS_2/dE \tag{3.28}$$

in thermal equilibrium. $dS/dE$ decreases upon increasing energy, so we define the temperature in statistical mechanics as

$$1/T = dS/dE. \tag{3.29}$$

Is the probability density $\rho(E_1)$ in equation 3.24 sharply peaked, as we have assumed? We can expand the numerator about the maximum $E_1 = E_1^*$, and use the fact that the temperatures balance to remove the terms linear in $E_1 - E_1^*$:

$$\begin{aligned}\Omega_1(E_1)\Omega_2(E-E_1) &= \exp\left(S_1(E_1)/k_B + S_2(E-E_1)/k_B\right) \\ &\approx \exp\Bigg(\Bigg(S_1(E_1^*) + \tfrac{1}{2}(E_1-E_1^*)^2\frac{\partial^2 S_1}{\partial E_1^2} \\ &\qquad + S_2(E-E_1^*) + \tfrac{1}{2}(E_1-E_1^*)^2\frac{\partial^2 S_2}{\partial E_2^2}\Bigg)\Big/k_B\Bigg) \\ &= \Omega_1(E_1^*)\Omega_2(E_2^*) \\ &\qquad \exp\left((E_1-E_1^*)^2\left(\frac{\partial^2 S_1}{\partial E_1^2} + \frac{\partial^2 S_2}{\partial E_2^2}\right)\Big/(2k_B)\right).\end{aligned} \tag{3.30}$$

Thus the energy fluctuations are Gaussian,

$$\rho(E_1) = \frac{1}{\sqrt{2\pi}\sigma_E} e^{-(E_1-E_1^*)^2/2\sigma_E^2} \tag{3.31}$$

with standard deviation $\sigma_E$ given by

$$k_B/\sigma_E^2 = \frac{\partial^2 S_1}{\partial E_1^2} + \frac{\partial^2 S_2}{\partial E_2^2}. \tag{3.32}$$

How large is $\frac{\partial^2 S}{\partial E^2}$ for a macroscopic system? It has units of inverse energy squared, but is the energy a typical system energy or an atomic energy? If it is a system-scale energy (scaling like the number of particles $N$) then the root-mean-square energy fluctuation $\sqrt{\langle (E_1 - E_1^*)^2\rangle}$ will be comparable to $E_1$ (enormous fluctuations). If it is an atomic-scale energy (going to a constant as $N \to \infty$) then the energy fluctuations will be independent of system size (microscopic). Quantities like the total energy which scale linearly with the system size are called *extensive*; quantities like temperature that go to a constant as the system grows large are called *intensive*. The entropy of a system is typically extensive,[30] so the second derivative $\frac{\partial^2 S}{\partial E^2} \approx [S]/[E^2] \approx N/N^2 \approx 1/N$. Hence, the energy fluctuations scale as $1/\sqrt{N}$ of the total energy.[31] Just as

[31] We can also calculate these fluctuations explicitly: see exercise 3.7.

[30] A large system can usually be decomposed into many small, weakly coupled subsystems, for which the entropies add. This will be simpler to show in the canonical ensemble (section 6.2. In the microcanonical ensemble, equation 3.30 can be used to show that the entropy of the total system is

$$S_{\rm tot}(E) = S_1(E_1^*) + S_2(E-E_1^*) + k_B \log(\sqrt{2\pi}\sigma_E) \tag{3.33}$$

which is extensive up to a small correction (due to the enhanced energy fluctuations by coupling the two subsystems). For systems with long-range forces like gravitation or with long-range correlations, breaking the system up into many weakly–coupled subsystems may not be possible, and entropy and energy need not be extensive.

for the configurations of the ideal gas, where the number of particles in half the box fluctuated very little, so also the energy $E_1$ fluctuates very little from the value $E_1^*$ at which the probability is maximum.[32] In both cases, the relative fluctuations scale as $1/\sqrt{N}$.

The inverse of the temperature is the cost of buying energy from the rest of the world. The lower the temperature, the more strongly the kinetic energy for the momentum component is pushed towards zero. Entropy is the currency being paid. For each unit $\delta E$ energy bought, we pay $\delta E/T = \delta E \cdot dS/dE = \delta S$ in reduced entropy of the world. Inverse temperature is the cost in entropy to buy a unit of energy.

The "rest of the world" is often called the *heat bath*; it is a source and sink for heat and fixes the temperature. All heat baths are equivalent, depending only on the temperature. More precisely, the equilibrium behavior of a system weakly coupled to the external world is independent of what the external world is made of – it depends only on the world's temperature. This is a deep truth.

[32] We will discuss fluctuations in detail in section 6.1, and in chapter 10.

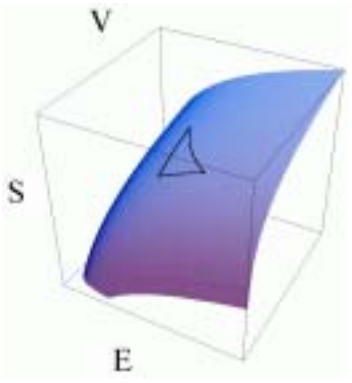



**Fig. 3.3 Entropy.** The entropy $S(E, V, N)$ as a function of energy $E$ and volume $V$ (at fixed number $N$). Viewed sideways, this surface also defines the energy $E(S, V, N)$. The three curves are lines at constant $S$,$E$, and $V$: the fact that they must close yields the relation $\left.\frac{\partial S}{\partial E}\right|_{V,N} \left.\frac{\partial E}{\partial V}\right|_{S,N} \left.\frac{\partial V}{\partial S}\right|_{E,N} = -1$ (see exercise 3.6).

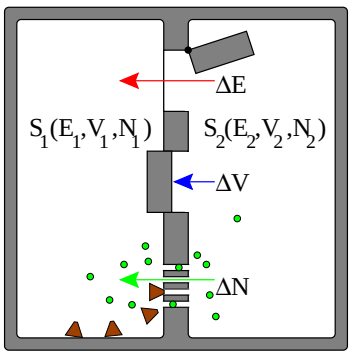



**Fig. 3.4 Two subsystems.** Two subsystems, isolated from the outside world, may exchange energy (open door through the insulation), volume (piston), or particles (tiny uncorked holes).

## 3.4 Pressure and Chemical Potential

The entropy $S(E, V, N)$ is our first example of a *thermodynamic potential*. In thermodynamics, all the macroscopic properties can be calculated by taking derivatives of thermodynamic potentials with respect to their arguments. It is often useful to think of thermodynamic potentials as surfaces: figure 3.3 shows the surface in $S, E, V$ space (at constant number of particles $N$). The energy $E(S, V, N)$ is another thermodynamic potential, completely equivalent to $S(E, V, N)$: it's the same surface with a different direction 'up'.

In section 3.3 we defined the temperature using $\left.\frac{\partial S}{\partial E}\right|_{V,N}$. What about the other two first derivatives, $\left.\frac{\partial S}{\partial V}\right|_{E,N}$ and $\left.\frac{\partial S}{\partial N}\right|_{E,V}$? That is, how does the entropy change when volume or particles are exchanged between two subsystems? The change in the entropy for a tiny shift $\Delta E$, $\Delta V$, and $\Delta N$ from subsystem 2 to subsystem 1 (figure 3.4) is

$$\Delta S = \left( \left.\frac{\partial S_1}{\partial E_1}\right|_{V,N} - \left.\frac{\partial S_2}{\partial E_2}\right|_{V,N} \right) \Delta E + \left( \left.\frac{\partial S_1}{\partial V_1}\right|_{E,N} - \left.\frac{\partial S_2}{\partial V_2}\right|_{E,N} \right) \Delta V + \left( \left.\frac{\partial S_1}{\partial N_1}\right|_{E,V} - \left.\frac{\partial S_2}{\partial N_2}\right|_{E,V} \right) \Delta N. \tag{3.34}$$

The first term is of course $(1/T_1 - 1/T_2)\Delta E$; exchanging energy to maximize the entropy sets the temperatures equal. Just as for the energy, if the two subsystems are allowed to exchange volume and number then the entropy will maximize itself with respect to these variables as well, with small fluctuations.[33] Equating the derivatives with respect to volume

[33] If the systems are at different temperatures and the piston is allowed to act, we would expect the pressures to equalize. Showing that this maximizes the entropy is complicated by the fact that the motion of a piston not only exchanges volume $\Delta V$

gives us our statistical mechanics definition of the pressure $P$:

$$P/T = \left.\frac{\partial S}{\partial V}\right|_{E,N} \tag{3.35}$$

and equating the derivatives with respect to number gives us the definition of the chemical potential $\mu$:

$$-\mu/T = \left.\frac{\partial S}{\partial N}\right|_{E,V}. \tag{3.36}$$

These definitions are a bit odd: usually we define pressure and chemical potential in terms of the change in energy $E$, not the change in entropy $S$. There is an important mathematical identity that we derive in exercise 3.6. If $f$ is a function of $x$ and $y$, then (see figure 3.3):[34]

$$\left.\frac{\partial f}{\partial x}\right|_y \left.\frac{\partial x}{\partial y}\right|_f \left.\frac{\partial y}{\partial f}\right|_x = -1. \tag{3.37}$$

[34] Notice that this is exactly minus the result you would have derived by cancelling $\partial f$, $\partial x$, and $\partial y$ from 'numerator' and 'denominator'.

Also, it's clear that if we keep all but one variable fixed, partial derivatives are like regular derivatives so

$$\left.\frac{\partial f}{\partial x}\right|_y = 1 \Big/ \left.\frac{\partial x}{\partial f}\right|_y. \tag{3.38}$$

Using this for $S(E,V)$ and fixing $N$, we find

$$-1 = \left.\frac{\partial S}{\partial V}\right|_{E,N} \left.\frac{\partial V}{\partial E}\right|_{S,N} \left.\frac{\partial E}{\partial S}\right|_{V,N} = \frac{P}{T}\left(1 \Big/ \left.\frac{\partial E}{\partial V}\right|_{S,N}\right) T \tag{3.39}$$

so

$$\left.\frac{\partial E}{\partial V}\right|_{S,N} = -P. \tag{3.40}$$

Thus the pressure is minus the energy cost per unit volume at constant entropy. Similarly,

$$-1 = \left.\frac{\partial S}{\partial N}\right|_{E,V} \left.\frac{\partial N}{\partial E}\right|_{S,V} \left.\frac{\partial E}{\partial S}\right|_{N,V} = -\frac{\mu}{T}\left(1 \Big/ \left.\frac{\partial E}{\partial N}\right|_{S,V}\right) T \tag{3.41}$$

so

$$\left.\frac{\partial E}{\partial N}\right|_{S,V} = \mu \tag{3.42}$$

the chemical potential is the energy cost of adding a particle at constant entropy.

between the two subsystems, but also changes the energy $\Delta E$ because of the work done. Equation 3.34 and 3.35 tell us that $\Delta S = \left(\frac{1}{T_1} - \frac{1}{T_2}\right)\Delta E + \left(\frac{P_1}{T_1} - \frac{P_2}{T_2}\right)\Delta V = 0$, implying that $\Delta E/\Delta V = (1-\lambda)P_1 + \lambda P_2$ with $\lambda = \frac{1}{1-T_2/T_1}$. If we hypothesize that the maximum entropy had $P_1 \neq P_2$, we would certainly expect that $\Delta E/\Delta V$ would lie between these two pressures, corresponding to $0 < \lambda < 1$, but if $T_2$ and $T_1$ are both positive and different then either $\lambda < 0$ or $\lambda > 1$. Hence the piston must move to equalize the pressure even when the temperatures do not agree.

The chemical potential will be unfamiliar to most of those new to statistical mechanics. We can feel pressure and temperature as our bodies exchange volume with balloons and heat with coffee cups. Most of us have not had comparable tactile experience with exchanging particles.[35] Your intuition will improve as you work with chemical potentials. They are crucial to the study of chemical reactions: whether a reaction will proceed depends in part on the relative cost of the products and the reactants, measured by the differences in their chemical potentials (section 6.6). The chemical potential is also central to noninteracting quantum systems, where the number of particles in each quantum state can vary (chapter 7).

[35] Our lungs exchange oxygen and carbon dioxide, but they don't have nerve endings that measure the chemical potentials.

Our familiar notion of pressure is from mechanics: the energy of a subsystem increases as the volume decreases, as $\Delta E = -P\Delta V$. What may not be familiar is that this energy change is measured at fixed entropy. With the tools we have now, we can show explicitly that the mechanical definition of pressure is the same as the statistical mechanics definition (equation 3.35): the argument is somewhat technical, but illuminating (footnote 36 at the end of this section).

We can also give a simpler argument, using properties of the entropy that we will discuss more fully in chapter 5. A mechanical measurement of the pressure must not exchange heat with the body. Changing the volume while adding heat to keep the temperature fixed, for example, is a different measurement. The mechanical measurement must also change the volume slowly. If the volume changes fast enough that the subsystem goes out of equilibrium (typically a piston moving near the speed of sound), then the energy needed to change the volume will include the energy for generating the sound and shock waves – energies not appropriate to include in a good measurement of the pressure. We call a process *adiabatic* if it occurs without heat exchange and sufficiently slowly that the system remains in equilibrium.

Consider the system comprising the subsystem and the mechanical device pushing the piston, under a cycle $V \to V+\Delta V \to V$. Because the subsystem remains in equilibrium at all times, the process of changing the volume is completely reversible: the entropy of the system at the end is the same as that at the beginning. Since entropy can only increase (chapter 5), the entropy of the system halfway through the cycle at $V + \Delta V$ must be the same as at $V$. The mechanical instrument can be made with few moving parts, so its entropy change can be neglected. Hence the entropy of the subsystem must be unchanged under an adiabatic change in volume. Thus a mechanical measurement of pressure is done at constant entropy.

Broadly speaking, the entropy of a system changing adiabatically (slowly and in thermal isolation) will be a constant. Indeed, you may view our detailed calculation (the following footnote) as providing a statistical mechanical derivation for this important truth.[36]

[36] We want to show that our statistical mechanics definition $P = T\left.\frac{\partial S}{\partial V}\right|_{E,N}$ corresponds to the everyday mechanical definition $P = -\Delta E/\Delta V$. We first must use

statistical mechanics to find a formula for the mechanical force per unit area $P$. Consider some general liquid or gas whose volume is changed smoothly from $V$ to $V + \Delta V$, and is otherwise isolated from the rest of the world. (A solid can support a shear stress. Because of this, it has not just a pressure, but a whole stress tensor, that can vary in space ...)

We can find the mechanical pressure if we can find out how much the energy changes as the volume changes. The initial system at $t = 0$ is an equilibrium ensemble at volume $V$, uniformly filling phase space in an energy range $E < \mathcal{H} < E + \delta E$ with density $1/\Omega(E, V)$. A member of this volume-expanding ensemble is a trajectory $\mathbb{P}(t), \mathbb{Q}(t)$ that evolves in time under the changing Hamiltonian $\mathcal{H}\big(\mathbb{P}, \mathbb{Q}, V(t)\big)$. The amount this particular trajectory changes in energy under the time-dependent Hamiltonian is

$$\frac{d\mathcal{H}\big(\mathbb{P}(t), \mathbb{Q}(t), V(t)\big)}{dt} = \frac{\partial \mathcal{H}}{\partial \mathbb{P}}\dot{\mathbb{P}} + \frac{\partial \mathcal{H}}{\partial \mathbb{Q}}\dot{\mathbb{Q}} + \frac{\partial \mathcal{H}}{\partial V}\frac{dV}{dt}. \tag{3.43}$$

A Hamiltonian for particles of kinetic energy $\frac{1}{2}\mathbb{P}^2/m$ and potential energy $U(\mathbb{Q})$ will have $\frac{\partial \mathcal{H}}{\partial \mathbb{P}} = \mathbb{P}/m = \dot{\mathbb{Q}}$ and $\frac{\partial \mathcal{H}}{\partial \mathbb{Q}} = V(\mathbb{Q}) = -\dot{\mathbb{P}}$, so the first two terms cancel on the right-hand side of equation 3.43. (You may recognize Hamilton's equations of motion; indeed, the first two terms cancel for any Hamiltonian system.) Hence the energy change for this particular trajectory is

$$\frac{d\mathcal{H}\big(\mathbb{P}(t), \mathbb{Q}(t), V(t)\big)}{dt} = \frac{\partial \mathcal{H}}{\partial V}\big(\mathbb{P}, \mathbb{Q}\big)\ \frac{dV}{dt}. \tag{3.44}$$

That is, the energy change of the evolving trajectory is the same as the expectation value of $\frac{\partial \mathcal{H}}{\partial t}$ at the static current point in the trajectory: we need not follow the particles as they zoom around.

We still must average this energy change over the equilibrium ensemble of initial conditions. This is in general not possible, until we make the second assumption involved in the adiabatic measurement of pressure: we assume that the potential energy turns on so slowly that the system remains in equilibrium at the current volume $V(t)$ and energy $E(t)$. This allows us to calculate the ensemble average energy change as an equilibrium thermal average:

$$\frac{d\langle \mathcal{H} \rangle}{dt} = \left\langle \frac{\partial \mathcal{H}}{\partial V} \right\rangle_{E(t), V(t)} \frac{dV}{dt}. \tag{3.45}$$

Since this energy change must equal $-P\frac{dV}{dt}$, we find

$$-P = \left\langle \frac{\partial \mathcal{H}}{\partial V} \right\rangle = \frac{1}{\Omega(E)} \int d\mathbb{P}\, d\mathbb{Q}\, \delta(E - \mathcal{H}(\mathbb{P}, \mathbb{Q}, V)) \frac{\partial \mathcal{H}}{\partial V}. \tag{3.46}$$

We now return to calculating the derivative

$$\left.\frac{\partial S}{\partial V}\right|_{E,N} = \frac{\partial}{\partial V} k_B \log(\Omega) = \frac{k_B}{\Omega} \left.\frac{\partial \Omega}{\partial V}\right|_{E,N}. \tag{3.47}$$

Using equation 3.4 to write $\Omega$ in terms of a derivative of the $\Theta$ function, we can change orders of differentiation

$$\begin{aligned}
\left.\frac{\partial \Omega}{\partial V}\right|_{E,N} &= \left.\frac{\partial}{\partial V}\right|_{E,N} \left.\frac{\partial}{\partial E}\right|_{V,N} \int d\mathbb{P}\, d\mathbb{Q}\, \Theta(E - \mathcal{H}(\mathbb{P}, \mathbb{Q}, V)) \\
&= \left.\frac{\partial}{\partial E}\right|_{V,N} \int d\mathbb{P}\, d\mathbb{Q}\, \frac{\partial}{\partial V}\, \Theta(E - \mathcal{H}(\mathbb{P}, \mathbb{Q}, V)) \\
&= -\left.\frac{\partial}{\partial E}\right|_{V,N} \int d\mathbb{P}\, d\mathbb{Q}\, \delta(E - \mathcal{H}(\mathbb{P}, \mathbb{Q}, V)) \frac{\partial \mathcal{H}}{\partial V}
\end{aligned} \tag{3.48}$$

But the phase–space integral in the last equation is precisely the same integral that appears in our formula for the pressure, equation 3.46: it is $\Omega(E)(-P)$. Thus

$$\begin{aligned}
\left.\frac{\partial \Omega}{\partial V}\right|_{E,N} &= \left.\frac{\partial}{\partial E}\right|_{V,N} (\Omega(E) P) \\
&= \left.\frac{\partial \Omega}{\partial E}\right|_{V,N} P + \Omega \left.\frac{\partial P}{\partial E}\right|_{V,N}
\end{aligned} \tag{3.49}$$

## 3.5 Entropy, the Ideal Gas, and Phase Space Refinements

Let's find the temperature and pressure for the ideal gas, using our microcanonical ensemble. We'll then introduce two subtle refinements to the phase space volume (one from quantum mechanics, and one for indistinguishable particles) which will not affect the temperature or pressure, but will be important for the entropy and chemical potential.

We derived the volume $\Omega(E)$ of the energy shell in phase space in section 3.2: it factored[37] into a momentum space volume from equation 3.15 and a configuration space volume $V^N$. Before our refinements, we have:

$$\begin{aligned}\Omega_{\text{crude}}(E) &= V^N \left(\frac{3N}{2E}\right) \pi^{\frac{3N}{2}} (2mE)^{\frac{3N}{2}} / \tfrac{3N}{2}! \\ &\approx V^N \pi^{\frac{3N}{2}} (2mE)^{\frac{3N}{2}} / \tfrac{3N}{2}!\end{aligned} \tag{3.52}$$

[37] It factors only because the potential energy is zero.

Notice that in the second line of 3.52 we have dropped the term $3N/2E$: it divides the phase space volume by a negligible factor (two-thirds the energy per particle).[38] The entropy and its derivatives are (before our

[38] Multiplying $\Omega(E)$ by a factor independent of the number of particles is equivalent to adding a constant to the entropy. The entropy of a typical system is so large (of order Avogadro's number times $k_B$) that adding a number-independent constant to it is irrelevant. Notice that this implies that $\Omega(E)$ is so large that *multiplying* it by a constant doesn't significantly change its value.

so

$$\begin{aligned}\left.\frac{\partial S}{\partial V}\right|_{E,N} &= \frac{\partial}{\partial V} k_B \log(\Omega) = \frac{k_B}{\Omega}\left(\left.\frac{\partial \Omega}{\partial E}\right|_{V,N} P + \Omega \left.\frac{\partial P}{\partial E}\right|_{V,N}\right) \\ &= \left.\frac{\partial k_B \log(\Omega)}{\partial E}\right|_{V,N} P + k_B \left.\frac{\partial P}{\partial E}\right|_{V,N} = \left.\frac{\partial S}{\partial E}\right|_{V,N} P + k_B \left.\frac{\partial P}{\partial E}\right|_{V,N} \\ &= P/T + k_B \left.\frac{\partial P}{\partial E}\right|_{V,N}\end{aligned} \tag{3.50}$$

Now, $P$ and $T$ are both intensive variables, but $E$ is extensive (scales linearly with system size). Hence $P/T$ is of order one for a large system, but $k_B \frac{\partial P}{\partial E}$ is of order $1/N$ where $N$ is the number of particles. (For example, we shall see that for the ideal gas, $PV = 2/3E = Nk_BT$, so $k_B \frac{\partial P}{\partial E} = \frac{2k_B}{3V} = \frac{2P}{3NT} = \frac{2}{3}\frac{P}{NT} \ll \frac{P}{T}$ for large $N$.) Hence the second term, for a large system, may be neglected, giving us the desired relation

$$\left.\frac{\partial S}{\partial V}\right|_{E,N} = P/T. \tag{3.51}$$

The derivative of the entropy $S(E, V, N)$ with respect to $V$ at constant $E$ and $N$ is thus indeed the mechanical pressure divided by the temperature. Adiabatic measurements (slow and without heat exchange) keep the entropy unchanged.

refinements)

$$S_{\text{crude}}(E) = k_B \log\left(V^N \pi^{\frac{3N}{2}} (2mE)^{\frac{3N}{2}} / \tfrac{3N}{2}!\right)$$
$$= Nk_B \log(V) + \tfrac{3}{2}Nk_B \log(2\pi mE) - k_B \log\left(\frac{3N}{2}!\right) \quad (3.53)$$
$$\frac{1}{T} = \left.\frac{\partial S}{\partial E}\right|_{V,N} = \frac{3Nk_B}{2E} \quad (3.54)$$
$$\frac{P}{T} = \left.\frac{\partial S}{\partial V}\right|_{E,N} = \frac{Nk_B}{V} \quad (3.55)$$
$$(3.56)$$

so the temperature and pressure are

$$k_B T = \frac{2E}{3N} \quad \text{and} \quad (3.57)$$
$$PV = Nk_B T. \quad (3.58)$$

The first line above is the temperature formula we promised in forming equation 3.21: the ideal gas has energy equal to $\frac{1}{2}k_BT$ per component of the velocity.[39]

The second formula is the *equation of state*[40] for the ideal gas. The equation of state is the relation between the macroscopic variables of an equilibrium system that emerges in the limit of large numbers of particles. The pressure $P(T, V, N)$ in an ideal gas will fluctuate around the value $Nk_BT/V$ given by the equation of state, with the magnitude of the fluctuations vanishing as the system size gets large.

In general, our definition for the energy shell volume in phase space needs two refinements. First, the phase space volume has units of ([length][momentum])$^{3N}$: the volume of the energy shell depends multiplicatively upon the units chosen for length, mass, and time. Changing these units will change the corresponding crude form for the entropy by a constant times $3N$. Most physical properties, like temperature and pressure above, are dependent only on derivatives of the entropy, so the overall constant won't matter: indeed, the zero of the entropy is undefined within classical mechanics. It is suggestive that [length][momentum] has units of Planck's constant $h$, and we shall see in chapter 7 that quantum mechanics in fact does set the zero of the entropy. We shall see in exercise 7.1 that dividing[41] $\Omega(E)$ by $h^{3N}$ nicely sets the entropy density to zero in equilibrium quantum systems at absolute zero.

Second, there is an important subtlety in quantum physics regarding *identical* particles. Two electrons, or two Helium atoms of the same isotope, are not just hard to tell apart: they really are completely and utterly the same (figure 7.3). We shall see in section 7.3 that the proper quantum treatment of identical particles involves averaging over possible states using Bose and Fermi statistics.

In classical physics, there is an analogous subtlety regarding *indistinguishable*[42] particles. For a system of two indistinguishable particles, the phase space points $(\mathbf{p}_A, \mathbf{p}_B, \mathbf{q}_A, \mathbf{q}_B)$ and $(\mathbf{p}_B, \mathbf{p}_A, \mathbf{q}_B, \mathbf{q}_A)$ should

[39] Since $k_BT = 2E/3N$, this means each particle on average has its share $E/N$ of the total energy, as it must.

[40] It is rare that the 'equation of state' can be written out as an explicit equation! Only in special cases (*e.g.*, noninteracting systems like the ideal gas) can one solve in closed form for the thermodynamic potentials, equations of state, or other properties.

[41] This is equivalent to using units for which $h = 1$.

[42] If we have particles that in principle are not identical, but our Hamiltonian and measurement instruments do not distinguish between them, then in classical statistical mechanics we may treat them with Maxwell-Boltzmann statistics as well: they are indistinguishable but not identical.

not both be counted: the volume of phase space $\Omega(E)$ should be half that given by a calculation for distinguishable particles. For $N$ indistinguishable particles, the phase space volume should be divided by $N!$, the total number of ways the labels for the particles can be permuted.

Unlike the introduction of the factor $h^{3N}$ above, dividing the phase space volume by $N!$ does change the predictions of statistical mechanics in important ways. We will see in section 5.2.1 that the entropy increase for joining containers of different kinds of particles should be substantial, while the entropy increase for joining containers filled with indistinguishable particles should be near zero. This result is correctly treated by dividing $\Omega(E)$ by $N!$ for each set of $N$ indistinguishable particles. We call the resulting ensemble *Maxwell-Boltzmann* statistics, to distinguish it from distinguishable statistics and from the quantum-mechanical Bose and Fermi statistics. We shall see in chapter 7 that identical fermions and bosons obey Maxwell-Boltzmann statistics at high temperatures – they become classical, but remain indistinguishable.

Combining these two refinements gives us for the ideal gas

$$\Omega(E) = (V^N/N!)\,(\pi^{\frac{3N}{2}}(2mE)^{\frac{3N}{2}}/\tfrac{3N}{2}!)\,(1/h)^{3N}. \tag{3.59}$$

$$S(E) = Nk_B \log\left[\frac{V}{h^3}(2\pi mE)^{3/2}\right] - k_B \log(N!\frac{3N}{2}!). \tag{3.60}$$

We can make our equation for the entropy more useful by using Stirling's formula $\log(N!) \approx N\log N - N$, valid at large $N$.

$$S(E,V,N) = \frac{5}{2}Nk_B + Nk_B \log\left[\frac{V}{Nh^3}\left(\frac{4\pi mE}{3N}\right)^{3/2}\right] \tag{3.61}$$

This is the standard formula for the entropy of an ideal gas. We can put it into a somewhat simpler form by writing it in terms of the particle density $\rho = N/V$

$$S = Nk_B\left(\frac{5}{2} - \log(\rho\lambda^3)\right). \tag{3.62}$$

where[43]

$$\lambda = h/\sqrt{4\pi mE/3N} \tag{3.63}$$

is called the *thermal de Broglie wavelength*, and will be physically significant for quantum systems at low temperature (chapter 7).

[43] de Broglie realized that matter could act as a wave: a particle of momentum $p$ had a wavelength $\lambda_{\rm quantum} = h/p$. The mean-square momentum of a particle in our gas is $p^2 = 2m(E/N)$, and would have a quantum wavelength of $h/\sqrt{2mE/3N} = \lambda_{\rm thermal}/\sqrt{2\pi/3}$ – close enough that we give de Broglie's name to $\lambda$ too.

## Exercises

Exercise 3.1 is the classic problem of planetary atmospheres. Exercise 3.3 is a nice generalization of the ideal gas law. Part (a) of exercise 3.4 is a workout in $\delta$-functions; parts (b) and (c) calculate the energy fluctuations for a mixture of two ideal gases, and could be assigned separately. Exercise 3.5 extends the calculation

of the density fluctuations from two subvolumes to $K$ subvolumes, and introduces the *Poisson* distribution. Finally, exercise 3.6 introduces some of the tricky partial derivative relations in thermodynamics (the triple product of equation 3.37 and the Maxwell relations) and applys them to the ideal gas.

(3.1) **Escape Velocity.** (Basic)

*Assuming the probability distribution for the z component of momentum given in equation 3.21,* $\rho(p_z) = 1/\sqrt{2\pi m k_B T}\, \exp(-{p_z}^2/2mk_BT)$*, give the probability density that an* $N_2$ *molecule will have a vertical component of the velocity equal to the escape velocity from the Earth (about 10 km/sec, if I remember right). Do we need to worry about losing our atmosphere?* Optional: Try the same calculation for $H_2$, where you'll find a substantial leakage.
(Hint:You'll want to know that there are about $\pi \times 10^7$ seconds in a year, and molecules collide (and scramble their velocities) many times per second. That's why Jupiter has hydrogen gas in its atmosphere, and Earth does not.)

(3.2) **Temperature and Energy.** (Basic)

What units [joules, millijoules, microjoules, nanojoules, ..., yoctojoules ($10^{-24}$ joules)] would we use to measure temperature if we used energy units instead of introducing Boltzmann's constant $k_B = 1.3807 \times 10^{-23}$ J/K?

(3.3) **Hard Sphere Gas** (Basic)

We can improve on the realism of the ideal gas by giving the atoms a small radius. If we make the potential energy infinite inside this radius ("hard spheres"), the potential energy is simple (zero unless the spheres overlap, which is forbidden). Let's do this in two dimensions.

A two dimensional $L \times L$ box with hard walls contains a gas of $N$ hard disks of radius $r \ll L$ (figure 3.5). The disks are dilute: the summed area $N\pi r^2 \ll L^2$. Let $A$ be the effective volume allowed for the disks in the box: $A = (L - 2r)^2$.

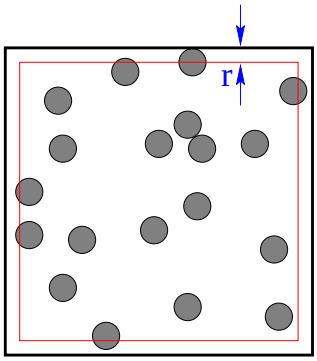



**Fig. 3.5** Hard Sphere Gas.

*(a) The area allowed for the second disk is* $A - \pi(2r)^2$ *(figure 3.6), ignoring the small correction when the excluded region around the first disk overlaps the excluded region near the walls of the box. What is the allowed* $2N$*-dimensional volume in configuration space, of allowed zero-energy configurations of hard disks, in this dilute limit? Ignore small corrections when the excluded region around one disk overlaps the excluded regions around other disks, or near the walls of the box. Remember the* $1/N!$ *correction for identical particles. Leave your answer as a product of* $N$ *terms.*

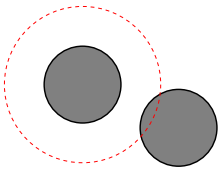


**Fig. 3.6** Excluded volume around a sphere.

*(b) What is the configurational entropy for the hard disks? Here, simplify your answer so that it does not involve a sum over* $N$ *terms, but valid to first order in the area of the disks* $\pi r^2$*. Show, for large* $N$*, that it is well approximated by* $S_{\mathbb{Q}} = Nk_B(1 + \log(A/N - b))$*, with* $b$ *representing the effective excluded area due to the other disks. (You may want to derive the formula* $\sum_{n=1}^{N} \log(A - (n-1)\epsilon) = N \log(A - (N-1)\epsilon/2) + O(\epsilon^2)$*.) What is the value of* $b$*, in terms of the area of the disk?*

*(c) Find the pressure for the hard-sphere gas in the large* $N$ *approximation of part (b). Does it reduce to the ideal gas law for* $b = 0$*?*

(3.4) **Connecting Two Macroscopic Systems.**

An isolated system with energy $E$ is composed of two macroscopic subsystems, each of fixed volume $V$ and number of particles $N$. The subsystems are weakly coupled, so the sum of their energies is $E_1 + E_2 = E$ (figure 3.4 with only the energy door open). We can use the Dirac delta function $\delta(x)$ to define the volume of the energy surface of a system with Hamiltonian $\mathcal{H}$ to be

$$\Omega(E) = \int d\mathbb{P}\, d\mathbb{Q}\, \delta\left(E - \mathcal{H}(\mathbb{P},\mathbb{Q})\right) \tag{3.64}$$

$$= \int d\mathbb{P}_1 d\mathbb{Q}_1 d\mathbb{P}_2 d\mathbb{Q}_2 \tag{3.65}$$
$$\delta\left(E - (\mathcal{H}_1(\mathbb{P}_1,\mathbb{Q}_1) + \mathcal{H}_2(\mathbb{P}_2,\mathbb{Q}_2))\right).$$

*(a) Derive formula 3.23 for the volume of the energy surface of the whole system using $\delta$-functions.* (Hint: Insert $\int \delta(E_1 - \mathcal{H}_1(\mathbb{P}_1,\mathbb{Q}_1))\, dE_1 = 1$ into equation 3.65.)

Consider a monatomic ideal gas (He) mixed with a diatomic ideal gas ($H_2$). We showed that a monatomic ideal gas of $N$ atoms has $\Omega_1(E_1) \propto E_1^{3N/2}$. A diatomic molecule has $\Omega_2(E_2) \propto E_2^{5N/2}$.[44]

*(b) Argue that the probability density of system 1 being at energy $E_1$ is the integrand of 3.23 divided by the whole integral, equation 3.24. For these two gases, which energy $E_1^{max}$ has the maximum probability?*

*(c) Use the saddle-point method [71, sect. 3.6] to approximate the integral 3.65 as the integral over a Gaussian. (That is, put the integrand into the form $\exp(f(E_1))$ and Taylor expand $f(E_1)$ to second order in $E_1 - E_1^{max}$.) Use the saddle-point integrand as a Gaussian approximation for the probability density $\rho(E_1)$ (valid, for large $N$, whenever $\rho(E_1)$ isn't absurdly small). In this approximation, what is the mean energy $\langle E_1\rangle$? What are the energy fluctuations per particle $\sqrt{\langle (E_1 - E_1^{max})^2\rangle}/N$?*

For subsystems with large numbers of particles $N$, temperature and energy density are well defined because $\Omega(E)$ for each subsystem grows extremely rapidly with increasing energy, in such a way that $\Omega_1(E_1)\Omega_2(E - E_1)$ is sharply peaked near its maximum.

(3.5) **Gauss and Poisson.** (Basic)

In section 3.2.1, we calculated the probability distribution for having $n = N_0 + m$ particles on the right–hand half of a box of volume $2V$ with $2N_0$ total particles. In section 10.3 we will want to know the number fluctuations of a small subvolume in an infinite system. Studying this also introduces the *Poisson distribution*.

Let's calculate the probability of having $n$ particles in a subvolume $V$, for a box with total volume $KV$ and a total number of particles $T = KN_0$. For $K = 2$ we will derive our previous result, equation 3.13, including the prefactor. As $K \to \infty$ we will derive the infinite volume result.

*(a) Find the exact formula for this probability: $n$ particles in $V$, with total of $T$ particles in $KV$.* (Hint: What is the probability that the *first* $n$ particles fall in the subvolume $V$, and the remainder $T - n$ fall outside the subvolume $(K-1)V$? How many ways are there to pick $n$ particles from $T$ total particles?)

The Poisson probability distribution

$$\rho_n = a^n e^{-a}/n! \tag{3.66}$$

arises in many applications. It arises whenever there is a large number of possible events $T$ each with a small probability $a/T$; the number of cars passing a given point during an hour on a mostly empty street, the number of cosmic rays hitting in a given second, etc.

*(b) Show that the Poisson distribution is normalized: $\sum_n \rho_n = 1$. Calculate the mean of the distribution $\langle n\rangle$ in terms of $a$. Calculate the variance (standard deviation squared) $\left\langle (n - \langle n\rangle)^2\right\rangle$.*

*(c) As $K \to \infty$, show that the probability that $n$ particles fall in the subvolume $V$ has the Poisson distribution 3.66. What is $a$?* (Hint: You'll need to use the fact that $e^{-a} = (e^{-1/K})^{Ka} \to (1 - 1/K)^{Ka}$ as $K \to \infty$, and the fact that $n \ll T$. Here don't assume that $n$ is large: the Poisson distribution is valid even if there are only a few events.)

From parts (b) and (c), you should be able to conclude that the variance in the number of particles found in a volume $V$ inside an infinite system should be equal to $N_0$, the expected number of particles in the volume:

$$\left\langle (n - \langle n\rangle)^2\right\rangle = N_0. \tag{3.67}$$

This is twice the squared fluctuations we found for the case where the volume $V$ was half of the total volume, equation 3.13. That makes sense, since the particles can fluctuate more freely in an infinite volume than in a doubled volume.

If $N_0$ is large, the probability $P_m$ that $N_0 + m$ particles lie inside our volume will be Gaussian for any $K$. (As a special case, if $a$ is large the Poisson distribution is well approximated as a Gaussian.) Let's derive this distribution for all $K$. First, as in section 3.2.1, let's use the weak

[44] In the range $\hbar^2/2I \ll k_B T \ll \hbar\omega$, where $\omega$ is the vibrational frequency of the stretch mode and $I$ is the moment of inertia. The lower limit makes the rotations classical; the upper limit freezes out the vibrations, leaving us with three classical translation modes and two rotational modes – a total of five degrees of freedom.

form of Stirling's approximation, equation 3.10 dropping the square root: $n! \sim (n/e)^n$.

*(d) Using your result from part (a), write the exact formula for* $\log(P_m)$*. Apply the weak form of Stirling's formula. Expand your result around* $m = 0$ *to second order in* $m$*, and show that* $\log(P_m) \approx -m^2/2\sigma_K^2$*, giving a Gaussian form*

$$P_m \sim e^{-m^2/2\sigma_K^2}. \tag{3.68}$$

*What is* $\sigma_K$*? In particular, what is* $\sigma_2$ *and* $\sigma_\infty$*?* Your result for $\sigma_2$ should agree with the calculation in section 3.2.1, and your result for $\sigma_\infty$ should agree with equation 3.67.

Finally, we should address the normalization of the Gaussian. Notice that the ratio of the strong and weak forms of Stirling's formula, (equation 3.10) is $\sqrt{2\pi n}$. We need to use this to produce the normalization $\frac{1}{\sqrt{2\pi}\sigma_K}$ of our Gaussian.

*(e) In terms of* $T$ *and* $n$*, what factor would the square–root term have contributed if you had kept it in Stirling's formula going from part (a) to part (d)?* (It should look like a ratio involving three terms like $\sqrt{2\pi X}$.) *Show from equation 3.68 that the fluctuations are small,* $m = n - N_0 \ll N_0$ *for large* $N_0$*. Ignoring these fluctuations, set* $n = N_0$ *in your factor, and give the prefactor multiplying the Gaussian in equation 3.68.* (Hint: your answer should be normalized.)

(3.6) **Microcanonical Thermodynamics** (Thermodynamics, Chemistry)

Thermodynamics was understood as an almost complete scientific discipline before statistical mechanics was invented. Stat mech can be thought of as the "microscopic" theory, which yields thermo as the "emergent" theory on long length and time scales where the fluctuations are unimportant.

The microcanonical stat mech distribution introduced in class studies the properties at fixed total energy $E$, volume $V$, and number of particles $N$. We derived the microscopic formula $S(N, V, E) = k_B \log \Omega(N, V, E)$. The principle that entropy is maximal led us to the conclusion that two weakly-coupled systems in thermal equilibrium would exchange energy until their values of $\frac{\partial S}{\partial E}|_{N,V}$ agreed, leading us to define the latter as the inverse of the temperature. By an analogous argument we find that systems that can exchange volume (by a thermally insulated movable partition) will shift until $\frac{\partial S}{\partial V}|_{N,E}$ agrees, and that systems that can exchange particles (by semipermeable membranes) will shift until $\frac{\partial S}{\partial N}|_{V,E}$ agrees.

How do we connect these statements with the definitions of pressure and chemical potential we get from thermodynamics? In thermo, one defines the pressure as minus the change in energy with volume $P = -\frac{\partial E}{\partial V}|_{N,S}$, and the chemical potential as the change in energy with number of particles $\mu = \frac{\partial E}{\partial N}|_{V,S}$; the total internal energy satisfies

$$dE = T\,dS - P\,dV + \mu\,dN. \tag{3.69}$$

*(a) Show by solving equation 3.69 for* $dS$ *that* $\frac{\partial S}{\partial V}|_{N,E} = P/T$ *and* $\frac{\partial S}{\partial N}|_{V,E} = -\mu/T$.

I've always been uncomfortable with manipulating $dX$'s. Let's do this the hard way. Our "microcanonical" equation of state $S(N, V, E)$ can be thought of as a surface embedded in four dimensions.

*(b) Show that, if* $f$ *is a function of* $x$ *and* $y$*, that* $\frac{\partial x}{\partial y}|_f \frac{\partial y}{\partial f}|_x \frac{\partial f}{\partial x}|_y = -1$*. (Draw a picture of a surface* $f(x, y)$ *and a triangular path with three curves at constant* $f$*,* $x$*, and* $y$ *as in figure 3.3. Specifically, draw a path that starts at* $(x_0, y_0, f_0)$ *and moves along a contour at constant* $f$ *to* $y_0 + \Delta y$*. The final point will be at* $(x_0 + \frac{\partial x}{\partial y}|_f \Delta y, y_0 + \Delta y, f_0)$*. Draw it at constant* $x$ *back to* $y_0$*, and then at constant* $y$ *back to* $(x_0, y_0, f_0)$*. How much must* $f$ *change to make this a single-valued function?) Applying this formula to* $S$ *at fixed* $E$*, derive the two equations in part (a) again.*

*(c)* **Ideal Gas Thermodynamics.** *Using the microscopic formula for the entropy of a monatomic ideal gas (from 3.61)*

$$S(N, V, E) = \frac{5}{2} N k_B + N k_B \log \left[ \frac{V}{N h^3} \left( \frac{4\pi m E}{3N} \right)^{3/2} \right], \tag{3.70}$$

*calculate* $\mu$*.*

**Maxwell Relations.** Imagine solving the microcanonical equation of state of some material (not necessarily an ideal gas) for the energy $E(S, V, N)$: it's the same surface in four dimensions, but looked at with a different direction pointing "up". One knows that the second derivatives of $E$ are symmetric: at fixed $N$, we get the same answer whichever order we take derivatives with respect to $S$ and $V$.

*(d) Use this to show the Maxwell relation*

$$\left.\frac{\partial T}{\partial V}\right|_{S,N} = -\left.\frac{\partial P}{\partial S}\right|_{V,N}. \tag{3.71}$$

*(This should take about two lines of calculus). Generate two other similar formulas by taking other second partial derivatives of* $E$*. There are many of these relations.*

*(e)* **Stat Mech check of the Maxwell relation.** *Using equation 3.70, write formulas for* $E(S, V, N)$*,* $T(S, V, N)$ *and* $P(S, V, N)$ *for the ideal gas (non trivial!). (This is different from* $T$ *and* $P$ *in part (c), which were functions of* $N$*,* $V$*, and* $E$*.) Show explicitly that the Maxwell relation equation 3.71 is satisfied.*

(3.7) **Microcanonical Energy Fluctuations.** (Basic)

We argued in section 3.3 that the energy fluctuations between two weakly coupled subsystems are of order $\sqrt{N}$. Let us calculate them explicitly.

Equation 3.30 showed that for two subsystems with energy $E_1$ and $E_2 = E - E_1$ the probability density of $E_1$ is a Gaussian with variance (standard deviation squared):

$$\sigma_{E_1}^2 = -k_B / \left(\partial^2 S_1/\partial E_1^2 + \partial^2 S_2/\partial E_2^2\right). \tag{3.72}$$

*(a) Show that*

$$\frac{1}{k_B}\frac{\partial^2 S}{\partial E^2} = -\frac{1}{k_B T}\frac{1}{N c_v T}. \tag{3.73}$$

*where $c_v$ is the inverse of the total specific heat at constant volume.* (The specific heat $c_v$ is the energy needed per particle to change the temperature by one unit: $N c_v = \frac{\partial E}{\partial T}\big|_{V,N}$.)

The denominator of equation 3.73 is the product of two energies. The second term $N c_v T$ is a system-scale energy: it is the total energy that would be needed to raise the temperature of the system from absolute zero, if the specific heat per particle $c_v$ were temperature independent. However, the first energy, $k_B T$, is an atomic-scale energy independent of $N$. The fluctuations in energy, therefore, scale like the geometric mean of the two, summed over the two subsystems in equation 3.30, and hence scale as $\sqrt{N}$: the total energy fluctuations per particle thus are roughly $1/\sqrt{N}$ times a typical energy per particle.

This formula is quite remarkable. Normally, to measure a specific heat one would add a small energy and watch the temperature change. This formula allows us to measure the specific heat of an object by watching the equilibrium fluctuations in the energy. These fluctuations are tiny in most experiments, but can be quite substantial in computer simulations.

*(b) If $c_v^{(1)}$ and $c_v^{(2)}$ are the specific heats per particle for the two subsystems, show using equation 3.72 and 3.73 that*

$$\frac{1}{c_v^{(1)}} + \frac{1}{c_v^{(2)}} = N k_B T^2 / \sigma_E^2. \tag{3.74}$$

We don't even need to couple two systems. The positions and momenta of a molecular dynamics simulation (atoms moving under Newton's laws of motion) can be thought of as two subsystems, since the kinetic energy does not depend on the configuration $\mathbb{Q}$, and the potential energy does not depend on the momenta $\mathbb{P}$. Let $E_1$ be the kinetic energy of a microcanonical molecular dynamics simulation with total energy $E$. Assume that the simulation is done inside a box with hard walls, so that total momentum and angular momentum are not conserved.[45]

*(c) Using the equipartition theorem, write the temperature in terms of the mean kinetic energy of the system. Show that $c_v^{(1)} = 3k_B/2$ for the momentum degrees of freedom. In terms of the mean and standard deviation of the kinetic energy, solve for the total specific heat of the molecular dynamics simulation (configurational plus kinetic).*

[45] If one uses periodic boundary conditions and the simulation is done with zero center-of-mass momentum, then the total specific heat for the momentum degrees of freedom is $N c_v^{(1)} = (N-3)k_B$, since the three center-of-mass degrees of freedom do not equilibrate with the rest of the system.

# Phase Space Dynamics and Ergodicity

# 4

So far, our justification for using the microcanonical ensemble was simple ignorance: all we know about the late time dynamics is that energy must be conserved, so we average over all states of fixed energy. Here we provide a much more convincing argument for the ensemble, and hence for equilibrium statistical mechanics as a whole. In section 4.1 we'll show for classical systems that averaging over the energy surface is consistent with time evolution: Liouville's theorem tells us that volume in phase space is conserved, so the trajectories only stir the energy surface around, they do not change the relative weights of different parts of the energy surface. In section 4.2 we introduce the concept of *ergodicity*: an ergodic system has an energy surface which is well stirred. Using Liouville's theorem and assuming ergodicity will allow us to show[1] that the microcanonical ensemble average gives the long-time average behavior that we call equilibrium.

[1] We do not aspire to rigor, but we will provide physical arguments for rigorously known results: see [63].

## 4.1 Liouville's Theorem

In chapter 3, we saw that treating all states in phase space with a given energy on an equal footing gave sensible predictions for the ideal gas, but we did not show that this democratic treatment was necessarily the correct one. Liouville's theorem, true for all Hamiltonian systems, will tell us that all states are created equal.

Systems of point particles obeying Newton's laws without dissipation are examples of Hamiltonian dynamical systems. Hamiltonian systems conserve energy. The Hamiltonian is the function $\mathcal{H}(\mathbb{P}, \mathbb{Q})$ that gives the energy for the system for any point in phase space: the equations of motion are given by

$$\begin{aligned} \dot{q}_\alpha &= \partial\mathcal{H}/\partial p_\alpha \\ \dot{p}_\alpha &= -\partial\mathcal{H}/\partial q_\alpha. \end{aligned} \tag{4.1}$$

where as usual $\dot{X} = \partial X/\partial t$. The standard example of a Hamiltonian, and the only example we will discuss in this text, is a bunch of particles interacting with a potential energy $V$:

$$\mathcal{H}(\mathbb{P}, \mathbb{Q}) = \sum_\alpha {p_\alpha}^2/2m_\alpha + V(q_1, \ldots, q_{3N}). \tag{4.2}$$

In this case, one immediately finds the expected equations of motion

$$\begin{aligned}\dot{q}_\alpha &= \partial\mathcal{H}/\partial p_\alpha = p_\alpha/m_\alpha \\ \dot{p}_\alpha &= -\partial\mathcal{H}/\partial q_\alpha = -\partial V/\partial q_\alpha = f_\alpha(q_1,\ldots,q_{3N}).\end{aligned} \tag{4.3}$$

where $f_\alpha$ is the force on coordinate $\alpha$. More general Hamiltonians[2] arise when studying, for example, the motions of rigid bodies or mechanical objects connected by hinges and joints, where the natural variables are angles or relative positions rather than points in space. Hamiltonians also play a central role in quantum mechanics.[3]

Hamiltonian systems have properties that are quite distinct from general systems of differential equations. Not only do they conserve energy: they also have many other unusual properties.[4] Liouville's theorem describes the most important of these properties.

Consider the evolution law for a general probability density in phase space

$$\rho(\mathbb{P},\mathbb{Q}) = \rho(q_1,...,q_{3N},p_1,...p_{3N}). \tag{4.4}$$

(As a special case, the microcanonical ensemble has $\rho$ equal to a constant in a thin range of energies, and zero outside that range.) This probability density $\rho$ is *locally conserved*: probability cannot be created or destroyed, it can only flow around in phase space. As an analogy, suppose a fluid of mass density $\rho_{3D}(x)$ in three dimensions has a velocity $v(x)$. Because mass density is locally conserved, $\rho_{3D}$ must satisfy the continuity equation $\partial\rho_{3D}/\partial t = -\nabla\cdot\mathbf{J}$, where $\mathbf{J} = \rho_{3D}\mathbf{v}$ is the mass current (figure 4.1). In the same way, the probability density in $6N$ dimensions has a phase-space probability current $(\rho\,\dot{\mathbb{P}}, \rho\,\dot{\mathbb{Q}})$ and hence satisfies a continuity equation

$$\begin{aligned}\frac{\partial\rho}{\partial t} &= -\sum_{\alpha=1}^{3N}\left(\frac{\partial(\rho\dot{q}_\alpha)}{\partial q_\alpha} + \frac{\partial(\rho\dot{p}_\alpha)}{\partial p_\alpha}\right) \\ &= -\sum_{\alpha=1}^{3N}\left(\frac{\partial\rho}{\partial q_\alpha}\dot{q}_\alpha + \rho\frac{\partial\dot{q}_\alpha}{\partial q_\alpha} + \frac{\partial\rho}{\partial p_\alpha}\dot{p}_\alpha + \rho\frac{\partial\dot{p}_\alpha}{\partial p_\alpha}\right)\end{aligned} \tag{4.5}$$

Now, it's clear what is meant by $\partial\rho/\partial q_\alpha$, since $\rho$ is a function of the $q_\alpha$'s and $p_\alpha$'s. But what is meant by $\partial\dot{q}_\alpha/\partial q_\alpha$? For our example of point particles, $\dot{q}_\alpha = p_\alpha/m$, which has no dependence on $q_\alpha$; nor does $\dot{p}_\alpha = f_\alpha(q_1,\ldots,q_{3N})$ have any dependence on the momentum $p_\alpha$.[5] Hence these two mysterious terms in equation 4.5 both vanish for Newton's laws for point particles. Indeed, in a general Hamiltonian system, using equation 4.1, we find that they cancel:

$$\begin{aligned}\partial\dot{q}_\alpha/\partial q_\alpha &= \partial(\partial\mathcal{H}/\partial p_\alpha)/\partial q_\alpha = \partial^2\mathcal{H}/\partial p_\alpha\partial q_\alpha = \partial^2\mathcal{H}/\partial q_\alpha\partial p_\alpha \\ &= \partial(\partial\mathcal{H}/\partial q_\alpha)/\partial p_\alpha = \partial(-\dot{p}_\alpha)/\partial p_\alpha = -\partial\dot{p}_\alpha/\partial p_\alpha.\end{aligned} \tag{4.6}$$

[2] You'll cover Hamiltonian dynamics in detail in most advanced courses in classical mechanics. For those who don't already know about Hamiltonians, rest assured that we won't use anything other than the special case of Newton's laws for point particles: you can safely ignore the more general case for our purposes.

[3] In section 7.1 we discuss the quantum version of Liouville's theorem.

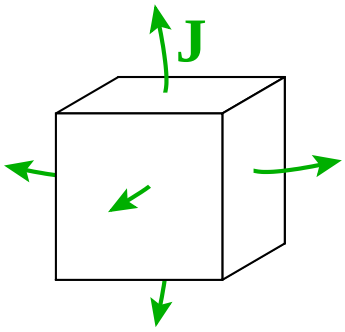



**Fig. 4.1 Conserved Current.** Think of the flow in and out of a small volume $\Delta V$ in space. The change in the density inside the volume $\partial\rho_{3D}/\partial t\,\Delta V$ must equal minus the flow of material out through the surface $-\int J\cdot dS$, which by Gauss' theorem equals $-\int\nabla\cdot J\,dV \sim -\nabla\cdot J\,\Delta V$.

[5] It would typically generally depend on the coordinate $q_\alpha$, for example.

[4] For the mathematically sophisticated reader, Hamiltonian dynamics preserves a *symplectic form* $\omega = dq_1\wedge dp_1 + \cdots + dq_{3N}\wedge dp_{3N}$: Liouville's theorem follows because the volume in phase space is $\omega^{3N}$.

This leaves us with the equation

$$\partial\rho/\partial t + \sum_{\alpha=1}^{3N} \frac{\partial\rho}{\partial q_\alpha}\dot{q}_\alpha + \frac{\partial\rho}{\partial p_\alpha}\dot{p}_\alpha = d\rho/dt = 0. \tag{4.7}$$

This is Liouville's theorem.

What is $d\rho/dt$, and how is it different from $\partial\rho/\partial t$? The former is called the *total derivative* of $\rho$ with respect to time: it's the evolution of $\rho$ seen by a particle moving with the flow. In a three dimensional flow, $d\rho_{3D}/dt = \partial\rho/\partial t + \mathbf{v}\cdot\nabla\rho = \frac{\partial\rho}{\partial t} + \sum_{i=1}^{3} \dot{x}_i \frac{\partial\rho}{\partial x_i}$; the first term is the change in $\rho$ due to the time evolution at fixed position, and the second is the change in $\rho$ that a particle moving with velocity $\mathbf{v}$ would see if the $\rho$ field didn't change in time. Equation 4.7 is the same physical situation, but in $6N$-dimensional phase space.

What does Liouville's theorem, $d\rho/dt = 0$, tell us about Hamiltonian dynamics?

- **Flows in phase space are incompressible.** In fluid mechanics, if the density $d\rho_{3D}/dt = 0$ it means that the fluid is incompressible. The density of a small element of fluid doesn't change as it moves around in the fluid: hence the small element is not compressing or expanding. In Liouville's theorem, it means the same thing: a small volume in phase space will evolve into a new shape, perhaps stretched, twisted, or folded, but with exactly the same volume.
- **There are no attractors.** In other dynamical systems, most states of the system are usually transient, and the system settles down onto a small set of states called the *attractor*. A damped pendulum will stop moving: the attractor has zero velocity and vertical angle (exercise 4.1). A forced, damped pendulum will settle down to oscillate with a particular amplitude: the attractor is a circle in phase space. The decay of these transients would seem closely related to equilibration in statistical mechanics, where at long times all initial states of a system will settle down into boring static equilibrium behavior.[6] Perversely, we've just proven that equilibration in statistical mechanics happens by a completely different mechanism! In equilibrium statistical mechanics all states are created equal: transient states are temporary only insofar as they are very unusual, so as time evolves they disappear, to arise again only as rare fluctuations.
- **Microcanonical ensembles are time independent.** An initial uniform density in phase space will stay uniform. More generally, since energy is conserved, a uniform density over a small shell of energies $(E, E+\delta E)$ will stay uniform.

Liouville's theorem tells us that the energy surface may get stirred around, but the relative weights of parts of the surface are given by their phase–space volumes (figure 3.1) and don't change. This is clearly a necessary condition for our microcanonical ensemble to describe the time–independent equilibrium state.

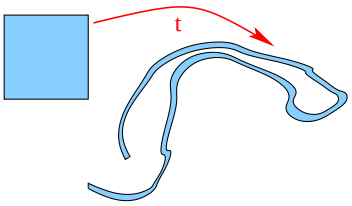



**Fig. 4.2** A small volume in phase space may be stretched and twisted by the flow, but Liouville's theorem shows that the volume stays unchanged.

[6] We'll return to the question of how irreversibility and damping emerge from statistical mechanics many times in the rest of this book. It will always involve introducing approximations to the microscopic theory.

## 4.2 Ergodicity

By averaging over the energy surface, statistical mechanics is making a hypothesis, first stated by Boltzmann. Roughly speaking, the hypothesis is that the energy surface is thoroughly stirred by the time evolution: it isn't divided into some kind of components that don't intermingle (see figure 4.3). A system which is thoroughly stirred is said to be *ergodic*.[7] The original way of defining ergodicity is due to Boltzmann. Adapting his definition,

> Definition 1: In an *ergodic* system, the trajectory of almost every[8] point in phase space eventually passes arbitrarily close[9] to every other point (position and momentum) on the surface of constant energy.

The most important consequence of ergodicity is that *time averages are equal to microcanonical averages*.[10] Intuitively, since the trajectory $(\mathbb{P}(t), \mathbb{Q}(t))$ covers the whole energy surface, the average of any property $A(\mathbb{P}(t), \mathbb{Q}(t))$ over time is the same as the average of $A$ over the energy surface.

This turns out to be tricky to prove, though. It's easier mathematically to work with another, equivalent definition of ergodicity. This definition roughly says the energy surface can't be divided into components which don't intermingle. Let's define an *ergodic component* $R$ of a set[11] $S$ to be a subset that remains invariant under the flow (so $r(t) \in R$ for all $r(0) \in R$).

> Definition 2: A time evolution in a set $S$ is *ergodic* if and only if all the ergodic components $R$ in $S$ either have zero volume or have a volume equal to the volume of $S$.

We can give an intuitive explanation of why these two definitions are equivalent (but of course it's hard to prove). A trajectory $r(t)$ of course must lie within a single ergodic component. If $r(t)$ covers the energy surface densely (definition 1), then there is 'no more room' for a second ergodic component with non-zero volume (definition 2).[12] If there is only one ergodic component $R$ with volume equal to $S$ (definition 2), then any trajectory starting in $R$ must get arbitrarily close to all points in $R$, otherwise the points in $R$ 'far' from the trajectory (outside the closure of the trajectory) would be an invariant set of non-zero volume.

Using this second definition of ergodic, we can argue that time averages must equal microcanonical averages. Let's denote the microcanonical average of an observable $A$ as $\langle A \rangle_S$, and let's denote the time average starting at initial condition $(\mathbb{P}, \mathbb{Q})$ as $\overline{A(\mathbb{P}, \mathbb{Q})}$.

Showing that the time average $\bar{A}$ equals the ensemble average $\langle A \rangle_S$ for an ergodic system (using this second definition) has three steps.

(1) **Time averages are constant on trajectories.** If $A$ is a nice function, (e.g. without any infinities on the energy surface), then

$$\overline{A(\mathbb{P}(t), \mathbb{Q}(t))} = \overline{A(\mathbb{P}(t+\tau), \mathbb{Q}(t+\tau))}; \tag{4.8}$$

[7] Mathematicians distinguish between ergodic (stirred) and *mixing* (scrambled); we only need to assume ergodicity here. See reference [63] for more information about ergodicity.

[8] What does "almost every" mean? Technically, it means all but a set of zero volume (measure zero). Basically, it's there to avoid problems with initial conditions like all the particles moving precisely at the same velocity in neat rows.

[9] Why not just assume that every point on the energy surface gets passed through? Boltzmann originally did assume this. However, it can be shown that a smooth curve (our time-trajectory) can't fill up a whole volume (the energy surface).

[10] If the system equilibrates (*i.e.*, doesn't oscillate forever), the time average behavior will be determined by the equilibrium behavior, and then ergodicity implies that the equilibrium properties are equal to the microcanonical averages.

[11] Here $S$ is the energy surface.

[12] Being careful in the definitions and proofs to exclude different invariant sets that are infinitely finely intertwined is probably why proving things is mathematically tricky.

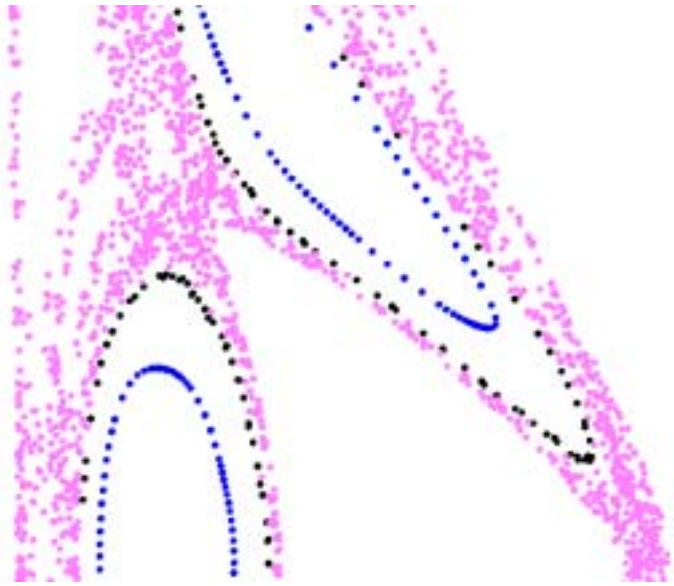


**Fig. 4.3 KAM tori and non-ergodic motion.** This is a (Poincaré) cross section of Earth's motion in the three-body problem (exercise 4.2), with Jupiter's mass set at almost 70 times its actual value. The closed loops correspond to trajectories that form tori in phase space, whose cross sections look like deformed circles in our view. The complex filled region is a single trajectory exhibiting chaotic motion, and represents an ergodic component. The tori, each an ergodic component, can together be shown to occupy non-zero volume in phase space, for small Jovian masses. Note that this system is not ergodic according to either of our definitions. The trajectories on the tori never explore the rest of the energy surface. The region $R$ formed by the chaotic domain is invariant under the time evolution; it has positive volume and the region outside $R$ also has positive volume.

the infinite time average doesn't depend on the values of $A$ during the finite time interval $(t, t+\tau)$. Thus the time average $\bar{A}$ is constant along the trajectory.[13]

(2) **Time averages are constant on the energy surface.** Now consider the subset $R_a$ of the energy surface where $\bar{A} < a$, for some value $a$. Since $\bar{A}$ is constant along a trajectory, any point in $R_a$ is sent under the time evolution to another point in $R_a$, so $R_a$ is an ergodic component. If our dynamics is ergodic on the energy surface, that means the set $R_a$ has either zero volume or the volume of the energy surface. This implies that $\bar{A}$ is a constant on the energy surface (except on a set of zero volume); its value is $a^*$, the lowest value where $R_{a^*}$ has the whole volume. Thus the equilibrium, time average value of our observable $A$ is independent of initial condition.

(3) **Time averages equal microcanonical averages.** Is this equilibrium value given by the microcanonical ensemble average over $S$? We need to show that the trajectories don't dawdle in some regions of the energy surface more than they should (based on the thickness of the energy shell, figure 3.1). Liouville's theorem in

[13] If we could show that $\bar{A}$ had to be a continuous function, we'd now be able to use the first definition of ergodicity to show that it was constant on the energy surface, since our trajectory comes close to every point on the surface. But it will not be continuous for Hamiltonian systems that are not ergodic. In figure 4.3, consider two initial conditions at nearby points, one just inside a chaotic region and the other on a KAM torus. The infinite time averages on the two trajectories for most quantities will be different: $\bar{A}$ will typically have a jump at the boundary.

section 4.1 told us that the microcanonical ensemble was time independent. But then the ensemble average equals the time average of the ensemble average, which equals the ensemble average of the time average. But the time average is constant (except on a set of zero volume), so the ensemble average equals the time average for all initial conditions (except perhaps a set of zero volume) in an ergodic system.[14]

Can we show that our systems are ergodic? Usually not. Ergodicity has been proven for the collisions of hard spheres, and for geodesic motion on finite surfaces with constant negative curvature,[15] but not for many systems of immediate practical importance. Indeed, many fundamental problems precisely involve systems which are not ergodic.

- **KAM tori and the three-body problem.** Generations of mathematicians and physicists have worked on the gravitational three-body problem.[16] The key challenge was showing that the interactions between the planets do not completely mess up their orbits over long times. One must note that "messing up their orbits" is precisely what an ergodic system must do! (There's just as much phase space at constant energy with Earth and Venus exchanging places, and a whole lot more with Earth flying out into interstellar space.) In the last century[17] the KAM theorem was proven, which showed that (for small interplanetary interactions and a large fraction of initial conditions) the orbits of the planets qualitatively stayed in weakly perturbed ellipses around the Sun (KAM tori, see figure 4.3). Other initial conditions, intricately intermingled with the stable ones, lead to chaotic motion. Exercise 4.2 investigates the KAM tori and chaotic motion in a numerical simulation.
  From the KAM theorem and the study of chaos in these systems we learn that Hamiltonian systems with small numbers of particles are often, even usually, not ergodic – there are commonly regions formed by tori of non-zero volume which do not mix with the rest of the energy surface.
- **Fermi, Pasta, Ulam and KdV.** You might think that this is a peculiarity of having only a few particles. Surely if there are lots of particles, such funny behavior has to go away? On one of the early computers developed for the Manhattan project, Fermi, Pasta and Ulam tested this [30]. They took a one-dimensional chain of atoms, coupled them with anharmonic potentials, and tried to look for thermalization. To quote them:

  > "Let us say here that the results of our computations were, from the beginning, surprising us. Instead of a continuous flow of energy from the first mode to the higher modes, all of the problems show an entirely different behavior. ... Instead of a gradual increase of all the higher modes, the energy is exchanged, essentially, among only a certain few. It is, therefore, very

[14] In formulas, $\langle A\rangle_S = \langle A(t)\rangle_S = \langle A(\mathbb{P}(t),\mathbb{Q}(t))\rangle_S$, where the average $\langle\rangle_S$ integrates over initial conditions $(\mathbb{P}(0),\mathbb{Q}(0))$ but evaluates $A$ at $(\mathbb{P}(t),\mathbb{Q}(t))$. Averaging over all time, and using the fact that $\bar{A} = a^*$ (almost everywhere), tells us

$$\begin{aligned}\langle A\rangle_S &= \lim_{T\to\infty}\frac{1}{T}\int_0^T \langle A(\mathbb{P}(t),\mathbb{Q}(t))\rangle_S\, dt \\ &= \left\langle \lim_{T\to\infty}\frac{1}{T}\int_0^T A(\mathbb{P}(t),\mathbb{Q}(t))\, dt\right\rangle_S \\ &= \langle \overline{A(\mathbb{P},\mathbb{Q})}\rangle_S = \langle a^*\rangle_S = a^*.\end{aligned} \tag{4.9}$$

[15] Geodesic motion on a sphere would be motion at a constant speed around great circles. Geodesics are the shortest paths between two points. In general relativity, falling bodies travel on geodesics in space-time.

[16] Newton solved the gravitational two-body problem, giving Kepler's ellipse.

[17] That is, the 20th century.

> hard to observe the rate of "thermalization" or mixing in our problem, and this was the initial purpose of the calculation." [30, p.978]

It turns out that their system, in the continuum limit, gave a partial differential equation (the Kortweg-de Vries equation) that was even weirder than planetary motion: it had an *infinite family* of conserved quantities, and could be exactly solved using a combination of fronts called "solitons".

The kind of non-ergodicity found in the Kortweg-de Vries equation was thought to arise in only rather special one–dimensional systems. The recent discovery of anharmonic localized modes in generic, three–dimensional systems by Sievers and Takeno [94, 88] suggests that non-ergodicity my arise in rather realistic lattice models.

- **Phase Transitions.** In systems with an infinite number of particles, one can have phase transitions. Often ergodicity breaks down in one of the phases. For example, a liquid may explore all of phase space with a given energy, but an infinite crystal (with a neat grid of atoms aligned in a particular orientation) will never fluctuate to change its orientation, or (in three dimensions) the registry of its grid.[18] The real system will explore only one ergodic component of the phase space (one crystal position and orientation), and we must do the same when making theories of the system.
- **Glasses.** There are other kinds of breakdowns of the ergodic hypothesis. For example, glasses fall out of equilibrium as they are cooled: they no longer ergodically explore all configurations, but just oscillate about one of many metastable glassy states. Certain models of glasses and disordered systems can be shown to break ergodicity - not just into a small family of macroscopic states as in normal symmetry-breaking phase transitions, but into an infinite number of different, disordered ground states. It is an open question whether real glasses truly break ergodicity when cooled infinitely slowly, or whether they are just sluggish, 'frozen liquids'.

[18] That is, a 3D crystal has broken *orientational* and *translational* symmetries: see chapter 9.

Should we be concerned that we cannot prove that our systems are ergodic? It is entertaining to point out the gaps in our derivations, especially since they tie into so many central problems in mathematics and physics (above). We emphasize that these gaps are for most purposes purely of academic concern. Statistical mechanics works phenomenally well in systems with large numbers of interacting degrees of freedom.

Indeed, the level of rigor here is unusual. In more modern applications[19] of statistical mechanics outside of equilibrium thermal systems there is rarely any justification of the choice of the ensemble comparable to that provided by Liouville's theorem and ergodicity.

[19] In disordered systems, disorder is heuristically introduced with Gaussian or discrete random variables. In stochastic systems Gaussian or white noise is added. In Bayesian statistics, the user is in charge of determining the prior model probability distribution, analogous to Liouville's theorem determining the measure on phase space.

# Exercises

(4.1) **The Damped Pendulum vs. Liouville's Theorem.** (Basic, Mathematics)

The damped pendulum has a force $-\gamma p$ proportional to the momentum slowing down the pendulum. It satisfies the equations

$$\begin{aligned} \dot{x} &= p/M \\ \dot{p} &= -\gamma p - K \sin(x). \end{aligned} \tag{4.10}$$

At long times, the pendulum will tend to an equilibrium stationary state, zero velocity at $x = 0$ (or more generally at the equivalent positions $x = 2m\pi$, for $m$ an integer): $(p, x) = (0, 0)$ is an attractor for the damped pendulum. An ensemble of damped pendulums is started with initial conditions distributed with probability $\rho(p_0, x_0)$. At late times, these initial conditions are gathered together near the equilibrium stationary state: Liouville's theorem clearly is not satisfied.

*(a) In the steps leading from equation 4.5 to equation 4.7, why does Liouville's theorem not apply to the damped pendulum? More specifically, what are $\partial\dot{p}/\partial p$ and $\partial\dot{q}/\partial q$?*

*(b) Find an expression for the total derivative $d\rho/dt$ in terms of $\rho$ for the damped pendulum. If we evolve a region of phase space of initial volume $A = \Delta p\, \Delta x$ how will its volume depend upon time?*

(4.2) **Jupiter! and the KAM Theorem** (Astrophysics, Mathematics)

See also the Jupiter Web pages [99].

The foundation of statistical mechanics is the *ergodic hypothesis*: any large system will explore the entire energy surface. We focus on large systems because it is well known that many systems with a few interacting particles are definitely not ergodic.

The classic example of a non-ergodic system is the Solar system. Jupiter has plenty of energy to send the other planets out of the Solar system. Most of the phase-space volume of the energy surface has eight planets evaporated and Jupiter orbiting the Sun alone: the ergodic hypothesis would doom us to one long harsh winter. So, the big question is: Why hasn't the Earth been kicked out into interstellar space?

Mathematical physicists have studied this problem for hundreds of years. For simplicity, they focused on the *three-body problem*: for example, the Sun, Jupiter, and the Earth. The early (failed) attempts tried to do perturbation theory in the strength of the interaction between planets. Jupiter's gravitational force on the Earth is not tiny, though: if it acted as a constant brake or accelerator, our orbit would be way out of whack in a few thousand years. Jupiter's effects must cancel out over time rather perfectly...

This exercise is mostly discussion and exploration: only a few questions need to be answered. Download the program Jupiter from the appropriate link at the bottom of reference [99]. (Go to the directory with the binaries and select Jupiter.) Check that Jupiter doesn't seem to send the Earth out of the Solar system. Try increasing Jupiter's mass to 35000 Earth masses. (If you type in a new value, you need to hit Enter to register it.)

Start the program over again (or reset Jupiter's mass back to 317.83 Earth masses). Shifting "View" to "Earth's trajectory", run for a while, and zoom in with the right mouse button to see the small effects of Jupiter on the Earth. (The left mouse button will launch new trajectories. Clicking with the right button will restore the original view.) Note that the Earth's position shifts depending on whether Jupiter is on the near or far side of the sun.

*(a) Estimate the fraction that the Earth's radius from the Sun changes during the first Jovian year (about 11.9 years). How much does this fractional variation increase over the next hundred Jovian years?*

Jupiter thus warps Earth's orbit into a kind of spiral around a tube. This orbit in physical three-dimensional space is a projection of the tube in 6N-dimensional phase space. The tube in phase space already exists for massless planets...

Let's start in the non-interacting planet approximation (where Earth and Jupiter are assumed to have zero mass). Both Earth's orbit and Jupiter's orbit then become circles, or more generally ellipses. The field of topology does not distinguish an ellipse from a circle: any stretched, wiggled rubber band is a circle so long as it forms a curve that closes into a loop. Similarly, a torus (the surface of a doughnut) is topologically equivalent to any closed surface with one hole in it (like the surface of a coffee cup, with the handle as the hole). Convince yourself in this

[20] Hint: plot the orbit in the $(x, y)$, $(x, p_x)$, and other planes. It should look like the projection of a circle along various axes.

non-interacting approximation that Earth's orbit remains topologically a circle in its six-dimensional phase space.[20]

*(b) In the non-interacting planet approximation, what topological surface is it in in the eighteen-dimensional phase space that contains the trajectory of the three bodies? Choose between (i) sphere, (ii) torus, (iii) Klein bottle, (iv) two-hole torus, (v) complex projective plane.*[21] *About how many times does Earth wind around this surface during each Jovian year? (This ratio of years is called the winding number).*

The mathematical understanding of the three-body problem was only solved in the past hundred years or so, by Kolmogorov, Arnol'd, and Moser. Their proof focuses on the topological integrity of this tube in phase space (called now the KAM torus). They were able to prove stability if the winding number (Jupiter year over Earth year) is sufficiently irrational. More specifically, they could prove in this case that for sufficiently small planetary masses that there is a distorted torus in phase space, near the unperturbed one, around which the planets spiral around with the same winding number.

*(c) About how large can you make Jupiter's mass before Earth's orbit stops looking like a torus? (You can hit "Clear" and "Reset" to put the planets back to a standard starting point. Otherwise, your answer will depend upon the location of Jupiter in the sky.) Admire the cool orbits when the mass becomes too heavy.*

Thus, for "small" Jovian masses, the trajectory in phase space is warped and rotated a bit, so that its toroidal shape is visible looking at Earth's position alone. (The circular orbit for zero Jovian mass is looking at the torus on edge.)

The fact that the torus isn't destroyed immediately is a serious problem for statistical mechanics! The orbit does not ergodically explore the entire allowed energy surface. This is a counterexample to Boltzmann's ergodic "theorem". That means that time averages are not equal to averages over the energy surface: our climate would be very unpleasant, on the average, if our orbit were ergodic.

Let's use a Poincaré section to explore these tori, and the chaotic regions between them. If a dynamical system keeps looping back in phase space, one can take a cross-section of phase space and look at the mapping from that cross section back into itself (see figure 4.4).

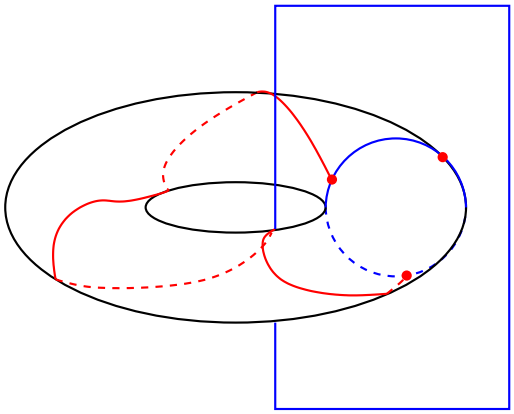


**Fig. 4.4** The Poincaré section of a torus is a circle. The dynamics on the torus becomes a mapping of the circle onto itself.

The Poincaré section shown in the figure is a planar cross section in a three-dimensional phase space. Can we reduce our problem to an interesting problem with three phase-space coordinates? The original problem has an eighteen dimensional phase space. In the center of mass frame it has twelve interesting dimensions. If we restrict the motion to a plane, it reduces to eight dimensions. If we assume the mass of the Earth is zero (the restricted planar three body problem) we have five relevant coordinates (Earth $xy$ positions and velocities, and the location of Jupiter along its orbit). If we remove one more variable by going to a rotating coordinate system that rotates with Jupiter, the current state of our model can be described with four numbers: two positions and two momenta for the Earth. We can remove another variable by confining ourselves to a fixed "energy". The true energy of the Earth isn't conserved (because Earth feels a periodic potential), but there is a conserved quantity which is like the energy in the rotating frame: more details described under "Help" or on the Web [99] under "Description of the Three Body Problem". This leaves us with a trajectory in three dimensions (so, for small Jovian masses, we have a torus embedded in a three-dimensional space). Finally, we take a Poincaré cross section: we plot a point of the trajectory every time Earth passes directly between Jupiter and the Sun. I plot the distance to Jupiter along the horizontal axis, and the velocity component towards Jupiter along the vertical axis; the perpendicular component of the velocity isn't shown (and is determined by the

[21] Hint: It's a circle cross a circle, parameterized by two independent angles – one representing the month of Earth's year, and one representing the month of the Jovian year. Feel free to look at part (c) before committing yourself, if pure thought isn't enough.

"energy").

Set the View to Poincaré. (You may need to expand the window a bit: sometimes the dot size is too small to see.) Set Jupiter's mass to 2000, and run for 1000 years. You should see two nice elliptical cross-sections of the torus. As you increase the mass (type in a mass, Enter, Reset and Run, repeat), watch the toroidal cross sections as they break down. Run for a few thousand years at $M_J = 22000 M_e$; notice the torus has broken into three circles.

Fixing the mass at $M_J = 22000 M_e$, let's explore the dependence of the planetary orbits on the initial condition. Select the preset for "Chaos" (or set $M_J$ to 22000 $M_e$, "View" to Poincaré, and Reset). Clicking on a point on the screen with the left mouse button will launch a trajectory with that initial position and velocity towards Jupiter; it sets the perpendicular component of the velocity to keep the current "energy". (If you click on a point where energy cannot be conserved, the program will tell you so.) You can thus view the trajectories on a two-dimensional cross-section of the three-dimensional constant "energy" surface.

Notice that many initial conditions slowly fill out closed curves. These are KAM tori that have been squashed and twisted like rubber bands.[22] Explore until you find some orbits that seem to fill out whole regions: these represent *chaotic* orbits.[23]

*(d) If you can do a screen capture, print out a Poincaré section with initial conditions both on KAM tori and in chaotic regions: label each.*[24] See figure 4.3 for a small segment of the picture you should generate.

It turns out that proving that Jupiter's effects cancel out depends on Earth's smoothly averaging over the surface of the torus. If Jupiter's year is a rational multiple of Earth's year, the orbit closes after a few years and you don't average over the whole torus: only a closed spiral. Rational winding numbers, we now know, leads to chaos when the interactions are turned on: the large chaotic region you found above is associated with an unperturbed orbit with a winding ratio of 3:1. Of course, the rational numbers are dense: between any two KAM tori there are chaotic regions, just because between any two irrational numbers there are rational ones. It's even worse: it turns out that numbers which are really, really close to rational (Liouville numbers like $1+1/10+1/10^{10}+1/10^{10^{10}}+\ldots$) also may lead to chaos. It was amazingly tricky to prove that lots of tori survive nonetheless. You can imagine why this took hundreds of years to understand (especially without computers).

(4.3) **Invariant Measures.** (Math, Complexity) (With Myers. [75])

**Reading:** Reference [49], Roderick V. Jensen and Christopher R. Myers, "Images of the critical points of nonlinear maps" *Physical Review A* **32**, 1222-1224 (1985).

Liouville's theorem tells us that all available points in phase space are equally weighted when a Hamiltonian system is averaged over all times. What happens for systems that evolve according to laws that are not Hamiltonian? Usually, the system does not continue to explore all points in its state space: at long times it is confined a subset of the original space known as the *attractor*.

We consider the behavior of the 'logistic' mapping from the unit interval $(0, 1)$ into itself.[25]

$$f(x) = 4\mu x(1 - x). \tag{4.11}$$

We talk of the trajectory of an initial point $x_0$ as the sequence of points $x_0$, $f(x_0)$, $f(f(x_0))$, $\ldots$, $f^{[n]}(x_0)$, $\ldots$. Iteration can be thought of as a time step (one iteration of a Poincaré return map of exercise 4.2 or one step $\Delta t$ in a time-step algorithm as in exercise 8.9).

**Attracting Fixed Point:** For small $\mu$, our mapping has an attracting fixed point. A fixed point of a mapping is a value $x^* = f(x^*)$; a fixed point is stable if under small perturbations shrink:

$$|f(x^* + \epsilon) - x^*| \approx |f'(x^*)|\epsilon < \epsilon, \tag{4.12}$$

[22] You can "Continue" if the trajectory doesn't run long enough to give you a complete feeling for the cross-section: also, increase the time to run). You can zoom in with the right mouse button, and zoom out by expanding the window or by using the right button and selecting a box which extends outside the window.

[23] Notice that the chaotic orbit doesn't throw the Earth out of the Solar system. The chaotic regions near infinity and near our initial condition are not connected. This may be an artifact of our simplified model: in other larger systems it is believed that all chaotic regions (on a connected energy surface) are joined through *Arnol'd diffusion*.

[24] At least under Linux, the "Print" feature is broken. Under Linux, try "gimp": File Menu, then Acquire, then Screen Shot. Under Windows, alt-Print Screen and then Paste into your favorite graphics program.

[25] We also study this map in exercises 5.11, 5.13, and 12.8.

which happens if the derivative $|f'(x^*)| < 1$.[26]

*(a)* **Iteration:** *Set $\mu = 0.2$; iterate $f$ for some initial points $0 < x_0 < 1$ of your choosing, and convince yourself that they all are attracted to zero. Plot $f$ and the diagonal $y = x$ on the same plot. Are there any fixed points other than $x = 0$? Repeat for $\mu = 0.4$, and 0.6. What happens?*

**Analytics:** *Find the non-zero fixed point $x^*(\mu)$ of the map 4.11, and show that it exists and is stable for $1/4 < \mu < 3/4$. If you're ambitious or have a computer algebra program, show that there is a stable period–two cycle for $3/4 < \mu < (1 + \sqrt{6})/4$.*

An attracting fixed point is the antithesis of Liouville's theorem: all initial conditions are transient except one, and all systems lead eventually to the same, time–independent state. (On the other hand, this is precisely the behavior we expect in statistical mechanics *on the macroscopic scale*: the system settles down into a time–independent equilibrium state! All microstates are equivalent, but the vast majority of accessible microstates have the same macroscopic behavior in most large systems). We could define a rather trivial "equilibrium ensemble" for this system, which consists of the single point $x^*$: any property $A(x)$ will have the long–time average $\langle A \rangle = A(x^*)$.

For larger values of $\mu$, more complicated things happen. At $\mu = 1$, the dynamics can be shown to fill the entire interval: the dynamics is ergodic, and the attractor fills the entire set of available states. However, unlike the case of Hamiltonian systems, not all states are weighted equally.

We can find time averages for functions of $x$ in two ways: by averaging over time (many iterates of the map) or by weighting an integral over $x$ by the *invariant density* $\rho(x)$. The invariant density $\rho(x)\,dx$ is the probability that a point on a long trajectory will lie between $x$ and $x + dx$. To find it numerically, we iterate a typical point[27] $x_0$ a thousand or so times ($N_{\text{transient}}$) to find a point $x_a$ on the attractor, and then collect a long trajectory of perhaps a million points ($N_{\text{cycles}}$). A histogram of this trajectory gives $\rho(x)$. Clearly averaging over this density is the same as averaging over the trajectory of a million points. We call $\rho(x)$ an invariant measure because it's left invariant under the mapping $f$: iterating our million–point approximation for $\rho$ once under $f$ only removes the first point $x_a$ and adds one extra point to the end.

*(b)* **Invariant Density:** *Set $\mu = 1$; iterate $f$ many times, and form a histogram of values giving the density $\rho(x)$ of points along the trajectory. You should find that points $x$ near the boundaries are approached more often than points near the center.*

**Analytics:** *Use the fact that the long time average $\rho(x)$ must be independent of time, verify for $\mu = 1$ that the density of points is*[28]

$$\rho(x) = \frac{1}{\pi\sqrt{x(1-x)}}. \tag{4.13}$$

*Plot this theoretical curve with your numerical histogram.* (Hint: The points in a range $dx$ of a point $x$ map under $f$ to a range $dy = f'(x)\,dx$ around the image $y = f(x)$. Each iteration maps two points $x_a$ and $x_b = 1 - x_a$ to $y$, and thus maps all the density $\rho(x_a)|dx_a|$ and $\rho(x_b)|dx_b|$ into $dy$. Hence the probability $\rho(y)dy$ must equal $\rho(x_a)|dx_a| + \rho(x_b)|dx_b|$, so

$$\rho(f(x_a)) = \rho(x_a)/|f'(x_a)| + \rho(x_b)/|f'(x_b)| \tag{4.14}$$

Plug equation 4.13 into equation 4.14. You'll need to factor a polynomial.)

Mathematicians call this probability density $\rho(x)dx$ the *invariant measure* on the attractor.[29] To get the long term average of any function $A(x)$, one can use

$$\langle A \rangle = \int A(x)\rho(x)dx \tag{4.15}$$

To a mathematician, a measure is a way of weighting different regions in doing integrals – precisely our $\rho(x)dx$. Notice that, for the case of an attracting fixed point, we would have $\rho(x) = \delta(x^*)$.[30]

[26] For many dimensional mappings, a sufficient criterion for stability is that all the eigenvalues of the Jacobian have magnitude smaller than one. A continuous time evolution $dy/dt = F(y)$, will be stable if $dF/dy$ is smaller than zero, or (for multidimensional systems) if the Jacobian $DF$ has eigenvalues whose real parts are all less than zero. This is all precisely analogous to discrete and continuous–time Markov chains, see section 8.3

[27] For example, we must not choose an unstable fixed point or unstable periodic orbit!

[28] You need not derive the factor $1/\pi$, which normalizes the probability density to one.

[29] There are actually many possible invariant measures on some attractors: this one is the SRB measure.[41]

[30] The case of a fixed point then becomes mathematically a measure with a point mass at $x^*$.

**Cusps in the invariant density:** At values of $\mu$ slightly smaller than one, our mapping has a rather complex invariant density.

*(c) Find the invariant density (as described above) for $\mu = 0.9$. Make your trajectory length $N_{\rm cycles}$ big enough and the bin size small enough to see the interesting structures. Notice that the attractor no longer fills the whole range $(0, 1)$: locate roughly where the edges are. Notice also the cusps in $\rho(x)$ at the edges of the attractor, and also at places inside the attractor (called* boundaries*, see reprint above). Locate some of the more prominent cusps.*

**Analytics of cusps:** Notice that $f'(\frac{1}{2}) = 0$. so by equation 4.14 we know that $\rho(f(x)) \geq \rho(x)/|f'(x)|$ must have a singularity near $x = \frac{1}{2}$: all the points near $x = \frac{1}{2}$ are squeezed together and folded to one side by $f$. Further iterates of this singularity produce more cusps: the crease after one fold stays a crease after being further stretched and kneaded.

*(d) Set $\mu = 0.9$. Calculate $f(\frac{1}{2})$, $f(f(\frac{1}{2}))$, ... and compare these iterates to the locations of the edges and cusps from part (c). (You may wish to include them both on the same plot.)*

**Bifurcation Diagram:** The evolution of the attractor and its invariant density as $\mu$ varies is plotted in the *bifurcation diagram*, which is shown for large $\mu$ in figure 4.5. One of the striking features in this plot are the sharp boundaries formed by the cusps.

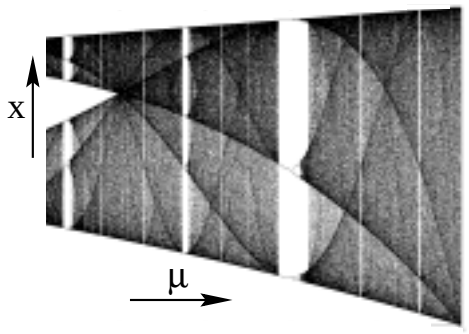



**Fig. 4.5 Bifurcation diagram in the chaotic region.** Notice the boundary lines threading through the diagram, images of the crease formed by the folding at $x = \frac{1}{2}$ in our map (see reprint above).

*(e)* **Bifurcation Diagram:** *Plot the attractor (duplicating figure 4.5) as a function of $\mu$, for $0.8 < \mu < 1$. (Pick regularly spaced $\delta\mu$, run $n_{\rm transient}$ steps, record $n_{\rm cycles}$ steps, and plot. After the routine is working, you should be able to push $n_{\rm transient}$ and $n_{\rm cycles}$ both larger than 100, and $\delta\mu < 0.01$.)*

*On the same plot, for the same $\mu$s, plot the first eight images of $x = \frac{1}{2}$, that is, $f(\frac{1}{2}), f(f(\frac{1}{2})), \ldots$. Are the boundaries you see just the cusps? What happens in the bifurcation diagram when two boundaries touch? (See the reprint above.)*

# Entropy

# 5

Entropy is the most influential topic to arise from statistical mechanics. What does it mean? Can we develop an intuition for it?

We shall see in this chapter that entropy has three related interpretations.[1] *Entropy measures the disorder in a system:* in section 5.2 we'll see this using the entropy of mixing and the residual entropy of glasses. *Entropy measures our ignorance about a system:* in section 5.3 we'll see this with examples from nonequilibrium systems and information theory. But we'll start in section 5.1 with the original interpretation, that grew out of the 19th century study of engines, refrigerators, and the end of the universe: *Entropy measures the irreversible changes in a system.*

[1] Equilibrium is a word with positive connotations, presumably because it allows us to compute properties easily. Entropy and the quantities it measures – disorder, ignorance, uncertainty – are words with negative connotations, presumably because entropy interferes with making efficient heat engines. Notice that these connotations are not always reliable: in information theory, for example, having high Shannon entropy is good, reflecting better compression of data.

## 5.1 Entropy as Irreversibility: Engines and Heat Death of the Universe

The early 1800's saw great advances in understanding motors and engines. In particular, scientists asked a fundamental question: How efficient can an engine be? The question was made more difficult because there were two relevant principles to be discovered: energy is conserved and entropy always increases.[2]

For some kinds of engines, only energy conservation is important. For example, there are electric motors that convert electricity into mechanical work (running an electric train), and generators that convert mechanical work (from a rotating windmill) into electricity.[3] For these electromechanical engines, the absolute limitation is given by the conservation of energy: the motor cannot generate more energy in mechanical work than is consumed electrically, and the generator cannot generate more electrical energy than is input mechanically.[4] An ideal electromechanical engine can convert all the energy from one form to another. Electric motors and generators are limited only by the conservation of energy.

Steam engines are more complicated. Scientists in the early 1800's were figuring out that heat is a form of energy. A steam engine, running a power plant or an old-style locomotive, transforms a fraction of the heat energy from the hot steam (the 'hot bath') into electrical energy or work, but some of the heat energy always ends up 'wasted' – dumped into the air or into the cooling water for the power plant (the 'cold bath'). In fact, if the only limitation on heat engines was conservation of energy, one would be able to make a motor using the heat energy from

[2] Some would be pedantic, and say only that entropy never decreases since a system in equilibrium has constant entropy. The phrase "entropy always increases" has a ring to it, though. We can justify using it by noticing that systems only equilibrate completely after an infinite time. For example, we'll see that Carnot cycles must be run infinitely slowly to be truly reversible.

[3] Electric motors are really the same as generators run in reverse: turning the shaft of a simple electric motor can generate electricity.

[4] Mechanical work (force times distance) is energy; electrical power (current times voltage) is energy per unit time.

a rock, getting both useful work and a very cold rock.

There is something fundamentally less useful about energy once it becomes heat. By spreading out the energy among all the atoms in a macroscopic chunk of material, not all of it can be retrieved again to do useful work. The energy is more useful for generating power when divided between hot steam and a cold lake, than in the form of water at a uniform, intermediate warm temperature. Indeed, most of the time when we use mechanical or electrical energy, the energy ends up as heat, generated from friction or other dissipative processes.

The equilibration of a hot and cold body to two warm bodies in an isolated system is *irreversible*: one cannot return to the original state without inputting some kind of work from outside the system. Carnot, publishing in 1824, realized that the key to producing the most efficient possible engine was to avoid irreversibility. A heat engine run in reverse is a refrigerator: it consumes mechanical work or electricity and uses it to pump heat from a cold bath to a hot one (extracting some of the heat as work). A reversible heat engine would be able to run forward generating work by transferring heat from the hot to the cold bath, and then run backward using the same work to pump the heat back into the hot bath.

If you had an engine more efficient than a reversible one, you could run it side-by-side with a reversible engine running as a refrigerator (figure 5.1). The pair of engines would generate work by extracting energy from the hot bath (as from our rock, above) without adding heat to the cold one. After we used this work, we could dump the extra heat from friction back into the hot bath, getting a perpetual motion machine that did useful work without consuming anything. *We postulate that such perpetual motion machines are impossible.* By calculating the properties of this reversible engine, Carnot placed a fundamental limit on the efficiency of heat engines and discovered what would later be called the entropy.

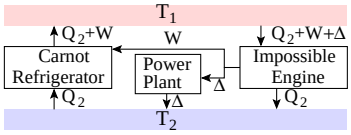



**Fig. 5.1** How to use an engine which produces $\Delta$ more work than the Carnot cycle to build a perpetual motion machine doing work $\Delta$ per cycle.

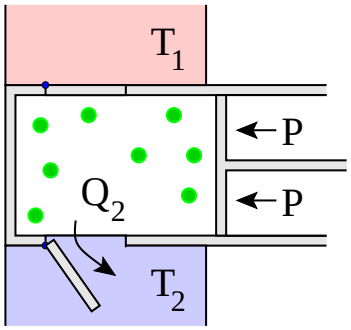



**Fig. 5.2 Prototype Heat Engine:** A piston with external exerted pressure $P$, moving through an insulated cylinder. The cylinder can be put into thermal contact with either of two heat baths: a hot bath at temperature $T_1$ (say, a coal fire in a power plant) and a cold bath at $T_2$ (say water from a cold lake). During one cycle of the piston in and out, heat energy $Q_1$ flows into the piston, mechanical energy $W$ is done on the external world by the piston, and heat energy $Q_2$ flows out of the piston into the cold bath.

Carnot considered a prototype heat engine (figure 5.2), given by a piston with external pressure $P$, two heat baths at a hot temperature $T_1$ and a cold temperature $T_2$, with some material inside the piston. During one cycle of his engine heat $Q_1$ flows out of the hot bath, heat $Q_2$ flows into our cold bath, and net work $W = Q_1 - Q_2$ is done by the piston on the outside world. To make his engine reversible Carnot must avoid (i) friction, (ii) letting hot things touch cold things, (iii) letting high pressures expand into low pressures, and (iv) moving the walls of the container too quickly (emitting sound or shock waves).

Carnot, a theorist, could ignore the practical difficulties. He imagined a frictionless piston run through a cycle at arbitrarily low velocities. The piston was used both to extract work from the system and to raise and lower the temperature. Carnot connected the gas thermally to each bath only when its temperature agreed with the bath, so his engine was fully reversible.

The Carnot cycle moves the piston in and out in four steps (figure 5.3).

- (a→b) The compressed gas is connected to the hot bath, and the piston

moves outward at a varying pressure; heat $Q_1$ flows in to maintain the gas at temperature $T_1$.

- (b→c) The piston expand further at varying pressure, cooling the gas to $T_2$ without heat transfer,
- (c→d) The expanded gas in the piston is connected to the cold bath and compressed; heat $Q_2$ flows out maintaining the temperature at $T_2$.
- (d→a) The piston is compressed, warming the gas to $T_1$ without heat transfer, returning it to the original state.

Energy conservation tells us that the net heat energy flowing into the piston, $Q_1 - Q_2$ must equal the work done on the outside world $W$:

$$Q_1 = Q_2 + W. \tag{5.1}$$

The work done by the piston is the integral of the force exerted times the distance. The force is the piston surface area times the pressure, and the distance times the piston surface area is the volume change, giving the simple result

$$W = \int F dx = \int (F/A)(A dx) = \int_{\text{cycle}} P\, dV = \text{Area inside PV Loop.} \tag{5.2}$$

That is, if we plot $P$ versus $V$ for the four steps of our cycle, the area inside the resulting closed loop is the work done by the piston on the outside world (figure 5.2).

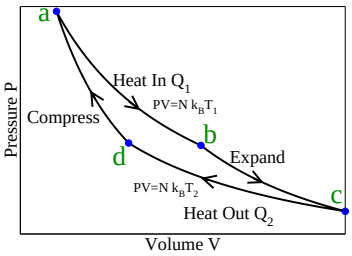



**Fig. 5.3 Carnot Cycle PV Diagram:** The four steps in the Carnot cycle: a→b heat in $Q_1$ at constant temperature $T_1$, b→c expansion without heat flow, c→d heat out $Q_2$ at constant temperature $T_2$, and d→a compression without heat flow to the original volume and temperature.

Carot realized that all reversible heat engines working with the same temperature baths had to produce exactly the same amount of work for a given heat flow from hot to cold (none of them could be more efficient than any other, since they all were the most efficient possible). This allowed him to fill the piston with the simplest possible material (a monatomic ideal gas), for which he knew the relation between pressure, volume, and temperature. We saw in section 3.5 that the ideal gas equation of state is

$$PV = Nk_BT \tag{5.3}$$

and that its total energy is its kinetic energy, given by the equipartition theorem

$$E = {}^3\!/_2 Nk_BT = {}^3\!/_2 PV. \tag{5.4}$$

Along a→b where we add heat $Q_1$ to the system, we have $P(V) = Nk_BT_1/V$. Using energy conservation (the first law),

$$Q_1 = E_b - E_a + W_{ab} = {}^2\!/_3 P_bV_b - {}^2\!/_3 P_aV_a + \int_a^b P\, dV \tag{5.5}$$

But $P_aV_a = Nk_BT_1 = P_bV_b$, so the first two terms cancel, and the last term simplifies

$$Q_1 = \int_a^b \frac{Nk_BT_1}{V}\, dV = Nk_BT_1 \log(V_b/V_a). \tag{5.6}$$

Similarly,

$$Q_2 = Nk_BT_2 \log(V_c/V_d). \tag{5.7}$$

For the other two steps in our cycle we need to know how the ideal gas behaves under expansion without any heat flow in or out. Again, using the first law on a small segment of the path, the work done for a small volume change $-PdV$ must equal the change in energy $dE$. Using equation 5.3, $-P\,dV =$

$-\frac{Nk_BT}{V}\,dV$, and using equation 5.4, $dE = {}^3\!/_2 Nk_B\,dT$, so $dV/V = -{}^3\!/_2 dT/T$. Integrating both sides from $b$ to $c$, we find

$$\int_b^c \frac{dV}{V} = \log(V_c/V_b) = \int_b^c -{}^3\!/_2 \frac{dT}{T} = -{}^3\!/_2 \log(T_2/T_1). \tag{5.8}$$

so $V_c/V_b = (T_1/T_2)^{3/2}$. Similarly, $V_d/V_a = (T_1/T_2)^{3/2}$. Thus $V_c/V_b = V_d/V_a$, and hence

$$\frac{V_c}{V_d} = \frac{V_c}{V_b}\frac{V_b}{V_d} = \frac{V_d}{V_a}\frac{V_b}{V_d} = \frac{V_b}{V_a}. \tag{5.9}$$

We can use the volume ratios from the insulated expansion and compression (equation 5.9) to substitute into the heat flow (equations 5.6 and 5.6) to find

$$Q_1/T_1 = Nk_B \log(V_b/V_a) = Nk_B \log(V_c/V_d) = Q_2/T_2. \tag{5.10}$$

This was Carnot's fundamental result: his cycle, and hence all reversible engines, satisfies the law

$$Q_1/T_1 = Q_2/T_2. \tag{5.11}$$

Later scientists decided to define[5] the *entropy change* to be this ratio of heat flow to temperature:

$$\Delta S_{\mathbf{thermo}} = Q/T. \tag{5.12}$$

For a reversible engine the entropy flow from the hot bath into the piston $Q_1/T_1$ equals the entropy flow from the piston into the cold bath $Q_2/T_2$: no entropy is created or destroyed. Any real engine will create[6] net entropy during a cycle: no engine can reduce the net amount of entropy in the universe.

The irreversible increase of entropy is not a property of the microscopic laws of nature. In particular, the microscopic laws of nature are *time reversal invariant*: the laws governing the motion of atoms are the same whether time is running backward or forward.[7] The microscopic laws do not tell us the *arrow of time*. The direction of time in which entropy increases is our definition of the *future*.[8]

This confusing point may be illustrated by considering the game of pool or billiards. Neglecting friction, the trajectories of the pool balls are also time-reversal invariant. If the velocities of the balls were reversed halfway through a pool shot, they would retrace their motions, building up all the velocity into one ball that then would stop as it hit the cue stick. In pool, the feature that distinguishes forward from backward time is the greater order at early times: all the momentum starts in one ball, and is later distributed among all the balls involved in collisions. Similarly, the only reason we can resolve the arrow of time – distinguish the future from the past – is that our universe started in an unusual, low entropy state,[9] and is irreversibly moving towards equilibrium.[10]

[5] The thermodynamic entropy is derived with a heat flow $\Delta E = Q$ at a fixed temperature $T$, so our statistical mechanics definition of temperature $1/T = \partial S/\partial E$ (from equation 3.29) is equivalent to the thermodynamics definition of entropy $\Delta S = Q/T$ (equation 5.12).

[6] For example, a small direct heat leak from the hot bath to the cold bath of $\delta$ per Carnot cycle would generate

$$\frac{Q_2+\delta}{T_2} - \frac{Q_1+\delta}{T_1} = \delta\left(\frac{1}{T_2} - \frac{1}{T_1}\right) > 0 \tag{5.13}$$

entropy per cycle.

[7] More correctly, the laws of nature are only invariant under CPT: changing the direction of time (T) along with inverting space (P) and changing matter to antimatter (C). Radioactive beta decay and other weak interaction forces are not invariant under time reversal. The basic conundrum for statistical mechanics is the same, though: we can't tell if we are matter beings living forward in time or antimatter beings living backward in time in a mirror. Time running backward would appear strange even if we were made of antimatter.

[8] In electromagnetism, the fact that waves radiate away from sources more often than they converge upon sources is a closely related distinction of past from future.

[9] The big bang was hot and probably close to equilibrium, but the volume per particle was small so the entropy was nonetheless low.

[10] Suppose some miracle (or waiting a really, really long time) produced a spontaneous fluctuation of an equilibrium system into a low-entropy, ordered state. Preceding that time, with extremely high probability, all of our laws of macroscopic physics would appear to run backward. The most probable route building up to an ordered state from equilibrium is the time reverse of the most probable decay of that ordered state back to equilibrium.

The cosmic implications of the irreversible increase of entropy were not lost on the intellectuals of the 19th century. In 1854, Helmholtz predicted the *heat death of the universe*: he suggested that as the universe ages all energy will become heat, all temperatures will become equal, and everything will "be condemned to a state of eternal rest". In 1895, H.G. Wells in *The Time Machine* [122, Chapter 11] speculated about the state of the Earth in the distant future:

> "...the sun, red and very large, halted motionless upon the horizon, a vast dome glowing with a dull heat...The earth had come to rest with one face to the sun, even as in our own time the moon faces the earth...There were no breakers and no waves, for not a breath of wind was stirring. Only a slight oily swell rose and fell like a gentle breathing, and showed that the eternal sea was still moving and living. ...the life of the old earth ebb[s] away..."

This gloomy prognosis has been re-examined recently: it appears that the expansion of the universe may provide loopholes. While there is little doubt that the sun and the stars will indeed die, it may be possible – if life can evolve to accommodate the changing environments – that civilization, memory, and thought could continue for an indefinite subjective time (*e.g.*, exercise 5.1).

## 5.2 Entropy as Disorder

A second intuitive interpretation of entropy is as a measure of the disorder in a system. Scientist mothers tell their children to lower the entropy by tidying their rooms; liquids have higher entropy than crystals intuitively because their atomic positions are less orderly.[11] We illustrate this interpretation by first calculating the *entropy of mixing*, and then discussing the *zero-temperature entropy of glasses*.

### 5.2.1 Entropy of Mixing: Maxwell's Demon and Osmotic Pressure

Scrambling an egg is a standard example of irreversibility: you can't re-separate the yolk from the white. A simple model for scrambling is given in figures 5.4 and 5.5: the mixing of two different types of particles. Here the entropy change upon mixing is a measure of increased disorder.

Consider a volume separated by a partition into two equal volumes of volume $V$. $N/2$ indistinguishable ideal gas white atoms are on one side of the partition, and $N/2$ indistinguishable ideal gas black atoms are on the other side. The configurational entropy of this system (section 3.5, ignoring the momentum space parts) is

$$S_{\text{unmixed}} = 2\,k_B \log(V^{N/2}/{}^{N}/_{2}!) \tag{5.14}$$

just twice the configurational entropy of $N/2$ atoms in a volume $V$. We assume that the black and white atoms have the same masses and the same total

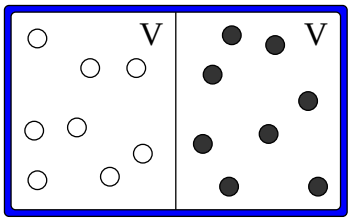



**Fig. 5.4** The pre-mixed state: $N/2$ white atoms on one side, $N/2$ black atoms on the other.

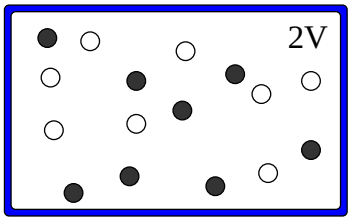



**Fig. 5.5** The mixed state: $N/2$ white atoms and $N/2$ black atoms scattered through the volume $2V$.

[11] There are interesting examples of systems that appear to develop more order as their entropy (and temperature) rises. These are systems where adding order of one, visible type (say, crystalline or orientational order) allows increased disorder of another type (say, vibrational disorder). Entropy is a precise measure of disorder, but is not the only possible or useful measure.

energy. Now consider the entropy change when the partition is removed, and the two sets of atoms are allowed to mix. Because the temperatures and pressures from both sides are equal, removing the partition does not involve any irreversible sound emission or heat transfer: any entropy change is due to the mixing of the white and black atoms. In the desegregated state,[12] the entropy has increased to

$$S_{\text{mixed}} = 2k_B \log((2V)^{N/2}/(N/2)!), \tag{5.15}$$

twice the entropy of $N/2$ indistinguishable atoms in a volume $2V$. Since $\log(2^m x) = m \log 2 + \log x$, the change in entropy due to the mixing is

$$\Delta S_{\text{mixing}} = S_{\text{mixed}} - S_{\text{unmixed}} = k_B \log 2^N = N k_B \log 2. \tag{5.16}$$

We gain $k_B \log 2$ in entropy every time we placed an atom into one of two boxes without looking which box we chose. More generally, we might define a counting entropy

$$S_{\text{counting}} = k_B \log(\text{Number of configurations}) \tag{5.17}$$

for systems with a discrete number of equally-likely configurations.

This kind of discrete choice arises often in statistical mechanics. In equilibrium quantum mechanics (for a finite system) the states are quantized: so adding a new (non-interacting) particle into one of $m$ degenerate states adds $k_B \log m$ to the entropy. In communications theory (section 5.3.2, exercises 5.6 and 5.7), each bit transmitted down your channel can be in one of two states, so a random stream of bits of length $N$ has $\Delta S = k_S N \log 2$.[13]

In more general cases, the states available to one particle depend strongly on the configurations of the other particles. Nonetheless, the equilibrium entropy still measures the logarithm of the number of different states that the total system could be in. For example, our equilibrium statistical mechanics entropy $S_{\text{equil}}(E) = k_B \log(\Omega(E))$ (equation 3.26) is the logarithm of the number of states of energy $E$, with phase-space volume $h^{3N}$ allocated to each state.

What would happen if we removed a partition separating $N/2$ black atoms on one side from $N/2$ indistinguishable black atoms on the other? The initial entropy is the same as above $S^{\text{BB}}_{\text{unmixed}} = 2\, k_B \log(V^{N/2}/(N/2)!)$, but the final entropy is now $S^{\text{BB}}_{\text{mixed}} = k_B \log((2V)^N/N!)$ Notice we have $N!$ rather than the $(N/2!)^2$ from equation 5.15, since all of our particles are now indistinguishable. Now $N! = (N) \cdot (N-1) \cdot (N-2) \cdot (N-3) \ldots$ and $(N/2!)^2 = (N/2) \cdot (N/2) \cdot (N-2)/2 \cdot (N-2)/2 \ldots$: they roughly differ by $2^N$, canceling the entropy change due to the volume doubling. Indeed, expanding the logarithm using Stirling's formula $\log n! \approx n \log n - n$ we find the entropy per atom is unchanged.[14] This is why we introduced the $N!$ term for indistinguishable particles in section 3.2.1: without it the entropy would decrease by $N \log 2$ whenever we split a container into two pieces.[15]

How can we intuitively connect this entropy of mixing with the thermodynamic entropy of pistons and engines in section 5.1? Can we use

[12] No social policy implications are implied by physics: the entropy of mixing for a few billion humans would not provide for an eye blink.

[13] Here it is natural to measure entropy not in units of temperature, but rather in base 2, so $k_S = 1/\log 2$. This means that $\Delta S = N$, for a random string of $N$ bits.

[14] If you keep Stirling's formula to higher order, you'll see that the entropy increases a bit when you remove the partition. This is due to the number fluctuations on the two sides that are now allowed.

[15] This is often called the Gibbs paradox.

our mixing entropy to do work? Clearly to do so we must discriminate between the two kinds of atoms. Suppose that the barrier separating the two walls in figure 5.4 was a membrane that was impermeable to black atoms but allowed white ones to cross. Since both black and white atoms are ideal gases, the white atoms would spread uniformly to fill the entire system, while the black atoms would remain on one side. This would lead to a pressure imbalance: if the semipermeable wall were used as a piston, work could be extracted as the black chamber was enlarged to fill the total volume.[16]

Suppose we had a more active discrimination? Maxwell introduced the idea of an intelligent 'finite being' (later termed *Maxwell's Demon*) that would operate a small door between the two containers. When a black atom approaches the door from the left or a white atom approaches from the right the demon would open the door: for the reverse situations the demon would leave the door closed. As time progresses, this active sorting would re-segregate the system, lowering the entropy. This is not a concern for thermodynamics, since of course running a demon is an entropy consuming process! Indeed, one can view this thought experiment as giving a fundamental limit on demon efficiency, putting a lower bound on how much entropy an intelligent being must create in order to engage in this kind of sorting process (see figure 5.6).

[16]Such semipermeable membranes are quite common not for gases but for dilute solutions of ions in water: some ions can penetrate and others cannot. The resulting force on the membrane is called *osmotic pressure*.

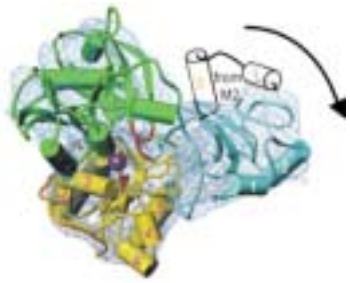


**Fig. 5.6 Ion pump.** An implementation of Maxwell's demon in biology is $Na^+/K^+$-ATPase, an enzyme located on the membranes of almost every cell in your body. This enzyme maintains extra potassium (K+) ions inside the cell and extra sodium ($Na^+$) ions outside the cell. The enzyme exchanges two $K^+$ ions from outside for three $Na^+$ ions inside, burning as fuel one $ATP$ (adenosine with three phosphates, the fuel of the cell) into $ADP$ (two phosphates). When you eat too much salt ($Na^+Cl^-$), the extra sodium ions in the blood increase the osmotic pressure on the cells, draw more water into the blood, and increase your blood pressure. The figure (©Nature 2000) shows the structure of the related enzyme calcium ATPase [118]: the arrow shows the shape change as the two $Ca^+$ ions are removed.

### 5.2.2 Residual Entropy of Glasses: The Roads Not Taken

In condensed-matter physics, glasses are the prototype of disordered systems. Unlike a crystal, in which each atom has a set position, a glass will have a completely different configuration of atoms each time it is formed. That is, the glass has a *residual entropy*: as the temperature goes to absolute zero, the glass entropy does not vanish, but rather equals $k_B \log \Omega_{\rm glass}$, where $\Omega_{\rm glass}$ is the number of zero-temperature configurations in which the glass might be trapped.

*What is a glass?*[17] Glasses are disordered like liquids, but are rigid like crystals. They are not in equilibrium: they are formed when liquids are cooled too fast to form the crystalline equilibrium state.[18] You are aware of glasses made from silica, like window glass,[19] and Pyrex.[TM20] You also know some molecular glasses, like hard candy (a glass made of sugar). Many other materials (even metals)[21] can form glasses when cooled quickly.

[17]See also section 12.3.3.

[18]The crystalline state must be *nucleated*, see section 11.2.

[19]Windows are made from soda-lime glass, with silica ($SiO_2$) mixed with sodium and calcium oxides.

[20]Pyrex[TM] is a borosilicate glass (boron and silicon oxides) with a low thermal expansion, used for making measuring cups that don't shatter when filled with boiling water.

[21]Most metals are polycrystalline. That is, the atoms sit in neat crystalline arrays but the metal is made up of many grains with different crystalline orientations separated by sharp grain boundaries.

[22] One can measure the entropy of the equilibrium liquid $S_{\rm liquid}(T_\ell)$ by slowly heating a crystal of the material from absolute zero and measuring $\int_0^{T_\ell} dQ/T$ flowing in.

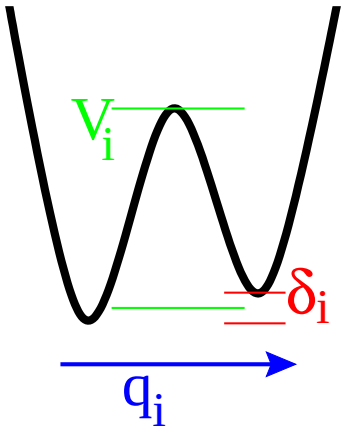



**Fig. 5.7 Double well potential.** A simple model for the potential energy for one coordinate $q_i$ in a glass: two states separated by a barrier $V_i$ and with a small energy difference $\delta_i$.

*How is the residual glass entropy measured?* First, one estimates the entropy of the equilibrium liquid;[22] then one measures the entropy flow $Q/T$ out from the glass as it is cooled from the liquid down to absolute zero. The difference

$$S_{\rm residual} = S_{\rm liquid}(T_\ell) - \int \frac{1}{T}\frac{dQ}{dt}dt = S_{\rm liquid}(T_\ell) - \int_0^{T_\ell} \frac{1}{T}\frac{dQ}{dT}dT \quad (5.18)$$

gives the residual entropy.

*How big is the residual entropy of a typical glass?* The residual entropy is on the order of $k_B$ per molecular unit of the glass ($SiO_2$ or sugar molecule, for example). This means that the number of glassy configurations $e^{S/k_B}$ is enormous (exercise 5.8 part (c)).

*How is it possible to measure the number of glass configurations the system didn't choose?* The glass is, after all, in one particular configuration. How can measuring the heat flow $Q(t)$ out of the liquid as it freezes into one glassy state be used to measure the number $\Omega_{\rm glass}$ of possible glassy states? In other words, how exactly is the statistical mechanics definition of entropy $S_{\rm stat} = k_B \log \Omega_{\rm glass}$ tied to the thermodynamic definition $\Delta S_{\rm thermo} = Q/T$?

We need first a simplified model of how a glass might fall out of equilibrium as it is cooled.[23] We view the glass as a collection of independent molecular units. Each unit has a double-well potential energy: along some internal coordinate $q_i$ there are two minima with an energy difference $\delta_i$ and separated by an energy barrier $V_i$ (figure 5.7). This internal coordinate might represent a rotation of a sugar molecule, or a shift in the location of an oxygen in a $SiO_2$ network.

Consider the behavior of one of these double-well degrees of freedom. As we cool our system, the molecular unit will be thermally excited over its barrier more and more slowly. So long as the cooling rate $\Gamma_{\rm cool}$ is small compared to this hopping rate, our systems will remain in equilibrium. However, at the temperature $T_i$ where the two rates cross for our unit

[23] The glass transition is not a sharp phase transition: the liquid grows thicker (more viscous) as it is cooled, with slower and slower dynamics, until the cooling rate becomes too fast for the atomic rearrangements needed to maintain equilibrium to keep up. At that point, there is a gradual, smeared–out transition over many degrees Kelvin as the viscosity effectively becomes infinite and the glass becomes bonded together. The fundamental nature of this transition remains controversial, and in particular we do not know why the viscosity diverges so rapidly in so many materials. There are at least three kinds of competing theories for the glass transition:

(1) It reflects an underlying equilibrium transition to an ideal, zero entropy glass state, which would be formed under infinitely slow cooling

(2) It is a purely dynamical transition (where the atoms or molecules jam together) with no thermodynamic signature.

(3) It is not a transition at all, but just a crossover where the liquid viscosity jumps rapidly (say, because of the formation of semipermanent covalent bonds).

Our simple model is not a good description of the glass transition, but is a rather accurate model for the continuing thermal rearrangements ($\beta$-relaxation) at temperatures below the glass transition, and an excellent model for the quantum dynamics (tunneling centers) which dominate many properties of glasses below a few degrees Kelvin.

the transitions between the wells will not keep up and our molecular unit will freeze into position. If the cooling rate $\Gamma_{\rm cool}$ is very slow compared to the attempt frequency $\Gamma_0$ (as it almost always is)[24] this transition will be fairly abrupt, and our model glass will freeze into the upper well with the probability given by the equilibrium distribution at $T_i$.

Our frozen molecular unit has a population in the upper well given by the Boltzmann factor $e^{-\delta_i/k_BT_i}$ times the population in the lower well. Hence, those centers with $\delta_i \gg k_BT_i$ will be primarily in the ground state (and hence already roughly in equilibrium). However, consider those $N$ centers with barriers high compared to the asymmetry, so $\delta_i \ll k_BT_i$. As the glass is cooled, one by one these units randomly freeze into one of two states (figure 5.8). For these centers, both states will be roughly equally populated when they fall out of equilibrium, so each will contribute about $k_B \log 2$ to the residual entropy. Thus, roughly speaking, the $N$ units with $T_i > \delta_i/k_B$ will contribute about $k_B \log 2 \sim k_B$ to the statistical entropy $S_{\rm stat}$.[25]

What about the thermodynamic entropy? Those centers with $k_BT_i \ll \delta_i$ which equilibrate into the lower well before they freeze will contribute the same amount to the entropy flow into the heat bath as they would in an equilibrium system. On the other hand, those centers with $k_BT \gg \delta_i$ will each fail (half the time) to release their energy $\delta_i$ to the heat bath, when compared to an infinitely slow equilibrium quench. Since in an equilibrium quench this heat would be transferred to the bath at a temperature around $\delta_i/k_B$, the missing entropy flow for that center is $\Delta Q/T \sim \delta_i/(\delta_i/k_B) \sim k_B$. Again, the $N$ units each contribute around $k_B$ to the experimentally measured thermodynamic residual entropy $S_{\rm thermo}$.

Thus the heat flow into a *particular* glass configuration counts the number of roads not taken by the glass on its cooling voyage.

[24] Atomic times like $1/\Gamma_0$ are around $10^{-12}$ per second (an atomic vibration time), and cooling times are typically between seconds and years, so the cooling rate is indeed slow compared to microscopic times.

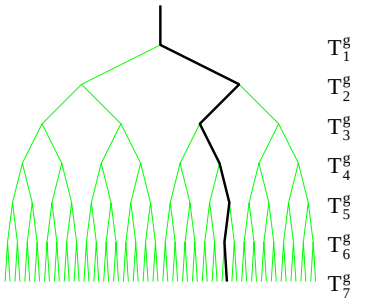



**Fig. 5.8 Roads Not Taken by the Glass.** The branching path of glassy states in our model. The entropy (both statistical and thermodynamic) is proportional to the number of branchings the glass chooses between as it cools. A particular glass will take one trajectory through this tree as it cools: nonetheless the thermodynamic entropy measures the total number of states.

[25] We're losing factors like $\log 2$ because we're ignoring those units with $k_BT_i \sim \delta_i$, which freeze into partly occupied states that aren't equally occupied, a case we haven't treated yet. In section 5.3.1 we'll introduce the general definition for the statistical entropy: using it, a (somewhat more complicated) calculation shows that the statistical and thermal definitions of the entropy agree without fudging.

## 5.3 Entropy as Ignorance: Information and Memory

The most general interpretation of entropy is as a measure of our *ignorance*[26] about a system. The equilibrium state of a system maximizes the entropy because we have lost all information about the initial conditions except for the conserved quantities: maximizing the entropy maximizes our ignorance about the details of the system. The entropy of a glass, or of our mixture of black and white atoms, is a measure of the number of arrangements the atoms could be in, given our ignorance.[27]

This interpretation – that entropy is not a property of the system, but of our knowledge about the system[28] (represented by the ensemble of possibilities) – cleanly resolves many otherwise confusing issues. The

[26] In information theory they use the alternative term *uncertainty*, which has misleading connotations from quantum mechanics: Heisenberg uncertainty has no associated entropy.

[27] Again, entropy is a precise measure of ignorance, but not necessarily a sensible one for all purposes. In particular, entropy does not distinguish the *utility* of the information. Isothermally compressing a mole of gas to half its volume decreases our ignorance by $10^{23}$ bits: a far larger change in entropy than would be produced by memorizing all the written works of human history.

[28] The entropy of an *equilibrium* system remains purely a property of the composition of the system, because our knowledge is fixed (at zero). Even systems out of equilibrium have bounds on the entropy given by the history of heat (entropy) exchange with the outside world (exercise 5.8).

atoms in a glass are in a definite configuration, which we could measure using some futuristic X-ray holographic technique. If we did so, our ignorance would disappear, and the residual entropy would become zero for us.[29] We could in principle use our knowledge of the glass atom positions to extract extra useful work out of the glass, not available before measuring the positions.

[29] Of course, the X-ray holographic process must create at least as much entropy during the measurement as the glass loses.

So far, we have confined ourselves to cases where our ignorance is maximal, where all allowed configurations are equally likely. What about systems where we have partial information, where some configurations are more probable than others? There is a powerful generalization of the definition of entropy to general probability distributions, which we will introduce in section 5.3.1 for traditional statistical mechanical systems. In section 5.3.2 we will show that this nonequilibrium entropy provides a generally useful measure of our ignorance about a wide variety of systems, with broad applications outside of traditional physics.

### 5.3.1 Nonequilibrium Entropy

So far, we have defined the entropy only for systems in equilibrium, where entropy is a constant. But the second law of thermodynamics tells us that entropy increases – presupposing some definition of entropy for non-equilibrium systems. In general, we may describe our partial knowledge about a system as a probability distribution $\rho$, defining the ensemble of states.

Let us start with a probability distribution among a discrete set of states. We know from section 5.2.1 that the entropy for $M$ equally likely states (equation 5.17) is $S(M) = k_B \log M$. In this case, the probability of each state is $p_i = 1/M$. If we write $S(M) = -k_B \log(1/M) = -k_B \langle \log(p_i) \rangle$, we get an appealing generalization for the counting entropy for cases where $p_i$ is not constant:

$$S_{\text{discrete}} = -k_B \langle \log p_i \rangle = -k_B \sum_i p_i \log p_i. \tag{5.19}$$

We shall see in section 5.3.2 and exercise 5.10 that this is clearly the correct generalization of entropy to systems out of equilibrium.

What about continuum distributions? Any non-equilibrium state of a classical Hamiltonian system can be described with a probability density $\rho(\mathbb{P}, \mathbb{Q})$ on phase space. The non-equilibrium entropy then becomes

$$\begin{aligned} S_{\text{nonequil}} &= -k_B \langle \log \rho \rangle = -k_B \int \rho \log \rho \\ &= -k_B \int_{E<\mathcal{H}(\mathbb{P},\mathbb{Q})<E+\delta E} d\mathbb{P} d\mathbb{Q}\, \rho(\mathbb{P}, \mathbb{Q}) \log \rho(\mathbb{P}, \mathbb{Q}). \end{aligned} \tag{5.20}$$

In the case of the microcanonical ensemble where $\rho_{\text{equil}} = \frac{1}{\Omega(E)\delta E}$, the non-equilibrium definition of the entropy is shifted from our equilibrium definition $S = k_B \log \Omega$ by a negligable amount $k_B \log(\delta E)/N$ per par-

ticle:[30]

$$S_{\text{micro}} = -k_B \log \rho_{\text{equil}} = k_B \log(\Omega(E)\delta E) \\ = k_B \log(\Omega(E)) + k_B \log(\delta E). \tag{5.21}$$

For quantum systems, the non-equilibrium entropy will written in terms of the density matrix $\boldsymbol{\rho}$ (section 7.1):

$$S_{\text{quantum}} = -k_B Tr(\boldsymbol{\rho} \log \boldsymbol{\rho}). \tag{5.22}$$

Finally, notice that $S_{\text{noneq}}$ and $S_{\text{quantum}}$ are defined for the microscopic laws of motion, which (section 5.1) are time-reversal invariant. We can thus guess that these microscopic entropies will be time–independent, since microscopically the system does not know in which direction of time entropy should increase.[31] No information is lost (in principle) by evolving a closed system in time. Entropy (and our ignorance) increases only in coarse-grained theories where we ignore or exclude some degrees of freedom (internal or external).

[31] You can show this explicitly in exercises 5.4 and 7.2).

### 5.3.2 Information Entropy

Understanding ignorance is central to many fields! Entropy as a measure of ignorance has been useful in everything from the shuffling of cards to reconstructing noisy images. For these other applications, the connection with temperature is unimportant, so we don't need to make use of Boltzmann's constant. Instead, we normalize the entropy with the constant $k_S = 1/\log(2)$:

$$S_S = -k_S \sum_i p_i \log p_i. \tag{5.23}$$

This normalization was introduced by Shannon [112], and the formula 5.23 is referred to as Shannon entropy in the context of *information theory*. Shannon noted that this entropy, applied to the ensemble of possible messages or images, can be used to put a fundamental limit on the amount they can be compressed[32] to efficiently make use of disk space or a communications channel (exercises 5.6 and 5.7). A low entropy data set is highly predictable: given the stream of data so far, we can predict the next transmission with some confidence. In language, twins and long-married couples can often complete sentences for one another. In image transmission, if the last six pixels were white the region being depicted is likely a white background, and the next pixel is also likely white. We need only transmit or store data that violates our prediction. The entropy measures our ignorance, how likely the best predictions about the rest of the message are to be wrong.

[32] Lossless compression schemes (files ending in gif, png, zip, and gz) remove the redundant information in the original files, and their efficiency is limited by the entropy of the ensemble of files being compressed. Lossy compression schemes (files ending in jpg, mpg, and mp3) also remove information that is thought to be unimportant for humans looking at or listening to the files.

Entropy is so useful in these various fields because it is the unique (continuous) function that satisfies three key properties.[33] In this sec-

[33] Unique, that is, up to the overall constant $k_S$ or $k_B$.

[30] The arbitrary choice of the width of the energy shell in the microcanonical ensemble thus is related to the arbitrary choice of the zero for the entropy of a classical system. Unlike the shift due to the units of phase space (section 3.5), this shift is microscopic.

tion, we will first explain what these three properties are and why they are natural for any function that measures ignorance. We will show our nonequilibrium Shannon entropy satisfies these properties; in exercise 5.10 you will show that this entropy is the only function to do so.

To take a tangible example of ignorance, suppose your roommate has lost their keys, and they are asking for your advice. You want to measure the roommate's progress in finding the keys by measuring your ignorance with some function $S_I$. Suppose there are $\Omega$ possible sites $A_k$ that they might have left the keys, which you estimate have probabilities $p_k = P(A_k)$, with $\sum_1^\Omega p_i = 1$.

What are the three key properties we want our ignorance function $S_I(p_1, \ldots, p_\Omega)$ to have?

*(1) Entropy is maximum for equal probabilities.* Without further information, surely the best plan is for your roommate to look first at the most likely site, which maximizes $p_i$. Your ignorance must therefore be maximal if all $\Omega$ sites have equal likelihood:

$$S_I\left(1/\Omega, \ldots, 1/\Omega\right) > S_I(p_1, \ldots, p_\Omega) \quad \text{unless } p_i = 1/\Omega \text{ for all } i. \tag{5.24}$$

Does the Shannon entropy obey property (1), equation 5.24?[34] We notice that the function $f(p) = -p \log p$ is concave (convex downward, figure 5.9). For a concave function $f$, the average value of $f(p)$ over a set of points $p_k$ is less than than or equal to $f$ evaluated at the average:[35]

$$1/\Omega \sum_k f(p_k) \le f\left(1/\Omega \sum_k p_k\right). \tag{5.26}$$

But this tells us that

$$\begin{aligned} S_S(p_1, \ldots, p_\Omega) &= -k_S \sum p_k \log p_k = k_S \sum f(p_k) \\ &\le k_S \Omega f\left(1/\Omega \sum_k p_k\right) = k_S \Omega f(1/\Omega) \\ &= -k_S \sum_1^\Omega 1/\Omega \log(1/\Omega) = S_S(1/\Omega, \ldots, 1/\Omega). \end{aligned} \tag{5.27}$$

[34] In exercise 6.7 you'll show that the Shannon entropy $S_S$ is an extremum when all probabilities are equal. Here we provide a proof that is a global maximum, using the convexity of $x \log x$ (figure 5.9).

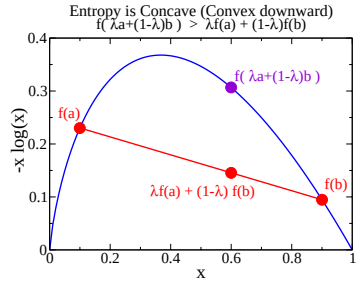



**Fig. 5.9 Entropy is Concave.** For $x \ge 0$, $f(x) = -x \log x$ is strictly convex downward (concave). That is, for $0 < \lambda < 1$, the linear interpolation lies below the curve:

$$f(\lambda a + (1-\lambda)b) \ge \lambda f(a) + (1-\lambda) f(b). \tag{5.28}$$

We know $f$ is concave because its second derivative, $-1/x$, is everywhere negative.

[35] Equation 5.26 can be proven by induction from the definition of concave (equation 5.28). For $\Omega = 2$, we use $\lambda = 1/2$, $a = p_1$, and $b = p_2$ to see that $f\left(\frac{p_1+p_2}{2}\right) \ge \frac{1}{2}\left(f(p_1) + f(p_2)\right)$. For general $\Omega$, we use $\lambda = (\Omega-1)/\Omega$, $a = (\sum_1^{\Omega-1} p_k)/(\Omega-1)$, and $b = p_\Omega$ to see

$$\begin{aligned} f\left(\frac{\sum_{k=1}^{\Omega} p_k}{\Omega}\right) &= f\left(\frac{\Omega-1}{\Omega}\frac{\sum_1^{\Omega-1} p_k}{\Omega-1} + 1/\Omega p_\Omega\right) \ge \frac{\Omega-1}{\Omega} f\left(\frac{\sum_1^{\Omega-1} p_k}{\Omega-1}\right) + 1/\Omega f(p_\Omega) \\ &\ge \frac{\Omega-1}{\Omega}\left(\sum_{k=1}^{\Omega-1} \frac{1}{\Omega-1} f(p_k)\right) + 1/\Omega f(p_\Omega) = 1/\Omega \sum_{k=1}^{\Omega} f(p_k) \end{aligned} \tag{5.25}$$

where in the third line we have used the truth of equation 5.26 for $\Omega - 1$ to inductively prove it for $\Omega$.

*(2) Entropy is unaffected by extra states of zero probability.* If there is no possibility that the keys are in your shoe (site $A_\Omega$), then your ignorance is no larger than it would have been if you hadn't included your shoe in the list of possible sites:

$$S_I(p_1, \ldots, p_{\Omega-1}, 0) = S_I(p_1, \ldots, p_{\Omega-1}). \tag{5.29}$$

The Shannon entropy obeys property (2) because $p_\Omega \log p_\Omega \to 0$ as $p_\Omega \to 0$.

*(3) Entropy change for conditional probabilities.* This last property for our ignorance function demands a new concept, *conditional probablility.*

To aid in the search, you'll likely ask the roommate where they were when they last saw the keys. Suppose there are $M$ locations $B_\ell$ that the roommate may have been (opening the apartment door, driving the car, in the basement laundry room, ...), with probabilities $q_\ell$. Surely the likelihood that the keys are currently in a coat pocket is larger if the roommate was outdoors when the keys were last seen. Let $r_{k\ell} = P(A_k \text{ and } B_\ell)$ be the probability the keys are at site $k$ and were last seen at location $\ell$, and[36]

$$P(A_k|B_\ell) = c_{k\ell} = r_{k\ell}/q_\ell \tag{5.30}$$

be the *conditional probability*, given that they were last seen at $B_\ell$ that the keys are at site $A_k$. Clearly

$$\sum_k P(A_k|B_\ell) = \sum_k c_{k\ell} = 1 : \tag{5.31}$$

wherever they were last seen, the keys are now *somewhere* with probability one.

Before you ask your roommate where the keys were last seen, you have ignorance $S_I(A) = S_I(p_1, \ldots, p_\Omega)$ about the site of the keys, and ignorance $S_I(B) = S_I(q_1, \ldots, q_M)$ about the location they were last seen. You have a joint ignorance about the two questions given by the ignorance function applied to all $\Omega \times M$ conditional probabilities:

$$\begin{aligned} S_I(AB) &= S_I(r_{11}, r_{12}, \ldots, r_{1M}, r_{21}, \ldots, r_{\Omega M}) \\ &= S_I(c_{11}q_1, c_{12}q_2, \ldots, c_{1M}q_M, c_{21}q_1, \ldots, c_{\Omega M}q_M). \end{aligned} \tag{5.32}$$

After the roommate answers your question, your ignorance about the location last seen is reduced to zero (decreased by $S_I(B)$). If the location last seen was in the laundry room (site $B_\ell$), the probability for the keys being at $A_k$ shifts to $c_{k\ell}$ and your ignorance about the site of the keys is now

$$S_I(A|B_\ell) = S_I(c_{1\ell}, \ldots, c_{\Omega\ell}). \tag{5.33}$$

So, your combined ignorance has decreased from $S_I(AB)$ to $S_I(A|B_\ell)$.

We can measure the usefulness of your question by the expected amount that it decreases your ignorance about where the keys reside.

[36] The conditional probability $P(A|B)$ [read "P of A given B"] times the probability of B is of course the probability of $A$ and $B$ both occurring, so $P(A|B)P(B) = P(A \text{ and} B)$, implying $c_{k\ell}q_\ell = r_{k\ell}$.

The expected ignorance after the question is answered is given by weighting the ignorance after each answer $B_\ell$ by the probability $q_\ell$ of that answer:

$$\langle S_I(A|B_\ell)\rangle_B = \sum_\ell q_\ell S_I(A|B_\ell). \tag{5.34}$$

This leads us to the third key property for an ignorance function. If we start with the joint distribution $AB$, and then measure $B$, it would be tidy if, on average, your joint ignorance declined by your original ignorance of $B$:

$$\langle S_I(A|B_\ell)\rangle_B = S_I(AB) - S_I(B). \tag{5.35}$$

Does the Shannon entropy satisfy equation 5.35, property (3)? The conditional probability $S_S(A|B_\ell) = -k_S \sum_\ell c_{k\ell} \log c_{k\ell}$, since $c_{k\ell}$ is the probability distribution for the $A_k$ sites given location $\ell$. So,

$$\begin{aligned}
S_S(AB) &= -k_S \sum_{k\ell} c_{k\ell} q_\ell \log(c_{k\ell} q_\ell) \\
&= -k_S \left( \sum_{k\ell} c_{k\ell} q_\ell \log(c_{k\ell}) + \sum_{k\ell} c_{k\ell} q_\ell \log(q_\ell) \right) \\
&= \sum_\ell q_\ell \left( -k_S \sum_k c_{k\ell} \log(c_{k\ell}) \right) - k_S \sum_\ell q_\ell \log(q_\ell) \cancel{\left( \sum_k c_{k\ell} \right)} \\
&= \sum_\ell q_\ell S_S(A|B_\ell) + S_S(B) \\
&= \langle S_S(A|B_\ell)\rangle_B + S_S(B)
\end{aligned} \tag{5.36}$$

and the Shannon entropy does satisfy condition (3).

If $A$ and $B$ are uncorrelated (for example, if they are measurements on uncoupled systems), then the probabilities of $A$ won't change upon measuring $B$, so $S_I(A|B_\ell) = S_I(A)$. Then our third condition implies $S_I(AB) = S_I(A) + S_I(B)$: our ignorance of uncoupled systems is additive. This is simply the condition that *entropy is extensive*. We argued that the entropy of weakly coupled subsystems in equilibrium must be additive in section 3.3. Our third condition implies that this remains true for uncorrelated systems in general.

## Exercises

*Entropy is an emergent property.* Unlike energy conservation, which is inherited from the microscopic theory, entropy is a constant for a closed system treated microscopically (for Newton's laws in exercise 5.4(a), for quantum mechanics in exercise 7.2). Entropy increases because information is lost – either to the outside world, to unimportant or ignored internal degrees of freedom (diffusion equation, 5.5, Markov chains 8.8), or to measurement in-

accuracies in the initial state (Lyapunov exponents 5.11, Poincaré cat map 5.4(b)).[37]

*Entropy is a general measure of ignorance,* useful far outside its traditional applications (6.7) in equilibrium systems. It is the unique function to have the appropriate properties to measure ignorance (5.10). It has applications to glasses (5.8) and to defining fractal dimensions (5.13). It is fascinating that entropy – our ignorance about the system – can exert real forces (*e.g.* in rubber bands, 5.9).

*Entropy provides fundamental limits* on engine efficiency (5.2, 5.3), data compression (5.6, 5.7), memory storage (where the limit is the formation of a black hole, 5.12(c)), and to intelligent life at the end of the universe (5.1).

(5.1) **Life and the Heat Death of the Universe.** (Basic, Astrophysics)

Freeman Dyson [28] discusses how living things might evolve to cope with the cooling and dimming we expect during the heat death of the universe.

Normally one speaks of living things as beings that consume energy to survive and proliferate. This is of course not correct: energy is conserved, and cannot be consumed. *Living beings intercept entropy flows*: they use low entropy sources of energy (e.g., high temperature solar radiation for plants, candy bars for us) and emit high entropy forms of the same energy (body heat).

Dyson ignores the survival and proliferation issues; he's interested in getting a lot of thinking in before the universe ends. He presumes that an intelligent being generates a fixed entropy $\Delta S$ per thought.[38] (This correspondence of information with entropy is a standard idea from computer science: see exercises 5.6 and 5.7.)

**Energy needed per thought.** Assume that the being draws heat $Q$ from a hot reservoir at $T_1$ and radiates it away to a cold reservoir at $T_2$.

*(a) What is the minimum energy $Q$ needed per thought, in terms of $\Delta S$ and $T_2$? You may take $T_1$ very large.* Related formulæ: $\Delta S = Q_2/T_2 - Q_1/T_1$; First Law: $Q_1 - Q_2 = W$ (energy is conserved).

**Time needed per thought to radiate energy.** Dyson shows, using theory not important here, that the power radiated by our intelligent–being–as–entropy–producer is no larger than $CT_2^3$, a constant times the cube of the cold temperature.[39]

*(b) Write an expression for the maximum rate of thoughts per unit time $dH/dt$ (the inverse of the time $\Delta t$ per thought), in terms of $\Delta S$, $C$, and $T_2$.*

**Number of thoughts for an ecologically efficient being.** Our universe is expanding: the radius $R$ grows roughly linearly in time $t$. The microwave background radiation has a characteristic temperature $\Theta(t) \sim R^{-1}$ which is getting lower as the universe expands: this redshift is due to the Doppler effect. An ecologically efficient being would naturally try to use as little heat as possible, and so wants to choose $T_2$ as small as possible. It cannot radiate heat at a temperature below $T_2 = \Theta(t) = A/t$.

*(c) How many thoughts $H$ can an ecologically efficient being have between now and time infinity, in terms of $\Delta S$, $C$, $A$, and the current time $t_0$?*

**Time without end: Greedy beings.** Dyson would like his beings to be able to think an infinite number of thoughts before the universe ends, but consume a finite amount of energy. He proposes that his beings need to be profligate in order to get their thoughts in before the world ends: he proposes that they radiate at a temperature $T_2(t) \sim t^{-3/8}$ which falls with time, but not as fast as $\Theta(t) \sim t^{-1}$.

*(d) Show that with Dyson's cooling schedule, the total number of thoughts $H$ is infinite, but the total energy consumed $U$ is finite.*

We should note that there are many refinements on Dyson's ideas. There are potential difficulties that may arise like to quantum limits to cooling or proton decay: how will we make bodies out of electrons and neutrinos? And there are different challenges depending on the expected future of the universe: a big crunch, for example, where the universe collapses back on itself, demands that we adapt to heat and pressure (but an infinite number of thoughts appears to remains possible before the end).

(5.2) **P-V Diagram.** (Basic, Thermodynamics)

A monatomic ideal gas in a piston is cycled around the path in the P-V diagram in figure 5.10 Leg **a** cools at constant volume by connecting to a heat bath at $T_c$; leg **b** heats at constant pressure by connecting to a heat bath at $T_h$; leg **c** compresses at constant temperature while remaining connected to the bath at $T_h$.

[37] The statistical mechanics arguments for the increase of entropy all are dependent in detail on the coarse-grained models involved. Our best and most general argument that any physical model must have entropy increase is Carnot's argument: if a system could consistently evolve to lower entropy states, one could (perhaps indirectly) use that system to extract useful work from a rock.

[39] The constant scales with the number of electrons in the being, so we can think of our answer $\Delta t$ as the time per thought per mole of electrons.

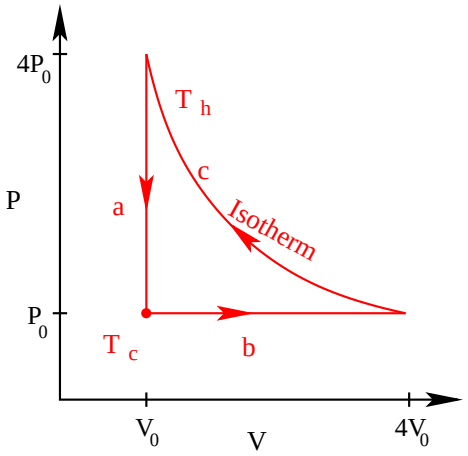



**Fig. 5.10** PV diagram

*Which of the following are true?*

*(T) (F) The cycle is reversible: no net entropy is created in the universe.*

*(T) (F) The cycle acts as a refrigerator, using work from the piston to draw energy from the cold bath into the hot bath, cooling the cold bath.*

*(T) (F) The cycle acts as an engine, transferring heat from the hot bath to the cold bath and doing positive net work on the outside world.*

*(T) (F) The work done per cycle has magnitude* $|W| = P_0V_0|4\log 4 - 3|$.

*(T) (F) The heat transferred into the cold bath,* $Q_c$ *has magnitude* $|Q_c| = (9/2)P_0V_0$.

*(T) (F) The heat transferred from the hot bath* $Q_h$*, plus the net work* $W$ *done by the piston onto the gas, equals the heat* $Q_c$ *transferred into the cold bath.*

Related formulæ: $PV = Nk_BT$, $U = (3/2)Nk_BT$, $\Delta S = Q/T$, $W = -\int PdV$, $\Delta U = Q + W$. Notice that the signs of the various terms depend on convention (heat flow out vs. heat flow in): you should figure the signs on physical grounds.

(5.3) **Carnot Refrigerator.** (Basic, Thermodynamics)

Our refrigerator is about $2m \times 1m \times 1m$, and has insulation about $3cm$ thick. The insulation is probably polyurethane, which has a thermal conductivity of about 0.02 W/(m K). Assume that the refrigerator interior is at 270K, and the room is at 300K.

*(a) How many watts of energy leak from our refrigerator through this insulation?*

Our refrigerator runs at 120 V, and draws a maximum of 4.75 amps. The compressor motor turns on every once in a while for a few minutes.

*(b) Suppose (i) we don't open the refrigerator door, (ii) the thermal losses are dominated by the leakage through the foam and not through the seals around the doors, and (iii) the refrigerator runs as a perfectly efficient Carnot cycle. How much power on average will our refrigerator need to operate? What fraction of the time will the motor run?*

(5.4) **Does Entropy Increase?** (Mathematics)

The second law of thermodynamics says that entropy always increases. Perversely, we can show that in an isolated system, no matter what non-equilibrium condition it starts in, entropy calculated with a complete microscopic description stays constant in time.

**Entropy is Constant: Classical.** [40] Liouville's theorem tells us that the total derivative of the probability density is zero: following the trajectory of a system, the local probability density never changes. The equilibrium states have probability densities that only depend on energy and number. Clearly something is wrong: if the density starts non-uniform, how can it become uniform?

*(a) Show for any function* $f(\rho)$ *that* $\partial f(\rho)/\partial t = -\nabla \cdot [f(\rho)\mathbb{V}] = -\sum_\alpha \partial/\partial p_\alpha\,(f(\rho)\dot{p}_\alpha) + \partial/\partial q_\alpha\,(f(\rho)\dot{q}_\alpha)$*, where* $\mathbb{V} = \left(\dot{\mathbb{P}}, \dot{\mathbb{Q}}\right)$ *is the* $6N$ *dimensional velocity in phase space. Hence, (by Gauss's theorem in* $6N$ *dimensions), show* $\int \partial f(\rho)/\partial t\, d\mathbb{P}d\mathbb{Q} = 0$*, assuming that the probability density vanishes at large momenta and positions and* $f(0) = 0$*. Show, thus, that the entropy* $S = -k_B \int \rho \log \rho$ *is constant in time.*

We will see that the quantum version of the entropy is also constant for a Hamiltonian system in exercise 7.2.

**The Arnol'd Cat.** Why do we think entropy increases? First, points in phase space don't just swirl in circles: they get stretched and twisted and folded back in complicated patterns – especially in systems where statistical mechanics seems to hold! Arnol'd, in a takeoff on Schrödinger's cat, suggested the following analogy. Instead of a continuous transformation of phase space onto itself preserving $6N$-dimensional volume, let's think of an area-preserving mapping of an $n \times n$ square in the plane into itself.[41]

[40] We'll see in exercise 7.2 that the non-equilibrium entropy is also constant in quantum systems.

[41] For our purposes, the Arnol'd cat just shows that volume preserving transformations can scramble a small region uniformly over a large one. More general, nonlinear

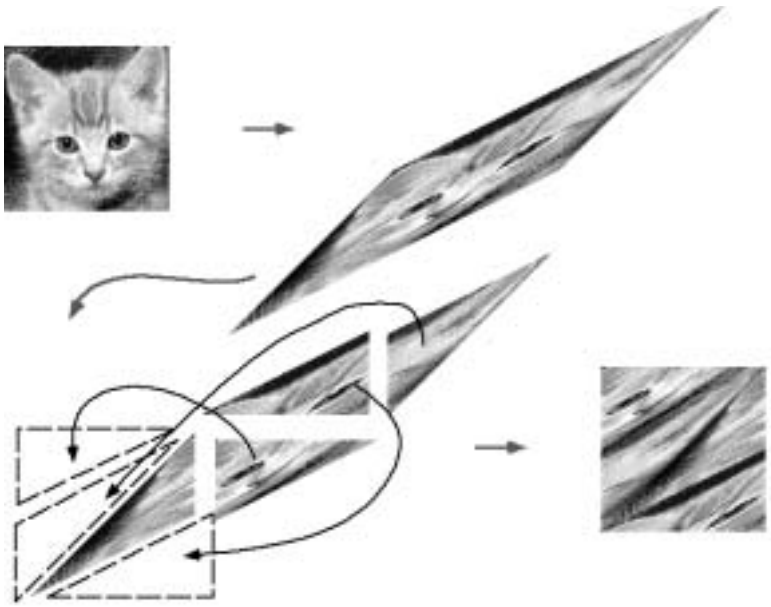


**Fig. 5.11** Arnol'd Cat Transform, from reference [83]; see movie too [92].

Consider the mapping

$$\Gamma \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x+y \\ x+2y \end{pmatrix} \bmod n. \tag{5.37}$$

See the map in figure 5.11

*(b) Check that* $\Gamma$ *preserves area. (It's basically multiplication by the matrix* $M = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$*. What is the determinant of* $M$*?). Show that it takes a square* $n \times n$ *(or a picture of* $n \times n$ *pixels) and maps it into itself with periodic boundary conditions. (With less cutting and pasting, you can view it as a map from the torus into itself.) As a linear map, find the eigenvalues and eigenvectors. Argue that a small neighborhood (say a circle in the center of the picture) will initially be stretched along an irrational direction into a thin strip (figure 5.12).*

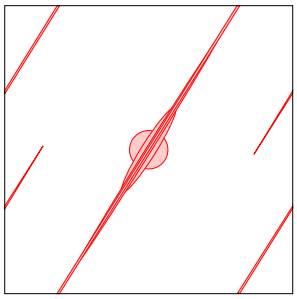


**Fig. 5.12** A small circular region stretches along an irrational angle under the Arnold cat map. The center of the figure is the origin $x = 0, y = 0$.

When this thin strip hits the boundary, it gets split into two; in the case of an $n \times n$ square, further iterations stretch and chop our original circle into a thin line uniformly covering the square. In the pixel case, there are always exactly the same number of pixels that are black, white, and each shade of gray: they just get so kneaded together that everything looks a uniform color. So, by

area–preserving maps of the plane are often studied as Hamiltonian–like dynamical systems. Area–preserving maps come up as Poincaré sections of Hamiltonian systems 4.2, with the area weighted by the inverse of the velocity with which the system passes through the cross–section. They come up in particular in studies of high–energy particle accelerators, where the mapping gives a snapshot of the particles after one orbit around the ring.

putting a limit to the resolution of our measurement (rounding errors on the computer, for example), or by introducing any tiny coupling to the external world, the final state can be seen to rapidly approach equilibrium, proofs to the contrary notwithstanding!

(5.5) **Entropy Increases: Diffusion.**

We saw that entropy technically doesn't increase for a closed system, for any Hamiltonian, either classical or quantum. However, we can show that entropy increases for most of the coarse-grained effective theories that we use in practice: when we integrate out degrees of freedom, we provide a means for the information about the initial condition to be destroyed. Here you'll show that entropy increases for the diffusion equation.

**Diffusion Equation Entropy.** Let $\rho(x,t)$ obey the one-dimensional diffusion equation $\partial\rho/\partial t = D\partial^2\rho/\partial x^2$. Assume that the density $\rho$ and all its gradients die away rapidly at $x = \pm\infty$.[42]

*Derive a formula for the time derivative of the entropy* $S = -k_B \int \rho(x) \log \rho(x)\, dx$ *and show that it strictly increases in time.* (Hint: integrate by parts. You should get an integral of a positive definite quantity.)

(5.6) **Information entropy.** (Basic, Computer Science, Mathematics, Complexity)

Entropy is a measure of your ignorance about a system: it is a measure of the lack of information. It has important implications in communication technologies: messages passed across the Ethernet communicate information, reducing the information entropy for the receiver. Shannon [112] worked out the use of entropy ideas in communications, focusing on problems where different messages have different probabilities. We'll focus on the simpler problem where all $N$ messages are equally likely. Shannon defines the *information entropy* of an unread message as being $\log_2 N = k_S \log N$, where $k_S = 1/(\log_e 2)$ is analogous to Boltzmann's constant, and changes from log-base-$e$ to log-base-2 (more convenient for computers, which think in base two.)

Your grandparent has sent you an e-mail message. From the header of the message, you know it contains 1000 characters. You know each character is made of 8 bits, which allows $2^8 = 256$ different letters or symbols per character. Assuming all possible messages from your grandparent are equally likely (a typical message would then look like G*me'!8V[beep]...), how many different messages $N$ could there be? This (unrealistic) assumption gives an upper bound for the information entropy $S_{max}$.

*(a) What* $S_{max}$ *for the unread message?*

Your grandparent writes rather dull messages: they all fall into the same pattern. They have a total of 16 equally likely messages. [43] After you read the message, you forget the details of the wording anyhow, and only remember these key points of information.

*(b) What is the actual information entropy change* $\Delta S_{Shannon}$ *you undergo when reading the message? If your grandparent writes one message per month, what is the minimum number of 8-bit characters per year that it would take to send your grandparent's messages? (You may lump multiple messages into a single character.)* (Hints: $\Delta S_{shannon}$ is the change in entropy from before you read the message to after you read which of 16 messages it was. The length of 1000 is *not* important for this part.)

Remark: This is an extreme form of data compression, like that used in gif images, zip files (Windows) and gz files (Unix). We are asking for the number of characters per year for an optimally compressed signal.

(5.7) **Shannon entropy.** (Computer Science)

Entropy can be viewed as a measure of the lack of information you have about a system. Claude Shannon [112] realized, back in the 1940's, that communication over telephone wires amounts to reducing the listener's uncertainty about the sender's message, and introduced a definition of an information entropy.

Most natural languages (voice, written English) are highly redundant; the number of intelligible fifty-letter sentences is many fewer than $26^{50}$, and the number of ten-second phone conversations is far smaller than the number of sound signals that could be generated with frequencies between up to 20,000 Hz.[44] Shannon, knowing

[42] Also, you may assume $\partial^n\rho/\partial x^n \log\rho$ goes to zero at $x = \pm\infty$, even though $\log\rho$ goes to $-\infty$.

[43] Each message mentions whether they won their bridge hand last week (a fifty-fifty chance), mentions that they wish you would write more often (every time), and speculates who will win the women's college basketball tournament in their region (picking at random one of the eight teams in the league).

[44] Real telephones don't span this whole frequency range: they are limited on the low end at 300–400 Hz, and on the high end at 3000–3500. You can still understand the words, so this crude form of data compression is only losing non-verbal nuances in the communication [36].

statistical mechanics, defined the entropy of an ensemble of messages: if there are $N$ possible messages that can be sent in one package, and message $m$ is being transmitted with probability $p_m$, then Shannon's entropy is

$$S_I = -k_S \sum_1^N p_m \log p_m \tag{5.38}$$

where instead of Boltzmann's constant, Shannon picked $k_S = 1/\log 2$.

This immediately suggests a theory for signal compression. If you can recode the alphabet so that common letters and common sequences of letters are abbreviated, while infrequent combinations are spelled out in lengthly fashion, you can dramatically reduce the channel capacity needed to send the data. (This is lossless compression, like zip and gz and gif).

An obscure language A'bç! for long-distance communication has only three sounds: a hoot represented by A, a slap represented by B, and a click represented by C. In a typical message, hoots and slaps occur equally often ($p = 1/4$), but clicks are twice as common ($p = 1/2$). Assume the messages are otherwise random.

*(a) What is the Shannon entropy in this language? More specifically, what is the Shannon entropy rate (entropy per sound, or letter, transmitted)?*

*(b) Show that a communication channel transmitting bits (ones and zeros) can transmit no more than one unit of Shannon entropy per bit.* (Hint: this should follow by showing that, for $N = 2^n$ messages, equation 5.38 is maximized by $p_m = 1/N$. You needn't prove it's a global maximum: check that it is a local extremum. You'll need either a Lagrange multiplier or will need to explicitly set $p_N = 1 - \sum_{m=1}^{N-1} p_m$.)

*(c) In general, argue that the Shannon entropy gives the minimum number of bits needed to transmit the ensemble of messages.* (Hint: compare the Shannon entropy of the $N$ original messages with the Shannon entropy of the $N$ (shorter) encoded messages.) Calculate the minimum number of bits per letter on average needed to transmit messages for the particular case of an A'bç! communication channel.

*(d) Find a compression scheme (a rule that converts a A'bç! message to zeros and ones, that can be inverted to give back the original message) that is optimal, in the sense that it saturates the bound you derived in part (b).* (Hint: Look for a scheme for encoding the message that compresses one letter at a time. Not all letters need to compress to the same number of bits.)

Shannon also developed a measure of the channel capacity of a noisy wire, and discussed error correction codes...

(5.8) **Entropy of Glasses.** [61]

Glasses aren't really in equilibrium. In particular they do not obey the third law of thermodynamics, that the entropy $S$ goes to zero at zero temperature. Experimentalists measure a "residual entropy" by subtracting the entropy change from the known entropy of the equilibrium liquid at a temperature $T_\ell$ at or above the crystalline melting temperature $T_c$ (equation 5.18):

$$S_{\text{residual}} = S_{\text{liquid}}(T_\ell) - \int_0^{T_\ell} \frac{1}{T}\frac{dQ}{dT} dT \tag{5.39}$$

where Q is the net heat flow out of the bath into the glass.

If you put a glass in an insulated box, it will warm up (very slowly) because of microscopic atomic rearrangements which lower the potential energy. So, glasses don't have a well-defined temperature or specific heat. In particular, the heat flow upon cooling and on heating $\frac{dQ}{dT}(T)$ won't precisely match (although their integrals will agree by conservation of energy).

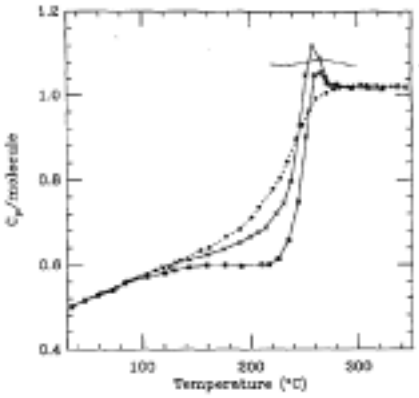



**Fig. 5.13** Specific heat of $B_2O_3$ glass measured while heating and cooling. The glass was first rapidly cooled from the melt (500°C → 50°C in a half hour), then heated from 33°C → 345°C in 14 hours (solid curve with squares), cooled from 345°C to room temperature in 18 hours (dotted curve with diamonds), and finally heated from 35°C → 325°C (solid curve with crosses). Figure from reference [117], see also [59].

Thomas and Parks in figure 5.13 are making the approximation that the specific heat of the glass is $dQ/dT$, the measured heat flow out of the glass divided by the temperature change of the heat bath. They find that the

[45] The fact that the energy lags the temperature near the glass transition, in linear response, leads to the study of "specific heat spectroscopy" [11].

specific heat defined in this way measured on cooling and heating disagree. [45] Consider the second cooling curve and the final heating curve, from 325°C to room temperature and back. Assume that the liquid at 325°C is in equilibrium both before cooling and after heating (and so has the same liquid entropy $S_{\text{liquid}}$).

*(a) Is the residual entropy, equation 5.39, experimentally larger on heating or on cooling in figure 5.13?* (Hint: Use the fact that the integrals under the curves, $\int_0^{T_\ell} \frac{dQ}{dT} dT$ give the heat flow, which by conservation of energy must be the same on heating and cooling. The heating curve shifts weight to higher temperatures: will that increase or decrease the integral in 5.39?)

*(b) By using the second law (entropy can only increase), show that when cooling and then heating from an equilibrium liquid the residual entropy measured on cooling must always be less than the residual entropy measured on heating.* Your argument should be completely general, applicable to any system out of equilibrium. (Hint: Consider the entropy flow into the outside world upon cooling the liquid into the glass, compared to the entropy flow from the outside world to heat the glass into the liquid again. The initial and final states of the liquid are both in equilibrium.)

The residual entropy of a typical glass is about $k_B$ per molecular unit. It's a measure of how many different glassy configurations of atoms the material can freeze into.

*(c) In a molecular dynamics simulation with one hundred indistinguishable atoms, and assuming that the residual entropy is $k_B \log 2$ per atom, what is the probability that two coolings to zero energy will arrive at equivalent atomic configurations (up to permutations)? In a system with $10^{23}$ molecular units, with residual entropy $k_B \log 2$ per unit, about how many coolings would be needed to arrive at the original configuration again, with probability 1/2?*

(5.9) **Rubber Band.** (Basic)

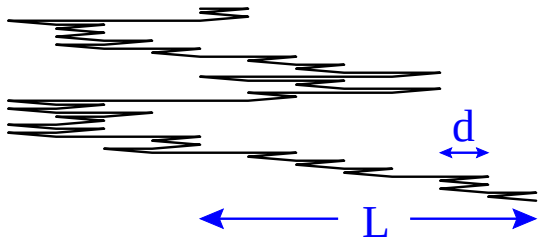



**Fig. 5.14** Simple model of a rubber band with $N = 100$ segments. The beginning of the polymer is at the top: the end is at the bottom; the vertical displacements are added for visualization.

Figure 5.14 shows a one–dimensional model for rubber. Rubber is formed of many long polymeric molecules, which undergo random walks in the undeformed material. When we stretch the rubber, the molecules respond by rearranging their random walk to elongate in the direction of the external stretch. In our model, the molecule is represented by a set of $N$ links of length $d$, which with equal energy point either parallel or antiparallel to the previous link. Let the total change in position to the right from the beginning of the polymer to the end be $L$.

As the molecule extent $L$ increases, the entropy of our rubber molecule decreases.

*(a) Find an exact formula for the entropy of this system in terms of $d$, $N$, and $L$.* (Hint: How many ways can one divide $N$ links into $M$ right–pointing links and $N - M$ left–pointing links, so that the total length is $L$?)

The external world, in equilibrium at temperature $T$, exerts a force pulling the end of the molecule to the right. The molecule must exert an equal and opposite *entropic* force $F$.

*(b) Find an expression for the force $-F$ exerted by the bath on the molecule in terms of the bath entropy.* Hint: the bath temperature $\frac{1}{T} = \frac{\partial S_{\text{bath}}}{\partial E}$, and force times distance is energy. *Using the fact that the length $L$ must maximize the entropy of the universe, write a general expression for $F$ in terms of the internal entropy $S$ of the molecule.*

*(c) Take our model of the molecule from part (a), the general law of part (b), and Stirling's formula 3.10 (dropping the square root), write the force law $F(L)$ for our molecule for large lengths $N$. What is the spring constant $K$ in Hooke's law $F = -KL$ for our molecule, for small $L$?*

Our model has no internal energy: this force is entirely entropic.

*(d) If we increase the temperature of our rubber band while it is under tension, will it expand or contract? Why?*

In a more realistic model of a rubber band, the entropy consists primarily of our configurational random–walk entropy plus a vibrational entropy of the molecules. If we stretch the rubber band without allowing heat to flow in or out of the rubber, the total entropy should stay approximately constant.[46]

*(e) True or false?*

[46] Rubber is designed to bounce well: little irreversible entropy is generated in a cycle of stretching and compression, so long as the deformation is not too abrupt.

*(T) (F) When we stretch the rubber band, it will cool: the configurational entropy of the random walk will decrease, causing the entropy in the vibrations to decrease, causing the temperature to decrease.*

*(T) (F) When we stretch the rubber band, it will cool: the configurational entropy of the random walk will decrease, causing the entropy in the vibrations to increase, causing the temperature to decrease.*

*(T) (F) When we let the rubber band relax, it will cool: the configurational entropy of the random walk will increase, causing the entropy in the vibrations to decrease, causing the temperature to decrease.*

*(T) (F) When we let the rubber band relax, there must be no temperature change, since the entropy is constant.*

This more realistic model is much like the ideal gas, which also had no configurational energy.

*(T) (F) Like the ideal gas, the temperature changes because of the net work done on the system.*

*(T) (F) Unlike the ideal gas, the work done on the rubber band is positive when the rubber band expands.*

You should check your conclusions experimentally: find a rubber band (thick and stretchy is best), touch it to your lips (which are very sensitive to temperature), and stretch and relax it.

(5.10) **Deriving Entropy.** (Mathematics)

In this exercise, you will show that the unique continuous function (up to the constant $k_B$) satisfying the three key properties (equations 5.24, 5.29, and 5.35):

$$S_I\left(\frac{1}{\Omega},\ldots,\frac{1}{\Omega}\right) > S_I(p_1,\ldots,p_\Omega) \quad \text{unless } p_i = \frac{1}{\Omega} \text{ for all } i, \tag{5.40}$$

$$S_I(p_1,\ldots,p_{\Omega-1},0) = S_I(p_1,\ldots,p_{\Omega-1}), \tag{5.41}$$

and

$$\langle S_I(A|B_\ell)\rangle_B = S_I(AB) - S_I(B). \tag{5.42}$$

where $S_I(A) = S_I(p_1,\ldots,p_\Omega)$, $S_I(B) = S_I(q_1,\ldots,q_M)$, $\langle S_I(A|B_\ell)\rangle_B = \sum_\ell q_\ell S_I(c_{1\ell},\ldots,c_{\Omega\ell})$ and $S_I(AB) = S_I(c_{11}q_1,\ldots,c_{\Omega M}q_M)$. The presentation is based on the proof in the excellent small book by Khinchin [51].

For convenience, define $L(g) = S_I(1/g,\ldots,1/g)$.

*(a) For any rational probabilities $q_\ell$, let $g$ be the least common multiple of their denominators, and let $q_\ell = g_\ell/g$ for integers $g_\ell$. Show that*

$$S_I(B) = L(g) - \sum_\ell q_\ell L(g_\ell). \tag{5.43}$$

(Hint: consider $AB$ to have $g$ possibilities of probability $1/g$, $B$ to measure which group of size $g_\ell$, and $A$ to measure which of the $g_\ell$ members of group $\ell$, figure 5.15.)

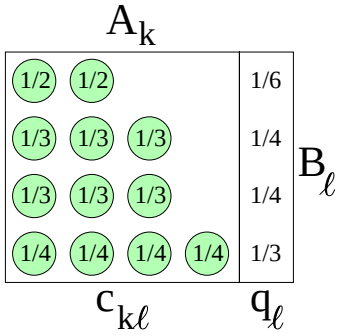



**Fig. 5.15 Rational Probabilities and Conditional Entropy.** Here the probabilities $q_\ell = (1/6, 1/3, 1/3, 1/2)$ of state $B_\ell$ are rational. We can split the total probability into $g = 12$ equal pieces (circles, each probability $r_{k\ell} = 1/12$), with $g_k = (2, 3, 3, 4)$ pieces for the corresponding measurement $B_\ell$. We can then write our ignorance $S_I(B)$ in terms of the (maximal) equal-likelihood ignorances $L(g) = S_I(1/g,\ldots)$ and $L(g_k)$, and use the entropy change for conditional probabilities property (equation 5.35) to derive our ignorance $S_I(B)$ (equation 5.43).

*(b) If $L(g) = k_S \log g$, show that equation 5.43 is the Shannon entropy 5.23.*

Knowing that $S_I(A)$ is the Shannon entropy for all rational probabilities, and assuming that $S_I(A)$ is continuous, makes $S_I(A)$ the Shannon entropy. So, we've reduced the problem to showing $L(g)$ is the logarithm up to a constant.

*(c) Show that $L(g)$ is monotone increasing with $g$.* (Hint: you'll need to use both of the first two properties.)

*(d) Show $L(g^n) = nL(g)$.* (Hint: consider $n$ independent probability distributions each of $g$ equally likely events. Use the third property recursively on $n$.)

*(e) If $2^m < s^n < 2^{m+1}$, using the results of parts (c) and (d) show*

$$\frac{m}{n} < \frac{L(s)}{L(2)} < \frac{m+1}{n}. \tag{5.44}$$

(Hint: how is $L(2^m)$ related to $L(s^n)$ and $L(2^{m+1})$?) *Show also using the same argument that $\frac{m}{n} < \frac{\log(s)}{\log(2)} < \frac{m+1}{n}$. Hence, show that $\left|\frac{L(s)}{L(2)} - \frac{\log(s)}{\log(2)}\right| < \frac{1}{n}$ and thus $L(s) = k \log s$ for some constant $k$.*

Hence our ignorance function $S_I$ agrees with the formula for the nonequilibrium entropy, uniquely up to an overall constant.

(5.11) **Chaos, Lyapunov, and Entropy Increase.** (Math, Complexity) (With Myers. [75])

Let's consider a dynamical system, given by a mappingfrom the unit interval $(0,1)$ into itself:[47]

$$f(x) = 4\mu x(1-x). \tag{5.45}$$

where the time evolution is given by iterating the map:

$$x_0, x_1, x_2, \cdots = x_0, f(x_0), f(f(x_0)), \ldots \tag{5.46}$$

In particular, for $\mu = 1$ it precisely folds the unit interval in half, and stretches it (non-uniformly) to cover the original domain.

The mathematics community lumps together continuous dynamical evolution laws and discrete mappings as both being dynamical systems. You can motivate the relationship using the Poincaré sections (figure 4.4), which connect a continuous recirculating dynamical system to the once–return map. The mapping 4.11 is not invertible, so it isn't directly given by a Poincaré section of a smooth differential equation[48] but the general stretching and folding exhibited by our map is often seen in driven physical systems without conservation laws.

In this exercise, we will focus on values of $\mu$ near one, where the motion is mostly *chaotic*. Chaos is sometimes defined as motion where the final position depends sensitively on the initial conditions. Two trajectories, starting a distance $\epsilon$ apart, will typically drift apart in time as $\epsilon e^{\lambda t}$, where $\lambda$ is the Lyapunov exponent for the chaotic dynamics.

*Start with $\mu = 0.9$ and two nearby points $x_0$ and $y_0 = x_0 + \epsilon$ somewhere between zero and one. Investigate the two trajectories $x_0, f(x_0), f(f(x_0)), \ldots f^{[n]}(x_0)$ and $y_0, f(y_0), \ldots$. How fast do they separate? Why do they stop separating? Estimate the Lyapunov exponent.* (Hint: $\epsilon$ can be a few times the precision of the machine (around $10^{-17}$ for double precision arithmetic), so long as you are not near the maximum value of $f$ at $x_0 = 0.5$.)

Many Hamiltonian systems are also chaotic. Two configurations of classical atoms or billiard balls, with initial positions and velocities that are almost identical, will rapidly diverge as the collisions magnify small initial deviations in angle and velocity into large ones. It is this chaos that stretches, folds, and kneads phase space (as in the Poincaré cat map of exercise 5.4) that is at root our explanation that entropy increases.[49]

(5.12) **Black Hole Thermodynamics.** (Astrophysics)

Astrophysicists have long studied *black holes*: the end state of massive stars which are too heavy to support themselves under gravity (see exercise 7.14). As the matter continues to fall into the center, eventually the escape velocity reaches the speed of light. After this point, the in-falling matter cannot ever communicate information back to the outside. A black hole of mass $M$ has radius[50]

$$R_s = G\frac{2M}{c^2}, \tag{5.47}$$

where $G = 6.67 \times 10^{-8}$ cm$^3$/g sec$^2$ is the gravitational constant, and $c = 3 \times 10^{10}$ cm/sec is the speed of light.

Hawking, by combining methods from quantum mechanics and general relativity, calculated the emission of radiation from a black hole.[51] He found a wonderful result: black holes emit perfect black–body radiation at a temperature

$$T_{bh} = \frac{\hbar c^3}{8\pi G M k_B}. \tag{5.48}$$

According to Einstein's theory, the energy of the black hole is $E = Mc^2$.

*(a) Calculate the specific heat of the black hole.*

The specific heat of a black hole is *negative*. That is, it gets cooler as you add energy to it. In a bulk material, this would lead to an instability: the cold regions would suck in more heat and get colder. Indeed, a population of black holes is unstable: the larger ones will eat the smaller ones.[52]

[47] We also study this map in exercises 4.3, 5.13, and 12.8.

[48] Remember the existence and uniqueness theorems from math class? The invertibility follows from uniqueness.

[49] There have been speculations by some physicists that entropy increases through information dropping into black holes – either real ones or tiny virtual black–hole fluctuations (see exercise 5.12. Recent work has cast doubt that the information is really lost even then: we're told it's just scrambled, presumably much as in chaotic systems.

[50] This is the Schwarzschild radius of the event horizon for a black hole with no angular momentum or charge.

[51] Nothing can leave a black hole: the radiation comes from vacuum fluctuations just outside the black hole that emit particles.

[52] A thermally insulated glass of ice water also has a negative specific heat! The surface tension at the curved the ice surface will decrease the coexistence temperature a slight amount (see section 11.2): the more heat one adds, the smaller the ice cube, the larger the curvature, and the lower the resulting temperature!

*(b) Calculate the entropy of the black hole, by using the definition of temperature* $\frac{1}{T} = \frac{\partial S}{\partial E}$ *and assuming the entropy is zero at mass* $M = 0$*. Express your result in terms of the surface area* $A = 4\pi R_s^2$*, measured in units of the Planck length* $L^* = \sqrt{\hbar G/c^3}$ *squared.*

As it happens, Bekenstein had deduced this formula for the entropy somewhat earlier, by thinking about analogies between thermodynamics, information theory, and statistical mechanics. On the one hand, when black holes interact or change charge and angular momentum, one can prove in classical general relativity that the area can only increase. So it made sense to assume that the entropy was somehow proportional to the area. He then recognized that if you had some waste material of high entropy to dispose of, you could ship it into a black hole and never worry about it again. Indeed, given that the entropy represents your lack of knowledge about a system, once matter goes into a black hole one can say that our knowledge about it completely vanishes.[53] (More specifically, the entropy of a black hole represents the inaccessibility of all information about what it was built out of.) By carefully dropping various physical systems into a black hole (theoretically) and measuring the area increase compared to the entropy increase,[54] he was able to deduce these formulas purely from statistical mechanics.

We can use these results to provide a fundamental bound on memory storage.

*(c) Calculate the maximum number of bits that can be stored in a sphere of radius one centimeter.*

(5.13) **Fractal Dimensions.** (Math, Complexity) (With Myers. [75])

There are many strange sets that emerge in science. In statistical mechanics, such sets often arise at continuous phase transitions, where self–similar spatial structures arise (chapter 12. In chaotic dynamical systems, the attractor (the set of points occupied at long times after the transients have disappeared) is often a fractal (called a *strange attractor*. These sets often are tenuous and jagged, with holes on all length scales: see figures 12.2, 12.3, and 12.14.

We often try to characterize these strange sets by a dimension. The dimensions of two extremely different sets can be the same: the path exhibited by a random walk (embedded in three or more dimensions) is arguably a two–dimensional set (note 6 on page 15), but does not locally look like a surface! However, if two sets have different spatial dimensions (measured in the same way) they surely are qualitatively different.

There is more than one way to define a dimension. Roughly speaking, strange sets are often spatially inhomogeneous, and what dimension you measure depends upon how you weight different regions of the set. In this exercise, we will calculate the *information dimension* (closely connected to the non-equilibrium entropy!), and the *capacity dimension* (originally called the *Hausdorff dimension*, also sometimes called the *fractal dimension*).

To generate our strange set – along with some more ordinary sets – we will use the logistic map[55]

$$f(x) = 4\mu x(1-x) \tag{5.49}$$

that we also study in exercises 5.11, 4.3, and 12.8. The attractor for the logistic map is a periodic orbit (dimension zero) at $\mu = 0.8$, and a chaotic, cusped density filling two intervals (dimension one)[56] at $\mu = 0.9$. At the onset of chaos at $\mu = \mu_\infty \approx 0.892486418$ (exercise 12.8) the dimension becomes intermediate between zero and one: the attractor is strange, self–similar set.

Both the information dimension and the capacity dimension are defined in terms of the occupation $P_n$ of cells of size $\epsilon$ in the limit as $\epsilon \to 0$.

*(a) Write a routine which, given* $\mu$ *and a set of bin sizes* $\epsilon$*,*

- *Iterates* $f$ *hundreds or thousands of times (to get on the attractor)*
- *Iterates* $f$ *many more times, collecting points on the attractor. (For* $\mu \leq \mu_\infty$*, you could just integrate* $2^n$ *times for* $n$ *fairly large.)*
- *For each* $\epsilon$*, use a histogram to calculate the probability* $P_n$ *that the points fall in the* $n^{\text{th}}$ *bin*
- *Return the set of vectors* $P_n[\epsilon]$*.*

[53] Except for the mass, angular momentum, and charge. This suggests that baryon number, for example, isn't conserved in quantum gravity. It has been commented that when the baryons all disappear, it'll be hard for Dyson to build his progeny out of electrons and neutrinos: see 5.1.

[54] In ways that are perhaps too complex to do here.

[55] We also study this map in exercises 4.3, 5.11, and 12.8.

[56] See exercise 4.3. The chaotic region for the logistic map isn't a strange attractor because it's confined to one dimension: period doubling cascades for dynamical systems in higher spatial dimensions likely will have fractal, strange attractors in the chaotic region.

You may wish to test your routine by using it for $\mu = 1$ (where the distribution should look like $\rho(x) = \frac{1}{\pi\sqrt{x(1-x)}}$, exercise 4.3(b)) and $\mu = 0.8$ (where the distribution should look like two $\delta$-functions, each with half of the points).

**The Capacity Dimension.** The definition of the capacity dimension is motivated by the idea that it takes at least

$$N_{\rm cover} = V/\epsilon^D \tag{5.50}$$

bins of size $\epsilon^D$ to cover a $D$-dimensional set of volume $V$.[57] By taking logs of both sides we find $\log N_{\rm cover} \approx \log V + D \log \epsilon$. The capacity dimension is defined as the limit

$$D_{\rm capacity} = \lim_{\epsilon \to 0} \frac{\log N_{\rm cover}}{\log \epsilon} \tag{5.51}$$

but the convergence is slow (the error goes roughly as $\log V / \log \epsilon$). Faster convergence is given by calculating the slope of $\log N$ versus $\log \epsilon$:

$$\begin{aligned} D_{\rm capacity} &= \lim_{\epsilon \to 0} \frac{d \log N_{\rm cover}}{d \log \epsilon} \\ &= \lim_{\epsilon \to 0} \frac{\log N_{i+1} - \log N_i}{\log \epsilon_{i+1} - \log \epsilon_i}. \end{aligned} \tag{5.52}$$

*(b) Use your routine from part (a), write a routine to calculate $N[\epsilon]$ by counting non-empty bins. Plot $D_{\rm capacity}$ from the fast convergence equation 5.52 versus the midpoint $\frac{1}{2}(\log \epsilon_{i+1} + \log \epsilon_i)$. Does it appear to extrapolate to $D = 1$ for $\mu = 0.9$?*[58] *Does it appear to extrapolate to $D = 0$ for $\mu = 0.8$? Plot these two curves together with the curve for $\mu_\infty$. Does the last one appear to converge to $D_1 \approx 0.538$, the capacity dimension for the Feigenbaum attractor gleaned from the literature? How small a deviation from $\mu_\infty$ does it take to see the numerical crossover to integer dimensions?*

**Entropy and the Information Dimension.** The entropy of a statistical mechanical system is given by equation 5.20, $S = -k_B {\rm Tr}(\rho \log \rho)$. In the chaotic regime this works fine. Our probabilities $P_n \approx \rho(x_n)\epsilon$, so converting the entropy integral into a sum $\int f(x)\,dx \approx \sum_n f(x_n)\epsilon$ gives

$$\begin{aligned} S &= -k_B \int \rho(x) \log(\rho(x))\,dx \\ &\approx -\sum_n P_n \log(P_n/\epsilon) = -\sum_n P_n \log P_n + \log \epsilon \end{aligned} \tag{5.53}$$

(setting the conversion factor $k_B = 1$ for convenience).

You might imagine that the entropy for a fixed point would be zero, and the entropy for a period-$n$ cycle would be $k_B \log n$. But this is incorrect: when there is a fixed point or a periodic limit cycle, the attractor is on a set of dimension zero (a bunch of points) rather than dimension one. The entropy must go to minus infinity – since we have precise information about where the trajectory sits at long times. To estimate the "zero–dimensional" entropy $k_B \log n$ on the computer, we would take the same bins as above but sum over bins $P_n$ instead of integrating over $x$:

$$S_{d=0} = -\sum_n P_n \log(P_n) = S_{d=1} - \log(\epsilon). \tag{5.54}$$

More generally, the 'natural' measure of the entropy for a set with $D$ dimensions might be defined as

$$S_D = -\sum_n P_n \log(P_n) + D \log(\epsilon). \tag{5.55}$$

Instead of using this formula to define the entropy, mathematicians use it to define the information dimension

$$D_{\rm inf} = \lim_{\epsilon \to 0} \Big( \sum P_n \log P_n \Big) / \log(\epsilon). \tag{5.56}$$

The information dimension agrees with the ordinary dimension for sets that locally look like $\mathbb{R}^D$. It's different from the capacity dimension because the information dimension weights each part (bin) of the attractor by the time spent in it. Again, we can speed up the convergence by noting that equation 5.55 says that $\sum_n P_n \log P_n$ is a linear function of $\log \epsilon$ with slope $D$ and intercept $S_D$. Measuring the slope directly, we find

$$D_{\rm inf} = \lim_{\epsilon \to 0} \frac{d \sum_n P_n(\epsilon) \log P_n(\epsilon)}{d \log \epsilon}. \tag{5.57}$$

*(c) As in part (b), write a routine that plots $D_{\rm inf}$ from equation 5.57 as a function of the midpoint $\log \epsilon$, as we increase the number of bins. Plot the curves for $\mu = 0.9$, $\mu = 0.8$, and $\mu_\infty$. Does the information dimension agree with the ordinary one for the first two? Does the last one appear to converge to $D_1 \approx 0.517098$, the information dimension for the Feigenbaum attractor from the literature?*

Most 'real world' fractals have a whole spectrum of different characteristic spatial dimensions: they are *multifractal*.)

[57] Imagine covering the surface of a sphere in 3D with tiny cubes: the number of cubes will go as the surface area [2D-volume] divided by $\epsilon^2$.

[58] In the chaotic regions, keep the number of bins small compared to the number of iterates in your sample, or you start finding empty bins between points and eventually get a dimension of zero.

# Free Energies

6

In this chapter, we explain how to study *parts* of statistical mechanical systems. If we ignore most of our system – agreeing not to ask questions about certain degrees of freedom – the statistical mechanical predictions about the remaining parts of our system are embodied in a new statistical ensemble and its associated *free energy*. These free energies usually make calculations easier and the physical behavior more comprehensible. What do we ignore?

*We ignore the external world.* Most systems are not isolated: they often can exchange energy, volume, or particles with an outside world in equilibrium (often called the *heat bath*). If the coupling to the external world is weak, we can remove it from consideration. The constant–temperature *canonical*[1] ensemble (section 6.1) and the *Helmholtz* free energy arise from a bath which can exchange energy; the *grand canonical* ensemble (section 6.3) and the grand free energy arise from baths which also exchange particles at fixed chemical potential. For large systems, these different ensembles predict the same average behavior (apart from tiny fluctuations): we could in principle do most calculations of interest in the isolated, microcanonical ensemble. However, calculations using the appropriate free energy can be much simpler (section 6.2).

[1] Webster's **canonical**: reduced to the simplest or clearest schema possible. The canonical ensemble will be simpler to compute with than the microcanonical one.

*We ignore unimportant internal degrees of freedom.* In studying (say) chemical reactions, magnets, or the motion of large mechanical objects, one is normally not interested in the motions of the individual atoms. To ignore them in mechanical systems, one introduces *friction* and *noise* (section 6.5). By ignoring them in magnets, one derives lattice models whose free energies depend only on the spin on each magnetic atom (section 8.2). By ignoring the atomic motions in chemical reactions, one derives reaction rate theory (section 6.6).

*We coarse grain.* Many systems are not homogeneous, because of initial conditions or boundary conditions; their properties vary in space and/or time. If these systems are locally near equilibrium, we can ignore the internal degrees of freedom in small volumes, keeping only the fields which describe the local state. That is, we remove the short distance fluctuations of the important degrees of freedom, *coarse-graining* to include only the long-wavelength fluctuations.[2] As an example, in section 6.7 we will calculate the free energy density for the ideal gas, and use it to (again) derive the diffusion equation.

[2] These are the *order parameter fields* that we will study in detail in chapter 9.

We will calculate free energies explicitly in several important cases in this chapter. Note that free energies are important tools, however, even for systems too complex to solve analytically. We provide these solvable

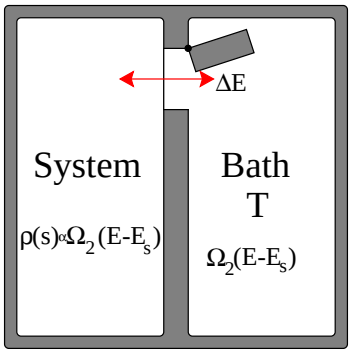



**Fig. 6.1 The canonical ensemble** describes equilibrium systems which can exchange energy with a heat bath. The bath is at temperature $T$. The probability of a state $s$ of the system with energy $E_s$ is $\rho(s) = \exp(-E_s/k_BT)/Z$. The thermodynamics of the canonical ensemble is embodied in the Helmholtz free energy $A(T,V,N) = E - TS$.

examples in part to motivate our later work on similar free energies for systems where microscopic calculations are not feasible.

## 6.1 The Canonical Ensemble

The canonical ensemble governs the equilibrium behavior of a system at fixed temperature. We defined the temperature in section 3.3 by considering a total system comprised of two weakly coupled parts, with phase-space coordinates $(\mathbb{P}_1, \mathbb{Q}_1)$ and $(\mathbb{P}_2, \mathbb{Q}_2)$ that can exchange energy. We will now focus on the first of these two parts (the *system*); the second part (the *heat bath*) we will assume is large. We are not interested in measuring any properties that depend upon the heat bath, and want a statistical ensemble for the system that averages over the relevant states of the bath.

How does the probability that our system is in a state $s$ depend upon its energy $E_s$? As we discussed in deriving equation 3.22 (see note 26), the probability density that our system will be in the particular state $s$ is proportional to the volume of the energy shell for our heat bath at bath energy $E - E_s$

$$\rho(s) \propto \Omega_2(E - E_s) = \exp\left(S_2(E - E_s)/k_B\right) \tag{6.1}$$

since a state $s$ gets a share of the microcanonical probability for each heat-bath partner it can coexist with at the fixed total energy $E$.

Let us compare the probability of two typical states $A$ and $B$ of our equilibrium system. We know that the energy fluctuations are small, and we assume that the heat bath is large. We can therefore assume that the inverse temperature $1/T_2 = \frac{\partial S_2}{\partial E_2}$ of the heat bath is constant in the range $(E - E_A, E - E_B)$. Hence,

$$\begin{aligned}\rho(s_B)/\rho(s_A) &= \Omega_2(E - E_B)/\Omega_2(E - E_A) \\ &= e^{(S_2(E-E_B)-S_2(E-E_A))/k_B} = e^{(E_A-E_B)\,(\partial S_2/\partial E)/k_B} \\ &= e^{(E_A-E_B)/k_BT_2}.\end{aligned} \tag{6.2}$$

This is the general derivation of the Boltzmann factor: the probability of a particular system state of energy $E_s$ is

$$\rho(s) \propto \exp(-E_s/k_BT). \tag{6.3}$$

We know that the probability is normalized, so

$$\begin{aligned}\rho(s) &= \exp(-E_s/k_BT) \Big/ \int d\mathbb{P}_1 d\mathbb{Q}_1 \exp(-\mathcal{H}_1(\mathbb{P}_1, \mathbb{Q}_1)/k_BT) \\ &= \exp(-E_s/k_BT) \Big/ \sum_n \exp(-E_n/k_BT) \\ &= \exp(-E_s/k_BT)/Z\end{aligned} \tag{6.4}$$

where the normalization factor

$$Z(T,N,V) = \sum_n \exp(-E_n/k_BT) = \int d\mathbb{P}_1 d\mathbb{Q}_1 \exp(-\mathcal{H}_1(\mathbb{P}_1, \mathbb{Q}_1)/k_BT) \tag{6.5}$$

is the *partition function*.[3]

Equation 6.4 is the definition of the canonical ensemble,[4] appropriate for calculating properties of systems which can exchange energy with an external world with temperature $T$.

The partition function $Z$ is just the normalization factor that keeps the total probability summing to one. It may surprise you to discover that this normalization factor plays a central role in the theory. Indeed, most quantities of interest can be calculated in two different ways: as an explicit sum over states (statistical mechanics) or in terms of derivatives of the partition function (thermodynamics, see section 6.4). Let's see how this works by using $Z$ to calculate the mean energy, the specific heat, and the entropy of a general system.

**Internal energy.** To calculate the average internal energy of our system[5] $\langle E \rangle$, we weight each state by its probability:

[3] To avoid blinding ourselves with integrals, we will write them as a 'continuous sum': $\int d\mathbb{P}_1 d\mathbb{Q}_1 \rightarrow \sum_n$ for the rest of this chapter. This notation foreshadows quantum mechanics (chapter 7), where for bound systems the energy levels are discrete; it also will be appropriate for lattice systems like the Ising model (section 8.2), where we've integrated away all the continuous degrees of freedom. No complications arise from translating the sums for the equations in this chapter back into integrals over phase space.

[5] The angle brackets represent canonical averages.

[4] A formal method of deriving the canonical ensemble is as a *partial trace*, removing the bath degrees of freedom from a microcanonical ensemble. To calculate the expectation of an operator $B$ that depends only on system coordinates $(\mathbb{P}_1, \mathbb{Q}_1)$, we start by averaging over the energy shell in the entire space, including both the system coordinates and the bath coordinates $(\mathbb{P}_2, \mathbb{Q}_2)$:

$$\begin{aligned}\Omega(E) &= \frac{1}{\delta E} \int_{E<\mathcal{H}_1+\mathcal{H}_2<E+\delta E} d\mathbb{P}_1 d\mathbb{Q}_1 d\mathbb{P}_2 d\mathbb{Q}_2 \\ &= \int dE_1 \Omega_1(E_1)\Omega_2(E-E_1). \qquad (6.6)\end{aligned}$$

$$\begin{aligned}\langle B \rangle &= \frac{1}{\Omega(E)\delta E} \int_{E<\mathcal{H}_1+\mathcal{H}_2<E+\delta E} d\mathbb{P}_1 d\mathbb{Q}_1 B(\mathbb{P}_1,\mathbb{Q}_1)\, d\mathbb{P}_2 d\mathbb{Q}_2 \\ &= \frac{1}{\Omega(E)} \int d\mathbb{P}_1 d\mathbb{Q}_1 B(\mathbb{P}_1,\mathbb{Q}_1)\, \Omega_2(E-\mathcal{H}_1(\mathbb{P}_1,\mathbb{Q}_1)). \qquad (6.7)\end{aligned}$$

Again, if the heat-bath is large the small variations $E_1 - \langle E_1 \rangle$ won't change its temperature; $1/T_2 = \frac{\partial S_2}{\partial E_2}$ being fixed implies $\frac{\partial \Omega_2(E-E_1)}{\partial E_1} = -\frac{1}{k_B T}\Omega_2$, and hence

$$\Omega_2(E-E_1) = \Omega_2(E-\langle E_1 \rangle) \exp(-(E_1 - \langle E_1 \rangle)/k_B T). \qquad (6.8)$$

This gives us

$$\begin{aligned}\Omega(E) &= \int dE_1 \Omega_1(E_1)\Omega_2(E-\langle E_1 \rangle) \exp(-(E_1-\langle E_1 \rangle)/k_B T) \\ &= \Omega_2(E-\langle E_1 \rangle) e^{\langle E_1 \rangle / k_B T} \int dE_1 \Omega_1(E_1) e^{-E_1/k_B T} \\ &= \Omega_2(E) \exp(-\langle E_1 \rangle / k_B T)\, Z \qquad (6.9)\end{aligned}$$

and

$$\begin{aligned}\langle B \rangle &= \frac{\int d\mathbb{P}_1\, d\mathbb{Q}_1 B(\mathbb{P}_1,\mathbb{Q}_1)\Omega_2(E-\langle E_1 \rangle) e^{-(\mathcal{H}_1(\mathbb{P}_1,\mathbb{Q}_1)-\langle E_1 \rangle)/k_B T}}{\Omega_2(E-\langle E_1 \rangle) e^{\langle E_1 \rangle / k_B T} Z} \\ &= \frac{1}{Z} \int d\mathbb{P}_1 d\mathbb{Q}_1 B(\mathbb{P}_1,\mathbb{Q}_1)\, \exp(-\mathcal{H}_1(\mathbb{P}_1,\mathbb{Q}_1)/k_B T). \qquad (6.10)\end{aligned}$$

By explicitly doing the integrals over $\mathbb{P}_2$ and $\mathbb{Q}_2$, we have turned a microcanonical calculation into the canonical ensemble (equation 6.4). Our calculation of the momentum distribution $\rho(p_1)$ in section 3.2.2 was precisely of this form: we integrated out all the other degrees of freedom, and were left with a Boltzmann distribution for the $x$-momentum of particle number one. This process is called *integrating out* the degrees of freedom for the heat bath, and is the general way of creating free energies.

$$\langle E\rangle = \sum_n E_n P_n = \frac{\sum_n E_n e^{-\beta E_n}}{Z} = -\frac{\partial Z/\partial\beta}{Z}$$
$$= -\partial \log Z/\partial\beta \tag{6.11}$$

This can be seen as a statistical mechanics derivation (involving sums over the ensemble) of a thermodynamics relation $E = -\frac{\partial \log Z}{\partial\beta}$.

**Specific Heat.** Let $c_v$ be the specific heat per particle at constant volume. (The specific heat is the energy needed to increase the temperature by one unit, $\frac{\partial\langle E\rangle}{\partial T}$.) Using equation 6.11, we get the thermodynamics formula for $c_v$ from the second derivative of $Z$:

$$Nc_v = \frac{\partial\langle E\rangle}{\partial T} = \frac{\partial\langle E\rangle}{\partial\beta}\frac{d\beta}{dT} = -\frac{1}{k_B T^2}\frac{\partial\langle E\rangle}{\partial\beta} = \frac{1}{k_B T^2}\frac{\partial^2 \log Z}{\partial\beta^2}. \tag{6.12}$$

We can expand the penultimate form of this formula into a statistical mechanics calculation, finding the intruiguing result

$$Nc_v = -\frac{1}{k_B T^2}\frac{\partial\langle E\rangle}{\partial\beta} = -\frac{1}{k_B T^2}\frac{\partial}{\partial\beta}\frac{\sum E_n e^{-\beta E_n}}{\sum e^{-\beta E_n}}$$
$$= -\frac{1}{k_B T^2}\left[\frac{\left(\sum E_n e^{-\beta E_n}\right)^2}{Z^2} + \frac{\sum -{E_n}^2 e^{-\beta E_n}}{Z}\right]$$
$$= \frac{1}{k_B T^2}\left[\langle E^2\rangle - \langle E\rangle^2\right] = {\sigma_E}^2/k_B T^2, \tag{6.13}$$

where $\sigma_E$ is the root-mean-square fluctuation[6] in the energy of our system at constant temperature. Equation 6.13 is a remarkable result: it is a relationship between a macroscopic susceptibility ($c_v$, the energy changes when the temperature is perturbed) to a microscopic fluctuation ($\sigma_E$, the energy fluctuation in thermal equilibrium). In general, fluctuations can be related to responses in this fashion. These relations are extremely useful, for example, in extracting susceptibilities from numerical simulations. No need to make small changes and try to measure the response: just watch it fluctuate in equilibrium (exercises 3.7 and 8.1).

Are results calculated using the canonical ensemble the same as those computed from our original microcanonical ensemble? Equation 6.13 tells us that the energy fluctuations are tiny (as we earlier argued in the microcanonical ensemble (section 3.3). The energy fluctuations per particle

$$\sigma_E/N = \sqrt{\langle E^2\rangle - \langle E\rangle^2}/N = \sqrt{(k_B T)(c_v T)}/\sqrt{N}. \tag{6.14}$$

is $1/\sqrt{N}$ times the geometric mean of two microscopic energies: $k_B T$ (two-thirds the kinetic energy per particle) and $c_v T$ (the energy per particle to heat from absolute zero, if the specific heat were temperature independent). Hence, for macroscopic systems the behavior in most regards is the same whether the system is completely isolated (microcanonical) or in thermal contact with the rest of the world (canonical).

[6] We've used the standard trick $\langle (E - \langle E\rangle)^2\rangle = \langle E^2\rangle - 2\langle E\langle E\rangle\rangle + \langle E\rangle^2 = \langle E^2\rangle - \langle E\rangle^2$, since $\langle E\rangle$ is just a constant that can be pulled out of the ensemble average.

**Entropy.** Using the general statistical mechanical formula[7] for the entropy 5.20, we find

$$\begin{aligned} S &= -k_B \sum P_n \log P_n = -k_B \sum \frac{\exp(-\beta E_n)}{Z} \log\left(\frac{\exp(-\beta E_n)}{Z}\right) \\ &= -k_B \frac{\sum \exp(-\beta E_n)(-\beta E_n - \log Z)}{Z} \\ &= k_B \beta \langle E \rangle + k_B \log Z \frac{\sum \exp(-\beta E_n)}{Z} = \langle E \rangle / T + k_B \log Z, \end{aligned} \tag{6.16}$$

another statistical mechanics derivation of a thermodynamic relation.

Notice that the formulas for $\langle E \rangle$, $c_v$, and $S$ all involve derivatives of $\log Z$. This motivates us to define a *free energy* for the canonical ensemble, called the *Helmholtz* free energy:

$$A(T, V, N) = -k_B T \log Z = \langle E \rangle - TS. \tag{6.17}$$

The entropy is minus the derivative of $A$ with respect to $T$. Explicitly,

$$\begin{aligned} \left.\frac{\partial A}{\partial T}\right|_{N,V} &= -\frac{\partial k_B T \log Z}{\partial T} = -k_B \log Z - k_B T \frac{\partial \log Z}{\partial \beta} \frac{\partial \beta}{\partial T} \\ &= -k_B \log Z - k_B T \langle E \rangle / (k_B T^2) = -k_B \log Z - \langle E \rangle / T \\ &= -S. \end{aligned} \tag{6.18}$$

Why is it called a free energy? First, $k_B T$ gives it the dimensions of an energy. Second, it is the energy available (free) to do work. A heat engine drawing energy $E = Q_1$ from a hot bath that must discharge an entropy $S = Q_2/T_2$ into a cold bath can do work $W = Q_1 - Q_2 = E - T_2 S$; $A = E - TS$ is the energy free to do useful work (section 5.1).

Why is this a natural definition? The total phase-space volume, or weight, associated with the system and bath consistent with the current values of $V$ and $N$ is $Z = \exp(-A(T, V, N)/k_B T)$. This is in precise analogy with the Boltzmann weight $\exp(-\mathcal{H}_1(\mathbb{P}_1, \mathbb{Q}_1)/k_B T)$, which is the bath phase-space volume consistent with the current position and momentum of the system. In general, free energies $F(\mathbb{X})$ will remove all degrees of freedom except for certain constraints $\mathbb{X}$. The phase-space volume consistent with the constraints $\mathbb{X}$ is $\exp(-F(\mathbb{X}))/k_B T)$.

[7] Alternatively, we could use the microcanonical definition of the entropy of the entire system and equation 6.8 to show

$$\begin{aligned} S &= k_B \log \int dE_1 \Omega_1(E_1) \Omega_2(E - E_1) \\ &= k_B \log \int dE_1 \Omega_1(E_1) \Omega_2(E - \langle E_1 \rangle) e^{-(E_1 - \langle E_1 \rangle)/k_B T} \\ &= k_B \log \Omega_2(E - \langle E_1 \rangle) + k_B \log\left(\exp(\langle E_1 \rangle / k_B T)\right) + k_B \log \int dE_1 \Omega_1(E_1) \exp(-E_1/k_B T) \\ &= k_B \log \Omega_2(E_2) + k_B \beta E_1 + k_B \log Z_1 \\ &= S_2 + E_1/T - A_1/T \end{aligned}$$

so

$$S_1 = E_1/T + k_B \log Z_1 = E_1/T - A_1/T, \tag{6.15}$$

avoiding the use of the non-equilibrium entropy to derive the same result.

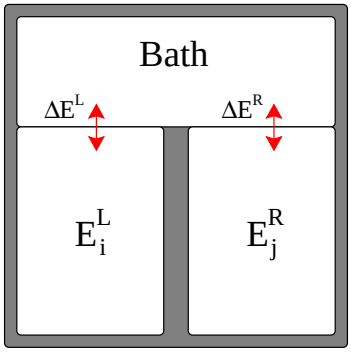



**Fig. 6.2 Uncoupled systems attached to a common heat bath.** Calculating the properties of two weakly coupled subsystems is easier in the canonical ensemble than in the microcanonical ensemble. This is because energy in one subsystem can be exchanged with the bath, and does not affect the energy of the other subsystem.

## 6.2 Uncoupled Systems and Canonical Ensembles

The canonical ensemble is much more convenient for doing calculations, because for systems in which the Hamiltonian splits into uncoupled components, the partition function factors into pieces that can be computed separately.

Suppose we have a system with two weakly interacting subsystems $L$ and $R$, both connected to a heat bath at $\beta = 1/k_B T$. The states for the whole system are pairs of states $(s_i^L, s_j^R)$ from the two subsystems, with energies $E_i^L$ and $E_j^R$ respectively. The partition function for the whole system is

$$\begin{aligned} Z &= \sum_{ij} \exp\left(-\beta(E_i^L + E_j^R)\right) = \sum_{ij} e^{-\beta E_i^L} e^{-\beta E_j^R} \\ &= \left(\sum_i e^{-\beta E_i^L}\right)\left(\sum_j e^{-\beta E_j^R}\right) \\ &= Z^L Z^R. \end{aligned} \tag{6.19}$$

Thus in the canonical ensemble of uncoupled systems the partition function factors. The Helmholtz free energy adds

$$A = -k_B T \log Z = -k_B T \log(Z^L \cdot Z^R) = A^L + A^R \tag{6.20}$$

as does the entropy, average energy, and other extensive properties that one expects to scale with the size of the system.

This is much simpler than the same calculation would be in the microcanonical ensemble! In a microcanonical ensemble, each subsystem would compete with the other for the available total energy. Even though two subsystems are uncoupled (the energy of one is independent of the state of the other) the microcanonical ensemble intermingles them in the calculation. By allowing each to draw energy from a large heat bath, the canonical ensemble allows uncoupled subsystems to become independent calculations.

We can now immediately do several important cases of uncoupled systems.

**Ideal Gas.** Since the different atoms in an ideal gas are uncoupled with one another, the partition function for the monatomic ideal gas of classical distinguishable particles of mass $m$ in a cubical box of volume $V = L^3$ factors into a product over each degree of freedom $\alpha$:

$$\begin{aligned} Z_{\text{ideal}}^{\text{dist}} &= \prod_{\alpha=1}^{3N} (1/h) \int_0^L dq_\alpha \int_{-\infty}^{\infty} dp_\alpha e^{-\beta p_\alpha^2/2m_\alpha} = \left(\frac{L}{h}\sqrt{\frac{2\pi m}{\beta}}\right)^{3N} \\ &= (L\sqrt{2\pi m k_B T/h^2})^{3N} = (L/\lambda)^{3N}. \end{aligned}$$

Here

$$\lambda = h/\sqrt{2\pi m k_B T} = \sqrt{2\pi\hbar^2/m k_B T} \tag{6.21}$$

is again the thermal de Broglie wavelength (equation 3.63).

The internal energy in the canonical ensemble is

$$\langle E \rangle = -\frac{\partial \log Z_{\text{ideal}}}{\partial \beta} = -\frac{\partial}{\partial \beta} \log(\beta^{-\frac{3N}{2}}) = 3N/2\beta = 3Nk_BT/2 \quad (6.22)$$

giving us the equipartition theorem for momentum without our needing to find volumes of spheres in $3N$ dimensions (section 3.2.2).

For $N$ indistinguishable particles, we have counted each real configuration $N!$ times for the different permutations of particles, so we must divide $Z_{\text{ideal}}^{\text{dist}}$ by $N!$ just as we did for the phase space volume $\Omega$ in section 3.5.

$$Z_{\text{ideal}}^{\text{indist}} = (L/\lambda)^{3N}/N! \quad (6.23)$$

This doesn't change the internal energy, but does change the Helmholtz free energy

$$\begin{aligned} A_{\text{ideal}}^{\text{indist}} &= -k_BT \log\left((L/\lambda)^{3N}/N!\right) \\ &= -Nk_BT \log(V/\lambda^3) - k_BT \log(N!) \\ &\sim -Nk_BT \log(V/\lambda^3) - k_BT(N \log N - N) \\ &= -Nk_BT \left(\log(V/N\lambda^3) + 1\right) \\ &= Nk_BT \left(\log(\rho\lambda^3) - 1\right). \end{aligned} \quad (6.24)$$

where $\rho = N/V$ is the average density, and we've used Stirling's formula $\log(N!) \sim N \log N - N$.

Finally, the entropy of $N$ indistinguishable particles, in the canonical ensemble, is

$$\begin{aligned} S &= -\frac{\partial A}{\partial T} = -Nk_B \left(\log(\rho\lambda^3) - 1\right) - Nk_BT \frac{\partial \log T^{-3/2}}{\partial T} \\ &= Nk_B \left(5/2 - \log(\rho\lambda^3)\right) \end{aligned} \quad (6.25)$$

as we derived (with much more effort) using the microcanonical ensemble (equation 3.62).

**Classical Harmonic Oscillator and the Equipartition Theorem.** Electromagnetic radiation, the vibrations of atoms in solids, and the excitations of many other systems near their equilibria can be approximately described as a set of uncoupled harmonic oscillator modes.[8] In

[8] For example, at temperatures low compared to the melting point a solid or molecule with an arbitrary many-body interaction potential $\mathcal{V}(\mathbb{Q})$ typically only makes small excursions about the minimum $\mathbb{Q}_0$ of the potential. We expand about this minimum, giving us

$$\mathcal{V}(\mathbb{Q}) \approx \mathcal{V}(\mathbb{Q}_0) + \sum_\alpha (\mathbb{Q} - \mathbb{Q}_0)_\alpha \partial_\alpha \mathcal{V} + \sum_{\alpha,\beta} \tfrac{1}{2}(\mathbb{Q} - \mathbb{Q}_0)_\alpha (\mathbb{Q} - \mathbb{Q}_0)_\beta \partial_\alpha \partial_\beta \mathcal{V} + \dots \quad (6.26)$$

Since the potential is a minimum at $\mathbb{Q}_0$, the gradient of the potential must be zero, so second term on the right-hand side must vanish. The third term is a big $3N \times 3N$ quadratic form, which we may diagonalize by converting to *normal modes* $q_k$. (If the masses of the atoms are not all the same, one must rescale the components of $\mathbb{Q} - \mathbb{Q}_0$ by the square root of the corresponding mass before diagonalizing.) In terms

the canonical ensemble, the statistical mechanics of these systems thus decomposes into a calculation for each mode separately.

A harmonic oscillator of mass $m$ and frequency $\omega$ has a total energy

$$\mathcal{H}(p,q) = p^2/2m + m\omega^2 q^2/2. \tag{6.29}$$

The partition function for one such oscillator is

$$\begin{aligned} Z = &\int_{-\infty}^{\infty} dq \int_{-\infty}^{\infty} dp (1/h) e^{-\beta(p^2/2m + m\omega^2 q^2/2)} = \frac{1}{h}\sqrt{\frac{2\pi}{\beta m\omega^2}}\sqrt{\frac{2\pi m}{\beta}} \\ =&\frac{1}{\beta\hbar\omega}. \end{aligned} \tag{6.30}$$

Hence the Helmholtz free energy for the classical oscillator is

$$A_\omega(T) = -k_B T \log Z = k_B T \log(\hbar\omega/k_B T), \tag{6.31}$$

the internal energy is

$$\langle E\rangle_\omega(T) = -\frac{\partial \log Z}{\partial \beta} = \frac{\partial}{\partial\beta}(\log\beta - \log\hbar\omega) = 1/\beta = k_B T, \tag{6.32}$$

and of course $c_v = \partial\langle E\rangle/\partial T = k_B$. This extends the equipartition theorem (section 3.2.2) to configuration space: for systems with quadratic potential energies the internal energy is $\frac{1}{2}k_B T$ per degree of freedom, where each harmonic oscillator has two degrees of freedom ($p$ and $q$).

**Classical Kinetic Energies.** One will notice both for the ideal gas and for the harmonic oscillator that each component of the momentum contributed a factor $\sqrt{\frac{2\pi m}{\beta}}$ to the partition function. As we promised in section 3.2.2, this will happen in any classical system where the momenta are uncoupled to the positions: that is, where the kinetic energy does not depend on the positions of the particles, and the potential energy is independent of the momenta. (Not all Hamiltonians have this form. For example, charged particles in magnetic fields will have terms that couple momenta and positions.) Thus the partition function for any classical interacting system of non-magnetic particles will be some configurational piece times $\prod_\alpha \sqrt{\frac{2\pi m_\alpha}{\beta}}$. This implies that the velocity distribution is always Maxwellian (equation 1.2) independent of what configuration the positions have.[9]

[9] This may be counterintuitive: an atom crossing a barrier has the same velocity distribution as it had in the bottom of the well (exercise 6.2). Such an atom does need to borrow some energy from the rest of the system, but in equilibrium it borrows equally from all modes, and so the crossing atom's velocity does not change in the limit of large system size.

of these normal modes, the Hamiltonian is a set of uncoupled harmonic oscillators

$$\mathcal{H} = \sum_k p_k^2/2m + m\omega_k^2 q_k^2/2. \tag{6.27}$$

At high enough temperatures that quantum mechanics can be ignored, we can then use equation 6.30 to find the total partition function for our harmonic system

$$Z = \prod_k Z_k = \prod_k (1/\beta\hbar\omega_k). \tag{6.28}$$

(In section 7.2 we'll do the quantum harmonic oscillator, which often gives an accurate theory for atomic vibrations at all temperatures below the melting point.)

## 6.3 Grand Canonical Ensemble

The grand canonical ensemble allows one to decouple the calculations of systems which can exchange both energy and particles with their environment.

Consider our system in a state $s$ with energy $E_s$ and number $N_s$, together with a bath with energy $E_2 = E - E_s$ and number $N_2 = N - N_s$ (figure 3.4). By analogy with equation 6.3 the probability density that the system will be in state $s$ is proportional to

$$\begin{aligned}\rho(s) &\propto \Omega_2(E - E_s, N - N_s) \\ &= \exp\left(\left(S_2(E - E_s, N - N_s)\right)/k_B\right) \\ &\propto \exp\left(\left(-E_s\frac{\partial S_2}{\partial E} - N_s\frac{\partial S_2}{\partial N}\right)/k_B\right) \\ &= \exp\left(-E_s/k_BT + N_s\mu/k_BT\right) \\ &= \exp\left(-(E_s - \mu N_s)/k_BT\right),\end{aligned} \tag{6.33}$$

where

$$\mu = -T\partial S/\partial N \tag{6.34}$$

is the *chemical potential*. Notice the factor of $-T$: this converts the entropy change into an energy change, so the chemical potential is the energy gain per particle for accepting particles from the bath. At low temperatures the system will fill with particles until the energy for the next particle reaches $\mu$.

Again, just as for the canonical ensemble, there is a normalization factor called the grand partition function

$$\Xi(T, V, \mu) = \sum_n e^{-(E_n - \mu N_n)/k_BT}; \tag{6.35}$$

the probability density of state $s_i$ is $\rho(s_i) = e^{-(E_i - \mu N_i)/k_BT}/\Xi$. There is a grand free energy

$$\Phi(T, V, \mu) = -k_BT\log(\Xi) = \langle E\rangle - TS - \mu N \tag{6.36}$$

analogous to the Helmholtz free energy $A(T, V, N)$. In exercise 6.4 you shall derive the Euler relation $E = TS - PV + \mu N$, and hence show that $\Phi(T, \mu, V) = -PV$.

**Partial Traces.**[10] Let us note in passing that we can write the grand canonical partition function as a sum over canonical partition functions. Let us separate the sum over states $n$ of our system into a double sum – an inner restricted sum[11] over states of fixed number of particles $M$ in

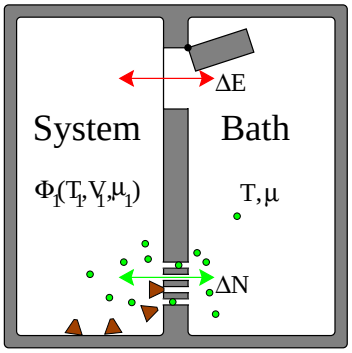



**Fig. 6.3 The grand canonical ensemble** describes equilibrium systems which can exchange energy and particles with a heat bath. The probability of a state $s$ of the system with $N_s$ particles and energy $E_s$ is $\rho(s) = \exp\left(-\frac{E_s + \mu N_s}{k_BT}\right)/Z$. The thermodynamics of the canonical ensemble is embodied in the grand free energy $\Phi(T, V, \mu)$.

[10] The classical mechanics integrals over phase space become traces over states in Hilbert space in quantum mechanics. Removing some of the degrees of freedom in quantum mechanics is done by a partial trace over the states. The name "partial trace" for removing some of the degrees of freedom has become standard also in classical statistical physics.

[11] This restricted sum is said to *integrate over* the internal degrees of freedom $\ell_M$.

the system and an outer sum over $M$. Let $s_{\ell_M,M}$ have energy $E_{\ell_M,M}$, so

$$\begin{aligned}\Xi(T,V,\mu) &= \sum_M \sum_{\ell_M} e^{-(E_{\ell_M,M}-\mu M)/k_BT} \\ &= \sum_M \left( \sum_{\ell_M} e^{-E_{\ell_M,M}/k_BT} \right) e^{\mu M/k_BT} \\ &= \sum_M Z(T,V,M) e^{\mu M/k_BT} \\ &= \sum_M e^{-(A(T,V,M)-\mu M)/k_BT}. \end{aligned} \tag{6.37}$$

Again, notice how the Helmholtz free energy in the last equation plays exactly the same role as the energy plays in equation 6.35: $\exp(-E_n/k_BT)$ is the probability of the system being in a particular system state $n$, while $\exp(-A(T,V,M)/k_BT)$ is the probability of the system having any state with $M$ particles.

**Using the grand canonical ensemble.** The grand canonical ensemble is particularly useful for non-interacting quantum systems (chapter 7). There each energy eigenstate can be thought of as a separate subsystem, independent of the others except for the competition between eigenstates for the particle number. A closely related ensemble emerges in chemical reactions (section 6.6).

For now, to illustrate how to use the grand canonical ensemble, let's compute the number fluctuations. The expected number of particles for a general system is

$$\langle N \rangle = \frac{\sum_m N_m e^{-(E_m-\mu N_m)/k_BT}}{\sum_m e^{-(E_m-\mu N_m)/k_BT}} = \frac{k_BT}{\Xi} \partial\Xi/\partial\mu = -\partial\Phi/\partial\mu. \tag{6.38}$$

Just as the fluctuations in the energy were related to the specific heat (the rate of change of energy with temperature, section 6.1), the number fluctuations are related to the rate of change of particle number with chemical potential.

$$\begin{aligned}\frac{\partial\langle N\rangle}{\partial\mu} &= \frac{\partial}{\partial\mu} \frac{\sum_m N_m e^{-(E_m-\mu N_m)/k_BT}}{\Xi} \\ &= -\frac{1}{\Xi^2} \frac{\left(\sum_m N_m e^{-(E_m-\mu N_m)/k_BT}\right)^2}{k_BT} \\ &\qquad + \frac{1}{k_BT} \frac{\sum_m {N_m}^2 e^{-(E_m-\mu N_m)/k_BT}}{\Xi} \\ &= \frac{\langle N^2\rangle - \langle N\rangle^2}{k_BT} = \frac{\langle (N-\langle N\rangle)^2\rangle}{k_BT} \end{aligned} \tag{6.39}$$

## 6.4 What is Thermodynamics?

Thermodynamics and statistical mechanics historically were closely tied, and often they are taught together. What is thermodynamics?

(0) **Thermodynamics**: *Physics that deals with the mechanical motions or relations of heat.* (Webster's)

(1) **Thermodynamics is the theory that emerges from statistical mechanics in the limit of large systems.** Statistical mechanics originated as a derivation of thermodynamics from an atomistic microscopic theory (somewhat before the existence of atoms was universally accepted). Thermodynamics can be viewed as statistical mechanics in the limit[12] as the number of particles $N \to \infty$. When we calculate the relative fluctuations in properties like the energy or the pressure and show that they vanish like $1/\sqrt{N}$, we are providing a microscopic justification for thermodynamics. Thermodynamics is the statistical mechanics of near-equilibrium systems when one ignores the fluctuations.

In this text, we will summarize many of the important methods and results of traditional thermodynamics in the exercises (3.6, 6.4, 6.6, 6.7, and 6.8). Our discussions of order parameters (chapter 9) will be providing thermodynamic laws, broadly speaking, for a wide variety of states of matter.

Statistical mechanics has a broader purview than thermodynamics. Particularly in applications to other fields like information theory, dynamical systems, and complexity theory, statistical mechanics describes many systems where the emergent behavior does not have a recognizable relation to thermodynamics.

(2) **Thermodynamics is a self-contained theory.** Thermodynamics can be developed as an axiomatic system. It rests on the so-called three laws of thermodynamics, which for logical completeness must be supplemented by a 'zeroth' law. Informally, they are:

(0) *Transitivity of equilibria:* If two systems are in equilibrium with a third, they are in equilibrium with one another.
(1) *Conservation of energy:* The total energy of an isolated system, including the heat energy, is constant.
(2) *Entropy always increases:* An isolated system may undergo irreversible processes, whose effects can be measured by a state function called the entropy.
(3) *Entropy goes to zero at absolute zero:* The entropy per particle of any two large equilibrium systems will approach the same value[13] as the temperature approaches absolute zero.

The zeroth law (transitivity of equilibria) becomes the basis for defining the temperature. Our statistical mechanics derivation of the temperature in section 3.3 provides the microscopic justification of the zeroth law: systems that can only exchange heat energy are in equilibrium with one another when they have a common value of $\frac{1}{T} = \frac{\partial S}{\partial E}\big|_{V,N}$.

The first law (conservation of energy) is now a fundamental principle of physics. Thermodynamics automatically inherits it from the microscopic theory. Historically, the thermodynamic understanding of how work transforms into heat was important in establishing that energy is conserved. Careful arguments about the energy transfer due to heat flow and mechanical work[14] are central to thermodynamics.

[12] The limit $N \to \infty$ is thus usually called the *thermodynamic limit*, even for systems like second–order phase transitions where the fluctuations remain important and thermodynamics *per se* is not applicable.

[13] This value is set to zero by dividing $\Omega(E)$ by $h^{3N}$, as in section 3.5.

[14] As we saw in our analysis of the Carnot cycle in section 5.1.

The second law (entropy always increases) is the heart of thermodynamics.[15] It is responsible for everything from forbidding perpetual motion machines to predicting the heat death of the universe (exercise 5.1). Entropy and its increase is not a part of our microscopic laws of nature, but is the foundation – an axiom – for our macroscopic theory of thermodynamics. The subtleties of how entropy and its growth emerges out of statistical mechanics was a theme of chapter 5 and is the focus of several exercises (5.4, 5.5, 7.2, and 8.8).

The third law (entropy goes to zero at $T = 0$, also known as Nernst's theorem), basically reflects the fact that quantum systems at absolute zero are in a ground state. Since the number of ground states of a quantum system typically is small[16] and the number of particles is large, systems at absolute zero have zero entropy per particle.[17]

The laws of thermodynamics have been written in many equivalent ways.[18] Carathéodory, for example, states the second law as *There are states of a system, differing infinitesimally from a given state, which are unattainable from that state by any quasi-static adiabatic*[19] *process.* The axiomatic form of the subject has attracted the attention of mathematicians.

In this text, we will not attempt to derive properties axiomatically or otherwise from the laws of thermodynamics: we focus on statistical mechanics.

(3) **Thermodynamics is a zoo of partial derivatives, transformations, and relations.** More than any other field of science, the thermodynamics literature seems filled with partial derivatives and tricky relations between varieties of physical quantities.

This is in part because there are several alternative free energies to choose between. For studying molecular systems one has not only the entropy (or the internal energy), the Helmholtz free energy, and the grand free energy, but also the Gibbs free energy, the enthalpy, and a number of others. There are corresponding free energies for studying magnetic systems, where instead of particles one studies the local magnetization or spin. There appears to be little consensus between textbooks on the symbols or even the names of these various free energies.

How do we transform from one free energy to another? Let's write out the Helmholtz free energy in more detail:

$$A(T, V, N) = E - TS(E, V, N). \tag{6.40}$$

The terms on the right-hand side of the equation involve four variables: $T$, $V$, $N$, and $E$. Why is $A$ only a function of three? Consider the derivative of $A = E_s - T_b S_s(E_s)$ with respect to the energy $E_s$ of the system, at fixed bath temperature $T_b$:

$$\partial A/\partial E_s = 1 - T_b \partial S_s/\partial E_s = 1 - T_b/T_s. \tag{6.41}$$

[15] In *The Two Cultures*, C. P. Snow suggests being able to describe the Second Law of Thermodynamics is to science as having read a work of Shakespeare is to the arts. (Some in non-English speaking cultures may wish to object.) Remembering which law is which number is not of great import, but the concept of entropy and its inevitable increase is indeed central.

[16] Some systems may have broken symmetry states, or multiple degenerate ground states, but the number of such states is typically independent of the size of the system, or at least does not grow exponentially with the number of particles, so the entropy per particle goes to zero.

[17] Systems like glasses that have not reached complete equilibrium can have non-zero residual entropy as their effective temperature goes to zero (section 5.2.2, exercise 5.8).

[18] Occasionally you hear them stated (1) You can't win, (2) You can't break even, and (3) You can't get out of the game. The connection between (3) and the third law is unclear.

[19] Carathéodory is using the term *adiabatic* just to exclude heat flow: we use it to also imply infinitely slow (quasi-static) transitions.

Since $A$ represents the system in equilibrium with the bath, the temperature of the system and the bath must agree, and hence $\frac{\partial A}{\partial E} = 0$: $A$ is independent of $E$. Physically, energy is transferred until $A$ is a minimum; $E$ is no longer an independent variable. This is an example of a *Legendre transformation* (see exercise 6.8). Legendre transformations allow one to change from one type of energy or free energy to another, by changing from one set of independent variables (here $E$, $V$, and $N$) to another ($T$, $V$, and $N$).

Thermodynamics seems cluttered in part also because it is so powerful. Almost any macroscopic property of interest can be found by taking derivatives of the free energy. First derivatives of the entropy, energy, or free energy give properties like the temperature and pressure. Thermodynamics introduces a condensed notation to help organize these derivatives. For example,[20]

$$dE = T\,dS - P\,dV + \mu\,dN. \tag{6.42}$$

basically asserts that $E(S, V, N)$ satisfies equations 3.29, 3.40, and 3.42:

$$\left.\frac{\partial E}{\partial S}\right|_{N,V} = T, \quad \left.\frac{\partial E}{\partial V}\right|_{N,S} = -P, \quad \text{and} \quad \left.\frac{\partial E}{\partial N}\right|_{V,S} = \mu. \tag{6.43}$$

The corresponding equation for the Helmholtz free energy $A(T, V, N)$ is

$$\begin{aligned} dA = d(E - TS) &= dE - T\,dS - S\,dT \\ &= -S\,dT - P\,dV + \mu\,dN. \end{aligned} \tag{6.44}$$

which satisfies

$$\left.\frac{\partial A}{\partial T}\right|_{N,V} = -S, \quad \left.\frac{\partial A}{\partial V}\right|_{N,T} = -P, \quad \text{and} \quad \left.\frac{\partial A}{\partial N}\right|_{V,T} = \mu. \tag{6.45}$$

Similarly, systems at constant temperature and pressure (for example, most biological and chemical systems) minimize the *Gibbs free energy* (figure 6.4)

$$G(T, P, N) = E - TS + PV \qquad dG = -S\,dT + V\,dP + \mu\,dN. \tag{6.46}$$

Systems at constant energy and pressure minimize the *enthalpy*

$$H(E, P, N) = E + PV \qquad dH = T\,dS + V\,dP + \mu\,dN, \tag{6.47}$$

and, as noted in section 6.3, systems at constant temperature, volume, and chemical potential are described by the grand free energy

$$\Phi(T, V, \mu) = E - TS - \mu N \qquad d\Phi = -S\,dT - P\,dV - N\,d\mu. \tag{6.48}$$

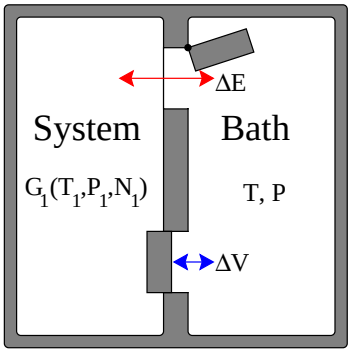



**Fig. 6.4 The Gibbs ensemble** $G(T, P, N)$ embodies the thermodynamics of systems that can exchange heat and volume with a bath. The **enthalpy** $H(E, P, N)$ is used for systems that only exchange volume.

[20] These formulas have precise meanings in differential geometry, where the terms $dX$ are *differential forms*. Thermodynamics distinguishes between *exact* differentials like $dS$ and *inexact* differentials like work and heat which are not derivatives of a state function, but path-dependent quantities. Mathematicians have *closed* and *exact* differential forms, which (in a simply connected space) both correspond to the exact differentials in thermodynamics. The relation between closed and exact differential forms is the basis for *cohomology theory*... These elegant topics are not central, though, to statistical mechanics and we will not pursue them here.

There are also many tricky, unintuitive relations in thermodynamics. The first derivatives must agree around a tiny triangle, yielding a tricky relation between their products (equation 3.37). Second derivatives of the free energy give properties like the specific heat, the bulk modulus, and the magnetic susceptibility. The second derivatives must be symmetric ($\frac{\partial^2}{\partial x \partial y} = \frac{\partial^2}{\partial y \partial x}$), giving tricky Maxwell relations between what naively seem different susceptibilities (exercise 3.6). There are further tricks involved with taking derivatives in terms of 'unnatural variables',[21] and there are many inequalities that can be derived from stability criteria.

Of course, statistical mechanics is not really different from thermodynamics in having a zoo of complex relationships. Indeed, statistical mechanics has its own collection of important relations that connect equilibrium fluctuations to transport and response, like the Einstein relation connecting fluctuations to diffusive transport in section 2.3 and the fluctuation-dissipation theorem we will derive in chapter 10. In statistical mecahnics, though, the focus of attention is usually not on the zoo of general relations, but on calculating the properties of specific systems.

[21] For example, with work you can take the derivative of $S(E, V, N)$ with respect to $P$ at constant $T$ without re-expressing it in the variables $P$ and $T$.

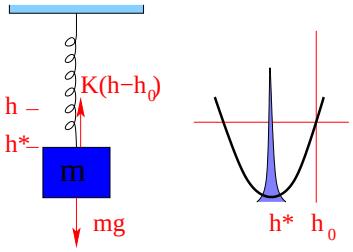



**Fig. 6.5** A mass on a spring in equilibrium sits very close to the minimum of the energy.

## 6.5 Mechanics: Friction and Fluctuations

A mass $M$ hangs on the end of a spring with spring constant $K$ and unstretched length $h_0$, subject to a gravitational field of strength $g$. How far does the spring stretch? We have all solved innumerable statics exercises of this sort in first-year mechanics courses. The spring stretches to a length $h^*$ where $-mg = K(h^* - h_0)$. At $h^*$ the forces balance and the energy is minimized.

What principle of physics is this? In physics, energy is conserved, not minimized! Shouldn't we be concluding that the mass will oscillate with a constant amplitude forever?

We have now come to the point in your physics education where we can finally explain why the mass appears to minimize energy. Here our system (the mass and spring)[22] is coupled to a very large number $N$ of internal atomic or molecular degrees of freedom. The oscillation of the mass is coupled to these other degrees of freedom (friction) and will share its energy with them. The vibrations of the atoms is heat: the energy of the pendulum is dissipated by friction into heat. Indeed, since the spring potential energy is quadratic we can use the equipartition theorem: in equilibrium $\frac{1}{2}K(h - h^*)^2 = \frac{1}{2}k_B T$. For a spring with $K = 10N/m$ at room temperature ($k_B T = 4 \times 10^{-21} J$), $\sqrt{\langle (h - h^*)^2 \rangle} = \sqrt{k_B T / K} = 2 \times 10^{-11} m = 0.2 \mathring{A}$. The energy is indeed minimized up to tiny thermal fluctuations. We'll return to this question in exercise 10.1.

How do we connect this statistical mechanics picture to the friction coefficient of the damped harmonic oscillator? A careful statistical mechanics treatment (exercise 10.5) gives a law of motion for the mass of

[22] We think of the subsystem as being just the macroscopic configuration of mass and spring, and the atoms comprising them as being part of the environment, the rest of the system.

the form

$$\ddot{h} = -\frac{K}{m}(h - h^\star) - \gamma\dot{h} + \xi(t) \tag{6.49}$$

where $\gamma$ represents the friction or *dissipation*, and $\xi(t)$ is a random, time-dependent noise force coming from the internal vibrational degrees of freedom of the system. This is an example of a *Langevin equation*. The strength of the noise $\xi$ depends on the dissipation $\gamma$ and the temperature $T$ so as to guarantee a Boltzmann distribution as the steady state. In general both $\xi$ and $\gamma$ can be frequency dependent: we'll study these issues in detail in chapter 10.

## 6.6 Chemical Equilibrium and Reaction Rates

In studying chemical reactions, one is often interested in the number of molecules of various types as a function of time, and not interested in observing properties depending on the positions or momenta of the molecules. In this section we develop a free energy to derive the laws of chemical equilibrium, and in particular the *law of mass-action*. We will then discuss more carefully the subtle question of exactly when the chemical reactions takes place, and motivate the *Arrhenius law* of thermally activated reaction rates.

Chemical reactions change one type of molecule into another. For example, ammonia ($NH_3$) can be produced from hydrogen and nitrogen through the reaction

$$3H_2 + N_2 \leftrightarrows 2NH_3. \tag{6.50}$$

All chemical reactions are in principle reversible, although the backward reaction rate may be very different from the forward rate. In *chemical equilibrium* at fixed volume,[23] the concentrations [X] of the various molecules X (in number per unit volume, say), satisfies the law of mass-action[24]

$$\frac{[NH_3]^2}{[N_2][H_2]^3} = K_{eq}(T). \tag{6.51}$$

The law of mass action can be motivated by imagining a chemical reaction arising from a simultaneous collision of all the reactants. The probability of one nitrogen and three hydrogen molecules colliding in a small reaction region is proportional to the nitrogen concentration and to the cube of the hydrogen concentration, so the forward reaction will occur with some rate per unit volume $K_F[N_2][H_2]^3$; similarly the backward reaction will occur with a rate per unit volume $K_B[NH_3]^2$ proportional to the probability that two $NH_3$ molecules will collide. Balancing these two rates to get a steady state gives us equation 6.51 with $K_{eq} = K_F/K_B$.

This motivating argument becomes unconvincing when one realizes that the actual reaction may proceed through several short-lived intermediate states – at no point is a multiple collision required.[25] How can we derive this law soundly from statistical mechanics?

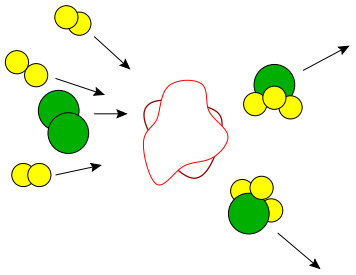


**Fig. 6.6 Ammonia collision.** The simple motivating argument for the law of mass action views the reaction as a simultaneous collision of all the reactants.

[23] Experimentally it is more common to work at constant pressure, which makes things more complicated but conceptually no more interesting.

[24] More generally, we can write a reaction as $\sum_i \nu_i A_i = 0$. Here the $\nu_i$ are the *stoichiometries*, giving the number of molecules of type $A_i$ changed during the reaction (with $\nu_i < 0$ for reactants and $\nu_i > 0$ for products). The law of mass-action in general states that $\prod_i [A_i]_i^{\nu} = K_{eq}$.

[25] The Haber-Bosch process used industrially for producing ammonia involves sev-

Since we are uninterested in the positions and momenta, at fixed volume and temperature our system is described by a Helmholtz free energy $A(T, V, N_{H_2}, N_{N_2}, N_{NH_3})$. When the chemical reaction takes place, it changes the number of the three molecules, and changes the free energy of the system:

$$\begin{aligned}\Delta A &= \frac{\partial A}{\partial N_{H_2}}\Delta N_{H_2} + \frac{\partial A}{\partial N_{N_2}}\Delta N_{N_2} + \frac{\partial A}{\partial N_{NH_3}}\Delta N_{NH_3} \\ &= -3\mu_{H_2} - \mu_{N_2} + 2\mu_{NH_3}. \end{aligned} \tag{6.52}$$

where $\mu_X = \frac{\partial A}{\partial X}$ is the chemical potential of molecule $X$. The reaction will proceed until the free energy is at a minimum,[26] so

$$-3\mu_{H_2} - \mu_{N_2} + 2\mu_{NH_3} = 0 \tag{6.53}$$

in equilibrium. In complete generality, we may deduce that chemical reactions proceed in the direction which reduces the net chemical potential of the system.

To derive the law of mass-action, we must now make an assumption: that the molecules are uncorrelated in space.[27] This makes each molecular species into a separate ideal gas. The Helmholtz free energy of the three gases are of the form

$$A(N, V, T) = Nk_BT\left[\log((N/V)\lambda^3) - 1\right] + N\Delta \tag{6.54}$$

where $\lambda = h/\sqrt{2\pi m k_B T}$ is the thermal deBroglie wavelength. The first two terms give the contribution to the partition function from the positions and momenta of the molecules (equation 6.24); the last term $N\Delta$ comes from the internal free energy of the molecules.[28] So, the chemical potential

$$\begin{aligned}\mu(N, V, T) = \frac{\partial A}{\partial N} &= k_BT\left[\log((N/V)\lambda^3) - 1\right] + Nk_BT(1/N) + \Delta \\ &= k_BT\log((N/V)\lambda^3) + \Delta \\ &= k_BT\log(N/V) + c + \Delta \end{aligned} \tag{6.55}$$

where the constant $c = k_BT\log(\lambda^3)$ is temperature and mass dependent, but independent of density. Using equation 6.55 in equation 6.53, dividing by $k_BT$, writing concentrations $[X] = N_X/V$, and pulling terms independent of concentrations to the right, we find the law of mass action:

$$\begin{aligned} -3\log[H_2] - \log[N_2] + 2\log[NH_3] &= \log(K_{eq}) \\ \Longrightarrow \quad \frac{[NH_3]^2}{[H_2]^3[N_2]} &= K_{eq}. \end{aligned} \tag{6.56}$$

We also find that the equilibrium rate constant depends exponentially on the net internal free energy difference[29] $\Delta_T = 3\Delta_{H_2} + \Delta_{N_2} - 2\Delta_{NH_3}$

[26] More precisely, the equilibrium probability that a state will be found with $\Delta N_{N_2}$ nitrogens away from this equilibrium is suppressed by a factor $\exp((-3\mu_{H_2} - \mu_{N_2} + 2\mu_{NH_3})\Delta N_{N_2}/k_BT)$.

[27] This assumption is often valid also for ions and atoms in solution: if the ion-ion interactions can be neglected and the solute (water) is not important, the ions are well described as ideal gases, with corrections due to integrating out the solvent degrees of freedom.

[28] If all the molecules remain in their ground states, $\Delta$ is just given by the molecular ground state energy.

[29] For the ammonia reaction, $\Delta_T = 92.4 kJ/mole$.

eral intermediate states. The the nitrogen and hydrogen molecules adsorb and disas-

between reactants and products:

$$K_{eq} \approx K_0 \exp(-\Delta_T/k_B T) \tag{6.57}$$

and the prefactor

$$K_0 = \frac{\lambda_{H_2}^9 \lambda_{N_2}^3}{\lambda_{NH_3}^6} = \frac{h^6 m_{NH_3}^3}{8 k_B^3 \pi^3 T^3 m_{H_2}^{9/2} m_{N_2}^{3/2}} \propto T^{-3}. \tag{6.58}$$

depends weakly on the temperature. The factor $e^{-\Delta_T/k_B T}$ represents the Boltzmann factor favoring a final state with molecular free energy $\Delta_T$ lower than the initial state, and $K_0$ is a much weaker factor representing the larger momentum–space entropy for the four initial molecules versus the two final molecules.

How do we define, during a messy collision, which atoms belong to which molecules? Exactly when during the trajectory do we say that the reaction has occurred? An $M$-atom chemical reaction (classically) is a trajectory in a $3M$ dimensional configuration space. It is traditional in chemistry to pick out one "reaction coordinate" $X$, and plot the energy (minimized with respect to the other $3M-1$ coordinates) versus $X$. Figure 6.7 shows this energy plot. [30] Notice the energy barrier $B$ separating the reactants from the products: the atomic configuration at the top of the barrier is called the *transition state*. Clearly atoms to the left of the barrier peak are reactants, and atoms to the right are products. This barrier, in $3M$-dimensional configuration space, is actually a saddlepoint. Dividing the reactants from the products demands the identification of a $3M-1$-dimensional transition state *dividing surface*. Our free energy $A(T, V, N_{H_2}, N_{N_2}, N_{NH_3})$ is properly a partial trace, with all configurations to the left of the transition state $B$ contributing to the free energy of the reactants and all configurations to the left of $B$ contributing as products.

How fast does our chemical reaction proceed, if we start out of equilibrium with extra reactants? Within the mass-action approximation that our molecules are not correlated in position, surely the forward reaction rate $K_F[H_2]^3[N_2]$ can't depend on the concentration of the product $[NH_3]$. If our reactions occur slowly enough so that the molecules remain in equilibrium at the current concentrations, we can estimate the *non-equilibrium* reaction rate by studying the *equilibrium* transitions from reactant to product at the same reactant concentration.

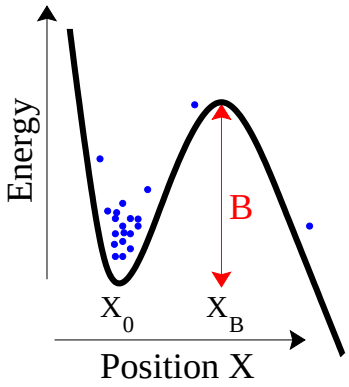



**Fig. 6.7 Barrier Crossing Potential.** Energy $E$ as a function of some reaction coordinate $X$ for a chemical reaction. The dots schematically represent how many atoms are at each position. (The "heights" of the dots are made up to suggest that they are confined in the well.) The reactants (left) are separated from the products (right) by an energy barrier of height $B$. One can estimate the rate of the reaction by calculating the number of reactants crossing the top of the barrier per unit time.

sociating into atoms on an iron substrate, the nitrogen atom picks up one hydrogen atom at a time, and finally the $NH_3$ molecule desorbs into the vapor. The iron acts as a *catalyst*. A catalyst, (or *enzyme*, for catalytic biological proteins) lowers the energy barrier separating reactants from products, speeding up the reaction rates without changing the equilibrium concentrations.

[30] The picture 6.7 is not completely appropriate for reactions, since in a collision the reactants and products come and go from infinity, not from the bottom of potential wells. The figure is applicable and less artificial for many other problems. For diffusion in crystals, replacing the continuous motions with discrete jumps by lattice vectors leads to lattice models for alloys, (section 8.2). For systems with double–well bistable atomic potentials (like glasses (section 5.2.2) and molecules (exercises 10.2 and 10.3)) this partial trace leaves us with a *two-state system*.

The reaction rate cannot be larger than the total number of atoms in equilibrium crossing past the energy barrier from reactant to product. It can be smaller if trajectories which do cross the barrier often immediately re-cross backward before equilibrating on the other side.[31] Such re-crossings are minimized by choosing the transition state dividing surface properly.[32] The density of particles at the top of the barrier is smaller than the density at the bottom of the well by a Boltzmann factor of $\exp(-B/k_BT)$. The rate of the reaction will therefore be of the thermally activated, or Arrhenius form

$$\Gamma = \Gamma_0 \exp(-B/k_BT) \tag{6.59}$$

with some prefactor $\Gamma_0$. By carefully calculating the population near the bottom of the well and the population and velocities near the top of the barrier, one can derive a formula for $\Gamma_0$ (see exercise 6.2).

This Arrhenius law for thermally activated motion governs not only chemical reaction rates, but also diffusion constants and more macroscopic phenomena like nucleation rates (section 11.2).[33]

## 6.7 Free Energy Density for the Ideal Gas

In section 2.2 we studied the dynamics of spatially varying densities of particles (or probability) using the diffusion equation. How is this connected with free energies and ensembles? Broadly speaking, inhomogeneous systems out of equilibrium can also be described by statistical mechanics, if the gradients in space and time are small enough that the system is close to a *local equilibrium*. We can then represent the local state of the system by *order parameter fields*, one field for each property (density, temperature, magnetization) needed to characterize the state of a uniform, macroscopic body. We can describe a spatially varying, inhomogeneous systems that is nearly in equilibrium using a *free energy density*, typically depending on the order parameter fields and their gradients.

We'll be discussing order parameter fields and free energy densities for a wide variety of complex systems in chapter 9. There we will use symmetries and gradient expansions to derive the form of the free energy density, because it will be too complex to compute directly. In this section, we will directly derive the free energy density for an inhomogeneous ideal gas, to give a tangible example of the general case. We will also use the free energy density of the ideal gas when we study correlation functions in section 10.3.

Again, the Helmholtz free energy of an ideal gas is nicely written (equation 6.24 in terms of the density $\rho = N/V$ and the thermal de-Broglie wavelength 3.63 $\lambda$:

$$A(N,V,T) = Nk_BT\left[\log(\rho\lambda^3) - 1\right], \tag{6.60}$$

Hence the free energy density for $n_j = \rho(\mathbf{x}_j)\Delta V$ atoms in a small volume

[31] This method of calculating rates is called *transition state theory*; the transition state is the saddlepoint between the reactants and the products. Recrossing is a *dynamical correction* to transition state theory: see reference [44].

[32] At low temperatures, it is mainly important that this surface be perpendicular to the unstable 'downward' eigendirection of the Hessian for the potential energy at the transition state.

[33] There are basically three ways in which slow processes arise in physics. (1) Large systems can respond slowly to external changes because communication from one end of the system to the other is sluggish: examples are the slow decay at long wavelengths in the diffusion equation (section 2.2) and Goldstone modes (section 9.3). (2) Systems like radioactive nuclei can respond slowly – decaying with lifetimes of billions of years – because of the slow rate of quantum tunneling through barriers. (3) Systems can be slow because they must thermally activate over barriers (with the Arrhenius rate of equation 6.59).

$\Delta V$ is

$$\mathcal{F}^{\rm ideal}(\rho(\mathbf{x}_j),T) = \frac{A(n_j,\Delta V,T)}{\Delta V} = \rho(\mathbf{x}_j)k_BT\left[\log(\rho(\mathbf{x}_j)\lambda^3)-1\right], \tag{6.61}$$

The probability density for a given particle density $\rho(x)$ is

$$P\{\rho\} = e^{-\beta\int dV\,\mathcal{F}^{\rm ideal}(\rho(\mathbf{x}))}/Z. \tag{6.62}$$

As usual, the free energy $F\{\rho\} = \int \mathcal{F}(\rho(\mathbf{x}))d\mathbf{x}$ acts just like the energy in the Boltzmann distribution. We've *integrated out* the microscopic degrees of freedom (positions and velocities of the individual particles) and replaced them with a coarse-grained field $\rho(x)$.

The free energy density of equation 6.61 can be used to determine any equilibrium property of the system that can be written in terms of the density $\rho(x)$. In chapter 10, for example, we will use it to calculate correlation functions $\langle\rho(x)\rho(x')\rangle$, and will discuss their relationship with susceptibilities and dissipation.

The free energy density also provides a framework for discussing the evolution laws for nonuniform densities. A system prepared with some nonuniform density will evolve in time $\rho(x,t)$; if in each small volume $\Delta V$ the system is close to equilibrium, then one may expect that its time evolution can be described by equilibrium statistical mechanics even though it is not globally in equilibrium. A non-uniform density will have a force which pushes it towards uniformity: the total free energy will decrease when particles flow from regions of high particle density to low density. We can use our free energy density to calculate this force, and then use the force to derive laws (depending on the system) for the time evolution.

The chemical potential for a uniform system is $\mu = \frac{\partial A}{\partial N} = \frac{\partial A/V}{\partial N/V} = \frac{\partial \mathcal{F}}{\partial \rho}$; the change in free energy for a change in the average density $\rho$. For a non-uniform system, the local chemical potential at a point $x$ is

$$\mu(x) = \frac{\delta\mathcal{F}}{\delta\rho} \tag{6.63}$$

the variational derivative of the free energy density with respect to $\rho$. Because our ideal gas free energy has no terms involving gradients of $\rho$, the variational derivative $\frac{\delta\mathcal{F}}{\delta\rho}$ equals the partial derivative $\frac{\partial\mathcal{F}}{\partial\rho}$:

$$\begin{aligned}\mu(x) &= \frac{\delta\mathcal{F}^{\rm ideal}}{\delta\rho} = \frac{\delta}{\delta\rho}\left(\rho k_BT\left[\log(\rho\lambda^3)-1\right]\right)\\ &= k_BT\left[\log(\rho\lambda^3)-1\right] + \rho k_BT/\rho\\ &= k_BT\log(\rho\lambda^3).\end{aligned} \tag{6.64}$$

The chemical potential is like a number pressure for particles: a particle can lower the free energy by moving from regions of high chemical potential to low chemical potential. The gradient of the chemical potential $-\frac{\partial\mu}{\partial x}$ is thus a pressure gradient, effectively the statistical mechanical force on an particle.

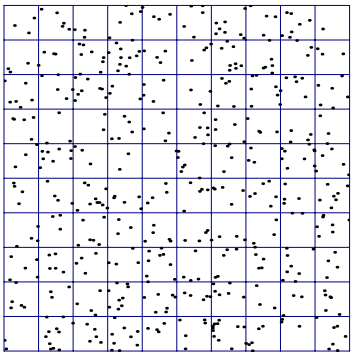


**Fig. 6.8 Density fluctuations in space.** If $n_j$ is the number of points in a box of size $\Delta V$ at position $\mathbf{x}_j$, then $\rho(\mathbf{x}_j) = n_j/\Delta V$.

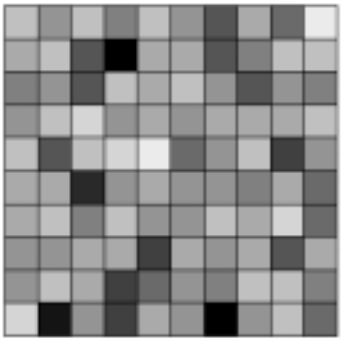


**Fig. 6.9 Coarse-grained density in space.** The density field $\rho(\mathbf{x})$ represents all configurations of points $\mathbb{Q}$ consistent with the average. Its free energy density $\mathcal{F}^{\rm ideal}(\rho(\mathbf{x}))$ contains contributions from all microstates $(\mathbb{P},\mathbb{Q})$ with the correct number of particles per box. The probability of finding this particular density is proportional to the integral of all ways $\mathbb{Q}$ that the particles in figure 6.8 can be arranged to give this density.

How will the particle density $\rho$ evolve in response to this force $\mu(x)$? This depends upon the problem. If our density was the density of the entire gas, the atoms would accelerate under the force – leading to sound waves.[34] There momentum is conserved as well as particle density. If our particles could be created and destroyed, the density evolution would include a term $\partial\rho/\partial t = -\eta\mu$ not involving a current. In systems that conserve (or nearly conserve) energy, the evolution will depend on Hamilton's equations of motion for the free energy density: in magnets, the magnetization responds to an external force by precessing; in superfluids, gradients in the chemical potential are associated with winding and unwinding the phase of the order parameter field (vortex motion)...

[34] In that case, we'd need to add the local velocity field into our description of the local environment.

Let's focusing the case of a small amount of perfume in a large body of still air. Here particle density is locally conserved, but momentum is strongly damped (since the perfume particles can scatter off of the air molecules). The velocity of our particles will be proportional to the effective force on them $\mathbf{v} = -\gamma\frac{\partial\mu}{\partial x}$, with $\gamma$ the mobility.[35] Hence the current $\mathbf{J} = \rho v$ of particles will will be

$$\begin{aligned}\mathbf{J} &= \gamma\rho(x)\left(-\frac{\partial\mu}{\partial x}\right) = -\gamma\rho(x)\frac{\partial k_B T \log(\rho\lambda^3)}{\partial x} \\ &= -\gamma\rho(x)\frac{k_B T}{\rho}\frac{\partial\rho}{\partial x} = -\gamma k_B T\frac{\partial\rho}{\partial x}\end{aligned} \tag{6.65}$$

and thus the rate of change of $\rho$ is given by the diffusion equation

$$\frac{\partial\rho}{\partial t} = -\nabla\cdot J = \gamma k_B T\frac{\partial^2\rho}{\partial x^2}. \tag{6.66}$$

[35] This is *linear response*. Systems that are nearly in equilibrium typically have currents proportional to gradients in their properties: examples include Ohm's law that the electrical current is proportional to the gradient of the electromagnetic potential $I = V/R = \frac{d\phi}{dx}/R$, thermal conductivity where the heat flow is proportional to the gradient in temperature $\mathbf{J} = \kappa\nabla T$, and viscous fluids, where the shear rate is proportional to the shear stress. We'll study linear response with more rigor in chapter 10.

Notice,

- We again have derived the diffusion equation (equation 2.7) $\frac{\partial\rho}{\partial t} = D\frac{\partial^2\rho}{\partial x^2}$, this time by starting with a free energy density from equilibrium statistical mechanics, and assuming a linear law relating velocity to force,
- We have rediscovered the Einstein relation (equation 2.22) $D = \gamma k_B T$,
- We have asserted that $-\frac{\partial\mu}{\partial x}$ acts just like an external force, even though $\mu$ comes from the ideal gas, which itself has no potential energy (see exercises 5.9 and 6.11).

Our free energy density for the ideal gas is simpler than the free energy density of a general system because the ideal gas has no stiffness to gradients in the density. Our derivation above worked by splitting space into little boxes: in general, these box regions will not be independent systems, and there will be a free energy difference that depends on the change in the coarse-grained fields between boxes – leading to terms involving gradients of the field.

# Exercises

(6.1) **Two–state system.** (Basic)

Consider the statistical mechanics of a tiny object with only two discrete states:[36] one of energy $E_1$ and the other of higher energy $E_2 > E_1$.

*(a)* **Boltzmann probability ratio.** *Find the ratio of the equilibrium probabilities $\rho_2/\rho_1$ to find our system in the two states, when weakly coupled to a heat bath of temperature $T$. What is the limiting probability as $T \to \infty$? As $T \to 0$?* Related formula: Boltzmann probability $= Z(T)\exp(-E/kT) \propto \exp(-E/kT)$.

*(b)* **Probabilities and averages.** *Use the normalization of the probability distribution (the system must be in one or the other state) to find $\rho_1$ and $\rho_2$ separately. (That is, solve for $Z(T)$ in the 'related formula' for part (A).) What is the average value of the energy $E$?*

(6.2) **Barrier Crossing.** (Basic, Chemistry)

In this exercise, we will derive the Arrhenius law (equation 6.59)

$$\Gamma = \Gamma_0 \exp(-E/k_B T) \tag{6.67}$$

giving the rate at which chemical reactions cross energy barriers.

The important exponential dependence on the barrier height $E$ is the relative Boltzmann probability that a particle is near the top of the barrier (and hence able to escape). Here we will do a relatively careful job of calculating the prefactor $\Gamma_0$.

Consider a system described by a coordinate $X$, with an energy $U(X)$ with a minimum at $X_0$ with energy zero and an energy barrier at $X_B$ with energy $U(X_B) = B$.[37] Let the temperature of the system be much smaller than $B/k_B$. To do our calculation, we will make some approximations. (1) We assume that the atoms escaping across the barrier to the right do not scatter back into the well. (2) We assume that the atoms deep inside the well are in equilibrium. (3) We assume that the particles crossing to the right across the barrier are given by the equilibrium distribution inside the well.

*(a) Let the probability that a particle has position $X$ be $\rho(X)$. What is the ratio of probability densities $\rho(X_B)/\rho(X_0)$ if the particles near the top of the barrier are assumed in equilibrium with those deep inside the well?* Related formula: Boltzmann distribution $\rho \propto \exp(-E/k_B T)$.

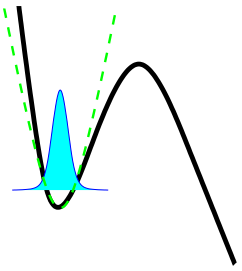


**Fig. 6.10 Well Probability Distribution.** The approximate probability distribution for the atoms still trapped inside the well.

If the barrier height $B \gg k_B T$, then most of the particles in the well stay near the bottom of the well. Often, the potential near the bottom is accurately described by a quadratic approximation $U(X) \approx \frac{1}{2} M\omega^2 (X - X_0)^2$, where $M$ is the mass of our system and $\omega$ is the frequency of small oscillations in the well.

*(b) In this approximation, what is the probability density $\rho(X)$ near the bottom of the well? (See figure 6.10.) What is $\rho(X_0)$, the probability density of having the system at the bottom of the well?* Related formula: Gaussian probability distribution $(1/\sqrt{2\pi\sigma^2})\exp(-x^2/2\sigma^2)$.

Hint: Make sure you keep track of the $2\pi$s.

[36] Visualize this as a tiny biased coin, which can be in the 'heads' or 'tails' state but has no other internal vibrations or center of mass degrees of freedom. Many systems are well described by large numbers of these two–state systems: some paramagnets, carbon monoxide on surfaces, glasses at low temperatures, ...

[37] This potential could describe a chemical reaction, with $X$ being a reaction coordinate. It could describe the escape of gas from a moon of Jupiter, with $X$ being the distance from the moon in Jupiter's direction.

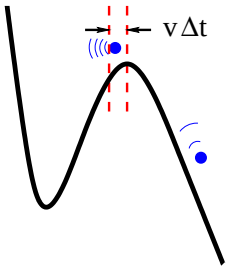



**Fig. 6.11 Crossing the Barrier.** The range of positions for which atoms moving to the right with velocity $v$ will cross the barrier top in time $\Delta t$.

Knowing the answers from (a) and (b), we know the probability density $\rho(X_B)$ at the top of the barrier. We need to also know the probability that particles near the top of the barrier have velocity $V$, because the faster-moving parts of the distribution of velocities contribute more to the flux of probability over the barrier (see figure 6.11). As usual, because the total energy is the sum of the kinetic and potential energy, the total Boltzmann probability factors: in equilibrium the particles will always have a velocity probability distribution $\rho(V) = 1/\sqrt{2\pi k_B T/M}\exp(-\frac{1}{2}MV^2/k_BT)$.

*(c) First give a formula for the decay rate* $\Gamma$ *(the probability per unit time that our system crosses the barrier towards the right), for an unknown probability density* $\rho(X_B)\rho(V)$ *as an integral over the velocity* $V$*. Then, using your formulas from parts (A) and (B), give your estimate of the decay rate for our system.* Related formula: $\int_0^\infty x\exp(-x^2/2\sigma^2)\,dx = \sigma^2$.

How could we go beyond this one-dimensional calculation? In the olden days, Kramers studied other one–dimensional models, changing the ways in which the system was coupled to the external bath. On the computer, one can avoid a separate heat bath and directly work with the full multidimensional configuration space, leading to *transition state theory*. The transition–state theory formula is very similar to the one you derived in part (c), except that the prefactor involves the product of all the frequencies at the bottom of the well and all the positive frequencies at the saddlepoint at the top of the barrier. (See reference [44].) Other generalizations arise when crossing multiple barriers [47] or in non–equilibrium systems [65].

(6.3) **Statistical Mechanics and Statistics.** (Mathematics)

Consider the problem of fitting a theoretical model to experimentally determined data. Let our model $M$ predict a time-dependent function $y^{(M)}(t)$. Let there be $N$ experimentally determined data points $y_i$ at times $t_i$ with errors of standard deviation $\sigma$. We assume that the experimental errors for the data points are independent and Gaussian distributed, so that the probability that our model actually generated the observed data points (the probability $P(D|M)$ of the data given the model) is

$$P(D|M) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{\left(y^{(M)}(t_i) - y_i\right)^2}{2\sigma^2}\right]. \tag{6.68}$$

*(a) True or false: This probability density corresponds to a Boltzmann distribution with energy* $H$ *and temperature* $T$*, with* $H = \sum_{i=1}^{N}(y^{(M)}(t_i) - y_i)^2/2$ *and* $k_BT = \sigma^2$.

There are two schools of statistics. Among a family of models, the frequentists will pick the model $M$ with the largest value of $P(D|M)$. The Bayesians take a different point of view. They argue that there is no reason to believe that all models have the same likelihood.[38] Suppose the intrinsic probability of the model (the *prior*) is $P(M)$. They use the theorem

$$P(M|D) = P(D|M)P(M)/P(D) = P(D|M)P(M) \tag{6.69}$$

where the last step notes that the probability that you measured the known data $D$ is presumably one.

The Bayesians often will pick the maximum of $P(M|D)$ as their model for the experimental data. But, given their perspective, it's even more natural to consider the entire *ensemble* of models, weighted by $P(M|D)$, as the best description of the data. This ensemble average then naturally provides error bars as well as predictions for various quantities.

Consider the problem of fitting a line to two data points. Suppose the experimental data points are at $t_1 = 0$, $y_1 = 1$ and $t_2 = 1$, $y_2 = 2$, where both $y$-values have uncorrelated Gaussian errors with standard deviation $\sigma = 1/2$, as assumed in equation (F.2.1) above. Our model $M(m, b)$ is $y(t) = mt+b$. Our Bayesian statistician knows that $m$ and $b$ both lie between zero and two, and assumes that the probability density is otherwise uniform: $P(m, b) = 1/4$ for $0 < m < 2$ and $0 < b < 2$.

*(b) Which of the contour plots in figure 6.12 accurately represent the probability distribution* $P(M|D)$ *for the model, given the observed data? (The spacing between the contour lines is arbitrary.)*

[38] There is no analogue of Liouville's theorem (chapter 4) in model space.

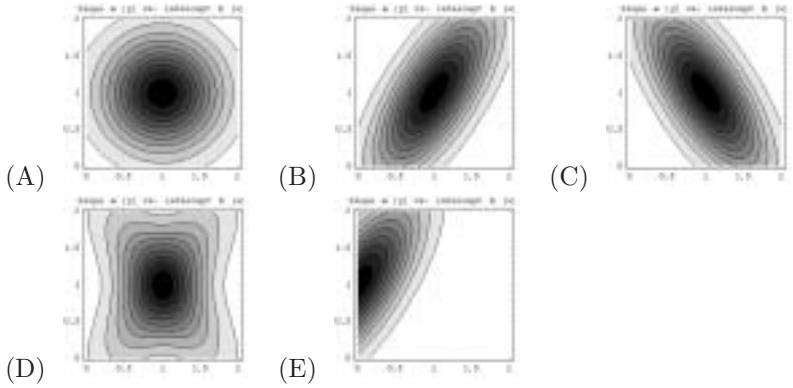



**Fig. 6.12**

(6.4) **Euler, Gibbs-Duhem, and Clausius-Clapeyron.** (Thermodynamics, Chemistry)

*(a) Using the fact that the entropy $S(N, V, E)$ is extensive, show that*

$$N \left.\frac{\partial S}{\partial N}\right|_{V,E} + V \left.\frac{\partial S}{\partial V}\right|_{N,E} + E \left.\frac{\partial S}{\partial E}\right|_{N,V} = S. \qquad (6.70)$$

*Show from this that in general*

$$S = (E + pV - \mu N)/T \qquad (6.71)$$

*and hence $E = TS - pV + \mu N$.* This is Euler's equation. As a state function, $S$ is supposed to depend only on $E$, $V$, and $N$. But equation 6.71 seems to show explicit dependence on $T$, $p$, and $\mu$ as well: how can this be?

*(b) One answer is to write the latter three as functions of $E$, $V$, and $N$. Do this explicitly for the ideal gas, using the ideal gas entropy equation 3.61*

$$S(N, V, E) = \frac{5}{2} N k_B + N k_B \log \left[ \frac{V}{N h^3} \left( \frac{4\pi m E}{3N} \right)^{3/2} \right], \qquad (6.72)$$

*and your (or the grader's) results for exercise 3.6(c), and verify equation 6.71 in that case.*

Another answer is to consider a small shift of all six variables. We know that $dE = TdS - pdV + \mu dN$, but if we shift all six variables in Euler's equation we get $dE = TdS - pdV + \mu dN + SdT - Vdp + Nd\mu$. This implies the Gibbs-Duhem relation

$$0 = SdT - Vdp + Nd\mu. \qquad (6.73)$$

It means that the intensive variables $T$, $p$, and $\mu$ are not all independent.

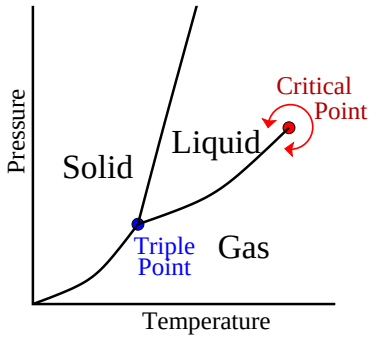



**Fig. 6.13** Generic phase diagram, showing the coexistence curves for solids, liquids, and gases.

**Clausius-Clapeyron equation.** Consider the phase diagram 6.13. Along an equilibrium phase boundary, the temperatures, pressures, and chemical potentials of the two phases must agree: otherwise a flat interface between the two phases would transmit heat, shift sideways, or leak particles, respectively (violating the assumption of equilibrium).

*(c) Apply the Gibbs-Duhem relation to both phases, for a small shift by $\Delta T$ along the phase boundary. Let $s_1$, $v_1$, $s_2$, and $v_2$ be the molecular entropies and volumes ($s = S/N$, $v = V/N$ for each phase); derive the Clausius-Clapeyron equation for the slope of the coexistence line on the phase diagram*

$$dP/dT = (s_1 - s_2)/(v_1 - v_2). \qquad (6.74)$$

It's hard to experimentally measure the entropies per particle: we don't have an entropy thermometer. But, as you will remember, the entropy difference upon a phase transformation $\Delta S = Q/T$ is related to the heat flow $Q$ needed

to induce the phase change. Let the latent heat $L$ be the heat flow per molecule.

*(d) Write a formula for $dP/dT$ not involving the entropy.*

(6.5) **Negative Temperature.** (Quantum)

A system of $N$ atoms can be in the ground state or in an excited state. For convenience, we set the zero of energy exactly in between, so the energies of the two states of an atom are $\pm\epsilon/2$. The atoms are isolated from the outside world. There are only weak couplings between the atoms, sufficient to bring them into internal equilibrium but without other effects.

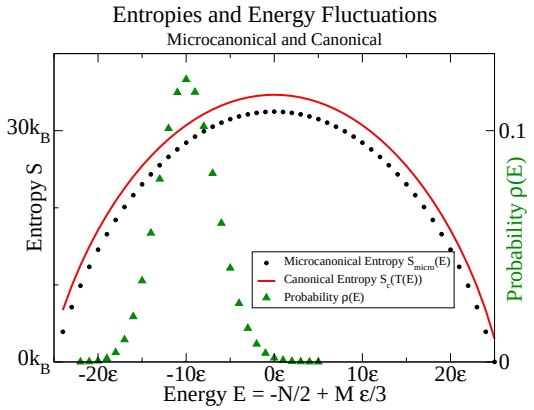



**Fig. 6.14** Entropies and energy fluctuations for this problem with $N = 50$. The canonical probability distribution for the energy is for $\langle E\rangle = -10\epsilon$, and $k_BT = 1.207\epsilon$. You may wish to check some of your answers against this plot.

*(a)* **Microcanonical Entropy.** *If the net energy is $E$ (corresponding to a number of excited atoms $m = E/\epsilon + N/2$), what is the microcanonical entropy $S_{\rm micro}(E)$ of our system? Simplify your expression using Stirling's formula, $\log n! \sim n\log n - n$.*

*(b)* **Negative Temperature.** *Find the temperature, using your simplified expression from part (a). (Why is it tricky to do it without approximation?) What happens to the temperature when $E > 0$?*

Having the energy $E > 0$ is a kind of population inversion. Population inversion is the driving mechanism for lasers. Microcanonical simulations can lead also to states with negative specific heats.

For many quantities, the thermodynamic derivatives have natural interpretations when viewed as sums over states. It's easiest to see this in small systems.

*(c)* **Canonical Ensemble: Explicit traces and thermodynamics.** *(i) Take one of our atoms and couple it to a heat bath of temperature $k_BT = 1/\beta$. Write explicit formulas for $Z_{\rm canon}$, $E_{\rm canon}$, and $S_{\rm canon}$ in the canonical ensemble, as a trace (or sum) over the two states of the atom. ($E$ should be the energy of each state multiplied by the probability $\rho_n$ of that state, $S$ should be the trace of $\rho_n \log \rho_n$.) (ii) Compare the results with what you get by using the thermodynamic relations. Using $Z$ from the trace over states, calculate the Helmholtz free energy $A$, $S$ as a derivative of $A$, and $E$ from $A = E - TS$. Do the thermodynamically derived formulas you get agree with the statistical traces? (iii) To remove some of the mystery from the thermodynamic relations, consider the thermodynamically valid formula $E = -\partial \log Z/\partial\beta = -(1/Z)\,\partial Z/\partial\beta$. Write out $Z$ as a sum over energy states, and see that this formula follows naturally.*

*(d) What happens to $E$ in the canonical ensemble as $T \to \infty$? Can you get into the regime discussed in part (b)?*

*(e)* **Canonical-Microcanonical Correspondence.** *Find the entropy in the canonical distribution for $N$ of our atoms coupled to the outside world, from your answer to part (c). How can you understand the value of $S(T = \infty) - S(T = 0)$ simply? Using the approximate form of the entropy from part (a) and the temperature from part (b), show that the canonical and microcanonical entropies agree, $S_{\rm micro}(E) = S_{\rm canon}(T(E))$. (Perhaps useful:* $\operatorname{arctanh}(x) = \frac{1}{2}\log\left((1+x)/(1-x)\right)$*.) Notice that the two are not equal in the figure above: the form of Stirling's formula we used in part (a) is not very accurate for $N = 50$. Explain in words why the microcanonical entropy is smaller than the canonical entropy.*

*(f)* **Fluctuations.** *Show in general that the root-mean-squared fluctuations in the energy in the canonical distribution $\langle (E - \langle E\rangle)^2\rangle = \langle E^2\rangle - \langle E\rangle^2$ is related to the specific heat $C = \partial E/\partial T$. (I find it helpful to use the formula from part (c.iii), $E = -\partial \log(Z)/\partial\beta$.) Calculate the root-mean-square energy fluctuations for $N$ of our atoms. Evaluate it at $T(E)$ from part (b): it should have a particularly simple form. For large $N$, are the fluctuations in $E$ small compared to $E$?*

(6.6) **Laplace.** (Thermodynamics) [39]

**Laplace Transform.** The Laplace transform of a function $f(t)$ is a function of $x$:

$$\mathcal{L}\{f\}(x) = \int_0^\infty f(t)e^{-xt}\,dt. \tag{6.75}$$

[39] Laplace (1749-1827). Math reference [71, sec. 4.3].

*Show that the canonical partition function $Z(\beta)$ can be written as the Laplace transform of the microcanonical volume of the energy shell $\Omega(E)$.*

(6.7) **Lagrange.** (Thermodynamics) [40]

**Lagrange Multipliers.** Lagrange multipliers allow one to find the extremum of a function $f(\mathbf{x})$ given a constraint $g(\mathbf{x}) = g_0$. One extremizes

$$f(\mathbf{x}) + \lambda(g(\mathbf{x}) - g_0) \tag{6.76}$$

as a function of $\lambda$ and $\mathbf{x}$. The derivative with respect to $\lambda$ being zero enforces the constraint and sets $\lambda$. The derivatives with respect to components of $\mathbf{x}$ then include terms involving $\lambda$, which act to enforce the constraint.

Let us use Lagrange multipliers to find the maximum of the nonequilibrium entropy

$$\begin{aligned} S &= -k_B \int \rho(\mathbb{P},\mathbb{Q}) \log \rho(\mathbb{P},\mathbb{Q}) \\ &= -k_B Tr(\rho \log \rho) \\ &= -k_B \sum p_i \log p_i \end{aligned} \tag{6.77}$$

constraining the normalization, energy, and number. You may use whichever form of the entropy you prefer: the first continuous form will demand some calculus of variations (see [71, ch. 12]); the last discrete form is mathematically the most straightforward.

*(a)* **Microcanonical:** *Using a Lagrange multiplier to enforce the normalization*

$$Tr(\rho) = \int_{\text{EnergySurface}} \rho(\mathbb{P},\mathbb{Q}) = 1, \tag{6.78}$$

*show that the probability distribution that extremizes the entropy is a constant (the microcanonical distribution).*

*(b)* **Canonical:** *Integrating over all $\mathbb{P}$ and $\mathbb{Q}$, use another Lagrange multiplier to fix the mean energy $\langle E \rangle = \int d\mathbb{P}d\mathbb{Q}\,\mathcal{H}(\mathbb{P},\mathbb{Q})\rho(\mathbb{P},\mathbb{Q})$. Show that the canonical distribution maximizes the entropy given the constraints of normalization and fixed energy.*

*(c)* **Grand Canonical:** *Summing over different numbers of particles $N$ and adding the constraint that the average number is $\langle N \rangle = \sum_N \int d\mathbb{P}d\mathbb{Q}\,N\rho_N(\mathbb{P},\mathbb{Q})$, show that you get the grand canonical distribution by maximizing the entropy.*

(6.8) **Legendre.** (Thermodynamics) [41]

**Legendre Transforms.** The Legendre transform of a function $f(t)$ is given by minimizing $f(x) - xp$ with respect to $p$, so that $p$ is the slope ($p = \frac{\partial f}{\partial x}$):

$$g(p) = \min_x \{f(x) - xp\}. \tag{6.79}$$

We saw in the text that in thermodynamics the Legendre transform of the energy is the Helmholtz free energy[42]

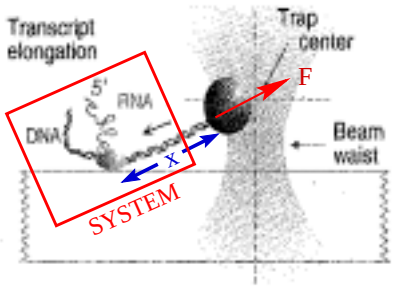
$$A(T,N,V) = \min_E \{E(S,V,N) - TS\}. \tag{6.80}$$

How do we connect this with the statistical mechanical relation of part (a), which related $\Omega = \exp(S/k_B)$ to $Z = \exp(-A/k_BT)$? Thermodynamics, roughly speaking, is statistical mechanics without the fluctuations.

*Using your Laplace transform of exercise 6.6, find an equation for $E_{max}$ where the integrand is maximized. Does this energy equal the energy which minimizes the Legendre transform 6.80? Approximate $Z(\beta)$ in your Laplace transform by the value of the integrand at this maximum (ignoring the fluctuations). Does it give the Legendre transform relation 6.80?*

(6.9) **Molecular Motors: Which Free Energy?** (Basic, Biology)

Figure 6.15 shows a study of the molecular motor that transcribes DNA into RNA. Choosing a good ensemble for this system is a bit involved. It is under two constant forces ($F$ and pressure), and involves complicated chemistry and biology. Nonetheless, you know some things based on fundamental principles. Let us consider the optical trap and the distant fluid as being part of the external environment, and define the "system" as the local region of DNA, the RNA, motor, and the fluid and local molecules in a region immediately enclosing the region, as shown in figure 6.15.




[40] Lagrange (1736-1813).
[41] Legendre (1752-1833).
[42] Actually, [6.40] in the text had $E$ as the independent variable. As usual in thermodynamics, we can solve $S(E,V,N)$ for $E(S,V,N)$.

**Fig. 6.15** An RNA polymerase molecular motor attached to a glass slide is pulling along a DNA molecule (transcribing it into RNA). The opposite end of the DNA molecule is attached to a bead which is being pulled by an optical trap with a constant external force $F$. Let the distance from the motor to the bead be $x$: thus the motor is trying to move to decrease $x$ and the force is trying to increase $x$.

*Without knowing anything further about the chemistry or biology in the system, which two of the following must be true on average, in all cases, according to basic laws of thermodynamics?*

*(T) (F) The total entropy of the universe (the system, bead, trap, laser beam ...) must increase or stay unchanged with time.*

*(T) (F) The entropy $S_s$ of the system must increase with time.*

*(T) (F) The total energy $E_T$ of the universe must decrease with time.*

*(T) (F) The energy $E_s$ of the system must decrease with time.*

*(T) (F) $G_s - Fx = E_s - TS_s + PV_s - Fx$ must decrease with time, where $G_s$ is the Gibbs free energy of the system.* Related formula: $G = E - TS + PV$.

Note: $F$ is a force, not the Helmholtz free energy. Precisely two of the answers are correct.

(6.10) **Michaelis-Menten and Hill** (Biology, Computation)

Biological systems often have reaction rates that are saturable: the cell needs to respond sensitively to the introduction of a new chemical S, but the response should not keep growing indefinitely as the new chemical concentration $[S]$ grows.[43] Other biological systems act as switches: they not only saturate, but they change sharply from one state to another as the concentration of a chemical S is varied. We shall analyze both of these important biological problems, and at the same time give tangible examples of how one develops effective dynamical theories by removing degrees of freedom: here, we remove an enzyme E from the equations to get an effective reaction rate, rather than coarse–graining some large statistical mechanical system.

The rate of a chemical reaction

$$NS + B \to C \tag{6.81}$$

where $N$ molecules of type S combine with a molecule of type B to make a molecule of type C will occur with a reaction rate given by a traditional chemical kinetics formula:

$$\frac{d[C]}{dt} = k[S]^N[B]. \tag{6.82}$$

If the reactants need all to be in a small volume $V$ in order to react, then $[S]^N[B]V^N$ is the probability that they are in location to proceed, and the rate constant $k$ divided by $V^N$ is the reaction rate of the confined molecules.[44]

**Saturation: the Michaelis–Menten equation.** Saturation is not seen in ordinary chemical reaction kinetics. Notice that the reaction rate goes as the $N^{\text{th}}$ power of the concentration $[S]$: far from saturating, the reaction rate grows linearly or faster with concentration.

The prototype example of saturation in biological systems is the *Michaelis–Menten* reaction form. A reaction of this form converting a chemical S (the substrate) into P (the product) has a rate given by the formula

$$\frac{d[P]}{dt} = \frac{v_{\mathbf{max}}[S]}{K_M + [S]}, \tag{6.83}$$

where $K_M$ is called the Michaelis constant (figure 6.16). This reaction at small concentrations acts like an ordinary chemical reaction with $N = 1$ and $k = v_{\mathbf{max}}/K_M$, but the rate saturates at $V_{\mathbf{max}}$ as $[S] \to \infty$. The Michaelis constant $K_M$ is the concentration $[S]$ at which the rate is equal to half of its saturation rate.

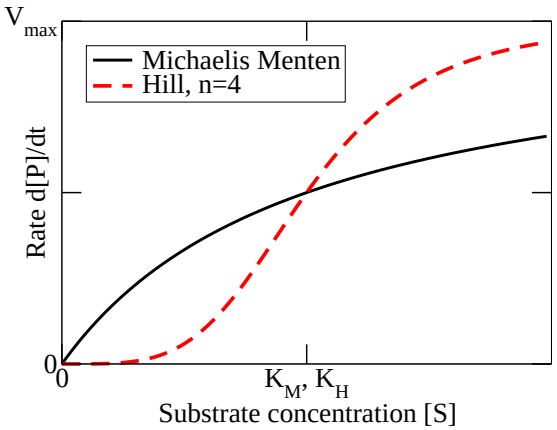



**Fig. 6.16 Michaelis–Menten and Hill equation forms.**

[43] $[S]$ is the concentration of $S$ (number per unit volume). S stands for *substrate*.

[44] The reaction will typically involve crossing an energy barrier $E$, and the rate will be given by a Boltzmann probability $k = k_0 \exp(-E/k_BT)$. The constant of proportionality $k_0$ can in principle be calculated using generalizations of the methods we used in exercise 6.2.

We can derive the Michaelis–Menten form by hypothesizing the existence of a catalyst or enzyme $E$, which is in short supply. The enzyme is presumed to be partly free and available for binding (concentration $[E]$) and partly bound to the substrate (concentration $[E:S]$, the colon denoting the dimer), helping it to turn into the product. The total concentration $[E]+[E:S]=E_{\rm tot}$ is fixed. The reactions are as follows:

$$E+S \quad \underset{k_1}{\overset{k_{-1}}{\leftrightarrows}} \quad E:S \quad \overset{k_{cat}}{\rightarrow} \quad E+P \tag{6.84}$$

We must then assume that the supply of substrate is large, so its concentration changes slowly with time. We can then assume that the concentration $[E:S]$ is in steady state, and remove it as a degree of freedom.

*(a) Assume the binding reaction (rates $k_1$, $k_{-1}$, and $k_{cat}$) in equation 6.84 are of traditional chemical kinetics form (equation 6.82), with $N=1$ or $N=0$ as appropriate. Write the equation for $d[E:S]/dt$, set it to zero, and use it to eliminate $[E]$ in the equation for $dP/dt$. What are $v_{\mathbf{max}}$ and $K_M$ in the Michaelis-Menten form (equation 6.83) in terms of the ks and $E_{\rm tot}$?*

We can understand this saturation intuitively: when all the enzyme is busy and bound to the substrate, adding more substrate can't speed up the reaction.

**Cooperativity and sharp switching: the Hill equation.** Hemoglobin is what makes blood red: this iron–containing protein can bind up to four molecules of oxygen in the lungs, and carries them to the tissues of the body where it releases them. If the binding of all four oxygens were independent, the $[O_2]$ concentration dependence of the bound oxygen concentration would have the Michaelis–Menten form (figure 6.16): to completely deoxygenate the Hemoglobin (Hb) would demand a very low oxygen concentration in the tissue.

What happens instead is that the Hb binding of oxygen looks much more sigmoidal – a fairly sharp transition between nearly 4 oxygens bound at high $[O_2]$ (lungs) to nearly none bound at low oxygen concentrations. This arises because the binding of the oxygens is enhanced by having other oxygens bound. This is not because the oxygens somehow stick to one another: instead, each oxygen deforms the Hb in a non–local *allosteric*[45] fashion, changing the configurations and and affinity of the other binding sites.

The Hill equation was introduced for hemoglobin to describe this kind of cooperative binding. Like the Michaelis–Menten form, it is also used to describe reaction rates, where instead of the carrier Hb we have an enzyme, or perhaps a series of transcription binding sites (see exercise 8.7). In the reaction rate form, the Hill equation is

$$\frac{d[P]}{dt} = \frac{v_{\mathbf{max}}[S]^n}{K_H^n + [S]^n}, \tag{6.85}$$

(see figure 6.16). For Hb, the concentration of the $n$-fold oxygenated form is given by the right-hand side of equation 6.85. In both cases, the transition becomes much more of a switch, with the reaction turning on (or the Hb accepting or releasing its oxygen) sharply at a particular concentration (figure 6.16). The transition can be made more or less sharp by increasing or decreasing $n$.

The Hill equation can be derived using a simplifying assumption that $n$ molecules bind in a single reaction:

$$E+nS \quad \underset{k_b}{\overset{k_u}{\leftrightarrows}} \quad E:(nS) \tag{6.86}$$

where E might stand for hemoglobin and S for the $O_2$ oxygen molecules. Again, there is a fixed total amount $E_{\rm tot}=[E]+[E:nS]$.

*(b) Assume that the two reactions in equation 6.86 have the chemical kinetics form (equation 6.82) with $N=0$ or $N=n$ as appropriate. Write the equilibrium equation for $E:(nS)$, and eliminate $[E]$ using the fixed total $E_{\rm tot}$.*

Usually, and in particular for hemoglobin, this cooperativity is not so rigid: the states with one, two, and three $O_2$ molecules bound also compete with the unbound and fully bound states. This is treated in an approximate way by using the Hill equation, but allowing $n$ to vary as a fitting parameter: for Hb, $n\approx 2.8$.

Both Hill and Michaelis–Menten equations are often used in biological reaction models even when there are no explicit mechanisms (enzymes, cooperative binding) known to generate them.

## (6.11) Pollen and Hard Squares.

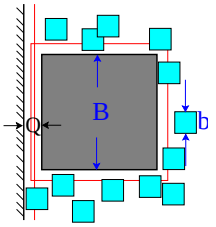



**Fig. 6.17** Square pollen grain in fluid of oriented square molecules, next to a wall. The thin lines represents the exclusion region around the pollen grain and away from the wall.

[45] Allosteric comes from *Allo* (other) and *steric* (structure or space). Allosteric interactions can be cooperative, as in hemoglobin, or inhibitory.

We model the entropic attraction between a pollen grain and a wall using a two-dimensional ideal gas of classical indistinguishable particles as the fluid. For convenience, we imagine that the pollen grain and the fluid are formed from square particles lined up with the axes of the box, of length $B$ and $b$, respectively. We assume *no interaction* between the ideal gas molecules (unlike in exercise 3.3); the potential energy is infinite, though, if the gas molecules overlap with the pollen grain or with the wall. The container as a whole has one pollen grain, $N$ gas molecules, and total area $L \times L$.

Assume the pollen grain is close to only one wall. Let the distance from the surface of the wall to the closest face of the pollen grain be $Q$. (A similar square-particle problem with interacting small molecules was studied in reference [33].)

*(a) What is the area $A(Q \gg 0)$ available for the gas molecules, in units of (length)$^2$, when the pollen grain is far from the wall? What is the overlap of the excluded regions, $A(0) - A(\infty)$, when the pollen grain touches the wall, $Q = 0$? Give formulas for $A(Q)$ as a function of $Q$ for the two relevant regions, $Q < b$ and $Q > b$.*

*(b) What is the configuration-space volume $\Omega(Q)$ for the gas, in units of (length)$^{2N}$? What is the configurational entropy of the ideal gas, $S(Q)$?* (Write your answers here in terms of $A(Q)$ to simplify grading.)

Your answers to part (b) can be viewed as giving a free energy for the pollen grain after integrating over the gas degrees of freedom (also known as a partial trace, or coarse-grained free energy).

*(c) What is the resulting coarse-grained free energy of the pollen grain, $\mathcal{F}(Q) = E - T\,S(Q)$, in the two regions $Q > b$ and $Q < b$? Use $\mathcal{F}(Q)$ to calculate the force on the pollen grain for $Q < b$. Is the force positive (away from the wall) or negative? Why?*

*(d) Directly calculate the force due to the ideal-gas pressure on the far side of the pollen grain, in terms of $A(Q)$. Compare it to the force from the partial trace in part (c). Why is there no balancing force from the other side? Effectively how "long" is the far side of the pollen grain?*

# Quantum Statistical Mechanics

7

Quantum statistical mechanics governs most of solid-state physics (metals, semiconductors, and glasses) and parts of molecular physics and astrophysics (white dwarves, neutron stars). It spawned the origin of quantum mechanics (Planck's theory of the black-body spectrum) and provides the framework for our understanding of other exotic quantum phenomena (Bose condensation, superfluids, and superconductors). Applications of quantum statistical mechanics are significant components of courses in these various subjects. We condense our treatment of this important subject into this one chapter in order to avoid overlap with other physics and chemistry courses, and also in order to keep our treatment otherwise accessible to outsiders uninitiated into the mysteries of quantum mechanics.

In this section, we will proceed from the abstract to the concrete, through a series of simplifications. We begin (section 7.1) by introducing *mixed states* for quantum ensembles, and the advanced topic of *density matrices* (for non-equilibrium quantum systems which are not mixtures of energy eigenstates). We illustrate mixed states in section 7.2 by solving the finite-temperature quantum harmonic oscillator. We discuss the statistical mechanics of identical particles (section 7.3). We then make the vast simplification of presuming that the particles are non-interacting (section 7.4), which leads us to the Bose-Einstein and Fermi distributions for the filling of single-particle eigenstates. We briefly contrast the Bose, Fermi, and Maxwell-Boltzmann statistics 7.5. We illustrate how amazingly useful the non-interacting particle picture is for quantum systems by solving the classic problems of black-body radiation and bose condensation (section 7.6), and for the behavior of metals (section 7.7).

## 7.1 Mixed States and Density Matrices

Classical statistical ensembles are probability distributions $\rho(\mathbb{P}, \mathbb{Q})$ in phase space. How do we generalize them to quantum mechanics? Two problems immediately arise. First, the Heisenberg uncertainty principle tells us that one cannot specify both position and momentum for a quantum system at the same time. The states of our quantum system will not be points in phase space. Second, quantum mechanics already has probability densities: even for systems in a definite state[1] $\Psi(\mathbb{Q})$

[1] Quantum systems with many particles have wavefunctions that are functions of all the positions of all the particles (or, in momentum space, all the momenta of all the particles).

the probability is spread among different configurations $|\Psi(\mathbb{Q})|^2$ (or momenta $|\widetilde{\Psi}(\mathbb{P})|^2$). In statistical mechanics, we need to introduce a second level of probability, to discuss an ensemble that has probabilities $p_n$ of being in a variety of quantum states $\Psi_n(\mathbb{Q})$. Ensembles in quantum mechanics are called *mixed states*: they are not superpositions of different wave functions, but incoherent mixtures.[2]

Suppose we want to compute the ensemble expectation of an operator $\mathbf{A}$. In a particular state $\Psi_n$, the quantum expectation is

$$\langle \mathbf{A} \rangle_{\text{pure}} = \int \Psi_n^*(\mathbb{Q}) \mathbf{A} \Psi_n(\mathbb{Q}) d^{3N}\mathbb{Q}. \tag{7.1}$$

So, in the ensemble the expectation is

$$\langle \mathbf{A} \rangle = \sum_n p_n \int \Psi_n^*(\mathbb{Q}) \mathbf{A} \Psi_n(\mathbb{Q}) d^{3N}\mathbb{Q}. \tag{7.2}$$

For most purposes, this is enough! Except for selected exercises in this chapter, one or two exercises in the rest of the book, and occasional specialized seminars, formulating the ensemble as a sum over states $\Psi_n$ with probabilities $p_n$ is perfectly satisfactory. Indeed, for all of the equilibrium ensembles, the $\Psi_n$ may be taken to be the energy eigenstates, and the $p_n$ either a constant in a small energy range (for the microcanonical ensemble), or $\exp(-\beta E_n)/Z$ (for the canonical ensemble), or $\exp\left(-\beta(E_n - N_n\mu)\right)/\Xi$ (for the grand canonical ensemble). For most practical purposes you may stop reading this section here, and proceed to the quantum harmonic oscillator.

**Advanced Topic: Density Matrices.** What do we gain from going beyond this picture? First, there are lots of mixed states that are not mixtures of energy eigenstates. Mixtures of energy eigenstates have time-independent properties, so any time-dependent non-equilibrium ensemble will be in this class. Second, although one can define the ensemble in terms of a set of states $\Psi_n$, the ensemble should be something one can look at in a variety of bases. Indeed, superfluids and superconductors show an energy gap when viewed in the energy basis, but show an exotic *off-diagonal long-range order* when looked at in position space (exercise 9.7). Third, we will see that the proper generalization of Liouville's theorem demands the more elegant, operator-based approach.

Our goal is to avoid carrying around the particular states $\Psi_n$. Instead, we will write the ensemble average (equation 7.2) in terms of $\mathbf{A}$ and an operator $\boldsymbol{\rho}$, the *density matrix*. For this section, it is convenient to use Dirac's bra-ket notation, in which the ensemble average can be written[3]

$$\langle \mathbf{A} \rangle = \sum_n p_n \langle \Psi_n | \mathbf{A} | \Psi_n \rangle. \tag{7.3}$$

Pick any complete orthonormal basis $\Phi_\alpha$. Then the identity operator

$$\mathbf{1} = \sum_\alpha |\Phi_\alpha\rangle\langle\Phi_\alpha| \tag{7.4}$$

[2] So, for example, if $|R\rangle$ is a right-circularly polarized photon, and $|L\rangle$ is a left-circularly polarized photon, then the superposition $\frac{1}{\sqrt{2}}(|R\rangle + |L\rangle)$ is a linearly polarized photon, while the mixture $\frac{1}{2}(|R\rangle\langle R| + |L\rangle\langle L|)$ is an unpolarized photon. The superposition is in *both* states, the mixture is in perhaps one or perhaps the other. See exercise 7.5(a).

[3] In Dirac's notation, $\langle\Psi|\mathbf{M}|\Phi\rangle = \int \Psi^* \mathbf{M} \Phi$. It is particularly useful when expressing operators in a basis $\Psi_m$; if the matrix elements are $M_{ij} = \langle\Psi_i|\mathbf{M}|\Psi_j\rangle$ then the operator itself can be written $\mathbf{M} = \sum_{ij} M_{ij}|\Psi_i\rangle\langle\Psi_j|$.

and, plugging the identity (equation 7.4) into equation 7.3 we find

$$\begin{aligned}\langle \mathbf{A} \rangle =& \sum_n p_n \langle \Psi_n | \left( \sum_\alpha |\Phi_\alpha\rangle\langle|\Phi_\alpha| \right) \mathbf{A} |\Psi_n\rangle \\ =& \sum_n p_n \sum_\alpha \langle \Phi_\alpha | \mathbf{A} \Psi_n \rangle \langle \Psi_n | \Phi_\alpha \rangle \\ =& \sum_\alpha \langle \Phi_\alpha \mathbf{A}| \left( \sum_n p_n |\Psi_n\rangle\langle\Psi_n| \right) |\Phi_\alpha\rangle \\ =& \mathrm{Tr}(\mathbf{A}\boldsymbol{\rho}) \end{aligned} \tag{7.5}$$

where[4]

$$\boldsymbol{\rho} = \left( \sum_n p_n |\Psi_n\rangle\langle\Psi_n| \right) \tag{7.6}$$

is the *density matrix*.

Some conclusions we can draw about the density matrix:

**Sufficiency.** In quantum mechanics, all measurement processes involve expectation values of operators. Our density matrix therefore suffices to embody everything we need to know about our quantum system.

**Pure states.** A pure state, with a definite wavefunction $\Phi$, has $\boldsymbol{\rho}_{\rm pure} = |\Phi\rangle\langle\Phi|$. In the position basis $|\mathbb{Q}\rangle$, this pure-state density matrix has matrix elements $\boldsymbol{\rho}_{\rm pure}(\mathbb{Q},\mathbb{Q}') = \langle\mathbb{Q}|\boldsymbol{\rho}_{\rm pure}|\mathbb{Q}'\rangle = \Phi^*(\mathbb{Q}')\Phi(\mathbb{Q})$. Thus in particular we can reconstruct[5] the wavefunction from a pure-state density matrix, up to an overall physically unmeasurable phase. Since our wavefunction is normalized $\langle\Phi|\Phi\rangle = 1$, we note also that the square of the density matrix for a pure state equals itself: ${\boldsymbol{\rho}_{\rm pure}}^2 = (|\Phi\rangle\langle\Phi|)(|\Phi\rangle\langle\Phi|) = \boldsymbol{\rho}_{\rm pure}$,

**Normalization.** The trace of a pure-state density matrix $\mathrm{Tr}\boldsymbol{\rho}_{\rm pure} = 1$, since we can pick an orthonormal basis with our wavefunction $\Phi$ as the first basis element, making the first term in the trace sum one and the others zero. The trace of a general density matrix is hence also one, since it is a probability distribution of pure-state density matrices:

$$\mathrm{Tr}\boldsymbol{\rho} = \mathrm{Tr}\left( \sum_n p_n |\Psi_n\rangle\langle\Psi_n| \right) = \sum_n p_n \mathrm{Tr}\left( |\Psi_n\rangle\langle\Psi_n| \right) = \sum_n p_n = 1. \tag{7.8}$$

**Canonical Distribution.** The canonical distribution can be written in terms of the Hamiltonian operator $\mathcal{H}$ as[6]

$$\boldsymbol{\rho}_{\mathrm{C}anon} = \frac{\exp(-\beta\mathcal{H})}{Z} = \frac{\exp(-\beta\mathcal{H})}{\mathrm{Tr}\exp(-\beta\mathcal{H})}. \tag{7.10}$$

Let $|E_n\rangle$ be the orthonormal many-body energy eigenstates. If we eval-

[4] The trace of a matrix is the sum of its diagonal elements, and is independent of what basis you write it in. The same is true of operators: we are summing the diagonal elements $\mathrm{Tr}(M) = \sum_\alpha \langle\Phi_\alpha|M|\Phi_\alpha\rangle$.

[5] In particular, since $\Phi$ is normalized $|\Phi^*(\mathbb{Q}')|^2 = \int d\mathbb{Q} |\rho(\mathbb{Q},\mathbb{Q}')|^2$ and thus

$$\Phi(\mathbb{Q}) = \frac{\rho(\mathbb{Q},\mathbb{Q}')}{\sqrt{\int d\widetilde{\mathbb{Q}} |\rho(\widetilde{\mathbb{Q}},\mathbb{Q}')|^2}} \tag{7.7}$$

up to the phase of $\phi^*(\mathbb{Q}')$.

[6] What is the exponential of a matrix $M$? We can define it in terms of a power series, $\exp(M) = \mathbf{1} + M + M^2/2! + M^3/3! + \ldots$, but it is usually easier to change basis to diagonalize $M$. In that basis, any function $f(M)$ is given by

$$f(\boldsymbol{\rho}) = \begin{pmatrix} f(\boldsymbol{\rho}_{11}) & 0 & 0 & \ldots \\ 0 & f(\boldsymbol{\rho}_{22}) & 0 & \ldots \\ \ldots & & & \end{pmatrix}. \tag{7.9}$$

At the end, change back to the original basis.

uate $\boldsymbol{\rho}_{\mathrm{C}anon}$ in the energy basis,

$$\begin{aligned}\rho_{mn} &= \langle E_m|\boldsymbol{\rho}_{\mathrm{C}anon}|E_n\rangle \\ &= \langle E_m|\Big(e^{-\beta\mathcal{H}}|E_n\rangle\Big)/Z \\ &= \langle E_m|\Big(e^{-\beta E_n}|E_n\rangle\Big)/Z \\ &= e^{-\beta E_n}\langle E_m|E_n\rangle/Z \\ &= e^{-\beta E_n}\delta_{mn}/Z \end{aligned} \tag{7.11}$$

so $\boldsymbol{\rho}_{\mathrm{C}anon}$ is diagonal in the energy basis

$$\boldsymbol{\rho} = \sum_n \frac{\exp(-\beta E_n)}{Z}|E_n\rangle\langle E_n| \tag{7.12}$$

and is given by the canonical weighting of each of the energy eigenstates, just as one would expect. Notice that the states $\Psi_n$ mixed to make the density matrix need not in general be eigenstates, or even orthogonal. For equilibrium statistical mechanics, though, life is simpler: the $\Psi_n$ can be chosen to be energy eigenstates, and the density matrix is diagonal in that basis.

**Entropy.** The entropy for a general density matrix will be

$$S = -k_B \mathrm{Tr}\left(\boldsymbol{\rho}\log\boldsymbol{\rho}\right). \tag{7.13}$$

**Time evolution for the density matrix.** The time evolution for the density matrix is determined by the time evolution of the pure states composing it:[7]

$$\frac{\partial\boldsymbol{\rho}}{\partial t} = \sum_n p_n \left(\frac{\partial|\Psi_n\rangle}{\partial t}\langle\Psi_n| + |\Psi_n\rangle\frac{\partial\langle\Psi_n|}{\partial t}\right). \tag{7.14}$$

[7] The $p_n$ are the probability that one *started* in the state $\Psi_n$, and thus clearly don't change with time.

Now, the time evolution of the 'ket' wavefunction $|\Psi_n\rangle$ is given by operating on it with the Hamiltonian

$$\frac{\partial|\Psi_n\rangle}{\partial t} = \frac{1}{i\hbar}\mathcal{H}|\Psi_n\rangle \tag{7.15}$$

and the time evolution of the 'bra' wavefunction $\langle\Psi_n|$ is given by the time evolution of $\Psi_n^*(\mathbb{Q})$:

$$\frac{\partial\Psi_n^*}{\partial t} = \left(\frac{\partial\Psi_n}{\partial t}\right)^* = \left(\frac{1}{i\hbar}\mathcal{H}\Psi_n\right)^* = -\frac{1}{i\hbar}\mathcal{H}\Psi_n^* \tag{7.16}$$

so since $\mathcal{H}$ is Hermitian,

$$\frac{\partial\langle\Psi_n|}{\partial t} = -\frac{1}{i\hbar}\langle\Psi_n|\mathcal{H}. \tag{7.17}$$

Hence[8]

$$\begin{aligned}\frac{\partial\boldsymbol{\rho}}{\partial t} &= \sum_n p_n \frac{1}{i\hbar}\Big(\mathcal{H}|\Psi_n\rangle\langle\Psi_n| - |\Psi_n\rangle\langle\Psi_n|\mathcal{H}\Big) = \frac{1}{i\hbar}\left(\mathcal{H}\boldsymbol{\rho} - \boldsymbol{\rho}\mathcal{H}\right) \\ &= \frac{1}{i\hbar}[\mathcal{H},\boldsymbol{\rho}]. \end{aligned} \tag{7.18}$$

[8] Notice that this is *minus* the formula one uses for the time-evolution of operators in the Heisenberg representation.

**Quantum Liouville Equation.** This time evolution law is the quantum version of Liouville's theorem. We can see this by using the equations of motion 4.1, $\dot{q}_\alpha = \partial\mathcal{H}/\partial p_\alpha$ and $\dot{p}_\alpha = -\partial\mathcal{H}/\partial q_\alpha$ and the definition of Poisson brackets

$$\{A, B\}_P = \sum_\alpha \frac{\partial A}{\partial q_\alpha}\frac{\partial B}{\partial p_\alpha} - \frac{\partial A}{\partial p_\alpha}\frac{\partial B}{\partial q_\alpha} \tag{7.19}$$

to rewrite Liouville's theorem that the total time derivative is zero (equation 4.7) into a statement about the partial time derivative:

$$0 = \frac{d\rho}{dt} = \frac{\partial\rho}{\partial t} + \sum_\alpha \frac{\partial\rho}{\partial q_\alpha}\dot{q}_\alpha + \frac{\partial\rho}{\partial p_\alpha}\dot{p}_\alpha = \frac{\partial\rho}{\partial t} + \sum_\alpha \left(\frac{\partial\rho}{\partial q_\alpha}\frac{\partial\mathcal{H}}{\partial p_\alpha} - \frac{\partial\rho}{\partial p_\alpha}\frac{\partial\mathcal{H}}{\partial q_\alpha}\right) \tag{7.20}$$

so

$$\frac{\partial\rho}{\partial t} = \{\mathcal{H}, \boldsymbol{\rho}\}_P. \tag{7.21}$$

Using the classical↔quantum correspondence between the Poisson brackets and the commutator $\{\ \}_P \leftrightarrow \frac{1}{i\hbar}[\ ]$ the time evolution law 7.18 is precisely the analogue of Liouville's theorem 7.21.

**Quantum Liouville and Statistical Mechanics.** The quantum version of Liouville's equation is not nearly as compelling an argument for statistical mechanics as was the classical version.

The classical theorem, you remember, stated that $d\rho/dt = 0$. Any equilibrium state must be time independent $\partial\rho/\partial t = 0$, so this implied that such a state must have $\rho$ constant along the trajectories. If the trajectory covers the energy surface (ergodicity), then the probability density had to be constant on the energy surface, justifying the microcanonical ensemble.

For an isolated quantum system, this argument breaks down. The condition that an equilibrium state must be time independent isn't very stringent! Indeed, $\partial\boldsymbol{\rho}/\partial t = [\mathcal{H}, \boldsymbol{\rho}] = 0$ for any mixture of many-body energy eigenstates! In principle, isolated quantum systems are very nonergodic, and one must couple them to the outside world to induce transitions between the many-body eigenstates and proceed to equilibrium.

This may seem less of a concern when one realizes just how peculiar a many-body eigenstates of a large system really is. Consider an atom in an excited state contained in a large box. We normally think of the atom as being in an energy eigenstate, which decays after some time into a ground state atom plus some photons. Clearly, the atom was only in an approximate eigenstate (or it would not decay): it is in a *resonance* that is an eigenstate if we ignore the coupling to the electromagnetic field. The true many-body eigenstates of the system are weird delicate superpositions of states with photons being absorbed by the atom and the atom emitting photons, carefully crafted to produce a stationary state. When one starts including more atoms and other interactions, the true many-body eigenstates are pretty useless things in most cases.[9] Tiny interactions with the outside world disrupt these many-body eigenstates, and usually lead efficiently to equilibrium.

[9] The many-body ground state and the lowest many-body excitations with given quantum numbers above the ground state are important: these states don't suffer from decays.

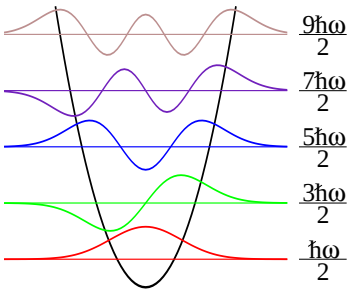



**Fig. 7.1** The quantum states of the harmonic oscillator are at equally spaced energies.

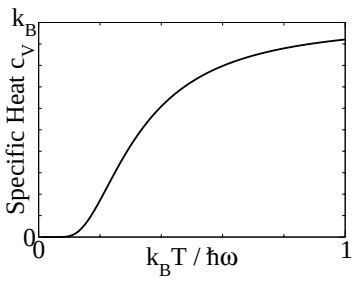



**Fig. 7.2** The specific heat for the quantum harmonic oscillator.

## 7.2 Quantum Harmonic Oscillator

The quantum harmonic oscillator is a great example of how statistical mechanics works in quantum systems. Consider a harmonic oscillator of frequency $\omega$. The energy eigenvalues are $E_n = (n + \frac{1}{2})\hbar\omega$ (figure 7.1). Hence its partition function is a geometric series $\sum x^n$, which we can sum to $1/(1-x)$:

$$\begin{aligned} Z_{\text{qho}} &= \sum_{n=0}^{\infty} e^{-\beta E_n} = \sum_{n=0}^{\infty} e^{-\beta\hbar\omega(n+\frac{1}{2})} \\ &= e^{-\beta\hbar\omega/2} \sum_{n=0}^{\infty} \left(e^{-\beta\hbar\omega}\right)^n = e^{-\beta\hbar\omega/2} \frac{1}{1-e^{-\beta\hbar\omega}} \\ &= \frac{1}{e^{\beta\hbar\omega/2} - e^{-\beta\hbar\omega/2}} = \frac{1}{2\sinh(\beta\hbar\omega/2)}. \end{aligned} \tag{7.22}$$

The average energy is

$$\begin{aligned} \langle E \rangle_{\text{qho}} &= -\frac{\partial \log Z_{\text{qho}}}{\partial\beta} = \frac{\partial}{\partial\beta}\left[\tfrac{1}{2}\beta\hbar\omega + \log\left(1 - e^{-\beta\hbar\omega}\right)\right] \\ &= \hbar\omega\left(\tfrac{1}{2} + \frac{e^{-\beta\hbar\omega}}{1-e^{-\beta\hbar\omega}}\right) = \hbar\omega\left(\tfrac{1}{2} + \frac{1}{e^{\beta\hbar\omega}-1}\right) \end{aligned} \tag{7.23}$$

which corresponds to an average excitation level

$$\langle n \rangle_{\text{qho}} = \frac{1}{e^{\beta\hbar\omega}-1}. \tag{7.24}$$

The specific heat is thus

$$c_V = \frac{\partial E}{\partial T} = k_B \left(\frac{\hbar\omega}{k_B T}\right)^2 \frac{e^{-\hbar\omega/k_B T}}{\left(1 - e^{-\hbar\omega/k_B T}\right)^2} \tag{7.25}$$

At high temperatures, $e^{-\hbar\omega/k_B T} \approx 1 - \hbar\omega/k_B T$, so $c_V \to k_B$ as we found for the classical harmonic oscillator (and as given by the equipartition theorem).

At low temperatures, $e^{-\hbar\omega/k_B T}$ becomes exponentially small, so the specific heat goes rapidly to zero as the energy asymptotes to the zero-point energy $\frac{1}{2}\hbar\omega$. More specifically, there is an energy gap[10] $\hbar\omega$ to the first excitation, so the probability of having any excitation of the system is suppressed by a factor of $e^{-\hbar\omega/k_B T}$.

## 7.3 Bose and Fermi Statistics

In quantum mechanics, indistinguishable particles are not just hard to tell apart – their quantum wavefunctions must be the same, up to an overall phase change,[11] when the coordinates are swapped (see fig-

[10] In solid state physics we call this the *energy gap*: the minimum energy needed to add an excitation to the system. In quantum field theory, where the excitations are particles, we call it the *mass* of the particle $mc^2$.

[11] In three dimensions, this phase change must be $\pm 1$. In two dimensions one can have any phase change, so one can have not only fermions and bosons but *anyons*. Anyons, with fractional statistics, arise as excitations in the fractional quantized Hall effect.

ure 7.3). In particular, for bosons[12] the wavefunction is unchanged under a swap, so

$$\Psi(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N) = \Psi(\mathbf{r}_2, \mathbf{r}_1, \ldots, \mathbf{r}_N) = \Psi(\mathbf{r}_{P_1}, \mathbf{r}_{P_2}, \ldots, \mathbf{r}_{P_N}) \quad (7.26)$$

for any permutation $P$ of the integers $1, \ldots, N$.[13] For fermions[14]

$$\Psi(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N) = -\Psi(\mathbf{r}_2, \mathbf{r}_1, \ldots, \mathbf{r}_N) = \sigma(P)\,\Psi(\mathbf{r}_{P_1}, \mathbf{r}_{P_2}, \ldots, \mathbf{r}_{P_N}). \quad (7.27)$$

The eigenstates for systems of identical fermions and bosons are a subset of the eigenstates of distinguishable particles with the same Hamiltonian

$$\mathcal{H}\Psi_n = E_n \Psi_n; \quad (7.28)$$

in particular, they are given by the distinguishable eigenstates which obey the proper symmetry properties under permutations. A non- symmetric eigenstate $\Phi$ with energy $E$ may be symmetrized to form a bose eigenstate

$$\Psi_{sym}(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N) = (\text{Normalization}) \sum_P \Phi(\mathbf{r}_{P_1}, \mathbf{r}_{P_2}, \ldots, \mathbf{r}_{P_N}) \quad (7.29)$$

or antisymmetrized to form a fermion eigenstate

$$\Psi_{asym}(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N) = (\text{Normalization}) \sum_P \sigma(P)\Phi(\mathbf{r}_{P_1}, \mathbf{r}_{P_2}, \ldots, \mathbf{r}_{P_N}) \quad (7.30)$$

if the symmetrization or antisymmetrization does not make the sum zero. These remain eigenstates of energy $E$, because they are sums of eigenstates of energy $E$.

Quantum statistical mechanics for identical particles is given by restricting the ensembles to sum over symmetric wavefunctions for bosons (or antisymmetric wavefunctions for fermions). So, for example, the partition function for the canonical ensemble is still

$$Z = \mathrm{Tr}\left(e^{-\beta H}\right) = \sum_n e^{-\beta E_n} \quad (7.31)$$

but now the trace is over a complete set of many-body symmetric (antisymmetric) states, and the sum is over the symmetric (antisymmetric) many-body energy eigenstates.

## 7.4 Non-Interacting Bosons and Fermions

Many-body quantum statistical mechanics is hard. We now make a huge approximation: we'll assume our quantum particles do not interact

[13] A permutation $\{P_1, P_2, \ldots P_N\}$ is just a reordering of the integers $\{1, 2, \ldots N\}$. The sign $\sigma(P)$ of a permutation is $+1$ if $P$ is an even permutation, and $-1$ if $P$ is an odd permutation. Swapping two labels, keeping all the rest unchanged, is an odd permutation. One can show that composing two permutations multiplies their signs, so odd permutations can be made by odd numbers of pair swaps, and even permutations are composed of even numbers of pair swaps.

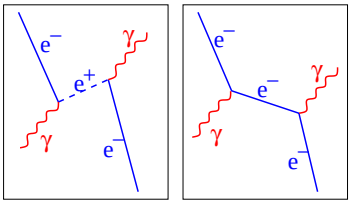



**Fig. 7.3 Feynman diagram: identical particles.** In quantum mechanics, two electrons, (or two atoms of the same isotope) are fundamentally identical. We can illustrate this with a peek at an advanced topic mixing quantum field theory and relativity. Here is a scattering event of a photon off an electron, viewed in two reference frames: time is vertical, a spatial coordinate is horizontal. On the left we see two 'different' electrons, one which is created along with an anti-electron or positron $e^+$, and the other which later annihilates the positron. At right we see the same event viewed in a different reference frame: here there is only one electron, which scatters two photons. (The electron is *virtual*, moving faster than light, between the collisions: this is allowed in intermediate states for quantum transitions.) The two electrons on the left are not only indistinguishable, they are the *same particle*! The antiparticle is also the electron, traveling backward in time.

[12] Examples of bosons include mesons, $He^4$, phonons, photons, gluons, $W^{\pm}$ and Z bosons, and (presumably) gravitons. The last four mediate the fundamental forces – the electromagnetic, strong, weak, and gravitational interactions. The spin-statistics theorem (not discussed here) states that bosons have integer spins. See exercise 7.9.

[14] Most of the common elementary particles are fermions: electrons, protons, neutrons, neutrinos, quarks, *etc.* Fermions have half-integer spins. Particles made up of even numbers of fermions are bosons.

with one another. Just as for the classical ideal gas, this will make our calculations straightforward.

The non-interacting Hamiltonian is a sum of single-particle Hamiltonians $H$:

$$\mathcal{H}^{NI} = \sum_{j=1}^{N} H(\mathbf{p}_j, \mathbf{r}_j) = \sum_{j=1}^{N} \frac{\hbar^2}{2m}\nabla_j^2 + V(\mathbf{r}_j). \tag{7.32}$$

Let $\psi_k$ be the single-particle eigenstates of $H$

$$H\psi_k(\mathbf{r}) = \epsilon_k \psi_k(\mathbf{r}). \tag{7.33}$$

For distinguishable particles, the many-body eigenstates can be written as a product of orthonormal single-particle eigenstates

$$\Psi^{NI}_{\text{dist}}(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N) = \prod_{j=1}^{N} \psi_{k_j}(\mathbf{r}_j). \tag{7.34}$$

where we say that particle $j$ is in the single-particle eigenstate $k_j$. The eigenstates for non-interacting bosons is given by symmetrizing over the coordinates $r_j$,

$$\Psi^{NI}_{\text{boson}}(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N) = (\text{Normalization}) \sum_P \prod_{j=1}^{N} \psi_{k_j}(\mathbf{r}_{P_j}), \tag{7.35}$$

and of course the fermion eigenstates are given by antisymmetrizing[15]

$$\Psi^{NI}_{\text{fermion}}(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N) = \frac{1}{\sqrt{N!}} \sum_P \sigma(P) \prod_{j=1}^{N} \psi_{k_j}(\mathbf{r}_{P_j}). \tag{7.37}$$

Let's consider two particles in orthonormal single-particle energy eigenstates $\psi_k$ and $\psi_\ell$. If the particles are distinguishable, there are two eigenstates $\psi_k(\mathbf{r_1})\psi_\ell(\mathbf{r_2})$ and $\psi_k(\mathbf{r_2})\psi_\ell(\mathbf{r_1})$. If the particles are bosons, the eigenstate is $\frac{1}{\sqrt{2}}\left(\psi_k(\mathbf{r_1})\psi_\ell(\mathbf{r_2}) + \psi_k(\mathbf{r_2})\psi_\ell(\mathbf{r_1})\right)$. If the particles are fermions, the eigenstate is $\frac{1}{\sqrt{2}}\left(\psi_k(\mathbf{r_1})\psi_\ell(\mathbf{r_2}) - \psi_k(\mathbf{r_2})\psi_\ell(\mathbf{r_1})\right)$.

What if the particles are in the same single-particle eigenstate $\psi_\ell$? For bosons, the eigenstate $\psi_k(\mathbf{r_1})\psi_\ell(\mathbf{r_2})$ is already symmetric and normalized.[16] For fermions, antisymmetrizing a state where both particles are in the same state gives zero: $\psi_\ell(\mathbf{r_1})\psi_\ell(\mathbf{r_2}) - \psi_\ell(\mathbf{r_2})\psi_\ell(\mathbf{r_1}) = 0$. This is the Pauli exclusion principle: you cannot have two fermions in the same quantum state.[17]

How do we do statistical mechanics for non-interacting fermions and bosons? Here it is most convenient to use the grand canonical ensemble (section 6.3), so we can think of each single-particle eigenstate $\psi_k$ as being filled independently from the other eigenstates. The grand partition function hence factors:

$$\Xi^{NI} = \prod_k \Xi_k. \tag{7.38}$$

The grand canonical ensemble thus allows us to separately solve the problem one eigenstate at a time, for non-interacting particles.

[15] This antisymmetrization can be written as

$$\frac{1}{\sqrt{N!}}\begin{vmatrix} \psi_{k_1}(\mathbf{r_1}) & \psi_{k_1}(\mathbf{r_2}) & \ldots & \psi_{k_1}(\mathbf{r_N}) \\ \psi_{k_2}(\mathbf{r_1}) & \psi_{k_2}(\mathbf{r_2}) & \ldots & \psi_{k_2}(\mathbf{r_N}) \\ \ldots & & \ldots & \\ \psi_{k_N}(\mathbf{r_1}) & \psi_{k_N}(\mathbf{r_2}) & \ldots & \psi_{k_N}(\mathbf{r_N}) \end{vmatrix} \tag{7.36}$$

called the Slater determinant.

[16] Notice that the normalization of the boson wavefunction depends on how many single-particle states are multiply occupied.

[17] Because the spin of the electron can be in two directions $\pm 1/2$, this means that two electrons can be placed into each single-particle spatial eigenstate.

**Bosons.** For bosons, all fillings $n_k$ are allowed. Each particle in eigenstate $\psi_k$ contributes energy $\epsilon_k$ and chemical potential $-\mu$, so

$$\Xi_k^{\text{boson}} = \sum_{n_k=0}^{\infty} e^{-\beta(\epsilon_k-\mu)n_k} = \sum_{n_k=0}^{\infty} \left(e^{-\beta(\epsilon_k-\mu)}\right)^{n_k} = \frac{1}{1-e^{-\beta(\epsilon_k-\mu)}} \tag{7.39}$$

so the boson grand partition function is

$$\mathbf{\Xi}_{\text{boson}}^{NI} = \prod_k \frac{1}{1-e^{-\beta(\epsilon_k-\mu)}}. \tag{7.40}$$

The grand free energy is a sum of single-state grand free energies

$$\mathbf{\Phi}_{\text{boson}}^{NI} = \sum_k \Phi_k^{\text{boson}} = \sum_k k_B T \log\left(1-e^{-\beta(\epsilon_k-\mu)}\right). \tag{7.41}$$

Because the filling of different states is independent, we can find out the expected number of particles in state $\psi_k$. From equation 6.38,

$$\langle n_k \rangle = -\frac{\partial \Phi_k^{\text{boson}}}{\partial \mu} = -k_B T \frac{-\beta e^{-\beta(\epsilon_k-\mu)}}{1-e^{-\beta(\epsilon_k-\mu)}} = \frac{1}{e^{\beta(\epsilon_k-\mu)}-1}. \tag{7.42}$$

This is called the *Bose-Einstein* distribution.

$$\langle n \rangle_{\text{BE}} = \frac{1}{e^{\beta(\epsilon-\mu)}-1}. \tag{7.43}$$

The Bose-Einstein distribution describes the filling of single-particle eigenstates by non-interacting bosons. For states with low occupancies, where $\langle n \rangle \ll 1$, $\langle n \rangle_{\text{BE}} \approx e^{-\beta(\epsilon-\mu)}$, and the boson populations correspond to what we would guess naively from the Boltzmann distribution.[18] The condition for low occupancies is $\epsilon_k - \mu \gg k_B T$, but perversely this often arises at high temperatures, when $\mu$ gets large and negative. Notice also that $\langle n \rangle_{\text{BE}} \to \infty$ as $\mu \to \epsilon_k$ since the denominator vanishes (and becomes negative for $\mu > \epsilon_k$); systems of non-interacting bosons always have $\mu$ less than or equal to the lowest of the single-particle energy eigenvalues.[19]

Notice that the average excitation $\langle n \rangle_{\text{qho}}$ of the quantum harmonic oscillator (equation 7.24) is given by the Bose-Einstein distribution (equation 7.43) with $\mu = 0$. We'll use this in exercise 7.9 to argue that one can treat excitations inside harmonic oscillators (vibrations) as particles obeying Bose statistics (phonons).

**Fermions.** For fermions, only $n_k = 0$ and $n_k = 1$ are allowed. The single-state fermion grand partition function is

$$\Xi_k^{\text{fermion}} = \sum_{n_k=0}^{1} e^{-\beta(\epsilon_k-\mu)n_k} = 1 + e^{-\beta(\epsilon_k-\mu)} \tag{7.44}$$

so the total fermion grand partition function is

$$\mathbf{\Xi}_{\text{fermion}}^{NI} = \prod_k \left(1+e^{-\beta(\epsilon_k-\mu)}\right). \tag{7.45}$$

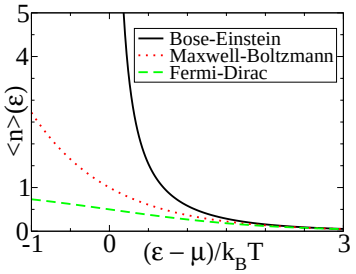



**Fig. 7.4 Bose-Einstein, Maxwell-Boltzmann, and Fermi-Dirac distributions**, $\langle n \rangle_{\text{BE}}(\epsilon)$ Occupation number for single-particle eigenstates as a function of energy $\epsilon$ away from the chemical potential $\mu$. The Bose-Einstein distribution diverges as $\mu$ approaches $\epsilon$; the Fermi-Dirac distribution saturates at one as $\mu$ gets small.

[18] We will formally discuss the Maxwell-Boltzmann distribution in section 7.5.

[19] When the river level gets up to the height of the fields, your farm gets flooded.

For summing over only two states, it's hardly worthwhile to work through the grand free energy to calculate the expected number of particles in a state:

$$\langle n_k \rangle = \frac{\sum_{n_k=0}^{1} n_k \exp(-\beta(\epsilon_k - \mu)n_k)}{\sum_{n_k=0}^{1} n_k \exp(-\beta(\epsilon_k - \mu)} = \frac{e^{-\beta(\epsilon_k-\mu)}}{1+e^{-\beta(\epsilon_k-\mu)}} = \frac{1}{e^{\beta(\epsilon_k-\mu)}+1}, \tag{7.46}$$

leading us to the *Fermi-Dirac* distribution

$$f(\epsilon) = \langle n \rangle_{\mathrm{FD}} = \frac{1}{e^{\beta(\epsilon-\mu)}+1} \tag{7.47}$$

where $f(\epsilon)$ is also known as the Fermi function. Again, when the occupancy of state $\psi_k$ is low, it is approximately given by the Boltzmann probability distribution, $e^{-\beta(\epsilon-\mu)}$. Here the chemical potential can be either greater than or less than any given eigenenergy $\epsilon_k$. Indeed, at low temperatures the chemical potential $\mu$ separates filled states $\epsilon_k < \mu$ from empty states $\epsilon_k > \mu$; only states within roughly $k_B T$ of $\mu$ are partially filled.

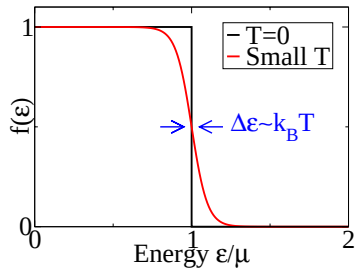



**Fig. 7.5 The Fermi distribution** $f(\epsilon)$ of equation 7.47. At low temperatures, states below $\mu$ are occupied, states above $\mu$ are unoccupied, and states within around $k_B T$ of $\mu$ are partially occupied.

The chemical potential $\mu$ is playing a large role in these calculations, and those new to the subject may wonder how one determines it. You will see in the exercises that one normally knows the expected number of particles $N$, and must vary $\mu$ until you reach that value. Hence $\mu$ very directly plays the role of a particle pressure from the outside world, which is varied until the system is correctly filled.

**The amazing utility of non-interacting bosons and fermions.** The classical ideal gas has been a great illustration of statistical mechanics, and does a good job of many gases, but nobody would suggest that it captures the main features of solids and liquids. The non-interacting approximation in quantum mechanics turns out to be far more powerful, for quite subtle reasons.

For bosons, the non-interacting approximation is quite accurate in three important cases: photons, phonons, and the dilute Bose gas. In section 7.6 we'll study two fundamental problems involving non-interacting bosons: black-body radiation and Bose condensation. The behavior of superconductors and superfluids share some common features with that of the bose gas.

For fermions, the non-interacting approximation would seem to rarely be useful. Electrons are charged, and the electromagnetic repulsion between the electrons in an atom, molecule, or material would seem to always be a major contribution to the energy. Neutrons interact via the strong interaction, so nuclei and neutron stars would seem also poor candidates for a non-interacting theory. Neutrinos are hard to pack into a box.[20] There are experiments on cold, dilute gases of fermion atoms[21] but naively non-interacting fermions would seem a foolish choice to focus on in an introductory course.

The truth is that the non-interacting Fermi gas describes all of these systems (atoms, metals, insulators, nuclei, and neutron stars) remarkably well. Interacting Fermi systems under most common circumstances

[20] Just in case you haven't heard, neutrinos are quite elusive. It is said that if you send neutrinos through a lead shield, more than half will penetrate until the thickness is roughly to the nearest star.

[21] These use the same techniques which led to the observation of Bose condensation.

behave very much like collections of non-interacting fermions in a modified potential.[22] The approximation is so powerful that in most circumstances we ignore the interactions: whenever we talk about exciting a '1S electron' in an oxygen atom, or an 'electron-hole' pair in a semiconductor, we are using this effective non-interacting electron approximation. The explanation for this amazing fact is called Landau Fermi liquid theory, and lies beyond the purview of this text.[23]

[22] In particular, the low-lying excitations above the ground state look qualitatively like fermions excited from below the Fermi energy to above the Fermi energy (electron-hole pairs in metals and semiconductors). It is not that these electrons don't significantly interact with those under the Fermi sea: it is rather that these interactions act to "dress" the electron with a screening cloud. These dressed electrons and holes, or *quasiparticles*, are what act so much like non-interacting particles.

[23] We will give a schematic example of a renormalization-group picture for the irrelevance of Coulomb repulsion in exercise 12.5.

## 7.5 Maxwell-Boltzmann "Quantum" Statistics

In classical statistical mechanics, we treated indistinguishable particles as distinguishable ones, except that we divided the phase-space volume, (or the partition function, in the canonical ensemble) by factor $N!$.

$$\begin{aligned} \Omega_N^{\mathrm{MB}} &= \frac{1}{N!}\Omega_N^{\mathrm{dist}} \\ Z_N^{\mathrm{MB}} &= \frac{1}{N!}Z_N^{\mathrm{dist}} \end{aligned} \tag{7.48}$$

This was important to get the entropy to be extensive (section 5.2.1). This approximation is also used in quantum statistical mechanics, although we should emphasize that it does not describe either bosons, fermions, or any physical system. These bogus particles are said to obey *Maxwell-Boltzmann* statistics.

What is the canonical partition function for the case of $N$ non-interacting distinguishable quantum particles?[24] If the partition function for one particle is

$$Z_1 = \sum_k e^{-\beta\epsilon_k} \tag{7.49}$$

then the partition function for two non-interacting, distinguishable (but otherwise similar) particles is

$$Z_2^{\mathrm{NI,dist}} = \sum_{k_1,k_2} e^{-\beta(\epsilon_{k_1}+\epsilon_{k_2})} = \left(\sum_{k_1} e^{-\beta\epsilon_{k_1}}\right)\left(\sum_{k_2} e^{-\beta\epsilon_{k_2}}\right) = {Z_1}^2. \tag{7.50}$$

and the partition function for $N$ such distinguishable, non-interacting particles is

$$Z_2^{\mathrm{NI,dist}} = \sum_{k_1,k_2,\ldots,k_n} e^{-\beta(\epsilon_{k_1}+\epsilon_{k_2}+\cdots+\epsilon_{k_N})} = \prod_{j=1}^{N}\left(\sum_{k_j} e^{-\beta\epsilon_{k_j}}\right) = {Z_1}^N. \tag{7.51}$$

So, the Maxwell-Boltzmann distribution for non-interacting particles is

$$Z_2^{\mathrm{NI,MB}} = {Z_1}^N/N!. \tag{7.52}$$

Let us illustrate the relation between these three distributions by considering the canonical ensemble of two non-interacting particles in three

[24] More particularly, we want particles which in principle are distinguishable, but which our Hamiltonian treats identically, so they have the same single-particle eigenstates. Two electrons with opposite spin in a Helium atom would be an example. Protons and neutrons in nuclear physics are often treated as identical particles with a different "isospin" quantum number.

possible states of energies $\epsilon_1$, $\epsilon_2$, and $\epsilon_3$. The Maxwell-Boltzmann partition function for such a system would be

$$\begin{aligned} Z_2^{\mathrm{NI,MB}} =& \frac{1}{2!}\left(e^{-\beta\epsilon_1}+e^{-\beta\epsilon_2}+e^{-\beta\epsilon_3}\right)^2 \\ =& \tfrac{1}{2}e^{-2\beta\epsilon_1}+\tfrac{1}{2}e^{-2\beta\epsilon_2}+\tfrac{1}{2}e^{-2\beta\epsilon_3} \\ &+e^{-\beta(\epsilon_1+\epsilon_2)}+e^{-\beta(\epsilon_1+\epsilon_3)}+e^{-\beta(\epsilon_2+\epsilon_3)}. \end{aligned} \tag{7.53}$$

[25] More precisely, we mean those many-body states where the single-particle states are all singly occupied or vacant.

The $1/N!$ fixes the weights of the singly-occupied states[25] nicely: each has weight one in the Maxwell-Boltzmann partition function. But the doubly occupied states, where both particles have the same wavefunction, have an unintuitive suppression by $\frac{1}{2}$ in the sum.

There are basically two ways to fix this. One is to stop discriminating against multiply occupied states, and to treat them all democratically. This gives us non-interacting bosons:

$$Z_2^{\mathrm{NI,boson}} = e^{-2\beta\epsilon_1}+e^{-2\beta\epsilon_2}+e^{-2\beta\epsilon_3}+e^{-\beta(\epsilon_1+\epsilon_2)}+e^{-\beta(\epsilon_1+\epsilon_3)}+e^{-\beta(\epsilon_2+\epsilon_3)}. \tag{7.54}$$

The other way is to squelch multiple occupancy altogether. This leads to fermions:

$$Z_2^{\mathrm{NI,fermion}} = e^{-\beta(\epsilon_1+\epsilon_2)}+e^{-\beta(\epsilon_1+\epsilon_3)}+e^{-\beta(\epsilon_2+\epsilon_3)}. \tag{7.55}$$

Thus the Maxwell-Boltzmann distribution treats multiple occupancy of states in an unphysical compromise between bosons (democratic) and fermions (forbidden).

We've been working in this section with the canonical distribution, fixing the number of particles to two. This is convenient only for small systems; normally we've used the grand canonical ensemble.[26] How does the grand canonical ensemble apply to particles with Maxwell-Boltzmann statistics? The grand partition function is a geometric series:[27]

[26] See exercise 7.6 for more details about the three ensembles and the four types of statistics.

[27] Notice the unusual appearance of $e^{e^x}$ in this formula.

$$\begin{aligned} \Xi^{\mathrm{NI,MB}} &= \sum_M \frac{1}{M!}\left(Z_M^{\mathrm{NI,MB}}\right)^M e^{M\beta\mu} = \sum_M \frac{1}{M!}\left(\sum_k e^{-\beta\epsilon_k}\right)^M e^{M\beta\mu} \\ &= \sum_M \frac{1}{M}\left(\sum_k e^{-\beta(\epsilon_k-\mu)}\right)^M = \sum_M e^{\sum_k e^{-\beta(\epsilon_k-\mu)}} \\ &= \prod_k e^{e^{-\beta(\epsilon_k-\mu)}}. \end{aligned} \tag{7.56}$$

The grand free energy is

$$\Phi^{\mathrm{NI,MB}} = -k_BT\log\Xi^{\mathrm{NI,MB}} = -k_BT\sum_k \Phi_k \tag{7.57}$$

with the single-particle grand free energy

$$\Phi_k = -k_BTe^{-\beta(\epsilon_k-\mu)}. \tag{7.58}$$

Finally, the expected[28] number of particles in a single-particle eigenstate with energy $\epsilon$ is

$$\langle n \rangle_{\rm MB} = -\frac{\partial \Phi}{\partial \mu} = e^{-\beta(\epsilon-\mu)}. \tag{7.60}$$

This is precisely the Boltzmann factor for filling the state that we expect for non-interacting distinguishable particles.

## 7.6 Black Body Radiation and Bose Condensation

### 7.6.1 Free Particles in a Periodic Box

For this section and the next section on fermions, we shall simplify even further. We consider particles which are not only non-interacting and identical, but are also free. That is, they are subject to no external potential, apart from being confined in a box of volume $L^3 = V$ with periodic boundary conditions. The single-particle quantum eigenstates of such a system are products of sine and cosine waves along the three directions – for example, for any three non-negative integers $n_i$,

$$\psi = (2/L)^{3/2} \sin\left(\frac{2\pi n_1}{L} x\right) \cos\left(\frac{2\pi n_2}{L} y\right) \sin\left(\frac{2\pi n_3}{L} z\right). \tag{7.61}$$

There are eight such states with the same energy, substituting cos for sin in all possible combinations along the three directions. These are more conveniently organized if we use the complex exponential instead of sine and cosine,

$$\psi_{\mathbf{k}} = (1/L)^{3/2} \exp(i\mathbf{k}\cdot\mathbf{r}) \tag{7.62}$$

with $\mathbf{k} = \frac{2\pi}{L}(n_1, n_2, n_3)$ and the $n_i$ can now be any integer.[29] The allowed single-particle eigenstates form a regular square grid in the space of wavevectors $k$, with an average density $(L/2\pi)^3$ per unit volume of $k$-space.

$$\text{Density of Plane Waves in } \mathbf{k}\text{-space} = V/8\pi^3 \tag{7.63}$$

For a large box volume $V$, the grid is extremely fine, and one can use a continuum approximation that the number of states falling into a $\mathbf{k}$-space region is given by its volume times the density (equation 7.63).

Basically, the continuum limit works because the shape of the box (which affects the arrangements of the allowed $\mathbf{k}$ vectors) is irrelevant to the physics so long as the box is large. For the same reason, the energy of the single-particle eigenstates is independent of direction: it will be

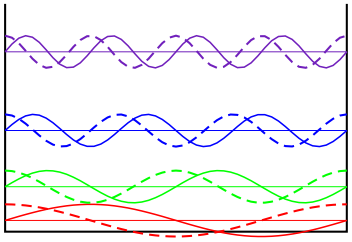


**Fig. 7.6** The quantum states of a particle in a one-dimensional box with periodic boundary conditions are sine and cosine waves $\psi_n$ with $n$ wavelengths in the box, $k_n = 2\pi n/L$. With a real box (zero boundary conditions at the walls) one would have only sine waves, but one at every half-wavelength, $k_n = \pi n/L$, giving the same net density of states.

[29] The eight degenerate states are now given by the choices of sign for the three integers.

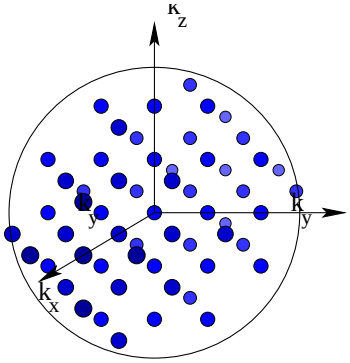



**Fig. 7.7** The allowed $\mathbf{k}$-space points for periodic boundary conditions form a regular grid. The points of equal energy lie on a sphere.

[28] It is amusing to note that non-interacting particles fill single particle energy states according to the same law

$$\langle n \rangle = \frac{1}{e^{\beta(\epsilon-\mu)} + c}, \tag{7.59}$$

with $c = -1$ for bosons, $c = 1$ for fermions, and $c = 0$ for Maxwell-Boltzmann statistics.

proportional to $|\mathbf{k}|$ for massless photons, and proportional to $\mathbf{k}^2$ for massive bosons and electrons (figure 7.7). This makes the calculations in the following sections tractable.

### 7.6.2 Black Body Radiation

Our first application is to electromagnetic radiation. Electromagnetic radiation has plane-wave modes similar to equation 7.62. Each plane-wave travels at the speed of light $c$, so its frequency is $\omega_k = c|\mathbf{k}|$. There are two modes per wavevector $\mathbf{k}$, one for each polarization. When one quantizes the electromagnetic field, each mode becomes a quantum harmonic oscillator.

Before quantum mechanics, people could not understand how electromagnetic radiation could come to equilibrium. The equipartition theorem suggested that if you could come to equilibrium, each mode would have $k_BT$ of energy. Since there are immensely more wavevectors in the ultraviolet and X-ray ranges than in the infrared and visible, this predicts that when you open your oven door you'd get a sun tan or worse (the so-called *ultraviolet catastrophe*). Simple experiments looking at radiation emitted from pinholes in otherwise closed boxes held at fixed temperature saw a spectrum which looked compatible with classical statistical mechanics for small frequency radiation, but was cut off at high frequencies.

Let us calculate the equilibrium energy distribution inside our box at temperature $T$. The number of single-particle plane-wave eigenstates $g(\omega)\,d\omega$ in a small range $d\omega$ is[30]

$$g(\omega)\,d\omega = (4\pi k^2)\left(\frac{d|\mathbf{k}|}{d\omega}d\omega\right)\left(\frac{2V}{(2\pi)^3}\right) \tag{7.64}$$

where the first term is the surface area of the sphere of radius $k$, the second term is the thickness of the sphere for a small $d\omega$, and the last is the density of single-particle plane-wave eigenstate wavevectors times two (because there are two photon polarizations per wavevector). Knowing $\mathbf{k}^2 = \omega^2/c^2$ and $d|\mathbf{k}|/d\omega = 1/c$, we find the density of plane-wave eigenstates per unit frequency

$$g(\omega) = \frac{V\omega^2}{\pi^2c^3}. \tag{7.65}$$

Now, the number of photons is not fixed: they can be created or destroyed, so their chemical potential $\mu$ is zero.[31] Their energy $\epsilon_k = \hbar\omega_k$. Finally, they are to an excellent approximation identical, non-interacting bosons, so the number of photons per eigenstate with frequency $\omega$ is $\langle n\rangle = \frac{1}{e^{\hbar\omega/k_BT}-1}$. This gives us a number of photons

$$(\#\text{ of photons})\,d\omega = \frac{g(\omega)}{e^{\hbar\omega/k_BT}-1}d\omega \tag{7.66}$$

[30] We're going to be sloppy and use $g(\omega)$ for photons to be eigenstates per unit frequency, and $g(\epsilon)$ later for single-particle eigenstates per unit energy = $\hbar\cdot$ frequency.

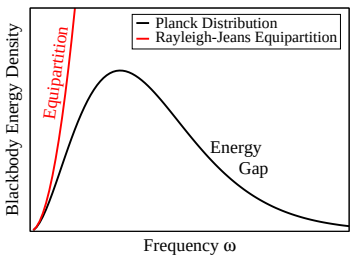



**Fig. 7.8** The Planck black-body radiation power spectrum, with the Rayleigh-Jeans approximation, valid for low frequency $\omega$.

[31] We can also see this from the fact that photons are excitations within a harmonic oscillator; in section 7.4 we noted that the excitations in a harmonic oscillator satisfy Bose statistics with $\mu = 0$.

and an electromagnetic (photon) energy per unit volume $u(\omega)$ given by

$$\begin{aligned} Vu(\omega)d\omega &= \frac{\hbar\omega g(\omega)}{e^{\hbar\omega/k_BT}-1}d\omega \\ &= \frac{V\hbar}{\pi^2c^3}\frac{\omega^3\,d\omega}{e^{\hbar\omega/k_BT}-1}. \end{aligned} \tag{7.67}$$

This is Planck's famous formula for black-body radiation.[32] At low frequencies, we can approximate $e^{\hbar\omega/k_BT-1} \approx \hbar\omega/k_BT$, yielding the Rayleigh-Jeans formula

$$\begin{aligned} Vu_{\mathrm{RJ}}(\omega)d\omega &= V\left(\frac{k_BT}{\pi^2c^3}\right)\omega^2\,d\omega \\ &= k_BTg(\omega) \end{aligned} \tag{7.68}$$

just as one would expect from equipartition: $k_BT$ per classical harmonic oscillator.

For modes with frequencies high compared to $k_BT/\hbar$, equipartition no longer holds. The energy gap $\hbar\omega$, just as for the low-temperature specific heat from section 7.2, leads to an excitation probability that is suppressed by the exponential Boltzmann factor $e^{-\hbar\omega/k_BT}$, as one can see from equation 7.67 by approximating $\frac{1}{e^{\hbar\omega/k_BT}-1} \approx e^{-\hbar\omega/k_BT}$. Planck's discovery that quantizing the energy averted the ultraviolet catastrophe was the origin of quantum mechanics, and led to his name being given to $\hbar$.

### 7.6.3 Bose Condensation

How does our calculation change when the non-interacting free bosons cannot be created and destroyed? Let us assume that our bosons are spinless, have mass $m$, and are non-relativistic so their energy is $\epsilon = p^2/2m = -\hbar^2\nabla^2/2m$. If we put them in our box with periodic boundary conditions, we can make the same continuum approximation to the density of states as we did in the case of black-body radiation. In equation 7.63, the number of plane-wave eigenstates per unit volume in $\mathbf{k}$-space is $V/8\pi^3$, so the density in momentum space $\mathbf{p} = \hbar\mathbf{k}$ is $V/(2\pi\hbar)^3$. For our massive particles $d\epsilon/d|\mathbf{p}| = |\mathbf{p}|/m = \sqrt{2\epsilon/m}$, so the number of

[32] Why is this called *black body* radiation? A black surface absorbs all radiation at all frequencies. In equilibrium, the energy it absorbs at a given frequency must equal the energy it emits, otherwise it would push the system out of equilibrium. (This is called *detailed balance*, section 8.3.) Since the emission at a surface is largely independent of the impinging energy, a black body emits a thermal distribution of photons, equation 7.67.

plane-wave eigenstates in a small range of energy $d\epsilon$ is

$$\begin{aligned} g(\epsilon)d\epsilon &= (4\pi p^2)\left(\frac{d|\mathbf{p}|}{d\epsilon}d\epsilon\right)\left(\frac{V}{(2\pi\hbar)^3}\right) \\ &= (4\pi(2m\epsilon))\left(\sqrt{\frac{m}{2\epsilon}}d\epsilon\right)\left(\frac{V}{(2\pi\hbar)^3}\right) \\ &= \frac{Vm^{3/2}}{\sqrt{2}\pi^2\hbar^3}\sqrt{\epsilon}\,d\epsilon. \end{aligned} \tag{7.69}$$

where the first term is the surface area of the sphere in $\mathbf{p}$ space, the second is the thickness of the sphere, and the third is the density of plane-wave eigenstates per unit volume in $\mathbf{p}$-space.

Now, we fill each of these single-particle plane-wave eigenstates with an expected number given by the Bose-Einstein distribution at chemical potential $\mu$, $1/(e^{(\epsilon-\mu)/k_BT}-1)$, so the total number of particles $N$ must be given by

$$N(\mu) = \int_0^\infty \frac{g(\epsilon)}{e^{(\epsilon-\mu)/k_BT}-1}\,d\epsilon. \tag{7.70}$$

We must vary $\mu$ in this equation to give us the correct number of particles $N$. For larger numbers of particles we raise $\mu$, forcing more particles into each of the single-particle states. There is a limit, however, to how hard we can push. As we noted in section 7.4, $\mu$ cannot be as large than the lowest single-particle eigenvalue, because at that point that state gets a diverging number of particles. However, for free bosons in three dimensions, the integral for $N(\mu)$ converges to a finite value.[33]

Thus the largest number of particles $N_{\mathrm{max}}^{\mathrm{cont}}$ we can fit into our box within our continuum approximation for the density of states is the value of equation 7.70 at $\mu = 0$,[34]

$$\begin{aligned} N_{\mathrm{max}}^{\mathrm{cont}} &= \int \frac{g(\epsilon)}{e^{\epsilon/k_BT}-1}\,d\epsilon. \\ &= \frac{Vm^{3/2}}{\sqrt{2}\pi^2\hbar^3}\int_0^\infty d\epsilon \frac{\sqrt{\epsilon}}{e^{\epsilon/k_BT}-1} \\ &= V\left(\frac{\sqrt{2\pi m k_B T}}{h}\right)^3 \frac{2}{\sqrt{\pi}}\int_0^\infty \frac{\sqrt{z}}{e^z-1}\,dz \\ &= \left(\frac{V}{\lambda^3}\right)\zeta(3/2). \end{aligned} \tag{7.71}$$

where $\zeta$ is the Riemann zeta function, with $\zeta(3/2) \approx 2.612$, and where $\lambda = h/\sqrt{2\pi m k_B T}$ is the thermal de Broglie wavelength we saw first in the canonical ensemble of the ideal gas, equation 3.63.[35] Thus something new has to happen at a critical density

$$\frac{N_{\mathrm{max}}^{\mathrm{cont}}}{V} = \frac{\zeta(3/2)}{\lambda^3} = \frac{2.612 \text{ particles}}{\text{deBroglie volume}}. \tag{7.72}$$

What happens when we try to cram more particles in? What happens is that our approximation of the distribution of eigenstates as a

[33] At $\mu = 0$, the denominator of the integrand in equation 7.70 is approximately $\epsilon/k_BT$ for small $\epsilon$, but the numerator goes as $\sqrt{\epsilon}$, so the integral converges at the lower end: $\int_0^X \epsilon^{-1/2} \sim 1/2\,\epsilon^{1/2}|_0^X = \sqrt{X}/2$.

[34] The $\zeta$ function $\zeta(s) = \frac{1}{(s-1)!}\int_0^\infty \frac{z^{s-1}}{e^s-1}\,ds$ is famous because it is related to the distribution of prime numbers, because it is the subject of the famous unproven Riemann hypothesis (about its zeros in the complex plane), and because the values in certain regions form excellent random numbers.

[35] This formula has an elegant interpretation: the quantum statistics of the particles begin to dominate the behavior when they are within a thermal de Broglie wavelength of one another.

continuum breaks down. Figure 7.9 shows a schematic of the first few single-particle eigenvalues. When the distance between $\mu$ and the bottom level becomes significantly smaller than the distance between the bottom and the next level, the continuum approximation (which roughly treats the single state $\epsilon_0$ as the integral halfway to $\epsilon_1$) becomes qualitatively wrong. This lowest state absorbs all the extra particles added to the system beyond $N_{\rm max}^{\rm cont}$.[36] This is called *Bose-Einstein condensation.*

Usually, one doesn't add particles at fixed temperature, one lowers the temperature at fixed density $N/V$, where Bose condensation occurs at temperature

$$k_B T_c^{\rm BEC} = \frac{h^2}{2\pi m}\left(\frac{N}{V\zeta(3/2)}\right)^{2/3}. \tag{7.73}$$

Bose condensation was first accomplished experimentally in 1995 (see exercise 7.11).

Bose condensation has also long been considered the underlying principle behind superfluidity.[37] Liquid $He^4$ undergoes an unusual transition at about 2.176K to a state without viscosity: it will swirl round a circular tube for as long as your refrigeration lasts. The quantitative study of the superfluid transition involves the interactions between the helium atoms, and uses the scaling methods we'll introduce in chapter 12. But it's interesting to note that the Bose condensation temperature for liquid $He^4$ (with $m = 6.65 \times 10^{-24}$gm and volume per particle $V/N = 27.6$ cm/mole) is 3.13K: quite close to the superfluid transition temperature.

[36] The next few states have quantitative corrections, but the continuum approximation is only off by small factors.

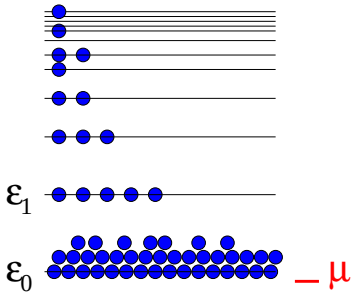



**Fig. 7.9** Bose condensation: the chemical potential $\mu$ is so close to the ground state energy $\epsilon_0$ that the continuum approximation to the density of states breaks down. The ground state is macroscopically occupied (that is, filled by a non-zero fraction of the total number of particles $N$).

[37] The connection is deep. The density matrix of a superfluid has an unusual property, called off-diagonal long-range-order, which is also found in the Bose condensate (see exercise 9.7).

## 7.7 Metals and the Fermi Gas

We claimed in section 7.4 that many systems of strongly-interacting fermions (metals, neutron stars, nuclei) are surprisingly well described by a model of non-interacting fermions. Let's solve for the properties of $N$ *free* non-interacting fermions in a box.

Let our particles be non-relativistic and spin 1/2. The single-particle eigenstates are the same as those for bosons except that there are two states (spin up, spin down) per plane wave. Hence the density of states is given by twice that of equation 7.69:

$$g(\epsilon) = \frac{\sqrt{2}Vm^{3/2}}{\pi^2\hbar^3}\sqrt{\epsilon}. \tag{7.74}$$

The number of fermions at chemical potential $\mu$ is given by integrating $g(\epsilon)$ times the expected number of fermions in a state of energy $\epsilon$, given by the Fermi function $f(\epsilon)$ of equation 7.47:

$$N(\mu) = \int_0^\infty g(\epsilon) f(\epsilon)\, d\epsilon = \int_0^\infty \frac{g(\epsilon)}{e^{(\epsilon-\mu)/k_BT}+1}\, d\epsilon. \tag{7.75}$$

What chemical potential will give us $N$ fermions? At non-zero temperature, one must do a self-consistent calculation, but at $T = 0$ one can find $N$ by counting the number of states below $\mu$. In the zero-temperature

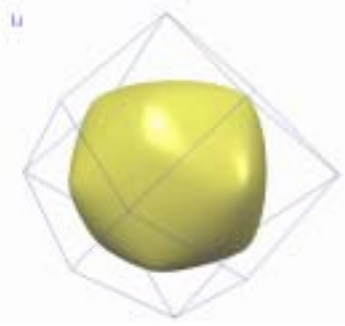


**Fig. 7.10 The Fermi surface for lithium**, from [22]. The Fermi energy for lithium is 4.74 eV, with one conduction electron outside a He closed shell. Note that for most metals the Fermi energy is much larger than $k_B$ times the melting point ($\epsilon_F$ =4.74 eV = 55,000 K, and the melting point is 453 K). Hence they are well described by the $T = 0$ Fermi surfaces here, slightly smeared by the Fermi function shown in figure 7.5.

limit (figure 7.5) the Fermi function is a step function $f(\epsilon) = \Theta(\mu - \epsilon)$; all states below $\mu$ are filled, and all states above $\mu$ are empty. The zero-temperature value of the chemical potential is called the *Fermi energy* $\epsilon_F$. We can find the number of fermions by integrating up to $\mu = \epsilon_F$:

$$N = \int_0^{\epsilon_F} g(\epsilon)\, d\epsilon = \frac{\sqrt{2}m^{3/2}}{\pi^2\hbar^3} V \int_0^{\epsilon_F} \sqrt{\epsilon}\, d\epsilon = \frac{(2\epsilon_F m)^{3/2}}{3\pi^2\hbar^3} V. \tag{7.76}$$

This formula becomes easier to understand if we realize that we're filling all states with wavevector $\mathbf{k} < k_F$, where the Fermi wavevector $k_F$ is the length of the wavevector whose eigenenergy equals the Fermi energy: $\hbar k_F^2/2m = p_F^2/2m = \epsilon_F$, so $k_F = \sqrt{2\epsilon_F m}/\hbar$. The resulting sphere of occupied states at $T = 0$ is called the Fermi sphere. The number of fermions inside the Fermi sphere is thus

$$N = \frac{{k_F}^3}{3\pi^2} = \left(\tfrac{4}{3}\pi {k_F}^3 V\right)\left(\frac{2V}{(2\pi)^3}\right) \tag{7.77}$$

the $\mathbf{k}$-space volume of the Fermi sphere times the $\mathbf{k}$-space density of states. Knowing the total number of electrons $N$ in volume $V$ at $T = 0$ thus tells us $k_F$, and hence $\epsilon_F$ and the chemical potential $\mu$.

We mentioned earlier that the independent fermion approximation was startlingly useful even though the interactions are not small. Ignoring the Coulomb repulsion between electrons in a metal, or the strong interaction between neutrons in a neutron star, gives an excellent description of their actual behavior. Our calculation above, though also assumed that the electrons are free particles, experiencing no external potential. This approximation isn't particularly accurate in general: the interactions with the atomic nuclei are important, and is primarily what makes one material different from another. In particular, the atoms in a crystal will form a periodic potential for the electrons. Rather than using the Coulomb potential for the nucleus, a better approximation is given by incorporating the effects of the *inner shell* electrons into the periodic potential, and filling the Fermi sea with the remaining *conduction* electrons. One can show that the single-particle eigenstates in a periodic potential can be chosen to be periodic functions times plane waves (equation 7.62) of exactly the same wave-vectors as in the free fermion case. The filling of the Fermi surface in the resulting $\mathbf{k}$-space described here is changed only insofar as the energies of these single-particle states is no longer isotropic. Some metals (particularly the alkali metals, like lithium in figure 7.10) have roughly spherical Fermi surfaces; many (like aluminum in figure 7.11) are quite intricate, with several pieces to them (reference [8, Ch. 9-11]).

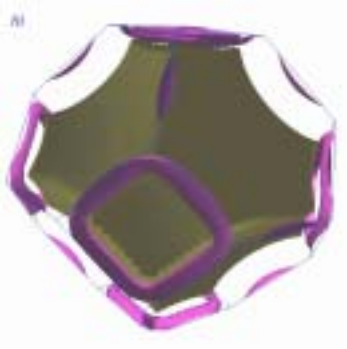


**Fig. 7.11 The Fermi surface for aluminum**, also from [22]. Aluminum has a Fermi energy of 11.7 eV, with three conduction electrons outside a Ne closed shell.

# Exercises

(7.1) **Phase Space Units and the Zero of Entropy.** (Quantum)

In classical mechanics, the entropy $S = k_B \log \Omega$ goes to minus infinity as the temperature is lowered to zero; in quantum mechanics the entropy per particle goes to zero,[38] because states are quantized and the ground state is the only one populated. This is Nernst's theorem, the third law of thermodynamics.

The classical phase-space volume $\Omega$ has units of $((\text{momentum}) \times (\text{distance}))^{3N}$. It's a little perverse to take the logarithm of a quantity with units. The obvious candidate with these dimensions is Planck's constant $h^{3N}$: if we measure phase-space volume in units of $h$ per dimension, $\Omega$ will be dimensionless. Of course, the correct dimension could be a constant times $h$, like $\hbar \ldots$

*(a)* **Arbitrary zero of the classical entropy.** *Show that the width of the energy shell $dE$ in the definition of the entropy does not change the classical entropy per particle $S/N$. Show that the choice of units in phase space does changes the classical entropy per particle.*

We need to choose the units of classical phase–space volume so that the entropy agrees with the high–temperature entropy for the quantum systems. That is, we need to find out how many quantum eigenstates per unit volume of classical phase space we should expect at high energies. We can fix these units by explicitly matching the quantum result to the classical one for a particular system. Let's start with a free particle.

*(b)* **Phase–space density of states for a particle in a one-dimensional box.** *Show, or note, that the quantum momentum–space density of states for a free quantum particle in a one–dimensional box of length $L$ with periodic boundary conditions is $L/h$. Draw a picture of the classical phase space of this box $(p, x)$, and draw a rectangle of length $L$ for each quantum eigenstate. Is the phase-space area per eigenstate equal to $h$, as we assumed in 3.5?*

This works also for $N$ particles in a three–dimensional box.

*(c)* **Phase–space density of states for $N$ particles in a box.** *Show that the density of states for $N$ free particles in a cubical box of volume $V$ with periodic boundary conditions is $V^N/h^{3N}$, and hence that the phase–space volume per state is $h^{3N}$.*

Can we be sure that the answer is independent of which system we use to match? Let's see if it also works for the harmonic oscillator.

*(d)* **Phase–space density of states for a harmonic oscillator.** *Consider a harmonic oscillator with Hamiltonian $\mathcal{H} = p^2/2m + \frac{1}{2}m\omega^2 q^2$. Draw a picture of the energy surface with energy $E$, and find the volume (area) of phase space enclosed.* (Hint: the area of an ellipse is $\pi r_1 r_2$ where $r_1$ and $r_2$ are the largest and smallest radii, corresponding to the major and minor axes.) *What is the volume per energy state, the volume between $E_n$ and $E_{n+1}$, for the eigenenergies $E_n = (n + \frac{1}{2})\hbar\omega$?*

Why must these two calculations agree? How can we derive this result in general, even for nasty systems of interacting particles? The two traditional methods for directly calculating the phase–space units in general systems – semiclassical quantization [56, ch. 48, p. 170] and the path–integral formulation of quantum statistical mechanics [31] – would be too distracting to present here.

Upon some thought, we realize that one cannot choose different units for the classical phase–space volume for different systems. They all must agree, because one can transform one into another. Consider $N$ interacting particles in a box, at high temperatures where classical statistical mechanics is valid. Imagine slowly and reversibly turning off the interactions between the particles (making them into our ideal gas). We carefully remain at high temperatures, and measure the entropy flow into or out of the system. The entropy difference will be given by classical statistical mechanics, whatever units one wishes to choose for the phase–space volume. The entropy of the interacting system is thus the entropy of the ideal gas (with volume $h$ per state) plus the classical entropy change – hence also must use the same phase–space units.[39]

(7.2) **Does Entropy Increase in Quantum Systems?** (Mathematics, Quantum)

We saw in exercise 5.4 that in classical Hamiltonian systems the non-equilibrium entropy $S_{nonequil} = -k_B \int \rho \log \rho$ is constant in a classical mechanical Hamiltonian system. We'll show here that it is constant also in a closed quantum Hamiltonian system.

A general ensemble in a quantum system is described by the density matrix $\boldsymbol{\rho}$. In most of statistical mechanics,

[38] If the ground state is degenerate, the entropy doesn't go to zero, but it typically stays finite as the number of particles $N$ gets big, so for large $N$ the entropy per particle goes to zero.

[39] In particular, if we cool the interacting system to zero temperature and remain in equilibrium, reaching the ground state, that its entropy will go to zero – the entropy flow out of the system on cooling is given by our classical formula with phase-space volume measured in units of $h$.

$\boldsymbol{\rho}$ is diagonal when we use a basis of energy eigenstates. Here, since each energy eigenstate is time independent except for a phase, any mixture of energy eigenstates will have a constant density matrix, and so will have a constant entropy.

*(a)* **Entropy is Constant: Mixtures of Energy Eigenstates.** *Prove that if $\boldsymbol{\rho}$ is a density matrix diagonal in the basis of energy eigenstates, that $\boldsymbol{\rho}$ is time independent. Hence, conclude that the entropy $S = \mathrm{Tr}\boldsymbol{\rho}\log\boldsymbol{\rho}$ is time-independent*

Thus, not only are the microcanonical and canonical ensembles time independent, but mixing energy eigenstates in any ratio would be time independent. To justify equilibration in quantum systems, one must couple the system to the outside world and induce transitions between eigenstates.

In the particular case of the entropy, the entropy is time independent even for general, time-dependent density matrices.

*(b)* **Entropy is Constant: General Density Matrices.** *Prove that $S = \mathrm{Tr}\,(\boldsymbol{\rho}\log\boldsymbol{\rho})$ is time-independent, where $\boldsymbol{\rho}$ is any density matrix.* (Hint: Show that $\mathrm{Tr}(\mathbf{ABC}) = \mathrm{Tr}(\mathbf{CAB})$ for any matrices $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$. Also you should know that an operator $\mathbf{M}$ commutes with any function $f(\mathbf{M})$.)

(7.3) **Phonons on a String.** (Quantum)

*One-dimensional Phonons.* A nano-string of length $L$ with mass per unit length $\mu$ under tension $\tau$ has a vertical, transverse displacement $u(x,t)$. The kinetic energy density is $(\mu/2)(\partial u/\partial t)^2$ and the potential energy density is $(\tau/2)(\partial u/\partial x)^2$.

*Write the kinetic energy and the potential energy in new variables, changing from $u(x,t)$ to normal modes $q_k(t)$ with $u(x,t) = \sum_n q_{k_n}(t)\sin(k_n x)$, $k_n = n\pi/L$. Show in these variables that the system is a sum of decoupled harmonic oscillators. Calculate the density of states per unit frequency $g(\omega)$, the number of normal modes in a frequency range $(\omega, \omega+\epsilon)$ divided by $\epsilon$, keeping $\epsilon$ large compared to the spacing between modes.*[40] *Calculate the specific heat of the string $c(T)$ per unit length in the limit $L \to \infty$, treating the oscillators quantum mechanically. What is the specific heat of the classical string?*

(7.4) **Crystal Defects.** (Quantum, Basic)

*Defects in Crystals.* A defect in a crystal has one on-center configuration with energy zero, and $M$ off-center configurations with energy $\epsilon$, with no significant quantum tunneling between the states. The Hamiltonian can be approximated by the $(M+1)\times(M+1)$ matrix

$$\mathcal{H} = \begin{pmatrix} 0 & 0 & 0 & \cdots \\ 0 & \epsilon & 0 & \cdots \\ 0 & 0 & \epsilon & \cdots \end{pmatrix} \tag{7.78}$$

There are $N$ defects in the crystal, which can be assumed stuck in position (and hence distinguishable) and assumed not to interact with one another.

*Write the canonical partition function $Z(T)$, the mean energy $E(T)$, the fluctuations in the energy, the entropy $S(T)$, and the specific heat $C(T)$ as a function of temperature. Plot the specific heat per defect $C(T)/N$ for $M = 6$; set the unit of energy equal to $\epsilon$ and $k_B = 1$ for your plot. Derive a simple relation between $M$ and the change in entropy between zero and infinite temperature. Check this relation using your formula for $S(T)$.*

(7.5) **Density Matrices.** (Quantum)

*(a)* **Density matrices for photons.** *Write the density matrix for a photon linearly traveling along $z$ and linearly polarized along $\hat{x}$, in the basis where $(1,0)$ and $(0,1)$ are polarized along $\hat{x}$ and $\hat{y}$. Write the density matrix for a right-handed polarized photon, $(1/\sqrt{2}, i/\sqrt{2})$, and the density matrix for unpolarized light. Calculate $\mathrm{Tr}(\boldsymbol{\rho})$, $\mathrm{Tr}(\boldsymbol{\rho}^2)$, and $S = -k_B\mathrm{Tr}(\boldsymbol{\rho}\log\boldsymbol{\rho})$. Interpret the values of the three traces physically: one is a check for pure states, one is a measure of information, and one is a normalization.*

*(b)* **Density matrices for a spin.** *(Adapted from Halperin's course, 1976.) Let the Hamiltonian for a spin be*

$$\mathcal{H} = -\frac{\hbar}{2}\mathbf{B}\cdot\hat{\vec{\sigma}} \tag{7.79}$$

*where $\hat{\vec{\sigma}} = (\sigma_x, \sigma_y, \sigma_z)$ are the three Pauli spin matrices, and $\mathbf{B}$ may be interpreted as a magnetic field, in units where the gyromagnetic ratio is unity. Remember that $\sigma_i\sigma_j - \sigma_j\sigma_i = 2i\epsilon_{ijk}\sigma_k$. Show that any $2\times 2$ density matrix may be written in the form*

$$\boldsymbol{\rho} = \tfrac{1}{2}(\mathbf{1} + \mathbf{p}\cdot\hat{\vec{\sigma}}). \tag{7.80}$$

*Show that the equations of motion for the density matrix $i\hbar\partial\boldsymbol{\rho}/\partial t = [\mathcal{H}, \boldsymbol{\rho}]$ can be written as $d\mathbf{p}/dt = -\mathbf{B}\times\mathbf{p}$.*

[40] This is the density of single–particle eigenstates per unit frequency. In exercise 7.11 we'll study the density of many-body energy eigenstates $g(E)$ in a trap with precisely three frequencies (where our $g(\omega)$ would be $= \delta(\omega-\omega_0) + 2\delta(\omega-\omega_1)$). Don't confuse the two.

(7.6) **Ensembles and Statistics: 3 Particles, 2 Levels.** (Quantum)

A system has two single-particle eigenfunctions, with energies (measured in degrees Kelvin) $E_0/k_B = -10$ and $E_2/k_B = 10$. Experiments are performed by adding three non-interacting particles to these two states, either identical spin 1/2 fermions, identical spinless bosons, distinguishable particles, or spinless identical particles obeying Maxwell-Boltzmann statistics. Please make a table for this exercise, giving your answers for the four cases (Fermi, Bose, Dist., and MB) for each of the three parts. Calculations may be needed, but only the answers will be graded.

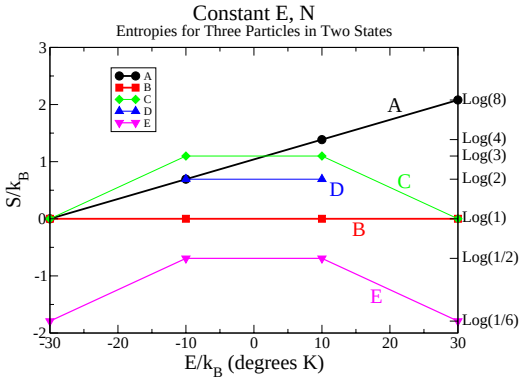



**Fig. 7.12**

*(a) The system is first held at constant energy. In figure 7.12 which curve represents the entropy of the fermions as a function of the energy? Bosons? Distinguishable particles? Maxwell-Boltzmann particles?*

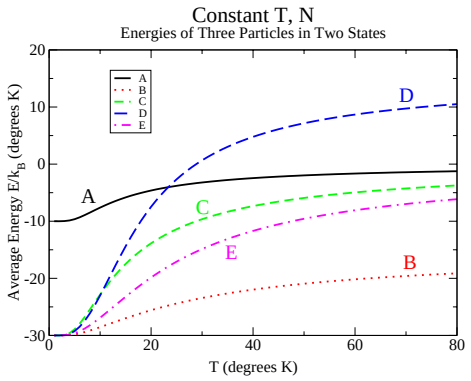



**Fig. 7.13**

*(b) The system is now held at constant temperature. In figure 7.13 which curve represents the mean energy of the fermions as a function of temperature? Bosons? Distinguishable particles? Maxwell-Boltzmann particles?*

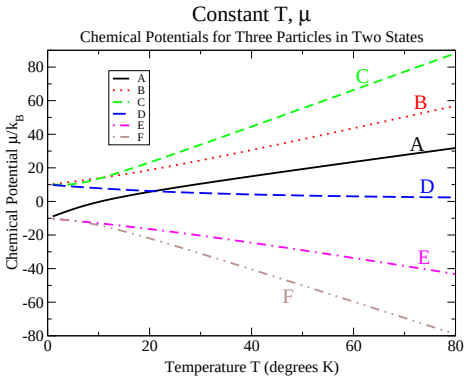



**Fig. 7.14**

*(c) The system is now held at constant temperature, with chemical potential set to hold the average number of particles equal to three. In figure 7.14, which curve represents the chemical potential of the fermions as a function of temperature? Bosons? Distinguishable? Maxwell-Boltzmann?*

(7.7) **Bosons are Gregarious: Superfluids and Lasers** (Quantum)

Many experiments insert a new particle into a many–body state: new electrons into a metal, new electrons or electron pairs into a superconductor, new bosonic atoms into a superfluid, new photons into a cavity already filled with light. These experiments explore how the "bare" inserted particle decomposes into the natural states of the many–body system. The cases of photons and bosons illustrate a key connection between laser physics and Bose condensates.

**Adding a particle to a Bose condensate.** Suppose we have a non–interacting system of bosonic atoms in a box with single–particle eigenstates $\psi_n$. Suppose the system begins in a Bose condensed state with all $N$ bosons in a state $\psi_0$, so

$$\Psi_N^{[0]}(\mathbf{r}_1, \ldots, \mathbf{r}_N) = \psi_0(\mathbf{r}_1) \ldots \psi_0(\mathbf{r}_N) \tag{7.81}$$

Suppose a new particle is gently injected into the system, into an equal superposition of the $M$ lowest single–

particle states.[41] That is, if it were injected into an empty box, it would start in state

$$\phi(\mathbf{r}_{N+1}) = \frac{1}{\sqrt{M}}\big(\psi_0(\mathbf{r}_{N+1}) + \psi_1(\mathbf{r}_{N+1}) + \cdots + \psi_{M-1}(\mathbf{r}_{N+1})\big) \tag{7.82}$$

The state $\Phi(\mathbf{r}_1, \ldots \mathbf{r}_{N+1})$ after the particle is inserted into the non–interacting Bose condensate is given by symmetrizing the product function $\Psi_N^{[0]}(\mathbf{r}_1, \ldots, \mathbf{r}_N)\phi(\mathbf{r}_{N+1})$ (equation 7.29).

*(a) Calculate the symmetrized initial state of the system with the injected particle. Show that the ratio of the probability that the new boson enters the ground state ($\psi_0$) is enhanced over that of its entering an empty state ($\psi_m$ for $0 < m < M$) by a factor $N + 1$.* (Hint: first do it for $N = 1$.)

So, if a macroscopic number of bosons are in one single–particle eigenstate, a new particle will be much more likely to add itself to this state than to any of the microscopically populated states.

Notice that nothing in your analysis depended on $\psi_0$ being the lowest energy state. If we started with a macroscopic number of particles in a single–particle state with wave-vector $\mathbf{k}$ (that is, a superfluid with a supercurrent in direction $\mathbf{k}$), new added particles, or particles scattered by inhomogeneities, will preferentially enter into that state. This is an alternative approach to understanding the persistence of supercurrents, complementary to the topological approach (exercise 9.4).

**Adding a photon to a laser beam.** In part (a), we saw that adding a boson to a single–particle eigenstate with $N$ existing bosons has a probability which is larger by a factor $N+1$ than adding a boson to an empty state. This chummy behavior between bosons is also the principle behind lasers.[42] If we think of an atom in an excited state, the photon it emits during its decay will prefer to join the laser beam than to go off into one of its other available modes. In this factor $N + 1$, the $N$ represents *stimulated emission*, where the existing electromagnetic field pulls out the energy from the excited atom, and the 1 represents *spontaneous emission* which occurs even in the absence of existing photons.

Imagine a single atom in an state with excitation energy energy $E$ and decay rate $\Gamma$, in a cubical box of volume $V$ with periodic boundary conditions for the photons. By the energy–time uncertainty principle, $\langle \Delta E \, \Delta t \rangle \geq \hbar/2$ the energy of the atom will be uncertain by an amount $\Delta E \propto \hbar\Gamma$. Assume for simplicity that, in a cubical box without pre–existing photons, the atom would decay at an equal rate into any mode in the range $E - \hbar\Gamma/2 < \hbar\omega < E + \hbar\Gamma/2$.

*(b) Assuming a large box and a small decay rate $\Gamma$, find a formula for the number of modes $M$ per unit volume $V$ competing for the photon emitted from our atom. Evaluate your formula for a laser with wavelength $\lambda = 619$ nm and the line-width $\Gamma = 10^4$ radians/sec.* (Hint: use the density of states, equation 7.65.)

Assume the laser is already in operation, so there are $N$ photons in the volume $V$ of the lasing material, all in one plane–wave state (a *single–mode* laser).

*(c) Using your result from part (a), give a formula for the number of photons per unit volume $N/V$ there must be in the lasing mode for the atom to have 50% likelihood of emitting into that mode.*

The main task in setting up a laser is providing a population of excited atoms! Amplification can occur if there is a *population inversion*, where the number of excited atoms is larger than the number of atoms in the lower energy state (clearly a non-equilibrium condition). This is made possible by *pumping* atoms in to the excited state by using one or two other single–particle eigenstates.

### (7.8) **Einstein's A and B** (Quantum, Mathematics)

Einstein deduced some basic facts about the interaction of light with matter very early in the development of quantum mechanics, by using statistical mechanics! In particular, he established that *stimulated emission* was demanded for statistical mechanical consistency, and found formulas determining the relative rates of absorption, spontaneous emission, and stimulated emission. (See [89, I.42-5]).

Consider a system consisting of non-interacting atoms weakly coupled to photons (electromagnetic radiation), in equilibrium at temperature $k_B T = 1/\beta$. The atoms have two energy eigenstates $E_1$ and $E_2$ with average populations $N_1$ and $N_2$: the relative population is given as

[41] For free particles in a cubical box of volume $V$, injecting a particle at the origin $\phi(\mathbf{r}) = \delta(\mathbf{r})$ would be a superposition of *all* plane–wave states of equal weight, $\delta(\mathbf{r}) = \frac{1}{V}\sum_{\mathbf{k}} e^{i\mathbf{k}\cdot x}$, appendix A. (In second–quantized notation, $a^\dagger(\mathbf{x} = 0) = \frac{1}{V}\sum_{\mathbf{k}} a^\dagger_{\mathbf{k}}$.) So, we "gently" add a particle at the origin by restricting this sum to low energy states. This is how quantum tunneling into condensed states (say, from Josephson junctions or scanning tunneling microscopes) is usually modeled.

[42] Laser is an acronym for Light Amplification by the Stimulated Emission of Radiation.

usual by the Boltzmann distribution

$$\left\langle \frac{N_2}{N_1} \right\rangle = e^{-\beta(E_2-E_1)}. \tag{7.83}$$

The energy density in the electromagnetic field is given by the Planck distribution (equation 7.67):

$$u(\omega) = \frac{\hbar}{\pi^2 c^3} \frac{\omega^3}{e^{\beta\hbar\omega} - 1}. \tag{7.84}$$

An atom in the ground state will absorb electromagnetic energy from the photons at a rate that is proportional to the energy density $u(\omega)$ at the excitation energy $\hbar\omega = E_2 - E_1$. Let us define this absorption rate per atom to be $2\pi B u(\omega)$.[43]

An atom in the excited state $E_2$, with no electromagnetic stimulation, will decay into the ground state with a rate $A$, emitting a photon. Einstein noted that neither $A$ nor $B$ should depend upon temperature.

Einstein argued that just these two rates would lead to an inconsistency.

*(a) Compute the long–time average ratio $N_2/N_1$ assuming only absorption and spontaneous emission. Even in the limit of weak coupling (small $A$ and $B$), show that this equation is incompatible with the statistical distributions [7.83] and [7.84].* (Hint: Write a formula for $dN_1/dt$, and set it equal to zero. Is $B/A$ temperature independent?)

Einstein fixed this by introducing stimulated emission. Roughly speaking, an atom experiencing an oscillating electromagnetic field is more likely to emit photons into that mode. Einstein found that the stimulated emission rate had to be a constant $2\pi B'$ times the energy density $u(\omega)$.

*(b) Write the equation for $dN_1/dt$, including absorption (a negative term) and spontaneous and stimulated emission from the population $N_2$. Assuming equilibrium, use this equation and equations 7.83 and 7.84 to solve for $B$, and $B'$ in terms of $A$.* These are generally termed the Einstein $A$ and $B$ coefficients.

Let's express the stimulated emission rate in terms of the number of excited photons per mode (see exercise 7.7b for an alternative derivation).

*(c) Show that the rate of decay of excited atoms $A + 2\pi B' u(\omega)$ is enhanced by a factor of $\langle n \rangle + 1$ over the zero temperature rate, where $\langle n \rangle$ is the expected number of photons in a mode at frequency $\hbar\omega = E_2 - E_1$.*

### (7.9) **Phonons and Photons are Bosons.** (Quantum)

Phonons and photons are the elementary, harmonic excitations of the elastic and electromagnetic fields. We've seen in 7.3 that phonons are decoupled harmonic oscillators, with a distribution of frequencies $\omega$. A similar analysis shows that the Hamiltonian of the electromagnetic field can be decomposed into harmonic normal modes called photons.

This exercise will explain why we think of phonons and photons as particles, instead of excitations of harmonic modes.

*(a) Show that the canonical partition function for a quantum harmonic oscillator of frequency $\omega$ is the same as the grand canonical partition function for bosons multiply filling a single state with energy $\hbar\omega$, with $\mu = 0$, up to a shift in the arbitrary zero of the total energy of the system.*

The Boltzmann filling of a harmonic oscillator is therefore the same as the Bose-Einstein filling of bosons into a single quantum state, except for an extra shift in the energy of $\hbar\omega/2$. This extra shift is called the *zero point energy*. The excitations within the harmonic oscillator are thus often considered particles with Bose statistics: the $n^{\text{th}}$ excitation is $n$ bosons occupying the oscillators quantum state.

This particle analogy becomes even more compelling for systems like phonons and photons where there are many harmonic oscillator states labeled by a wavevector $k$ (see exercise 7.3). Real, massive bose particles like $He^4$ in free space have single–particle quantum eigenstates with a dispersion relation $\epsilon_k = \hbar^2 k^2/2m$. Phonons and photons have one harmonic oscillator for every $k$, with an excitation energy $\epsilon_k = \hbar\omega_k$. If we treat them, as in part (a), as bosons filling these as single-particle states we find that they are completely analogous to ordinary massive particles. The only difference is that the relation between energy and wave-vector (called the *dispersion relation*) is different: for photons, $\epsilon_k = \hbar\omega_k = \hbar c|k|$.[44]

*(b) Do phonons or photons Bose condense at low temperatures? Can you see why not? Can you think of a*

[43] The literature uses $u_{\text{cycles}}(f)$ where $f = \omega/2\pi$ is in cycles per second, and has no factor of $2\pi$. Since $u_{\text{cycles}}(f)df = u(\omega)d\omega$, the absorption rate $Bu_{\text{cycles}}(f) = Bu(\omega)d\omega/df = 2\pi B u(\omega)$.

[44] If massive particles are moving fast, their energies are $\epsilon = \sqrt{m^2c^4 - \mathbf{p}^2c^2}$. This formula reduces to $p^2/2m + mc^2 = \hbar^2k^2/2m + mc^2$ if the kinetic energy is small compared to the rest mass $mc^2$. For massless particles, $\epsilon = |\mathbf{b}|c = \hbar|\mathbf{k}|c$, precisely the relation we find for photons (and for phonons at low frequencies). So actually even the dispersion relation is the same: photons and phonons are massless bosons.

*non–equilibrium Bose condensation of photons, where a macroscopic occupation of a single frequency and momentum state occurs?*

Be careful not to get confused when we put real, massive bosons into a harmonic oscillator potential (exercise 7.11). There it is best to think of each harmonic oscillator as being many separate eigenstates being filled by the atoms.

(7.10) **Bose Condensation in a Band.** (Basic, Quantum)

The density of states $g(E)$ of a system of non-interacting bosons forms a band: the single-particle eigenstates are confined to an energy range $E_{min} < E < E_{max}$, so $g(E)$ is non-zero in this range and zero otherwise. The system is filled with a finite density of bosons. Which of the following is necessary for the system to undergo Bose condensation at low temperatures?

*(a)* $g(E)/(e^{\beta(E-E_{min})}+1)$ *is finite as* $E \to E_{min}^{-}$.

*(b)* $g(E)/(e^{\beta(E-E_{min})}-1)$ *is finite as* $E \to E_{min}^{-}$.

*(c)* $E_{min} \geq 0$.

*(d)* $\int_{E_{min}}^{E} g(E')/(E'-E_{min})\,dE'$ *is a convergent integral at the lower limit* $E_{min}$.

*(e) Bose condensation cannot occur in a system whose states are confined to an energy band.*

(7.11) **Bose Condensation in a Parabolic Potential.** (Quantum)

[45]

Wieman and Cornell in 1995 were able to get a dilute gas of rubidium-87 atoms to Bose condense [4].

*(a) Is rubidium-87 (37 protons and electrons, 50 neutrons) a boson or a fermion?*

*(b) At their quoted maximum number density of* $2.5 \times 10^{12}/cm^3$*, at what temperature* $T_c^{\text{predict}}$ *do you expect the onset of Bose condensation in free space? They claim that they found Bose condensation starting at a temperature of* $T_c^{\text{measured}} = 170nK$*. Is that above or below your estimate?* (Useful constants: $h = 6.6262 \times 10^{-27}$ erg sec, $m_n \sim m_p = 1.6726 \times 10^{-24}$ gm, $k_B = 1.3807 \times 10^{-16}$ erg/K.)

The trap had an effective potential energy that was harmonic in the three directions, but anisotropic with cylindrical symmetry. The frequency along the cylindrical axis was $f_0$ =120Hz so $\omega_0 \sim$ 750Hz, and the two other frequencies were smaller by a factor of $\sqrt{8}$: $\omega_1 \sim$ 265Hz. The Bose condensation was observed by abruptly removing the trap potential,[46] and letting the gas atoms spread out: the spreading cloud was imaged 60ms later by shining a laser on them and using a CCD to image the shadow.

For your convenience, the ground state of a particle of mass $m$ in a one-dimensional harmonic oscillator with frequency $\omega$ is $\psi_0(x) = \left(\frac{m\omega}{\pi\hbar}\right)^{1/4} e^{-m\omega x^2/2\hbar}$, and the momentum-space wave function is $\tilde{\psi}_0(k) = \left(\frac{\hbar}{\pi m\omega}\right)^{1/4} e^{-\hbar k^2/2m\omega}$.

*(c) What is the ground-state wave-function for one rubidium-87 atom in this potential? What is the wavefunction in momentum space? The probability distribution of the momentum? What is the ratio of the velocity widths along the axis and perpendicular to the axis for the ground state? For the classical thermal distribution of velocities? If the potential is abruptly removed, what will the shape of the distribution of positions look like 60ms later, (ignoring the small width of the initial distribution in space)? Compare your predicted anisotropy to the false-color images above. If the x axis goes mostly right and a bit up, and the y axis goes mostly up and a bit left, which axis corresponds to the axial frequency and which corresponds to one of the two lower frequencies?*

Their Bose condensation isn't in free space: the atoms are in a harmonic oscillator potential. In the calculation in free space, we approximated the quantum states as a continuum density of states $g(E)$. That's only sensible if $k_BT$ is large compared to the level spacing near the ground state.

*(d) Compare* $\hbar\omega$ *to* $k_BT$ *at the Bose condensation point* $T_c^{\text{measured}}$ *in their experiment.*

For bosons in a one-dimensional harmonic oscillator of frequency $\omega_0$, it's clear that $g(E) = 1/(\hbar\omega_0)$: the number of states in a small range $\Delta E$ is the number of $\hbar\omega_0$'s it contains.

*(e) Compute the density of states*

$$g(E) = \int_0^\infty d\epsilon_1 d\epsilon_2 d\epsilon_3\, g_1(\epsilon_1) g_2(\epsilon_2) g_3(\epsilon_3) \\ \delta\left(E - (\epsilon_1 + \epsilon_2 + \epsilon_3)\right) \tag{7.85}$$

*for a three-dimensional harmonic oscillator, with one frequency* $\omega_0$ *and two of frequency* $\omega_1$*. Show that it's equal to* $1/\Delta$ *times the number of states in* $\vec{\epsilon}$ *space between energies* $E$ *and* $E+\Delta$*. Why is this triangular slab not of thickness* $\Delta$*?*

[45] "Observation of Bose-Einstein Condensation in a Dilute Atomic Vapor", M.H. Anderson, J.R. Ensher, M.R. Matthews, C.E. Wieman, and E.A. Cornell, *Science* **269**, 198 (1995). http://jilawww.colorado.edu/bec/.

[46] Actually, they first slowly reduced it by a factor of 75 and then abruptly reduced it from there; I'm not sure why, but let's ignore that complication.

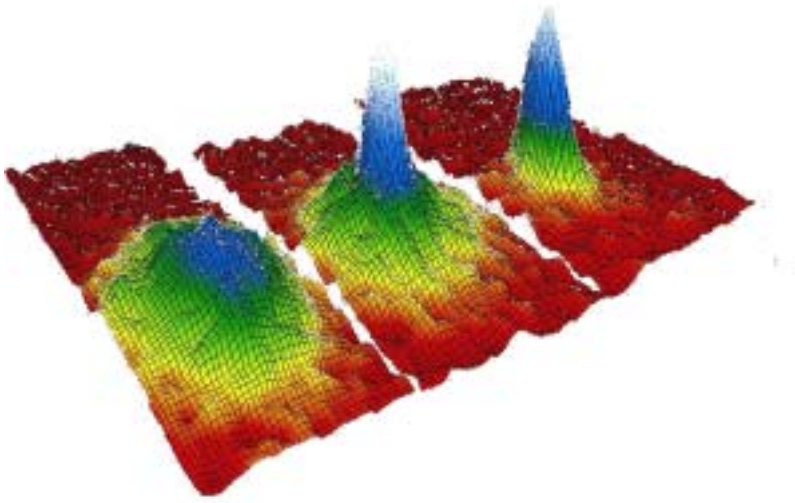


**Fig. 7.15 Bose-Einstein Condensation** at 400, 200, and 50 nano-Kelvin, from reference [4]. The pictures are spatial distributions 60ms after the potential is removed; the field of view of each image is $200\mu m \times 270\mu m$. The left picture is roughly spherically symmetric, and is taken before Bose condensation; the middle has an elliptical Bose condensate superimposed on the spherical thermal background; the right picture is nearly pure condensate. I believe this may not be the same experiment as described in their original paper.

Their experiment has $N = 2 \times 10^4$ atoms in the trap as it condenses.

*(f) By working in analogy with the calculation in free space, find the maximum number of atoms that can occupy the three-dimensional harmonic oscillator potential in part (e) without Bose condensation at temperature $T$. (You'll want to know $\int_0^\infty z^2/(e^z - 1)\, dz = 2\,\zeta(3) = 2.40411$.) According to your calculation, at what temperature $T_c^{\rm HO}$ should the real experimental trap have Bose condensed?*

(7.12) **Light Emission and Absorption.** (Quantum, Basic)

The experiment that Planck was studying did not directly measure the energy density per unit frequency, equation 7.67 inside a box. It measured the energy radiating out of a small hole, of area $A$. Let us assume the hole is on the upper face of the cavity, perpendicular to the $z$ axis.

What is the photon distribution just inside the boundary of the hole? Clearly there are few photons coming into the hole from the outside, so the distribution is depleted for those photons with $v_z < 0$. However, the photons with $v_z > 0$ to an excellent approximation should be unaffected by the hole – since they were emitted from far distant walls of the cavity, where the existence of the hole is a negligible perturbation. So, presuming the relevant photons just inside the hole are distributed in the same way as in the box as a whole (equation 7.67), how many leave in a time $dt$?

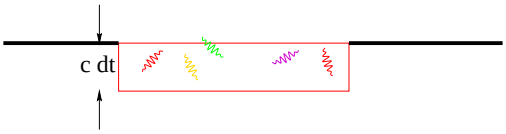



**Fig. 7.16** The photons leaving a cavity in a time $dt$ are those within $v_z dt$ of the hole.

As one can see geometrically (figure 7.16), those photons within $v_z dt$ of the boundary will escape in time $dt$. The vertical velocity $v_z = c\cos(\theta)$, where $\theta$ is the photon velocity angle with respect to the vertical. The Planck distribution is isotropic, so the probability that a photon will be moving at an angle $\theta$ is the perimeter of the $\theta$ circle on the sphere divided by the area of the sphere, $\frac{2\pi \sin(\theta)\, d\theta}{4\pi} = \frac{1}{2}\sin(\theta)\, d\theta$.

*(a) Show that the probability density*[47] $\rho(v_z)$ *for a particular photon to have velocity* $v_z$ *is independent of* $v_z$ *in the range* $(-c, c)$*, and thus is* $\frac{1}{2c}$. (Hint: $\rho(v_z)\,\Delta v_z = \rho(\theta)\,\Delta\theta$.)

Clearly, an upper bound on the energy emitted from a hole of area $A$ is given by the energy in the box as a whole (eq. 7.67) times the fraction $\frac{Ac\,dt}{V}$ of the volume within $c\,dt$ of the hole.

*(b) Show that the actual energy emitted is* 1/4 *of this upper bound.* (Hint: You'll need to integrate $\int_0^c \rho(v_z) v_z\,dv_z$.)

Hence the power per unit area emitted from the small hole in equilibrium is

$$P_{\text{black}}(\omega, T) = \left(\frac{c}{4}\right) \frac{\hbar}{\pi^2 c^3} \frac{\omega^3\,d\omega}{e^{\hbar\omega/k_B T} - 1}. \tag{7.86}$$

Why is this called black–body radiation? Certainly a small hole in a large (cold) cavity looks black – any light entering the hole bounces around inside until it is absorbed by the walls. Suppose we placed a black object – a material that absorbed radiation at all frequencies and angles – capping the hole. This object would absorb radiation from the cavity, rising in temperature until it came to equilibrium with the cavity – emitting just as much radiation as it absorbs. Thus the overall power per unit area emitted by our black object in equilibrium at a given temperature must equal that of the hole. This must also be true if we place a selective filter between the hole and our black body, passing through only particular types of photons. Thus the emission and absorption of our black body must agree with the hole for every photon mode individually, an example of the *principle of detailed balance* we will discuss in more detail in section 8.3.

How much power per unit area $P_{\text{colored}}(\omega, T)$ is emitted in equilibrium at temperature $T$ by a red or maroon body? A white body? A mirror? These objects are different in the fraction of incident light they absorb at different frequencies and angles $a(\omega, \theta)$. We can again use the principle of detailed balance, by placing our colored object next to a black body and matching the power emitted and absorbed for each angle and frequency:

$$P_{\text{colored}}(\omega, T, \theta) = P_{\text{black}}(\omega, T) a(\omega, \theta) \tag{7.87}$$

Finally, we should calculate $Q_{\text{tot}}(T)$, the total power per unit area emitted from a black body at temperature $T$, by integrating 7.86 over frequency.

*(c) Using the fact that* $\int_0^\infty x^3/(e^x - 1)\,dx = \pi^4/15$*, show that*

$$Q_{\text{tot}}(T) = \int_0^\infty P_{\text{black}}(\omega, T)\,d\omega = \sigma T^4. \tag{7.88}$$

*and give a formula for the* Stefan–Boltzmann *constant* $\sigma$. The value is $\sigma = 5.67 \times 10^{-5}$ erg cm$^{-2}$ K$^{-4}$ s$^{-1}$. (Hint: use this to check your answer.)

(7.13) **Fermions in Semiconductors.** (Quantum)

Let's consider a caricature model of a doped semiconductor [8, ch. 28]. Consider a crystal of phosphorous-doped silicon, with $N - M$ atoms of silicon and $M$ atoms of phosphorous. Each silicon atom contributes one electron to the system, and has two states at energies $\pm\Delta/2$, where $\Delta = 1.16$eV is the energy gap. Each phosphorous atom contributes *two* electrons and two states, one at $-\Delta/2$ and the other at $\Delta/2 - \epsilon$, where $\epsilon = 0.044$eV is much smaller than the gap.[48] (Our model ignores the quantum mechanical hopping between atoms that broadens the levels at $\pm\Delta/2$ into the conduction band and the valence band. It also ignores spin and chemistry: each silicon really contributes four electrons and four levels, and each phosphorous five electrons and four levels.) To summarize, our system has $N + M$ spinless electrons (maximum of one electron per state), $N$ valence band states at energy $-\Delta/2$, $M$ impurity band states at energy $\Delta/2 - \epsilon$, and $N - M$ conduction band states at energy $\Delta/2$.

*(a) Derive a formula for the number of electrons as a function of temperature* $T$ *and chemical potential* $\mu$ *for the energy levels of our system.*

*(b) What is the limiting occupation probability for the states as* $T \to \infty$*, where entropy is maximized and all states are equally likely? Using this, find a formula for* $\mu(T)$ *valid at large* $T$*, not involving* $\Delta$ *or* $\epsilon$*.*

*(c) Draw an energy level diagram showing the filled and empty states at* $T = 0$*. Find a formula for* $\mu(T)$ *in the low temperature limit* $T \to 0$*, not involving the variable* $T$*.* (Hint: Balance the number of holes in the impurity band with the number of electrons in the conduction band. Why can you ignore the valence band?)

*(d) In a one centimeter cubed sample, there are* $M = 10^{16}$ *phosphorous atoms; silicon has about* $N = 5 \times 10^{22}$ *atoms per cubic centimeter. Find* $\mu$ *at room temperature (1/40*

[47] We're being sloppy again, using the same name $\rho$ for the probability densities per unit velocity and per unit angle.

[48] The phosphorous atom is neutral when both of its states are filled: the upper state can be thought of as an electron bound to a phosphorous positive ion. The energy shift $\epsilon$ represents the Coulomb attraction of the electron to the phosphorous ion: it's small because the dielectric constant is large (see A&M above).

*eV) from the formula you derived in part (a). (Probably trying various $\mu$ is easiest: set up a program on your calculator or computer.) At this temperature, what fraction of the phosphorous atoms are ionized (have their upper energy state empty)? What is the density of holes (empty states at energy $-\Delta/2$)?*

Phosphorous is an *electron donor*, and our sample is doped $n$-type, since the dominant carriers are electrons: $p$-type semiconductors are doped with holes.

(7.14) **White Dwarves, Neutron Stars, and Black Holes.** (Astrophysics,Quantum)

As the energy sources in large stars are consumed, and the temperature approaches zero, the final state is determined by the competition between gravity and the chemical or nuclear energy needed to compress the material.

A simplified model of ordinary stellar matter is a Fermi sea of non-interacting electrons, with enough nuclei to balance the charge. Let's model a white dwarf (or black dwarf, since we assume zero temperature) as a uniform density of $He^4$ nuclei and a compensating uniform density of electrons. Assume Newtonian gravity. Assume the chemical energy is given solely by the energy of a gas of non-interacting electrons (filling the levels to the Fermi energy).

*(a) Assuming non-relativistic electrons, calculate the energy of a sphere with $N$ zero-temperature non-interacting electrons and radius $R$.*[49] *Calculate the Newtonian gravitational energy of a sphere of $He^4$ nuclei of equal and opposite charge density. At what radius is the total energy minimized?*

A more detailed version of this model was studied by Chandrasekhar and others as a model for white dwarf stars. Useful numbers: $m_p = 1.6726 \times 10^{-24}$ gm, $m_n = 1.6749 \times 10^{-24}$ gm, $m_e = 9.1095 \times 10^{-28}$ gm, $\hbar = 1.05459 \times 10^{-27}$ erg sec, $G = 6.672 \times 10^{-8}$ cm$^3$/(gm s$^2$), 1 eV $= 1.60219 \times 10^{-12}$ erg, $k_B = 1.3807 \times 10^{-16}$ erg / K, and $c = 3 \times 10^{10}$ cm/s.

*(b) Using the non-relativistic model in part (a), calculate the Fermi energy of the electrons in a white dwarf star of the mass of the Sun, $2 \times 10^{33}$ gm, assuming that it is composed of helium. (i) Compare it to a typical chemical binding energy of an atom. Are we justified in ignoring the electron-electron and electron-nuclear interactions (i.e., chemistry)? (ii) Compare it to the temperature inside the star, say $10^7 K$. Are we justified in assuming that the electron gas is degenerate (roughly zero temperature)? (iii) Compare it to the mass of the electron. Are we roughly justified in using a non-relativistic theory? (iv) Compare it to the mass difference between a proton and a neutron.*

The electrons in large white dwarf stars are relativistic. This leads to an energy which grows more slowly with radius, and eventually to an upper bound on their mass.

*(c) Assuming extremely relativistic electrons with $\epsilon = pc$, calculate the energy of a sphere of non-interacting electrons. Notice that this energy cannot balance against the gravitational energy of the nuclei except for a special value of the mass, $M_0$. Calculate $M_0$. How does your $M_0$ compare with the mass of the Sun, above?*

A star with mass larger than $M_0$ continues to shrink as it cools. The electrons (note *(b.iv)* above) combine with the protons, staying at a constant density as the star shrinks into a ball of almost pure neutrons (a *neutron star*, often forming a *pulsar* because of trapped magnetic flux). Recent speculations [85] suggests that the "neutronium" will further transform into a kind of quark soup with many strange quarks, forming a transparent insulating material.

For an even higher mass, the Fermi repulsion between quarks can't survive the gravitational pressure (the quarks become relativistic), and the star collapses into a black hole. At these masses, general relativity is important, going beyond the purview of this course. But the basic competition, between degeneracy pressure and gravity, is the same.

[49] You may assume that the single–particle eigenstates have the same energies and **k**–space density in a sphere of volume $V$ as they do for a cube of volume $V$: just like fixed versus periodic boundary conditions, the boundary doesn't matter to bulk properties.

# Calculation and Computation

# 8

Most statistical mechanical systems cannot be solved using paper and pencil. Statistical mechanics provides general relationships and organizing principles (temperature, entropy, free energy, thermodynamic relations) even when a solution is not available. But there are times when specific answers about specific models or experiments are needed.

There are two basic tools for extracting answers out of statistical mechanics for realistic systems. The first is to use *perturbation theory*, adding corrections to a simple, solvable model, or to a complex model in a limit (like low or high temperatures) where its properties are possible to calculate. Section 8.1 discusses the deep connection between the convergence of perturbation theory and the existence of *phases*.

The second tool is *simulation*. Sometimes one simply simulates the microscopic theory. For example, a molecular dynamics simulation will move the atoms according to Newton's laws. To emulate a microscopic system connected to a heat bath, one can add friction and noise to the microscopic theory, in the correct proportions so as to lead to proper thermal equilibration.[1] We will not be discussing such microscopic simultions in detail here, but see exercise 10.5.

If one is not interested in the detailed dynamical trajectories of the system, one can use *Monte-Carlo* simulation methods to extract the equilibrium properties from a model. We introduce these methods in section 8.2 in the context of the *Ising model*, the most well-studied of the lattice models in statistical mechanics. The theory underlying the Monte-Carlo method is the mathematics of *Markov Chains*, which we discuss in section 8.3: using more general Markov chains allows for the development of faster algorithms for reaching equilibrium (exercise 8.5).

[1] For the diffusion equation, the mobility (friction) $\gamma$ is related to the diffusion constant (noise) $D$ by the Einstein relation 2.22 $D/\gamma = k_B T$. We will discuss these relations in more general contexts in section 10.5.

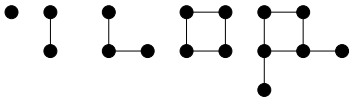


**Fig. 8.1** The high and low temperature expansions for the Ising and other lattice models involve terms which correspond to stick–figure Feynman diagrams on the lattice. The first diagram on the left (the single point) gives the Curie law at high temperatures, and the energy–gap formula at low temperatures that you derive in exercise 8.1.

## 8.1 What is a Phase? Perturbation theory.

Much of a typical advanced course in statistical mechanics will be devoted to perturbation theory. Lattice theories at high and low temperatures $T$ have perturbative expansions in powers of $T$ and $1/T$, with Feynman diagrams involving all ways of drawing stick figures on lattice

points (figure 8.1).[2] Gases at high temperatures and low densities have *virial expansions*. Metals at low temperatures have *Fermi liquid theory*, where the electron–electron interactions are perturbatively incorporated by dressing the electrons into quasiparticles. Properties of systems near continuous phase transitions can be explored by perturbing in the dimension of space, giving the $\epsilon$-*expansion*. Let's illustrate these expansions in the particular case of the ideal Bose and Fermi gases, and then discuss some basic issues about phases and perturbation theory.

In figure 8.2 we plot the specific heat of the ideal Fermi gas, which we studied in section 7.7.[3] At high temperatures, the specific heat per particle goes to $\frac{3}{2}k_B$, as we derived for the classical ideal gas. One can do perturbation theory valid at high temperatures to get the virial expansion corrections to this classical result. The natural expansion parameter is the number of particles per de Broglie volume $\Delta = \frac{\lambda^3 N}{gV}$ where $\lambda = \frac{h}{\sqrt{2\pi m k_B T}}$ is the thermal deBroglie wavelength from equation 3.63 and $g = 2$ is the number of spin states of our Fermions:

$$C_v^{\text{fermion}} = \tfrac{3}{2}Nk_B\left(1 - 0.0884\Delta + 0.0066\Delta^2 - 0.0004\Delta^3 + \dots\right). \quad (8.2)$$

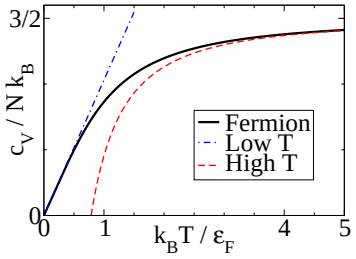



**Fig. 8.2** The specific heat of the ideal Fermi gas, along with the low–temperature 8.3 expansion and the high–temperature 8.2 expansion kept to first order in $\Delta = \frac{4}{3\sqrt{\pi}}\left(\frac{\epsilon_F}{k_B T}\right)^{3/2}$.

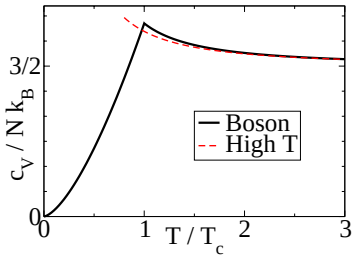



**Fig. 8.3** The specific heat of the ideal Bose gas, along with the high–temperature expansion 8.4 to first order in $\Delta = \zeta(3/2)(T_c/T)^{3/2}$. Notice the cusp at the Bose condensation temperature $T_c$.

At low temperatures, the specific heat of the Fermi gas vanishes linearly in the temperature. At low temperatures one works with the *elementary excitations*, which here are excitations of particles and holes away from the Fermi surface.[4] At low temperatures we can again do perturbation theory, giving us the *Sommerfeld expansion*, whose first term for the specific heat is

$$C_v^{\text{fermion}} = \frac{\pi^2}{2}k_B^2 T N g(\epsilon_F) + \cdots = \frac{\pi^2}{2}\frac{Nk_B^2 T}{\epsilon_F} + \dots \quad (8.3)$$

In figure 8.3 we plot the specific heat of the ideal Bose gas, which we studied in section 7.6.3. At high temperatures we again have a virial expansion

$$C_v^{\text{boson},T>T_c} = \tfrac{3}{2}Nk_B\left(1 + 0.0884\Delta + 0.0066\Delta^2 + 0.0004\Delta^3 + \dots\right), \quad (8.4)$$

[2] The low–temperature expansion of the magnetization per spin of the three–dimensional Ising model on a cubic lattice starts out

$$\begin{aligned} m =& 1 - 2t^3 - 12t^5 + 14t^6 - 90t^7 + 192t^8 - 492t^9 + 2148t^{10} - 7716t^{11} \\ &+ 23262t^{12} - 79512t^{13} + 252054t^{14} - 846628t^{15} + 2753520t^{16} \\ &- 9205800t^{17} + 30371124t^{18} - 101585544t^{19} + 338095596t^{20} + \dots \end{aligned} \quad (8.1)$$

(references [82, 27]). where $t = e^{-4J/k_B T}$. The mean magnetization per spin is zero above $T_c$: the magnetization has a singularity (no Taylor series) at the critical point.

[3] This section does discuss results from quantum mechanical calculations, but the ideas are relevant to all phases.

[4] Only states within $k_B T$ of $\epsilon_F$ are thermally excited (figure 7.5) so the energy shift is roughly the number of excitable states $Ng(\epsilon_F)k_B T$ (equation 7.74) times their excitation energy $k_B T$, or $E(T) - E(0) \sim Nk_B^2 T^2 g(\epsilon_F)$. This gives a specific heat of $2Nk_B^2 T/\epsilon_F$, linear in temperature. Indeed, the coefficient we get from this rough argument is only a factor of $\pi^2/4$ away from the correct first–order term.

where here $g = 1$ for spinless bosons. At low temperatures, the behavior of the ideal Bose gas is particularly simple: the specific heat is exactly

$$C_v^{\text{boson},T<T_c} = \frac{15}{4}\zeta(5/2)/\Delta, \tag{8.5}$$

What is the *fundamental* difference between the ideal Fermi and Bose gases? *Only one has a phase transition.* For the Fermi gas, perturbation theory links the high and low temperature regions. For the Bose gas, the extrapolation from high temperatures does not describe the behavior below the critical point $T_c$. The Bose gas has two distinct regions: the normal phase and the condensed phase.

What is a phase? *Perturbation theory works inside phases.* More precisely, inside a phase the properties are analytic (have convergent Taylor expansions) as functions of the external conditions. If the system is inside a phase at temperature $T_0$, the specific heat per particle (for example) will have a Taylor series $c_v(T - T_0) = \sum a_n(T - T_0)^n$ which converges in a region about $T_0$.[5] Inside a phase, the specific heat, the pressure $P$, the density $\rho$, the elastic constants $c_{ijk\ell}$ ... *all* properties are smooth functions (with convergent power series) of all external parameters. Phase boundaries occur at parameter values where the properties aren't smooth – where the continuation of the properties on one side does not predict the behavior on the other. We could almost define phases as regions where perturbation theory works – except for the awkward problem that we don't want liquids and gases to be called part of the same fluid 'phase', even though they are connected by paths going around the critical point (figure 8.7).[6]

This leads to an important experimental method. Suppose you've invented a new exotic liquid crystal. How can you tell if it is in an already known phase? You look for an experimental path, mixing materials and changing external conditions, for smoothly changing your phase to the known one. For example, are oil and water both in the same (fluid) phase? Can we go from one to the other smoothly, without passing through a phase transition?[7] You can't mix oil and water, but you can mix oil and alcohol, and certainly can mix alcohol and water. Changing the concentrations smoothly starting from oil, going through pure alcohol, and ending at water demonstrates that these two fluids are part of the same phase. (see figure 8.4). This argument is also the basis for much of theoretical physics: if you can go smoothly from A (your theory) to B (the experiment) by adding corrections, then A and B are in the same phase: publish![8]

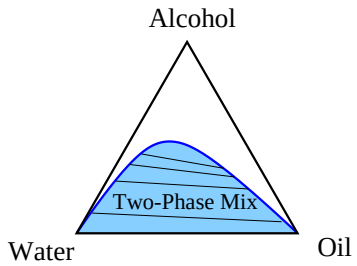



**Fig. 8.4** A schematic ternary phase diagram for mixtures of oil, water, and alcohol. Each point in the triangle represents a mixture of percentages of the three, with the corners being pure water, oil, and alcohol. The shaded region shows where phase separation occurs: relative concentrations in the shaded region will separate into a two–phase mixture given by the endpoints of the tie–line passing through that point. Oil and water basically don't dissolve in one another: a mixture of the two will separate into the two separate fluids. You can go smoothly from one to the other, though, by first adding alcohol.

[5] The simplest and most useful series are usually expanding around the boundaries of phases (infinite temperature, zero temperature, zero force, small field ...). Some of these perturbation series have zero radius of convergence for that reason: they are *asymptotic expansions* (see exercise 1.4). For example, Hooke's law in elastic theory is the first term in a nonlinear elastic series, which at non-zero temperature has zero radius of convergence [16, 17].

[6] You can also go from the ferromagnetic phase to the paramagnetic phase in the Ising model, by first turning on an external field $H$, then raising the temperature to $T > T_c$, and then lowering the field back to zero (figure 8.8). Introducing this field is clearly cheating, because it breaks the same up–down symmetry that the Ising model spontaneously breaks at low temperatures. Going from liquids to gases, you don't even need to cheat.

[7] This process is sometimes called *adiabatic continuity*. Phases can also be thought of as universality classes for attracting fixed points: see chapter 12.

[8] Some unperturbed theories are better than others, even if they are in the same phase. The BCS theory is deemed the correct theory of superconductors, despite the fact that (earlier) theories of Bose condensation of pre-formed electron pairs are not separated from BCS theory by a phase transition. The Cooper pairs in most superconductors are large compared to their separation, so they overlap many other pairs. The BCS theory uses this fact to do calculations that are almost exact for many systems, without including perturbative corrections.

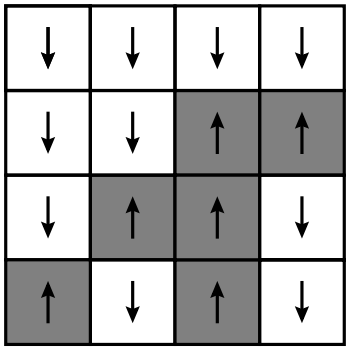


**Fig. 8.5** The 2D square-lattice Ising model. It is traditional to denote the values $s_i = \pm 1$ as up and down, or as two different colors.

[9] Ising's name is pronounced "Eesing", but the model is usually pronounced "Eyesing" with a long I sound.

[10] Unlike a true quantum spin 1/2 particle there are no terms in the Ising Hamiltonian that lead to superpositions of states with different spins.

[11] We shall use boldface $\mathbf{M}$ to denote the total magnetization, and (especially in the exercises) will also refer to $M = \mathbf{M}/N$, the average magnetization per spin.

[12] Thus $\sigma_j^z$ commutes with $\mathcal{H}$, and every spin configuration is an energy eigenstate. The Ising model (equation 8.6) can be interpreted as a classical Hamiltonian with discrete states, as a free energy for a system with discrete states (as for the binary alloy in section 8.2.2), or as a fully quantum Hamiltonian for particles of spin ½ coupling only via $\sigma^z$. *Transverse* Ising models have couplings including other components of $\sigma$, and are challenging quantum systems.

[14] "Ferromagnetic" is named after iron (Fe), which is the most common material which has a spontaneous magnetization.

## 8.2 The Ising Model

Lattice models are a big industry within statistical mechanics. Placing some degrees of freedom on each site of a regular grid and forming a Hamiltonian or a dynamical evolution law to equilibrate or evolve the resulting system forms a centerpiece of computational statistical mechanics (as well as the focus of much of the analytical work). Critical phenomena and phase transitions [12], lattice QCD and quantum field theories, quantum magnetism and models for high temperature superconductors, phase diagrams for alloys [8.2.2], the behavior of systems with dirt or disorder, and non–equilibrium systems exhibiting avalanches and crackling noise [12], all make important use of lattice models.

In this section, we will introduce the most studied of these models, the Ising model.[9]

The Ising model is a lattice of sites $i$ with a single, two–state degree of freedom $s_i$ on each site that may take values $\pm 1$. This degree of freedom is normally called a 'spin'.[10] We will be primarily interested in the Ising model in two dimensions on a square lattice, see figure 8.5. The Hamiltonian for the Ising model is

$$\mathcal{H} = -\sum_{\langle ij \rangle} J s_i s_j - H \sum_i s_i. \tag{8.6}$$

Here the sum $\langle ij \rangle$ is over all pairs of spins on nearest-neighbor sites, and $J$ is the *coupling* between these neighboring spins. (There are four neighbors per spin on the 2D square lattice.) Usually one refers to $H$ as the external field, and the sum $\mathbf{M} = \sum_i s_i$ as the magnetization, in reference to the Ising model's original application to magnetic systems.[11] We'll usually assume the model has $N$ spins forming a square, with periodic boundary conditions.

### 8.2.1 Magnetism

As a model for magnetism, our spin $s_i = 2\sigma_i^z$, the z-component of the net spin of a spin 1/2 atom in a crystal.[12] The interactions between spins is $J s_i s_j = 4J\sigma_i^z \sigma_j^z$.[13] The coupling of the spin to the external magnetic field is microscopically $-H s_i = gH \cdot \sigma_i^z$, where $g$ is the gyromagnetic ratio for the spin (close to two for the electron).

The energy of two spins $-J s_i s_j$ is $-J$ if the spins are parallel, and $+J$ if they are antiparallel. Thus if $J > 0$ the model favors parallel spins: we say that the interaction is *ferromagnetic*.[14] At low temperatures, the spins will organize themselves to mostly point up or down, forming a *ferromagnetic phase*. As the temperature approaches zero, all the spins in the Ising model become parallel and the magnetization

[13] The interaction between spins is usually better approximated by the dot product $\sigma_i \cdot \sigma_j = \sigma_i^x \sigma_i^x + \sigma_i^y \sigma_i^y + \sigma_i^z \sigma_i^z$, used in the more realistic Heisenberg model. Some materials have anisotropic crystal structures which make the Ising model at least approximately valid.

per spin will approach $M = \pm 1$. If $J < 0$ we call the interaction *antiferromagnetic*; for our square lattice the spins will tend to align in a checkerboard fashion at low temperatures (an *antiferromagnetic* phase). At high temperatures, we expect entropy to dominate; the spins will fluctuate wildly in a *paramagnetic phase* and the magnetization per spin $M$ for a large system will be near zero.

### 8.2.2 Binary Alloys

The Ising model is quite a convincing model for binary alloys.[15] Imagine a square lattice of atoms, which can be either of type $A$ or $B$ (figure 8.6).[16] We set the spin values $A = +1$ and $B = -1$. Let the number of the two kinds of atoms be $N_A$ and $N_B$, with $N_A + N_B = N$. Let the interaction energy between two neighboring atoms be $E_{AA}$, $E_{BB}$, and $E_{AB}$; these can be thought of as the bond strength – the energy needed to break the bond. Let the total number of $AA$ nearest–neighbor bonds be $N_{AA}$, and similarly for $N_{BB}$ and $N_{AB}$. Then the Hamiltonian for our binary alloy is

$$\mathcal{H}_{\text{binary}} = -E_{AA}N_{AA} - E_{BB}N_{BB} - E_{AB}N_{AB}. \tag{8.7}$$

How is this the Ising model? Let's start by adding a constant $-CN$ to the Ising model, and plugging in our new variables:

$$\begin{aligned}\mathcal{H}_{\text{ising}} &= -J\sum_{\langle ij\rangle} s_i s_j - H\sum_i s_i - CN \\ &= -J\left(N_{AA} + N_{BB} - N_{AB}\right) - H\left(N_A - N_B\right) - CN,\end{aligned} \tag{8.8}$$

since $N_A - N_B = \mathbf{M}$ the sum of the spins, $N_{AA} + N_{BB}$ is the number of parallel neighbors, and $N_{AB}$ is the number of antiparallel neighbors. There are two bonds per spin, so $N_{AA}+N_{BB}+N_{AB} = 2N$; we substitute $N = \frac{1}{2}(N_{AA}+N_{BB}+N_{AB})$. For every $A$ atom there must be four bonds ending with an $A$, and similarly for every $B$ atom there must be four bonds ending with a $B$. Each $AA$ bond gives half an $A$ atom worth of 'bond ends', and each $AB$ bond gives a quarter, so

$$\begin{aligned}N_A &= \tfrac{1}{2}N_{AA} + \tfrac{1}{4}N_{AB} \qquad \text{and similarly} \\ N_B &= \tfrac{1}{2}N_{BB} + \tfrac{1}{4}N_{AB}\end{aligned} \tag{8.9}$$

and we may substitute $N_A - N_B = \frac{1}{2}(N_{AA} - N_{BB})$. We now find

$$\begin{aligned}\mathcal{H}_{\text{ising}} &= -J\left(N_{AA} + N_{BB} - N_{AB}\right) - H\left(\tfrac{1}{2}(N_{AA} - N_{BB})\right) \\ &\quad - C\left(\tfrac{1}{2}(N_{AA} + N_{BB} + N_{AB})\right) \\ &= -(J + \tfrac{1}{2}H + \tfrac{1}{2}C)N_{AA} - (J - \tfrac{1}{2}H + \tfrac{1}{2}C)N_{BB} \\ &\quad - (-J + \tfrac{1}{2}C)N_{AB}.\end{aligned} \tag{8.10}$$

This is just of the form of the binary alloy Hamiltonian 8.7, with $J = \frac{1}{4}(E_{AA} + E_{BB} - 2E_{AB})$, $H = E_{AA} - E_{BB}$, and $C = \frac{1}{2}(E_{AA} + E_{BB} + 2E_{AB})$.

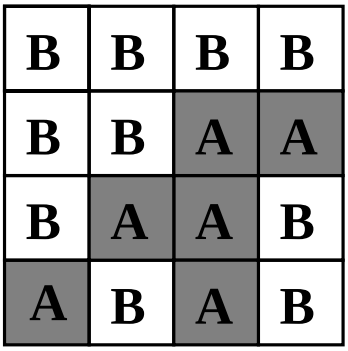



**Fig. 8.6** The Ising model as a binary alloy. Atoms in crystals naturally sit on a regular grid: alloys have more than one type of element which can sit on the lattice sites (here, types $A$ and $B$).

[15] Indeed, any classical system on a lattice with local interactions can be mapped onto an Ising–like model.

[16] A realistic alloy might mix roughly half copper and half zinc to make $\beta$-brass. At low temperatures, the copper and zinc atoms sit each on a cubic lattice, with the zincs in the middle of the copper cubes, together forming a body–centered cubic (bcc) lattice. At high temperatures, the zincs and coppers freely interchange on the two lattices. The transition temperature is about 733C.

Now, our model just contains atoms on their lattice sites. Surely if one kind of atom is larger than the other, it'll push neighboring atoms off their sites? We simply include these reshufflings into the energies in our Hamiltonian 8.7.

What about the vibrations of the atoms about their equilibrium positions? We can imagine doing a partial trace, as we discussed in section 6.6. Just as in exercise 6.2, one can incorporate the entropy due to the local atomic motions $S\{s_i\}$ about their lattice sites into an effective free energy for each atomic configuration[17]

$$\begin{aligned}\mathcal{F}_{\text{binary}}\{s_i\} &= -k_B T \log\left(\int d\mathbb{P} \int_{r_i \text{ on site } s_i} d\mathbb{Q}\, e^{-\mathcal{H}(\mathbb{P},\mathbb{Q})/k_B T}\right) \\ &= \mathcal{H}_{\text{binary}}\{s_i\} - TS\{s_i\}. \end{aligned} \tag{8.11}$$

Again, as in section 6.3, we're doing a partial trace over states. If we ignore the configurations where the atoms are not near lattice sites, we can recover the total partition function by summing over spin configurations

$$Z = \sum_{\{s_i\}} e^{-\mathcal{F}_{\text{binary}}\{s_i\}/k_B T} \tag{8.12}$$

$$= \sum_{\{s_i\}} \int d\mathbb{P} \int_{r_i \text{ on site } s_i} d\mathbb{Q}\, e^{-\mathcal{H}(\mathbb{P},\mathbb{Q})/k_B T} \approx \int d\mathbb{P} \int d\mathbb{Q}\, e^{-\mathcal{H}(\mathbb{P},\mathbb{Q})/k_B T}. \tag{8.13}$$

Insofar as the entropy in the free energy $\mathcal{F}_{\text{binary}}\{s_i\}$ can be approximated as a sum of pair energies, we again get an Ising model, but now with temperature dependent parameters.

More elaborate Ising models (with three–site and longer–range interactions, for example) are commonly used to compute realistic phase diagrams for alloys (reference [125]). Sometimes, though, the interactions introduced by relaxations and thermal fluctuations off lattice sites have important long–range pieces, which can lead to qualitative changes in the behavior – for example, turning the transition from continuous to abrupt.

[17] This nasty–looking integral over configurations where the atom hasn't shifted too far past its lattice site would normally be approximated by a Gaussian integral over phonon vibrations, similar to that described in exercise 6.2(b), figure 6.10.

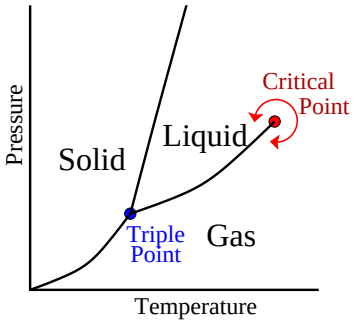



**Fig. 8.7** A schematic phase diagram for a typical material. There is a solid phase at high pressures and low temperatures, a gas phase at low pressures and high temperatures, and a liquid phase in a region in between. The solid-liquid phase boundary corresponds to a change in symmetry, and cannot end. The liquid-gas phase boundary typically does end: one can go continuously from the liquid phase to the gas phase by increasing the pressure above $P_c$, then the temperature above $T_c$, and then lowering the pressure again.

### 8.2.3 Lattice Gas and the Critical Point

The Ising model is also used as a model for the liquid-gas transition. In this *lattice gas* interpretation, up-spins ($s_i = +1$) count as atoms and down-spins count as a site without an atom. The gas is the phase with mostly down spins (negative 'magnetization'), with only a few up-spin atoms in the vapor. The liquid phase is mostly atoms (up-spins), with a few vacancies.

The Ising model description of the gas phase seems fairly realistic. The liquid, however, seems much more like a crystal, with atoms sitting on a regular lattice. Why do we suggest that this model is a good way of studying transitions between the liquid and gas phase?

Unlike the binary alloy problem, the Ising model is not a good way to get quantitative phase diagrams for fluids. What it is good for is to

understand the properties near the *critical point*. As shown in figure 8.7, one can go continuously between the liquid and gas phases: the phase boundary separating them ends at a critical point $T_c, P_c$, above which the two phases blur together seamlessly, with no jump in the density separating them.

The Ising model, interpreted as a lattice gas, also has a line $H = 0$ along which the density (magnetization) jumps, and a temperature $T_c$ above which the properties are smooth as a function of $H$ (the paramagnetic phase). The phase diagram 8.8 looks only topologically like the real liquid-gas coexistence line 8.7, but the behavior near the critical point in the two systems is remarkably similar. Indeed, we will find in chapter 12 that in many ways the behavior at the liquid–gas critical point is described *exactly* by the three-dimensional Ising model.

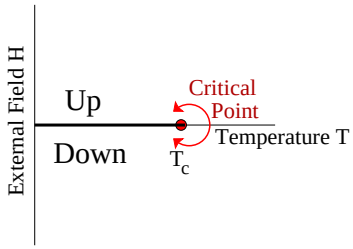



**Fig. 8.8** The phase diagram for the Ising model. Below the critical temperature $T_c$, the $H = 0$ line separates two 'phases', an up-spin and a down-spin phase. Above $T_c$ the behavior is smooth as a function of $H$; below $T_c$ there is a jump in the magnetization as one crosses $H = 0$.

### 8.2.4 How to Solve the Ising Model.

How do we solve for the properties of the Ising model?

(1) Solve the one–dimensional Ising model, as Ising did.[18]
(2) Have an enormous brain. Onsager solved the two–dimensional Ising model in a bewilderingly complicated way. Since Onsager, many great minds have found simpler, elegant solutions, but all would take at least a chapter of rather technical and unilluminating manipulations to duplicate. Nobody has solved the three-dimensional Ising model.
(3) Do Monte Carlo on the computer.[19]

[18] This is a typical homework exercise in a course like ours: with a few hints, you can do it too.

[19] Or, do high temperature expansions, low temperature expansions, transfer matrix methods, exact diagonalization of small systems, $1/N$ expansions in the number of states per site, $4 - \epsilon$ expansions in the dimension of space, ...

The Monte Carlo[20] method involves doing a kind of random walk through the space of lattice configurations. We'll study these methods in great generality in section 8.3. For now, let's just outline the Heat Bath Monte Carlo method.

**Heat Bath Monte Carlo for the Ising Model**

- Pick a site $i = (x, y)$ at random.
- Check how many neighbor spins are pointing up:

$$m_i = \sum_{j:\langle ij \rangle} s_j = \begin{cases} 4 & (4 \text{ neighbors up}) \\ 2 & (3 \text{ neighbors up}) \\ 0 & (2 \text{ neighbors up}) \\ -2 & (1 \text{ neighbor up}) \\ -4 & (0 \text{ neighbors up}) \end{cases} \tag{8.14}$$

- Calculate $E_+ = -Jm_i - H$ and $E_- = +Jm_i + H$, the energy for spin $i$ to be $+1$ or $-1$ given its current environment.
- Set spin $i$ up with probability $e^{-\beta E_+}/(e^{-\beta E_+} + e^{-\beta E_-})$ and down with probability $e^{-\beta E_-}/(e^{-\beta E_+} + e^{-\beta E_-})$.
- Repeat.

[20] Monte Carlo is a gambling center in Monaco. Lots of random numbers are generated there.

The heat-bath algorithm just thermalizes one spin at a time: it sets the spin up or down with probability given by the thermal distribution given that its neighbors are fixed. Using it, and fast modern computers, you can simulate the Ising model fast enough to explore its behavior rather thoroughly, as we will in a variety of exercises.

## 8.3 Markov Chains

The heat-bath Monte-Carlo algorithm is an example of a Markov chain. Markov models are important not only for lattice Monte Carlo, but are generally useful in computational mathematics. In particular, they have become important in bioinformatics and speech recognition, where one attempts to deduce the *hidden Markov model* which describes the patterns and relations in speech or the genome.

A *Markov chain* has a finite set of states $\{\alpha\}$, through which the system evolves in a discrete series of steps $n$.[21] The probabilities of moving to different new states in a Markov chain depends *only* on the current state.[22] That is, the system has no memory of the past evolution.

For example, an $N$-state Ising model has $2^N$ states $\mathbb{S} = \{s_i\}$. A Markov chain for the Ising model has a transition rule, which at each step shifts the current state $\mathbb{S}$ to a state $\mathbb{S}'$ with probability $P_{\mathbb{S}'\Leftarrow\mathbb{S}}$.[23] For the heat–bath algorithm, $P_{\mathbb{S}'\Leftarrow\mathbb{S}}$ is equal to zero unless $\mathbb{S}'$ and $\mathbb{S}$ are the same except for at most one spin flip.

There are many problems outside of mainstream statistical mechanics that can be formulated in this general way. For example, exercise 8.3 discusses a model with 1001 states (different numbers $\alpha$ of red bacteria), and transition rates $P_{\alpha+1\Leftarrow\alpha}$, $P_{\alpha-1\Leftarrow\alpha}$, and $P_{\alpha\Leftarrow\alpha}$. In general, we want to understand the probability of finding different states after long times. Under what circumstances will an algorithm, defined by our matrix $P$, take our system into thermal equilibrium?

Let the probabilities of being in various states $\alpha$ at step $n$ be arranged in a vector $\vec{\rho}_\alpha(n)$. Then the rates[24] $P_{\beta\alpha}$ for moving from $\alpha$ to $\beta$ in a general Markov chain have the following properties:

- **Time evolution:** The probability vector at step $n+1$ is

$$\vec{\rho}_\beta(n+1) = \sum_\alpha P_{\beta\alpha}\vec{\rho}_\alpha(n),$$
$$\vec{\rho}(n+1) = P \cdot \vec{\rho}(n). \tag{8.15}$$

- **Positivity:** The matrix elements are probabilities, so

$$0 \le P_{\beta\alpha} \le 1. \tag{8.16}$$

- **Conservation of Probability:** The state $\alpha$ must go somewhere, so

$$\sum_\beta P_{\beta\alpha} = 1 \tag{8.17}$$

- **Not symmetric!** Typically $P_{\beta\alpha} \neq P_{\alpha\beta}$.

[21] There are analogues of Markov chains that are continuous in time and/or continuous in space.

[22] More generally, systems which lack memory are called Markovian.

[23] We put the subscripts in this order because we will use $P$ as a matrix, which will take a probability vector $P_{\beta\alpha}$ from one step ($\alpha$) to the next ($\beta$), equation 8.15.

[24] We heretofore leave out the left arrow.

This last point isn't a big surprise: it should be much more likely to go from a high–energy state to a low one than from low to high. However, this asymmetry means that much of our mathematical intuition and many of our tools, carefully developed for symmetric and Hermitian matrices, won't apply to our transition matrix $P_{\alpha\beta}$. (In particular, we cannot assume in general that we can diagonalize our matrix.)

What can we know about the Markov chain and its asymmetric matrix $P$? We'll outline the relevant mathematics, proving what is convenient and illuminating and simply asserting other truths.

It is true in great generality that our matrix $P$ will have eigenvalues. Also, it is true that for each distinct eigenvalue there will be at least one right eigenvector[25]

$$P \cdot \vec{\rho}^{\lambda} = \lambda \vec{\rho}^{\lambda} \tag{8.18}$$

and one left eigenvector

$$\vec{\sigma}^{\lambda T} \cdot P = \lambda \vec{\sigma}^{\lambda T}. \tag{8.19}$$

However, for degenerate eigenvalues there may not be multiple eigenvectors, and the left and right eigenvectors usually will not be equal to one another.[26]

For the particular case of our transition matrix $P$, we can go further. If our Markov chain reaches an equilibrium state at long times, that state must be unchanged under the time evolution $P$. That is, $P \cdot \vec{\rho} = \vec{\rho}$, and thus the equilibrium probability density is a right eigenvector with eigenvalue one. We can show that our Markov chain transition matrix $P$ has such a right eigenvector.

**Theorem 8.1.** *$P$ has at least one right eigenvector $\vec{\rho}^*$ with eigenvalue one.*

Sneaky Proof: $P$ has a left eigenvector $\vec{\sigma}^*$ with eigenvalue one: the vector all of whose components are one, $\vec{\sigma}^{*T} = (1, 1, 1, \ldots, 1)$:

$$(\vec{\sigma}^{*T} \cdot P)_\alpha = \sum_\beta \sigma^*_\beta P_{\beta\alpha} = \sum_\beta P_{\beta\alpha} = 1 = \sigma^*_\beta. \tag{8.20}$$

Hence $P$ must have an eigenvalue equal to one, and hence it must also have a right eigenvector with eigenvalue one.

We can also show that all the other eigenvalues have right eigenvectors that sum to zero, since $P$ conserves probability: [27]

**Theorem 8.2.** *Any right eigenvector $\vec{\rho}^{\lambda}$ with eigenvalue $\lambda$ different from one must have components that sum to zero.*

[25] For example, the matrix $\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ has a double eigenvalue of zero, but only one left and right eigenvector with eigenvalue zero.

[26] We won't prove these truths. They perhaps may be motivated by stating another truth we won't demonstrate: a general matrix $M$ can be put into *Jordan canonical form* by a suitable change of basis $S$: $M = SJS^{-1}$. The matrix $J$ is block diagonal, with one eigenvalue $\lambda$ associated with each block (but perhaps multiple blocks per $\lambda$). A given (say, $3 \times 3$) block will be of the form $\begin{pmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{pmatrix}$ with $\lambda$ along the diagonal and 1 in the elements immediately above the diagonal. The first column of the block is associated with the right eigenvector for $\lambda$; the last row is associated with the left eigenvector. (The word canonical here means "simplest form", and doesn't indicate a connection with the canonical ensemble.)

[27] One can also view this theorem as saying that all the right eigenvectors except $\rho^*$ are orthogonal to the left eigenvector $\sigma^*$.

Proof: $\vec{\rho}^{\lambda}$ is a right eigenvector, $P \cdot \vec{\rho}^{\lambda} = \lambda \vec{\rho}^{\lambda}$. Hence

$$\lambda \sum_{\beta} \rho_{\beta}^{\lambda} = \sum_{\beta} \left(\lambda \rho_{\beta}^{\lambda}\right) = \sum_{\beta} \left( \sum_{\alpha} P_{\beta\alpha} \rho_{\alpha}^{\lambda} \right) = \sum_{\alpha} \left( \sum_{\beta} P_{\beta\alpha} \right) \rho_{\alpha}^{\lambda}$$
$$= \sum_{\alpha} \rho_{\alpha}^{\lambda}. \tag{8.21}$$

This implies that either $\lambda = 1$ or $\sum_{\alpha} \rho_{\alpha}^{\lambda} = 0$.

Markov chains can have more than one stationary probability distribution.[28] They can have *transient* states, which the system eventually leaves never to return.[29] They can also have *cycles*, which are probability distributions which like a clock $1 \to 2 \to 3 \to \cdots \to 12 \to 1$ shift through a finite number of distinct classes of states before returning to the original one. All of these are obstacles in our quest for finding the equilibrium states in statistical mechanics. We can bypass all of them by studying *ergodic* Markov chains.[30] A finite–state Markov chain is ergodic if its transition matrix to some power $n$ has all positive (non-zero) matrix elements: ${P^n}_{\beta\alpha} > 0$ for all states $\alpha$ and $\beta$.[31]

We use a famous theorem, without proving it here:

**Theorem 8.3 (Perron–Frobenius Theorem).** *Let $A$ be a matrix with all non–negative matrix elements such that $A^n$ has all positive elements. Then $A$ has a positive eigenvalue $\lambda_0$, of multiplicity one, whose corresponding right and left eigenvectors have all positive components. Furthermore any other eigenvalue $\lambda$ of $A$ must be smaller, $|\lambda| < \lambda_0$.*

For an ergodic Markov chain, we can use theorem 8.2, to see that the Perron–Frobenius eigenvector with all positive components must have eigenvalue $\lambda_0 = 1$. We can rescale this eigenvector to sum to one, proving that an ergodic Markov chain has a unique time–independent probability distribution $\vec{\rho}^{*}$.

What's the difference between our definition of ergodic Markov chains, and the definition of ergodic we used in section 4.2 in reference to trajectories in phase space? Clearly the two concepts are related: ergodic in phase space meant that we eventually come close to all states on the energy surface, and for finite Markov chains it is the stronger condition that we have non-zero probability of getting between all states in the chain after precisely $n$ steps. Indeed, one can show for finite state Markov chains that if one can get from every state $\alpha$ to every other state $\beta$ by a sequence of moves (that is, the chain is *irreducible*), and if the chain is not cyclic, then it is ergodic (proof not given here). Any algorithm that has a finite probability for each state to remain unchanged ($P_{\alpha\alpha} > 0$ for all states) is automatically free of cycles (clocks which lose time will get out of synchrony).

It is possible to show that an ergodic Markov chain will take any initial probability distribution $\vec{\rho}(0)$ and converge to equilibrium, but the proof in general is rather involved. We can simplify it by specializing one more time, to Markov chains that satisfy *detailed balance*.

[28] A continuum example of this is given by the KAM theorem of exercise 4.2. There is a probability density smeared over each KAM torus which is time-independent.

[29] Transient states are important in dissipative dynamical systems, where they are all the states not on the attractors.

[30] We're compromising here between the standard Markov-chain usage in physics and in mathematics. Physicists usually ignore cycles, and call algorithms which can reach every state ergodic (what mathematicians call irreducible). Mathematicians use the term ergodic to exclude cycles and exclude probability running to infinity (not important here, where we have a finite number of states). They also allow ergodic chains to have transient states: only the "attractor" need be connected. Chains with $P^n$ everywhere positive, that we're calling ergodic, are called by the mathematicians *regular* Markov chains. In the exercises, to prove a system is ergodic just show that it can reach everywhere (irreducible) and doesn't have cycles.

[31] That is, after $n$ steps every state has non–zero probability to reach every other state.

A Markov chain satisfies *detailed balance* if there is some probability distribution $\vec{\rho}^*$ such that[32]

$$P_{\alpha\beta}\rho^*_\beta = P_{\beta\alpha}\rho^*_\alpha \tag{8.22}$$

for each state $\alpha$ and $\beta$. In words, the probability flux from state $\alpha$ to $\beta$ (the rate times the probability of being in $\alpha$) balances the probability flux back, in detail (*i.e.*, for every pair of states).

[32] There is an elegant equivalent definition of detailed balance directly in terms of $P$ and not involving the equilibrium probability distribution $\vec{\rho}^*$: see exercise 8.4.

If a physical system is time reversal invariant (no dissipation, no magnetic fields), and its states are also invariant under time reversal (no states with specified velocities or momenta) then its dynamics automatically satisfy detailed balance. This is true because the equilibrium state is also the equilibrium state under time reversal, so the probability flow from $\beta \to \alpha$ must equal the time–reversed flow from $\alpha \to \beta$. Quantum systems undergoing transitions between energy eigenstates in perturbation theory usually satisfy detailed balance, since the eigenstates are time-reversal invariant. Most classical models (like the binary alloy in 8.2.2) have states involving only configurational degrees of freedom, which again satisfy detailed balance.

Detailed balance allows us to find a complete set of eigenvectors and right eigenvalues for our transition matrix $P$. One can see this with a simple transformation. If we divide both sides of equation 8.22 by $\sqrt{\rho^*_\beta \rho^*_\alpha}$, we create a symmetric matrix $Q_{\alpha\beta}$

$$\begin{aligned} Q_{\alpha\beta} &= P_{\alpha\beta}\sqrt{\frac{\rho^*_\beta}{\rho^*_\alpha}} = P_{\alpha\beta}\rho^*_\beta / \sqrt{\rho^*_\alpha \rho^*_\beta} \\ &= P_{\beta\alpha}\rho^*_\alpha / \sqrt{\rho^*_\alpha \rho^*_\beta} = P_{\beta\alpha}\sqrt{\frac{\rho^*_\alpha}{\rho^*_\beta}} = Q_{\beta\alpha}. \end{aligned} \tag{8.23}$$

This particular symmetric matrix has eigenvectors $Q \cdot \tau^\lambda = \lambda \tau^\lambda$ which can be turned into right eigenvectors of $P$ when rescaled[33] by $\sqrt{\rho^*}$:

$$\rho^\lambda_\alpha = \tau^\lambda_\alpha \sqrt{\rho^*_\alpha}\,: \tag{8.24}$$

$$\begin{aligned} \sum_\alpha P_{\beta\alpha}\rho^\lambda_\alpha &= \sum_\alpha P_{\beta\alpha}(\tau^\lambda_\alpha \sqrt{\rho^*_\alpha}) = \sum_\alpha \left( Q_{\beta\alpha}\sqrt{\frac{\rho^*_\beta}{\rho^*_\alpha}} \right)(\tau^\lambda_\alpha \sqrt{\rho^*_\alpha}) \\ &= \sum_\alpha \left( Q_{\beta\alpha}\tau^\lambda_\alpha \right)\sqrt{\rho^*_\beta} = \lambda(\tau^\lambda_\beta \sqrt{\rho^*_\beta}) = \lambda \rho^\lambda_\beta. \end{aligned} \tag{8.25}$$

[33] This works in reverse to get the right eigenvectors of $P$ from $Q$. One multiplies $\tau^\lambda_\alpha$ by $\sqrt{\rho^*_\alpha}$ to get $\rho^\lambda_\alpha$, and divides to get $\sigma^\lambda_\alpha$, so if detailed balance holds, $\sigma^\lambda_\alpha = \rho^\lambda_\alpha / \rho^*_\alpha$. In particular, $\sigma^1 = \sigma^* = (1, 1, 1, \dots)^T$, as we saw in theorem 8.1.

Now for the main theorem underlying the algorithms for equilibrating lattice models in statistical mechanics.

**Theorem 8.4 (Main Theorem).** *A system with a finite number of states can be guaranteed to converge to an equilibrium distribution $\vec{\rho}^*$ if the computer algorithm*

- *is Markovian (has no memory),*

- *is ergodic (can reach everywhere and is acyclic) and*
- *satisfies detailed balance.*

Proof: Let $P$ be the transition matrix for our algorithm. Since the algorithm satisfies detailed balance, $P$ has a complete set of eigenvectors $\vec{\rho}^{\lambda}$. Since our algorithm is ergodic there is only one right eigenvector $\vec{\rho}^{1}$ with eigenvalue one, which we can choose to be the stationary distribution $\vec{\rho}^{*}$; all the other eigenvalues $\lambda$ have $|\lambda| < 1$. Decompose the initial condition $\vec{\rho}(0) = a_1\vec{\rho}^{*} + \sum_{|\lambda|<1} a_\lambda \vec{\rho}^{\lambda}$. Then[34]

$$\vec{\rho}(n) = P \cdot \vec{\rho}(n-1) = P^n \cdot \vec{\rho}(0) = a_1\vec{\rho}^{*} + \sum_{|\lambda|<1} a_\lambda \lambda^n \vec{\rho}^{\lambda}. \tag{8.26}$$

Since the (finite) sum in this equation decays to zero, the density converges to $a_1\vec{\rho}^{*}$. This implies both that $a_1 = 1$ and that our system converges to $\vec{\rho}^{*}$ as $n \to \infty$.

Thus, to develop a new equilibration algorithm, one need only ensure that it is Markov, ergodic, and satisfies detailed balance (exercise 8.5).

[34] The eigenvectors closest to one will be the slowest to decay. You can get the slowest characteristic time $\tau$ for a Markov chain by finding the largest $|\lambda_{\max}| < 1$ and setting $\lambda^n = e^{-n\tau}$.

# Exercises

Exercise 8.1 introduces the Ising model, its phase diagram, magnetization and magnetization fluctuations, and low and high temperature behavior. Exercises 8.2 and 8.3 introduce simple examples of Markov chains. Exercise 8.5 explores the heat bath, Metropolis, and Wolff cluster-flip algorithm for equilibrating the Ising model. Exercises 8.6 and 8.7 explore the Gillespie algorithm for stochastic simulations of chemical reactions in small systems like biological cells where number fluctuations can be important. In exercise 8.8 you will show that entropy increases for Markov chains. Exercise 8.9 introduces methods for solving ordinary differential equations that are used for molecular dynamics simulations, emphasizing the three themes of accuracy, stability, and fidelity. Finally, exercises 8.10 and 8.11 introduce the simulation of *networks* and their connectivity, using the classic examples of small-world and percolation networks.

(8.1) **The Ising Model.** (Computational)
You'll need the program **ising**, available on the Web [105]. The Ising Hamiltonian is

$$\mathcal{H} = -J\sum_{\langle ij\rangle} S_i S_j - H\sum_i S_i, \tag{3.7.1}$$

where $S_i = \pm 1$ are "spins" on a square lattice, and the sum $\sum_{\langle ij\rangle}$ is over the four nearest-neighbor bonds (each pair summed once). It's conventional to set the coupling strength $J = 1$ and Boltzmann's constant $k_B = 1$, which amounts to measuring energies and temperatures in units of $J$. The constant $H$ is called the external field, and $\mathbf{M} = \sum_i S_i$ is called the magnetization.

As noted in class, the Ising model can be viewed as an anisotropic magnet with $S_i$ being $2\sigma_z$ for the spin at site $i$, or it can represent the occupancy of a lattice site (atom or no atom for a lattice gas simulation, copper or gold for a binary alloy, ...). As a lattice gas, $\mathbf{M}$ gives the net concentration, and $H$ corresponds to a chemical potential. Our simulation doesn't conserve the number of spins up, so it's not a natural simulation for a bulk lattice gas. You can think of it as a grand canonical ensemble, or as a model for a lattice gas on a surface exchanging atoms with the vapor above.

Play with it. At high temperatures, the spins should not be strongly correlated. At low temperatures the spins should align all parallel, giving a large magnetization. Can you roughly locate the phase transition? Can you see growing clumps of aligned spins as $T \to T_c+$ (*i.e.*, $T$ approaching $T_c$ from above)?

*(a)* **Phase diagram.** *Draw a rough phase diagram in the $(H, T)$ plane, showing* (i) *the "spin up" phase where*

$\langle \mathbf{M} \rangle > 0$, (ii) *the "spin down" phase with* $\langle \mathbf{M} \rangle < 0$, (iii) *the paramagnetic phase line where* $\mathbf{M} = 0$, (iv) *the ferromagnetic phase line where* $|\langle \mathbf{M} \rangle| > 0$ *for large systems even though* $H = 0$, *and* (v) *the critical point, where at* $H = 0$ *the system develops a non-zero magnetization.*

**Correlations and Susceptibilities: Analytical.** The partition function for the Ising model is $Z = \sum_n \exp(-\beta E_n)$, where the states $n$ run over all $2^N$ possible configurations of the spins, and the free energy $\mathcal{F} = -kT \log Z$.

*(b) Show that the average of the magnetization* $\mathbf{M}$ *equals* $-(\partial \mathcal{F}/\partial H)|_T$. *Derive the formula writing the susceptibility* $\chi_0 = (\partial \mathbf{M}/\partial H)|_T$ *in terms of* $\langle (\mathbf{M} - \langle \mathbf{M} \rangle)^2 \rangle = \langle \mathbf{M}^2 \rangle - \langle \mathbf{M} \rangle^2$. *(Hint: remember our derivation of formula 6.13* $\langle (\mathbf{E} - \langle \mathbf{E} \rangle)^2 \rangle = k_B T^2 C$*?)*

Notice that the program outputs, at each temperature and field, averages of several quantities: $\langle |M| \rangle$, $\langle (M - \langle M \rangle)^2 \rangle$, $\langle E \rangle$, $\langle (E - \langle E \rangle)^2 \rangle$. Unfortunately, $E$ and $M$ in these formulas are *measured per spin*, while the formulas in the class and the exercise are measured for the *system as a whole*. You'll need to multiply the squared quantities by the number of spins to make a comparison. To make that easier, change the system size to 100×100, using *configure*. While you're doing that, increase *speed* to ten or twenty to draw the spin configuration fewer times. To get good values for these averages, equilibrate for a given field and temperature, "reset", and then start averaging.

*(c)* **Correlations and Susceptibilities: Numerical.** *Check the formulas for* $C$ *and* $\chi$ *from part (b) at* $H = 0$ *and* $T = 3$, *by measuring the fluctuations and the averages, and then changing by* $\Delta H = 0.02$ *or* $\Delta T = 0.1$ *and measuring the averages again. Check them also for* $T = 2$, *where* $\langle \mathbf{M} \rangle \neq 0$.[35]

There are systematic series expansion for the Ising model at high and low temperatures, using Feynman diagrams (see section 8.1). The first terms of these expansions are both famous and illuminating.

**Low Temperature Expansion for the Magnetization.** At low temperatures we can assume all spins flip alone, ignoring clusters.

*(d) What is the energy for flipping a spin antiparallel to its neighbors? Equilibrate at low temperature* $T = 1.0$, *and measure the magnetization. Notice that the primary excitations are single spin flips. In the low temperature approximation that the flipped spins are dilute (so we may ignore the possibility that two flipped spins touch or overlap), write a formula for the magnetization. (Remember, each flipped spin changes the magnetization by 2.) Check your prediction against the simulation.* (Hint: see equation 8.1.)

The magnetization (and the specific heat) are exponentially small at low temperatures because there is an *energy gap* to spin excitations in the Ising model,[36] just as there is a gap to charge excitations in a semiconductor or an insulator.

**High Temperature Expansion for the Susceptibility.** At high temperatures, we can ignore the coupling to the neighboring spins.

*(e) Calculate a formula for the susceptibility of a free spin coupled to an external field. Compare it to the susceptibility you measure at high temperature* $T = 100$ *for the Ising model (say,* $\Delta M/\Delta H$ *with* $\Delta H = 1$. *Why is* $H = 1$ *a small field in this case?)*

Your formula for the high-temperature susceptibility is known more generally as Curie's law.

(8.2) **Coin Flips and Markov Chains.** (Mathematics, Basic)

A physicist, testing the laws of chance, flips a coin repeatedly until it lands tails.

*(a) Treating the two states of the physicist ("still flipping" and "done") as states in a Markov process. The current probability vector then is* $\vec{\rho} = \begin{pmatrix} \rho_{flipping} \\ \rho_{done} \end{pmatrix}$. *Write the transition matrix* $\mathcal{P}$, *giving the time evolution* $\mathcal{P} \cdot \vec{\rho}_n = \vec{\rho}_{n+1}$, *assuming that the coin is fair.*

*(b) Find the eigenvalues and right eigenvectors of* $\mathcal{P}$. *Which eigenvector is the steady state* $\rho^*$*? Call the other eigenvector* $\tilde{\rho}$. *For convenience, normalize* $\tilde{\rho}$ *so that its first component equals one.*

*(c) Assume an arbitrary initial state is written* $\rho_0 = A\rho^* + B\tilde{\rho}$. *What are the conditions on* $A$ *and* $B$ *needed to make* $\rho_0$ *a valid probability distribution? Write* $\rho_n$ *as a function of* $A$ *and* $B$, $\rho^*$ *and* $\tilde{\rho}$.

(8.3) **Red and Green Bacteria** (Mathematics) (From Princeton. [119])

A growth medium at time $t = 0$ has 500 red bacteria and 500 green bacteria. Each hour, each bacterium divides

[35] Be sure to wait until the state is equilibrated before you start! Below $T_c$ this means the state should not have red and black 'domains', but be all in one ground state. You may need to apply a weak external field for a while to remove stripes at low temperatures.

[36] Not all real magnets have a gap: if there is a spin rotation symmetry, one can have gapless *spin waves*, like phonons for spins.

in two. A color-blind predator eats exactly 1000 bacteria per hour.[37]

*(a) After a very long time, what is the probability distribution for the number $\alpha$ of red bacteria in the growth medium?*

*(b) Roughly how long will it take to reach this final state?*[38]

*(c) Assume that the predator has a 1% preference for green bacteria (implemented as you choose). Roughly how much will this change the final distribution?*

(8.4) **Detailed Balance.** (Basic)

In an equilibrium system, for any two states $\alpha$ and $\beta$ with equilibrium probabilities $\rho^*_\alpha$ and $\rho^*_\beta$, detailed balance states (equation 8.22) that

$$P_{\beta\Leftarrow\alpha}\rho^*_\alpha = P_{\alpha\Leftarrow\beta}\rho^*_\beta, \tag{8.27}$$

that is, the equilibrium flux of probability from $\alpha$ to $\beta$ is the same as the flux backward from $\beta$ to $\alpha$. It's both possible and elegant to reformulate the condition for detailed balance so that it doesn't involve the equilibrium probabilities. Consider three states of the system, $\alpha$, $\beta$, and $\gamma$.

*(a) Assume that each of the three types of transitions among the three states satisfies detailed balance. Eliminate the equilibrium probability densities to write the unknown rate $P_{\alpha\Leftarrow\beta}$ in terms of the five other rates.* (Hint: see equation below for answer.)

If we view the three states $\alpha$, $\beta$, and $\gamma$ to be around a circle, you've derived a relationship between the rates going clockwise and the rates going counter-clockwise around the circle,

$$P_{\alpha\Leftarrow\beta}P_{\beta\Leftarrow\gamma}P_{\gamma\Leftarrow\alpha} = P_{\alpha\Leftarrow\gamma}P_{\gamma\Leftarrow\beta}P_{\beta\Leftarrow\alpha}. \tag{8.28}$$

It is possible to show conversely that if every triple of states in a Markov chain satisfies the condition you derived then it satisfies detailed balance (*i.e.*, that there is at least one probability density $\rho^*$ which makes the probability fluxes between all pairs of states equal). The only complication arises because some of the rates can be zero.

*(b) Suppose $P$ is the transition matrix for some Markov process satisfying the condition 8.28 for every triple of states $\alpha$, $\beta$, and $\gamma$. Assume for simplicity that there is a state $\alpha$ with non-zero transition rates from all other states $\delta$. Construct a probability density $\rho^*_\delta$ that demonstrates that $P$ satisfies detailed balance (equation 8.27).* (Hint: If you assume a value for $\rho^*_\alpha$, what must $\rho^*_\delta$ be to ensure detailed balance for the pair? Show that this candidate distribution satisfies detailed balance for any two states.)

(8.5) **Heat Bath, Metropolis, and Wolff.** (Mathematics, Computation)

There are a number of different methods for equilibrating lattice simulations like the Ising model. They give the model different dynamics, but keep the equilibrium properties unchanged. This is guaranteed by the theorem we asserted in class on Markov processes: if they are ergodic and obey detailed balance, they converge to the equilibrium distribution. We'll first look at the two most common algorithms. We'll then consider the most sophisticated, sneaky use of the theorem I know of.

The simulation **ising** in exercise 8.1 uses the *heat-bath* algorithm, which thermalizes one spin at a time:

Heat Bath

(a) Pick a spin at random,
(b) Calculate the energies $\mathbf{E}_\uparrow$ and $\mathbf{E}_\downarrow$ for the spin being up or down given its current environment.
(c) Thermalize it: place it up with probability $e^{-\beta\mathbf{E}_\uparrow}/(e^{-\beta\mathbf{E}_\uparrow} + e^{-\beta\mathbf{E}_\downarrow})$, down with probability $e^{-\beta\mathbf{E}_\downarrow}/(e^{-\beta\mathbf{E}_\uparrow} + e^{-\beta\mathbf{E}_\downarrow})$.

Another popular choice is the Metropolis algorithm, which also flips a single spin at a time:

Metropolis

(a) Pick a spin at random,
(b) Calculate the energy $\Delta\mathbf{E}$ for flipping the spin.
(c) If $\Delta\mathbf{E} < 0$ flip it; if $\Delta\mathbf{E} > 0$, flip it with probability $e^{-\beta\,\Delta\mathbf{E}}$.

*(a) Show that Heat Bath and Metropolis satisfy detailed balance. Note that they are ergodic and Markovian (no memory), and hence argue that they will lead to thermal equilibrium. Is Metropolis more efficient (fewer random numbers needed to get to equilibrium)? Why?*

[37] This question is purposely open–ended, and rough answers to parts (b) and (c) within a factor of two are perfectly acceptable. Numerical and analytical methods are both feasible.

[38] Within the accuracy of this question, you may assume either that one bacterium reproduces and then one is eaten 1000 times per hour, or that at the end of each hour all the bacteria reproduce and then 1000 are consumed. The former method is more convenient for analytical work finding eigenvectors; the latter can be used to motivate approaches using the diffusion of probability with an $\alpha$–dependent diffusion constant.

Near the critical point $T_c$ where the system develops a magnetization, any single-spin-flip dynamics becomes very slow. Wolff (*Phys. Rev. Lett.* **62**, 361 (1989)), improving on ideas of Swendsen and Wang (*Phys. Rev. Lett.* **58**, 86 (1987)), came up with a clever method to flip whole clusters of spins.

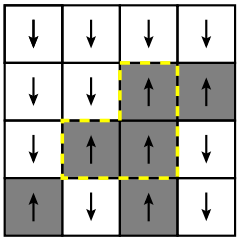


**Fig. 8.9 Cluster Flip.** The region inside the dotted line is flipped in one Wolff move. Let this configuration be A.

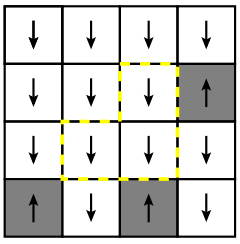


**Fig. 8.10 Cluster Flip.** Let this configuration be B. Let the cluster flipped be C. Notice that the boundary of C has $n\uparrow = 2$, $n\downarrow = 6$.

Wolff Cluster Flips

(a) Pick a spin at random, remember its direction $D = \pm 1$, and flip it.

(b) For each of the four neighboring spins, *if* it is in the direction $D$, flip it with probability $p$.

(c) For each of the new flipped spins, recursively flip their neighbors as in (2).

Because with finite probability you can flip any spin, the Wolff algorithm is ergodic. It's obviously Markovian when viewed as a move which flips a cluster. Let's see that it satisfies detailed balance, when we pick the right value of $p$ for the given temperature.

*(b) Show for the two configurations shown above that $E_B - E_A = 2(n_\uparrow - n_\downarrow)J$. Argue that this will be true for flipping any cluster of up spins to down.*

The cluster flip can start at any site $\alpha$ in the cluster C. The ratio of rates $\Gamma_{A\to B}/\Gamma_{B\to A}$ depends upon the number of times the cluster chose *not* to grow on the boundary. Let $P_\alpha^C$ be the probability that the cluster grows internally from site $\alpha$ to the cluster $C$ (ignoring the moves which try to grow outside the boundary). Then

$$\Gamma_{A\to B} = \sum_\alpha P_\alpha^C (1-p)^{n\uparrow}, \tag{8.29}$$

$$\Gamma_{B\to A} = \sum_\alpha P_\alpha^C (1-p)^{n\downarrow}, \tag{8.30}$$

since the cluster must refuse to grow $n_\uparrow$ times when starting from the up-state A, and $n_\downarrow$ times when starting from B.

*(c) What value of p lets the Wolff algorithm satisfy detailed balance at temperature T?*

Find a Windows machine. Download the Wolff simulation [106]. Using the parameter reset (top left) reset the temperature to 2.3, the algorithm to Heat Bath, and the height and width to 512. Watch the slow growth of the characteristic cluster sizes. Now change to Wolff, and see how much faster the code is. Also notice that each sweep almost completely rearranges the pattern: the correlation time is much smaller for the Wolff algorithm. (See [78, secs. 4.2 and 4.3] for more details on the Wolff algorithm.)

(8.6) **Stochastic Cells.** (Biology, Computation) (With Myers. [75])

Living cells are amazingly complex mixtures of a variety of complex molecules (RNA, DNA, proteins, lipids ...) that are constantly undergoing reactions with one another. This complex of reactions has been compared to computation: the cell gets input from external and internal sensors, and through an intricate series of reactions produces an appropriate response. Thus, for example, receptor cells in the retina 'listen' for light and respond by triggering a nerve impulse.

The kinetics of chemical reactions are usually described using differential equations for the concentrations of the various chemicals, and rarely are statistical fluctuations considered important. In a cell, the numbers of molecules of a given type can be rather small: indeed, there is (often) only one copy of the relevant part of DNA for a given reaction. It's an important question whether and when we may describe the dynamics inside the cell using continuous concentration variables, even though the actual numbers of molecules are always integers.

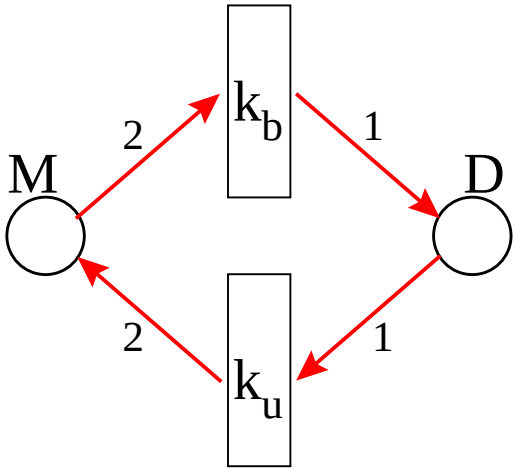



**Fig. 8.11 Dimerization reaction.** A Petri net diagram for a dimerization reaction, with dimerization rate $k_b$ and dimer dissociation rate $k_u$.

Consider a dimerization reaction: a molecule $M$ (called the "monomer") joins up with another monomer and becomes a dimer $D$: $2M \longleftrightarrow D$. Proteins in cells often form dimers: sometimes (as here) both proteins are the same (homodimers) and sometimes they are different proteins (heterodimers). Suppose the forward reaction rate is $k_u$ and the backward reaction rate is $k_d$. Figure 8.11 shows this as a Petri net [39] with each reaction shown as a box with incoming arrows showing species that are consumed by the reaction, and outgoing arrows showing species that are produced by the reaction: the number consumed or produced (the *stoichiometry*) is given by a label on each arrow.[39] There are thus two reactions: the backward unbinding reaction rate per unit volume is $k_u[D]$ (each dimer disassociates with rate $k_u$), and the forward binding reaction rate per unit volume is $k_b[M]^2$ (since each monomer must wait for a collision with another monomer before binding, the rate is proportional to the monomer concentration squared).[40]

The brackets [] denote concentrations. We assume, as does reference [29], that the volume per cell is such that one molecule per cell is 1nM ($10^{-9}$ moles per liter). For convenience, we shall pick nanomoles as our unit of concentration, so $[M]$ is also the number of monomers in the cell. Assume $k_b = 1\,nM^{-1}s^{-1}$ and $k_u = 2\,s^{-1}$, and that at $t = 0$ all $N$ monomers are unbound.

*(a)* Continuum dimerization. *Write the differential equation for $dM/dt$ treating $M$ and $D$ as continuous variables.* (Hint: remember that two $M$ molecules are consumed in each reaction.) *What are the equilibrium concentrations for $[M]$ and $[D]$ for $N = 2$ molecules in the cell, assuming these continuous equations and the values above for $k_b$ and $k_u$? For $N = 90$ and $N = 10100$ molecules? Numerically solve your differential equation for $M(t)$ for $N = 2$ and $N = 90$, and verify that your solution settles down to the equilbrium values you found.*

For large numbers of molecules in the cell, we expect that the continuum equations may work well, but for just a few molecules there surely will be relatively large fluctuations. These fluctuations are called *shot noise*, named in early studies of electrical noise at low currents due to individual electrons in a resistor. We can implement a Monte–Carlo algorithm to simulate this shot noise.[41] Suppose the reactions have rates $\Gamma_i$, with total rate $\Gamma_{\rm tot} = \sum_i \Gamma_i$. The idea is that the expected time to the next reaction is $1/\Gamma_{\rm tot}$, and the probability that the next reaction will be $j$ is $\Gamma_j/\Gamma_{\rm tot}$. To simulate until a final time $t_f$, the algorithm runs as follows:

(a) Calculate a list of the rates of all reactions in the system.

(b) Find the total rate $\Gamma_{\rm tot}$.

(c) Pick a random time $t_{\rm wait}$ with probability distribution $\rho(t) = \Gamma_{\rm tot} \exp(-\Gamma_{\rm tot}\, t)$.

(d) If the current time $t$ plus $t_{\rm wait}$ is bigger than $t_f$, no further reactions will take place: return.

(e) Otherwise,

- Increment $t$ by $t_{\rm wait}$,
- Pick a random number $r$ uniformly distributed in the range $[0, \Gamma_{\rm tot})$,
- Pick the reaction $j$ for which $\sum_{i<j} \Gamma_i \le r < \sum_{i<j+1} \Gamma_i$ (that is, $r$ lands in the $j^{th}$ interval of the sum forming $\Gamma_{\rm tot}$).
- Execute that reaction, by incrementing each chemical involved by its stoichiometry.

(f) Repeat.

[39] An enzyme that is necessary but not consumed is shown with an incoming and outgoing arrow.

[40] In the discrete case, the rate will be proportional to $M(M-1)$, since a monomer cannot collide with itself.

[41] In the context of chemical simulations, this algorithm is named after Gillespie [35]; the same basic approach was used just a bit earlier in the Ising model by Bortz, Kalos and Lebowitz [13], and is called *continuous–time Monte Carlo* in that context.

As mentioned earlier, the binding reaction rate for $M$ total monomers binding is no longer $k_b M^2$ for discrete molecules: it's $k_b M(M-1)$ (where again $[M] \approx M$ for a one nanoliter cell, when using concentrations in nanomolar).[42]

*(b)* Stochastic dimerization. *Implement this algorithm for the dimerization reaction of part (a). Simulate for $N = 2$, $N = 90$, and $N = 10100$ and compare a few stochastic realizations with the continuum solution. How large a value of $N$ do you need for the individual reactions to be well described by the continuum equations (say, fluctuations less than $\pm 20\%$ at late times)?*

Measuring the concentrations in a single cell is often a challenge. Experiments often average over many cells. Such experiments will measure a smooth time evolution even though the individual cells are noisy. Let's investigate whether this ensemble average is well described by the continuum equations.

*(c)* Average Stochastic dimerization. *Find the average of many realizations of your stochastic dimerization in part (b), for $N = 2$ and $N = 90$, and compare with your deterministic solution. How much is the long–term average shifted by the stochastic noise? How large a value of $N$ do you need for the ensemble average of $M(t)$ to be well described by the continuum equations (say, shifted by less than 5% at late times)?*

(8.7) **The Repressilator.** (Biology, Computation) (With Myers. [75])

**Reading:** Reference [29], Michael B. Elowitz and Stanislaw Leibler, "A synthetic oscillator network of transcriptional regulators" *Nature* **403**, 335-338 (2000).

The 'central dogma' of molecular biology is that the flow of information is from DNA to RNA to proteins: DNA is *transcribed* into RNA, which then is *translated* into protein.

Now that the genome is sequenced, it is thought that we have the parts list for the cell. 'All' that remains is to figure out how they work together. The proteins, RNA, and DNA form a complex network of interacting chemical reactions, which governs metabolism, responses to external stimuli, reproduction (*proliferation*), *differentiation* into different cell types, and (when the system perceives itself to be breaking down in dangerous ways) programmed cell death, or *apoptosis*.

Our understanding of the structure of these interacting networks is growing rapidly, but our understanding of the dynamics is still rather primitive. Part of the difficulty is that the cellular networks are not neatly separated into different modules: a given protein may participate in what would seem to be several separate regulatory pathways. In this exercise, we will study a model gene regulatory network, the *Repressilator*. This experimental system involves three proteins each of which inhibits the formation of the next. They were added to the bacterium *E. coli*, with hopefully minimal interactions with the rest of the biological machinery of the cell. We will implement the stochastic model that the authors used to describe their experimental system [29], in order to

- Implement in a tangible system an example both of the central dogma and of *transcriptional regulation*: the control by proteins of DNA expression into RNA,
- Introduce sophisticated Monte–Carlo techniques for simulations of stochastic reactions,
- Introduce methods for automatically generating continuum descriptions from reaction rates, and
- Illustrate the *shot noise* fluctuations due to small numbers of molecules and the *telegraph noise* fluctuations due to finite rates of binding and unbinding of the regulating proteins.

Figure 8.12 shows the biologist's view of the repressilator network. Three proteins (TetR, $\lambda$CI, and LacI) each repress the formation of the next. We shall see that, under appropriate circumstances, this can lead to spontaneous oscillations: each protein peaks in turn, suppressing the suppressor of its suppressor, leading to its own later decrease.

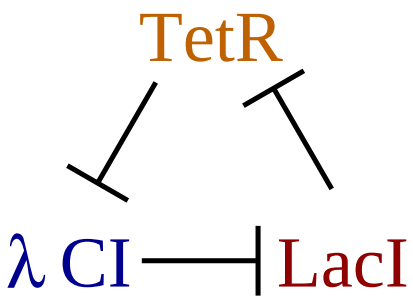



**Fig. 8.12** The biologist's view of the Repressilator network. The T-shapes are blunt arrows, signifying that the protein at the tail (bottom of the T) suppresses the production of the protein at the head. Thus LacI (pronounced lack-eye) suppresses TetR (tet are), which suppresses $\lambda$ CI (lambda-see-one). This simple description summarizes a complex series of interactions (see figure 8.13).

[42] Without this change, if you start with an odd number of cells your concentrations can go negative!

The biologists' notation summarizes a much more complex picture. The LacI protein, for example, can bind to one or both of the *transcriptional regulation* or *operator* sites ahead of the gene that codes for the tetR mRNA.[43] When bound, it largely blocks the translation of DNA into tetR.[44] The level of tetR will gradually decrease as it degrades; hence less TetR protein will be translated from the tetR mRNA. The resulting network of ten reactions is depicted in figure 8.13, showing one third of the total repressilator network. The biologist's shorthand (figure 8.12 does not specify the details of how one protein represses the production of the next. The larger diagram, for example, includes two operator sites for the repressor molecule to bind to, leading to three states ($P_0$, $P_1$, and $P_2$) of the promotor region depending upon how many LacI proteins are bound.

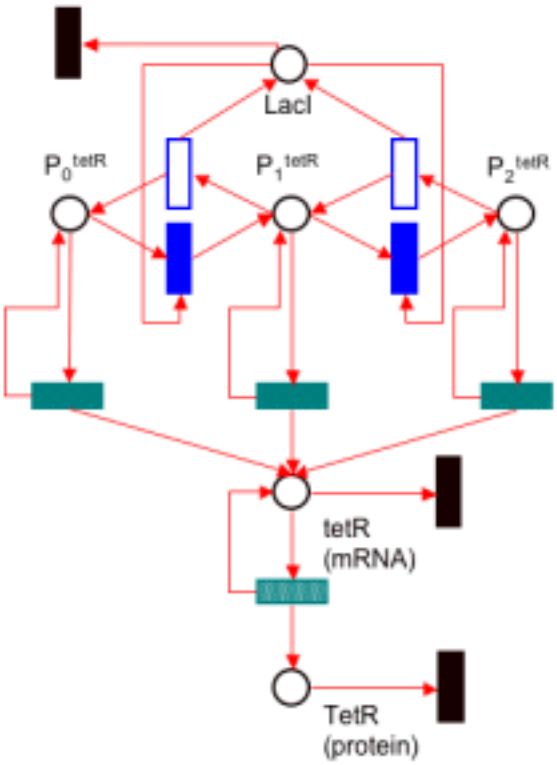



**Fig. 8.13** The Petri net version [39] of one–third of the Repressilator network (the LacI repression of TetR). (Thanks to Myers [75]). The solid lighter vertical rectangles represent binding reactions $A + B \to C$, with rate $k_b[A][B]$; the open vertical rectangles represent unbinding $C \to A + B$, with rate $k_u[C]$. The horizonal rectangles represent catalyzed synthesis reactions $C \to C + P$, with rate $\gamma[C]$; the darker ones represent transcription (formation of mRNA), and the lighter one represent translation (formation of protein). The black vertical rectangles represent degredation reactions, $A \to$ nothing with rate $k_d[A]$. (The stoichiometry of all the arrows is one.) The LacI protein (top) can bind to the DNA in two *promoter sites* ahead of the gene coding for tetR: when bound, it largely blocks the transcription (formation) of tetR mRNA. $P_0$ represents the promotor without any LacI bound; $P_1$ represents the promotor with one site blocked, and $P_2$ represents the doubly-bound promotor. LacI can bind to one or both of the promotor sites, changing $P_i$ to $P_{i+1}$, or correspondingly unbind: the unbinding rate for the protein is modeled in reference [29] to be faster when only one site is occupied. The unbound $P_0$ state transcribes tetR mRNA quickly, and the bound states transcribe it slowly (leaky repression). The tetR mRNA then catalyzes the formation of the TetR protein.[45]

If you are not provided with it, you may retrieve a simulation package for the Repressilator from the book Web site [108].

*(a) Run the simulation for at least 6000 seconds and plot the protein, RNA, and promotor states as a function of time. Notice that*

- *The protein levels do oscillate, as in figure 1(c) in reference [29],*
- *There are significant noisy–looking fluctuations,*
- *There are many more proteins than RNA*

We will study this noise in parts (c) and (d); it will be due to the low numbers of RNA molecules in the cell, and to the discrete fluctuations between the three states of the promotor sites. Before we do this, we should (a) increase the efficiency of the simulation, and (b) compare it to the continuum simulation that would be obtained if there were no fluctuations.

To see how important the fluctuations are, we should compare the stochastic simulation to the solution of the continuum reaction rate equations (as we did in exercise 8.6). In reference [29], the authors write a set of

[43] Messenger RNA (mRNA) codes for proteins. Other forms of RNA can serve as enzymes or parts of the machinery of the cell.

[44] *RNA polymerase*, the molecular motor responsible for transcribing DNA into RNA, needs to attach to the DNA at a *promotor site*. By binding to the adjacent operator sites, our repressor protein inhibits this attachment and hence partly blocks transcription. The residual transcription is called 'leakiness'.

[45] Proteins by convention have the same names as their mRNA, but start with capitals where the mRNA start with small letters.

six differential equations giving a continuum version of the stochastic simulation. These equations are simplified: they both 'integrate out' or coarse–grain away the promotor states from the system, deriving a Hill equation (see exercise 6.10) for the mRNA production, and they also rescale their variables in various ways. Rather than typing in their equations and sorting out these rescalings, it is convenient and illuminating to write a routine to generate the continuum differential equations directly from our reaction rates.

*(b) Write a DeterministicRepressilator, derived from Repressilator just as StochasticRepressilator was. Write a routine dcdt(c, t), that*

- *Sets the chemical amounts in the reaction network to the values in the array c,*
- *Sets a vector dcdt (of length the number of chemicals) to zero,*
- *For each reaction,*
  - *compute its rate*
  - *for each chemical whose stoichiometry is changed by the reaction, add the stoichiometry change times the rate to the corresponding entry of dcdt.*

*Call a routine to integrate the resulting differential equation (as described in the last part of exercise 8.9, for example), and compare your results to those of the stochastic simulation.*

The stochastic simulation has significant fluctuations away from the continuum equation. Part of these fluctuations are due to the fact that the numbers of proteins and mRNAs are small: in particular, the mRNA numbers are significantly smaller than the protein numbers.

*(c) Write a routine that creates a stochastic repressilator network that multiplies the mRNA concentrations by RNAFactor without otherwise affecting the continuum equations. (That is, multiply the initial concentrations and the transcription rates by RNAFactor, and divide the translation rate by RNAFactor.) Try boosting the RNAFactor by ten and one hundred. Do the RNA and protein fluctuations become significantly smaller?* This noise, due to the discrete, integer values of chemicals in the cell, is analogous to the *shot noise* seen in electrical circuits due to the discrete quantum of electric charge. It scales, as do most fluctuations, as the square root of the number of molecules.

A continuum description of the binding of the proteins to the operator sites on the DNA seems particularly dubious: a variable that must be zero or one is replaced by a continuous evolution between these extremes. (Such noise in other contexts is called *telegraph noise* – in analogy to the telegraph, which is either silent or sending as the operator taps the key.) The continuum description is accurate in the limit where the binding and unbinding rates are fast compared to all of the other changes in the system: the protein and mRNA variations then see the average, local equilibrium concentration. On the other hand, if the rates are slow compared to the response of the mRNA and protein, the latter can have a switching appearance.

*(d) Incorporate a telegraphFactor into your stochastic repressilator routine, that multiplies the binding and unbinding rates. Run for 1000 seconds with RNAFactor=10 (to suppress the shot noise) and telegraphFactor = 0.001. Do you observe features in the mRNA curves that appear to switch as the relevant proteins unbind and bind?*

**Advanced Algorithms:** The simulation you will be given implements the Gillespie algorithm discussed in exercise 8.6. At each step, the rates of all possible reactions are calculated, in order to randomly choose when and which the next reaction will be. For a large, loosely connected system of reactions there is no need to recalculate each rate – only the rates which have changed due to the previous reaction. To do so, we must keep track of the dependency network (which chemical amounts affect which reactions change the amounts of which chemicals [74]).

*(e) Alter the reaction network to store the current reaction rates. Add a function UpdateRates(reac) to the reaction network, which for each chem whose stoichiometry is changed by reac, updates the rates for each reaction affected by the amount of chem. Alter the Step method of the stochastic repressilator simulation to use the stored current reaction rates (rather than recomputing them) and to call UpdateRates with the chosen reaction before returning. Time your new routine, and compare to the speed of the old one.* A network of thirty reactions for fifteen chemical components is rather small on biological scales. The dependency network algorithm should be significantly faster for large systems.

(8.8) **Entropy Increases! Markov chains.** (Math)

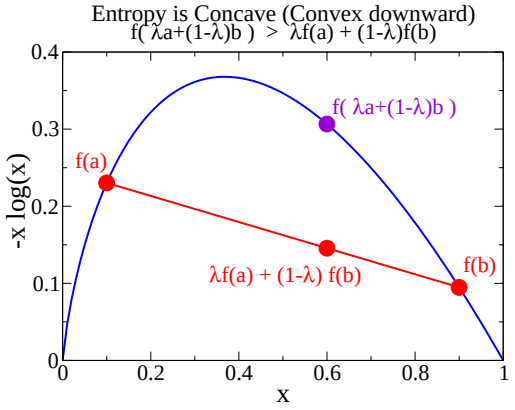



**Fig. 8.14** For $x \geq 0$, $f(x) = -x \log x$ is strictly convex downward (concave) as a function of the probabilities: for $0 < \lambda < 1$, the linear interpolation lies below the curve.

Convexity arguments are a basic tool in formal statistical mechanics. The function $f(x) = -x \log x$ is strictly concave (convex downward) for $x \geq 0$ (figure 8.14): this is easily shown by noting that its second derivative is negative in this region.

*(a) Convexity for sums of many terms. If $\sum_\alpha \mu_\alpha = 1$, and if for all $\alpha$ both $\mu_\alpha \geq 0$ and $x_\alpha \geq 0$, show by induction on the number of states $M$ that if $g(x)$ is concave for $x \geq 0$,*

$$g(\sum_{\alpha=1}^{M} \mu_\alpha x_\alpha) \geq \sum_{\alpha=1}^{M} \mu_\alpha g(x_\alpha). \tag{8.31}$$

(Hint: In the definition of concave, $f(\lambda a + (1-\lambda)b) \geq \lambda f(a) + (1-\lambda) f(b)$, take $(1-\lambda) = \mu_{M+1}$ and $b = x_{M+1}$. Then $a$ is a sum of $M$ terms, rescaled from their original values. Do the coefficients of $x_\alpha$ in $a$ sum to one? Can we apply induction?)

**Microcanonical Entropy is Maximum.** In exercise 6.7, you showed that the microcanonical ensemble was an extremum of the entropy, using Lagrange multipliers. We can use the convexity of $-x \log x$ to show that it's actually a global maximum.

*(b) Using equation 8.31 for $g(x) = -x \log x$ and $\mu_\alpha = 1/M$, show that the entropy for a system of $M$ states $-k_B \sum_\alpha \rho_\alpha \log \rho_\alpha \leq k_B \log M$, the entropy of the (uniform) microcanonical ensemble.*

**Markov Chains: Entropy Increases!** In exercise 5.4 you noticed that, formally speaking, entropy doesn't increase in Hamiltonian systems. Let us show that it does increase for Markov chains.[46]

The Markov chain is implicitly exchanging energy with a heat bath at the temperature $T$. Thus to show that the entropy for the world as a whole increases, we must show that $\Delta S - \Delta E/T$ increases, where $\Delta S$ is the entropy of our system and $\Delta E/T$ is the entropy flow from the heat bath. Hence, showing that entropy increases for our Markov process is equivalent to showing that the free energy $E - TS$ decreases.

Let $P_{\alpha\beta}$ be the transition matrix for a Markov process, satisfying detailed balance with energy $E_\alpha$ at temperature $T$. The current probability of being in state $\alpha$ is $\rho_\alpha$. The free energy

$$F = E - TS = \sum_\alpha \rho_\alpha E_\alpha + k_B T \sum_\alpha \rho_\alpha \log \rho_\alpha. \tag{8.32}$$

*(c) Show that the free energy decreases for a Markov process. In particular, using equation 8.31, show that the free energy for $\rho_\beta^{(n+1)} = \sum_\alpha P_{\beta\alpha} \rho_\alpha^{(n)}$ is less than or equal to the free energy for $\rho^{(n)}$. You may use the properties of the Markov transition matrix $P$, ($0 \leq P_{\alpha\beta} \leq 1$ and $\sum_\alpha P_{\alpha\beta} = 1$), and detailed balance ($P_{\alpha\beta}\rho_\beta^* = P_{\beta\alpha}\rho_\alpha^*$, where $\rho_\alpha^* = \exp(-E_\alpha/k_B T)/Z$).* (Hint: you'll want to use $\mu_\alpha = P_{\alpha\beta}$ in equation 8.31, but the entropy will involve $P_{\beta\alpha}$, which is *not* the same. Use detailed balance to convert from one to the other.)

(8.9) **Solving ODE's: The Pendulum** (Computational) (With Myers. [75])

**Reading:** Numerical Recipes [84], chapter 16.

Physical systems usually evolve continuously in time: their laws of motion are differential equations. Computer simulations must approximate these differential equations using discrete time steps. In this exercise, we will introduce some common methods for simulating differential equations using the example of the pendulum:

$$\frac{d^2\theta}{dt^2} = \ddot{\theta} = -(g/L)\sin(\theta). \tag{8.33}$$

This equation gives the motion of a pendulum with a point mass at the tip of a massless rod[47] of length $L$: rederive it using a free body diagram.

Go to our Web site [108] and download the pendulum files for the language you'll be using. The animation

[46] We know that the Markov chain eventually evolves to the equilibrium state, and we argued that the latter minimizes the free energy. What we're showing here is that the free energy goes continuously downhill for a Markov chain.

[47] We'll depict our pendulum emphasizing the rod rather than the mass: the equation for a physical rod without an end mass is similar.

should show a pendulum oscillating from an initial condition $\theta_0 = 2\pi/3$, $\dot\theta = 0$; the equations being solved have $g = 9.8\text{m/s}^2$ and $L = 1\text{m}$.

There are three independent criteria for picking a good algorithm for solving differential equations: *fidelity*, *accuracy*, and *stability*.

**Fidelity.** Notice that in our time step algorithm, we did not do the straightforward choice – using the current $(\theta(t), \dot\theta(t))$ to produce $(\theta(t+\delta), \dot\theta(t+\delta))$. Rather, we used $\theta(t)$ to calculate the acceleration and update $\dot\theta$, and then used $\dot\theta(t+\delta)$ to calculate $\theta(t+\delta)$.

$$\begin{aligned}\dot\theta(t+\delta) &= \dot\theta(t) + \ddot\theta(t)\,\delta \\ \theta(t+\delta) &= \theta(t) + \dot\theta(t+\delta)\,\delta\end{aligned} \tag{8.34}$$

Wouldn't it be simpler and make more sense to update $\theta$ and $\dot\theta$ simultaneously from their current values, so $\theta(t+\delta) = \theta(t) + \dot\theta(t)\,\delta$? (This simplest of all time-stepping schemes is called the *Euler method*, and should not be used for ordinary differential equations (although it is sometimes used in partial differential equations.)

*(a) Try it. First, see why reversing the order of the updates to $\theta$ and $\dot\theta$,*

$$\begin{aligned}\theta(t+\delta) &= \theta(t) + \dot\theta(t)\,\delta \\ \dot\theta(t+\delta) &= \dot\theta(t) + \ddot\theta(t)\,\delta\end{aligned} \tag{8.35}$$

*in our loop would give us a simultaneous update. Swap these two lines in the code, and watch the pendulum swing for several turns, until it starts looping the loop. Is the new algorithm as good as the old one? (Make sure you switch the two lines back afterwards.)*

The simultaneous update scheme is just as accurate as the one we chose, but it is not as faithful to the physics of the problem: its fidelity is not as good. For subtle reasons we won't explain here, updating first $\dot\theta$ and then $\theta$ allows our algorithm to exactly conserve an approximation to the energy: it's called a *symplectic* algorithm.[48] Improved versions of this algorithm – like the Verlet algorithms below – are often used to simulate systems that conserve energy (like molecular dynamics) because they exactly[49] simulate the dynamics for an approximation to the Hamiltonian – preserving important physical features not kept by just approximately solving the dynamics.

**Accuracy.** Most computational methods for solving differential equations (and many other continuum problems like integrating functions) involve a step size $\delta$, and become more accurate as $\delta$ gets smaller. What is most important is not the error in each time step, but the accuracy of the answer after a fixed time $T$, which is the accumulated error after $T/\delta$ time steps. If this accumulated error varies as $\delta^n$, we say that the algorithm has $n^{\text{th}}$ order cumulative accuracy. Our algorithm is not very high order!

*(b) Plot the pendulum trajectory $\theta(t)$ for time steps $\delta = 0.1$, $0.01$, and $0.001$. Zoom in on the curve at one of the coarse points (say, $t = 1$) and compare the values from the three time steps. Does it appear that this time is converging[50] as $\delta \to 0$? From your measurement, what order accuracy is our method?*

We can write higher–order symplectic algorithms. The approximation to the second derivative

$$\ddot\theta \approx \left(\theta(t+\delta) - 2\theta(t) + \theta(t-\delta)\right)/\delta^2 \tag{8.36}$$

(which you can verify with a Taylor expansion is correct to $O(\delta^4)$) motivates the *Verlet Algorithm*

$$\theta(t+\delta) = 2\theta(t) - \theta(t-\delta) + \ddot\theta\,\delta^2. \tag{8.37}$$

This algorithm is a bit awkward to start up since you need to initialize[51] $\theta(t-\delta)$; it's also often convenient to know the velocities as well as the positions. The *Velocity Verlet* algorithm fixes both of these problems; it is motivated by the constant acceleration formula $x(t) = x_0 + v_0 t + \frac{1}{2}at^2$:

$$\begin{aligned}\theta(t+\delta) &= \theta(t) + \dot\theta(t)\,\delta + \tfrac{1}{2}\ddot\theta(t)\,\delta^2 \\ \dot\theta(t+\delta/2) &= \dot\theta(t) + \tfrac{1}{2}\ddot\theta(t)\,\delta \\ \dot\theta(t+\delta) &= \dot\theta(t+\delta/2) + \tfrac{1}{2}\ddot\theta(t+\delta)\,\delta.\end{aligned} \tag{8.38}$$

The trick that makes this algorithm so good is to cleverly split the velocity increment into two pieces, half for the

[48] It conserves a *symplectic form*. In non-mathematician's language, this means our time-step perfectly simulates a Hamiltonian system satisfying Liouville's theorem and energy conservation, but with an approximation to the true energy.

[49] Up to rounding errors

[50] You may note that the approximate answers seem to extrapolate nicely to the correct answer. Using this can allow one to converge more quickly to the correct answer. This is called Richardson extrapolation and is the basis for the Bulirsch–Stoer methods.

[51] Since we start with $\dot\theta = 0$, the simulation is symmetric under reversing the sign of time and you can get away with using $\theta(t-\delta) = \theta(t) + \frac{1}{2}\ddot\theta + O(\delta^4)$.

acceleration at the old position and half for the new position.[52] (You'll want to initialize $\ddot{\theta}$ once before starting the loop.)

*(c) Pick one of the Verlet algorithms, implement it, and plot the trajectory for time steps* $\delta = 0.1$, $0.01$, *and* $0.001$. *You should see a dramatic improvement in convergence. What cumulative order accuracy does Verlet have?*[53]

**Stability.** In many cases high accuracy is not crucial. What prevents us from taking enormous time steps? In a given problem, there is usually a typical fastest time scale: a vibration or oscillation period (as in our exercise) or a growth or decay rate. When our time step becomes a substantial fraction of this fastest time scale, algorithms like ours usually become *unstable*: the first few time steps may be fairly accurate, but small errors build up until the errors become unacceptable (indeed, often ones first warning of problems are machine overflows).

*(d) Plot the pendulum trajectory* $\theta(t)$ *for time steps* $\delta = 0.1, 0.2, \ldots, 0.8$, *using a small amplitude oscillation* $\theta_0 = 0.01$, $\dot{\theta}_0 = 0.0$, *up to* $t_{\max} = 10$. *At about what* $\delta_c$ *does it go unstable? Looking at the first few points of the trajectory, does it seem like sampling the curve at steps much larger than* $\delta_c$ *would miss the oscillations? At* $\delta_c/2$, *how accurate is the amplitude of the oscillation?* (You'll need to observe several periods in order to estimate the maximum amplitude of the solution.)

In solving the properties of large, nonlinear systems (*e.g.*, partial differential equations (PDEs) and molecular dynamics) stability tends to be the key difficulty. The maximum stepsize depends on the local configuration, so highly nonlinear regions can send the system unstable before one might expect. The maximum safe stable stepsize often has accuracy far higher than needed; indeed, some algorithms become less stable if the stepsize is decreased![54]

**ODE packages**: higher order, variable stepsize, stiff systems ...

The Verlet algorithms are not hard to code, and we use higher–order symplectic algorithms in Hamiltonian systems mostly in unusual applications (planetary motion) where high accuracy is demanded, because they are typically significantly less stable. In systems of differential equations where there is no conserved energy or Hamiltonian, or even in Hamiltonian systems (like high–energy collisions) where accuracy at short times is more crucial than fidelity at long times, we use general purpose methods.

The general–purpose solvers come in a variety of basic algorithms (Runge–Kutta, predictor–corrector, ...), and methods for maintaining and enhancing accuracy (variable step size, Richardson extrapolation). There are also *implicit* methods for *stiff* systems. A system is stiff if there is a large separation between the slowest and fastest relevant time scales: implicit methods often allow one to take time steps much larger than the fastest time scale (unlike the explicit Verlet methods you studied in part (d), which go unstable). Large, sophisticated packages have been developed over many years for solving differential equations – switching between algorithms and varying the time steps to most efficiently maintain a given level of accuracy. They solve $d\mathbf{y}/dt = \mathbf{dydt}(\mathbf{y}, t)$, where for us $\mathbf{y} = [\theta, \dot{\theta}]$ and $\mathbf{dydt} = [\dot{\theta}, \ddot{\theta}]$. They typically come in the form of subroutines or functions, which need as arguments

- Initial conditions $\mathbf{y}_0$,
- The right–hand side **dydt**, a function of the vector $\mathbf{y}$ and time $t$, which returns a vector giving the current rate of change of $\mathbf{y}$, and
- The initial and final times, and perhaps intermediate times, at which the trajectory $\mathbf{y}(t)$ is desired.

They often have options that

- Ask for desired accuracy goals, typically a relative (fractional) accuracy and an absolute accuracy, sometimes set separately for each component of $\mathbf{y}$,
- Ask for and return derivative and time step information from the end of the last step (to allow efficient restarts after intermediate points),
- Ask for a routine that computes the derivatives of **dydt** with respect to the current components of $\mathbf{y}$ (for use by the stiff integrator), and

[52] You may check that both Verlet algorithms give exactly the same values for $\theta(t_0 + n\delta)$.

[53] The error in each time step of the Verlet algorithm is of order $\delta^4$. It's usually said that the Verlet algorithms have third order accuracy, naively assuming that running for a time $T$ should have errors bounded by the number of time steps $T/\delta$ times the error per time step $\delta^4$. However, one can check that the errors in successive time steps build up quadratically at short times (*i.e.*, the velocity errors build up linearly with time), so after $T/\delta$ time steps the accumulated error is $\delta^4 (T/\delta)^2 \propto \delta^2$. We'll use "cumulative order" of the algorithm to distinguish it from the naive order.

[54] For some partial differential equations, decreasing the spacing $\Delta x$ between points can lead to instabilities unless the time step is also decreased.

- Return information about the methods, time steps, and performance of the algorithm.

You will be supplied with one of these general–purpose packages, and instructions on how to use it.

*(e) Write the function* **dydt**, *and use the general purpose solver to solve for the motion of the pendulum as in parts (a)-(c), and informally check that the trajectory is accurate.*

(8.10) **Small World Networks.** (Complexity, Computation) (With Myers. [75])

Many interesting problems arise from studying properties of randomly generated networks. A network is a collection of *nodes* and *edges*, with each edge connected to two nodes, but with each node potentially connected to any number of edges. A random network is constructed probabilistically according to some definite rules; studying such a random network usually is done by studying the entire ensemble of networks, each weighted by the probability that it was constructed. Thus these problems naturally fall within the broad purview of statistical mechanics.

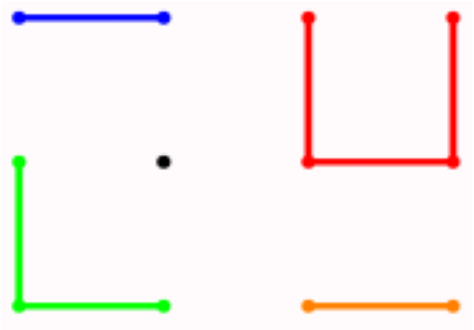


**Fig. 8.15** A network is a collection of nodes (circles) and edges (lines between the circles).

One of the more popular topics in random network theory is the study of how connected they are. "Six degrees of separation" is the phrase commonly used to describe the interconnected nature of human acquaintances: various somewhat uncontrolled studies have shown that any random pair of people in the world can be connected to one another by a short chain of people (typically around six), each of whom knows the next fairly well. If we represent people as nodes and acquaintanceships as neighbors, we reduce the problem to the study of the relationship network.

In this exercise, we will generate some random networks, and calculate the distribution of distances between pairs of points. We'll study *small world networks* [121, 76], a theoretical model that suggests how a small number of shortcuts (unusual international and intercultural friendships, ) can dramatically shorten the typical chain lengths. Finally, we'll study how a simple, universal scaling behavior emerges for large networks with few shortcuts.

On the Web site for this book [108], you'll find some hint files and graphic routines to facilitate working this exercise, for a variety of languages and systems (currently Python under Unix and Windows).

**Constructing a small world network.** The $L$ nodes in a small world network are arranged around a circle. There are two kinds of edges. Each node has $Z$ short edges connecting it to its nearest neighbors around the circle (up to a distance $Z/2$). In addition, there are $p \times L \times Z/2$ shortcuts added to the network, which connect nodes at random (see figure 8.16). (This is a more tractable version [76] of the original model [121], which rewired a fraction $p$ of the $LZ/2$ edges.)

*(a) Define a network object on the computer. For this exercise, the nodes will be represented by integers. Implement a network class, with five functions:*

*(1)* `HasNode(node)`, *which checks to see if a node is already in the network,*

*(2)* `AddNode(node)`, *which adds a new node to the system (if it's not already there),*

*(3)* `AddEdge(node1, node2)`, *which adds a new edge to the system,*

*(4)* `GetNodes()`, *which returns a list of existing nodes, and*

*(5)* `GetNeighbors(node)`, *which returns the neighbors of an existing node.*

*Write a routine to construct a small–world network, which (given $L$, $Z$, and $p$) adds the nodes and the short edges, and then randomly adding the shortcuts. Use the software provided to draw this small world graph, and check that you've implemented the periodic boundary conditions correctly (each node $i$ should be connected to nodes $(i - Z/2) \mathrm{mod} L, \ldots, (i + Z/2) \mathrm{mod} L$).*

**Fig. 8.16** Small world network, with $L = 20$, $Z = 4$, and $p = 0.2$.[55]

**Measuring the minimum distances between nodes.** The most studied property of small world graphs is the distribution of shortest paths between nodes. Without the long edges, the shortest path between $i$ and $j$ will be given by hopping in steps of length $Z/2$ along the shorter of the two arcs around the circle: there will be no paths of length longer than $L/Z$ (halfway around the circle), and the distribution $\rho(\ell)$ of path lengths $\ell$ will be constant for $0 < \ell < L/Z$. When we add shortcuts, we expect that the distribution will be shifted to shorter path lengths.

*(b) Write three functions to find and analyze the path length distribution:*

*(1)* `FindPathLengthsFromNode(graph, node)`*, which returns for each* `node2` *in the graph the shortest distance from* `node` *to* `node2`*. An efficient algorithm is a* breadth first traversal *of the graph, working outward from* `node` *in shells. There will be a* `currentShell` *of nodes whose distance will be set to $\ell$ unless they have already been visited, and a* `nextShell` *which will be considered after the current one is finished (looking sideways before forward, breadth–first):*

- *Initialize $\ell = 0$, the distance from* `node` *to itself to zero, and* `currentShell = [node]`
- *While there are nodes in the new* `currentShell`*:*
  - *Start a new empty* `nextShell`
  - *For each neighbor of each node in the current shell, if the distance to* `neighbor` *has not been set, add the node to* `nextShell` *and set the distance to $\ell + 1$*
  - *Add one to $\ell$, and set the current shell to* `nextShell`
- *Return the distances*

This will sweep outward from `node`, measuring the shortest distance to every other node in the network. (Hint: Check your code with a network with small $N$ and small $p$, comparing a few paths to hand–calculations from the graph image generating as in part (a).)

*(2)* `FindPathLengthHistogram(graph)`*, which computes the probability $\rho(\ell)$ that a shortest path will have length $\ell$, by using FindPathLengthsFromNode repeatedly to find the mean over all pairs of nodes. Check your function by testing that the histogram of path lengths at $p = 0$ is constant for $0 < \ell < L/Z$, as advertised. Generate graphs at $L = 1000$ and $Z = 2$ for $p = 0.02$ and $p = 0.2$: display the circle graphs and plot the histogram of path lengths. Zoom in on the histogram: how much does it change with $p$? What value of $p$ would you need to get 'six degrees of separation'?*

*(3)* `FindAveragePathLength(graph)`*, which similarly computes the mean $\langle \ell \rangle$ over all pairs of nodes. Compute $\ell$ for $Z = 2$, $L = 100$, and $p = 0.1$ a few times: your answer should be around $\ell = 10$. Notice that there are substantial statistical fluctuations in the value from sample to sample. Roughly how many long bonds are there in this system? Would you expect fluctuations in the distances?*

*(c) Plot the average path length between nodes $\ell(p)$ divided by $\ell(p = 0)$ for $Z = 2$, $L = 50$, with $p$ on a semi-log plot from $p = 0.001$ to $p = 1$. Compare with figure 2 of Watts and Strogatz [121]. You should find roughly the same curve, with the values of $p$ shifted by a factor of 100. (They do $L = 1000$ and $Z = 10$).*

**Large $N$ and the emergence of a continuum limit.** We can understand the shift in $p$ of part (c) as a continuum limit of the problem. In the limit where the number of nodes $N$ becomes large and the number of short cuts $pLZ/2$ stays fixed, this network problem has a nice limit where distance is measured in radians $\Delta\theta$ around the circle. Dividing $\ell$ by $\ell(p = 0) \approx L/(2Z)$ essentially does this, since $\Delta\theta = \pi Z \ell / L$.

*(d) Create and display a circle graph of your geometry from part (c) [$Z = 2$, $L = 50$] at $p = 0.1$; create and display circle graphs of Watts and Strogatz' geometry [$Z = 10$, $L = 1000$] at $p = 0.1$ and $p = 0.001$. Which*

[55] There are seven new shortcuts, where $pLZ/2 = 8$; one of the added edges overlapped an existing edge or connected a node to itself.

*of their systems looks statistically more similar to yours? Plot (perhaps using the scaling collapse routine provided) the rescaled average path length $\pi Z\ell/L$ versus the total number of shortcuts $pLZ/2$, for a range $0.001 < p < 1$, for $L = 100$ and 200 and $Z = 2$ and 4.*

In this limit, the average bond length $\langle \Delta\theta \rangle$ should be a function only of $M$. Since reference [121] ran at a value of $ZL$ a factor of 100 larger than ours, our values of $p$ are a factor of 100 larger to get the same value of $M = pLZ/2$. Newman and Watts [79] derive this continuum limit with a renormalization–group analysis (chapter 12).

*(e)* **Real Networks.** *From the book Web site [108], or through your own researches, find a real network*[56] *and find the mean distance and histogram of distances between nodes.*

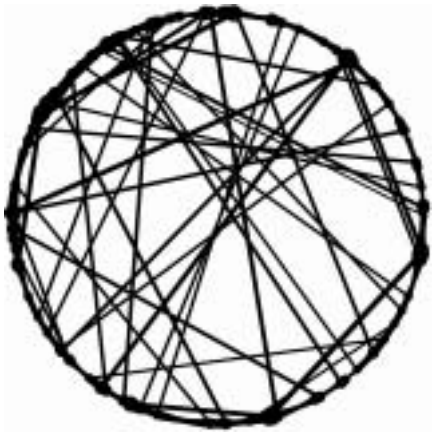


**Fig. 8.17** Small world network with $L = 500$, $K = 2$ and $p = 0.1$, with node and edge sizes scaled by the square root of their betweenness.

In the small–world network, a few long edges are crucial to efficient transfer through the system (transfer of information in a computer network, transfer of disease in a population model, ...). It is often useful to measure how crucial a given node or edge is to these shortest paths. We say a node or edge is "between" two other nodes if it is along a shortest path between them. We measure the "betweenness" of a node or edge as the total number of such shortest paths passing through, with (by convention) the initial and final nodes included in the 'between' nodes; see figure 8.17. (If there are $K$ multiple shortest paths of equal length between two nodes, each path adds $1/K$ to its intermediates.) The efficient algorithm to measure betweenness is a depth–first traversal quite analogous to the shortest–path–length algorithm discussed above.

*(f)* **Betweenness.** *(Advanced) Read references [77] and [37] , discussing the algorithms for finding the betweenness. Implement them on the small world network, and perhaps the real–world network you analyzed in part (e). Visualize your answers by using the graphics software provided on the book Web site [108].*

(8.11) **Building a Percolation Network.** (Complexity,Computation) (With Myers. [75])

Figure 8.18 shows what a large sheet of paper, held at the edges, would look like if small holes were successively punched out at random locations. Here the ensemble averages over the different choices of random locations for the holes; this figure shows the sheet just before it fell apart. Of course, certain choices of hole positions would cut the sheet in two far earlier (a straight line across the center) or somewhat later (checkerboard patterns), but for the vast majority of members of our ensemble the paper will have the same kinds of hole patterns seen here. Again, it is easier to analyze all the possible patterns of punches than to predict a particular pattern.

Percolation theory is the study of the qualitative change in connectivity of a large system as its components are randomly removed. Outside physics, it has become a prototype of criticality at continuous transitions, presumably because the problem is simple to state and the analysis does not demand a background in equilibrium statistical mechanics.[57] In this exercise, we'll study bond percolation (figure 8.18) and site percolation (8.19) in two dimensions.

[56] Noteworthy examples include movie-actor costars, "Six degrees of Kevin Bacon" or baseball players who played on the same team.

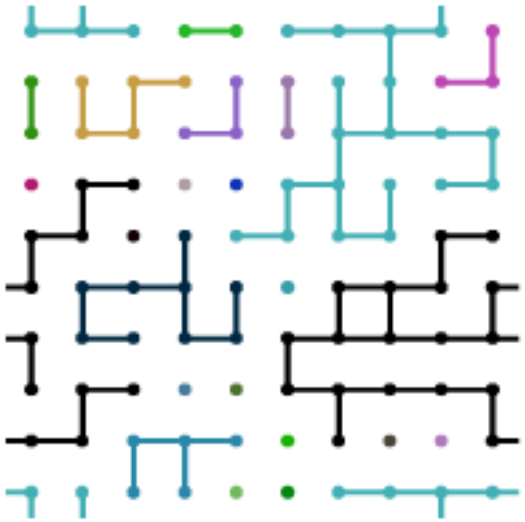


**Fig. 8.18 Bond Percolation network.** Each bond on a $10 \times 10$ square lattice is present with probability $p = 0.4$. This is below the percolation threshold $p = 0.5$ for the infinite lattice, and indeed the network breaks up into individual clusters (each shaded separately). Note the periodic boundary conditions. Note there are many small clusters, and only a few large ones, here twelve clusters of size $S = 1$, three of size $S = 2$, and one cluster of size $S = 29$ (black). For a large lattice near the percolation threshold the probability distribution of cluster sizes $\rho(S)$ forms a power law (exercise 12.9).

On the Web site for this book [108], you'll find some hint files and graphic routines to facilitate working this exercise, for a variety of languages and systems (currently Python under Unix and Windows).

**Bond percolation on a square lattice.**

*(a) Define a 2D bond percolation network with periodic boundary conditions on the computer, for size $L \times L$ and bond probability $p$. For this exercise, the nodes will be represented by pairs of integers $(i, j)$. You'll need the method* `GetNeighbors(node)`*, which returns the neighbors of an existing node. Use the bond-drawing software provided to draw your bond percolation network for various $p$ and $L$, and use it to check that you've implemented the periodic boundary conditions correctly.* (There are two basic approaches. You can start with an empty network and use `AddNode` and `AddEdge` in loops to generate the nodes, vertical bonds, and horizontal bonds (see exercise 8.10). Alternatively, and more traditionally, you can set up a 2D array of vertical and horizontal bonds, and implement `GetNeighbors(node)` by constructing the list of neighbors from the bond networks when the site is visited.)

**The percolation threshold and duality.** In most continuous phase transitions, one of the challenges is to find the location of the transition. We chose bond percolation on the square lattice because one can argue, in the limit of large systems, that the percolation threshold $p_c = 1/2$. The argument makes use of the *dual lattice*.

The nodes of the dual lattice are the centers of the squares between nodes in the original lattice. The edges of the dual lattice are those which do not cross an edge of the original lattice. Since every potential dual edge crosses exactly one edge of the original lattice, the probability $p^*$ of having bonds on the dual lattice is $1 - p$ where $p$ is the probability of bonds for the original lattice. If we can show that the dual lattice percolates if and only if the original lattice does not, then $p_c = 1/2$. This is easiest to see graphically:

*(b) Generate and print a small lattice with $p = 0.4$, picking one where the largest cluster does not span across either the vertical or the horizontal direction (or print figure 8.18). Draw a path on the dual lattice spanning the system from top to bottom and from left to right. (You'll be emulating a rat running through a maze.) Is it clear for large systems that the dual lattice will percolate if and only if the original lattice does not?*

**Finding the clusters.** *(c) Write two functions that together find the clusters in the percolation network:*

*(1)* `FindClusterFromNode(graph, node, visited)`*, which returns the cluster in* `graph` *containing* `node`*, and marks the sites in the cluster as having been* `visited`*. The cluster is of course the union of* `node`*, the neighbors, the neighbors of the neighbors, etc. The trick is to use the set of* `visited` *sites to avoid going around in circles. The efficient algorithm is a* breadth first traversal *of the graph, working outward from* `node` *in shells. There will be a* `currentShell` *of nodes whose neighbors have not yet been checked, and a* `nextShell` *which will be considered after the current one is finished (breadth–first):*

- *Initialize* `visited[node]=True`*,* `cluster=[node]`*, and* `currentShell=graph.GetNeighbors(node)`*.*
- *While there are nodes in the new* `currentShell`*:*
  - *Start a new empty* `nextShell`
  - *For each node in the current shell, if the node has not been visited,*
    - *add the node to the cluster,*
    - *mark the node as visited,*
    - *and add the neighbors of the node to the* `nextShell`
  - *Set the current shell to* `nextShell`
- *Return the cluster*

*(2)* `FindAllClusters(graph)`*, which sets up the* `visited` *set to be False for all nodes, and calls* `FindClusterFromNode(graph, node, visited)` *on all nodes that haven't been visited, collecting the resulting clusters. Optionally, you may want to order the clusters from largest to smallest, for convenience in graphics (and in finding the largest cluster).*

*Check your code by running it for small L and using the graphics software provided. Are the clusters, drawn in different colors, correct?*

**Site percolation on a triangular lattice.** Universality states that the statistical behavior of the percolation clusters at long length scales should be independent of the microscopic detail. That is, removing bonds from a square lattice should leave the same fractal patterns of holes, near $p_c$, as punching out circular holes in a sheet just before it falls apart. Nothing about your algorithms from part (c) depended on their being four neighbors of a node, or their even being nodes at all sites. Let's implement site percolation on a triangular lattice (figure 8.19): nodes are occupied with probability $p$, with each node connected to any of its six neighbor sites that are also filled (punching out hexagons from a sheet of paper). The triangular site lattice also has a duality transformation, so again $p_c = 0.5$.

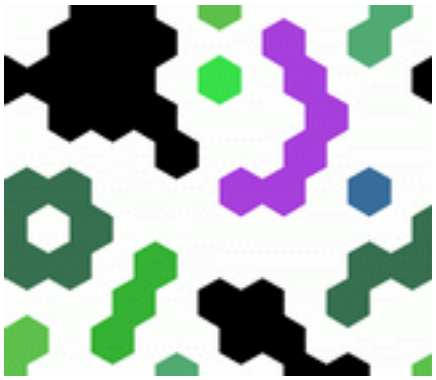


**Fig. 8.19 Site Percolation network.** Each site on a $10 \times 10$ triangular lattice is present with probability $p = 0.5$, the percolation threshold for the infinite lattice. Note the periodic boundary conditions at the sides, and the shifted periodic boundaries at the top and bottom.

It is computationally convenient to label the site at $(x, y)$ on a triangular lattice by $[i, j]$, where $x = i + j/2$ and $y = \frac{\sqrt{3}}{2} j$. If we again use periodic boundary conditions with $0 \leq i < L$ and $0 \leq j < L$, we cover a region in the shape of a 60° rhombus.[58] Each site $[i, j]$ has six neighbors, at $[i, j] + e$ with $e = [1, 0], [0, 1], [-1, 1]$ upward and to the right and minus the same three downward and left.

*(d) Generate a site percolation network on a triangular lattice. You can treat the sites one at a time, using AddNode with probability p, and check HasNode(neighbor) to bond to all existing neighbors. Alternatively, you can start by generating a whole matrix of random numbers in one sweep to determine which sites are occupied by nodes, add those nodes, and then fill in the bonds. Check your resulting network by running it for small L and using the graphics software provided. (Notice the shifted periodic boundary conditions at the top and bottom, see figure 8.19.) Use your routine from part (c) to generate the clusters, and check these (particularly at the periodic boundaries) using the graphics software.*

*(e) Generate a small square–lattice bond percolation cluster, perhaps* $30 \times 30$*, and compare with a small triangular–lattice site percolation cluster. They should look rather different in many ways. Now generate a large*[59] *cluster of each, perhaps* $1000 \times 1000$ *(or see figure 12.9). Stepping back and blurring your eyes, do the two look substantially similar?*

Chapter 12 and 12.9 will discuss percolation theory in more detail.

(8.12) **Hysteresis Model: Computational Methods.** (Complexity)

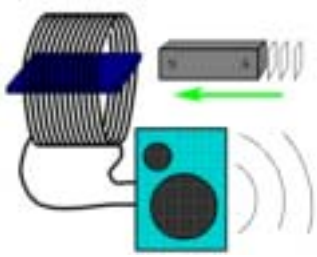


**Fig. 8.20** Barkhausen noise experiment.

[58] The graphics software uses the periodic boundary conditions to shift this rhombus back into a rectangle.

[59] Your code, if written properly, should run in a time of order $N$, the number of nodes. If it seems to slow down more than a factor of 4 when you increase the length of the side by a factor of two, check for inefficiencies.

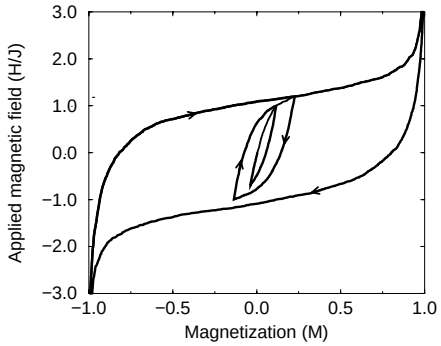



**Fig. 8.21** Hysteresis loop with subloops.

**Fig. 8.22** Tiny jumps: Barkhausen noise.

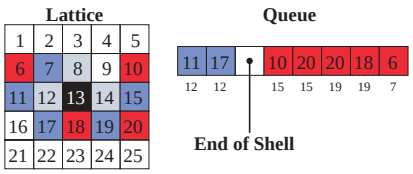



**Fig. 8.23** Avalanche propagation in the hysteresis model.

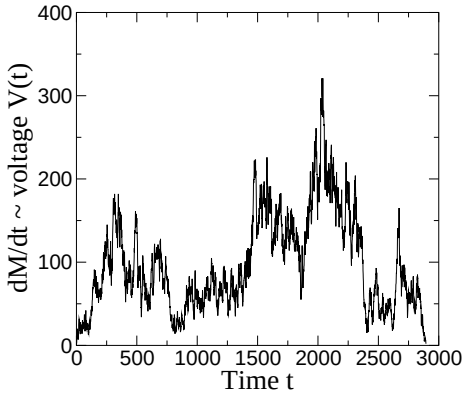



**Fig. 8.24** Number of domains flipped per time step for the avalanche shown in figure 12.2. Notice how the avalanche almost stops several times: if the forcing were slightly smaller compared to the disorder, the avalanche would have separated into smaller ones. The fact that the disorder is just small enough to keep the avalanche growing is the criterion for the phase transition, and the cause of the self–similarity. At the critical point, a partial avalanche of size $S$ will on average trigger another one of size $S$.

# Order Parameters, Broken Symmetry, and Topology

# 9

This chapter is slightly modified from a lecture given at the Santa Fe institute [96].

As a kid in elementary school, I was taught that there were three states of matter: solid, liquid, and gas. The ancients thought that there were four: earth, water, air, and fire, which was considered sheer superstition. In junior high, I remember reading a book called *The Seven States of Matter*. At least one was "plasma", which made up stars and thus most of the universe,[1] and which sounded rather like fire to me.

[1] They hadn't heard of dark matter back then.

The original three, by now, have become multitudes. In important and precise ways, magnets are a distinct form of matter. Metals are different from insulators. Superconductors and superfluids are striking new states of matter. The liquid crystal in your wristwatch is one of a huge family of different liquid crystalline states of matter [26] (nematic, cholesteric, blue phase I, II, and blue fog, smectic A, B, C, C*, D, I, ...). There are over 200 qualitatively different types of crystals, not to mention the quasicrystals (figure 9.1). There are disordered states of matter like spin glasses, and states like the fractional quantum hall effect with excitations of charge $e/3$ like quarks. Particle physicists tell us that the vacuum we live within has in the past been in quite different states: in the last vacuum but one, there were four different kinds of light [24] (mediated by what is now the photon, the $W^+$, the $W^-$, and the Z particle).

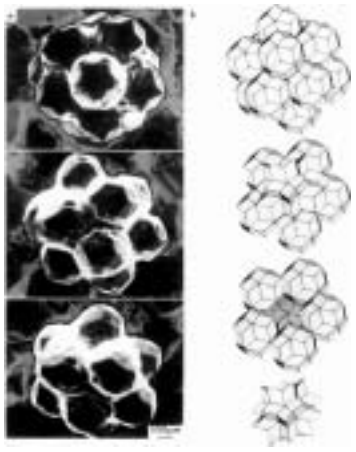


**Fig. 9.1 Quasicrystals.** Much of this chapter will discuss the properties of crystals. Crystals are surely the oldest known of the broken–symmetry phases of matter, and remain the most beautiful illustrations. It's amazing that in the past few years, we've uncovered an entirely new class of crystals. Shown here is a photograph of a quasicrystalline metallic alloy, with icosahedral symmetry. Notice that the facets are pentagonal: our old notions of crystals had to be completely revised to include this type of symmetry.

When there were only three states of matter, we could learn about each one and then turn back to learning long division. Now that there are multitudes, though, we've had to develop a system. Our system is constantly being extended and modified, because we keep finding new phases which don't fit into the old frameworks. It's amazing how the 500th new state of matter somehow screws up a system which worked fine for the first 499. Quasicrystals, the fractional quantum hall effect, and spin glasses all really stretched our minds until (1) we understood why they behaved the way they did, and (2) we understood how they fit into the general framework.

In this chapter, I'm going to tell you the system.

The system consists of four basic steps [73]. First, you must identify the broken symmetry. Second, you must define an order parameter. Third, you are told to examine the elementary excitations. Fourth, you classify the topological defects. Most of what I say in this chapter I take from Mermin [73], Coleman [24], and deGennes [26], and I heartily recommend these excellent articles. We take each step in turn.

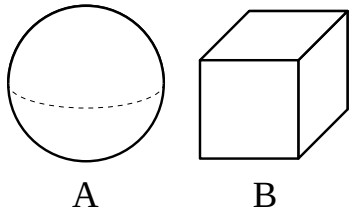



**Fig. 9.2 Which is more symmetric?** (A) The cube has many symmetries. It can be rotated by 90°, 180°, or 270° about any of the three axes passing through the faces. It can be rotated by 120° or 240° about the corners and by 180° about an axis passing from the center through any of the 12 edges. (B) The sphere, though, can be rotated by *any* angle. The sphere respects rotational invariance: all directions are equal. The cube is an object which breaks rotational symmetry: once the cube is there, some directions are more equal than others.

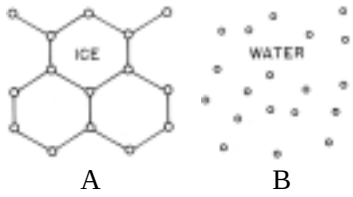



**Fig. 9.3 Which is more symmetric?** At first glance, water seems to have much less symmetry than ice. (A) The picture of "two–dimensional" ice clearly breaks the rotational invariance: it can be rotated only by 120° or 240°. It also breaks the translational invariance: the crystal can only be shifted by certain special distances (whole number of lattice units). (B) The picture of water has no symmetry at all: the atoms are jumbled together with no long–range pattern at all. Water, though, isn't a snapshot: it is better to think of it as a combination (or ensemble) of all possible snapshots. Water has a complete rotational and translational symmetry: the pictures will look the same if the container is tipped or shoved.

## 9.1 Identify the Broken Symmetry

What is it which distinguishes the hundreds of different states of matter? Why do we say that water and olive oil are in the same state (the liquid phase), while we say aluminum and (magnetized) iron are in different states? Through long experience, we've discovered that most phases differ in their symmetry.[2]

Consider figures 9.2, showing a cube and a sphere. Which is more symmetric? Clearly, the sphere has many more symmetries than the cube. One can rotate the cube by 90° in various directions and not change its appearance, but one can rotate the sphere by any angle and keep it unchanged.

In figure 9.3, we see a 2-D schematic representation of ice and water. Which state is more symmetric here? Naively, the ice looks much more symmetric: regular arrangements of atoms forming a lattice structure. The water looks irregular and disorganized. On the other hand, if one rotated figure 9.3B by an arbitrary angle, it would still look like water! Ice has broken rotational symmetry: one can rotate figure 9.3A only by multiples of 60°. It also has a broken translational symmetry: it's easy to tell if the picture is shifted sideways, unless one shifts by a whole number of lattice units. While the snapshot of the water shown in the figure has no symmetries, water as a phase has complete rotational and translational symmetry.

## 9.2 Define the Order Parameter

Particle physics and condensed–matter physics have quite different philosophies. Particle physicists are constantly looking for the building blocks. Once pions and protons were discovered to be made of quarks, they became demoted into engineering problems. Now quarks and electrons and photons seem to be made of strings, and strings are hard to study experimentally (so far). Condensed–matter physicists, on the other hand, try to understand why messy combinations of zillions of electrons and nuclei do such interesting simple things. To them, the fundamental question is not discovering the underlying quantum mechanical laws, but in understanding and explaining the new laws that emerge when many particles interact.[3]

As one might guess, we don't always keep track of all the electrons and protons. We're always looking for the important variables, the important degrees of freedom. In a crystal, the important variables are the

[2] This is not to say that different phases always differ by symmetries! Liquids and gases have the same symmetry. It is safe to say, though, that if the two materials have different symmetries, they are different phases.

[3] The particle physicists use order parameter fields too. Their order parameter fields also hide lots of details about what their quarks and gluons are composed of. The main difference is that they don't know of what their fields are composed. It ought to be reassuring to them that we don't always find our greater knowledge very helpful.

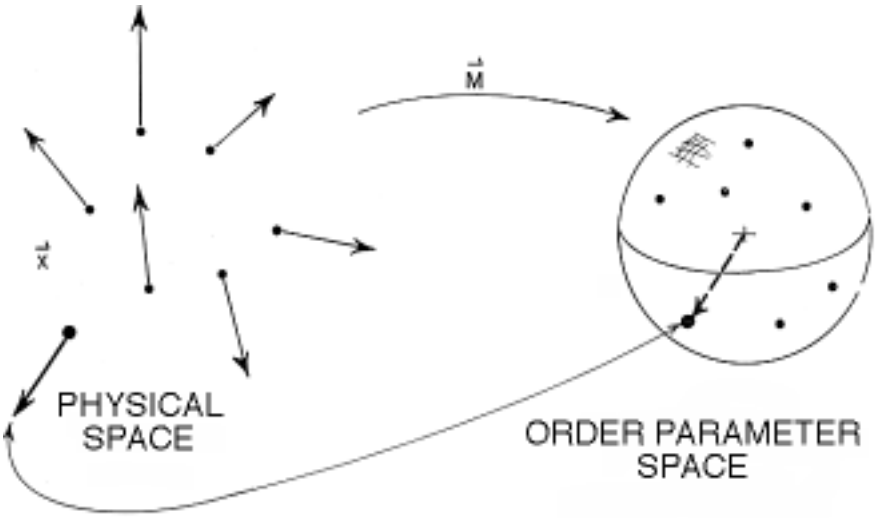



**Fig. 9.4 Magnet.** We take the magnetization $\vec{M}$ as the order parameter for a magnet. For a given material at a given temperature, the amount of magnetization $|\vec{M}| = M_0$ will be pretty well fixed, but the energy is often pretty much independent of the direction $\hat{M} = \vec{M}/M_0$ of the magnetization. (You can think of this as a arrow pointing to the north end of each atomic magnet.) Often, the magnetization changes directions smoothly in different parts of the material. (That's why not all pieces of iron are magnetic!) We describe the current state of the material by an order parameter field $\vec{M}(\mathbf{x})$.
The order parameter field is usually thought of as an arrow at each point in space. It can also be thought of as a function taking points in space $\mathbf{x}$ into points on the sphere $|\vec{M}| = M_0$. This sphere $\mathbb{S}^2$ is the order parameter space for the magnet.

motions of the atoms away from their lattice positions. In a magnet, the important variable is the local direction of the magnetization (an arrow pointing to the "north" end of the local magnet). The local magnetization comes from complicated interactions between the electrons, and is partly due to the little magnets attached to each electron and partly due to the way the electrons dance around in the material: these details are for many purposes unimportant.

The important variables are combined into an "order parameter field". In figure 9.4, we see the order parameter field for a magnet.[4] At each position $\mathbf{x} = (x, y, z)$ we have a direction for the local magnetization $\vec{M}(\mathbf{x})$. The length of $\vec{M}$ is pretty much fixed by the material, but the direction of the magnetization is undetermined. By becoming a magnet, this material has broken the rotational symmetry. The order parameter $\vec{M}$ labels which of the various broken symmetry directions the material has chosen.

The order parameter is a field: at each point in our magnet, $\vec{M}(\mathbf{x})$ tells the local direction of the field near $\mathbf{x}$. Why do we do this? Why would the magnetization point in different directions in different parts of the magnet? Usually, the material has lowest energy when the order parameter field is uniform, when the symmetry is broken in the same way throughout space. In practice, though, the material often doesn't break symmetry uniformly. Most pieces of iron don't appear magnetic, simply because the local magnetization points in different directions at different places. The magnetization is already there at the atomic level: to make a magnet, you pound the different domains until they line up. We'll see in this chapter that most of the interesting behavior we can

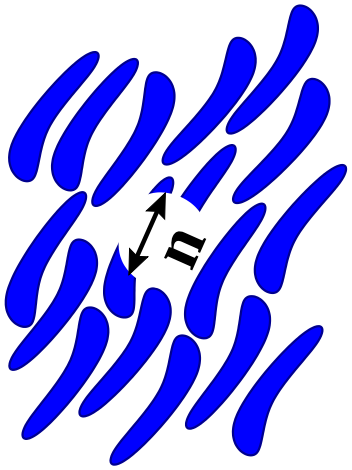



**Fig. 9.5 Nematic liquid crystals** are made up of long, thin molecules that prefer to align with one another. (Liquid crystal watches are made of nematics.) Since they don't care much which end is up, their order parameter is not a vector $\hat{n}$ along the axis of the molecules, but is instead a unit vector up to the equivalence $\hat{n} \equiv -\hat{n}$.

[4] Most magnets are crystals, which already have broken the rotational symmetry. For some "Heisenberg" magnets, the effects of the crystal on the magnetism is small. Magnets are really distinguished by the fact that they break time–reversal symmetry: if you reverse the arrow of time, the magnetization changes sign.

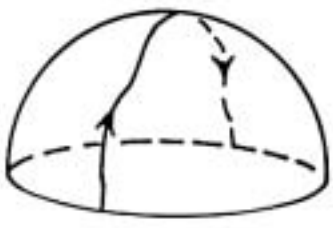


**Fig. 9.6 The nematic order parameter space** is a half-sphere, with antipodal points on the equator identified. Thus, for example, the path shown over the top of the hemisphere is a closed loop: the two intersections with the equator correspond to the same orientations of the nematic molecules in space.

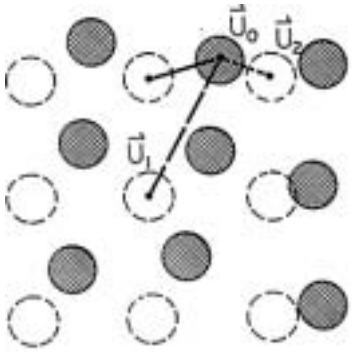



**Fig. 9.7 Two dimensional crystal.** A crystal consists of atoms arranged in regular, repeating rows and columns. At high temperatures, or when the crystal is deformed or defective, the atoms will be displaced from their lattice positions. The displacements $\vec{u}$ are shown. Even better, one can think of $u(\mathbf{x})$ as the local translation needed to bring the ideal lattice into registry with atoms in the local neighborhood of $\mathbf{x}$. Also shown is the ambiguity in the definition of $u$. Which "ideal" atom should we identify with a given "real" one? This ambiguity makes the order parameter $u$ equivalent to $u + ma\hat{x} + na\hat{y}$. Instead of a vector in two dimensions, the order parameter space is a square with periodic boundary conditions.

study involves the way the order parameter varies in space.

The order parameter field $\vec{M}(\mathbf{x})$ can be usefully visualized in two different ways. On the one hand, one can think of a little vector attached to each point in space. On the other hand, we can think of it as a mapping from real space into order parameter space. That is, $\vec{M}$ is a function which takes different points in the magnet onto the surface of a sphere (figure 9.4). As we mentioned earlier, mathematicians call the sphere $\mathbb{S}^2$, because it locally has two dimensions. (They don't care what dimension the sphere is embedded in.)

Choosing an order parameter is an art. Usually it's a new phase which we don't understand yet, and guessing the order parameter is a piece of figuring out what's going on. Also, there is often more than one sensible choice. In magnets, for example, one can treat $\vec{M}$ as a fixed–length vector in $\mathbb{S}^2$, labeling the different broken symmetry states. This *topological order parameter* is the best choice at low temperatures, where we study the elementary excitations and topological defects. For studying the transition from low to high temperatures, when the magnetization goes to zero, it is better to consider $\vec{M}$ as a 'soft–spin' vector of varying length (a vector in $\mathbb{R}^3$), as in exercise 9.5. Finding the simplest description for your needs is often the key to the problem.

Before varying our order parameter in space, let's develop a few more examples. The liquid crystal in LCD displays (like those in digital watches) are nematics. Nematics are made of long, thin molecules which tend to line up so that their long axes are parallel. Nematic liquid crystals, like magnets, break the rotational symmetry. Unlike magnets, though, the main interaction isn't to line up the north poles, but to line up the axes. (Think of the molecules as American footballs: the same up and down.) Thus the order parameter isn't a vector $\vec{M}$ but a headless vector $\vec{n} \equiv -\vec{n}$. The order parameter space is a hemisphere, with opposing points along the equator identified (figure 9.6). This space is called $\mathbb{R}P^2$ by the mathematicians (the projective plane), for obscure reasons.

For a crystal, the important degrees of freedom are associated with the broken translational order. Consider a two-dimensional crystal which has lowest energy when in a square lattice, but which is deformed away from that configuration (figure 9.7). This deformation is described by an arrow connecting the undeformed ideal lattice points with the actual positions of the atoms. If we are a bit more careful, we say that $\vec{u}(\mathbf{x})$ is that displacement needed to align the ideal lattice in the local region onto the real one. By saying it this way, $\vec{u}$ is also defined between the lattice positions: there still is a best displacement which locally lines up the two lattices.

The order parameter $\vec{u}$ isn't really a vector: there is a subtlety. In general, which ideal atom you associate with a given real one is ambiguous. As shown in figure 9.7, the displacement vector $\vec{u}$ changes by a multiple of the lattice constant $a$ when we choose a different reference atom:

$$\vec{u} \equiv \vec{u} + a\hat{x} = \vec{u} + ma\hat{x} + na\hat{y}. \tag{9.1}$$

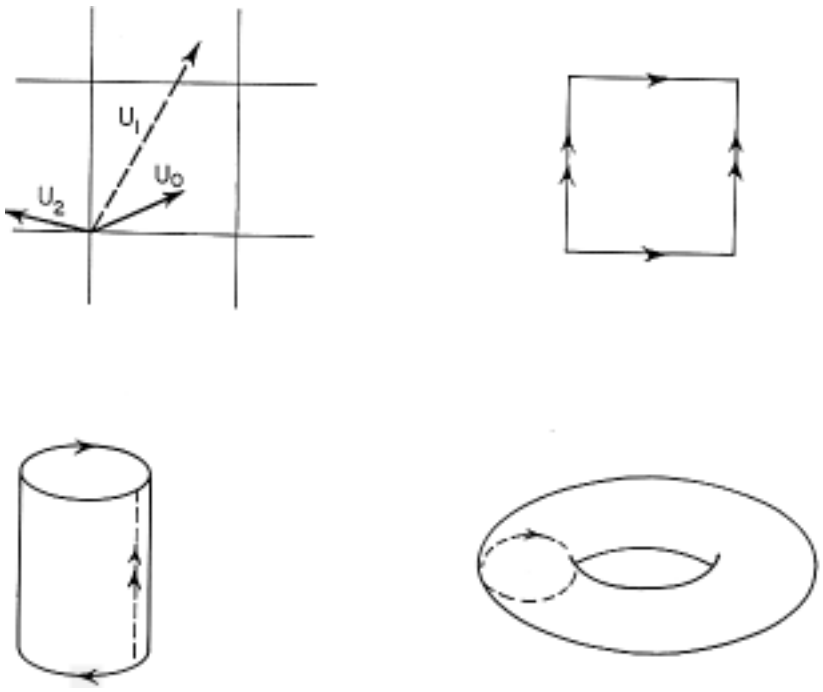



**Fig. 9.8 Order parameter space for a two-dimensional crystal.** Here we see that a square with periodic boundary conditions is a torus. (A torus is a surface of a doughnut, inner tube, or bagel, depending on your background.)

The set of distinct order parameters forms a square with periodic boundary conditions. As figure 9.8 shows, a square with periodic boundary conditions has the same topology as a torus, $\mathbb{T}^2$. (The torus is the surface of a doughnut, bagel, or inner tube.)

Finally, let's mention that guessing the order parameter (or the broken symmetry) isn't always so straightforward. For example, it took many years before anyone figured out that the order parameter for superconductors and superfluid Helium 4 is a complex number $\psi$. The order parameter field $\psi(\mathbf{x})$ represents the "condensate wave function", which (extremely loosely) is a single quantum state occupied by a large fraction of the Cooper pairs or helium atoms in the material. The corresponding broken symmetry is closely related to the number of particles. In "symmetric", normal liquid helium, the local number of atoms is conserved: in superfluid helium, the local number of atoms becomes indeterminate (exercise 9.7)! (This is because many of the atoms are condensed into that delocalized wave function.) Anyhow, the magnitude of the complex number $\psi$ is a fixed function of temperature, so the topological order parameter space is the set of complex numbers of magnitude $|\psi|$. Thus the order parameter space for superconductors and superfluids is a circle $\mathbb{S}^1$.

Now we examine small deformations away from a uniform order parameter field.

**Fig. 9.9** The order parameter field for a one–dimensional crystal is the local displacement $u(x)$. Long–wavelength waves in $u(x)$ have low frequencies, and cause sound.
Crystals are rigid because of the broken translational symmetry. Because they are rigid, they fight displacements. Because there is an underlying translational symmetry, a uniform displacement costs no energy. A nearly uniform displacement, thus, will cost little energy, and thus will have a low frequency. These low–frequency elementary excitations are the sound waves in crystals.

[5]We argue here that low frequency excitations come from spontaneously broken symmetries. They can also come from conserved quantities: since air cannot be created or destroyed, a long–wavelength density wave cannot relax quickly.

[6]Terms with high derivatives become small when you look on long length and time scales. If the displacement $u$ varies on a characteristic length scale $D$, for example, the $n^{th}$ derivative $\partial^n u/\partial x^n \sim 1/D^n$. (Test this: take the 400th derivative of $u(x) = \cos(2\pi x/D)$.) Higher powers of the displacement gradients $\frac{\partial u}{\partial x}^n \sim 1/D^n$ are also small.

## 9.3 Examine the Elementary Excitations

Its amazing how slow human beings are. The atoms inside your eyelash collide with one another a million million times during each time you blink your eye. It's not surprising, then, that we spend most of our time in condensed–matter physics studying those things in materials that happen slowly. Typically only vast conspiracies of immense numbers of atoms can produce the slow behavior that humans can perceive.

A good example is given by sound waves. We won't talk about sound waves in air: air doesn't have any broken symmetries, so it doesn't belong in this chapter.[5] Consider instead sound in the one-dimensional crystal shown in figure 9.9. We describe the material with an order parameter field $u(x)$, where here $x$ is the position within the material and $x-u(x)$ is the position of the reference atom within the ideal crystal.

Now, there must be an energy cost for deforming the ideal crystal. There won't be any cost, though, for a uniform translation: $u(x) \equiv u_0$ has the same energy as the ideal crystal. (Shoving all the atoms to the right doesn't cost any energy.) So, the energy will depend only on derivatives of the function $u(x)$. The simplest energy that one can write looks like

$$\mathcal{E} = \int dx \, \tfrac{1}{2}\kappa \left(\frac{du}{dx}\right)^2. \tag{9.2}$$

Higher derivatives won't be important for the low frequencies that humans can hear.[6] Now, you may remember Newton's law $F = ma$. The force here is given by the derivative of the energy $F = -(d\mathcal{E}/du)$. The mass is represented by the density of the material $\rho$. Working out the math (a variational derivative and an integration by parts, for those who are interested) gives us the equation

$$\rho\ddot{u} = \kappa(d^2u/dx^2). \tag{9.3}$$

The solutions to this equation

$$u(x,t) = u_0 \cos(2\pi(x/\lambda - \nu_\lambda t)) \tag{9.4}$$

represent phonons or sound waves. The wavelength of the sound waves is $\lambda$, and the frequency is $\nu_\lambda$. Plugging 9.4 into 9.3 gives us the relation

$$\nu_\lambda = \sqrt{\kappa/\rho}/\lambda. \tag{9.5}$$

The frequency gets small only when the wavelength gets large. This is the vast conspiracy: only huge sloshings of many atoms can happen slowly. *Why does the frequency get small?* Well, there is no cost to a uniform translation, which is what 9.4 looks like for infinite wavelength. *Why is there no energy cost for a uniform displacement?* Well, there is a translational symmetry: moving all the atoms the same amount doesn't change their interactions. *But haven't we broken that symmetry?* That is precisely the point.

Long after phonons were understood, Jeremy Goldstone started to think about broken symmetries and order parameters in the abstract. He

found a rather general argument that, whenever a continuous symmetry (rotations, translations, $SU(3)$, ...) is broken, long–wavelength modulations in the symmetry direction should have low frequencies. The fact that the lowest energy state has a broken symmetry means that the system is stiff: modulating the order parameter will cost an energy rather like that in equation 9.2. In crystals, the broken translational order introduces a rigidity to shear deformations, and low frequency phonons (figure 9.9). In magnets, the broken rotational symmetry leads to a magnetic stiffness and spin waves (figure 9.10). In nematic liquid crystals, the broken rotational symmetry introduces an orientational elastic stiffness (they pour, but resist bending!) and rotational waves (figure 9.11).

**Fig. 9.10 Magnets: spin waves.** Magnets break the rotational invariance of space. Because they resist twisting the magnetization locally, but don't resist a uniform twist, they have low energy spin wave excitations.

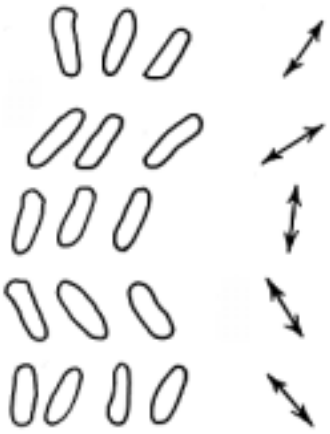


**Fig. 9.11 Nematic liquid crystals: rotational waves.** Nematic liquid crystals also have low–frequency rotational waves.

In superfluids, the broken gauge symmetry leads to a stiffness which results in the superfluidity. Superfluidity and superconductivity really aren't any more amazing than the rigidity of solids. Isn't it amazing that chairs are rigid? Push on a few atoms on one side, and $10^9$ atoms away atoms will move in lock–step. In the same way, decreasing the flow in a superfluid must involve a cooperative change in a macroscopic number of atoms, and thus never happens spontaneously any more than two parts of the chair ever drift apart.

The low–frequency Goldstone modes in superfluids are heat waves! (Don't be jealous: liquid helium has rather cold heat waves.) This is often called second sound, but is really a periodic modulation of the temperature which passes through the material like sound does through a metal.

O.K., now we're getting the idea. Just to round things out, what about superconductors? They've got a broken gauge symmetry, and have a stiffness to decays in the superconducting current. What is the low energy excitation? It doesn't have one. But what about Goldstone's theorem? Well, you know about physicists and theorems ...

That's actually quite unfair: Goldstone surely had conditions on his theorem which excluded superconductors. Actually, I believe Goldstone was studying superconductors when he came up with his theorem. It's just that everybody forgot the extra conditions, and just remembered that you always got a low frequency mode when you broke a continuous symmetry. We condensed-matter physicists already knew why there isn't a Goldstone mode for superconductors: my Ph.D. advisor (P.W. Anderson) had shown that it was related to the Meissner effect. The high energy physicists forgot, though, and had to rediscover it for themselves. Now we all call the loophole in Goldstone's theorem the Higgs mechanism, because (to be truthful) Higgs and his high–energy friends found a simpler and more elegant explanation than we condensed-matter physicists had.

I'd like to end this section, though, by bringing up another exception to Goldstone's theorem: one we've known about even longer, but which we don't have a nice explanation for. What about the orientational order in crystals? Crystals break both the continuous translational order and the continuous orientational order. The phonons are the Goldstone modes for the translations, but *there are no orientational Goldstone*

**Fig. 9.12 Dislocation in a crystal.** Here is a topological defect in a crystal. We can see that one of the rows of atoms on the right disappears halfway through our sample. The place where it disappears is a defect, because it doesn't locally look like a piece of the perfect crystal. It is a topological defect because it can't be fixed by any local rearrangement. No reshuffling of atoms in the middle of the sample can change the fact that five rows enter from the right, and only four leave from the left! The Burger's vector of a dislocation is the net number of extra rows and columns, combined into a vector (columns, rows).

[8]The next fashion, catastrophe theory, never became particularly important.

*modes.*[7] I think understanding this simply and elegantly is one of the most interesting unsolved basic questions in the subject.

## 9.4 Classify the Topological Defects

When I was in graduate school, the big fashion was topological defects. Everybody was studying homotopy groups, and finding exotic systems to write papers about. It was, in the end, a reasonable thing to do.[8] It is true that in a typical application you'll be able to figure out what the defects are without homotopy theory. You'll spend forever drawing pictures to convince anyone else, though. Most important, homotopy theory helps you to think about defects.

A defect is a tear in the order parameter field. A topological defect is a tear that can't be patched. Consider the piece of 2-D crystal shown in figure 9.12. Starting in the middle of the region shown, there is an extra row of atoms. (This is called a dislocation.) Away from the middle, the crystal locally looks fine: it's a little distorted, but there is no problem seeing the square grid and defining an order parameter. Can we rearrange the atoms in a small region around the start of the extra row, and patch the defect?

No. The problem is that we can tell there is an extra row without ever coming near to the center. The traditional way of doing this is to traverse a large loop surrounding the defect, and count the net number of rows crossed on the path. In the path shown, there are two rows going up and three going down: no matter how far we stay from the center, there will naturally always be an extra row on the right.

How can we generalize this basic idea to a general problem with a broken symmetry? Remember that the order parameter space for the 2-D square crystal is a torus (see figure 9.8). Remember that the order parameter at a point is that translation which aligns a perfect square grid to the deformed grid at that point. Now, what is the order parameter far to the left of the defect (a), compared to the value far to the right (d)? Clearly, the lattice to the right is shifted vertically by half a lattice constant: the order parameter has been shifted halfway around the torus. As shown in figure 9.13, along the top half of a clockwise loop the order

[7]In two dimensions, crystals provide another loophole in a well-known result, known as the Mermin-Wagner theorem. Hohenberg, Mermin, and Wagner, in a series of papers, proved in the 1960's that two-dimensional systems with a continuous symmetry cannot have a broken symmetry at finite temperature. At least, that's the English phrase everyone quotes when they discuss the theorem: they actually prove it for several particular systems, including superfluids, superconductors, magnets, and translational order in crystals. Indeed, crystals in two dimensions do not break the translational symmetry: at finite temperatures, the atoms wiggle enough so that the atoms don't sit in lock-step over infinite distances (their translational correlations decay slowly with distance). But the crystals do have a broken orientational symmetry: the crystal axes point in the same directions throughout space. (Mermin discusses this point in his paper on crystals.) The residual translational correlations (the local alignment into rows and columns of atoms) introduce long-range forces which force the crystalline axes to align, breaking the continuous rotational symmetry.

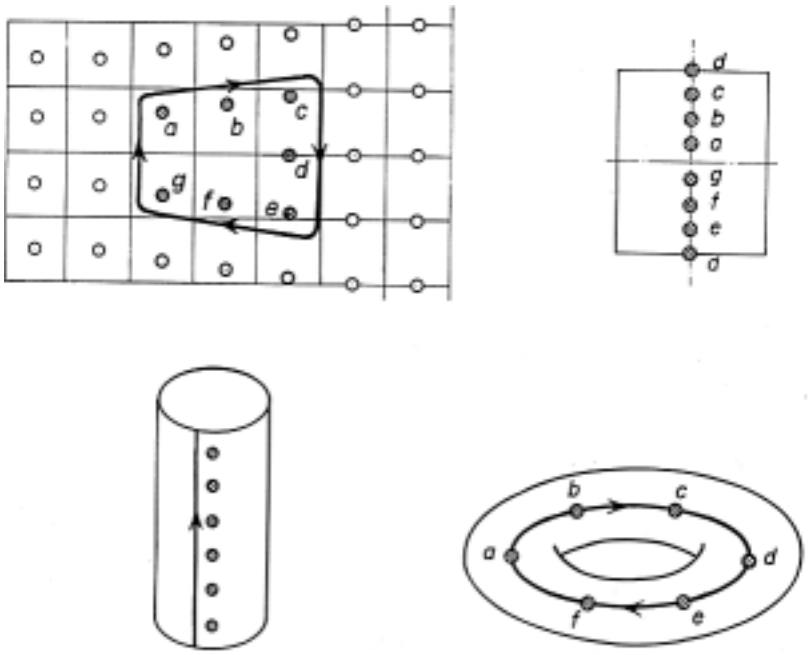



**Fig. 9.13 Loop around the dislocation mapped onto order parameter space.** How do we think about our defect in terms of order parameters and order parameter spaces? Consider a closed loop around the defect. The order parameter field $u$ changes as we move around the loop. The positions of the atoms around the loop with respect to their local "ideal" lattice drifts upward continuously as we traverse the loop. This precisely corresponds to a loop around the order parameter space: the loop passes once through the hole in the torus. A loop *around* the hole corresponds to an extra column of atoms.
Moving the atoms slightly will deform the loop, but won't change the number of times the loop winds through or around the hole. Two loops which traverse the torus the same number of times through and around are equivalent. The equivalence classes are labeled precisely by pairs of integers (just like the Burger's vectors), and the first homotopy group of the torus is $\mathbb{Z} \times \mathbb{Z}$.

parameter (position of the atom within the unit cell) moves upward, and along the bottom half, again moves upward. All in all, the order parameter circles once around the torus. The winding number around the torus is the net number of times the torus is circumnavigated when the defect is orbited once.

This is why they are called topological defects. Topology is the study of curves and surfaces where bending and twisting is ignored. An order parameter field, no matter how contorted, which doesn't wind around the torus can always be smoothly bent and twisted back into a uniform state. If along any loop, though, the order parameter winds either around the hole or through it a net number of times, then enclosed in that loop is a defect which cannot be bent or twisted flat: the winding number can't change by an integer in a smooth and continuous fashion.

How do we categorize the defects for 2-D square crystals? Well, there are two integers: the number of times we go around the central hole, and the number of times we pass through it. In the traditional description, this corresponds precisely to the number of extra rows and columns of atoms we pass by. This was called the Burger's vector in the old days, and nobody needed to learn about tori to understand it. We now call it the first Homotopy group of the torus:

$$\Pi_1(\mathbb{T}^2) = \mathbb{Z} \times \mathbb{Z} \tag{9.6}$$

where $\mathbb{Z}$ represents the integers. That is, a defect is labeled by two

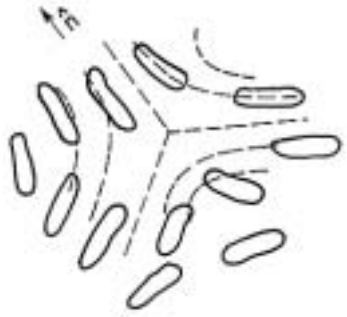


**Fig. 9.14 Defect line in a nematic liquid crystal.** You can't lasso the sphere, but you can lasso a hemisphere! Here is the defect corresponding to the path shown in figure 9.6. As you pass clockwise around the defect line, the order parameter rotates counterclockwise by 180°.
This path on figure 9.6 would actually have wrapped around the right–hand side of the hemisphere. Wrapping around the left–hand side would have produced a defect which rotated clockwise by 180°. (Imagine that!) The path in figure 9.6 is halfway in between, and illustrates that these two defects are really not different topologically.

integers $(m, n)$, where $m$ represents the number of extra rows of atoms on the right-hand part of the loop, and $n$ represents the number of extra columns of atoms on the bottom.

Here's where in the chapter I show the practical importance of topological defects. Unfortunately for you, I can't enclose a soft copper tube for you to play with, the way I do in the lecture. They're a few cents each, and machinists on two continents have been quite happy to cut them up for my demonstrations, but they don't pack well into books. Anyhow, most metals and copper in particular exhibits what is called work hardening. It's easy to bend the tube, but it's amazingly tough to bend it back. The soft original copper is relatively defect–free. To bend, the crystal has to create lots of line dislocations, which move around to produce the bending.[9] The line defects get tangled up, and get in the way of any new defects. So, when you try to bend the tube back, the metal becomes much stiffer. Work hardening has had a noticeable impact on the popular culture. The magician effortlessly bends the metal bar, and the strongman can't straighten it ... Superman bends the rod into a pair of handcuffs for the criminals ...

[9]This again is the mysterious lack of rotational Goldstone modes in crystals.

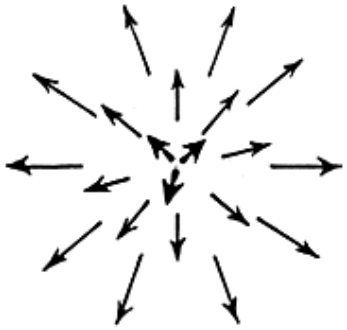


**Fig. 9.15 Hedgehog defect.** Magnets have no line defects (you can't lasso a basketball), but do have point defects. Here is shown the hedgehog defect, $\vec{M}(\mathbf{x}) = M_0\,\hat{x}$. You can't surround a point defect in three dimensions with a loop, but you can enclose it in a sphere. The order parameter space, remember, is also a sphere. The order parameter field takes the enclosing sphere and maps it onto the order parameter space, wrapping it exactly once. The point defects in magnets are categorized by this *wrapping number*: the second Homotopy group of the sphere is $\mathbb{Z}$, the integers.

[10]The zeroth homotopy group classifies domain walls. The third homotopy group, applied to defects in three-dimensional materials, classifies what the condensed matter people call textures and the particle people sometimes call skyrmions. The fourth homotopy group, applied to defects in space–time path integrals, classifies types of instantons.

Before we explain why these curves form a group, let's give some more examples of topological defects and how they can be classified. Figure 9.15 shows a "hedgehog" defect for a magnet. The magnetization simply points straight out from the center in all directions. How can we tell that there is a defect, always staying far away? Since this is a point defect in three dimensions, we have to surround it with a sphere. As we move around on this sphere in ordinary space, the order parameter moves around the order parameter space (which also happens to be a sphere, of radius $|\vec{M}|$). In fact, the order parameter space is covered exactly once as we surround the defect. This is called the *wrapping number*, and doesn't change as we wiggle the magnetization in smooth ways. The point defects of magnets are classified by the wrapping number:

$$\Pi_2(\mathbb{S}^2) = \mathbb{Z}. \tag{9.7}$$

Here, the 2 subscript says that we're studying the second Homotopy group. It represents the fact that we are surrounding the defect with a 2-D spherical surface, rather than the 1-D curve we used in the crystal.[10]

You might get the impression that a strength 7 defect is really just seven strength 1 defects, stuffed together. You'd be quite right: occasionally, they do bunch up, but usually big ones decompose into small ones. This doesn't mean, though, that adding two defects always gives a bigger one. In nematic liquid crystals, two line defects are as good as none! Magnets didn't have any line defects: a loop in real space never surrounds something it can't smooth out. Formally, the first homotopy group of the sphere is zero: you can't loop a basketball. For a nematic liquid crystal, though, the order parameter space was a hemisphere (figure 9.6). There is a loop on the hemisphere in figure 9.6 that you can't get rid of by twisting and stretching. It doesn't look like a loop, but you have to remember that the two opposing points on the equator really represent the same nematic orientation. The corresponding defect has

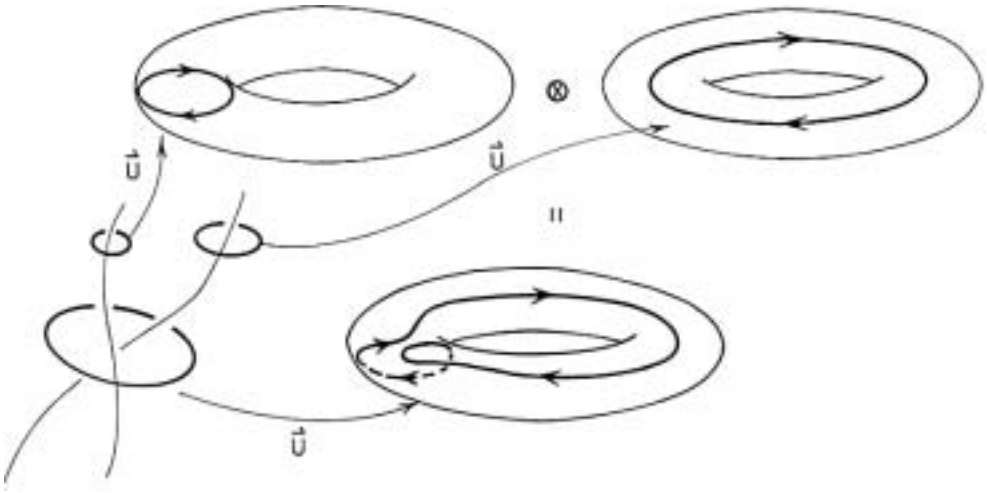



**Fig. 9.16 Multiplying two loops.** The product of two loops is given by starting from their intersection, traversing the first loop, and then traversing the second. The inverse of a loop is clearly the same loop traveled backward: compose the two and one can shrink them continuously back to nothing. This definition makes the homotopy classes into a group.
This multiplication law has a physical interpretation. If two defect lines coalesce, their homotopy class must of course be given by the loop enclosing both. This large loop can be deformed into two little loops, so the homotopy class of the coalesced line defect is the product of the homotopy classes of the individual defects.

a director field $n$ which rotates 180° as the defect is orbited: figure 9.14 shows one typical configuration (called an $s = -1/2$ defect). Now, if you put two of these defects together, they cancel. (You can draw the pictures in exercise 9.2.) Nematic line defects add modulo 2, like clock arithmetic in elementary school:

$$\Pi_1(\mathbb{R}P^2) = \mathbb{Z}_2. \tag{9.8}$$

Two parallel defects can coalesce and heal, even though each one individually is stable: each goes halfway around the sphere, and the whole loop can be shrunk to zero.

Finally, why are these defect categories a group? A group is a set with a multiplication law, not necessarily commutative, and an inverse for each element. For the first homotopy group, the elements of the group are equivalence classes of loops: two loops are equivalent if one can be stretched and twisted onto the other, staying on the manifold at all times.[11] For example, any loop going through the hole from the top (as in the top right-hand torus in figure 9.16) is equivalent to any other one. To multiply a loop $u$ and a loop $v$, one must first make sure that they meet at some point (by dragging them together, probably). Then one defines a new loop $u \otimes v$ by traversing first the loop $u$ and then $v$.[12]

The inverse of a loop $u$ is just the loop which runs along the same path in the reverse direction. The identity element consists of the equivalence class of loops which don't enclose a hole: they can all be contracted

[11] A loop is a continuous mapping from the circle into the order parameter space: $\theta \to u(\theta)$, $0 \le \theta < 2\pi$. When we encircle the defect with a loop, we get a loop in order parameter space as shown in figure 9.4: $\theta \to \vec{x}(\theta)$ is the loop in real space, and $\theta \to u(\vec{x}(\theta))$ is the loop in order parameter space. Two loops are equivalent if there is a continuous one-parameter family of loops connecting one to the other: $u \equiv v$ if there exists $u_t(\theta)$ continuous both in $\theta$ and in $0 \le t \le 1$, with $u_0 \equiv u$ and $u_1 \equiv v$.

[12] That is, $u \otimes v(\theta) \equiv u(2\theta)$ for $0 \le \theta \le \pi$, and $\equiv v(2\theta)$ for $\pi \le \theta \le 2\pi$.


**Fig. 9.17 Defect entanglement.** Can a defect line of class $\alpha$ pass by a line of class $\beta$, without getting topologically entangled?


**Fig. 9.18** We see that we can pass by if we leave a trail: is the connecting double line topologically trivial? Encircle the double line by a loop. The loop can be wiggled and twisted off the double line, but it still circles around the two legs of the defects $\alpha$ and $\beta$.

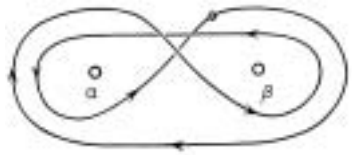



**Fig. 9.19** The homotopy class of the loop is precisely $\beta\alpha\beta^{-1}\alpha^{-1}$, which is trivial precisely when $\beta\alpha = \alpha\beta$. Thus two defect lines can pass by one another if their homotopy classes commute!

smoothly to a point (and thus to one another). Finally, the multiplication law has a direct physical implication: encircling two defect lines of strength $u$ and $v$ is completely equivalent to encircling one defect of strength $u \otimes v$.

This all seems pretty trivial: maybe thinking about order parameter spaces and loops helps one think more clearly, but are there any real uses for talking about the group structure? Let me conclude this chapter with an amazing, physically interesting consequence of the multiplication laws we described.

Can two defect lines cross one another? Figure 9.17 shows two defect lines, of strength (homotopy type) $\alpha$ and $\beta$, which are not parallel. Suppose there is an external force pulling the $\alpha$ defect past the $\beta$ one. Figure 9.18 shows the two line defects as we bend and stretch one to pass by the other. There is a trail left behind of two parallel defect lines. $\alpha$ can really leave $\beta$ behind only if it is topologically possible to erase the trail. Can the two lines annihilate one another? Only if their net strength is zero, as measured by the loop in 9.18.

Now, get two wires and some string. Bend the wires into the shape found in figure 9.18. Tie the string into a fairly large loop, surrounding the doubled portion. Wiggle the string around, and try to get the string out from around the doubled section. You'll find that you can't completely remove the string, (No fair pulling the string past the cut ends of the defect lines!) but that you can slide it downward into the configuration shown in 9.19.

Now, in 9.19 we see that each wire is encircled once clockwise and once counterclockwise. Don't they cancel? Not necessarily! If you look carefully, the order of traversal is such that the net homotopy class is $\beta\alpha\beta^{-1}\alpha^{-1}$, which is only the identity if $\beta$ and $\alpha$ *commute*. Thus the physical entanglement problem for defects is directly connected to the group structure of the loops: commutative defects can pass through one another, noncommutative defects entangle.

I'd like to be able to tell you that the work hardening in copper is due to topological entanglements of defects. It wouldn't be true. The homotopy group of dislocation lines in fcc copper is commutative. (It's rather like the 2-D square lattice: if $\alpha = (m, n)$ and $\beta = (o, p)$ with $m, n, o, p$ the number of extra horizontal and vertical lines of atoms, then $\alpha\beta = (m + o, n + p) = \beta\alpha$.) The reason dislocation lines in copper don't pass through one another is energetic, not topological. The two dislocation lines interact strongly with one another, and energetically get stuck when they try to cross. Remember at the beginning of the chapter, I said that there were gaps in the system: the topological theory can only say when things are impossible to do, not when they are difficult to do.

I'd like to be able to tell you that this beautiful connection between the commutativity of the group and the entanglement of defect lines is nonetheless is important in lots of other contexts. That too would not be true. There are two types of materials I know of which are supposed to suffer from defect lines which topological entangle. The first are biaxial

nematics, which were thoroughly analyzed theoretically before anyone found one. The other are the metallic glasses, where David Nelson has a theory of defect lines needed to relieve the frustration. Nelson's defects don't commute, and so can't cross one another. He originally hoped to explain the freezing of the metallic glasses into random configurations as an entanglement of defect lines. Nobody has ever been able to take this idea and turn it into a real calculation, though.

## Exercises

(9.1) **Topological Defects in the XY Model.** (Math)

Let the order parameter field $\mathbf{m}(x, y)$ of a two–dimensional XY model be a unit vector field in the plane (order parameter space $\mathbb{S}^1$). The topological defects are characterized by their *winding number* $s$, since $\Pi_1(\mathbb{S}^1) = \mathbb{Z}$. The winding number is the number of counter–clockwise orbits around the order parameter space for a single counter–clockwise path around the defect.

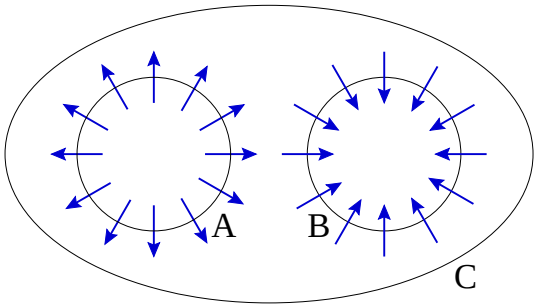



**Fig. 9.20** Two topological defects, circumscribed by loops $A$ and $B$ running counter–clockwise; the pair of defects is circumscribed by a path $C$ also going counter–clockwise. The path in order parameter space mapped to from $C$, as a homotopy group element, is the group product of the paths from $A$ and $B$.

*(a) What are the winding numbers of the two defects surrounded by paths $A$ and $B$ in figure 9.20? What should the winding number be around path $C$, according to the group multiplication law for $\Pi_1(\mathbb{S}^1)$?*

*(b) Copy the figure onto a separate sheet of paper, and fill in the region around $A$ and $B$ past $C$ with a smooth, nonsingular, non-vanishing order parameter field of unit vectors.* (Hint: You can use your answer for (b) to check your answer for (a).)

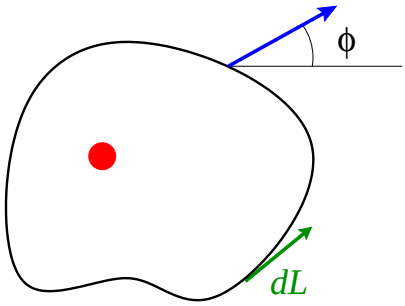



**Fig. 9.21** The curve $D$ encircles the defect once; $d\ell$ is a unit vector tangent to $D$ running counter–clockwise. Define $\phi$ to be the angle between the unit vector $\mathbf{m}$ and the $x$-axis.

For this model, we can find a formula for the winding number as an integral around the defect. Let $D$ encircle a defect counter–clockwise once (figure 9.21).

*(c) Show that the winding number is given by the line integral around the curve $D$*

$$s = \frac{1}{2\pi} \oint \sum_{j=1}^{2} (m_1 \partial_j m_2 - m_2 \partial_j m_1) d\ell_j \qquad (9.9)$$

*where the two coordinates are $x_1$ and $x_2$, $\partial_j = \frac{\partial}{\partial x_j}$, and $\ell_j$ is the tangent vector to the contour being integrated around (so the integral is of the form $\oint \mathbf{v} \cdot d\ell$).* (Hint: $\mathbf{m} = (\cos(\phi), \sin(\phi))$, and the integral of a directional derivative $\nabla f \cdot d\ell$ is the difference in $f$ between the two endpoints.)

One can often find formulas of this kind (e.g., the wrapping number of a vector order parameter around the

sphere $\Pi_2(\mathbb{S}^2)$ is given by an integral of the Jacobian of the mapping).[13]

(9.2) **Topological Defects in Nematic Liquid Crystals.** (Soft Condensed Matter)

The winding number $S$ of a defect is $\theta_{net}/2\pi$, where $\theta_{net}$ is the net angle of rotation that the order parameter makes as you circle the defect. The winding number is positive if the order parameter rotates in the same direction as the traversal (left 9.22), and negative if you traverse in opposite directions (right).

As you can deduce topologically (middle figure 9.22), the winding number is *not* a topological invariant in general. It is for superfluids $\mathbb{S}^1$ and crystals $\mathbb{T}^D$, but not for Heisenberg magnets or nematic liquid crystals (shown). If we treat the plane of the figure as the equator of the hemisphere, you can see that the $S = 1/2$ defect rotates around the sphere around the left half of the equator, and the $S = -1/2$ defect rotates around the right half of the equator. These two paths can be smoothly deformed into one another: the path shown on the order parameter space is about half-way between the two.

*Which part of figure 9.23 represents the defect configuration in real space halfway between $S = 1/2$ and $S = -1/2$, corresponding to the intermediate path shown in the middle above?* (The changing shapes denote rotations into the third dimension.)

(9.3) **Defect Energetics and Total Divergence Terms.** (Soft Condensed Matter)

**Fig. 9.24** Hedgehog defect.

A hypothetical liquid crystal is described by a unit-vector order parameter $\hat{\mathbf{n}}$, representing the orientation of the long axis of the molecules. (Think of it as a nematic liquid crystal where the heads of the molecules all line up as well.)[14] The free energy density is normally written

$$\mathcal{F}_{bulk}[\hat{\mathbf{n}}] = \frac{K_{11}}{2}(\operatorname{div}\hat{\mathbf{n}})^2 + \frac{K_{22}}{2}(\hat{\mathbf{n}}\cdot\operatorname{curl}\hat{\mathbf{n}})^2 + \frac{K_{33}}{2}(\hat{\mathbf{n}}\times\operatorname{curl}\hat{\mathbf{n}})^2. \tag{9.10}$$

(9.11)

Assume a spherical droplet of radius $R_0$ contains a hedgehog defect (shown above) in its center, with order parameter field $\hat{\mathbf{n}}(\mathbf{r}) = \hat{\mathbf{r}} = \mathbf{r}/|r| = (x, y, z)/\sqrt{x^2 + y^2 + z^2}$. The hedgehog is a topological defect, which wraps around the sphere once.

*(a) Show that* $\operatorname{curl}\hat{\mathbf{n}} = 0$ *for the hedgehog. Calculate the free energy of the hedgehog, by integrating $\mathcal{F}[\hat{\mathbf{n}}]$ over the sphere. Compare the free energy to the energy in the same volume with $\hat{\mathbf{n}}$ constant (say, in the $\hat{x}$ direction).*

There are other terms allowed by symmetry that are usually not considered as part of the free energy density, because they are total divergence terms. Any term in the free energy which is a divergence of a vector field, by Gauss' theorem, is equivalent to the flux of the vector field out of the boundary of the material. For periodic boundary conditions such terms vanish, but our system has a boundary.

*(b) Consider the effects of an additional term $\mathcal{F}_{div}[\hat{n}] = K_0(\operatorname{div}\hat{\mathbf{n}})$, allowed by symmetry, in the free energy $\mathcal{F}[\hat{\mathbf{n}}]$. Calculate its contribution to the energy of the hedgehog, both by integrating it over the volume of the sphere and by using Gauss' theorem to calculate it as a surface integral. Compare the total energy $\int \mathcal{F}_{bulk} + \mathcal{F}_{div}\, d^3r$ with that of the uniform state with $\hat{\mathbf{n}} = \hat{x}$, and with the anti-hedgehog, $\hat{\mathbf{n}}(\mathbf{r}) = -\hat{\mathbf{r}}$. Which is lowest, for large $R_0$? How does the ground state for large $R_0$ depend on the sign of $K_0$?*

Clearly, the term $K_0$ from part (b) is not negligible! Liquid crystals in many cases appear to have *strong pinning* boundary conditions, where the relative angle of the order parameter and the surface is fixed by the chemical treatment of the surface. Some terms like $K_0$ are not included in the bulk energy because they become too *large*: they rigidly constrain the boundary conditions and become otherwise irrelevant.

(9.4) **Superfluid Order and Vortices.** (Quantum)

Superfluidity in helium is closely related to Bose condensation of an ideal gas: the strong interactions between

[13] They can be useful, for example, in path integrals for changing the weights of different topological sectors, or in Landau theories to add counter-terms in the free energy that change the core energies of defect lines.

[14] The order parameter is the same as the Heisenberg antiferromagnet, but the latter has a symmetry where the order parameter can rotate independently from the spatial rotations, which isn't true of liquid crystals.

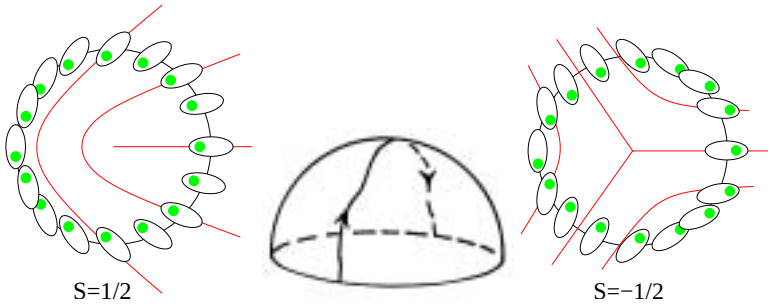



**Fig. 9.22 Defects in Nematic Liquid Crystals.** The molecules in nematic liquids have long-range order in the orientation of their long axes, but the direction of the heads of the molecules do not order. The dots on each molecule are not physical, they are are a guide to help you trace the orientations: we paint them starting on the left and moving clockwise around the defect. Note in the pictures in 9.23 that the changing shapes represent the rotations of the long axes of the molecules out of the plane of the figure.

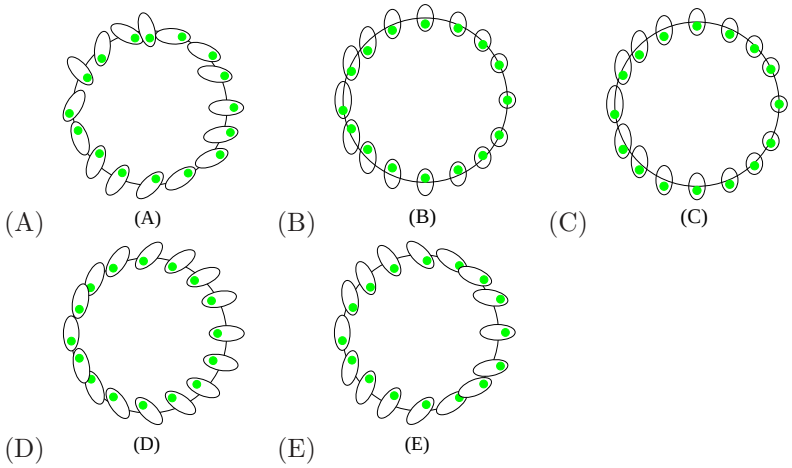



**Fig. 9.23**

the helium atoms quantitatively change things, but many properties are shared. In particular, we describe the superfluid in terms of a complex number $\psi(\mathbf{r})$, which we think of as a wavefunction which is occupied by a large fraction of all the atoms in the fluid.

*(a) If $N$ non-interacting bosons reside in a state $\chi(\mathbf{r})$, write an expression for the net current density $J(\mathbf{r})$.*[15] *Write the complex field $\chi(\mathbf{r})$ in terms of an amplitude and a phase, $\chi(\mathbf{r}) = |\chi(\mathbf{r})| \exp(i\phi(\mathbf{r}))$. We write the superfluid density as $n_s = N|\chi|^2$. Give the current $J$ in terms of $\phi$ and $n_s$. What is the resulting superfluid velocity, $v = J/n_s$? (It should be independent of $n_s$.)*

The Landau order parameter in superfluids $\psi(\mathbf{r})$ is traditionally normalized so that the amplitude is the square root of the superfluid density: in part (a), $\psi(\mathbf{r}) = \sqrt{N}\chi(\mathbf{r})$.

In equilibrium statistical mechanics, the macroscopically occupied state is always the ground state, which is real and hence has no current. We can form non-equilibrium states, however, which macroscopically occupy other quantum states. For example, an experimentalist might cool a torus filled with helium while it's moving: the ground state in the moving reference frame obviously has a current in the unmoving laboratory frame. More commonly, the helium is prepared in a rotating state.

*(b) Consider a torus filled with an ideal Bose gas at $T = 0$ with the hole along the vertical axis: the superfluid is condensed into a state which is rotating around the hole. Using your formula from part (a) and the fact that $\phi + 2n\pi$ is indistinguishable from $\phi$ for any integer $n$, show that the circulation $\oint v \cdot dr$ around the hole is quantized. What is the quantum of circulation?*

Superfluid helium can't swirl except in quantized units! Notice that you've now explained why superfluids have no viscosity. The velocity around the torus is quantized, and hence it cannot decay continuously to zero: if it starts swirling with non–zero $n$ around the torus, it must swirl forever.[16] This is why we call them superfluids.

In bulk helium this winding number labels line defects called *vortex lines*.

*(c) Treat $\phi(\mathbf{r})$, the phase of the superconducting wave function, as the topological order parameter of the superfluid. Is the order parameter a closed loop (topologically a circle, $\mathbb{S}^1$)? Classify the types of vortex lines in a superfluid. (That is, either give the first Homotopy group of the order parameter space, or give the allowed values of the quantized circulation around a vortex.)*

### (9.5) Landau Theory for the Ising model.

This chapter has focused on the topological order parameter, which labels the different ground states of the system when there is a spontaneously broken symmetry. To study the defect cores, interfaces, and high temperatures near phase transitions, one would like an order parameter which can vary in magnitude as well as direction.

In section 6.7, we explicitly computed a free energy for the ideal gas as a function of the density. Can we use symmetry and gradient expansions to derive free energy densities for more realistic systems – even systems that we don't understand microscopically? Lev Landau used the approach we discuss here to develop theories of magnets, superconductors, and superfluids – before the latter two were understood in microscopic terms.[17] In this exercise, you will develop a Landau[18] theory for the Ising model.

Here we outline the general recipe, and ask you to implement the details for the Ising model. Along the way, we will point out places where the assumptions made in Landau theory can break down – often precisely in the cases where the theory is most useful.

#### (1) Pick an Order Parameter Field.

Remember that the Ising model had a high-temperature paramagnetic phase with zero magnetization per spin $m$, and a low-temperature ferromagnetic phase with a net magnetization per spin $\pm m(T)$ that went to one at $T = 0$.

[15] You can use the standard quantum mechanics single-particle expression $J = (i\hbar/2m)(\psi\nabla\psi^* - \psi^*\nabla\psi)$ and multiply by the number of particles, or you can use the many particle formula $J(\mathbf{r}) = (i\hbar/2m) \int d^3\mathbf{r_1} \ldots d^3\mathbf{r_N} \sum_\ell \delta(\mathbf{r}_\ell - \mathbf{r})(\Psi\nabla_\ell\Psi^* - \Psi^*\nabla_\ell\Psi)$ and plug in the condensate wave function $\Psi(\mathbf{r_1} \ldots \mathbf{r_N}) = \prod_n \chi(\mathbf{r_n})$.

[16] Or at least until a dramatic event occurs which changes $n$, like a vortex line passing across the torus, demanding an activation energy proportional to the width of the torus. See also exercise 7.7.

[17] Physicists call this Landau theory. Rather similar formalisms have been developed in various other fields of physics and engineering, from liquid crystals to "rational mechanics" treatments of martensites (see exercises 11.6 and 11.5). The vocabulary is often different (Frank, Ericsen and Leslie instead of Landau, constitutive relations rather than free energies, and internal state variables rather than order parameters) but the basic approach is similar.

[18] More properly, a Ginsburg-Landau theory, because we include gradient terms in the free energy density, which Landau first did in collaboration with Ginsburg.

The Ising model picks one of the two equilibrium states (up or down): we say it spontaneously breaks the up-down symmetry of the Hamiltonian[19]

Hence the natural[20] order parameter is the scalar $m(\mathbf{x}, t)$, the local magnetization averaged over some volume $\Delta V$. This can be done by averaging the magnetization in small boxes, as in section 6.7.

*(a) If there are $n$ spins in the volume $\Delta V$, what value will $m(\mathbf{x})$ take at temperatures high compared to the interaction $J$ in the Ising model? What values will it take at temperatures very low compared to $J$?*

(2) **Write a General Local[21] Free Energy Density, for long wavelengths and translational symmetry.**

A local free energy is one which depends on the order parameter field and its gradients:

$$\mathcal{F}^{\text{ising}}\{m, T\} = \mathcal{F}(\mathbf{x}, m, \partial_j m, \partial_j \partial_k m, \dots) \tag{9.12}$$

As in section 9.3, we Taylor expand in gradients.[22] Keeping terms with up to two gradients of $m$ (and, for simplicity, no gradients of temperature), we find

$$\begin{aligned}\mathcal{F}^{\text{ising}}\{m, T\} = A(m, T) + V_i(m, T)\partial_i m \\ + B_{ij}(m, T)\partial_i \partial_j m + C_{ij}(m, T)(\partial_i m)(\partial_j m).\end{aligned} \tag{9.13}$$

*(b) What symmetry tells us that the unknown functions $A$, $B$, and $C$ do not depend on position $\mathbf{x}$? If the magnetization varies on a length scale $D$, how much smaller would a term involving three derivatives be than the terms $B$ and $C$ that we have kept?*

(3) **Impose the other symmetries of the problem.**

The Ising model has an up-down symmetry[23] so the free energy density $\mathcal{F}^{\text{ising}}\{m\} = \mathcal{F}\{-m\}$. Hence the coefficients $A$ and $C$ are functions of $m^2$, and the functions $V_i(m, T) = m v_i(m^2, T)$ and $B_{ij}(m) = m\, b_{ij}(m)$.

The two-dimensional Ising model on a square lattice is symmetric under 90° rotations. This tells us that $v_i = 0$ because no vector is invariant under 90° rotations. Similarly, $b$ and $C$ must commute with these rotations, and so must be multiples of the identity matrix.[24] Hence we have

$$\begin{aligned}\mathcal{F}^{\text{ising}}\{m, T\} = A(m^2, T) + m\, b(m^2, T)\nabla^2 m \\ + C(m^2, T)(\nabla m)^2.\end{aligned} \tag{9.15}$$

Many systems are isotropic: the free energy density is invariant under all rotations. For isotropic systems, the material properties (like the functions $A$, $B_{ij}$, and $C_{ij}$ in equation 9.13) must be invariant under rotations. All

[19] Real magnets break *time-reversal* symmetry. A magnetic field changes sign under time reversal ($M \to -M$ as $t \to -t$). One can see this by considering a current loop generating a field: under time reversal the current reverses direction.

[20] Landau has a more systematic approach, based on group representation theory, which can be quite useful in more complex systems.

[21] What about things like the Coulomb interaction, which depends on fields emitted from distant regions? Treat the electric fields explicitly! If the system has long-range interactions, one should incorporate the relevant (electric, elastic, gravitational, ...) fields into the free energy as order parameters. Broadly speaking, for a complete description the order parameter should incorporate long-range fields, conserved quantities, and all broken symmetries.

[22] Notice that this Taylor expansion will not be valid if the order parameter varies quickly on the microscopic length scales – such as at sharp interfaces or in the cores of topological defects. Landau theory is often nonetheless used to study such problems: it may not be quantitatively valid, but it provides a solvable if uncontrolled approximation to the real behavior.

[23] The equilibrium state may not, but the model – and hence the free energy density – certainly does.

[24] Let's see this explicitly. Under a 90° rotation $R = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ about the point $\mathbf{x}$, $m\, v_i(m)\partial_i m$ goes to $m\, v_i(m) R_{ij} \partial_j m$. These must be equal for all fields $m$, implying that $(v_1, v_2) = (v_1, v_2)\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} = (-v_2, v_1)$, showing that $v = 0$. Similarly, the matrix $C$ rotates to

$$RCR^{-1} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}\begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix}\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} C_{22} & -C_{12} \\ -C_{21} & C_{11} \end{pmatrix} \tag{9.14}$$

so for $C$ to be invariant it must satisfy $C_{11} = C_{22}$ and $C_{12} = C_{21} = 0$: $C$ is a multiple of the identity. Similar arguments show that any two-index tensor with hexagonal symmetry (dimension $D = 2$), cubic symmetry ($D = 3$), or hypercubic symmetry ($D > 3$) must be a multiple of the identity as well.

terms in a local free energy for an isotropic system must be writable in terms of dot and cross products of the gradients of the order parameter field.

*(c) Would the free energy density of equation 9.15 change for a magnet that had a continuous rotational symmetry?*

(4) **Simplify using total divergence terms.**

Free energy densities are intrinsically somewhat arbitrary. If one adds to $\mathcal{F}$ a gradient of any smooth vector function $\nabla \cdot \boldsymbol{\xi}(m)$, the integral will differ only by a surface term $\int \nabla \cdot \boldsymbol{\xi}(m)\, dV = \int \boldsymbol{\xi}(m) \cdot dS$. In many circumstances, surface terms may be ignored. (i) If the system has periodic boundary conditions, then the integral $\int \boldsymbol{\xi}(m) \cdot dS = 0$ because the opposite sides of the box will cancel. (ii) Large systems will have surface areas which are small compared to their volumes, so often the surface terms can often be ignored, $\int \nabla \cdot \boldsymbol{\xi}(m) dV = \int \boldsymbol{\xi}(m) \cdot dS \sim L^2 \ll \int \mathcal{F} dV \sim L^3$. (iii) Total divergence terms can be interchanged for changes in the surface free energy, which depends upon the orientation of the order parameter with respect to the boundary of the sample.[25] What this means in practice is that we can integrate any of our terms by parts, changing the free energy only by a ignorable surface term. This allows us to integrate terms in the free energy by parts: schematically, by subtracting a total divergence $\nabla(uv)$ from the free energy we can exchange a term $u\,\nabla v$ for a term $-v\,\nabla u$. For example, we can subtract a term $-\nabla \cdot \left(m\, b(m^2, T) \nabla m\right)$ from the free energy 9.15

$$\begin{aligned}
\mathcal{F}^{\text{ising}}\{m, T\} &= A(m^2, T) + m\, b(m^2, T) \nabla^2 m \\
&\quad + C(m^2, T)(\nabla m)^2 - \nabla \left(m\, b(m^2, T) \cdot \nabla m\right) \\
&= A(m^2, T) + C(m^2, T)(\nabla m)^2 - \nabla \left(m\, b(m^2, T)\right) \cdot \nabla m \\
&= A(m^2, T) + C(m^2, T)(\nabla m)^2 \\
&\quad - \left(b(m^2, T) + 2m^2 b'(m^2, T)\right)(\nabla m)^2 \\
&= A(m^2, T) \qquad (9.16) \\
&\quad + \left(C(m^2, T) - b(m^2, T) - 2m^2 b'(m^2, T)\right)(\nabla m)^2,
\end{aligned}$$

replacing $\left(m\, b(m^2, T)\right)(\nabla^2 m)$ with the equivalent term $-(\nabla m)(\nabla(m\, b(m^2, T) \nabla m) \cdot \nabla m)$. Thus we may absorb the $b$ term proportional to $\nabla^2 m$ into an altered $c = C(m^2, T) - b(m^2, T) - 2m^2 b'(m^2, T)$ term times $(\nabla m)^2$:

$$\mathcal{F}^{\text{ising}}\{m, T\} = A(m^2, T) + c(m^2, T)(\nabla m)^2. \qquad (9.17)$$

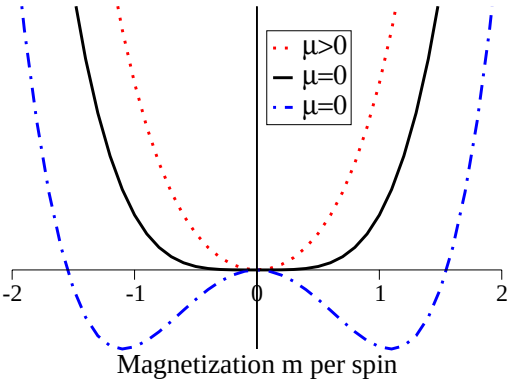



**Fig. 9.25** The Landau free energy density for the Ising model 9.18, at positive, zero, and negative values of the quadratic term $\mu$.

(5) **(Perhaps) Assume the order parameter is small.**[26]

If we assume $m$ is small, we may Taylor expand $A$ and $c$ in powers of $m^2$:[27] $A(m^2, T) = f_0 + \frac{\mu(T)}{2} m^2 + \frac{g}{4!} m^4$ and $c(m^2, T) = \frac{1}{2} K$, leading to the traditional Landau free energy for the Ising model

$$\mathcal{F}^{\text{ising}} = \tfrac{1}{2} K (\nabla m)^2 + f_0 + (\mu(T)/2) m^2 + (g/4!) m^4 \qquad (9.18)$$

where $f_0$, $g$ and $K$ can also depend upon $T$.

The free energy density of equation 9.18 is one of the most extensively studied models in physics. The field theorists use $\phi$ instead of $m$ for the order parameter, and call it the $\phi^4$ model. Ken Wilson added fluctuations to this model in developing the renormalization group (chapter 12).

Notice that the Landau free energy density has a qualitative change at $\mu = 0$. For positive $\mu$ it has a single minimum at $m = 0$; for negative $\mu$ it has two minima at $m = \pm\sqrt{-6\mu/g}$. Is this related to the transition in the Ising model from paramagnetic phase $m = 0$ to ferromagnetic phase at $T_c$?

The free energy density already incorporates (by our assumptions) fluctuations in $m$ on length-scales small compared to the coarse-graining length $W$. *If we ignored*

[25] See exercise 9.3 and reference [60]. One must also be wary of total divergence terms for systems with topological defects, which count as internal surfaces: see reference [95].

[26] Notice that this approximation is not valid for abrupt phase transitions, where the order parameter is large until the transition and zero afterward. Again, Landau theories for abrupt transitions are often qualitatively illuminating anyhow.

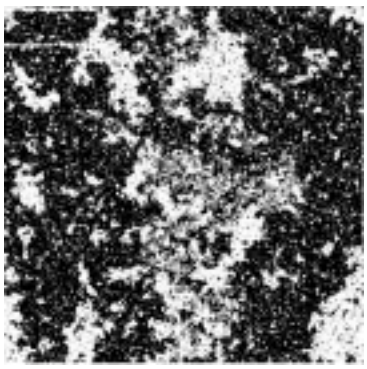
[27] The symbols used varies. The factors of $\frac{1}{2}$ and $1/4!$ are traditional in the field theory community: they make later calculations simpler.

*fluctuations on scales larger than $W$* then the free energy of the whole system[28] would be given by the volume times the free energy density, and the magnetization at a temperature $T$ would be given by minimizing the free energy density. The quadratic term $\mu(T)$ would vanish at $T_c$, and if we expand $\mu(T) \sim a(T - T_c) + \ldots$ we find $m = \pm\sqrt{6a/g}\sqrt{T_c - T}$ near the critical temperature.

This is qualitatively correct, but quantitatively wrong. The magnetization does vanish at $T_c$ with a power law $m \sim (T_c - T)^\beta$, but the exponent $\beta$ is not generally 1/2: in two dimensions it is $\beta_{2d} = 1/8$ and in three dimensions it is $\beta_{3d} = 0.31$. These exponents (particularly the presumably non-rational one in 3D) cannot be fixed by keeping more or different terms in the analytic Landau expansion:

*(d) Show that the power-law $\beta_{\text{Landau}} = 1/2$ is unchanged in the limit $T \to T_c$ by keeping additional terms in the Landau expansion ($\frac{h}{6!}m^6$ in equation 9.18). (That is, show that $m(T)/(T - T_c)^\beta$ goes to a constant as $T \to T_c$.)* (Hint: you should get a quadratic equation for $m^2$. Keep the root that vanishes at $T = T_c$, and expand in powers of $h$.) *Explore also the alternative, non-generic phase transition where $g \equiv 0$ but $h > 0$: what is $\beta$ for that transition?*

As we see in figure 9.26 there is no length $W$ above which the Ising model near $T_c$ looks smooth and uniform. The Landau free energy density gets corrections on all length scales: for the infinite system the free energy has a singularity at $T_c$ (making our power-series expansion for $\mathcal{F}^{\text{ising}}$ inadequate). The Landau free energy density is only be a starting point for studying continuous phase transitions;[29] we must use the renormalization-group methods of chapter 12 to explain and predict these singularities.



[28] The free energy density $\mathcal{F}$ shown in figure 9.25 is qualitatively different from the total free energy $F$ in two other ways. (1) We expand $\mathcal{F}$ as a power series in the order parameters and gradients, but one should not expect generally to be able to expand $F$ in this way. You may have wondered how we knew that we could expand $\mathcal{F}$. How did we know that the free energy density didn't vary as $\sqrt{m}$, or $|m|$? (The free energy of the infinite system below $T_c$, for example, has a term $M(T)|H|$, which has a cusp because $M$ changes sign abruptly at $H = 0$. The free energy of a finite system will not have a cusp, since both signs of the total magnetization occur with finite, smooth dependence on $H$.)

Because we're coarse-graining over a small region of size $W^3$, our free energy density is much the same as that of a finite sample of size $W^3$. One can show that the free energy is an analytic function of $\beta = 1/k_B T$ for a finite system, since it's a rapidly convergent sum of exponentials $e^{-E_n/k_B T}$. Similarly, one expects that the free energy of a finite system will be analytic in other external parameters or constraints. (At least for sensible choices: if one chose $y = m^6$ as the order parameter, the free energy would involve powers of $y^{1/3}$.) *There are no phase transitions in finite systems at non-zero temperature*, since phase transitions are precisely when power series and analyticity breaks down.

(2) The total free energy density has no barrier between the two signs of magnetization! A typical macroscopic system $L \times L \times L$ at low temperatures with zero net magnetizationis composed of two equal-sized domains, one magnetized upward and one downward. Its total free energy is of order a surface tension (the free energy per unit area for an interface) $\sigma$ times the cross sectional area, $\sigma L^2$, and so contributes a free energy per spin of $\sigma/L$ – an amount that vanishes for large systems. The free energy $F$ of the macroscopic system is *convex*: $\lambda F(m_0) + (1-\lambda)F(m_1) \geq F(\lambda m_0 + (1-\lambda)m_1)$ by exactly this argument: the left-hand side is the free energy of a mixture of two subregions of fractional size $\lambda$ and $1 - \lambda$ with the same average magnetization as the right-hand side.

[29] An important exception to this is superconductivity, where the Cooper pairs are large compared to their separation. Because they overlap so many neighbors, the fluctuations in the order parameter field are suppressed, and Landau theory is valid even very close to the phase transition.

**Fig. 9.26** A snapshot of the Ising model at $T_c$. Notice that there are fluctuations on all length scales.

In many cases Landau theory is used even when can't be justified quantitatively starting from the microscopic theory. For properties that are independent of the microscopic theory (universal properties near critical points, hydrodynamic properties at long wavelengths) one can use Landau theory as the basic theoretical starting point for modeling: the free energy 9.18 (commonly written in terms of $\phi$ and termed $\phi^4$ theory) is the starting point of many field theory discussions and the Hamiltonian used by Wilson to develop the renormalization group (chapter 12). For other properties (calculating energies of different phases, or calculating the energies and structures of defects and interfaces) Landau theory provides useful insights even though it can't be trusted quantitatively.

(9.6) **Bloch walls in Magnets.**

The free energy density of an Ising magnet below $T_c$ can be roughly approximated as a double well potential (equation 9.18), with two minima at $\pm m_0$:

$$\mathcal{F} = \tfrac{1}{2}K(\nabla m)^2 + (\mu(T)/2)m^2 + (g/4!)m^4 \tag{9.19}$$
$$= \tfrac{1}{2}K(\nabla m)^2 + (g/4!)(m^2 - m_0^2)^2. \tag{9.20}$$

This exercise studies the structure of the domain wall, or Bloch wall, separating a region of positive magnetization from one of negative magnetization.

Consider a magnetization $m(x)$ varying only along the $x$ direction, with $m(-\infty) = -m_0$ and $m(\infty) = m_0$. In between, it must pass through a barrier region with $m \approx 0$. The stiffness $K$ penalizes sharp gradients in the magnetization; $g$ penalizes regions with magnetization away from the equilibria at $\pm m_0$. In part (a), we give a rough argument for the width of the Bloch wall, in terms of $K$, $m_0$, and $g$, by balancing the gradient cost of a thin wall against the barrier cost of a thick wall.

The second term in $\mathcal{F}$ is a double-wall potential, with a barrier $B$ separating two wells, with units energy per unit volume. An interface between $m = -m_0$ and $m = +m_0$ with width $\Delta$ will have a energy cost $\sim B \times \Delta$ per unit area due to the barrier, which wants $\Delta$ to be as small as possible. The first term in $\mathcal{F}$ is a stiffness of the order parameter against rapid changes in $m$, adding an energy per unit area $\sim K\Delta \times (m_0/\Delta)^2$.

*(a) Using these rough estimates find $B$, minimize the sum, and give a rough value for the energy per unit area of the Bloch wall in terms of $K$, $m_0$, and $g$.*

The rest of this exercise will lead you through a variational calculation of the shape of the Bloch wall (see [71, Chapter 12], for information about the calculus of variations).

*(b) Find the equation satisfied by that $m(x)$ which minimizes $F = \int \mathcal{F}\, dx$, given the boundary conditions.* (This is the Euler-Lagrange equation from the calculus of variations.)

*(c) Show that the solution $m(x)$ has the property that the combination*

$$E = (K/2)(\partial m/\partial x)^2 - (g/4!)(m^2 - m_0^2)^2 \tag{9.21}$$

*is independent of $x$.* (Hint: what is $\partial E/\partial x$?)

$E$ is analogous to the energy of a particle in an inverted potential well, with $x$ playing the role of time and the potential being the negative of the $g$ term in the original free energy density, with the double-well at $\pm m_0$ becoming a potential with two hills. (The conservation law comes from the symmetry of the Bloch wall under translations.) Solving for the minimum $m(x)$ is finding the classical trajectory in the inverted potential: it rolls down one hill and rolls back up the second one.

*(d) Argue from the boundary conditions that $E = 0$. Using that, find the minimum free energy path $m(x)$ satisfying the boundary conditions $m(\pm\infty) = \pm m_0$. Was your wall thickness estimate of part (a) roughly correct?* (Hint: if you know $dy/dx = f(y)$, you know $\int dy/f(y) = \int dx$.)

(9.7) **Superfluids: Density Matrices and ODLRO.** (Quantum) [30]

**Density Matrices.** We saw in the last problem that a Bose-condensed ideal gas can be described in terms of a complex number $\psi(\mathbf{r})$ representing the eigenstate which is macroscopically occupied. For superfluid helium, the atoms are in a strongly interacting liquid state when it goes superfluid. We can define the order parameter $\psi(\mathbf{r})$ even for an interacting system using the reduced density matrix.

Suppose our system is in a mixture of many-body states $\Psi_\alpha$ with probabilities $P_\alpha$. The full density matrix in the position representation, you will remember, is

$$\hat{\rho}(\mathbf{r}_1', \ldots, \mathbf{r}_N', \mathbf{r}_1, \ldots, \mathbf{r}_N) \tag{9.22}$$
$$= \sum_\alpha P_\alpha \Psi^*(\mathbf{r}_1', \ldots, \mathbf{r}_N')\Psi(\mathbf{r}_1, \ldots, \mathbf{r}_N).$$

(Properly speaking, these are the matrix elements of the density matrix in the position representation: rows are labeled by $\{\mathbf{r}_i'\}$, columns are labeled by $\{\mathbf{r}_j\}$.) The reduced

[30] See references [5, 6].

density matrix $\hat{\rho}(\mathbf{r}', \mathbf{r})$ (which I'll call the density matrix thereafter) is given by setting $\mathbf{r}'_j = \mathbf{r}_j$ for all but one of the particles and integrating over all possible positions, multiplying by $N$:

$$\hat{\rho}_2(\mathbf{r}', \mathbf{r}) = N \int d^3 r_2 \dots d^3 r_N \hat{\rho}(\mathbf{r}', \mathbf{r}_2 \dots, \mathbf{r}_N, \mathbf{r}, \mathbf{r}_2, \dots, \mathbf{r}_N). \tag{9.23}$$

(For our purposes, the fact that it's called a matrix is not important: think of $\hat{\rho}_2$ as a function of two variables.)

*(a) What does the reduced density matrix $\rho_2(\mathbf{r}', \mathbf{r})$ look like for a zero–temperature Bose condensate of non-interacting particles, condensed into a normalized single–particle state $\chi(\mathbf{r})$.*

An alternative, elegant formulation for this density matrix is to use second-quantized creation and annihilation operators instead of the many-body wavefunctions. These operators $a^\dagger(\mathbf{r})$ and $a(\mathbf{r})$ add and remove a boson at a specific place in space. They obey the commutation relations

$$\begin{aligned} [a(\mathbf{r}), a^\dagger(\mathbf{r}')] &= \delta(\mathbf{r} - \mathbf{r}') \\ [a(\mathbf{r}), a(\mathbf{r}')] &= [a^\dagger(\mathbf{r}), a^\dagger(\mathbf{r}')] = 0. \end{aligned} \tag{9.24}$$

and the vacuum has no particles, so

$$\begin{aligned} a(\mathbf{r})|0\rangle &= 0 \\ \langle 0|a^\dagger(\mathbf{r}) &= 0. \end{aligned} \tag{9.25}$$

*(b) Let's calculate formula 9.23 for a pure state. First we need to create a wave-function using the creation and annihilation operators as a ket:*

$$|\Psi\rangle = (1/\sqrt{N!}) \int d^3\mathbf{r}_1 \dots d^3\mathbf{r}_N \Psi(\mathbf{r}_1, \dots \mathbf{r}_N) a^\dagger(\mathbf{r}_1) \dots a^\dagger(\mathbf{r}_N)|0\rangle. \tag{9.26}$$

(9.27)

*Show that the ket is normalized if the symmetric Bose wavefunction $\Psi$ is normalized.* (Hint: Using equation 9.24 to pull the $a$'s to the right through the $a^\dagger$'s, you should get a sum of $N!$ terms, each a product of $N$ $\delta$-functions, setting different permutations of $\mathbf{r}_1 \dots \mathbf{r}_N$ equal to $\mathbf{r}'_1 \dots \mathbf{r}'_N$.) *Show that $\langle\Psi|a^\dagger(\mathbf{r}')a(\mathbf{r})|\Psi\rangle$, the overlap of $a(\mathbf{r})|\Psi\rangle$ with $a(\mathbf{r}')|\Psi\rangle$ for the pure state $|\Psi\rangle$ gives the the reduced density matrix 9.23.*

Since this is true of all pure states, it's true of mixtures of pure states as well: hence the reduced density matrix is the same as the expectation value $\langle a^\dagger(\mathbf{r}')a(\mathbf{r})\rangle$.

In a non-degenerate Bose gas, in a system with Maxwell-Boltzmann statistics, or in a Fermi system, one can calculate $\hat{\rho}_2(\mathbf{r}', \mathbf{r})$ and show that it rapidly goes to zero as $|\mathbf{r}' - \mathbf{r}| \to \infty$. This makes sense: in a big system, $a(\mathbf{r})|\Psi(\mathbf{r})\rangle$ leaves a state with a missing particle localized around $\mathbf{r}$, which will have no overlap with $a(\mathbf{r}')|\Psi\rangle$ which has a missing particle at the distant place $\mathbf{r}'$.

**ODLRO and the superfluid order parameter.** This is no longer true in superfluids: just as in the condensed Bose gas of part (a), interacting, finite-temperature superfluids have a reduced density matrix with off-diagonal long-range order (ODLRO):

$$\hat{\rho}_2(\mathbf{r}', \mathbf{r}) \to \psi^*(\mathbf{r}')\psi(\mathbf{r}) \quad \text{as } |\mathbf{r}' - \mathbf{r}| \to \infty. \tag{9.28}$$

It's called long-range order because there are correlations between distant points: it's called off-diagonal because the diagonal of this density matrix in position space is $\mathbf{r} = \mathbf{r}'$. The order parameter for the superfluid is $\psi(\mathbf{r})$, describing the long-range piece of this correlation.

*(c) Show that equation 9.28 determines $\psi(\mathbf{r})$ up to an overall constant multiplicative phase* $\exp(i\phi_0)$. Hint: $\rho(a, r)$ is $\psi(r)$ up to a constant $\psi^*(a)$, so long as $r$ is far from $a$. If $b$ and $c$ are far from $a$, show that $|\psi^*(a)| = \sqrt{\rho(a, c)\rho(b, a)/\rho(b, c)}$. Finally, you'll need to patch the region where $r$ is near $a$. *What is $\psi(\mathbf{r})$ for the non-interacting Bose condensate of part (a), in terms of the condensate wave function $\chi(\mathbf{r})$?*

This correlation function is analogous in many ways to the density-density correlation function we defined for gases $C(\mathbf{r}', \mathbf{r}) = \langle\rho(\mathbf{r}')\rho(\mathbf{r})\rangle$ and the correlation function for magnetization $\langle M(\mathbf{r}')M(\mathbf{r})\rangle$. The fact that $\hat{\rho}_2$ is long range is analogous to the fact that $\langle M(\mathbf{r}')M(\mathbf{r})\rangle \sim \langle M\rangle^2$ as $\mathbf{r}' - \mathbf{r} \to \infty$: the long-range order in the direction of magnetization is the analog of the long-range phase relationship in superfluids.

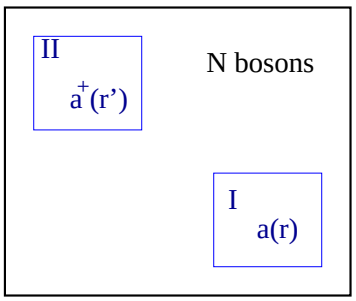



**Fig. 9.27** Particles in superfluids are delocalized: the number of particles in a subvolume is not well defined. Annihilating a boson at $\mathbf{r}$ in region I, insofar as the boson comes out of the condensate, is equivalent to annihilating it at $\mathbf{r}'$. The probability overlap between these two states is precisely $\hat{\rho}_2(\mathbf{r}', r) = \psi^*(\mathbf{r}')\psi(\mathbf{r})$.

**Number conservation and $\psi$.** Figure 9.27 illustrates the fact that the local number of particles in a subvolume

of a superfluid is indeterminate. Our ground state locally violates conservation of particle number.[31] If the number of particles in a local region isn't well defined, perhaps we can think of the local state as some kind of superposition of states with different particle number? Then we could imagine factoring the off-diagonal long-range order $\langle a^\dagger(\mathbf{r}')a(\mathbf{r})\rangle \sim \psi^*(\mathbf{r}')\psi(\mathbf{r})$ into order $\langle a^\dagger(\mathbf{r}')\rangle\langle a(\mathbf{r})\rangle$, with $\psi(\mathbf{r}) = \langle a\rangle$. (This is obviously zero in a closed system, since $a(\mathbf{r})$ changes the total number of particles.) The immediate question is how to set the relative phases of the parts of the wave-function with differing numbers of particles. Let's consider a region small enough that we can ignore the spatial variations.

*(d) Consider a zero-temperature Bose condensate of $N$ non-interacting particles in a local region. Let the state into which the bosons condense, $\chi(\mathbf{r}) = \chi = |\chi|\exp(i\phi)$, be spatially uniform. What is the phase of the $N$-particle bose-condensed state?*

The phase $\exp(i\phi(\mathbf{r}))$ is the relative phase between the components of the local Bose condensates with $N$ and $N-1$ particles. The superfluid state is a *coherent superposition of states with different numbers of particles* in local regions. How odd!

Momentum conservation comes from translational symmetry; energy conservation comes from time translational symmetry; angular momentum conservation comes from rotational symmetry. What symmetry leads to number conservation?

*(e) Consider the Hamiltonian $\mathcal{H}$ for a system that conserves the total number of particles, written in second quantized form (in terms of creation and annihilation operators). Argue that the Hamiltonian is invariant under a global symmetry which multiplies all of the creation operators by $\exp(i\zeta)$ and the annihilation operators by $\exp(-i\zeta)$. (This amounts to changing the phases of the $N$-particle parts of the wave-function by $\exp(iN\zeta)$.)* (Hint: note that all terms in $\mathcal{H}$ have an equal number of creation and annihilation operators.)

The magnitude $|\psi(\mathbf{r})|^2$ describes the superfluid density $n_s$. As we saw above, $n_s$ is the whole density for a zero temperature non-interacting Bose gas; it's about one percent of the density for superfluid helium, and about $10^{-8}$ for superconductors. If we write $\psi(\mathbf{r}) = \sqrt{n_s(\mathbf{r})}\exp(i\phi(\mathbf{r}))$, then the phase $\phi(\mathbf{r})$ labels which of the broken-symmetry ground states we reside in.[32]

**Broken Gauge invariance.** We can draw a deep connection with quantum electromagnetism by promoting this global symmetry into a local symmetry. Consider the effects of shifting by a spatially dependent phase $\zeta(x)$. It won't change the potential energy terms, but will change the kinetic energy terms because they involve gradients. Consider the case of a single-particle pure state. Our wave-function $\chi(x)$ changes into $\tilde{\chi} = \exp(i\zeta(x))\chi(x)$, and $\frac{p^2}{2m}\tilde{\chi} = \frac{\left(\frac{\hbar}{i}\nabla\right)^2}{2m}\tilde{\chi}$ includes terms involving $\nabla\zeta$.

*(f) Show that this single-particle Hamiltonian is invariant under a transformation which changes the phase of the wavefunction by $\exp(i\zeta(x))$ and simultaneously replaces $p$ with $p - \hbar\nabla\zeta$.*

This invariance under multiplication by a phase is closely related to gauge invariance in electromagnetism. Remember in classical electromagnetism the vector potential $\mathbf{A}$ is arbitrary up to adding a gradient of an arbitrary function $\Lambda$: changing $\mathbf{A} \to \mathbf{A} + \nabla\Lambda$ leaves the magnetic field unchanged, and hence doesn't change anything physical. There choosing a particular $\Lambda$ is called choosing a *gauge*, and this arbitrariness is called *gauge invariance*. Also remember how we incorporate electromagnetism into the Hamiltonian for charged particles: we change the kinetic energy for each particle of charge $q$ to $(p - \frac{q}{c}A)^2/2m$, using the "covariant derivative" $\frac{\hbar}{i}\nabla - \frac{q}{c}A$.

In quantum electrodynamics, particle number is not conserved, but charge is conserved. Our local symmetry, stemming from number conservation, is analogous to the symmetry of electrodynamics when we multiply the wave function by $\exp(i(q/e)\zeta(x))$, where $-e$ is the charge on an electron.

*(g) Consider the Hamiltonian for a charged particle in a vector potential $H = (\frac{\hbar}{i}\nabla - \frac{q}{c}A)^2/2m + V(x)$. Show that this Hamiltonian is preserved under a transformation which multiplies the wave-function by $\exp(i(q/e)\zeta(x))$ and performs a suitable gauge transformation on $A$. What is the required gauge transformation?*

To summarize, we found that superconductivity leads to a state with a local indeterminacy in the number of particles. We saw that it is natural to describe local regions of superfluids as coherent superpositions of states with different numbers of particles. The order parameter $\psi(\mathbf{r}) = \langle a(\mathbf{r})\rangle$ has amplitude given by the square root of the superfluid density, and a phase $\exp(i\phi(\mathbf{r}))$ giving the relative quantum phase between states with

[31] This isn't just the difference between canonical and grand canonical ensembles. Grand canonical ensembles are probability mixtures between states of different numbers of particles: superfluids have a coherent superposition of wave-functions with different numbers of particles.

[32] $\psi(\mathbf{r})$ is the Landau order parameter; the phase $\phi(\mathbf{r})$ is the topological order parameter.

different numbers of particles. We saw that the Hamiltonian is symmetric under uniform changes of $\phi$: the superfluid ground state breaks this symmetry just as a magnet might break rotational symmetry. Finally, we saw that promoting this global symmetry to a local one demanded changes in the Hamiltonian completely analogous to gauge transformations in electromagnetism: number conservation comes from a gauge symmetry. Superfluids spontaneously break gauge symmetry!

In the references to this exercise, you can find more along these lines. In particular, number $N$ and phase $\phi$ turn out to be conjugate variables. The implied equation $i\hbar\dot{N} = [\mathcal{H}, N] = i\partial\mathcal{H}/\partial\phi$ gives the Josephson current, and is also related to the the equation for the superfluid velocity we derived in exercise 9.4, and so on...

# Correlations, Response, and Dissipation

10

**Fig. 10.1** The 2-D square–lattice Ising model, quenched (abruptly cooled) from high temperatures to zero temperature [103]. The model quickly separates into local blobs of up and down spins, which grow and merge, coarsening to larger blob sizes (section 11.3.1).

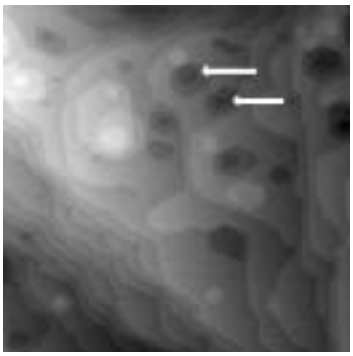


**Fig. 10.2** An STM image of a surface, created by bombarding a close–packed gold surface with noble–gas atoms, and then allowing the irregular surface to thermally anneal (Tatjana Curcic and Barbara H. Cooper) [25]. The figure shows individual atomic–height steps; the arrows show a single step pit inside another pit. The characteristic sizes of the pits and islands grow as the surface evolves and flattens.

This chapter focuses on the space and time dependence of statistical mechanical systems, and the relations between correlations, response, and dissipation. We introduce and derive several famous, powerful, and central results in this (otherwise somewhat dry) discussion. More information can be found in references [20, chapter 8] (classical), [32] (quantum), and [70].

We start by studying the spatial and temporal cohesiveness of a system, using *correlation functions* (section 10.1). In section 10.2 we note that diffraction experiments measure correlation functions in a system.

We then specialize to studying correlations in equilibrium, where many subtle relations can be derived. We'll focus in the text on the special case of the ideal gas; in some extended footnotes we'll give the formalism for more general order parameter fields. In section 10.3 we illustrate how to use equilibrium statistical mechanics to calculate correlations at equal times. Since there are no equal–time correlations in a classical ideal gas, we find *white noise*. In section 10.4 we use Onsager's *Regression Hypothesis* to derive the time–dependence of the correlations in our ideal gas.

We then examine how systems respond, in space and time, to a small external perturbation. This linear response is described in general by a *susceptibility*, which we introduce in section 10.5. We note there that the imaginary part of the frequency–dependent susceptibility is related to dissipation, and the real part to elastic response. We show that the static susceptibility is proportional to the equal–time correlation function. We derive the *Fluctuation–Dissipation Theorem*, giving the dynamic susceptibility in terms of the time–dependent correlation function. Finally, in section 10.6 we use causality – the fact that the response cannot precede the perturbation – to derive the *Kramers–Krönig* relation connecting the imaginary (dissipative) part of the AC susceptibility to the real (responsive) part.

## 10.1 Correlation Functions: Motivation

We've learned how to derive the laws giving the equilibrium states of a system (figure 10.3) and the evolution laws of systems as they approach equilibrium (figures 10.1 and 10.2). How, though, do we characterize the

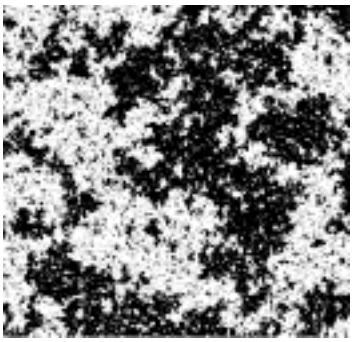


**Fig. 10.3** The 2-D square–lattice Ising model near the critical temperature $T_c$. Here the 'islands' come in all sizes, and the equilibrium fluctuations happen on all time scales: see chapter 12.

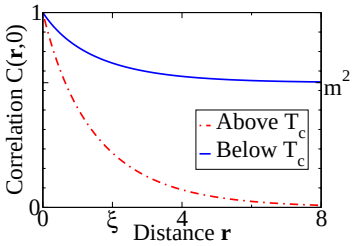



**Fig. 10.4** A schematic equal–time correlation function $C(r, \tau = 0)$ for the 2D Ising model at a temperature just above and just below the critical temperature $T_c$. At $r = 0$ the correlation function is $\langle S^2 \rangle = 1$. The distance $\xi(T)$ after which the correlation function decays exponentially to its long–distance value (zero or $m^2$) is the *correlation length*. At $T_c$ the correlation length diverges, leading to fluctuations on all scales.

resulting behavior? How do we extract from our ensemble of systems some testable numbers or functions that measure the patterns in space and time, that we can use to compare experiment and theory?

In figures 10.1 and 10.2 we see two systems that can be roughly characterized as blobs which gradually merge and grow with time. The first is the Ising model at low temperature, showing the spin $S(\mathbf{x}, t)$ at position $\mathbf{x}$ and time $t$: the up–spin and down–spin regions are competing to determine which will take over as the broken–symmetry ground state. The second is a gold surface that is thermally flattening from an irregular initial shape, showing the height $h(\mathbf{x}, t)$; here the blobs are pits and islands. One should of course directly study (and publish) snapshots like these, and make available animations of the evolution [25, 103]. Such visual images incorporate a full, rich description of individual members of the ensemble of models – but it's hard to quantify whether experiments and theory agree by comparing snapshots of a random environment. How can we characterize the ensemble as a whole, pulling out functions and numbers to characterize, say, the typical blob size as a function of time?

Figure 10.3 shows the Ising model at $T_c$, where fluctuations occur on all length and time scales. The visual pictures are cool, but how do we pull out functions that describe how likely a black region will extend a distance $\mathbf{r}$, or survive for a time $\tau$?

We typically characterize such systems using *correlation functions*. Consider the alignment of two Ising spins $S(\mathbf{x}, t)$ and $S(\mathbf{x} + \mathbf{r}, t)$ in the coarsening figure 10.1: spins at the same time $t$, but separated by a distance $\mathbf{r}$. If $|\mathbf{r}|$ is much larger than a typical blob size $L(t)$, the spins will have a 50/50 chance of being aligned or misaligned, so their average product will be near zero. If $|\mathbf{r}|$ is much smaller than a typical blob size $L$, the spins will typically be aligned parallel to one another (both $+1$ or both $-1$), so their average product will be near one. The equal–time spin–spin correlation function

$$C_{coar}(\mathbf{r}) = \langle S(\mathbf{x}, t) S(\mathbf{x} + \mathbf{r}, t) \rangle \tag{10.1}$$

will go from one at $\mathbf{x} = \mathbf{0}$ to zero at $\mathbf{x} \gg L(t)$, and will cross $1/2$ at a characteristic blob size $L(t)$. In non–equilibrium problems like this one, the system is evolving in time, so the equal–time correlation function also evolves. We'll discuss coarsening in more detail in section 11.3.1 and exercises 11.3 and 12.1.

The correlation function in general contains more information than just the typical blob size. For the equilibrium Ising model, it also contains information about how much a spin influences its distant neighbors. Even at high temperatures, if a spin is up its immediate neighbors are more likely to point up than down. As the temperature approaches the ferromagnetic transition temperature $T_c$, this preference extends to further neighbors, roughly extending to a distance $\xi(T)$ (figure 10.4). Below $T_c$ we have *long–range order*: even very distant neighbors will tend to align with our spin, since the two broken–symmetry equilibrium states each have net magnetization per spin $m$. Above $T_c$ the correlation function goes to zero at long distances $r$; below $T_c$ it goes to $m^2$, since

the two spins become uncorrelated

$$\langle S(\mathbf{x},t)S(\mathbf{x}+\mathbf{r},t)\rangle \sim \langle S(\mathbf{x},t)\rangle\langle S(\mathbf{x}+\mathbf{r},t)\rangle = m^2. \tag{10.2}$$

What happens at the critical temperature? As we approach $T_c$ the correlation length $\xi(T)$ diverges. At $T_c$ the equal–time correlation function $C(r,0)$ is a power law at long distances (figure 10.5), representing the fact that there are correlations at all length scales. Similarly, the (equal–position) spin–spin correlation function in time $C(\mathbf{0},\tau) = \langle s(t)s(t+\tau)\rangle$ gets small rapidly after times $\tau$ larger than the typical lifetime of a correlated region; at the critical point this lifetime diverges (giving *critical slowing down*, see exercise 8.5), and the spin–spin correlation function also decays as a power law, $C(\mathbf{0},\tau) \sim \tau^{-(d-2+\eta)/z}$ at $T_c$. We'll see how to explain these power laws in chapter 12.

In other systems, one might study the atomic density–density correlation functions $C(\mathbf{r},\tau) = \langle \rho(\mathbf{x}+\mathbf{r},t+\tau)\rho(\mathbf{x},t)\rangle$,[1] or the height–height correlation function for a surface (figure 10.2), or the phase–phase correlations of the superfluid order parameter (exercise 9.7), ...

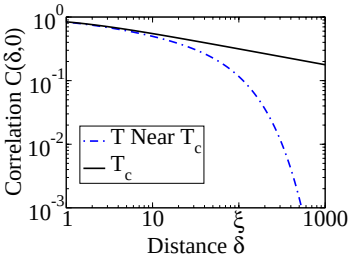



**Fig. 10.5** The schematic Ising equal–time correlation function on a log–log plot, at $T_c$ (straight line, representing the power law $C \sim r^{-(d-2+\eta)}$, chapter 12), and just above $T_c$ (where the dependence shifts to $C \sim e^{-r/\xi(T)}$ at distances beyond the correlation length $\xi(T)$).

[1] Here $\rho(\mathbf{x},t) = \sum_j \boldsymbol{\delta}(\mathbf{x}-\mathbf{x}_j)$ just measures the positions of the centers for each atom; no coarse–graining is implied.

## 10.2 Experimental Probes of Correlations

Many scattering experiments directly measure correlation functions. X-rays measure the electron density–density correlation function, neutrons can measure spin–spin correlation functions, and so on. Elastic scattering gives the equal–time correlation functions, while inelastic scattering can give the time–dependent correlation functions.

Let's briefly summarize how this works for X-ray elastic scattering. In X-ray diffraction experiments[2] a plane–wave beam of wave–vector $\mathbf{k}_0$ scatters off of the sample, with the emitted radiation along wave–vector $\mathbf{k}_0+\mathbf{k}$ proportional to $\widetilde{\rho}_e(\mathbf{k})$, the Fourier transform of the electron density $\rho_e(\mathbf{x})$ in the sample. The intensity of the scattered beam $|\widetilde{\rho}_e(\mathbf{k})|^2$ can be measured, for example, by exposing photographic film with different geometries and experimental conditions. But this square is given by the Fourier transform of the equal–time electron[3] density–density correlation function $C_{ee}(\mathbf{r}) = \langle \rho_e(\mathbf{x}+\mathbf{r},t)\rho_e(\mathbf{x},t)\rangle$, as seen by a general and useful

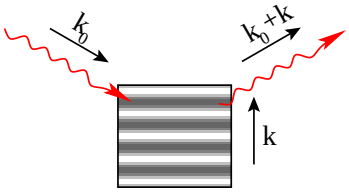



**Fig. 10.6 X-ray Scattering.** A beam of wave-vector $\mathbf{k}_0$ scatters off a density variation $\rho(\mathbf{x})$ with wave-vector $\mathbf{k}$ to a final wave-vector $\mathbf{k}_0+\mathbf{k}$; the intensity of the scattered beam is proportional to $|\widetilde{\rho}(\mathbf{k})|^2$: see [8, chapter 6].

[2] Medical X-rays and CAT scans measure the penetration of X-rays, not their diffraction.

[3] Since the electrons are mostly tied to atomic nuclei, $C_{ee}(\mathbf{r})$ is writable in terms of the corresponding atom–atom correlation function, called *pair distribution function* $C(\mathbf{r})$. This is done using *form factors*: see [8, chapter 6].

calculation (equation A.20):

$$\begin{aligned}
|\widetilde{\rho}(\mathbf{k})|^2 = \widetilde{\rho}(\mathbf{k})^* \widetilde{\rho}(\mathbf{k}) &= \int d\mathbf{x}'\, e^{-i\mathbf{k}\cdot\mathbf{x}'} \rho(\mathbf{x}') \int d\mathbf{x}\, e^{i\mathbf{k}\cdot\mathbf{x}} \rho(\mathbf{x}) \\
&= \int d\mathbf{x} d\mathbf{x}'\, e^{i\mathbf{k}\cdot(\mathbf{x}-\mathbf{x}')} \rho(\mathbf{x}')\rho(\mathbf{x}) = \int d\mathbf{r}\, e^{i\mathbf{k}\cdot\mathbf{r}} \int d\mathbf{x}'\, \rho(\mathbf{x}')\rho(\mathbf{x}'+\mathbf{r}) \\
&= \int d\mathbf{r}\, e^{i\mathbf{k}\cdot\mathbf{r}} V \langle \rho(\mathbf{x})\rho(\mathbf{x}+\mathbf{r})\rangle = V \int d\mathbf{r}\, e^{i\mathbf{k}\cdot\mathbf{r}} C(\mathbf{r}) \\
&= V\widetilde{C}(\mathbf{k})
\end{aligned} \tag{10.3}$$

In the same way, other scattering experiments also measure two–point correlation functions, averaging over the entire illuminated sample.

Real space microscopy experiments and $k$-space diffraction experiments provide complementary information about a system. The real–space images are direct and easily appreciated and comprehended by the human mind. They are invaluable for studying unusual events (which would be swamped in a bulk average), distributions of local geometries (individual ensemble elements rather than averages over the ensemble), and physical dynamical mechanisms. The $k$-space methods, on the other hand, by averaging over the entire sample, can provide great precision, and have close ties with calculational methods and analytical methods (as presented in this chapter). Indeed, often one will computationally Fourier transform measured real–space data in order to generate correlation functions.

## 10.3 Equal–Time Correlations in the Ideal Gas

For the rest of this chapter we will consider systems which are in equilibrium and close to equilibrium, where we can relate the correlations, response, and dissipation functions. Let's start by calculating the equal–time correlation function. We focus on the ideal gas, which is both the simplest and the most difficult case. In the exercises, you can calculate correlation functions that are algebraically more challenging (exercise 10.4) but the ideal gas case is both conceptually subtle and fundamental. In particular, the ideal gas is the special case called *white noise* where there are no correlations between different points in space.

The Helmholtz free energy density of the ideal gas is

$$\mathcal{F}^{\text{ideal}}(\rho(\mathbf{x}), T) = \rho(\mathbf{x}) k_B T \left[\log(\rho(\mathbf{x})\lambda^3) - 1\right] \tag{10.4}$$

(equation 6.61). The probability $P\{\rho(\mathbf{x})\}$ of finding a particular density profile $\rho(\mathbf{x})$ as a fluctuation is proportional to

$$P\{\rho(\mathbf{x})\} \propto e^{-\beta \int \mathcal{F}(\rho(\mathbf{x}))d\mathbf{x}}. \tag{10.5}$$

Let's assume the fluctuations are small, and linearize about $\langle\rho\rangle = \rho_0$:

$$\begin{aligned}\mathcal{F}(\rho(\mathbf{x})) &= \mathcal{F}_0 + \left.\frac{\delta\mathcal{F}}{\delta\rho}\right|_{\rho_0}(\rho-\rho_0) + \tfrac{1}{2}\left.\frac{\delta^2\mathcal{F}}{\delta\rho^2}\right|_{\rho_0}(\rho-\rho_0)^2 \\ &= \mathcal{F}_0 + \mu_0(\rho-\rho_0) + \tfrac{1}{2}\alpha(\rho-\rho_0)^2 \end{aligned} \tag{10.6}$$

where $\mu_0 = \left.\frac{\delta\mathcal{F}}{\delta\rho}\right|_{\rho_0}$ is the chemical potential and the coefficient of the quadratic term is

$$\alpha = \left.\frac{\partial^2\mathcal{F}}{\partial\rho^2}\right|_{\rho_0} = k_BT/\rho_0 = P_0/\rho_0^2 \tag{10.7}$$

(since the pressure $P_0 = Nk_BT/V = \rho k_BT$). Only the integral of the free energy matters, so

$$\int \mathcal{F}(\rho(\mathbf{x}))\,d\mathbf{x} = V\mathcal{F}_0 + \mu_0\cancel{\int(\rho-\rho_0)\,d\mathbf{x}} + \int \tfrac{1}{2}\alpha(\rho-\rho_0)^2\,d\mathbf{x} \tag{10.8}$$

where $\mu_0$ drops out because the average of $\rho$ equals $\rho_0$. We can also drop $\mathcal{F}_0$ because it changes the free energy of all configurations by a constant, and doesn't change their relative probabilities.[4] So the effective free energy of the ideal gas, for small density fluctuations, is

$$\mathcal{F}(\rho) = \tfrac{1}{2}\alpha(\rho-\rho_0)^2 \tag{10.9}$$

and the probability of finding a density fluctuation is

$$P\{\rho(\mathbf{x})\} \propto e^{-\beta\int \frac{1}{2}\alpha(\rho-\rho_0)^2\,d\mathbf{x}}. \tag{10.10}$$

We can now calculate the expectation value of the equal–time density–density correlation function

$$C^{\mathbf{Ideal}}(\mathbf{r},0) = \langle\rho(\mathbf{x},t)\rho(\mathbf{x}+\mathbf{r},t)\rangle - \rho_0^2 = \langle(\rho(\mathbf{x},t)-\rho_0)(\rho(\mathbf{x}+\mathbf{r},t)-\rho_0)\rangle \tag{10.11}$$

where we subtract off the square of the average density, so that we measure the correlations between the fluctuations of the order parameter about its mean value. (Subtracting the means gives us the *connected* correlation function.) If we break up the ideal gas into tiny boxes of size $\Delta V$, the probability of having density $\rho(\mathbf{x}_j)$ in volume $j$ is

$$P_j(\rho) \propto e^{-\frac{\frac{1}{2}\alpha(\rho-\rho_0)^2\Delta V}{k_BT}} = e^{-\frac{(\rho-\rho_0)^2}{2/(\beta\alpha\,\Delta V)}}. \tag{10.12}$$

This is a Gaussian of root-mean-square width $\sigma = \sqrt{1/(\beta\alpha\,\Delta V)}$, so the mean square fluctuations inside a single box is $\langle(\rho-\rho_0)^2\rangle = \frac{1}{\beta\alpha\,\Delta V}$. The density fluctuations in different boxes are uncorrelated. This means $C^{\mathbf{Ideal}}(\mathbf{r},0) = 0$ for $\mathbf{r}$ reaching between two boxes, and $C^{\mathbf{Ideal}}(\mathbf{0},0) = \frac{1}{\beta\alpha\,\Delta V}$ within one box.

What does it mean for $C^{\mathbf{Ideal}}$ to depend on the box size $\Delta V$? The fluctuations become stronger as the box gets smaller. We're of course familiar with this: in section 3.2.1 and equation 3.67 we saw using the

[4] Each Boltzmann factor shifts by $e^{-\beta V\mathcal{F}_0}$, so $Z$ shifts by the same factor, so the ratio giving the probability is independent of $\mathcal{F}_0$.

microcanonical ensemble that the square of the number fluctuations in a small subvolume of ideal gas was equal to the expected number of particles $\langle (N - \langle N \rangle)^2 \rangle \sim N$.[5]

How do we write the correlation function, though, in the limit $\Delta V \to 0$? It must be infinite at $\mathbf{r} = \mathbf{0}$, and zero for all non-zero $\mathbf{r}$. As usual, it is proportional the Dirac $\delta$-function, this time in three dimensions $\boldsymbol{\delta}(\mathbf{r}) = \delta(r_x)\delta(r_y)\delta(r_z)$. The equal–time connected correlation function for the ideal gas is

$$C^{\mathbf{Ideal}}(\mathbf{r}, 0) = \frac{1}{\beta\alpha}\boldsymbol{\delta}(\mathbf{r}). \tag{10.14}$$

*The correlation function of an uncorrelated system is white noise.*[6]

We can see that the constant outside is indeed $1/\beta\alpha$, by using equation 10.14 to check that the mean square fluctuations of the integral of $\rho$ inside the box of volume $\Delta V$ is indeed $\frac{1}{\beta\alpha\,\Delta V}$:

$$\begin{aligned}
\langle(\rho - \rho_0)^2\rangle_{\text{box}} &= \left\langle \left( \frac{1}{\Delta V} \int_{\Delta V} (\rho(\mathbf{x}) - \rho_0)\, d\mathbf{x} \right)^2 \right\rangle \\
&= \frac{1}{(\Delta V)^2} \int_{\Delta V} d\mathbf{x} \int_{\Delta V} d\mathbf{x}' \langle (\rho(\mathbf{x}) - \rho_0)(\rho(\mathbf{x}') - \rho_0) \rangle \\
&= \frac{1}{(\Delta V)^2} \int_{\Delta V} d\mathbf{x} \int_{\Delta V} d\mathbf{x}' \frac{1}{\beta\alpha} \boldsymbol{\delta}(\mathbf{x} - \mathbf{x}') \\
&= \frac{1}{\beta\alpha(\Delta V)^2} \int_{\Delta V} d\mathbf{x} = \frac{1}{\beta\alpha\,\Delta V}.
\end{aligned} \tag{10.15}$$

[6] White light is a mixture of all frequencies of light with equal amplitude and random phases (exercise A.7). Our noise has the same property. The Fourier transform of $C(\mathbf{r}, 0)$, $\widetilde{C}(\mathbf{k}, t = 0) = \frac{1}{V}|\widetilde{\rho}(k)|^2$ (as in equation 10.3), is constant, independent of the wavevector $k$. Hence all modes have equal weight. To show that the phases are random, we can express the free energy (equation 10.8) in Fourier space, where it is a sum over uncoupled harmonic modes: hence in equilibrium they have random relative phases (see exercise 10.4).

## 10.4 Onsager's Regression Hypothesis and Time Correlations

Equilibrium statistical mechanics does not determine the dynamics: air has sound waves and perfume in still air diffuses, but both are good ideal gases. In studying the time evolution of correlations, we could work with the microscopic laws: indeed, most treatments of this topic start from quantum mechanics [32]. Instead, we will rely on the macroscopic evolution laws to specify our dynamics.

How are the density fluctuations in an ideal gas of perfume correlated in time? In particular, suppose at $t = 0$ there is a rare fluctuation, increasing the density of perfume at one point (figure 10.7). How will it decay to a more typical density profile as time passes?

Macroscopically our perfume obeys the diffusion equation of chapter 2. There we derived the evolution laws for imposed initial non–equilibrium

[5] So, using equation 10.7 and the ideal gas law $P_0 V = N k_B T = N/\beta$, the density fluctuations

$$\begin{aligned}
\langle(\rho - \rho_0)^2\rangle &= \frac{(N - \langle N \rangle)^2}{(\Delta V)^2} = \frac{N}{(\Delta V)^2} = \frac{\rho_0}{\Delta V} \\
&= \frac{1}{(\rho_0/P_0)(P_0/\rho_0^2)\Delta V} = \frac{1}{\frac{N}{P_0 V}\alpha \Delta V} = \frac{1}{\beta\alpha\Delta V}
\end{aligned} \tag{10.13}$$

just as we computed from $\mathcal{F}^{\text{ideal}}$.

density profiles, and ignored the spontaneous thermal fluctuations. A macro–scale initial condition (figure 10.8) will evolve according to the diffusion equation $\frac{\partial \rho}{\partial t} = D\nabla^2\rho$. The confusing point about the microscopic density (figure 10.7) is that it introduces new spontaneous thermal fluctuations while it flattens old ones.

In this text, we have been rather casual in denoting averages, using the same symbol $\langle\rangle$ for averages over time and space, and averages over microcanonical, canonical, and grand canonical ensembles. In this section and in some of the exercises we'll be doing several different kinds of averages, and we need to distinguish between them. Our microcanonical, canonical, and grand canonical ensemble averages cannot be used to calculate quantities depending on more than one time (because the equilibrium ensembles are independent of dynamics). Let's write $\langle\rangle_{\rm eq}$ for these equal–time equilibrium averages. The time–time correlation functions are defined by an equilibrium ensemble of time evolutions, which may include noise from the environment.[7] Let us denote these averages by $\langle\rangle_{\rm ev}$. Thus $\langle\rangle_{\rm ev}$ generalizes $\langle\rangle_{\rm eq}$ to operators at non-equal times. Finally, let's write $[A]_{A_{\rm i}}$ for the noisy evolution average of $A(\mathbf{x},t)$, fixing the initial condition $A(\mathbf{x},0) = A_{\rm i}(\mathbf{x})$ at time zero. This averages over all the new spontaneous density fluctuations, allowing us to examine the decay of an initial spontaneous density fluctuation, or perhaps an initial imposed density profile.

We will assume that this last average fixing the initial condition obeys the same diffusion equation that governs the macroscopic time evolution. For our diffusion of perfume, this means

$$\frac{\partial}{\partial t}\left[\rho(\mathbf{x},t)\right]_{\rho_{\rm i}} = D\nabla^2\left[\rho(\mathbf{x},t)\right]_{\rho_{\rm i}}. \tag{10.16}$$

Generalizing this formula to arbitrary systems and putting it into words, leads us to Onsager's *Regression Hypothesis* [81],

> ... we may assume that a spontaneous deviation from the equilibrium decays according to the same laws as one that has been produced artificially.

It is clear now that we may calculate the density–density correlation function $\langle\rho(\mathbf{r}+\mathbf{r}',t+t')\rho(\mathbf{r}',t')\rangle_{\rm ev}$ by taking our evolution ensemble for fixed initial condition $[\rho(\mathbf{x},t)]_{\rho_{\rm i}}$ and then taking a thermal average over initial conditions $\rho(\mathbf{x},0) = \rho_{\rm i}(\mathbf{x})$. We may use the fact that our system is homogeneous and time–independent to measure our correlation function starting at the origin:

$$\begin{aligned} C(\mathbf{r},\tau) &= \langle(\rho(\mathbf{x}+\mathbf{r},t+\tau)-\rho_0)(\rho(\mathbf{x},t)-\rho_0)\rangle_{\rm ev} \\ &= \langle(\rho(\mathbf{r},\tau)-\rho_0)(\rho(\mathbf{0},0)-\rho_0)\rangle_{\rm ev} \\ &= \Big\langle([\rho(\mathbf{r},\tau)]_{\rho_{\rm i}}-\rho_0)(\rho_{\rm i}(\mathbf{0})-\rho_0)\Big\rangle_{\rm eq}. \end{aligned} \tag{10.17}$$

In words, averaging over both initial conditions and noise $\langle\rangle_{\rm ev}$ is the same as first averaging over noise $[]_{\rm i}$ and then over initial conditions

[7] Our perfume molecules will be buffeted by the surrounding air molecules, which we are not explicitly integrating over. Bath–induced noise like this is often modeled using *Langevin equations* (exercise 10.5).

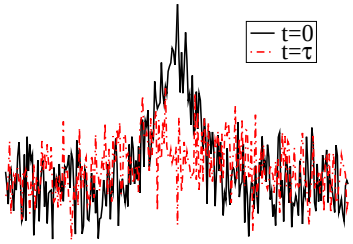



**Fig. 10.7** An unusual fluctuation at $t = 0$ will slowly decay to a more typical thermal configuration at a later time $\tau$.

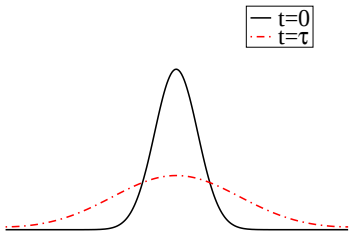



**Fig. 10.8** An initial condition with the same density will slowly decay to zero.

$\langle\rangle_{\rm eq}$. We know from Onsager's regression hypothesis that

$$\begin{aligned}\frac{\partial C^{\mathbf{Ideal}}}{\partial t} &= \left\langle \frac{\partial}{\partial t}\left[\rho(\mathbf{r},t)\right]_{\rho_{\rm i}}\left(\rho_{\rm i}(\mathbf{0})-\rho_0\right)\right\rangle_{\rm eq} \\ &= \left\langle D\nabla^2\left[\rho(\mathbf{r},t)\right]_{\rho_{\rm i}}\left(\rho_{\rm i}(\mathbf{0})-\rho_0\right)\right\rangle_{\rm eq} \\ &= D\nabla^2\left\langle\left[\rho(\mathbf{r},t)\right]_{\rho_{\rm i}}\left(\rho_{\rm i}(\mathbf{0})-\rho_0\right)\right\rangle_{\rm eq} \\ &= D\nabla^2\left\langle(\rho(\mathbf{r},t)-\rho_0)(\rho(\mathbf{0},0)-\rho_0)\right\rangle_{\rm ev} \\ &= D\nabla^2 C^{\mathbf{Ideal}}(\mathbf{r},t). \end{aligned} \tag{10.18}$$

The correlation function $C$ obeys the same equation as the decays of imposed initial conditions. This is true in general.[8]

Thus to solve in a general system for the correlation function $C$, we must calculate as the initial condition the instantaneous correlations $C(\mathbf{x},0)$ using equilibrium statistical mechanics, and evolve it according to the macroscopic evolution law. In the case of the ideal perfume gas, the equal–time correlations (equation 10.14) are $C^{\mathbf{Ideal}}(\mathbf{x},0)=\frac{1}{\beta\alpha}\boldsymbol{\delta}(\mathbf{x})$, and the evolution law is given by the diffusion equation. We know how an initial $\delta$ -function distribution evolves under the diffusion equation: it's given by the Green's function from section 2.4.2. The Greens function for the diffusion equation in one dimension (equation 2.33) is $G(x,t)=\frac{1}{\sqrt{4\pi Dt}}\,e^{-x^2/4Dt}$. In three dimensions we take the product along $x$, $y$,

[8] We can use Onsager's regression hypothesis to calculate the correlation function $C$ for a general order parameter $s(\mathbf{x},t)$? Suppose that the macroscopic time evolution of $s(\mathbf{x},t)$, to linear order in deviations away from its average value $\bar{s}$, is given by some Greens function (section 2.4.2):

$$s_{\rm macro}(\mathbf{x},t)=\bar{s}-\int G(\mathbf{x}-\mathbf{x}',t)(s_{\rm macro}(\mathbf{x}',0)-\bar{s})\,d\mathbf{x}'. \tag{10.19}$$

For convenience, let's set $\bar{s}=0$. This convolution simplifies if we Fourier transform in position $\mathbf{x}$ but not in time $t$, using the convolution theorem for Fourier transforms (equation A.22):

$$\widehat{s}_{\rm macro}(\mathbf{k},t)=\widehat{G}(\mathbf{k},t)\widehat{s}(\mathbf{k},0), \tag{10.20}$$

where we use a hat to denote the Fourier transform in position space. Onsager's regression hypothesis says that a spontaneous thermal fluctuation will evolve according to the same law,

$$\left[\widehat{s}(\mathbf{k},t)\right]_{\widehat{s}_{\rm i}}=\widehat{G}(\mathbf{k},t)\widehat{s}_i(\mathbf{k})=\widehat{G}(\mathbf{k},t)\widehat{s}(\mathbf{k},t=0) \tag{10.21}$$

so the connected correlation function

$$C(\mathbf{r},t)=\langle s(\mathbf{r},t)s(\mathbf{0},0)\rangle_{\rm ev}=\left\langle\left[s(\mathbf{r},t)\right]_{s_{\rm i}}s_i(\mathbf{0})\right\rangle_{\rm eq} \tag{10.22}$$

evolves by

$$\begin{aligned}\widehat{C}(\mathbf{k},t) &= \left\langle\left[\widehat{s}(\mathbf{k},t)\right]_{s_{\rm i}}s_i(\mathbf{0})\right\rangle_{\rm eq}=\left\langle\widehat{G}(\mathbf{k},t)\widehat{s}(\mathbf{k},0)s_i(\mathbf{0})\right\rangle_{\rm eq} \\ &= \widehat{G}(\mathbf{k},t)\widehat{C}(\mathbf{k},0) \end{aligned} \tag{10.23}$$

Again, the correlation function obeys the same evolution law as the decay of an imposed initial condition.

and $z$ to get $G$, and then divide by $\beta\alpha$, to get the correlation function

$$C^{\mathbf{Ideal}}(\mathbf{r},t) = \frac{1}{\beta\alpha}G(\mathbf{r},t) = \frac{1}{\beta\alpha}\left(\frac{1}{\sqrt{4\pi Dt}}\right)^3 e^{-\mathbf{r}^2/4Dt}. \tag{10.24}$$

This is the correlation function for an ideal gas satisfying the diffusion equation.

## 10.5 Susceptibility and the Fluctuation–Dissipation Theorem

How will our system ring or yield when we kick it? The *susceptibility* $\chi(\mathbf{r},\tau)$ gives the response at a distance $\mathbf{r}$ and time $\tau$ from the kick. Let us formulate susceptibility for a general order parameter $s(\mathbf{x},t)$, kicked by an external field $f(\mathbf{x},t)$. That is, we assume that $f$ appears in the free energy functional

$$F = F_0 + F_f \tag{10.25}$$

as a term

$$F_f(t) = -\int d\mathbf{x}\, f(\mathbf{x},t)s(\mathbf{x},t). \tag{10.26}$$

You can think of $f$ as a force density pulling $s$ upward. If $s$ is the particle density $\rho$, then $f$ is minus an external potential $-V(\mathbf{x})$ for the particles; if $s$ is the magnetization $M$ of an Ising model, then $f$ is the external field $H$; if $s$ is the polarization $\mathbf{P}$ of a dielectric material, then $f$ is an externally applied vector electric field $\mathbf{E}(\mathbf{x},t)$. For convenience, we will assume in this section that $\bar{s} = 0$.[9]

[9] Or rather, we define the order parameter $s$ as the deviation from the equilibrium value in the absence of a field.

How will the order parameter field $s$ respond to the force $f$? If the force is a weak perturbation, we can presume a linear response, but perhaps one which is non–local in space and time. So, $s(\mathbf{x},t)$ will depend upon $f(\mathbf{x}',t')$ at all earlier times $t' < t$:

$$s(\mathbf{x},t) = \int d\mathbf{x}' \int_{-\infty}^{t} dt'\, \chi(\mathbf{x}-\mathbf{x}', t-t')f(\mathbf{x}',t'). \tag{10.27}$$

This non–local relation becomes much simpler if we Fourier transform[10] $s$, $f$, and $\chi$ in space and time. The AC susceptibility[11] $\widetilde{\chi}(\mathbf{k},\omega)$ satisfies

$$\widetilde{s}(\mathbf{k},\omega) = \widetilde{\chi}(\mathbf{k},\omega)\widetilde{f}(\mathbf{k},\omega), \tag{10.28}$$

since as usual the Fourier transform of the convolution is the product of the Fourier transforms (equation A.22). The function $\chi$ is the *susceptibility* of the order parameter $s$ to the external field $f$. For example, the polarization versus field is defined in terms of the polarizability $\alpha$:[12] $\widetilde{\mathbf{P}}(\mathbf{k},\omega) = \widetilde{\alpha}(\mathbf{k},\omega)\widetilde{\mathbf{E}}(\mathbf{k},\omega)$, the magnetization from an external field is $\widetilde{\mathbf{M}}(\mathbf{k},\omega) = \widetilde{\chi}(\mathbf{k},\omega)\widetilde{\mathbf{H}}(\mathbf{k},\omega)$, and so on.

[10] We will use $\widetilde{A}(\mathbf{k},\omega)$ to represent the Fourier transform of the function $A(\mathbf{x},t)$ with respect to both space and time. Later in this section we will use $\tilde{B}(\mathbf{k})$ to represent the Fourier transform of the static function $B(\mathbf{x})$ with respect to space, and $\widehat{A}(\mathbf{k},t)$ to represent the Fourier transform of $A(\mathbf{x},t)$ in space $\mathbf{x}$ alone.

[11] AC stands for 'alternating current', the kind of electricity that is used in most buildings except in rural areas: the voltage fluctuates periodically in time. The current from batteries is DC or direct current, which doesn't vary in time. Somehow we've started using AC for frequency dependent properties in general.

[12] Because $\mathbf{P}$ and $\mathbf{E}$ are vectors, $\alpha$ will be a matrix in anisotropic materials.

### 10.5.1 Dissipation and the imaginary part $\chi''(\omega)$

In general the real–space susceptibility $\chi(\mathbf{x},t)$ is real, but the AC susceptibility

$$\widetilde{\chi}(\mathbf{k},\omega) = \int d\mathbf{x}dt\, e^{i\omega t}e^{i\mathbf{k}\cdot\mathbf{x}}\chi(\mathbf{x},t) = \chi'(\mathbf{k},\omega) + i\chi''(\mathbf{k},\omega) \tag{10.29}$$

has a real part $\chi' = Re[\widetilde{\chi}]$ and an imaginary part $\chi'' = Im[\widetilde{\chi}]$.[13] It is clear from the definition that $\widetilde{\chi}(-\mathbf{k},-\omega) = \widetilde{\chi}^*(\mathbf{k},\omega)$; for a system with inversion symmetry $\mathbf{x} \leftrightarrow -\mathbf{x}$ we see further that $\chi(\mathbf{x},t) = \chi(-\mathbf{x},t)$ and hence $\widetilde{\chi}(\mathbf{k},-\omega) = \widetilde{\chi}^*(\mathbf{k},\omega)$, so $\chi'$ is even in $\omega$ and $\chi''$ is odd. $\chi'$ gives the in–phase response to a sinusoidal force, and $\chi''$ gives the response that lags in phase.[14]

The imaginary part $\chi''$ in general gives the amount of *dissipation* induced by the external field. The dissipation can be measured directly (for example, by measuring the resistance as a function of frequency of a wire) or by looking at the decay of waves in the medium (optical and ultrasonic attenuation and such). We know that 'energy' is the integral of 'force' $f$ times 'distance' $\partial s$, or force times velocity $\partial s/\partial t$ integrated over time. Ignoring the spatial dependence for simplicity, the time average of the power $p$ dissipated per unit volume is

$$p = \lim_{T\to\infty}\frac{1}{T}\int_0^T f(t)\frac{\partial s}{\partial t}\,dt = \lim_{T\to\infty}\frac{1}{T}\int_0^T -s(t)\frac{\partial f}{\partial t}\,dt \tag{10.31}$$

where we've averaged over a time $T$ and integrated by parts, assuming the boundary terms are negligible for $T\to\infty$. Assuming an AC force $f(t) = Re[f_\omega e^{-i\omega t}] = \frac{1}{2}(f_\omega e^{-i\omega t} + f_\omega^* e^{i\omega t})$,

$$p(\omega) = \lim_{T\to\infty}\frac{1}{T}\int_0^T s(t)\frac{i\omega}{2}(f_\omega e^{-i\omega t} - f_\omega^* e^{i\omega t})\,dt, \tag{10.32}$$

where the motion $s(t)$ is in turn due to the forcing at earlier times

$$\begin{aligned} s(t) &= \int_{-\infty}^{\infty} dt'\,\chi(t-t')f(t') \\ &= \int_{-\infty}^{\infty} d\tau\,\chi(\tau)f(t-\tau) \\ &= \int_{-\infty}^{\infty} d\tau\,\frac{\chi(\tau)}{2}(f_\omega e^{-i\omega(t-\tau)} + f_\omega^* e^{i\omega(t-\tau)}). \end{aligned} \tag{10.33}$$

Plugging [10.33] into [10.32], we get

$$\begin{aligned} p(\omega) &= \lim_{T\to\infty}\frac{1}{T}\int_0^T dt\int_{-\infty}^{\infty} d\tau\frac{i\omega\chi(\tau)}{4} \\ &\qquad (f_\omega e^{-i\omega(t-\tau)} + f_\omega^* e^{i\omega(t-\tau)})(f_\omega e^{-i\omega t} - f_\omega^* e^{i\omega t}) \\ &= \int_{-\infty}^{\infty} d\tau\frac{i\omega\chi(\tau)}{4}\lim_{T\to\infty}\frac{1}{T}\int_0^T dt \\ &\qquad \left[f_\omega^2 e^{-i\omega(2t-\tau)} - f_\omega^{*2}e^{i\omega(2t-\tau)} + |f_\omega|^2(e^{-i\omega\tau} - e^{i\omega\tau})\right]. \end{aligned} \tag{10.34}$$

[13] Some use the complex conjugate of our formulas for the Fourier transform (see section A.1), substituting $-i$ for $i$ in all the formulas. Their $\chi''$ is the same as ours, because they define it to be *minus* the imaginary part of their Fourier transformed susceptibility.

[14] If we apply $f(t) = \cos(\omega t)$, so $\widetilde{f}(\omega) = \frac{1}{2}(\delta(\omega) + \delta(-\omega))$, then then the response is $\widetilde{s}(\omega) = \widetilde{\chi}(\omega)\widetilde{f}(\omega)$ so

$$\begin{aligned} s(t) &= \frac{1}{2\pi}\int e^{-i\omega t}\widetilde{s}(\omega)\,d\omega \\ &= \frac{1}{4\pi}\left(e^{-i\omega t}\chi(\omega) + e^{i\omega t}\chi(-\omega)\right) \\ &= \frac{1}{4\pi}\big(e^{-i\omega t}(\chi'(\omega) + i\chi''(\omega)) \\ &\qquad + e^{i\omega t}(\chi'(\omega) - i\chi''(\omega))\big) \\ &= \frac{1}{2\pi}\left(\chi'(\omega)\cos(\omega t) + \chi''(\omega)\sin(\omega t)\right). \end{aligned} \tag{10.30}$$

Hence $\chi'$ gives the immediate in–phase response, and $\chi''$ gives the out–of–phase delayed response.

The first and second term are zero, and the third gives a sine,[15] so

$$\begin{aligned} p(\omega) =& \frac{\omega|f_\omega|^2}{2}\int_{-\infty}^{\infty} d\tau \chi(\tau)\sin(\omega\tau) = \frac{\omega|f_\omega|^2}{2} Im[\chi(\omega)] \\ =& \frac{\omega|f_\omega|^2}{2}\chi''(\omega). \end{aligned} \tag{10.35}$$

Let's interpret this formula in the familiar case of electrical power dissipation in a wire. Under a (reasonably low frequency) AC voltage $V\cos(\omega t)$, a wire of resistance $R$ dissipates power $V^2/R$. A wire of length $L$ and cross-section $A$ has electric field $E = V/L$ and resistance $R = \frac{L}{\sigma A}$ where $\sigma$ is the conductivity of the metal. So the dissipated power per unit volume $p = \frac{1}{2}\frac{(EL)^2}{L/\sigma A}\frac{1}{LA} = \frac{1}{2}\sigma E^2$; at general frequencies the conductivity is defined so that $p(\omega) = \frac{1}{2}\sigma(\omega)|E(\omega)|^2$. Equation 10.35 thus tells us in this context that $\sigma(\omega) = \alpha''(\omega)$; the imaginary part of the polarizability divided by the frequency.

### 10.5.2 Static susceptibility $\widetilde{\chi}_0(\mathbf{k})$

In many cases, we are interested in how a system responds to a static external force – rather than kicking a system, we lean on it. Under a point-like force, the dimple formed in the order parameter field is described by the *static susceptibility* $\chi_0(\mathbf{r})$.

If the external force is time independent (so $f(\mathbf{x}', t') = f(\mathbf{x}')$) the system will reach a perturbed equilibrium, and we may use equilibrium statistical mechanics to find the resulting static change in the average order parameter field $s(\mathbf{x})$. The non-local relation between $s$ and a small field $f$ is given by the *static susceptibility*, $\chi_0$:

$$s(\mathbf{x}) = \int d\mathbf{x}'\, \chi_0(\mathbf{x}-\mathbf{x}')f(\mathbf{x}'). \tag{10.36}$$

If we take the Fourier series of $s$ and $f$, we may represent this relation in terms of the Fourier transform of $\chi_0$ in analogy to equation 10.28:

$$\widetilde{s}_{\mathbf{k}} = \widetilde{\chi}_0(\mathbf{k})\widetilde{f}_{\mathbf{k}}. \tag{10.37}$$

As an example, the free energy density for the ideal gas in the linearized approximation of section 10.3, is $\mathcal{F} = \frac{1}{2}\alpha(\rho-\rho_0)^2$. For a spatially varying static external potential $f(\mathbf{x}) = -V(\mathbf{x})$, this is minimized by $\rho(\mathbf{x}) = \rho_0 + f(\mathbf{x})/\alpha$, so (comparing to equation 10.36) we find the static susceptibility is

$$\chi_0^{\mathbf{Ideal}}(\mathbf{r}) = \delta(\mathbf{r})/\alpha, \tag{10.38}$$

and in Fourier space it is $\widetilde{\chi}_0(\mathbf{k}) = 1/\alpha$. $\chi_0$ is the "spring constant" giving the response to a constant external force.

[15] In particular, $\lim_{T\to\infty}(1/T)\int_0^T dt e^{\pm 2i\omega t} = 0$, $\lim_{T\to\infty}(1/T)\int_0^T dt e^{0i\omega t} = 1$, and $i(e^{-i\omega\tau} - e^{i\omega\tau}) = 2\sin(\omega\tau)$.

Notice for the ideal gas that the static susceptibility $\chi_0^{\mathbf{Ideal}}(\mathbf{r})$ and the equal–time correlation function $C^{\mathbf{Ideal}}(\mathbf{r},0) = \frac{\delta(\mathbf{r})}{\beta\alpha}$ are proportional to one another: $\chi_0^{\mathbf{Ideal}}(\mathbf{r}) = \beta C^{\mathbf{Ideal}}(\mathbf{r},0)$. This can be shown to be true in general for equilibrium systems:[16]

$$\chi_0(\mathbf{r}) = \beta C(\mathbf{r},0) \tag{10.42}$$

the $\omega = 0$ static susceptibility $\chi_0$ is given by dividing the instantaneous correlation function $C$ by $k_B T$ – both in real space and also in Fourier space:

$$\widetilde{\chi}_0(\mathbf{k}) = \beta \widehat{C}(\mathbf{k},0). \tag{10.43}$$

This *fluctuation-response* relation should be intuitively reasonable: a system or mode which is easy to perturb will also have big fluctuations.

[16] To calculate the static susceptibility for a general system, we formally find the expectation value $\langle \widetilde{s}_{\mathbf{k}} \rangle_{\mathrm{eq}}$ for a given $\widetilde{f}_{\mathbf{k}}$, and then take the derivative with respect to $\widetilde{f}_{\mathbf{k}}$ to get $\widetilde{\chi}_0(\mathbf{k})$. The interaction term in the free energy equation 10.26 reduces in the case of a static force to

$$F_f = -\int d\mathbf{x}\, f(\mathbf{x}) s(\mathbf{x}) = -V \sum_{\mathbf{k}} \widetilde{f}_{\mathbf{k}} \widetilde{s}_{-\mathbf{k}} \tag{10.39}$$

where $V$ is the size of the system (we assume periodic boundary conditions) and the sum over $\mathbf{k}$ is the sum over allowed wave-vectors in the box (appendix A). (We use Fourier series here instead of Fourier transforms because it makes the calculations more intuitive: we get factors of the volume rather than $\delta$-function infinities.) The expectation value of the order parameter in the field is

$$\langle \widetilde{s}_{\mathbf{k}} \rangle_{\mathrm{eq}} = \frac{\mathrm{Tr}\left[\widetilde{s}_{\mathbf{k}} e^{-\beta\left(F_0 - V\sum_{\mathbf{k}} \widetilde{f}_{\mathbf{k}} \widetilde{s}_{-\mathbf{k}}\right)}\right]}{\mathrm{Tr}\left[e^{-\beta\left(F_0 - V\sum_{\mathbf{k}} \widetilde{f}_{\mathbf{k}} \widetilde{s}_{-\mathbf{k}}\right)}\right]} = \frac{1}{\beta}\frac{\partial \log Z}{\partial f_k} \tag{10.40}$$

where Tr integrates over all order parameter configurations $s$. The susceptibility is given by differentiating equation 10.40:

$$\begin{aligned}
\widetilde{\chi}_0(\mathbf{k}) = & \left.\frac{\partial \langle \widetilde{s}_{\mathbf{k}} \rangle_{\mathrm{eq}}}{\partial \widetilde{f}_{\mathbf{k}}}\right|_{f=0} \\
= & \left.\frac{\mathrm{Tr}\left[\widetilde{s}_{\mathbf{k}} (\beta V \widetilde{s}_{-\mathbf{k}}) e^{-\beta\left(F_0 - V\sum_{\mathbf{k}} \widetilde{f}_{\mathbf{k}} \widetilde{s}_{-\mathbf{k}}\right)}\right]}{\mathrm{Tr}\left[e^{-\beta\left(F_0 - V\sum_{\mathbf{k}} \widetilde{f}_{\mathbf{k}} \widetilde{s}_{-\mathbf{k}}\right)}\right]}\right|_{f=0} \\
& - \left.\frac{\mathrm{Tr}\left[\widetilde{s}_{\mathbf{k}} e^{-\beta\left(F_0 - V\sum_{\mathbf{k}} \widetilde{f}_{\mathbf{k}} \widetilde{s}_{-\mathbf{k}}\right)}\right] \mathrm{Tr}\left[(\beta V \widetilde{s}_{-\mathbf{k}}) e^{-\beta\left(F_0 - V\sum_{\mathbf{k}} \widetilde{f}_{\mathbf{k}} \widetilde{s}_{-\mathbf{k}}\right)}\right]}{\mathrm{Tr}\left[e^{-\beta\left(F_0 - V\sum_{\mathbf{k}} \widetilde{f}_{\mathbf{k}} \widetilde{s}_{-\mathbf{k}}\right)}\right]^2}\right|_{f=0} \\
= & \beta V \left(\langle \widetilde{s}_{\mathbf{k}} \widetilde{s}_{-\mathbf{k}} \rangle - \langle \widetilde{s}_{\mathbf{k}} \rangle \langle \widetilde{s}_{-\mathbf{k}} \rangle\right) = \beta V \left\langle (\widetilde{s}_{\mathbf{k}} - \langle \widetilde{s}_{\mathbf{k}} \rangle)^2 \right\rangle \\
= & \beta \widehat{C}(\mathbf{k},0).
\end{aligned} \tag{10.41}$$

where the last equation is the Fourier equality of the correlation function to the absolute square of the fluctuation (equation A.20, except that, because we're using Fourier series instead of Fourier transforms, there are two extra factors of $V$ and the $\langle s_{\mathbf{k}} \rangle$ subtraction gives us the connected correlation function, with $\bar{s}$ subtracted off).

Note that in equation 10.40 that $\langle \widetilde{s}_{\mathbf{k}} \rangle = \frac{1}{\beta}\frac{\partial \log Z}{\partial \widetilde{f}_{\mathbf{k}}}$ and in equation 10.41 $\widehat{C}(\mathbf{k},0) = \frac{1}{\beta^2}\frac{\partial \log Z}{\partial \widetilde{f}_{\mathbf{k}}^2}$: as usual, everything is calculated by taking derivatives of the partition function. The higher connected correlation functions can be gotten in turn by taking higher derivatives of $\log Z$. This is a common theoretical technique: to calculate correlations in an ensemble, add a force coupled to the corresponding field and take derivatives.

How is the static susceptibility $\chi_0(\mathbf{r})$ related to our earlier dynamic susceptibility $\chi(\mathbf{r},t)$? We can use the dynamic susceptibility (equation 10.27) in the special case of a time–independent force

$$\begin{aligned} s(\mathbf{x},t) &= \int d\mathbf{x}' \int_{-\infty}^{t} dt'\, \chi(\mathbf{x}-\mathbf{x}', t-t') f(\mathbf{x}') \\ &= \int d\mathbf{x}' \int_{0}^{\infty} d\tau\, \chi(\mathbf{x}-\mathbf{x}', \tau) f(\mathbf{x}'), \end{aligned} \tag{10.44}$$

to derive a formula for $\chi_0$:

$$\chi_0(\mathbf{r}) = \int_0^\infty dt\, \chi(\mathbf{r},t) = \int_{-\infty}^\infty dt\, \chi(\mathbf{r},t). \tag{10.45}$$

Here we use the fact that the physical world obeys *causality* (effects cannot precede their cause) to set $\chi(\mathbf{r},t) = 0$ for $t < 0$ (see figure 10.10). By integrating over time in equation 10.45 we extract the $\omega = 0$ Fourier component, so the $\mathbf{k}$ dependent static susceptibility is the zero–frequency limit of the AC susceptibility:

$$\widetilde{\chi}_0(\mathbf{k}) = \widetilde{\chi}(\mathbf{k}, \omega = 0). \tag{10.46}$$

Often, one discusses the uniform susceptibility of a system – the response to an external field uniform in space and time. The specific heat of section 6.1 and the magnetic susceptibility of exercise 8.1 are the uniform $\mathbf{k} = \mathbf{0}$ value of the static susceptibility to changes in temperature and field. For the uniform static susceptibility, $s = \int d\mathbf{x}' \chi_0(\mathbf{x}-\mathbf{x}') f = \widetilde{\chi}_0(\mathbf{k}=\mathbf{0}) f$, so the uniform susceptibility is given by $\widetilde{\chi}_0(\mathbf{k})$ at $\mathbf{k} = \mathbf{0}$. Knowing $\widetilde{\chi}_0(\mathbf{k}) = \beta \widehat{C}(\mathbf{k}, t=0)$ (equation 10.43), we can relate the uniform susceptibility to the $\mathbf{k} = \mathbf{0}$ component of the equal-time correlation function. But at $\mathbf{k} = \mathbf{0}$, the correlation function is given by the average square of the order parameter, so

$$\begin{aligned} k_B T \widetilde{\chi}_0(\mathbf{k}=\mathbf{0}) &= \widehat{C}(\mathbf{k}=\mathbf{0}, t=0) = \int d\mathbf{r} \langle s(\mathbf{r}+\mathbf{x}) s(\mathbf{x}) \rangle \\ &= \int d\mathbf{r} \frac{1}{V} \left\langle \int d\mathbf{x}\, s(\mathbf{r}+\mathbf{x}) s(\mathbf{x}) \right\rangle = \frac{1}{V} \left\langle \int d\mathbf{r}' s(\mathbf{r}') \int d\mathbf{x}\, s(\mathbf{x}) \right\rangle \\ &= \frac{1}{V} \left\langle \left( \int s(\mathbf{x}) d\mathbf{x} \right)^2 \right\rangle \end{aligned} \tag{10.47}$$

We've thus connected a uniform linear response to the fluctuations of the whole system. We've done this in special cases twice before, in exercise 8.1(b) where the fluctuations in magnetization gave the susceptibility in the Ising model, and equation 6.13 where the energy fluctuations were related to the specific heat.[17]

[17] The energy fluctuations in equation 6.13 have a slightly different formula. The inverse temperature $\beta$ already couples to the energy fluctuations, so there is no separate conjugate force.

### 10.5.3 $\chi(\mathbf{r},t)$ and Fluctuation–Dissipation

Now we turn to computing the dynamic susceptibility. It too is related to the correlation function, via the *fluctuation-dissipation* theorem.

How can we compute $\chi(\mathbf{r},t)$ the space-time evolution after we kick the system at $\mathbf{r}=t=0$? We know the time evolution starting from an imposed *initial condition* is given by the Greens function $G(\mathbf{r},t)$. We can impose an initial condition using a static force $f(\mathbf{x},t)=f(\mathbf{x})$ for $t<0$, and release it at $t=0$ so $f(\mathbf{x},t)=0$ for $t>0$. We can then match[18] the Green's function time evolution $s(\mathbf{x},t)=\int dx' G(\mathbf{x}-\mathbf{x}',t)s(\mathbf{x}',0)$ with that given by the susceptibility $s(\mathbf{x},t)=\int_{-\infty}^{0}dt'\int d\mathbf{x}' f(\mathbf{x}')\chi(\mathbf{x}-\mathbf{x}',t-t')$.

Let's work it out for the ideal gas, where $\chi_0(\mathbf{r})=\delta(\mathbf{r})/\alpha$ (equation 10.38), so $\rho(\mathbf{x},0)=f(\mathbf{x})/\alpha$. The subsequent time evolution is given by the Greens function $G(\mathbf{x},t)$, which we've seen for the ideal gas gives the correlation function $C^{\mathbf{Ideal}}(\mathbf{x},t)=G(\mathbf{x},t)/(\beta\alpha)$ by Onsager's

[18] We can do this for a general order parameter field $s(\mathbf{x},t)$. We start with an initial condition defined by a static external field $f(\mathbf{x})$, which is given by

$$\widehat{s}(\mathbf{k},t=0)=\widetilde{\chi}_0(\mathbf{k})\widetilde{f}(\mathbf{k}). \tag{10.48}$$

The subsequent time evolution is given by convolving with the Greens function $G(\mathbf{x},t)$ (equation 2.36), which is the same as multiplying by $\widehat{G}(\mathbf{k},t)$:

$$\widehat{s}(\mathbf{k},t)=\widetilde{\chi}_0(\mathbf{k})\widetilde{f}(\mathbf{k})\widehat{G}(\mathbf{k},t) \tag{10.49}$$

We can also find an equation for $\widehat{s}(\mathbf{k},t)$ by using the dynamic susceptibility, equation 10.27, and the fact that $f(t')=0$ for $t'>0$:

$$\begin{aligned} s(\mathbf{x},t) &= \int d\mathbf{x}' \int_{-\infty}^{t} dt' \chi(\mathbf{x}-\mathbf{x}',t-t')f(\mathbf{x}',t') \\ &= \int d\mathbf{x}' \int_{-\infty}^{0} dt' \chi(\mathbf{x}-\mathbf{x}',t-t')f(\mathbf{x}') \\ &= \int d\mathbf{x}' \int_{t}^{\infty} dt' \chi(\mathbf{x}-\mathbf{x}',\tau)f(\mathbf{x}') \end{aligned} \tag{10.50}$$

so

$$\widehat{s}(\mathbf{k},t)=\int_{t}^{\infty} d\tau \widehat{\chi}(\mathbf{k},\tau)\widetilde{f}(\mathbf{k}). \tag{10.51}$$

This is true for any $\widetilde{f}(\mathbf{k})$, so with equation 10.49, we find

$$\int_{t}^{\infty} d\tau \widehat{\chi}(\mathbf{k},\tau)=\widetilde{\chi}_0(\mathbf{k})\widehat{G}(\mathbf{k},t) \tag{10.52}$$

Now from the last section, equation 10.43, we know $\widetilde{\chi}_0(\mathbf{k})=\beta\widehat{C}(\mathbf{k},0)$. From the Onsager regression hypothesis, the Greens function $\widehat{G}(\mathbf{k},t)$ for $s$ has the same evolution law as is obeyed by the correlation function $C$ (equation 10.23), so

$$\widehat{C}(\mathbf{k},0)\widehat{G}(\mathbf{k},t)=\widehat{C}(\mathbf{k},t) \tag{10.53}$$

Hence

$$\begin{aligned} \int_{t}^{\infty} d\tau \widehat{\chi}(\mathbf{k},\tau) &= \beta\widehat{C}(\mathbf{k},0)\widehat{G}(\mathbf{k},t) \\ &= \beta\widehat{C}(\mathbf{k},t) \end{aligned} \tag{10.54}$$

Differentiating both sides with respect to time yields the fluctuation-dissipation theorem in $\mathbf{k}$-space:

$$\widehat{\chi}(\mathbf{k},t)=-\beta\frac{\partial\widehat{C}(\mathbf{k},t)}{\partial t}. \tag{10.55}$$

regression hypothesis (equation 10.24):

$$\begin{aligned}\rho(\mathbf{x},t) &= \int d\mathbf{x}' \rho(\mathbf{x}',0) G(\mathbf{x}-\mathbf{x}',t) \\ &= \int d\mathbf{x}' \frac{f(\mathbf{x}')}{\alpha} G(\mathbf{x}-\mathbf{x}',t) \\ &= \int d\mathbf{x}' f(\mathbf{x}') \beta C^{\mathbf{Ideal}}(\mathbf{x}-\mathbf{x}',t)\end{aligned} \tag{10.56}$$

We match against $\rho(\mathbf{x},t)$ written using the dynamical susceptibility. Since $f(\mathbf{x},t) = 0$ for $t > 0$ the formula involves integrals up to time zero: we change variables to $\tau = t - t'$:

$$\begin{aligned}\rho(\mathbf{x},t) &= \int d\mathbf{x}' \int_{-\infty}^{0} dt' f(\mathbf{x}') \chi(\mathbf{x}-\mathbf{x}', t-t') \\ &= \int d\mathbf{x}' f(\mathbf{x}') \int_{t}^{\infty} d\tau \chi(\mathbf{x}-\mathbf{x}', \tau).\end{aligned} \tag{10.57}$$

Comparing these two formulas, we see that

$$\beta C^{\mathbf{Ideal}}(\mathbf{r},t) = \int_{t}^{\infty} d\tau \chi(\mathbf{r},\tau). \tag{10.58}$$

Taking the derivative of both sides, we derive one form of the *Fluctuation–Dissipation Theorem*:

$$\chi^{\mathrm{Ideal}}(\mathbf{r},t) = -\beta \frac{\partial C^{\mathbf{Ideal}}}{\partial t} \qquad (t>0) \tag{10.59}$$

The fluctuation–dissipation theorem in this form is true in general for the linear response of classical equilibrium systems (see note 18). The linear dynamic susceptibility $\chi$ of a general order parameter field $s(\mathbf{x},t)$ with correlation function $C(\mathbf{x},t)$ is given by

$$\chi(\mathbf{x},t) = -\beta \frac{\partial C(\mathbf{x},t)}{\partial t} \quad (t>0). \tag{10.60}$$

What happens for $t < 0$? The correlation function must be symmetric in time (figure 10.9) since the equilibrium state is invariant under time reversal symmetry:

$$\begin{aligned}C(\mathbf{r},\tau) &= \langle s(\mathbf{x},t) s(\mathbf{x}+\mathbf{r}, t+\tau)\rangle \\ &= \langle s(\mathbf{x},t) s(\mathbf{x}+\mathbf{r}, t-\tau)\rangle = C(\mathbf{r},-\tau).\end{aligned} \tag{10.61}$$

But $\chi$ must be zero for $t < 0$ (figure 10.10) by causality:

$$\chi(\mathbf{r},t) = 0 \qquad (t<0). \tag{10.62}$$

We can see why it is called the fluctuation-*dissipation* theorem by looking at the AC version of the law. Again, for convenience, we ignore

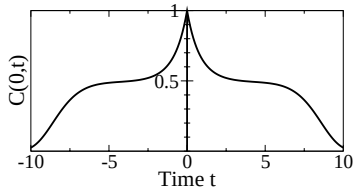



**Fig. 10.9** The time–time correlation function $C(\mathbf{r} = \mathbf{0}, \tau)$ for a hypothetical system with two characteristic relaxation times.

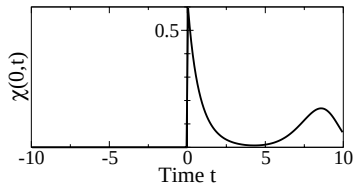



**Fig. 10.10** The susceptibility $\chi(\mathbf{r} = \mathbf{0}, t)$ for the same hypothetical system as in figure 10.4. $\chi(\mathbf{r},t)$ gives the response to an impulse a time $t$ in the past; causality requires that there be no response *preceding* the impulse, so $\chi(\mathbf{r},t) = 0$ for $t < 0$.

the spatial degrees of freedom. Using equations 10.60 and 10.62, and integrating by parts, we find

$$\begin{aligned}\widetilde{\chi}(\omega) &= \int_{-\infty}^{\infty} dt \chi(t) e^{i\omega t} = -\beta \int_0^{\infty} dt \frac{\partial C}{\partial t} e^{i\omega t} \\ &= -\beta C(t) e^{i\omega t}\big|_0^{\infty} + i\omega\beta \int_0^{\infty} dt C(t) e^{i\omega t}. \end{aligned} \tag{10.63}$$

Now, the first term is real and $C(t) = C(-t)$, so we may write the imaginary part of the susceptibility as

$$\begin{aligned}\chi''(\omega) = Im[\widetilde{\chi}(\omega)] &= \beta\omega \int_0^{\infty} dt C(t) \cos(\omega t) \\ &= \frac{\beta\omega}{2} \int_{-\infty}^{\infty} dt C(t) e^{i\omega t} = \frac{\beta\omega}{2} \widetilde{C}(\omega). \end{aligned} \tag{10.64}$$

Using this result and equation 10.35 relating the dissipation to $\chi''$ and putting back the spatial pieces,

$$\begin{aligned} p(\omega) &= \frac{\omega |f_\omega|^2}{2} \chi''(\omega) = \frac{\omega |f_\omega|^2}{2} \frac{\beta\omega}{2} \widetilde{C}(\omega) \\ &= \frac{\beta\omega^2 |f_\omega|^2}{4} \widetilde{C}(\omega). \end{aligned} \tag{10.65}$$

This tells us that the power dissipated, $p(\omega)$ under an external forcing $f_\omega$ is given in terms of the correlation function of the spontaneous fluctuations $C(\omega)$: hence the name fluctuation–dissipation theorem.

We've ignored quantum mechanics in our derivation[19] and indeed there are quantum-mechanical correction of order $\hbar$. The fully quantum version of the fluctuation–dissipation theorem is

$$\chi''(\mathbf{k}, \omega) = Im[\widetilde{\chi}(\mathbf{k}, \omega)] = \frac{1}{2\hbar}(1 - e^{-\beta\hbar\omega}) \widetilde{C}(\mathbf{k}, \omega). \tag{10.66}$$

At high temperatures, $1 - e^{-\beta\hbar\omega} \sim \beta\hbar\omega$ and we regain our classical result, equation 10.64.

[19] Usually, one derives the fluctuation–dissipation theorem quantum mechanically, and then uses it to derive the Onsager regression hypothesis [32].

## 10.6 Causality and Kramers Krönig

The susceptibility $\chi(t)$ (again, dropping the positions for simplicity) is a real-valued function on the half-line $t > 0$. The frequency-dependent susceptibility is composed of two real-valued functions $\chi'(\omega)$ and $\chi''(\omega)$ on the entire line. We can use the symmetries $\widetilde{\chi}(-\omega) = -\widetilde{\chi}^*(\omega)$ to reduce this to two real-valued functions on the half-line $\omega > 0$, but it still seems like $\widetilde{\chi}(\omega)$ contains twice the information of $\chi(t)$. It makes it plausible that $\chi'$ and $\chi''$ might be related somehow. Suppose we measure the frequency–dependent absorption of the material, and deduce $\chi''(\mathbf{k}, \omega)$. Can we find the real part of the susceptibility $\chi'(\mathbf{k}, \omega)$?

It's a remarkable fact that we can find a formula for $\chi'(\omega)$ in terms of $\chi''(\omega)$. This relation is called the *Kramers–Krönig relation*, and it

follows from *causality*. For this argument, you'll need to know some complex analysis.[20]

We know that $\chi(t) = 0$ for $t < 0$, because the laws of nature are causal: the response cannot precede the perturbation. What does this imply about $\chi(\omega) = \int dt\, \chi(t) e^{i\omega t}$? Consider the function $\chi$ as a function of a complex frequency $\omega = u + iv$:

$$\chi(u + iv) = \int dt \chi(t) e^{iut} e^{-vt} \tag{10.67}$$

converges nicely for $v > 0$, but looks dubious for $v < 0$. In the complex $\omega$ plane, this implies that $\chi(\omega)$ is analytic in the upper half plane ($v > 0$).[21] Also, it seems likely that there will be singularities (*e.g.* poles) in $\widetilde{\chi}(\omega)$ in the lower half-plane.

We now apply deep results of complex analysis. *Cauchy's theorem* states that the line integral of a complex function around a contour is zero, $\oint_C f(\omega')d\omega' = 0$, if the function $f$ is analytic everywhere inside the contour. *Cauchy's integral theorem* [22] states under the same conditions that if the contour $C$ circles a point $\omega$ counterclockwise once, then

$$\oint_C \frac{f(\omega')}{\omega' - \omega}\, d\omega' = 2\pi i\, f(\omega). \tag{10.69}$$

Now consider the contour in figure 10.11, and the ordinary line integral $\int_{-\infty}^{\infty} \frac{\chi(\omega')}{\omega' - \omega}\, d\omega'$. If we change this integral to avoid the pole at $\omega = \omega'$ by deforming it in a tiny semicircle above the real axis as shown, and close the contour in the upper half plane, then (since $\chi$ is analytic in the upper half plane) Cauchy's theorem says the integral is zero. If we deform it in a tiny semicircle *below* the real axis our path circles the pole counterclockwise once, and hence the integral is $2\pi i \chi(\omega)$. An expansion about the point $\omega' = \omega$ tells us that if we plow directly through the pole on the real axis[23] then the integral around the semicircle is the average of these two results, $\pi i \chi(\omega)$. But the integral back along the large semicircle can be shown to vanish as the radius goes to infinity, so we find an equation for the line integral giving $\chi(\omega)$ in terms of itself:

$$\chi(\omega) = \frac{1}{\pi i} \int_{-\infty}^{\infty} \frac{\chi(\omega')}{\omega' - \omega} d\omega'. \tag{10.70}$$

[20] If you haven't heard of Cauchy's theorem, read on – you'll be getting a preview of the key result in complex analysis.

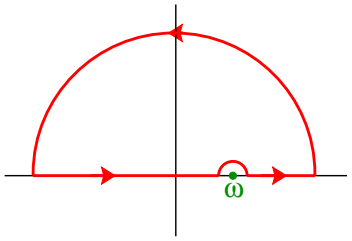



**Fig. 10.11** A contour $C$ in the complex $\omega'$ plane. The horizontal axis is $Re(\omega')$ and the vertical axis is $Im(\omega')$. The integration contour runs along the real axis from $-\infty$ to $\infty$ with a tiny semicircular detour near a pole at $\omega$. The contour is closed with a semicircle back at infinity, where $\chi(\omega')$ is vanishes rapidly. The contour encloses no singularities, so Cauchy's theorem tells us the integral around it is zero. Detouring the tiny semicircle below the axis gives the full residue of the pole at $\omega$: plowing straight through, giving the integral along the real $\omega'$ axis, gives half of the residue of the pole.

[23] We must take the principle value, symmetrically approaching the pole from both sides

[21] Analytic means that Taylor series about points with $v > 0$ converge. It's amazing how many functions in physics are analytic: it seems we almost always can assume power series make sense. We've discussed in section 8.1 that all properties of materials are analytic functions of the parameters inside phases; we've discussed in exercise 9.5 how the free energy for finite systems and for finite coarse–grainings is an analytic function of the natural variables in a system. Here we find yet another excuse for finding analytic functions: causality!

[22] The integral theorem looks like a complete miracle, but it follows pretty simply from Cauchy's theorem: one deforms the path using Cauchy's theorem to a small circle $L = \omega + \epsilon\, e^{i\theta}$ around the singularity at $\omega$, where the integral

$$\begin{aligned} \oint_L f(\omega')/(\omega' - \omega)\, d\omega' &\approx f(\omega) \oint_L 1/(\omega' - \omega)\, d\omega' \\ &= f(\omega) \log(\epsilon\, \exp(i\theta))|_0^{2\pi} = 2\pi i f(\omega) \end{aligned} \tag{10.68}$$

is well approximated by a log. Cauchy's theorem, on the other hand, is a miracle.

Notice the $i$ in the denominator. This implies that the real part of the integral gives the imaginary part of $\chi$ and vice versa. In particular,

$$\chi'(\omega) = Re[\chi(\omega)] = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{Im[\chi(\omega')]}{\omega' - \omega} d\omega' = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{\chi''(\omega')}{\omega' - \omega} d\omega'. \tag{10.71}$$

This is the *Kramer's Krönig relation.* It's traditional to simplify it a bit more, by noticing that $\chi''(\omega) = -\chi''(-\omega)$, so

$$\begin{aligned}\chi'(\omega) &= \frac{1}{\pi}\int_0^{\infty} \chi''(\omega') \left(\frac{1}{\omega' - \omega} - \frac{1}{-\omega' - \omega}\right) d\omega' \\ &= \frac{2}{\pi}\int_0^{\infty} \chi''(\omega') \frac{\omega'}{\omega'^2 - \omega^2} d\omega'. \end{aligned} \tag{10.72}$$

Hence in principle one can measure the imaginary, dissipative part of a frequency–dependent susceptibility and do a simple integral to get the real, reactive part. In practice this is a challenge, since one must know (or approximate) the imaginary part at all frequencies, from deep infrared to X-ray.

# Exercises

(10.1) **Fluctuations in Damped Oscillators.**

Let us explore further the fluctuating mass-on-a-spring example of section 6.5. We saw there that the coupling of the macroscopic motion to the internal degrees of freedom will eventually damp the oscillations: the remaining motions were microscopic for macroscopic springs and masses. These remanent thermal fluctuations can be important, however, for nanomechanical and biological systems. In addition, the damped harmonic oscillator is a classic model for many other physical processes, such as dielectric loss in insulators. (See [70] for a nice treatment by an originator of this subject.)

Consider a damped, simple harmonic oscillator, forced with an external force $f$, obeying the equation of motion

$$\frac{d^2\theta}{dt^2} = -\omega_0^2\theta - \gamma\frac{d\theta}{dt} + f(t)/m. \tag{10.73}$$

*(a)* **Susceptibility.** *Find the AC susceptibility $\widetilde{\chi}(\omega)$ for the oscillator. Plot $\chi'$ and $\chi''$ for $\omega_0 = m = 1$ and $\gamma = 0.2$, 2, and 5.* (Hint: Fourier transform the equation of motion, and solve for $\widetilde{\theta}$ in terms of $\widetilde{f}$.)

*(b)* **Causality and Critical Damping.** *Check, for positive damping $\gamma$, that your $\chi(\omega)$ is causal ($\chi(t) = 0$ for $t < 0$), by examining the singularities in the complex $\omega$ plane (section 10.6). At what value of $\gamma$ do the poles begin to sit on the imaginary axis?* The system is overdamped, and the oscillations disappear, when the poles are on the imaginary axis.

At this point, it would be natural to ask you to verify the Kramers-Krönig relation, equation 10.72, and show explicitly that you can write $\chi'$ in terms of $\chi''$. That turns out to be tricky both analytically and numerically, though: if you're ambitious, try it.

*(c)* **Dissipation and Susceptibility.** *Given a forcing $f(t) = A\cos(\omega t)$, solve the equation and calculate $\theta(t)$. Calculate the average power dissipated by integrating your resulting formula for $f\, d\theta/dt$. Do your answers for the power and $\chi''$ agree with the general formula for power dissipation, equation 10.35?*

*(d)* **Correlations and Thermal Equilibrium.** *Use the fluctuation–dissipation theorem to calculate the correlation function $\tilde{C}(\omega)$ from $\chi''(\omega)$, (see equation 10.64), where*

$$C(t - t') = \langle\theta(t)\theta(t')\rangle. \tag{10.74}$$

*Find the equal-time correlation function $C(0) = \langle\theta^2\rangle$, and show that it satisfies the equipartition theorem.* (Hints: our oscillator is in a potential well $V(\theta) = \frac{1}{2}m\omega_0^2\theta^2$. You will need to know the integral $\int_{-\infty}^{\infty} \frac{1}{\omega^2 + (1-\omega^2)^2}\, d\omega = \pi$.)

(10.2) **Telegraph Noise and RNA Unfolding.** (Biology)

RNA, ribonucleic acid, is a long polymer like DNA, with many functions in living cells. It has four monomer units (A, U, C, and G: DNA has T instead of U[24]) The sequence of monomers can encode information for building proteins, and can also cause the RNA to fold into shapes that are important to its function.

Michelle Wang's group here studies the strength of these hairpins by pulling on them (see data from a different group, above). Under a sufficiently large force, the hairpin will unzip. Near the threshold for unzipping, the RNA is found to jump between the zipped and unzipped states, giving *telegraph noise*, figure 10.14. Just as the current in a telegraph signal is either on or off, these systems are bistable and make transitions from one state to the other. When these systems are in equilibrium, they provide a prototype statistical mechanics problem: a system in a double well. (See exercise 6.2.)

**Fig. 10.12 Hairpins in RNA,** from reference [64]. A length of RNA attaches to an inverted, complementary strand immediately following.

The two RNA configurations presumably have different energies $(E_z, E_u)$ entropies $(S_z, S_u)$ and volumes $(V_z, V_u)$ for the local region around the zipped and unzipped states, respectively. The environment is at temperature $T$ and pressure $P$. Let $\rho_z$ be the fraction of the time our molecule is zipped at a given external force $F$, and $\rho_u = 1 - \rho_z$ be the unzipped fraction of time.


**Fig. 10.13 Experimental setup** from reference [64], used to apply a force $F$ to the RNA: the two ends are attached, via complementary DNA strands, to two beads which are then pulled apart. The RNA has two metastable states over a range of forces: the hairpin, zipped configuration shown and an unzipped configuration. Let the difference in length between the zipped and unzipped configuration be $L$.

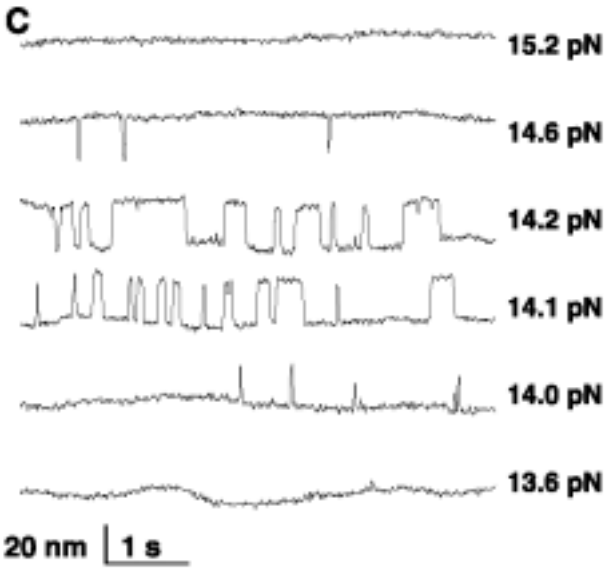



**Fig. 10.14 Telegraph Noise in RNA unzipping,** from reference [64]. As the force increases, the fraction of time spent in the zipped state decreases.

*Of the following statements, which are true, assuming that the pulled RNA is in equilibrium?*

*(T) (F)* $\rho_z/\rho_u = \exp((S_z^{tot} - S_u^{tot})/k_B)$*, where* $S_z^{tot}$ *and* $S_u^{tot}$ *are the total entropy of the universe when the RNA is in the zipped and unzipped states, respectively.*

*(T) (F)* $\rho_z/\rho_u = \exp(-(E_z - E_u)/k_B T)$.

*(T) (F)* $\rho_z/\rho_u = \exp(-(G_z - G_u)/k_B T)$*, where* $G_z = E_z - TS_z + PV_z$ *and* $G_u = E_u - TS_u + PV_u$ *are the Gibbs energies in the two states.*

[24] Adenine, Uracil, Cytosine, and Guanine; DNA has Thymine [80].

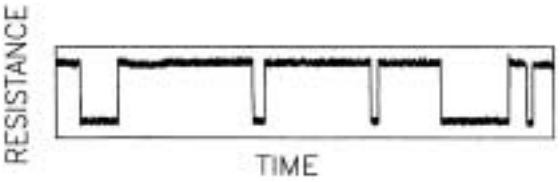



**Fig. 10.15 Telegraph Noise in a metallic nanobridge.** Resistance versus time $R(t)$ for a copper constriction, from reference [87]. We label $\alpha$ the state with low resistance $R_\alpha$, and $\beta$ the state with high resistance $R_\beta$. The two states probably represent a local shift of an atom or a small group of atoms in the constriction from one metastable state to another.

*(T) (F)* $\rho_z/\rho_u = \exp(-(G_z - G_u + FL)/k_BT)$*, where* $L$ *is the extra length of the unzipped RNA and* $F$ *is the applied force.*

(10.3) **Telegraph Noise in Nanojunctions.**

Many systems in physics exhibit *telegraph noise*.

The junction has two states, $\alpha$ and $\beta$. It makes transitions at random, with rate $P_{\beta\leftarrow\alpha} = P_{\beta\alpha}$ from $\alpha$ to $\beta$ and rate $P_{\alpha\beta}$ from $\beta$ to $\alpha$.

**Master Equation.** Consider an ensemble of many identical copies of this system. Let the state of this ensemble at time $t$ be given by $\vec{\rho}(t) = (\rho_\alpha, \rho_\beta)$, a vector of probabilities that the system is in the two states. This vector thus evolves according to the *master equation*

$$d\vec{\rho}/dt = M \cdot \vec{\rho}. \tag{10.75}$$

*(a) What is the* $2 \times 2$ *matrix* $M$ *for our system? (It's different from the discrete-time case we studied earlier. Notice under* $P_{\alpha\beta}$*,* $\rho_\alpha$ *increases and* $\rho_\beta$ *decreases.) At long times, what fraction of the time will our system be in the* $\alpha$ *state,* $\langle\rho_\alpha\rangle = \lim_{t\to\infty} \rho_\alpha(t)$*?*

*(b) Find the eigenvalue–eigenvector pairs for* $M$*.*[25] *Which corresponds to the stationary state* $\vec{\rho}(\infty)$ *from part (a)? Suppose that at* $t = 0$ *the system is known to be in the* $\alpha$ *state,* $\vec{\rho}(0) = (1, 0)$*. Write this initial condition in the basis of eigenvectors, and hence give a formula for the subsequent time evolution* $\rho_\alpha(t)$*. What is the rate of decay to the stationary state?*

Let's call your answer for $\rho_\alpha(t) = P_{\alpha\alpha}(t)$ to emphasize the fact that it is the probability of being in the $\alpha$ state at time $t' + t$, given that it is in the $\alpha$ state at time $t'$. You may wish to check that $P_{\alpha\alpha}(0) = 1$, and that $P_{\alpha\alpha}(t) \to \langle\rho_\alpha\rangle$ as $t \to \infty$.

*(c)* **Correlation function.** *Let* $R(t)$ *be the resistance as a function of time, hopping between* $R_\alpha$ *and* $R_\beta$*, as shown in figure 10.15, and let* $\bar{R}$ *be the time average of the resistance. Write a formula for the connected correlation function* $C(t) = \langle(R(t') - \bar{R})(R(t'+t) - \bar{R})\rangle$ *in terms of* $P_{\alpha\alpha}(t)$*? For ease of grading, please do not plug in your answer for the latter from part (b).* (Hint: what is $\langle(R(t')-R_\beta)(R(t'+t)-R_\beta)\rangle$, in terms of $P_{\alpha\alpha}(t)$? What is it in terms of $C(t)$?)

You may wish to check that your $C(t) \to 0$ as $t \to \infty$, and that $C(0) = \langle\rho_\alpha\rangle(R_\alpha - \bar{R})^2 + \langle\rho_\beta\rangle(R_\beta - \bar{R})^2$.

Nanojunctions, especially at higher temperatures, often show more than two metastable states in the experimental bandwidth (fluctuating neither too fast or too slowly to measure). Usually these form independent two-level fluctuators (atomic rearrangements too far apart to interact substantially), but sometimes more complex behavior is seen. Figure 10.16 shows three resistance states, which we label $\alpha$, $\beta$, and $\gamma$ from lowest resistance to highest. We notice from figure 10.16 that the rates $P_{\gamma\beta}$ and $P_{\beta\gamma}$ are the highest, followed by the rates $P_{\alpha\gamma}$ and $P_{\gamma\alpha}$. There are no transitions seen going between states $\alpha$ and $\beta$, at least as far as I can see.

There is a large current flowing through the nanojunction, allowing the resistance to be measured. Whether these transitions are equilibrium fluctuations, perhaps with a field–dependent effective temperature, or whether they are non–equilibrium transitions induced by the external field, could be tested if these last two rates could be measured. If detailed balance is violated, the system is out of equilibrium.

*(d)* **Detailed Balance.** *What is the ratio between the two rates* $P_{\alpha\beta}$ *and* $P_{\beta\alpha}$*, assuming that the system satisfies detailed balance, in terms of these four rates?* (Hint: see exercise 8.4.)

(10.4) **Coarse-Grained Magnetic Dynamics.** [26]

[25] More specifically, the right eigenvectors $M \cdot \vec{\rho}_\lambda = \lambda\vec{\rho}_\lambda$

[26] Math reference: [71, sec. 4.1, ch. 12].

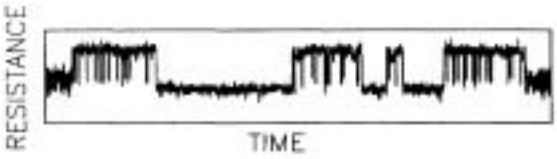



**Fig. 10.16 Telegraph Noise in a metallic nanobridge with three metastable states.** From reference [86].

A one-dimensional classical magnet above its critical point is described by a free energy density

$$\mathcal{F}[M] = (C/2)(\nabla M)^2 + (B/2)M^2 \tag{10.76}$$

where $M(x)$ is the variation of the magnetization with position along the single coordinate $x$. The average magnetization is zero, and the total free energy of the configuration $M(x)$ is $F[M] = \int \mathcal{F}[M]dx$.

The methods we used to find the correlation functions and susceptibilities for the diffusion equation can be applied with small modifications to this (mathematically more challenging) magnetic system.

Assume for simplicity that the magnet is of length $L$, and that it has periodic boundary conditions. We can then write $M(x)$ in a Fourier series (equation A.2)

$$M(x) = \sum_{n=-\infty}^{\infty} \widetilde{M}_n \exp(-ik_n x) \tag{10.77}$$

with $k_n = 2\pi n/L$ and (equation A.1)

$$\widetilde{M}_n = (1/L)\int_0^L M(x)\exp(ik_n x). \tag{10.78}$$

*(a) Show that (as always, for linear systems with translation invariance) the free energy $F[M]$, when written in terms of $\widetilde{M}_n$, becomes an independent sum over modes, with a quadratic energy in each mode.*[27] (Hint: The only subtle case is $\widetilde{M}_n^2$: break it into real and imaginary parts.) *What is $\left\langle |\widetilde{M}_n|^2 \right\rangle_{\rm eq}$, by equipartition? Argue that*

$$\left\langle \widetilde{M}_{-m}\widetilde{M}_n \right\rangle_{\rm eq} = \frac{k_B T}{L(Ck_n^2 + B)}\delta_{mn} \tag{10.79}$$

*(b) Calculate the equal-time correlation function for the magnetization in equilibrium, $C(x,0) = \langle M(x,0)M(0,0)\rangle_{\rm eq}$.* (First, find the formula for the magnet of length $L$, in terms of a sum over $n$. Then convert the sum to an integral: $\int dk \leftrightarrow \sum_k \delta k = 2\pi/L\sum_k$.) You'll want to know the integral

$$\int_{-\infty}^{\infty} e^{iuv}/(1+a^2u^2)\,du = (\pi/a)\exp(-|v|/a). \tag{10.80}$$

Assume the magnetic order parameter is not conserved, and is overdamped, so the time derivative of $[M]_{M_{\rm i}}$ is given by the mobility $\gamma$ times the variational derivative of the free energy:

$$\frac{\partial [M]_{M_{\rm i}}}{\partial t} = -\gamma\frac{\delta \mathcal{F}}{\delta M}. \tag{10.81}$$

$M$ evolves in the direction of the total force on it.[28] The term $\delta\mathcal{F}/\delta M$ is the *variational derivative:*[29]

$$\begin{aligned}\delta F &= F[M+\delta M] - F[M] \\ &= \int \mathcal{F}[M+\delta M] - \mathcal{F}[M]\,dx \\ &= \int (\delta\mathcal{F}/\delta M)\delta M\,dx.\end{aligned} \tag{10.82}$$

*(c) Calculate $\delta\mathcal{F}/\delta M$.* As in the derivation of the Euler-Lagrange equations, you'll need to integrate one term by parts to factor out the $\delta M$.

*(d) From your answer to part (c), calculate the Greens function for $[M]_{M_{\rm i}}$, $G(x,t)$ giving the time evolution of an initial condition $M_{\rm i}(x) = M(x,0) = G(x,0) = \delta(x)$.* (Hint: You can solve this with Fourier transforms.)

The Onsager regression hypothesis tells us that the time evolution of a spontaneous fluctuation (like those giving $C(x,0)$ in part (b)) is given by the same formula as the

[27] Notice that this implies that the phases of the different modes are uncorrelated in equilibrium.

[28] Remember that the average $[M]_{M_{\rm i}}$ is over all future evolutions given the initial condition $M_{\rm i}(x) = M(x,0)$.

[29] This formula is analogous to taking the gradient of a scalar function of a vector, $f(\vec{y}+\vec{\delta}) - f(\vec{y}) \approx \nabla f \cdot \vec{\delta}$, with the dot product for vector gradient replaced by the integral over $x$ for derivative in function space.

evolution of an imposed initial condition (given by the Greens function of part (d)).

$$\begin{aligned} C(x,t) &= \langle M(x,t)M(0,0)\rangle_{\rm ev} && (10.83)\\ &= \Big\langle [M(x,t)]_{M_{\rm i}}\, M(0,0)\Big\rangle_{\rm eq} && (10.84)\\ &= \left\langle \int_{-\infty}^{\infty} M(x',0)G(x-x',t)\,dx'\, M(0,0)\right\rangle_{\rm eq}\\ &= \int_{-\infty}^{\infty} C(x',0)G(x-x',t)\,dx'. \end{aligned}$$

*(e) Using the Onsager regression hypothesis calculate the space-time correlation function* $C(x,t) = \langle M(x,t)M(0,0)\rangle_{\rm ev}$. *(This part is a challenge: your answer will involve the error function.) If it's convenient, plot it for short times and for long times: does it look like* $\exp(-|y|)$ *in one limit and* $\exp(-y^2)$ *in another?*

The fluctuation–dissipation theorem can be used to relate the susceptibility $\chi(x,t)$ to the time dependent impulse to the correlation function $C(x,t)$ (equation 10.60). Let $\chi(x,t)$ represent the usual response of $M$ to an external field $H(x',t')$ (equation 10.27) with the interaction energy being given by $\int M(x)H(x)\,dx$.

*(f) Calculate the susceptibility* $\chi(x,t)$ *from* $C(x,t)$. *Start by giving the abstract formula, and then plug in your answer from part (e).*

(10.5) **Noise and Langevin equations.**

In this chapter, we have never explicitly discussed how the energy removed from a system by damping is returned to the system to maintain thermal equilibrium. This energy input is through the thermal fluctuation noise introduced through the coupling to the heat bath. In this exercise we will derive a *Langevin equation* incorporating both noise and dissipation (see also [20, section 8.8]).

We start with a system with coordinate $\mathbb{P},\mathbb{Q}$ and internal potential energy $V(\mathbb{Q})$, coupled linearly to a heat bath through some coupling term $\mathbb{Q}\cdot\mathbb{F}$:

$$\mathcal{H} = \frac{\mathbb{P}^2}{2m} + V(\mathbb{Q}) + \mathcal{H}_{\rm bath}(y_1,y_2,y_3,\dots) - \mathbb{Q}\cdot\mathbb{F}(y_1,\dots). \tag{10.85}$$

In the absence of the coupling to our system, assume that the bath would contribute an external noise $\mathbb{F}_b(t)$ with mean zero. In the presence of the coupling to the system, the mean value of the force will develop a non-zero expectation value

$$\langle \mathbb{F}(t)\rangle = \int_{-\infty}^{t} dt'\,\chi_b(t-t')\mathbb{Q}(t'), \tag{10.86}$$

where $\chi_b(t-t')$ is the susceptibility of the bath to the motion of the system $\mathbb{Q}(t)$. Our system then has an equation of motion with a random noise $\mathbb{F}$ and a time-retarded interaction due to $\chi_b$:

$$m\ddot{\mathbb{Q}} = -\frac{\partial V}{\partial \mathbb{Q}} + \mathbb{F}_b + \int_{-\infty}^{t} dt'\,\chi_b(t-t')\mathbb{Q}(t'). \tag{10.87}$$

We can write this susceptibility in terms correlation function of the noise in the absence of the system

$$C_b(t-t') = \langle \mathbb{F}_b(t)\mathbb{F}_b(t')\rangle \tag{10.88}$$

using the fluctuation-dissipation theorem

$$\chi_b(t-t') = -\beta\frac{\partial C_b}{\partial t}. \qquad t>t'. \tag{10.89}$$

*(a) Integrating by parts and keeping the boundary terms, show that the equation of motion has the form*

$$m\ddot{\mathbb{Q}} = -\frac{\partial \bar{V}}{\partial \mathbb{Q}} + \mathbb{F}_b - \beta\int_{-\infty}^{t} dt'\,C_b(t-t')\dot{\mathbb{Q}}(t'). \tag{10.90}$$

*What is the 'potential of mean force'* $\bar{V}$, *in terms of* $V$ *and* $C_b$*?*

*(b) If the correlations in the bath are short-lived compared to the time-scales of the system, we can approximate* $\dot{\mathbb{Q}}(t') \approx \dot{\mathbb{Q}}(t)$ *in equation 10.90, leading to a viscous friction force* $-\gamma\dot{\mathbb{Q}}$. *What is the formula for* $\gamma$*?*

*(c) Conversely, for a model system with a perfect viscous friction law* $-\gamma\dot{\mathbb{Q}}$ *at temperature* $T$, *derive the equation for correlation function for the noise* $C_b(t-t')$. Notice that viscous friction implies a memoryless, Markovian heat bath, and vice-versa.

Langevin equations are useful both in analytic calculations, and as one method for maintaining a constant temperature in molecular dynamics simulations.

(10.6) **Fluctuations, Correlations, and Response: Ising**

This exercise again uses the program **ising**, downloadable from the Web [105].

This time we'll consider the Ising Hamiltonian in a time-dependent external field $H(t)$,

$$\mathcal{H} = -J\sum_{\langle ij\rangle} S_iS_j - H(t)\sum_i S_i, \tag{10.91}$$

and look at the fluctuations and response of the time-dependent magnetization $M(t) = \sum_i S_i(t)$. Our Ising model simulation outputs both the time-dependent magnetization per spin $m(t) = (1/N)\sum_i S_i$ and the time-time correlation function of the magnetization per spin,

$$c(t) = \Big\langle (\mathbf{m}(0) - \langle\mathbf{m}\rangle_{\rm eq})(\mathbf{m}(t) - \langle\mathbf{m}\rangle_{\rm eq})\Big\rangle_{\rm ev}. \tag{10.92}$$

(select *autocorrelation* in the menu above the plot). We'll be working above $T_c$, so $\langle \mathbf{m} \rangle_{\text{eq}} = 0$. Note, as before, that most of our formulas in class are in terms of the total magnetization $M = Nm$ and its correlation function $C = N^2 c$.

The time-time correlation function will start non-zero, and should die to zero over time. Suppose we start with a non-zero small external field, and turn it off at $t = 0$: $H(t) = H_0 \Theta(-t)$? The magnetization $M(t)$ will be non-zero at $t = 0$, but will decrease to zero over time. By the Onsager regression hypothesis, these two time decays should look the same.

Run the Ising model, changing the size to $200 \times 200$. Equilibrate at $T = 3$ and $H = 0$, *reset*, do a good measurement of the time-time autocorrelation function, and *copy graph*. Rescale the graph using *configure* to focus on the short times where things are interesting. Now equilibrate at $T = 3$, $H = 0.05$, set $H = 0$ and *reset*, and run for a short time.

*(a) Does the shape and the time-scale of the magnetization decay look the same as that of the autocorrelation function? Note down the values for $c(0)$, $C(0)$, $m(0)$, and $M(0)$.*

**Response Functions and the Fluctuation-Dissipation Theorem.** The response function $\chi(t)$ gives the change in magnetization due to an infinitesimal impulse in the external field $H$. By superposition, we can use $\chi(t)$ to generate the linear response to any external perturbation. If we impose a small time-dependent external field $H(t)$, the average magnetization

$$\mathbf{M}(t) - \langle \mathbf{M} \rangle_{\text{eq}} = \int_{-\infty}^{t} dt' \chi(t - t') H(t'), \qquad (10.93)$$

where $\langle \mathbf{M} \rangle_{\text{eq}}$ is the equilibrium magnetization without the extra field $H(t)$ (zero for us, above $T_c$).

*(b) Using equation 10.93, write $M(t)$ for the step down $H(t) = H_0 \Theta(-t)$, in terms of $\chi(t)$.*

The fluctuation-dissipation theorem states that this response

$$\chi(t) = -\beta dC(t)/dt. \qquad (10.94)$$

where $C(t) = \Big\langle (\mathbf{M}(0) - \langle \mathbf{M} \rangle_{\text{eq}})(\mathbf{M}(t) - \langle \mathbf{M} \rangle_{\text{eq}}) \Big\rangle_{\text{ev}}$ is the time-time correlation function for the total magnetization.

*(c) Use equation 10.94 and your answer to part (b) to predict the relationship between the demagnetization $M(t)$ and the correlation $C(t)$ you measured in part (a). How does your analytical ratio compare with the $t = 0$ ratio you noted down in part (a)?*

(10.7) **Spin Correlation Functions and Susceptibilities.** (From Halperin. [43])

A spin in a solid has two states, $s_z = \pm 1/2$ with magnetizations $M = g\mu_B s_z$, where $g\mu_B$ is a constant[30] Due to fluctuations in its environment, the spin can flip back and forth thermally between these two orientations. In an external magnetic field $H$, the spin has energy $-M \cdot H$. Let $M(t)$ be the magnetization of the spin at time $t$. Given a time-dependent small external field $H(t)$ along $z$, the expectation value of $M$ satisfies

$$d\left[M(t)\right]_{M_{\text{i}}}/dt = -\Gamma \left[M(t)\right]_{M_{\text{i}}} + \Gamma \chi_0 H(t) \qquad (10.95)$$

where $\Gamma$ is the spin equilibration rate and $\chi_0$ is the static magnetic susceptibility, and the averaging $[\,]_{M_{\text{i}}}$ is over the noise provided by the environment, fixing the initial condition $M_{\text{i}} = M(0)$.

*(a) In the case that the field $H$ is time independent, you can use equilibrium statistical mechanics to determine $M(H)$. Using this formula for small $H$, determine $\chi_0$ (which should be independent of $H$ but dependent on temperature).*

*(b) Use the Onsager regression hypothesis to compute $C(t) = \langle M(t)M(0) \rangle_{\text{ev}}$ at zero external field $H = 0$. What should it be for times $t < 0$? What is $\widetilde{C}(\omega)$, the Fourier transform of $C(t)$?*

*(c) Assuming the classical fluctuation-dissipation theorem (classical if $\hbar\omega \ll T$), derive the frequency dependent susceptibility $\chi(t)$ and $\widetilde{\chi}(\omega)$.*

*(d) Compute the energy dissipated by the oscillator for an external magnetic field $H(t) = H_0 \cos(\omega t)$.*

[30] $g$ is the gyromagnetic ratio for the spin, and $\mu_B = e\hbar/2m_e$ is the Bohr magneton.

# Abrupt Phase Transitions

# 11

Water remains liquid as it is cooled, until at $0C$ it abruptly freezes into ice. Water remains liquid as it is heated, until at $100C$ it abruptly turns to vapor. The transitions from crystal to liquid, and liquid to gas are abrupt. Both transition temperatures depend on the pressure, leading to the familiar phase diagram 11.1.

Most phase transitions are abrupt. They occur with discontinuities in most physical properties: the density, compressibility, viscosity, specific heat, dielectric constant, and thermal conductivity all jump to new values. Furthermore, in most cases these transitions occur without any precursors, any hints that a change is about to occur: pure water does not first turn to slush and then to ice,[1] and water vapor at $101C$ has no little droplets of water inside.[2]

There is nothing abrupt, however, in boiling away a pan of water.[3] This is because one is not controlling the temperature directly, but rather is adding energy at a constant rate. Consider an insulated, flexible container of $H_2O$ at fixed pressure, as we slowly add energy to it. When the system first reaches the liquid–gas transition, a small bubble of gas will form at the top: this bubble will gradually grow, inflating and filling the container over a range of energies. The transition from liquid to gas at fixed energy passes through an intermediate *two–phase region*: the temperature of the system stays constant until the last liquid is gone. It's more typical to measure this two–phase region at fixed temperature as the volume is varied, yielding the phase diagram figure 11.2.

To avoid these two–phase mixtures, we choose to work in the variables $P$ and $T$, so we use the Gibbs free energy[4]

$$G(T, P, N) = E - TS + PV. \tag{11.1}$$

As usual (section 6.5), whichever state minimizes $G$ wins.[5] The Euler relation (exercise 6.4) $E = TS - PV + \mu N$ tells us that $G = \mu N$. So, the state with the lower chemical potential will be favored, and the phase transition will occur when $\mu_{\text{liq}} = \mu_{\text{gas}}$. That makes sense: at a liquid–vapor boundary, one can exchange energy, volume, or particles. Energy will exchange until the temperatures are equal, volume will exchange

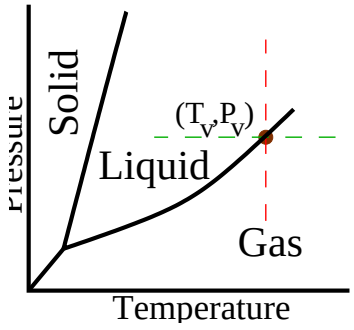



**Fig. 11.1 PT Phase Diagram**. Schematic phase diagram for a typical material. The liquid–gas line gives the vapor pressure $P_v(T)$ or the liquid–gas transition temperature $T_v(P)$. (Ice is (atypically) less dense than water, so the solid–liquid boundary for $H_2O$ will slope to the left.)

[1] Water with impurities like salt does form slush. Slush is formed by ice that is mostly pure $H_2O$ surrounded by rather salty water. The salt lowers the freezing point: as the ice freezes it expels the salt, lowering the freezing point of the remaining liquid.

[2] More precisely, there are only exponentially rare droplet fluctuations, which will be important for nucleation, section 11.2.

[3] The *latent heat* at a phase transition is the energy change between the two phases (or, more specifically, the change in enthalpy $E + PV$). The latent heat of boiling for water is 2500 J/gm = 600 calories/gm and the calorie is defined to make the specific heat of water one calorie per gram per C. Hence, at a constant energy input from your stove, it takes six times as long to boil away the water as it takes to raise it from freezing to boiling ((1 cal/gm C) × 100 C), which should seem about right to the cooks among the readers.

[4] We could also avoid the two–phase mixtures by using the grand free energy, $\Phi(T, V, N) = E - TS - \mu N$. The grand partition function allows the total number of particles to vary, so when the liquid turns to gas the molecules in the extra volume are simply removed.

[5] Minimizing $G$ for the system maximizes the entropy for the universe as a whole (the system plus the bath with which it exchanges energy and volume): see section 6.5.

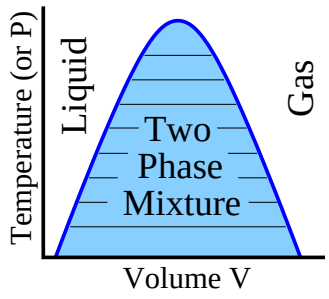



**Fig. 11.2 TV Phase Diagram.** A schematic liquid–gas phase diagram at fixed $N$, showing the two–phase region. The horizontal tie–lines represent the phase separation: a system with volume and temperature at a point in this region will separate into regions of liquid and gas on the two endpoints of the corresponding tie-line. See figure 12.7 for real experimental coexistence lines, and figure 8.4 for a similar phase diagram as a function of concentration.

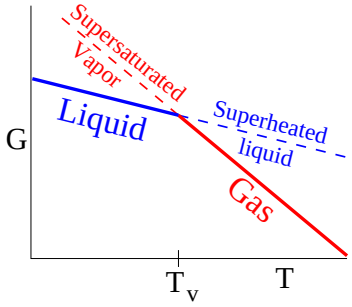



**Fig. 11.3 Gibbs free energy: liquid–gas.** The Gibbs free energy for the liquid and gas phases along an isobar (constant pressure, horizontal dashed line in figure 11.1). The phase transition occurs when the two curves cross.

[7] That's precisely what occurs when the relative humidity goes beyond 100%.

[8] The fact that the transition is abrupt is implemented in these 'mean field' theories by having two separate free energy minima: the transition arises when one minimum passes the other, but both exist before and after the transition.

[9] Just as the lifetime of a resonance in quantum mechanics is related to the imaginary part of its energy $E + i\hbar\Gamma$, so similarly is the rate per unit volume of nucleation of the new phase (section 11.2) related to the imaginary part of the free energy density.

until the pressures are equal, and then particles will move from liquid to gas until the chemical potentials are equal.

Remembering the shorthand thermodynamic relation $dE = TdS - PdV + \mu dN$, we find $dG = -SdT + VdP + \mu dN$. Varying the temperature at fixed pressure and number of particles, we thus learn that

$$\left.\frac{\partial G}{\partial T}\right|_{P,N} = -S. \tag{11.2}$$

Figure 11.3 shows the Gibbs free energy versus temperature for the liquid and gas. At the phase boundary, the two free energies agree. The difference in slopes of the two lines is given by the difference in entropies between the liquid and the gas (equation 11.2). The thermodynamic definition of the entropy $S = dQ/T$ (section 5.1) tells us that the entropy difference is given by the latent heat times the number of particles over the transition temperature,

$$\Delta S = L/T_v. \tag{11.3}$$

The fact that the Gibbs free energy has a kink at the phase transition reflects the jump in the entropy between liquid and gas – abrupt phase transitions will have jumps in the first derivatives of their free energies. This led early workers in the field to term these transitions *first–order transitions*.[6]

Notice that we continue to draw the free energy curves for the liquid and vapor on the 'wrong' sides of the phase boundary. It is a common experimental fact that one can supercool vapors significantly beyond their condensation point.[7] With careful experiments on pure liquids, one can also significantly superheat the liquids. Theoretically the issue is subtle. Some theories of these transitions have well–defined metastable phases of the "wrong" type.[8] However, there certainly is no *equilibrium* vapor state below $T_v$. More sophisticated approaches (not discussed here) give an imaginary part to the free energy density of the metastable phase (see [58, 16]).[9]

## 11.1 Maxwell Construction.

Figure 11.4 shows the pressure versus volume as we expand our material at constant temperature. The liquid turns metastable as the volume increases, when the pressure reaches the vapor pressure for that temperature. The gas becomes metastable at that same pressure when the volume decreases. The metastable states are well defined only near the vapor pressure, where nucleation is slow (section 11.2) and lifetimes are reasonably large. The dashed line shows a region which is completely

[6] We avoid using this term, and the analogous term *second–order* for continuous phase transitions. This is not only because their origin is obscure: but also because in the latter case it is misleading: the thermodynamic functions at a continuous phase transition have power–law singularities or essential singularities, not plain discontinuities in the second derivative (chapter 12).

unstable: a mixture of molecules prepared uniformly in space in this region will spontaneously separate into finely intertangled networks of the two phases[10] (section 11.3.1).

How did we know to draw the coexistence line at the pressure we chose in figure 11.4? How can we find the vapor pressure at which the liquid and gas coexist at this temperature? We know from the last section that the coexistence line occurs when the chemical potentials agree $\mu_{\rm liq} = \mu_{\rm gas}$, and hence when their Gibbs free energies agree $G_{\rm liq} = N\mu_{\rm liq} = N\mu_{\rm gas} = G_{\rm gas}$. Since $dG = -SdT + VdP + \mu dN$, at constant temperature and number $\frac{\partial G}{\partial P}\big|_{T,N} = V$. Hence, we know that

$$\Delta G = \int_{P_{\rm liq}}^{P_{\rm gas}} V(P)\, dP = 0. \tag{11.4}$$

Now this integral may seem trivial, because the limits of integration are both equal to the vapor pressure, $P_{\rm liq} = P_{\rm gas} = P_v(T)$. But this formula represents the sum of four pieces (figure 11.5):

$$\Delta G = \int_{P_{\rm liq}}^{P_{\rm min}} V(P)\, dP + \int_{P_{\rm min}}^{P_{\rm liq}} V(P)\, dP + \int_{P_{\rm gas}}^{P_{\rm max}} V(P)\, dP + \int_{P_{\rm max}}^{P_{\rm gas}} V(P)\, dP \tag{11.5}$$

remembering that $P_{\rm liq} = P_{\rm gas}$. Notice that the first and last terms are negative, since $P_{\rm liq} > P_{\rm min}$ and $P_{\rm max} > P_{\rm gas}$. These three integrals have a nice graphical interpretation, shown in figure 11.5: the first two subtract to give the area with stripes up and to the right and the last two subtract to give minus the area with stripes down and to the right. These two areas must agree at the vapor pressure. The transition pressure equalizes the areas in the PV diagram. This is the *Maxwell equal–area construction.*[11]

[10] This spontaneous separation is termed *spinodal decomposition.* In the past, the endpoints of the dashed curve were called *spinodal points*, but there is reason to doubt that there is any clear transition between nucleation and spontaneous separation, except in mean-field theories.

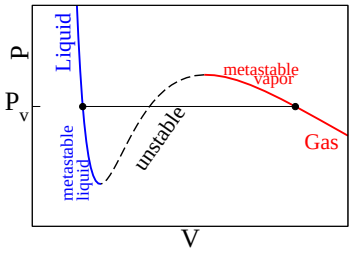



**Fig. 11.4** The pressure versus volume curve along an isotherm (dashed vertical line at constant temperature in figure 11.1). At low volumes the material is liquid: as the volume crosses into the two phase region in figure 11.2 the liquid becomes metastable. At high volumes the gas phase is stable, and again the metastable gas phase extends into the two–phase region. The dots represent the coexistence point where the pressures and chemical potentials of the two phases are equal.

## 11.2 Nucleation: Critical Droplet Theory.

On a humid night, as the temperature drops, the air may become supersaturated with water vapor. How does this metastable vapor turn into drops of dew (near the ground) or into the tiny water droplets that make up clouds (farther up)?

We've seen (figure 11.3) that the Gibbs free energy difference between the gas and the liquid grows as the temperature decreases below $T_v$. We can estimate the chemical potential difference driving the formation of drops: we know $\partial G/\partial T = -S$ (equation 11.2), and $\Delta S = L/T_v$ (equation 11.3), so

$$\begin{aligned}\Delta\mu = (G_{\rm gas} - G_{\rm liq})/N &= \left( \frac{\partial (G_{\rm gas} - G_{\rm liq})}{\partial T}\bigg|_{P,N} \Delta T \right) \Big/ N \\ &= \Delta S \Delta T/N = (L/T_v)(\Delta T/N) = \ell \Delta T/T_v \end{aligned} \tag{11.6}$$

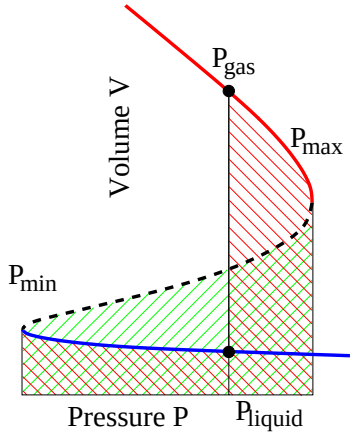



**Fig. 11.5 Maxwell Equal Area Construction.** The Gibbs free energy difference between two points at equal pressures is given by the difference in areas scooped out in the PV curves shown (upward diagonal area minus downward diagonal area).

[11] The Maxwell construction only makes sense, one must remember, for theories like mean-field theories where one has an unstable branch for the $P(V)$ curve.

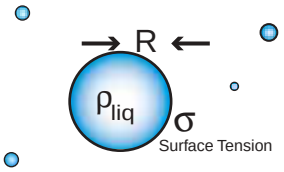



**Fig. 11.6** The nucleation of a new phase happens through a rare thermal fluctuation, where a droplet of the new phase forms of sufficient size to get it over the free energy barrier of figure 11.7.

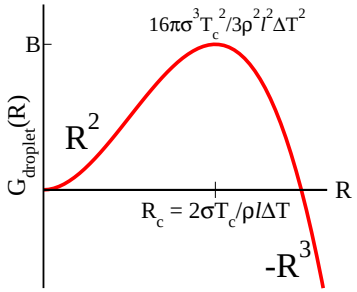



**Fig. 11.7** Free energy barrier for droplet formation.

where $\ell$ is the latent heat per particle.

What is the obstacle impeding the formation of the droplets? It is the *surface tension* $\sigma$ between the liquid and the gas phase. The surface tension is the Gibbs free energy per unit area of the interface between liquid and gas. Like tension in a rope, the interface between two phases can exert a force, pulling inward to minimize its area.[12]

To make a large droplet you must grow it from zero radius (figure 11.6). Since cost of the surface tension cost grows with the surface area $A$ and the bulk free energy gain grows as the volume $V$, tiny droplets will cost the system more than they gain. Consider the energy of a spherical droplet of radius $R$. The surface Gibbs free energy is $\sigma A$. If the liquid has $\rho_{\rm liq}$ particles per unit volume, and each particle provides a Gibbs free energy gain of $\Delta\mu = \ell\Delta T/T_v$, the bulk free energy gain is $V\rho_{\rm liq}\Delta\mu$. Hence

$$G_{\rm droplet}(R) = \sigma A - V\rho_{\rm liq}\Delta\mu = 4\pi R^2\sigma - (\tfrac{4}{3}\pi R^3)\rho_{\rm liq}(\ell\Delta T/T_v). \quad (11.7)$$

This free energy is plotted in figure 11.7. Notice that at small $R$ where surface tension dominates it rises quadratically, and at large $R$ where the bulk chemical potential difference dominates it drops as the cube of $R$. The gas will stay a gas until a rare thermal fluctuation pays the free energy cost to reach the top of the barrier, making a *critical droplet*. The critical droplet radius $R_c$ and the free energy barrier $B$ are found by finding the maximum of $G(R)$:

$$\left.\frac{\partial G_{\rm droplet}}{\partial R}\right|_{R_c} = 8\pi\sigma R_c - 4\pi\rho_{\rm liq}(\ell\Delta T/T_v)R_c^2 = 0 \quad (11.8)$$

$$R_c = \frac{2\sigma T_v}{\rho_{\rm liq}\ell\Delta T} \quad (11.9)$$

$$B = \frac{16\pi\sigma^3 T_v^2}{3\rho_{\rm liq}^2\ell^2}\frac{1}{(\Delta T)^2}. \quad (11.10)$$

The probability of finding a critical droplet per unit volume is given by $\exp(-B/k_BT)$ times a prefactor. The nucleation rate per unit volume is the net flux of droplets passing by $R_c$, which is the velocity over the barrier times a correction for droplets recrossing, times the probability of being on top of the barrier.[13] The prefactors are important for detailed

[13] See exercise 6.2 for the similar problem of calculating chemical reaction rates.

[12] How do we define the surface tension? Suppose we treat the two side boundaries of a constant pressure system (in zero gravity) so that one (like wax) prefers the gas and the other prefers the liquid phase. (We'll have periodic boundary conditions in the other two directions, or use a large, thin container.) The equilibrium state will then be forced to create an interface somewhere between, of area $A$ given by the cross-sectional area of our container. We can compare its Gibbs free energy $G_{lg}$ with that of system prepared with two liquid–loving surfaces $G_{ll}$ and two gas–loving surfaces $G_{gg}$. The surface tension is then $\sigma = \frac{G_{lg} - \frac{1}{2}(G_{ll}+G_{gg})}{A}$, to cancel out the interaction free energies with the side-walls. It doesn't matter where the interface lies, since the bulk free energies per particle agree in the two phases $\mu_{\rm liq} = \mu_{\rm gas}$

For curved interfaces, there will be a pressure difference between the two phases, so we need to use the grand potential $\Phi(T, V, \mu)$ instead of the Gibbs free energy $G(T, P, N)$ (see note 4 on page 219). There are also subtle questions about how to define the location of curved interfaces [57, Chapter XV].

theories, but the main experimental dependence is the exponentially small probability for having a critical droplet: our net droplet nucleation rate per unit volume $\Gamma$ has the form

$$\Gamma = (\text{Prefactors}) e^{-B/k_B T}. \tag{11.11}$$

Notice:

- The critical droplet radius $R_c \propto 1/\Delta T$: if you undercool the gas only a tiny amount, you need a big droplet to overcome the surface tension.
- The barrier height $B \propto 1/(\Delta T)^2$. The energy barrier for nucleation diverges at $T_v$.
- The droplet nucleation rate $\Gamma \propto \exp(-C/(\Delta T)^2)$. It can be really tiny for small undercoolings.[14]

The rates we have calculated are for *homogeneous nucleation*: the rate of forming the new phase in a perfectly clean system without boundaries. In practice, this is rarely the situation. Because nucleation is so strongly suppressed by the surface tension, the system will almost always be able to find a way of bypassing at least part of the energy barrier. That's why dewdrops form on grass (or your windshield), rather than always forming in the air and dropping to the ground: the surface tension between water and grass is much lower than that between water and air, so a roughly hemispherical droplet can form – dividing the free energy barrier $B$ in two. In the upper atmosphere, nucleation occurs on small dust particles – again, lowering the interfacial area needed to get a droplet of a given curvature.[15]

Finally, we should mention that the nucleation of crystalline phases won't proceed with precisely spherical droplets. Because crystals have anisotropic surface tension, the maximum number of particles for a given surface energy is given not by a sphere, but by the *equilibrium crystal shape* (the same shape that a crystal will form in equilibrium at constant number of particles, figure 11.8).

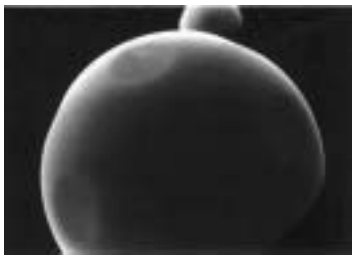


**Fig. 11.8** Equilibrium shape of a lead crystal at about 300 C [90, 98]. Notice the flat facets, which correspond to low–energy high–symmetry surfaces. There are three interesting statistical mechanics problems associated with these facets that we will not discuss in detail. (1) The crystalline orientations with flat facets are in a different phase than the rounded regions: they are below their *roughening transition*. (2) The equilibrium crystal shape, which minimizes the free energy, can be viewed as a Legendre transform of that free energy. (3) At lower temperatures the entire equilibrium crystal shape can be faceted (below the *edge* and *corner rounding* transitions): we predict that the coarsening length will grow only logarithmically in time in this phase (figure 11.14).

[14] The decay rate has an essential singularity at $T_v$: it is zero to all orders in perturbation theory in $\Delta T$. In some ways, this is why one can study the metastable states – perturbation theory naturally ignores the fact that they are unstable.

[15] Actually, in many clouds the temperature is low enough that ice crystals nucleate, rather than water droplets. Certain plant pathogens (*Pseudomonas syringae*) make proteins that are designed to efficiently nucleate ice crystals: the bacteria use the frost damage on the plants to invade. Humans use these proteins in snow–making machines at ski resorts.

## 11.3 Morphology of abrupt transitions.

What happens after the phase transition is nucleated (or when the undercooling is so large that the transition occurs immediately)? This question leads us into a gigantic, rich subject that mostly belongs to geology, engineering, and materials science rather than to statistical mechanics. We will give a brief introduction, with emphasis on topics where statistical mechanics is useful.

### 11.3.1 Coarsening.

What do salad dressing, cast iron, and rocks have in common? *Coarsening* is crucial to all three. When you shake up the salad dressing, the oil and vinegar get jumbled together in small droplets. When you stop

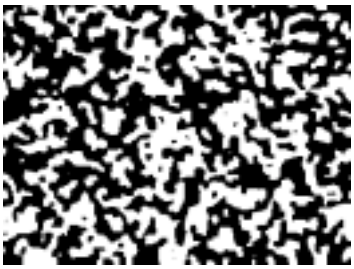


**Fig. 11.9** The spin configuration of the Ising model at $T = 0$ with non-conserved dynamics, after roughly a time of twenty sweeps through the lattice.

**Fig. 11.10** The same model as figure 11.9, after roughly 200 sweeps. Notice the characteristic morphologies look similar, except that the later picture has a length scale roughly three times larger ($\sqrt{10} \approx 3$).

shaking, the tiny droplets merge into bigger ones, gradually making for a coarser and coarser mixture until all the oil is on the top.

Molten iron, before it is cast, has a fair percentage of carbon dissolved in it. As it cools, this dissolved carbon precipitates out (with many nuclei forming as in section 11.2 and then growing until the carbon runs out), staying dispersed through the iron in particles whose size and number depend on the cooling schedule. The hardness and brittleness of cast iron depends on the properties of these carbon particles, and thus depends on how the iron is cooled.

Rocks often have lots of tiny grains of different materials: quartz, alkali feldspar, and plagioclase in granite; plagioclase feldspar and calcium-rich pyroxene in basalt, ... Different rocks have different sizes of these grains. Rocks formed from lava of erupting volcanoes have very fine grains: rocks deep underground cooled from magma over eons form large grains. For a particular grain to grow, the constituent atoms must diffuse through neighboring grains of other materials – a process that gets very slow as the grains get larger. Polycrystals form also from cooling single materials: different liquid regions will nucleate crystalline grains in different orientations, which then will grow and mush together. Here the grains can grow by stealing one another's molecules, rather than waiting for their brand of molecules to come from afar.

One can also see coarsening on the computer. Figures 11.9 and 11.10 show two snapshots of the Ising model, quenched to zero temperature, at times differing by roughly a factor of ten. Notice that the patterns of up and down spins look statistically similar in the two figures, except that the overall length scale $L(t)$ has grown larger at later times. This is the characteristic feature that underlies all theories of coarsening: the system is statistically similar to itself at a later time, except for a time–dependent length scale $L(t)$.

The basic results about coarsening can be derived by arguments that are almost simplistic. Consider a snapshot of a coarsening system (figure 11.11). In this snapshot, most features observed have a characteristic length scale $R \sim L(t)$. The coarsening process involves the smaller features shrinking to zero, so as to leave behind only features on larger scales. Thus we need to understand how long it takes for spheres and protrusions on a scale $R$ to vanish, to derive how the size $L(t)$ of the remaining features grows with time. $L(t)$ is the size $R_0$ of the smallest original feature that has not yet shrunk to zero.

The driving force behind coarsening is surface tension: the system can lower its free energy by lowering the interfacial area between the different domains. We'll focus on the evolution of a sphere as a solvable case. The surface tension energy for a sphere of radius $R$ is $\mathcal{F}_{\mathbf{surface}} = 4\pi R^2\sigma$, so there is an inward force per unit area, (or *traction*) $\tau$:

$$\tau = \frac{\partial \mathcal{F}_{\mathbf{surface}}}{\partial R}/(4\pi R^2) = 2\sigma/R. \tag{11.12}$$

A general surface has two radii of curvature $R_1$ and $R_2$ which can be positive or negative: the traction $\tau$ is perpendicular to the surface and given

by the same formula 11.12 with $1/R$ replaced with the mean curvature $\frac{1}{2}(\frac{1}{R_1} + \frac{1}{R_2})$.

There are two broad classes of coarsening problems: ones with *conserved* and *non-conserved* order parameters. Oil and vinegar, cast iron and granite have conserved order parameters; to grow a domain one must pull molecules through the other materials. The single–component polycrystals and the Ising model shown in figures 11.9 and 11.10 are non–conserved: spins are free to flip from one orientation to the other, and molecules are free to shift from one grain orientation to another.

For a non–conserved order parameter, the interface will generally move with a velocity proportional to the traction, with some coefficient $\gamma$:

$$\frac{dR}{dt} = -\gamma\tau = -\gamma\frac{2\sigma}{R}. \tag{11.13}$$

We can solve for the time $t_f$ it takes for the sphere to disappear, and hence find out how $L(t)$ grows for the non–conserved case:

$$\begin{aligned} \int_{R_0}^{0} R dR &= \int_{0}^{t_f} -2\sigma\gamma dt \\ R_0^2/2 &= 2\sigma\gamma t_f \\ L(t) \sim R_0 &= \sqrt{4\sigma\gamma t} \propto \sqrt{t}. \end{aligned} \tag{11.14}$$

More complex geometries with protrusions and necks and such are not possible to solve explicitly, but in general features with a length scale $R$ evolve on a time scale $t \propto R^2$, so the typical length scales grow as $L(t) \sim t^\beta$ with $\beta = 1/2$.

The argument for the case of a conserved order parameter is quite similar in spirit (figure 11.12). Here the curvature sets up a gradient in the chemical potential $\frac{\partial\mu}{\partial x}$ which causes molecules to diffuse from regions of high positive curvature to regions of low or negative curvature. The current per unit area will be given by the diffusion constant times the chemical potential gradient:

$$J = D\frac{\partial\mu}{\partial x}. \tag{11.15}$$

The chemical potential change for moving a molecule from our sphere of radius $R$ to some flat interface is just the free energy change for removing one particle: if $\rho$ is the density of particles per unit volume, then the number of particles in our sphere is $N = \frac{4}{3}\pi R^3\rho$ and

$$\Delta\mu = \frac{d\mathcal{F}_{\mathbf{surface}}}{dR}\Big/\frac{dN}{dR} = (8\pi\sigma R)/(4\pi R^2\rho) = \frac{2\sigma}{R\rho}. \tag{11.16}$$

The distance $\Delta R$ from our sphere to another region of the same phase is (by our assumption of only one characteristic length scale) also of order $R$, so

$$J \sim D\frac{\Delta\mu}{\Delta R} \sim \frac{2D\sigma}{\rho R^2}. \tag{11.17}$$

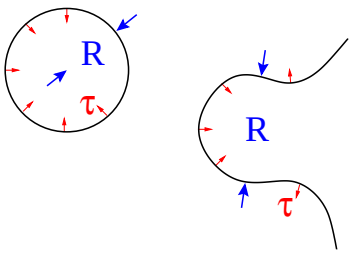



**Fig. 11.11 Curvature–driven interface motion.** The surface tension $\sigma$ at the interface produces a traction (force per unit area) $\tau = 2\sigma\kappa$ that is proportional to the local mean curvature of the surface $\kappa$ at that point. The coarsening morphology has a characteristic length $R$, so it has a characteristic mean curvature $\kappa \sim 1/R$. For non–conserved order parameters, these forces will lead to a length scale $L(t) \sim t^{1/2}$.

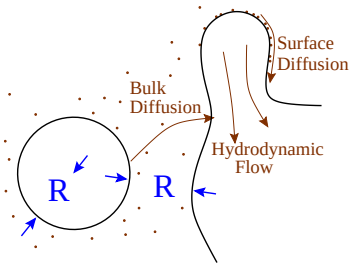



**Fig. 11.12 Coarsening for conserved order parameter.** Differences in local mean curvature drives the growth in the case of a conserved order parameter. Atoms will diffuse from regions of high positive curvature to regions of low or negative curvature. Bulk diffusion dominates on long length scales ($L(t) \sim t^{1/3}$): surface diffusion can be important when the scales are small $L(t) \sim t^{1/4}$). For liquids, hydrodynamic flow makes things more complicated [114].

**Fig. 11.13 Anisotropic Coarsening: Logarithmic Growth.** A cross section of a $80^3$ Ising model with a second–neighbor antiferromagnetic bond, quenched to low temperatures [113]. Notice the large facets along the coordinate axes. This is reflected, for example, in an anisotropy in the late–time correlation function: the correlations along the axes are significantly shorter–range than those along the face or body diagonals.

**Fig. 11.14 Logarithmic Growth of an Interface.** A simplified model of an interface perpendicular to a body diagonal in the Ising model of figure 11.13, [113]. Notice that the interface is lowering its energy by poking out into facets along the cubic directions (a kind of facet coarsening). This process gets much slower as the faces get longer, because the energy barrier needed to flip a face grows linearly with its length.

The rate of change of volume of the droplet is the current density $J$ times the area of the droplet:

$$\begin{aligned}
\frac{dV_{\text{droplet}}}{dt} &= {}^{4}\!/_{3}\pi \left(3R^2 \frac{dR}{dt}\right) \qquad (11.18)\\
&= -A_{\mathbf{droplet}} J = -(4\pi R^2)\frac{2D\sigma}{\rho R^2} = -8\pi D\sigma/\rho \\
\frac{dR}{dt} &= -(2D\sigma/\rho)(1/R^2) \\
\int_{R_0}^{0} R^2 dR &= \int_{0}^{t_f} -2D\sigma/\rho \, dt \\
R_0^3/3 &= (2D\sigma/\rho)t_f \\
L(t) \sim R_0 &= \left(\frac{6D\sigma}{\rho}t\right)^{1/3} \propto t^{1/3}. \qquad (11.19)
\end{aligned}$$

This crude calculation – almost dimensional analysis – leads us to the correct conclusion that conserved order parameters should coarsen with $L(t) \sim t^{\beta}$ with $\beta = {}^{1}\!/_{3}$, if bulk diffusion dominates the transport.

The subject of coarsening has many further wrinkles.

- **Surface Diffusion.** Often the surface diffusion rate is much higher than the bulk diffusion rate: the activation energy to hop on a surface is much lower than to remove a molecule completely. The current in surface diffusion goes as $J$ times a perimeter (single power of $R$) instead of $JA$; repeating the analysis above gives $L(t) \sim t^{1/4}$. In principle, surface diffusion will always be less important than bulk diffusion as $t \to \infty$, but often it will dominate in the experimental range of interest.
- **Hydrodynamics.** In fluids, there are other important mechanisms for coarsening. For example, in binary liquid mixtures (oil and water) near 50/50, the two phases form continuous interpenetrating networks. Different regions of the network can have different curvatures and pressures, leading to coarsening via hydrodynamic flow [114].
- **Non-universality.** Much of the motivation for studying coarsening by physicists has been the close analogies with the scaling and power laws seen in continuous phase transitions (chapter 12 and exercise 12.1). However, there are important differences. The power laws in coarsening are simpler and in a way more universal – the ½ and ⅓ power laws we derived above are independent of the dimension of space, for example. However, for coarsening in crystals the scaling behavior and morphology is not universal: they will depend upon the anisotropic free energies and mobilities of the interfaces [91]. For example, adding next–neighbor antiferromagnetic bonds to the three-dimensional Ising model leads to a strongly cubic anisotropy (figure 11.13). Basically each combination of materials and temperatures will have different scaling functions at late times. In retrospect this is reassuring: there is

such a bewildering variety of microstructures in materials science and mineralogy that it made no sense for one scaling function to rule them all.

- **Glassy logarithmic coarsening.** A hidden assumption in much of the coarsening theory is that rearrangements can happen one molecule at a time. For example, if you need to remove a whole layer at a time, the dynamics may slow down dramatically as the sizes of the layers grows. This is precisely what happens in the 3D Ising model with antiferromagnetic next–neighbor bonds mentioned above (figure 11.14, [113]).[16] The energy barrier needed to flip a layer of spins grows proportionally to $L$, leading to a logarithmic growth law $L(t) \sim \log(t)$.

### 11.3.2 Martensites.

Many crystals will undergo abrupt structural rearrangements as they are cooled – phase transitions between different crystal structures. A good example might be a cubic crystal stretching along one axis and shrinking along the other two. These transitions often are problematic: when part of the sample has transformed and the rest has not, the tearing stress at the interface often shatters the crystal.

In many materials[17] the crystalline shape transition bypasses large–scale stress buildup in the crystal by developing intricate mixed structures. Figure 11.15 shows a picture of a martensite, showing how it forms a patterned microstructure in order to stretch locally without an overall net strain.

The tools used to study martensites are not statistical mechanics, but mathematics.[18] Their basic goal, however, is to minimize the kinds of non–convex free energies that we introduced in exercise 9.5 for fixed boundary conditions; see exercises 11.5 and 11.6.

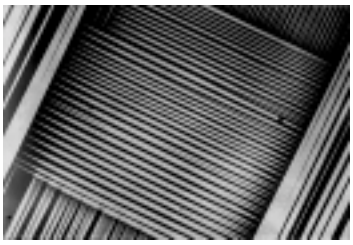


**Fig. 11.15 Martensite.** A particularly nice example of a martensitic structure, from [23]. The light and dark stripes represent two different martensitic variants – that is, the crystal going through the phase transition can stretch in different directions, and the two colors indicate that the local lattice is stretching along two different axes. The tilted square region occupying most of the photograph couldn't change its overall shape without putting incompatible strains on the neighboring domains. By making this striped pattern, or *laminate*, the martensite can form an average of the different stretching directions that gives zero net strain.

### 11.3.3 Dendritic Growth.

Why do snowflakes form? The nucleation of new ice crystals in the upper atmosphere initially forms roughly spherical crystals of ice. As the ice crystals continue to grow, however, an instability develops. The tips of the ice crystals that extend furthest into the surrounding supersaturated vapor will grow fastest, both because they see the highest concentration of water vapor and because the heat released by freezing diffuses away fastest at the tips. The characteristic six–fold patterns arise because each snowflake is a single crystal with six–fold symmetry, and different

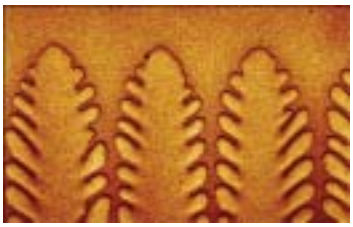


**Fig. 11.16 Dendrites**. Crystals growing into a melt will typically not grow compactly: the tips grow faster than the grooves. Here is shown some crystals growing into a melt in a thin film being pulled through a temperature gradient [12].

[16] This slowdown happens below the "corner rounding transition" described in the caption to figure 11.8.

[17] Notably including iron! Steel has both the complications of carbon particle coarsening and martensitic domain structure, both of which are important for its structural properties and both of which depend in detail on the heating and beating it undergoes during its manufacture.

[18] All the pathological functions you find in real analysis – continuous but nowhere differentiable functions – are practical tools for studying martensites.

crystal surface orientations grow at different rates. The immense variety of snowflake shapes reflects the different thermal and humidity variations experienced by each snowflake.

[19] Dendron is Greek for tree.

The same kind of branched patterns, called *dendrites*,[19] also form in other growing crystals, for precisely the same reasons. Frost on your window is one obvious example (again water condensing from vapor); figure 11.16 shows a single (cubic) solvent crystal growing into a mixture of solvent and polymer. Here instead of heat being trapped in the grooves, the slowly–diffusing polymer is being trapped there and slowing down the solidification process. Many practical metals and alloys used in manufacturing are composed microscopically of tiny dendritic structures packed together.

# Exercises

(11.1) **van der Waals Water.** (Chemistry)

The van der Waals (vdW) equation

$$(P + N^2a/V^2)(V - Nb) = Nk_BT \tag{11.20}$$

is often applied as an approximation to real liquids and gases. It's qualitatively quite useful, but we'll see that in detail it doesn't work well, either far from the transition or near the transition.

*(a) Figure 11.17 shows the vdW pressure versus volume curves for one mole of $H_2O$. Detach the page or trace over it. By hand, roughly implement the Maxwell construction for each curve, and sketch the region in the $P-V$ plane where liquid and gas can coexist. (At constant pressure, the phase with lower Gibbs free energy wins, but at constant volume and temperature the liquid evaporates until the pressure rises to the coexistence, or vapor pressure.)*

**The Critical Point.** The top of the coexistence curve is the pressure, density, and temperature at which the distinction between liquid and gas disappears. It's the focus of much study, as the prime example of a critical point, with self-similar fluctuations and scaling behavior.

*(b) Draw this point in on your sketch from part (a). The vdW constants are fit to the critical temperature $T_c = 647.3K$ and pressure $P_c = 22.09MPa = 220.9 \times 10^6 dyne/cm^2$; check that your estimate for the critical point roughly agrees with the values quoted. I've found few references that quote the critical volume per mole, and the two I've found disagree: one says around 50 $cm^3$/mole and one says around 55. Plot the true critical point on the figure. Is the location of the critical density of water predicted well by the vdW equation of state?*

*(c) Your sketch may not be precise enough to tell this, but the vdW phase boundaries meet at the critical point with a quadratic maximum. Your plot thus shows $1/\rho_\ell - 1/\rho_g \sim (P - P_c)^{1/2}$, where $\rho_\ell$ and $\rho_g$ are the densities on the coexistence boundary (moles per volume) at the pressure $P$. Similarly, one can show that the vdW equation of state implies that*

$$\rho_\ell - \rho_g \sim (T_c - T)^{1/2} \sim (-t)^{1/2}. \tag{11.21}$$

*Compare this latter prediction with Yeomans figure 2.2. What critical exponent $\beta$ does the van der Waals equation predict, assuming equation 11.21?*

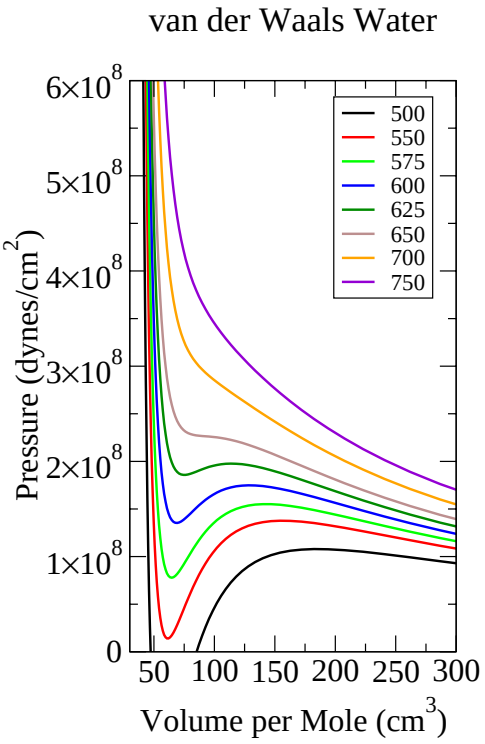



**Fig. 11.17** Pressure versus volume for a van der Waals approximation to $H_2O$, with $a = 0.5507$ Joules meter cubed per mole squared ($a = 1.51957 \times 10^{-35}$ ergs cm$^3$ per molecule), and $b = 3.04\times10^{-5}$ meter cubed per mole ($b = 5.04983\times10^{-23}$ cm$^3$ per molecule), from reference [120].

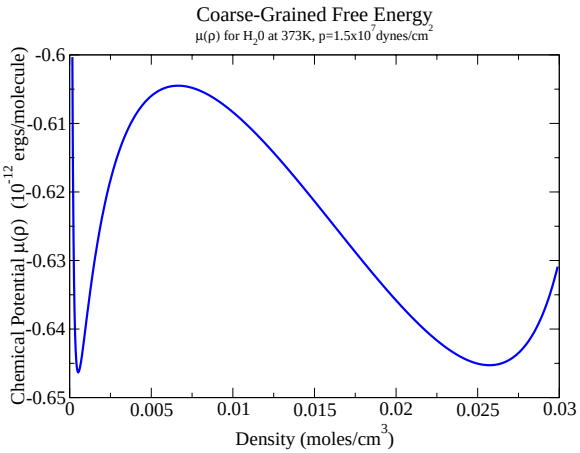



**Fig. 11.18** Chemical potential $\mu[\rho]$ of water fit with the van der Waals equation, as in figure 11.17, at $T = 373K$ and $P = 1.5 \times 10^7 dynes/cm^2$.

**Interfaces and the chemical potential $\mu[\rho(x)]$.** The chemical potential per particle for the vdW equation of state is

$$\mu[\rho] = -k_BT + P/\rho - a\rho + k_BT\log(\lambda^3\rho) - k_BT\log(1-b\rho) \tag{11.22}$$

where $\rho = N/V$ is the density.

*(d) Show that $\mu$ is minimized when $\rho$ satisfies the vdW equation of state.*

*(e) According to the caption to figure 11.18, what is the vdW approximation to the vapor pressure at $373K = 100C$? Atmospheric pressure is around one bar $= 0.1MPa = 10^6 dynes/cm^2$. How close is the vdW approximation to the true vapor pressure of water?* (Hint: what happens when the vapor pressure hits atmospheric pressure?)

We can view figure 11.18 as a kind of free energy barrier for the formation of a liquid-gas interface. If $\mu_0$ is the common chemical potential shared by the water and the vapor at this temperature, the extra Gibbs free energy for a density fluctuation $\rho(x)$ is

$$\Delta G = \int \rho(x)\,(\mu[\rho(x)] - \mu_0)\; d^3x \tag{11.23}$$

since $\rho(x)\,d^3x$ is the number of particles that suffer the chemical potential rise $\mu[\rho(x)]$ in the volume $d^3x$.

*(f) At room temperature, the interface between water and water vapor is very sharp: perhaps a molecule thick. This of course makes the whole idea of using a coarse-grained free energy problematical. Nonetheless, assuming an interfacial width of two or three Ångstroms, use the vdW model for the chemical potential (figure 11.18) and equation 11.23 to roughly estimate the surface tension of water (the extra Gibbs free energy per unit area, roughly the barrier height times thickness). (One mole $= 6.023 \times 10^{23}$ molecules.) How does your answer compare with the measured value at the boiling point, 59 dynes/cm?*

## (11.2) Nucleation in the Ising Model.

The Ising model is not only our prototype for a second-order phase transition and the first triumph of the renor-

malization group; it is also our best model for nucleation and for the dynamics of phase separation.

The Ising model can be used to study the nucleation of one phase inside another. Supercooling water and waiting for an ice crystal nucleus to form can be shown to be quite analogous to changing a magnet from external field $H_{\rm ext} > 0$ to $H_{\rm ext} < 0$ at a temperature $T < T_c$. The analogy with changing the temperature or pressure of a gas and waiting for a raindrop is even better.

Start up the Ising model simulation, available on the Web [105]. Run at $T = 1.5$ and the default size $40 \times 40$. Set *initial conditions* to $M = 1$, and hit the *init* button to start all spins up. Set $H_{\rm ext} = -0.3$ and watch the spins. It should flip the system over to $M = -1$, with the down phase starting in a small, roughly circular cluster of spins, which then grows to fill the system. (You may need to hit "run" more than once: the nucleation time at $H = -0.3$ is roughly equal to the default number of sweeps.) The cluster may start near a boundary or corner: the system has periodic boundary conditions.

*(a) Using the right-hand graph of magnetization versus time, measure the average time it takes to cross zero (which we will call the time to nucleate the down phase) by averaging ten measurements. I recommend the sequence init, reset, run. You can set speed (the number of sweeps per draw) up to 10 or 100 until the bottleneck is the numerics. You can also copy graph on one run, and steal data on subsequent runs, to put all the magnetizations on one graph: zoom in with the right mouse button and hold down the left mouse button to pull off numerical $(x, y)$ data points. Similarly measure the average time to nucleate the down phase for ($H_{\rm ext} = -0.2$), changing number of sweeps to $10^4$ and speed to 100. Since the nucleation center can be located at any site on the lattice, the nucleation rate scales with the number of spins in the system. Calculate, for both fields, the nucleation rate per spin $\Gamma_{exp}(H)$.*

The nucleation rate is often estimated using "critical droplet theory". Small droplets of the stable phase will shrink due to surface tension $\sigma$: large ones grow due to the free energy difference per unit area $H_{\rm ext}\Delta M(T)$, where $\Delta M$ is the magnetization difference between the two states. Presuming that the temperature is high and the droplet large and the times long (so that continuum theories are applicable), one can estimate the critical radius $R_c$ for nucleation.

*(b) Give the formula for the free energy of a flipped cluster of radius $R$ as a function of $\sigma$, $H$, and $\Delta M$. Give formulas for $R_c$ (the critical droplet size where the free energy is a local maximum), the resulting barrier $B$ to nucleation, and the predicted rate $\Gamma_{\rm theory} = \exp(-B/T)$ (assuming a prefactor of roughly one attempt per sweep per spin). At low temperatures, $\sigma \sim 2J \equiv 2$ and $\Delta M \approx 2$, since the system is almost fully magnetized and $\sigma$ is the number of broken bonds ($2J$ each) per unit length of interface. Make a table with rows for the two fields you simulated and with columns for $H$, $R_c$, $B$, $\Gamma_{\rm theory}$, and $\Gamma_{\rm exp}$ from (a).*

This should work pretty badly. Is the predicted droplet size large enough (several lattice constants) so that the continuum theory should be valid?

We can test these ideas better by starting with droplets of down spins (white) in an up background. Use a small system (40×40). You can make such a droplet by setting the *initial conditions* in the upper left to *M=1*, and then flipping a circle of spins in the center by clicking on them with the left mouse button (the only one on Macs). After making the circle, you can *store* and then *recall* it for re-use many times. You'll want to set *speed* (the number of sweeps per draw) to one, since the droplet will grow or shrink rather quickly.

*(c) Start with $H = -0.2$, $T = 1.5$ and a down-spin droplet of radius five (diameter of ten). Does it grow more often than it shrinks, or vice-versa? (Testing this should be fast. Set speed to 100, number of sweeps to 1000, reset, then recall/run/repeat ten times. On the magnetization curve, count the fraction $f$ of times out of ten that the system flipped down.) Make a table of the values of $H$ and $f$ you measure. Vary the field $H$ until the probabilities roughly match: find the field for $R_c = 5$ to within 0.1. For what field is the theoretical critical droplet radius $R_c = 5$ at $T = 1.5$?*

In part (b) we found that critical droplet theory worked badly for predicting the nucleation rate. In part (c) we found that it worked rather well (within a factor of two) at predicting the relationship between the critical droplet size and the external field. This is mostly because the nucleation rate depends exponentially on the barrier, so a small error in the barrier (or critical droplet radius) makes a big error in the nucleation rate. You'll notice that theory papers rarely try to predict rates of reactions. They will almost always instead compare theoretical and experimental barrier heights (or here, critical droplet radii). This avoids embarrassment!

This free energy barrier to nucleation is what allows supercooled liquids and supersaturated vapor to be stable for long periods.

### (11.3) Coarsening and Criticality in the Ising Model.

If you suddenly cool water vapor at constant volume far below the transition, you can go into a region of the phase diagram where there is no barrier to nucleation. In this region, one observes *spinodal decomposition*: in the absence of gravity, gas and liquid phase separate slowly, forming

elaborate intertwining regions of the two phases. This can be observed in salad dressings (oil and water after shaking) and is an important source of the morphology of many materials (from rocks to Pyrex glassware). Similar behavior is seen even for weak undercoolings when the nucleation rate is large.

More generally, we call *coarsening* the process by which phases separate from one another: the surface tension drives the tiny fingers and droplets to shrink, leading to a characteristic length scale that grows with time. In our Ising model, the magnetization (or the number of up-spin "atoms", viewed as a lattice gas) is not conserved, so the coarsening behavior is somewhat different from that seen in spinodal decomposition.

Start up the Ising model again [105]. Using the *configure* menu, set the lattice width and height to a reasonably large value (280×280 works well on my computer). Demagnetize the sample by selecting *T=inf* from the initial conditions. Now set $T$ below $T_c$ (say, $T = 1.5$), set the external field to zero, and let it run with *speed* one (one sweep at a time). (As the pattern develops, the changes occur more slowly and you may want to increase *speed* to 10 or more.) The pattern looks statistically the same at different times, except for a typical *coarsening length* that is growing. How can we define and measure the typical length scale $L(t)$ of this pattern?

*(a) Argue that at zero temperature the total energy above the ground state energy is proportional to the perimeter separating up-spin and down-spin regions. (At finite temperatures, there is a contribution from thermally flipped spins, which shouldn't really count as perimeter for coarsening.) Argue that the inverse of the perimeter per unit area is a reasonable definition for the length-scale of the pattern.*

*(b) Initialize at T=inf, set temperature and external field to zero and number of sweeps to one, run for one sweep, and measure the mean energy per unit area $\langle E \rangle$ (displayed on the upper right). Measure the mean energy as a function of time for $t = 2, 4, 8, \ldots, 1024$: reset/run/measure $\langle E \rangle$/double number of sweeps/repeat. (You're measuring the average perimeter length over the last half of the time interval, but that scales in the same way as the perimeter does.) Feel free to turn speed up in later stages. Make a table with columns for $t$, $\langle E(t) \rangle$, and $L(t) \propto 1/(\langle E \rangle + 2)$. Make a log-log plot of your estimate of the coarsening length $L(t) \propto 1/(\langle E \rangle + 2)$ versus time. What power law does it grow with? What power law did we expect?*

**Self-Similarity at the Critical Point.** We'll use the (numerically easier) hysteresis model as our example of a system with critical fluctuations and scaling, but we should at least see the Ising model critical point.

Run a large system (280 × 280 or so) at zero external field and $T = T_c = 2/\log(1+\sqrt{2}) = 2.26919$. (Set *speed* large enough so that graphics is not the bottleneck, and run for at least a couple hundred sweeps to equilibrate.) You should see a fairly self-similar structure, with fractal-looking up-spin clusters inside larger down-spin structures inside ... *(c) Can you find a nested chain of three clusters? Four?*

## (11.4) Nucleation of Dislocation Pairs.

Consider a two-dimensional crystal under shear shown in figure 11.19.[20] The external force is being relieved by the motion of the upper half of the crystal to the left with respect to the bottom half of the crystal by one atomic spacing $a$. If the crystal is of length $L$, the energy released by this shuffle when it is complete will be $|F|a = \sigma_{xy} L a$. This shuffling has only partially been completed: only the span $R$ between the two edge dislocations has been shifted (the dislocations are denoted by the conventional "tee" representing the end of the extra column of atoms). Thus the strain energy released by the dislocations so far is

$$|F|aR/L = \sigma_{xy} R a. \tag{11.24}$$

This energy is analogous to the bulk free energy gained for vapor droplet in a superheated fluid (http://www.lassp.cornell.edu/sethna/Nucleation/).

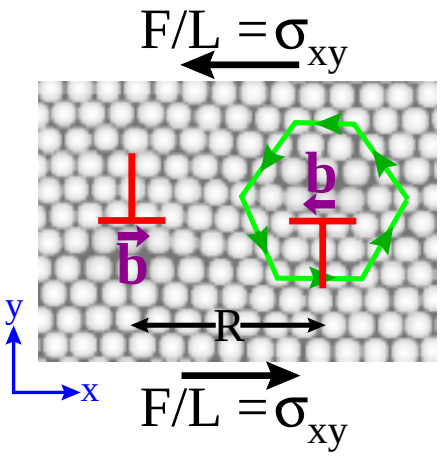



**Fig. 11.19 Dislocation Pair in a 2D Hexagonal Crystal**, (Nick Bailey). The loop around the defect on the right

[20] A similar problem, for superfluids, was studied in references [2, 3], see also [116]. The complete solution is made more complex by the effects of other dislocation pairs renormalizing the elastic constants at high temperatures.

shows an extra atom along the bottom row. By the conventions used in materials physics (assuming here that the dislocation "points" up out of the paper, see Hirth and Lothe, *Theory of Dislocations*, second edition, figure 1.20, p. 23). this *edge dislocation* has Burgers vector $\mathbf{b} = a\hat{x}$, where $a$ is the distance between neighboring atoms. Similarly, the defect on the left has an extra atom in the top row, and has $\mathbf{b} = -a\hat{x}$. The defects are centered at the same height, separated by a distance R. The crystal is under a shear stress $\sigma_{xy} = F/L$, where the force $F = \pm\sigma_{xy}L\hat{y}$ is applied to the top and bottom as shown (but the crystal is kept from rotating).

The dislocations, however, cost energy (analogous to the surface tension of the vapor droplet). They have a fixed core energy $C$ that depends on the details of the interatomic interaction, and a long-range interaction energy which, for the geometry shown in figure 11.19, is

$$\frac{\mu}{2\pi(1-\nu)}a^2\log(R/a). \tag{11.25}$$

Here $\mu$ is the 2D shear elastic constant[21] and $\nu$ is Poisson's ratio. For this exercise, assume the temperature is low (so that the energies given by equations 11.24 and 11.25 are good approximations for the appropriate free energies). By subtracting the energy gained from the dislocation from the energy cost, one finds in analogy to other critical droplet problems a critical radius $R_c$ and a barrier height for thermally nucleated dislocation formation $B$.

*Of the following statements, which are true?*

*(T) (F) The critical radius $R_c$ is proportional to $1/\sigma_{xy}$.*

*(T) (F) The energy barrier to thermal nucleation is proportional to $1/\sigma_{xy}^2$.*

*(T) (F) The rate $\Gamma$ of thermal nucleation of dislocations predicted by our critical droplet calculation is of the form $\Gamma = \Gamma_0(T)\,(\sigma_{xy}/\mu)^{D/k_BT}$, for a suitable material-dependent function $\Gamma_0(T)$ and constant $D$.*

Dislocations mediate plastic shear. For a small sample, each pair of dislocations nucleated will travel to opposite boundaries of the system and lead to a net shear of one lattice constant. Thus, at any non-zero temperature and external stress, a (two-dimensional) crystal will shear at a non-zero rate. How is the crystal, then, different in its response from a liquid? To be specific,

*(T) (F) According to our calculation, the response of a two-dimensional crystal under stress is indistinguishable from that of a liquid: even at low temperatures, the strain rate due to an external shear force is proportional to the stress.*

(11.5) **Oragami Microstructure.** (Mathematics)

Figure 11.15 shows the domain structure in a thin sheet of material that has undergone a *martensitic* phase transition. These phase transitions change the shape of the crystalline unit cell: for example, the high temperature phase might be cubic, and the low temperature phase might be stretched along one of the three axes and contracted along the other two. These three possibilities are called *variants*. A large single crystal at high temperatures thus can transform locally into any one of the three variants at low temperatures.

The order parameter for the martensitic transition is a *deformation field* $\mathbf{y}(\mathbf{x})$, representing the final position $\mathbf{y}$ in the martensite of an original position $\mathbf{x}$ in the undeformed, unrotated austenite. The variants differ by their *deformation gradients* $\nabla\mathbf{y}$ representing the stretch, shear, and rotation of the unit cells during the crystalline shape transition.

In this exercise, we develop an analogy due to Richard James [48] between martensites and paper folding. Consider a piece of graph paper, white on one side and lined on the other, lying flat on a table. This piece of paper has two distinct low energy states, one variant with white side up and one variant with lined side up.

The (free) energy density for the paper is independent of rotations, but grows quickly when the paper is stretched or sheared. The paper, like martensites, can be represented as a deformation field $\mathbf{y}(\mathbf{x})$, representing the final position $\mathbf{y}$ of a point $\mathbf{x}$ of the paper placed horizontally on the table with the lined–side up. Clearly, $\mathbf{y}(\mathbf{x})$ must be a continuous function to avoid ripping the paper. Since the energy is independent of an overall translation of the paper on the table, it can depend only on gradients of the deformation field. To lowest order,[22] the energy density can be written in terms of the deformation gradient $\nabla\mathbf{y} = \partial_j y_i$:

$$\mathcal{F} = \alpha|(\nabla\mathbf{y})^T\nabla\mathbf{y} - \mathbb{I}|^2 = \alpha(\partial_i y_j\partial_i y_k - \delta_{jk})^2 \tag{11.26}$$

The constant $\alpha$ is large, since paper is hard to stretch. In this problem, we'll be interested in the zero–energy ground states for the free energy.

*(a) Show that the zero–energy ground states of the paper free energy density (equation 11.26) include the two variants and rotations thereof, as shown in figure 11.20. Specifically, show that any rotation $y_i(x_j) = R_{ij}x_j$ of*

[21] The $2D$ elastic constants $\mu$ and $\nu$ can be related to their 3D values; in our notation $\mu$ has units of energy per unit area.

[22] Including higher derivatives of the deformation field into the energy density would lead to an energy per unit length for the creases.

*the lined-side-up position is a ground state, where* $R_{ij} = \begin{pmatrix} \cos\theta_\ell & -\sin\theta_\ell \\ \sin\theta_\ell & \cos\theta_\ell \end{pmatrix}$, *and also that flipping the paper to the white-side-up and then rotating* $y_i(x_j) = R_{ik}P_{kj}x_j = \begin{pmatrix} \cos\theta_w & -\sin\theta_w \\ \sin\theta_w & \cos\theta_w \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$ *gives a ground state.*

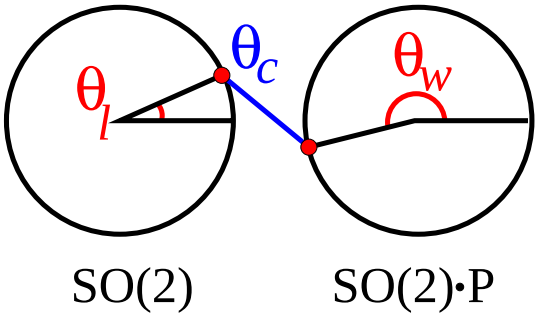



**Fig. 11.20** The allowed zero-energy deformation gradients for a piece of paper lying flat on a table. Let $\theta$ be the angle between the $x$-axis of the graph paper and the near edge of the table. The paper can be rotated by any angle $\theta_\ell$ (so the deformation gradient is a pure rotation in the group[23] SO(2)). Or, it can be flipped over horizontally ($(x, y) \to (x, -y)$, multiplying by $P = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$) and then rotated by $\theta_w$ (deformation gradient in the set SO(2)·P). An interface between two of these ground states is a straight crease at angle $\theta_c$ (figure 11.21).

In the real martensite, there are definite rules for boundaries between variants: given one variant, only certain special orientations are allowed for the boundary and the other variant. A boundary in our piece of paper between a lined-up and white-up variant lying flat on the table is simply a crease (figure 11.21).

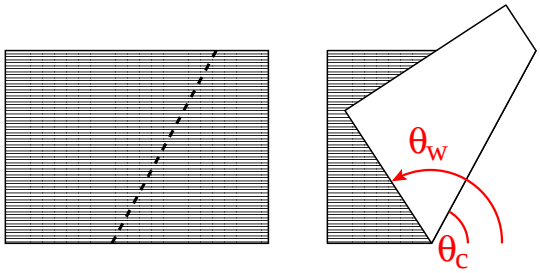



**Fig. 11.21 Crease.** An interface between two ground states $\theta_\ell = 0$ and $\theta_w$ for our paper on the table is a straight crease with angle $\theta_c$.

*(b) Place a piece of paper long-edge downward on the table. Holding the left end fixed* $\theta_\ell = 0$, *try folding it along crease lines at different angles* $\theta_c$. *Find a definite relation between the crease angle* $\theta_c$ *and the angle* $\theta_w$ *of the right-hand portion of the paper.*

Suppose the crease is along an axis $\hat{\mathbf{c}}$. We can derive the general relation governing a crease by noting that $\mathbf{y}$ along the crease must agree for the white and the lined faces, so the directional derivative $D\mathbf{y} \cdot \mathbf{c} = (\hat{\mathbf{c}} \cdot \nabla)\mathbf{y}$ must agree.

*(c) Given the relation you deduced for the geometry in part (b), show that the difference in the directional derivatives* $(D\mathbf{y}^\ell - D\mathbf{y}^w)$ *is zero along* $\mathbf{c}$, $(D\mathbf{y}^\ell - D\mathbf{y}^w) \cdot \mathbf{c} = (\partial_j y_i^\ell - \partial_j y_i^w)c_j = \mathbf{0}$. (Hints: $D\mathbf{y}^\ell$ is the identity. $\cos(2\theta) = \cos^2\theta - \sin^2\theta$, $\sin(2\theta) = 2\sin\theta\cos\theta$.)

In general, two variants with deformation gradients $A$ and $B$ of a martensite can be connected together along a flat boundary perpendicular to $\mathbf{n}$ if there are rotation matrices $R_1$ and $R_2$ such that[24]

$$R_2 B - R_1 A = \mathbf{a} \otimes \mathbf{n}. \tag{11.27}$$

This ensures that the directional derivatives of $y$ along the boundary directions (perpendicular to $\mathbf{n}$) will be the same for the two variants, and hence that the deformation field is continuous at the boundary.

As can be seen in figure 11.15, the real martensite did not transform by stretching uniformly along one axis. Instead, it formed multiple thin layers of two of the variants.

[23] A matrix $M$ is *orthogonal* if $M^T M = \mathbb{I}$, the matrix times its transpose is the identity. The set of all $n \times n$ orthogonal matrices forms a group, $O(n)$. Since $\det(AB) = \det(A)\det(B)$, and $\det(M^T) = \det(M)$, orthogonal matrices $M$ either have $\det(M) = 1$ (so-called *special orthogonal* matrices, in the group $SO(n)$) or $\det(M) = -1$ (in which case they are the product of a special orthogonal matrix times the reflection $P$). Thus $O(n)$ as a set or manifold always comes in two distinct components. Hence in part (a) you're showing that all elements of $O(2)$ are ground states for the paper.

[24] That is, the difference is a *rank one* matrix, with zero eigenvalues along all directions perpendicular to $\mathbf{n}$.

It can do so for a modest energy cost because the surface energy of the boundary between two variants is low.

The martensite is driven to this laminated structure to satisfy *boundary conditions*. Steels go through a martensitic transition: as the blacksmith cools the horseshoe, local crystalline regions of the iron stretch along one of several possible axes. The red-hot horseshoe does not change shape overall as it is plunghed into the water, though. This is for two reasons. First, if part of the horseshoe started stretching before the rest, there would be big stresses at the boundary between the transformed and untransformed regions. Second, a horseshoe is made up of many different crystalline grains, and the stretching is along different axes in different grains. Instead, the horseshoe, to a good approximation, picks a local mixture between the different variants that overall produces no net average stretch.

This is done by creating finely divided structures, like the laminated structure seen in figure 11.15.[25] At the boundaries of the square region, the martensite mustn't stretch, so it produces a fine laminated structure where the stretching in one domain cancels the contraction for its neighbors.

Our paper folding example forms a similar microstructure when we insist that the boundary lie along a curve other than the natural one.

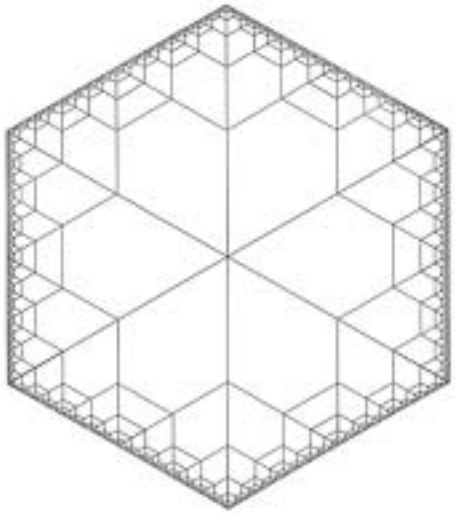


**Fig. 11.22** Two–dimensional oragami example of microstructure formation by Richard James [48].

*(d) Go to the book Web site [108] and print out a full-sized version of figure 11.22. Cut out the hexagon, and fold along the edges. Where does the boundary go?*[26]

The mathematicians and engineers who study these problems take the convenient limit where the energy of the boundaries between the variants (the crease energy in our exercise) goes to zero. In that limit, the microstructures can become infinitely fine, and only quantities like the relative mixtures between variants are well defined. It's a wonderful (and rare) example where the pathological functions of real analysis describe important physical phenomena.

(11.6) **Minimizing Sequences and Microstructure.** (Mathematics)

The martensitic morphology seen in figure 11.15 is a finely divided mixture between two different crystal variants. This layered structure (or *laminate*) is produced by the

[25] The laminated microstructure of the real martensite is mathematically even more strange than that of the paper. The martensite, in the limit where the boundary energy is ignored, has a deformation gradient which is discontinuous everywhere in the region; our folded paper has a deformation gradient which is discontinuous only everywhere along the boundary. See exercise 11.6

[26] The proof that the diagram can be folded along the creases is a special case of a general theorem [48], that any network of creases where all nodes have four edges and opposite opening angles add up to 180° can be folded onto the plane, a condition which is possible, but challenging, to derive from equation 11.27. Deducing the final shape of the boundary can be done by considering how the triangles along the edge overlap after being folded.

[27] Except on the boundaries between domains, which although dense still technically have measure zero.

material to minimize the strain energy needed to glue the different domains together. If the interfacial energy needed to produce the boundaries between the domains were zero, the layering could become infinitely fine, leading to a mathematically strange function. The displacement field $\mathbf{y}(\mathbf{x})$ in this limit would be continuous, and at each point[27] $\mathbf{x}$ it would have a gradient which agrees with one of the ground states. However, the gradient would be discontinuous everywhere, jumping from one variant to the next each time a boundary between domains is crossed.

It is in this weird limit that the theory of martensites becomes elegant and comprehensible. If you are thinking that no such function $\mathbf{y}(\mathbf{x})$ exists, you are of course correct: one can approach zero strain energy with finer and finer laminates, but no function $\mathbf{y}(\mathbf{x})$ actually can have zero energy. Just as for the function in figure 11.23, the greatest lower bound of the martensitic energy *exists*, but is not *attained*.

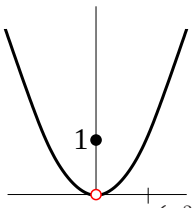



**Fig. 11.23** The function $g(x) = \begin{cases} x^2 & x \neq 0 \\ 1 & x = 0 \end{cases}$ has a minimum value $g = 0$, but never attains that minimum.

A *minimizing sequence* for a function $g(x)$ with lower bound $g_0$ is a sequence of arguments $x_1, x_2, \ldots$ for which $g(x_n) > g(x_{n+1})$ and $\lim g(x_n) = g_0$.

*(a) Find a minimizing sequence for the somewhat silly function g in figure 11.23.*

This kind of microstructure often arises in systems with *non-convex* free energy densities. Consider a problem where the energy of a function $y(x)$ is given by

$$\mathcal{F}[y] = \int_0^1 (y'^2 - 1)^2 + y^2 \, dx. \tag{11.28}$$

with boundary conditions $y(0) = y(1) = 0$. This energy is low if $y(x)$ stays near zero and the slope $dy/dx = y'(x)$ stays near $\pm 1$. The latter is why it is non-convex: there are two values of the slope which have low energy density, but intermediate values of the slope have higher energy density.[28] This free energy is similar to that for the two-dimensional paper folding exercise 11.5; you could think of it as the folding of a one-dimensional sheet of paper ($y' = \pm 1$ representing face-up and face-down states) in a potential $y^2$ pulling all points to the origin, forcing the paper to crumple into a small ball.

**Microstructure Theorem 1.** $\mathcal{F}[y]$ of equation 11.28 does not attain its minimum.

*(b) Prove Microstructure Theorem 1. In particular,*

- *Show that zero is a lower bound for the energy $\mathcal{F}$.*
- *Construct a minimizing sequence of functions $y_n(x)$ for which $\lim \mathcal{F}[y_n] = 0$.*
- *Show that the second term of $\mathcal{F}[y]$ is zero only for $y(x) = 0$, which does not minimize $\mathcal{F}$.*

(Advanced.) **Young Measures.** It is intuitively clear that any minimizing sequence for the free energy of equation 11.28 must have slopes that approach $y' \approx \pm 1$, and yet have values that approach $y \approx 0$. Mathematically, we introduce a probability distribution (the *Young measure*) $\nu_x(S)$ giving the probability of having slope $S = y'(x+\epsilon)$ for points $x + \epsilon$ near $x$.

*(c) Argue that the Young measure which describes minimizing sequences for the free energy in equation 11.28 is $\nu_x(S) = \frac{1}{2}\delta(S-1) + \frac{1}{2}\delta(S+1)$.* (Hint: The free energy is the sum of two squares. Use the first term to argue that the Young measure is of the form $\nu_x(S) = a(x)\delta(S-1) + (1-a(x))\delta(S+1)$. Then write $\langle y(x) \rangle$ as an integral involving $a(x)$, and use the second term in the free energy to show $a(x) = \frac{1}{2}$.)

[28] A function $f[x]$ is convex if $f[\lambda a + (1-\lambda)b] \leq \lambda f[a] + (1-\lambda)f[b]$; graphically, the straight line segment between the two points $(a, f[a])$ and $(b, f[b])$ lies below $f$ if $f$ is convex. The free energy $\mathcal{F}$ in equation 11.28 is non-convex as a function of the slope $y'$.

# Continuous Transitions

# 12

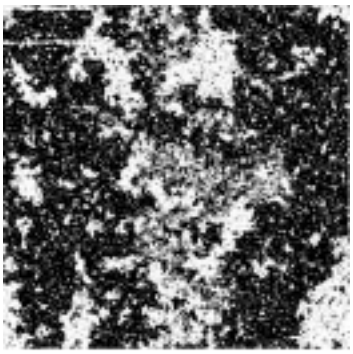


**Fig. 12.1** The Ising model at the critical temperature $T_c$, separating the magnetized phase $T < T_c$ from the zero–magnetization phase $T > T_c$. The white and black regions represent positive and negative magnetizations $s = \pm 1$. Unlike the abrupt transitions studied in chapter 11, here the magnetization goes to zero continuously as $T \to T_c$ from below.

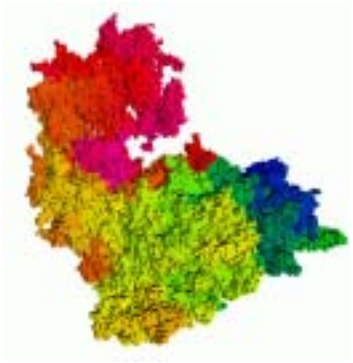


**Fig. 12.2** A medium–sized avalanche (flipping 282,785 domains) in a model of avalanches and hysteresis in magnets (see exercises 8.12, 12.10 and figure 12.14). The shading depicts the time evolution: the avalanche started in the dark region in the back, and the last spins to flip are in the upper, front region. The sharp changes in shading are real, and represent sub–avalanches separated by times where the avalanche almost stops (see figure 8.24).

Continuous phase transitions are fascinating. In figure 12.1 we again show the Ising model at its critical temperature. Above the critical point $T_c$, the Ising model has zero net magnetization; below $T_c$ it has a non–zero magnetization $\pm M(T)$ which has one of two possible signs. Unlike the case of an abrupt transition (chapter 11), the magnetization vanishes as $T$ is raised to $T_c$. At $T_c$, we see large fluctuations in the magnetization in space (figure 12.1) and time: instead of picking one of the up–spin, down–spin, or zero–magnetization states, the Ising model at $T_c$ is a kind of fractal[1] blend of all three.

This fascinating behavior is not confined to equilibrium phase transitions. Figure 12.3 shows the *percolation transition*. An early paper which started the widespread study of this topic [62] described punching holes at random places in a conducting sheet of paper and measuring the conductance. Let the probability of punching a hole in a given region be $(1 - p)$; then for $p$ near one (no holes) the conductivity is large, but decreases as $p$ decreases. After enough holes are punched (at a particular $p_c$), the sheet falls apart and, naturally, the conductance goes to zero. The conductivity they measured fell to a very small value as the number of holes approached the critical concentration, because the conducting paths were few and tortuous just before the sheet fell apart. Thus this model too shows a continuous transition – a qualitative change in behavior at $p_c$ at which the properties are singular but continuous. Figure 12.3 shows a percolation model implemented on the computer (exercise 8.11) where bonds between grid points are removed rather than circular holes.

Many physical systems involve events of a wide range of sizes (figure 12.2). We often find such systems particularly important, since the largest events often are catastrophic. Figure 12.4 shows the energy released in earthquakes in 1995 versus time. The earth's crust responds to the slow motion of the tectonic plates in continental drift through a series of sharp, impulsive earthquakes. The same kind of crackling noise arises in many other systems, from crumpled paper [46] to Rice Krispies™ [53, 109]. The number $D(S)$ of these impulsive events (called *avalanches*) of size $S$ often forms a power law over many decades of sizes (figure 12.5).

In the last few decades, it has been recognized that many of these

[1] The term *fractal* was coined to describe sets which have characteristic dimensions that are not integers: it roughly corresponds to non-integer Hausdorff dimensions in mathematics. The term has entered the popular culture, and is associated with strange, rugged sets like those depicted in the figures here.

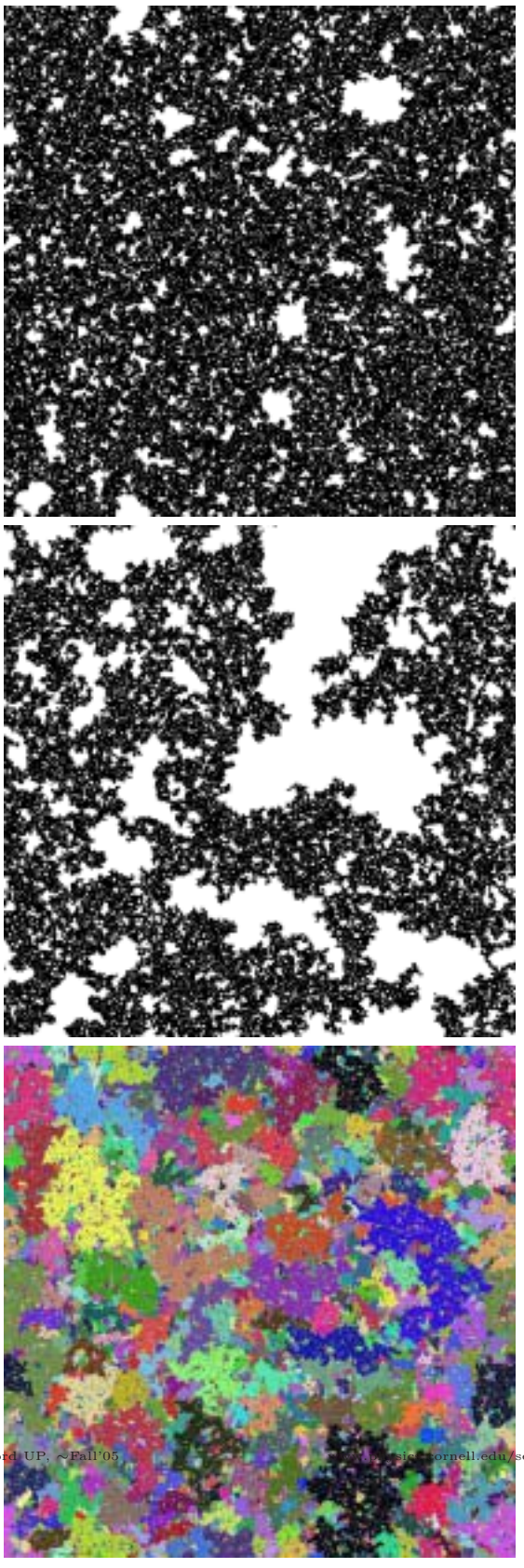


**Fig. 12.3 Percolation.** If we take a square lattice of sites, with bonds between nearest neighbor sites, and remove bonds with probability $1 - p = 1/2$, a largest cluster for a 1024x1024 lattice is shown in the center picture. It happens that $p_c = 1/2$ for 2D bond percolation on the square lattice, and one can see the biggest cluster just barely hangs together, with holes on all length scales. At larger probabilities of retaining bonds $p = 0.51$, the largest cluster is intact with only small holes (top): at smaller $p = 0.49$ the sheet falls into small fragments (bottom, shades denote clusters). Percolation has a phase transition at $p_c$, separating a connected phase from a fragmented phase. See exercises 8.11 and 12.9.

systems can be studied as critical points – continuous transitions between qualitatively different states. There are many theoretical models of earthquakes with this property, one of which is shown in figure 12.6. There are also many critical point models of other systems exhibiting crackling noise. Figure 12.2 shows the domains flipped in a single avalanche in a model of hysteresis studied by the author ([109], exercises 8.12 and 12.10). We'll see that this model also has a continuous transition between qualitatively different behaviors, and that we can understand most of the properties of large avalanches in the system using the same tools developed for studying equilibrium phase transitions.

The renormalization–group and scaling methods we use to study these critical points are deep and powerful. Much of the history and practice in the field revolves around complex perturbative schemes to implement these methods (approximately) for various specific systems. In this chapter, we will focus on the key ideas most useful in exploring experimental systems and new theoretical models, and will not discuss the perturbative methods for calculating critical exponents.

In section 12.1 we will examine the striking phenomenon of *universality*: two systems, microscopically completely different, can exhibit the precisely the same critical behavior near their phase transitions. We'll provide a theoretical rationale for universality in terms of a *renormalization–group* flow in a space of all possible systems.

In section 12.2 we'll explore the characteristic *self–similar* structures found at continuous transitions. Self–similarity is the explanation for the fractal–like structures seen at critical points: a system at its critical point looks the same when rescaled in length (and time). We'll show that *power laws* and *scaling functions* are simply explained from the assumption of self–similarity.

Finally, in section 12.3 we'll give an overview of the wide variety of types of systems that are being understood using renormalization–group and scaling methods.

## 12.1 Universality.

Quantitative theories of physics are possible because macroscale phenomena are often independent of microscopic details.[2] We saw in chapter 2 that the diffusion equation was independent of many details of the underlying random collision processes. Fluid mechanics relies upon the emergence of simple laws – the Navier–Stokes equations – from complex

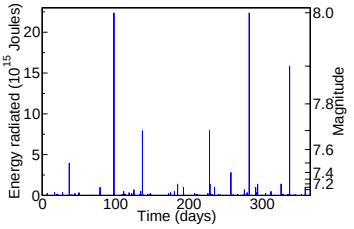



**Fig. 12.4** Earthquake energy release on the Earth versus time for 1995. This time series, when sped up, sounds quite similar to the noise formed by crumpling paper, or the electromagnetic waves emitted by iron as it is magnetized [53].

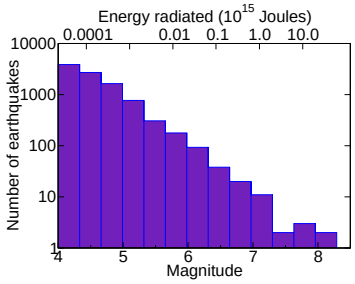



**Fig. 12.5** Histogram of the number of earthquakes in 1995 as a function of their size $S$ (see figure 12.4). Notice the logarithmic scales: the smallest earthquakes shown are a million times smaller and a thousand times more probable than the largest earthquakes. The fact that this distribution is well described by a power law (straight line on the log–log plot) is the Gutenberg–Richter law $\sim S^{-\frac{2}{3}}$.

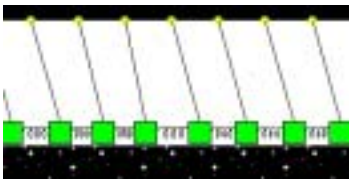


**Fig. 12.6** The Burridge–Knopoff model of earthquakes, with the earthquake fault modeled by blocks pulled from above and sliding with friction on a surface below. It was later realized by Carlson and Langer [19] that this model evolves into a state with a large range of earthquake sizes even for regular arrays of identical blocks.

[2] One should be careful here: much of theoretical particle physics is based on a kind of anti–universality: in their experience, only a few microscopic theories can possibly yield the behavior they observe at low energies (corresponding to our long length and time scales). One must note, however, that the mundane types of universality remain underpinnings for particle physics. For example, the utility of numerical lattice simulations (lattice QCD, for example) depends upon the fact that a suitably chosen simulation which breaks translational, rotational, and Lorentz symmetries can flow to a fixed point that exhibits these symmetries – many different lattice regularizations will flow to the same low–energy, long–wavelength theory.

underlying microscopic interactions: if the macroscopic fluid motions depended in great detail on the shapes and interactions of the constituent molecules, we could not write simple continuum laws. Similarly, our everyday use of quantum mechanics relies on the fact that the behavior of electrons, nuclei, and photons are largely independent of the microscopic details – quantum electrodynamics is an effective theory for low energies which emerges out of more complicated unified theories.[3]

[3]Indeed, the first theories of renormalization grew out of understanding how this emergence happens in relativistic quantum systems.

The behavior near continuous transitions is unusually independent of the microscopic details of the system. Figure 12.7 shows that the liquid and gas densities $\rho_\ell(T)$ and $\rho_g(T)$ for a variety of atoms and small molecules appear quite similar when rescaled to the same critical density and temperature. This similarity is partly for simple reasons: the interactions between the molecules is roughly the same in the different systems up to overall scales of energy and distance. Hence argon and carbon monoxide satisfy

$$\rho^{\mathrm{CO}}(T) = A\rho^{\mathrm{Ar}}(BT) \tag{12.1}$$

for some overall changes of scale $A$, $B$. However, figure 12.8 shows a completely different physical system – interacting electronic spins in nickel, going through a ferromagnetic transition. The magnetic and liquid–gas theory curves through the data are the same if we allow ourselves to not only rescale $T$ and the order parameter ($\rho$ and $M$, respectively), but also allow ourselves to tilt the temperature axis using a more general coordinate change

$$\rho^{\mathrm{Ar}}(T) = A(M(BT), T) \tag{12.2}$$

where $B = T_c^M/T_c^{\ell g}$ is as usual the rescaling of temperature and $A(M,T) = a_1 M + a_2 + a_3 T = (\rho_c\rho_0/M_0)\,M + \rho_c(1+s) - \left(\rho_c s/T_c^{\ell g}\right) T$ is a simple shear coordinate transformation from $(\rho, T^{\ell g})$ to $(M, T^M)$ which untilts the axis.[4]

This would perhaps not be a surprise if these two phase diagrams had parabolic tops: the local maximum of an analytic curve generically[5] looks parabolic. But the jump in magnetization and density near $T_c$ both vary as $(T_c - T)^\beta$ for an exponent $\beta \approx 0.325$, distinctly different from the square–root singularity $\beta = 1/2$ of a generic analytic function.

[5]The term *generic* is a mathematical term which roughly translates as 'except for accidents of zero probability', like finding a function with zero second derivative at the maximum. We'll see that the $\rho(T)$ curve has a non-analytic power-law singularity at its peak.

Also, there are many other properties (susceptibility, specific heat, correlation lengths) which have power–law singularities at the critical point, and all of these power laws for the liquid–gas systems agree with the corresponding exponents for the magnets. We call this *universality*. When two different systems have the same singular properties at their

[4]Nature doesn't anticipate our choice of $\rho$ and $T$ for variables: at the liquid–gas critical point the natural measure of density is temperature dependent: $A(M,T)$ is the coordinate change to the natural coordinates. As it happens, there is another correction proportional to $(T_c - T)^{(1-\alpha)}$, where $\alpha \sim 0.1$ is the specific heat exponent. It can also be seen as a kind of tilt, from a pressure–dependent effective Ising–model coupling strength. It's small for the simple molecules in figure 12.7, but significant for liquid metals [38]. Both the tilt and this $1-\alpha$ correction are *subdominant*, meaning that they vanish faster as we approach $T_c$ than the order parameter $(T_c - T)^\beta$.

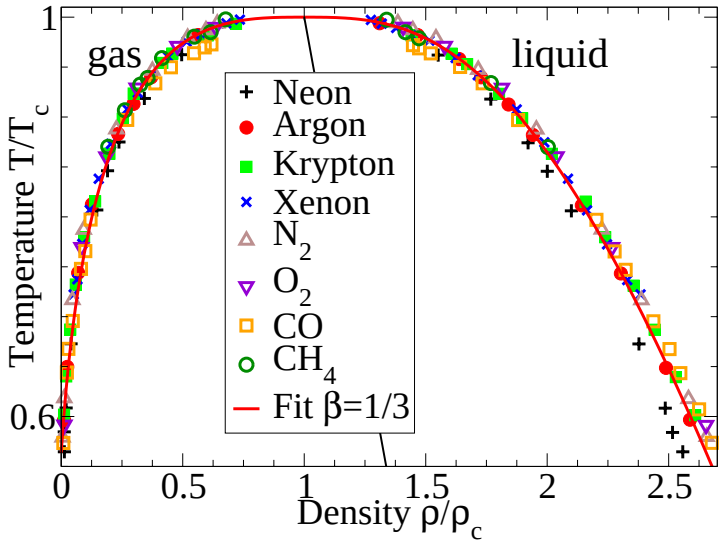



**Fig. 12.7 Universality: Liquid–Gas Critical Point.** The liquid–gas coexistence lines ($\rho(T)/\rho_c$ versus $T/T_c$) for a variety of atoms and small molecules, near their critical points $(T_c, \rho_c)$ [42]. The curve is a fit to the argon data, $\rho/\rho_c = 1 + s(1 - T/T_c) \pm \rho_0(1 - T/T_c)^\beta$ with $s = 0.75$, $\rho_0 = 1.75$, and $\beta = 1/3$ (reference [42]).

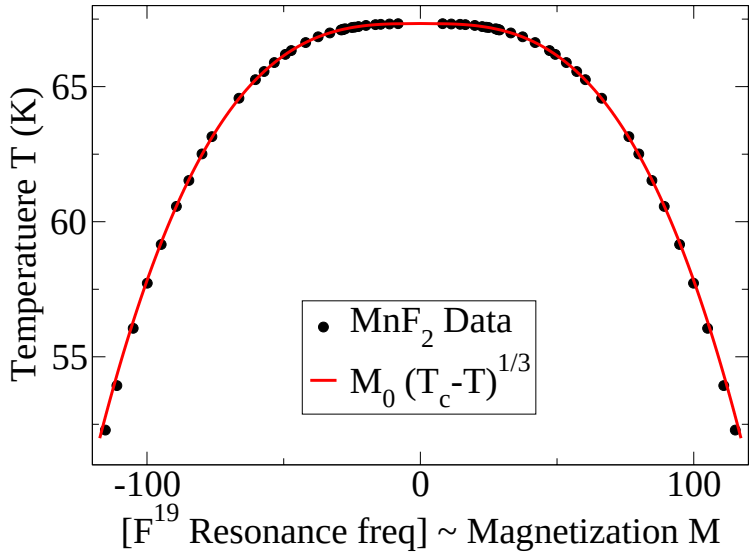



**Fig. 12.8 Universality: Ferromagnetic-Paramagnetic Critical Point.** Magnetization versus temperature for a uniaxial antiferromagnet $MnF_2$ [45]. We've shown both branches $\pm M(T)$ and swapped the axes so as to make the analogy with the liquid–gas critical point apparent (figure 12.7). Notice that both the magnet and the liquid–gas critical point have order parameters that vary as $(1 - T/T_c)^\beta$ with $\beta \approx 1/3$. (The liquid–gas coexistence curves are tilted: the two theory curves would align if we defined an effective magnetization for the liquid gas critical point $\rho_{\text{eff}} = \rho - 0.75\rho_c(1 - T/T_c)$ (thin midline, figure 12.7).) This is not an accident: both are in the same universality class, along with the three-dimensional Ising model, with the current estimate for $\beta = 0.325 \pm 0.005$ [124, Chapter 28].

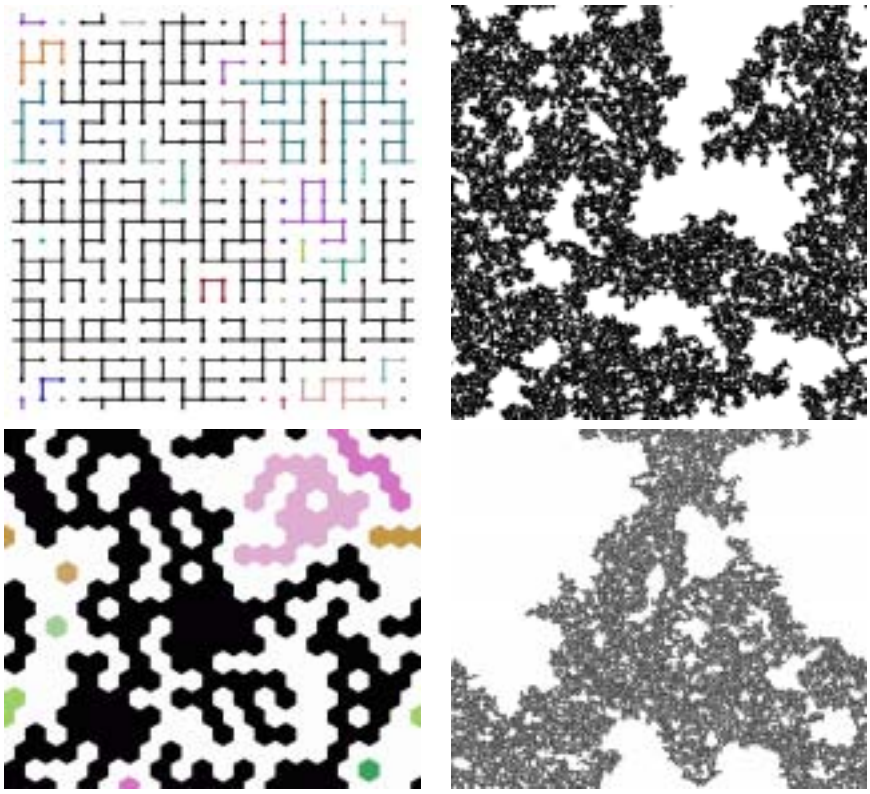


**Fig. 12.9 Universality in Percolation.** Universality suggests that the entire morphology of the percolation cluster at $p_c$ should be independent of microscopic details. On the top, we have bond percolation, where the bonds connecting nodes on a square lattice are occupied at random with probability $p$: the top right shows the infinite cluster on a $1024 \times 1024$ lattice at $p_c = 0.5$. On the bottom, we have site percolation on a triangular lattice, where it is the hexagonal sites that are occupied with probability $p = p_c = 0.5$. Even though the microscopic lattices and occupation rules are completely different, the resulting clusters look statistically identical. (One should note that the site percolation cluster is slightly less dark. Universality holds up to overall scale changes, here up to an change in the density.)

critical points, we say they are in the same *universality class*. Importantly, the theoretical Ising model (despite its drastic simplification of the interactions and morphology) is also in the same universality class as these experimental uniaxial ferromagnets and liquid–gas systems – allowing theoretical physics to be directly predictive in real experiments.

To get a more clear feeling about how universality arises, consider site and bond percolation in figure 12.9. Here we see two microscopically different systems (left) from which basically the same behavior emerges (right) on long length scales. Just as the systems approach the threshold of falling apart, they become similar to one another! In particular, all signs of the original lattice structure and microscopic rules have disappeared.[6]

Thus we observe in these cases that different microscopic systems look the same near critical points, if we ignore the microscopic details and confine our attention to long length scales. To study this systematically, we need a method to take a kind of continuum limit, but in systems which remain inhomogeneous and fluctuating even on the largest scales. This systematic method is called the *renormalization–group.*[7]

The renormalization group starts with a remarkable abstraction: it works in an enormous "system space". Different points in system space represent different materials under different experimental conditions, and different physical models of these materials with different interactions and evolution rules. So, for example, in figure 12.10 we can consider the space of all possible models for hysteresis and avalanches in three dimensional systems. There is a different dimension in this system space for each possible parameter in theoretical models (disorder, coupling, next–neighbor coupling, dipolar fields, etc.) and also for each parameter in an experiment (chemical composition, temperature, and annealing time). A given theoretical model, will traverse a line in system space as a parameter is varied: the line at the top of figure might represent our model of hysteresis and avalanches (exercise 8.12) as the strength of the disorder $R$ is varied.

The renormalization group studies the way in which system space maps into itself under *coarse–graining*. The coarse–graining operation shrinks the system and removes microscopic degrees of freedom. Ignoring the microscopic degrees of freedom yields a new physical system with identical long–wavelength physics, but with different (renormalized) values of the parameters.[8]

Under coarse–graining, we often find a fixed point $S^*$ to this mapping in system space. All the systems that flow into this fixed point under coarse–graining will share the same long–wavelength properties, and will hence be in the same universality class. The renormalization group

[6] Notice in particular the *emergent symmetries* in the problem. The large percolation clusters at $p_c$ are statistically both translation invariant and rotation invariant, independent of the grids that underly them. In addition, we'll see that there is an emergent *scale invariance* – a kind of symmetry connecting different length scales (as we also saw for random walks, figure 2.2).

[7] The word renormalization grew out of quantum electrodynamics, where the effective charge on the electron changes size (norm) as a function of length scale. The word group is usually thought to refer to the family of coarse–graining operations that underly the method (with the group product being repeated coarse–graining). However, there is no inverse operation to coarse–graining, so the renormalization group does not satisfy the definition of a mathematical group.

[8] Many detailed mathematical techniques have been developed to implement this coarse–graining operation – $\epsilon$-expansions, real–space renormalization groups, Monte–Carlo renormalization groups, *etc.* These methods are both approximate and technically challenging. Except for the simple case of the random walk in the exercises (12.6, 12.7) we will not cover the methods used to implement coarse–graining, focusing instead on the behavior that we can explain by hypothesizing the existence of renormalization–group flows and fixed points.

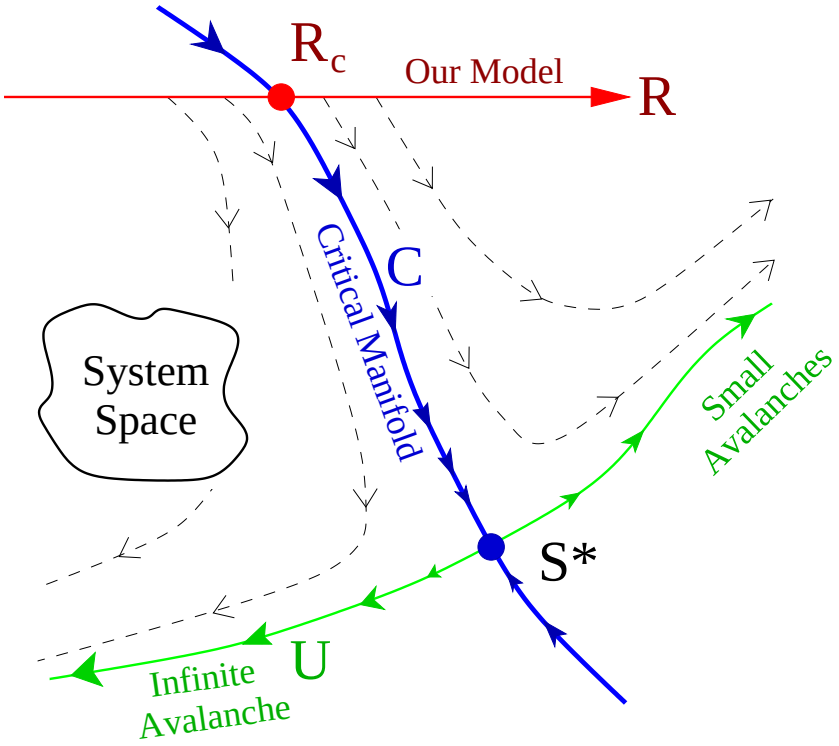



**Fig. 12.10 Renormalization–Group Flow.** The renormalization–group is the theory of how coarse–graining to longer length scales introduces a mapping from the space of physical systems into itself. Consider the space of all possible models of avalanches in hysteresis [109]. Each model can be coarse–grained, removing some fraction of the microscopic degrees of freedom and introducing more complicated rules so that the remaining domains still flip at the same external fields. This defines a mapping from system space into itself. A fixed point $S^*$ under this mapping will be self–similar (figure 12.14) because it maps into itself under a coarse–graining change in length scale. Points like $R_c$ that flow into $S^*$ will also show this self–similar behavior on scales large enough that they've flowed close to $S^*$: they all share the same *universality class*. Notice also that as our system approaches near $R_c$, the coarse–grained versions of the system flow alongside the unstable curve $U$ flowing out of $S^*$: for this reason, systems close to their critical point share universal properties too (figure 12.15).

explains the physical phenomenon of universality as the attraction of different physical systems to the same long–wavelength fixed-point theory, under a coarse–graining transformation.

Figure 12.10 depicts the flows in system space. It's a two–dimensional picture of an infinite–dimensional space. You can think of it as a planar cross–section in system space, which we've chosen to include the line for our model and the fixed point $S^*$: in this interpretation the arrows and flows denote projections, since the real flows will point somewhat out of the plane. Alternatively, you can think of it as the curved surface swept out by our model in system space as it coarse–grains, in which case you should ignore the parts of the figure below the roughly horizontal curve $U$.

Figure 12.10 shows the case of a fixed–point $S^*$ that has one unstable direction, leading outward along the curve $U$.[9] Points deviating from $S^*$ in that direction will not flow to it under coarse–graining, but rather will flow away from it. Fixed points with unstable directions correspond to continuous transitions between qualitatively different states. In the case of hysteresis and avalanches, there is a phase consisting of models where all the avalanches remain small, and another phase consisting of models where one large avalanche sweeps through the system, flipping most of the domains. The surface $C$ which flows into $S^*$ represents systems at their critical points: hence our model exhibits avalanches of all scales at $R_c$ where it crosses $C$.[10]

[9] The unstable manifold of the fixed point.

Other cases , with two tuning parameters to set to find the critical point (like the liquid–gas transition $[T_c, P_c]$), will have fixed–points with two unstable directions in system space. What happens when we have no unstable directions? The fixed point $S_a$ in figure 12.11 represents an entire region of system space that shares long–wavelength properties: it represents a phase of the system. Usually phases don't show fluctuations on all scales: such fluctuations arise near transitions because the system doesn't know which of the available neighboring phases to prefer. However, there are cases where the fluctuations persist even inside phases, leading to *generic scale invariance*. A good example is the case of the random walk (section 2.1, exercises 12.6 and 12.7), where a broad range of microscopic rules lead to the same long–wavelength random walks, where fluctuations remain important on all scales without tuning any parameters.

[10] Because $S^*$ has only one unstable direction, $C$ has one less dimension than system space (mathematically we say $C$ has *co-dimension* one) and hence can divide system space into two phases. $C$ is the *stable manifold*, or *basin of attraction*, for $S^*$.

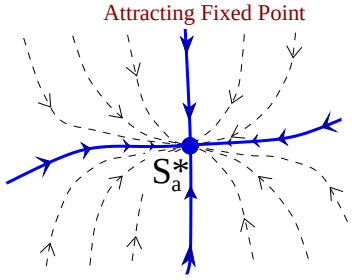



**Fig. 12.11 Attracting fixed point.** Often there will be fixed points that attract in all directions. These fixed points describe phases rather than phase transitions. Most phases are rather simple, with fluctuations that die away on long length scales. When fluctuations remain important, they will exhibit self–similarity and power laws called *generic scale invariance*.

Sometimes the external conditions acting on a system naturally drive it to stay near or at a critical point, allowing one to spontaneously observe fluctuations on all scales. A good example is provided by certain models of earthquake fault dynamics. The upper part of figure 12.12 shows the renormalization–group flows for these models of earthquakes. The horizontal axis represents the external stress on the earthquake fault. For small external stresses, the faults remain stuck, and there are no earthquakes. For strong external stresses, the faults slide with an average velocity $v$, with some irregularities but no large events. The earthquake fixed point $S^*_{\mathbf{eq}}$ describes the transition between the stuck and sliding phases, and shows earthquakes of all scales. The earth,

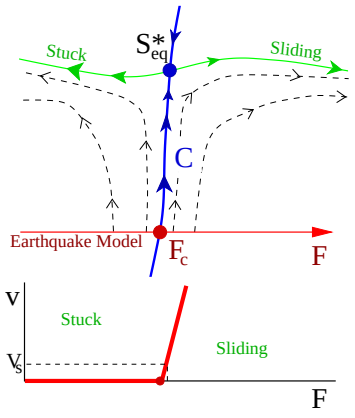



**Fig. 12.12 Self–Organized Criticality.** (Top) The renormalization–group flows in a space of earthquake models. The critical manifold $C$ separates a phase of stuck faults from a phase of sliding faults, with the transition due to the external stress $F$ across the fault. Only along $C$ does one find self–similar behavior and a spectrum of earthquakes. (Bottom) The velocity of the fault will vary as a power law $v \sim (F - F_c)^\beta$ near the critical force $F_c$. The motion of the continental plates, however, drives the fault at a constant, very slow velocity $v_s$, automatically setting $F$ to $F_c$ and yielding earthquakes of all sizes.

however, does not apply a constant stress to the fault: rather, continental drift applies a constant, extremely small velocity $v_s$ (of the order of centimeters per year). The lower part of figure 12.12 shows the velocity versus external force for this transition, and illustrates how forcing at a small external velocity naturally sets the earthquake model at its critical point – allowing spontaneous generation of critical fluctuations, called *self–organized criticality*.

## 12.2 Scale Invariance

The other striking feature of continuous phase transitions is the common occurrence of self–similarity, or scale invariance. We can see this vividly in the snapshots of the critical point in the Ising model (figure 12.1), percolation (12.3), and the avalanche in our hysteresis model (12.2). Each shows roughness, irregularities, and holes on all scales at the critical point. This roughness and fractal–looking structure stems at root from a hidden symmetry in the problem: these systems are (statistically) invariant under a change in *length scale*.

Consider the figures in 2.2, 12.13 and 12.14, depicting self–similarity in random walks, the 2D Ising model, and a cross section of our model of avalanches and hysteresis in 3D, respectively. In each set, the upper right figure shows a large system, and each succeeding picture zooms in by another factor of two. In the Ising model, some figures may look slightly more black (-1) than white (+1) and others vice versa, representing the inability of the Ising model at $T_c$ to decide on a ground state. All have patterns of white and black which have similar irregular boundaries, and similar holes within holes. In the hysteresis model, all the figures show a large avalanche spanning the system (black), with a variety of smaller avalanches of various sizes, each with the same kind of irregular boundary seen in 3D in figure 12.2. If you blur your eyes a bit, all the figures should look roughly alike. This rescaling and eye-blurring process are, at root, the renormalization–group transformation we discussed in section 12.1.

This scale–invariance can be thought of as a new symmetry that emerges at critical points. Almost as the expectation of any function of two positions $x_1$, $x_2$ in a translation–invariant system can be written in terms of the separation between the two points $\langle g(x_1, x_2)\rangle = \mathcal{G}(x_2 - x_2)$, scale invariance will allow us to write functions of $N$ variables in terms of *scaling functions* of $N-1$ variables – except that these scaling functions are typically multiplied by power laws in one of the variables.

Let's begin with the case of functions of one variable. Consider the avalanche size distribution $D(S)$ for a model, say real earthquakes (figure 12.4) or our model for hysteresis, at the critical point. Imagine taking the same system, but increasing the units of length with which we measure the system – stepping back, blurring our eyes, and looking at the system on a *coarse–grained* level.[11] Imagine that we multiply the spacing between markings on our rulers by a small amount $B = (1 + \epsilon)$.

[11] Figure 12.14 shows this done repeatedly by a factor of two: starting at the lower right at the most fine–grained level, and ending at the upper left at the coarsest.

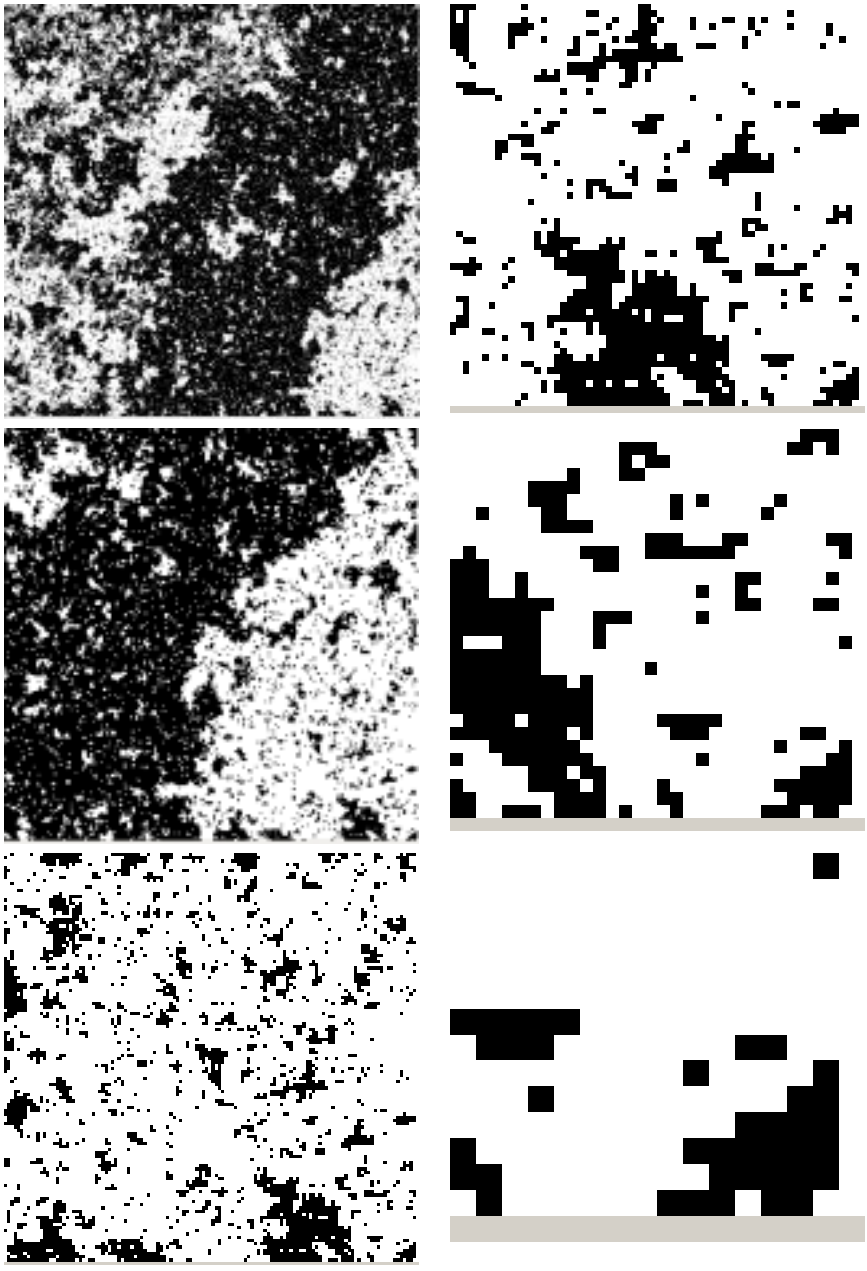


**Fig. 12.13 Ising Model at $T_c$: Scale Invariance.** Magnifications of a snapshot of the two dimensional Ising model 12.1 at its critical point, each one doubling the magnification ($B = 2$). (Each plot is the lower right–hand quarter of the previous.) At the largest magnifications, the individual lattice sites become apparent, but all effects of the lattice quickly disappear at larger magnifications. The larger scales look the same, in a statistical sense.

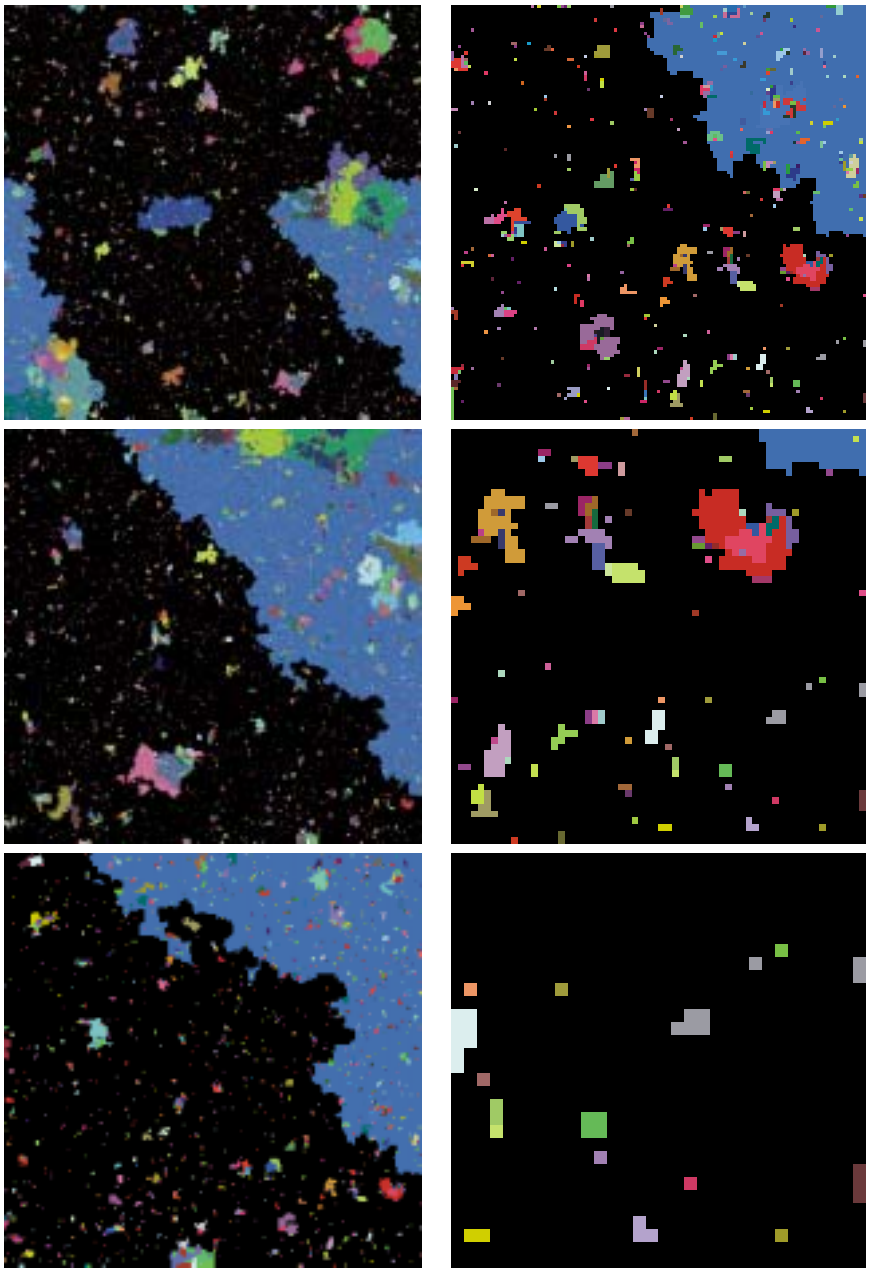


**Fig. 12.14 Avalanches: Scale Invariance.** Magnifications of a cross section of all the avalanches in a run of our hysteresis model, each one the lower right–hand quarter of the previous. The system started with a billion domains ($1000^3$). Each avalanche is shown in a different shade. Again, the larger scales look statistically the same. See exercises 8.12 and 12.10

After coarsening, any length scales in the problem (like the correlation length $\xi$) will be divided by $B$. The avalanche sizes $S$ after coarse–graining will also be smaller by some factor $C = (1 + c\epsilon)$.[12] Finally, the overall scale of $D(S)$ after coarse–graining will be rescaled by some factor $A = (1 + a\epsilon)$.[13] Hence under the coarse–graining

$$\begin{aligned} \xi' &= \xi/B = \xi/(1+\epsilon) \\ S' &= S/C = S/(1+c\epsilon) \\ D' &= AD = D(1+a\epsilon). \end{aligned} \tag{12.3}$$

Now the probability that the coarse–grained system has an avalanche of size $S'$ is given by the rescaled probability that the original system had an avalanche of size $S = (1 + c\epsilon)S'$:

$$D'(S') = AD(CS') = (1+a\epsilon)D\big((1+c\epsilon)S'\big). \tag{12.4}$$

$D'(S')$ is the distribution measured with the new ruler: a smaller avalanche with a larger relative probability. Because we are at a self–similar critical point, the coarse–grained distribution $D'(S')$ should equal $D(S')$. Making $\epsilon$ infinitesimal leads us to a differential equation

$$\begin{aligned} D(S') = D'(S') &= (1+a\epsilon)D\big((1+c\epsilon)S'\big) \\ 0 &= a\epsilon D + c\epsilon S' \frac{dD}{dS} \\ \frac{dD}{dS} &= -\frac{aD}{cS} \end{aligned} \tag{12.5}$$

which has the general solution[14]

$$D = D_0 S^{-a/c}. \tag{12.6}$$

Because the properties shared in a universality class only hold up to overall scales, the constant $D_0$ is system dependent. However, the exponents $a$, $c$, and $a/c$ are *universal* – independent of experiment within large universality classes of systems. Some of these exponents have standard names: the exponent $c$ giving the fractal dimension of the avalanche is usually called $d_f$ or $1/\sigma\nu$. The exponent $a/c$ giving the size distribution decay is called $\tau$ in percolation and in most models of avalanches in magnets.[15] and is related to the Gutenberg–Richter exponent in the earthquake community[16] (figure 12.5).

Most measured quantities depending on one variable will have similar power–law singularities at the critical point. Thus the correlation function of the Ising model at $T_c$ (figure 10.4) decays with distance $x$ in dimension $d$ as $C(x) \propto x^{-(d-2+\eta)}$ and the distance versus time for random walks (section 2.1) grows as $t^{1/2}$, both because these systems are self–similar.[17]

[12] This factor is precisely the kind of quantity that one calculates using the various technical implementations of the renormalization group. If the size of the avalanche were the cube of its length, then $c$ would equal 3 since $(1+\epsilon)^3 = 1+3\epsilon+O(\epsilon^2)$. The fact that $c$ can be less than the dimension tells us that in general the avalanches can be fractal (figure 12.2): $c$ will be the fractal dimension of the avalanche.

[13] The same avalanches occur independent of your measuring instrument, but the probability density $D(S)$ changes, because the fraction of large avalanches depends upon how many small avalanches you measure, and because the fraction per unit $S$ changes as the scale of $S$ changes.

[14] Since $\int dD/D = -a/c \int dS/S$, $\log D = K - (a/c)\log S$ for some integration constant $K = \log D_0$

[15] Except ours, where we used $\tau$ to denote the avalanche size decay at the critical field and disorder: integrated over the hysteresis loop $D_{\rm int} \propto S^{-\bar\tau}$ with $\bar\tau = \tau + \sigma\beta\delta$.

[16] We must not pretend that we've explained the Gutenberg–Richter law. There are many different models that give exponents $\approx 2/3$, but it remains controversial which of these, if any, are correct for real world earthquakes.

[17] Power laws are the only self–similar function! If $f(x) = x^{-\alpha}$, then on a new scale multiplying $x$ by $B$, $f(Bx) = B^{-\alpha}x^{-\alpha} \propto f(x)$.

Universality is expected also near to the critical point. Here as one coarsens the length scale a system will be statistically similar to itself *at a different set of parameters*. Thus a system undergoing phase separation (section 11.3.1, exercise 12.1), when coarsened, is similar to itself at an earlier time (when the domains were smaller), a percolation cluster just above $p_c$ (top figure 12.3) when coarsened is similar to one generated further from $p_c$ (hence with smaller holes), and so on.

For a magnet slightly below $T_c$, a system coarsened by a factor $B = (1+\epsilon)$ will be similar to one farther from $T_c$ by a factor $E = (1+\epsilon/\nu)$ (figure 12.15). Similar to the case of the avalanche size distribution, the coarsened system must have its magnetization rescaled upward by $A = (1+\beta\epsilon/\nu)$ to match that of the lower–temperature original magnet (figure 12.15):

$$\begin{aligned} M'(T_c - t) =& AM(T_c - t) = M(T_c - Et) \\ (1+\beta\epsilon/\nu)\, M(T_c - t) =& M\Big(T_c - t(1+\epsilon/\nu)\Big). \end{aligned} \tag{12.7}$$

[18] Notice that increasing $t$ decreases $T$ here.

Again, taking $\epsilon$ infinitesimal leads[18] us to the conclusion that $M \propto t^\beta$, providing a rationale for the power–laws seen in magnetism and the liquid–gas transition (figures 12.7 and 12.8). Similarly, the specific heat, correlation length, correlation time, susceptibility, and surface tension of an equilibrium system will have power–law divergences $(T - T_c)^{-X}$, where by definition $X$ is $\alpha$, $\nu$, $z\nu$, $\gamma$, and $-2\nu$, respectively. One can also vary the field $H$ away from the critical point and measure the resulting magnetization, which varies as $H^{1/\delta}$.

To specialists in critical phenomena, these exponents are central: whole conversations often rotate around various combinations of Greek letters. We know how to calculate critical exponents from the various analytical approaches,[19] and they are also simple to measure (although hard to measure well).

[19] They can be derived from the eigenvalues of the linearization of the renormalization–group flow around the fixed point $S^*$ in figure 12.10, see exercise 12.7.

Critical exponents are not everything, however. Many other scaling predictions are easily extracted from numerical simulations. Universality should extend even to those properties that we haven't been able to write formulas for. In particular, there are an abundance of functions of two and more variables that one can measure. Figure 12.16 shows the distribution of avalanche sizes $D_{\text{int}}(S, R)$ in our model of hysteresis, integrated over the hysteresis loop (figure 8.21), at various disorders $R$ above $R_c$ (exercise 8.12). Notice that only at $R_c \approx 2.16$ do we get a power–law distribution of avalanche sizes; at larger disorders there are extra small avalanches, and a strong decrease in the number of avalanches beyond a certain size $S_{\mathbf{max}}(R)$.

Let's derive the scaling form for $D_{\text{int}}(S, R)$. We first change from $a$ and $b$ to the traditional exponent names in our scaling equations 12.3:

$$\begin{aligned} S' &= S \Big/ \left(1 + \frac{1}{\sigma\nu}\epsilon\right) \\ D' &= D\left(1 + \frac{\bar{\tau}}{\sigma\nu}\epsilon\right) \end{aligned} \tag{12.8}$$

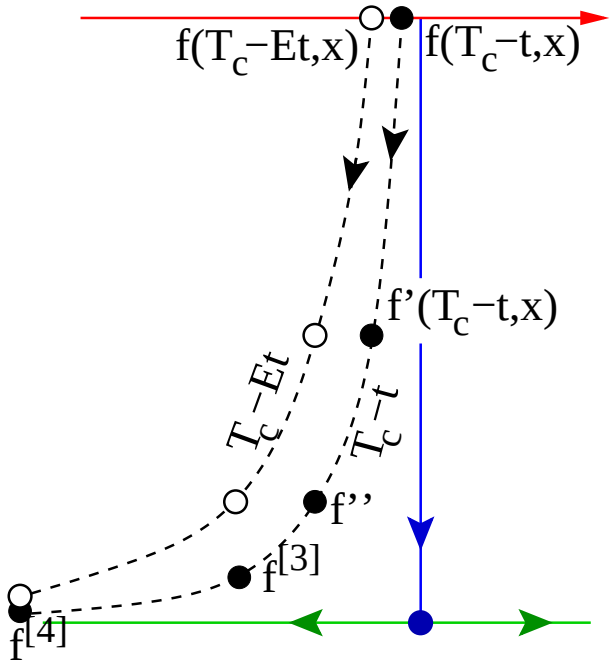



**Fig. 12.15 Scaling Near Criticality.** If two points in system space flow towards one another under coarse–graining, their behavior must be similar on long length scales. Here we measure a function $f(x)$ at two temperatures for our system (top line) at two different temperatures, $T_c - t$ and $T_c - Et$. The dots represent successive coarse–grainings by a factor $B$: under this renormalization group $f \to f' \to f'' \to f^{[3]} \ldots$. Here $f(T_c - t, x)$ after four coarse–grainings maps to nearly the same system as $f(T_c - Et, x)$ after three coarse–grainings. We thus know, on long length scales, that $f'(T_c - t, x)$ must agree with $f(T_c - Et, x)$: the system is similar to *itself* at a different set of external parameters. In particular, each coarse–graining step changes $x$ by a factor $B$ and $f$ by some factor $A$, so $f^{[n]}(T_c - t, x) = A^n f(T_c - t, B^n x)$. Equating $f^{[n]}(T_c - t, x) = f^{[n-1]}(T_c - Et, x)$ for large $n$, we find $A^n f(T_c - t, B^n x) = A^{n-1} f(T_c - Et, B^{n-1} x)$ implying that $Af(T_c - t, By) = f(T_c - Et, y)$ for large distances $y = B^{n-1} x$.

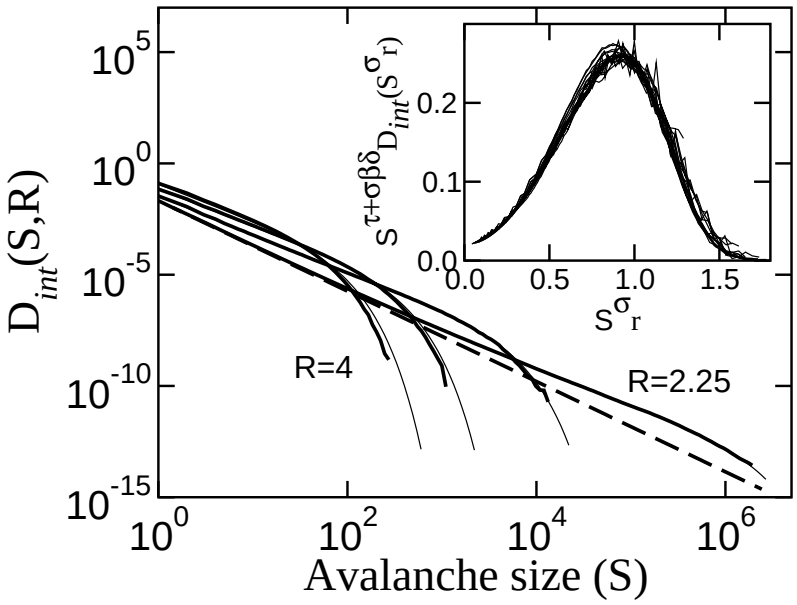



**Fig. 12.16 Avalanche Size Distribution:** The distribution of avalanche sizes in our model for hysteresis. Notice the logarithmic scales. (We can measure $D(S)$ of $10^{-14}$ by running billions of spins and binning over ranges $\Delta S$ $10^5$.) Although only at $R_c \approx 2.16$ do we get a pure power law ($D(S) \propto S^{-\bar{\tau}}$, straight dashed line) we get two decades of avalanche sizes even a factor of two away from the critical point. Notice that the power law at $R_c$ does not describe the data well except very close to the critical point: be warned that a power law measured over two decades of size (although often publishable [66]) will not yield a reliable exponent. Notice that the scaling curves (thin lines) work well even far from $R_c$. *Inset:* We plot $D(S)$ rescaled by $S^{\bar{\tau}}$ versus $S^{\sigma}(R - R_c)/R$ to extract the universal scaling curve $\mathcal{D}(X)$ (equation 12.15). Varying the critical exponents and $R_c$ to get a good collapse allows us to measure the exponents long before the power laws are reliable, see exercise 12.9, part (g).

A system at $R = R_c + r$ after coarse–graining will be similar to a system further from the critical disorder, at $R = R_c + Er = R_c + r(1+\epsilon/\nu)$, so

$$D(S', R_c + Er) = D'(S', R_c + r) = AD(CS', R_c + r) \tag{12.9}$$

$$D(S', R_c + (1+\epsilon/\nu)r) = \left(1 + \frac{\bar\tau}{\sigma\nu}\epsilon\right) D\left(\left(1 + \frac{1}{\sigma\nu}\epsilon\right) S', R_c + r\right) \tag{12.10}$$

To facilitate deriving the scaling form for multiparameter functions, it is helpful to change coordinates to *scaling variables*. Consider the combination $X = S^\sigma r$. After coarse–graining $S' = S/C$ and shifting to the higher disorder $r' = Er$ this combination is unchanged:

$$\begin{aligned} S'^\sigma r' &= (S/C)^\sigma (Er) = \left(S/\left(1 + \frac{1}{\sigma\nu}\epsilon\right)\right)^\sigma ((1+\epsilon/\nu)r) \\ &= S^\sigma r \left(\frac{1+\epsilon/\nu}{(1+\epsilon/\sigma\nu)^\sigma}\right) = S^\sigma r + O(\epsilon^2). \end{aligned} \tag{12.11}$$

Let $\bar D(S, X) = D(S, R)$ be the size distribution as a function of $S$ and $X$. Then $\bar D$ coarse–grains much like a function of one variable, since $X$ stays fixed. Equation 12.10 now becomes

$$\bar D(S', X') = \bar D(S', X) = \left(1 + \frac{\bar\tau}{\sigma\nu}\epsilon\right) \bar D\left(\left(1 + \frac{1}{\sigma\nu}\epsilon\right) S', X\right) \tag{12.12}$$

so

$$\frac{\bar\tau}{\sigma\nu}\bar D = -\frac{1}{\sigma\nu}\frac{\partial \bar D}{\partial S} \tag{12.13}$$

and hence

$$\bar D(S, X) = S^{-\bar\tau}\mathcal{D}(X) \tag{12.14}$$

for some *scaling function* $\mathcal{D}(X)$. This function corresponds to the (non–universal) constant $D_0$ in equation 12.6, except here the scaling function is another universal prediction of the theory (up to overall scale factors in $X$ and $D$). Rewriting things in terms of the original variables, we find the scaling form for the avalanche size distribution:

$$D(S, R) \propto S^{-\bar\tau}\mathcal{D}(S^\sigma(R - R_c)). \tag{12.15}$$

We can use a *scaling collapse* of the experimental or numerical data to extract this universal function, by multiplying $D$ by the power $S^\tau$ and plotting against $X = S^\sigma(R - R_c)$: the inset of figure 12.16 shows this scaling collapse.[20]

Similar universal scaling functions appear in many contexts. Just for the equilibrium properties of magnets near $T_c$, there are scaling functions for the magnetization $M(H, T) = (T_c - T)^\beta \mathcal{M}\left(H/(T_c - T)^{\beta\delta}\right)$,

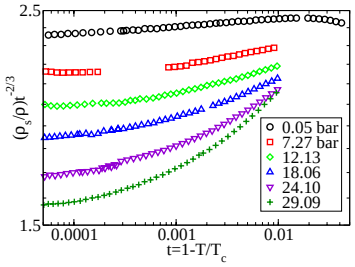



**Fig. 12.17 Superfluid density in Helium: scaling plot.** This classic experiment [40, 1] in 1980 measured the superfluid density $\rho_s(T)$ in helium to great precision to within a small distance from the critical temperature $T_c$. Notice the logarithmic scale on the horizontal axis: the lowest pressure data (saturated vapor pressure $\approx 0.0504$ bar) spans three decades of temperature shift from $T_c$. This plot emphasizes the deviations from the expected power law.

[20] Depending on the application, one may choose different forms of the scaling variable, or pull out powers of different parameters. Hence we could have used $Sr^{1/\sigma} = X^{1/\sigma}$ instead of $X$ as our scaling variable, and we also can write $D(S, R) \propto S^{-\bar\tau}(S^\sigma r)^{\bar\tau/\sigma}\frac{\mathcal{D}(S^\sigma r)}{(S^\sigma r)^{\bar\tau/\sigma}} = r^{\bar\tau/\sigma}\widetilde{\mathcal{D}}(S^\sigma r)$.

for the correlation function in space and time at different temperatures $C(x,t,T) = x^{-(2-d+\eta)}\mathcal{C}\Big(x/|T-T_c|^{-\nu}, t/|T-T_c|^{-z\nu}\Big)$, and a finite–size scaling function for the magnetization in a box of size $L$ with, say, periodic boundary conditions $M(T,L) = (T_c-T)^{\beta}\mathcal{M}\Big(L/(T_c-T)^{-\nu}\Big)$.

## 12.3 Examples of Critical Points.

Ideas from statistical mechanics have found broad applicability in sciences and intellectual endeavors far from their roots in equilibrium thermodynamic systems. The scaling and renormalization–group methods introduced in this chapter have seen a particularly broad range of applications: we will touch upon a few in this conclusion to our text.

### 12.3.1 Traditional Equilibrium Criticality: Energy versus Entropy.

Scaling and renormalization–group methods have of course been central to the study of continuous phase transitions in equilibrium systems. Ising models, Potts models,[21] Heisenberg models, phase transitions in liquid crystals (like the nematic to smectic-A transition), wetting transitions, equilibrium crystal shapes, two–dimensional melting, roughening (figure 11.8) and other phase transitions on surfaces – these are the grindstone problems on which our renormalization–group tools were sharpened.

The transition in all of these systems represents the competition between energy and entropy, with energy favoring order at low temperatures and entropy destroying it at high temperatures. Figure 12.17 shows the results of a classic, amazing experiment – the analysis of the superfluid transition in helium (the same order parameter, and hence the same universality class, as the XY model). The superfluid density is expected to have the form

$$\rho_s \propto (T_c - T)^{\beta}(1 + d(T_c - T)^{x}) \tag{12.16}$$

where $x$ is a universal, subdominant *correction to scaling*. Since $\beta \approx 2/3$, they plot $\rho_s/(T-T_c)^{2/3}$ so that deviations from the simple expectation are highlighted. The slope in the top, roughly straight curve reflects the difference between their measured value of $\beta = 0.6749 \pm 0.0007$ and their multiplier $2/3$. The other curves show the effects of the subdominant correction, whose magnitude $d$ increases with increasing pressure. Current experiments improving on these results are being done in the space shuttle, in order to reduce the effects of gravity.

### 12.3.2 Quantum Criticality: Zero-point fluctuations versus energy.

Thermal fluctuations do not exist at zero temperature, but there are many well studied *quantum phase transitions* which arise from the com-

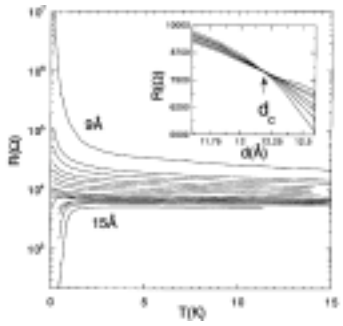



**Fig. 12.18 The superconductor–insulator transition.** [69] Thin films of amorphous bismuth are insulators (resistance grows to infinity at zero temperature), while films above about 12 Åwent superconducting (resistance goes to zero at a temperature above zero).

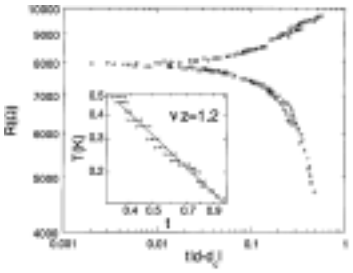



**Fig. 12.19 Scaling collapse at the superconductor–insulator transition.** [69] Resistance plotted against the temperature, rescaled by a factor $t$ to get a good collapse. Notice the characteristic pair of scaling curves, the top one $\mathcal{F}_-$ for the insulators $d < d_c$ and the bottom one $\mathcal{F}_+$ for the superconductors $d > d_c$.

[21] Potts models are Ising–like models with $N$ states per site rather than two.

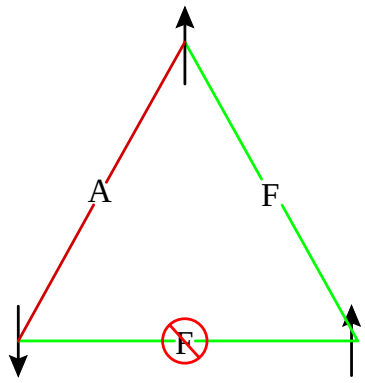



**Fig. 12.20 Frustration.** A spin glass has a collection of magnetic ions with interactions of random sign. Here we see a triangle of Ising $\pm 1$ spins with one antiferromagnetic bond – one of the three bonds must be unsatisfied in any spin configuration: the system is said to be *frustrated*. These systems go through a phase transition from a disordered paramagnet to a state that is disordered in space, but has long–range order in *time*: hence the name spin glass.

petition of potential energy and quantum fluctuations (from the kinetic energy). Many of the earliest studies focused on the metal–insulator transition and the phenomenon of *localization*, where disorder can lead to insulators even when there are states at the Fermi surface. Scaling and renormalization–group methods played a central role in this early work, and the states near the *mobility edge* separating localized from extended states are self–similar and fractal. The Kondo effect, macroscopic quantum tunneling (testing the fundamentals of quantum measurement theory), transitions between quantum Hall plateaus, and superconductor–normal metal transitions (the Bose Glass) are other active areas of research. Figures 12.18 and 12.19 show a recent experiment showing a transition directly from a superconductor to an insulator, as the thickness of the film is varied. The resistance is expected to have the scaling form

$$R(d,T) = R_c \mathcal{F}_{\pm}\Big((d-d_c)T^{-1/\nu z}\Big) \tag{12.17}$$

and they plot $R(d,T)$ versus $t(d-d_c)$, varying $t$ until it gives a nice collapse (main figure 12.19).

The resulting curve for $t(T)$ (inset, figure 12.19) forms a nice power–law, allowing them to measure $\nu z = 1.2 \pm 0.2$. While it is clear that scaling and renormalization–group ideas are applicable to this problem, we should note that as of the time this text was written, no theory yet convincingly explains these observations.

### 12.3.3 Glassy Systems: Random but Frozen.

Many materials form glasses. Window glasses are made of silicon dioxide and various other oxides, but sugars will also form glasses when cooled quickly (hard candy): you can even form metallic glasses by cooling metal alloys quickly. Glasses are out of equilibrium: their relaxation times diverge as they are cooled, and they stop rearranging at a typical temperature known as the glass transition temperature. Many other disordered systems – spin glasses, random–field Ising models, precursor tweed phases in martensites – also appear to be glassy, in that their relaxation times get very slow as they are cooled, and they freeze in to disordered configurations.[22] This freezing process is sometimes described as developing long–range order in time, or as a *broken ergodicity* (see section 4.2).

The basic reason that many of the glassy systems freeze into random states is *frustration*. Frustration was defined first for spin glasses, which are formed by randomly substituting magnetic atoms into a nonmagnetic host. The magnetic spins are coupled to one another at random – some pairs prefer to be parallel (ferromagnetic couplings) and some antiparallel (antiferromagnetic). Whenever strongly–interacting

[22] Glasses are different from disordered systems. The randomness in disordered systems is fixed, and occurs in both the high and low temperature phases: the disorder in the traditional *configurational* glasses freezes in as it cools.

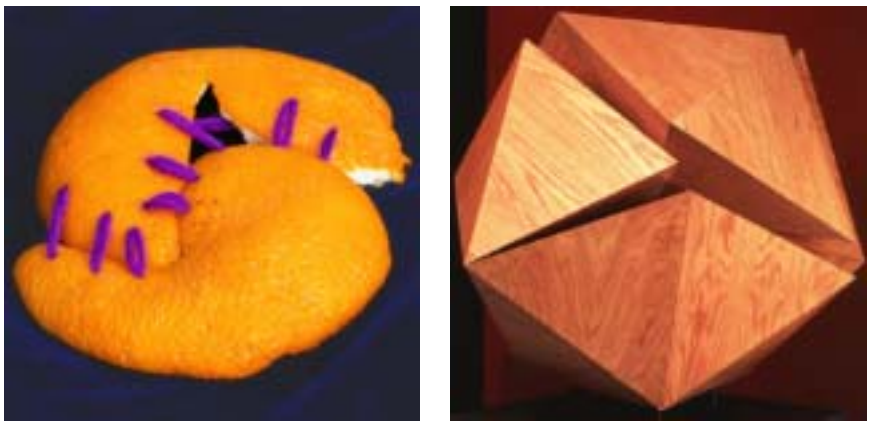


**Fig. 12.21 Frustration and Curvature.** Many materials without disorder are frustrated as well: sometimes this leads to glassy ground states [107], sometimes to exotic phases with complex structures [97]. One kind of frustration arises when the energetically favorable local packing of atoms or molecules is incompatible with the demands of building large structures: here we show two artistic renditions (courtesy Pamela Davis [52]). On the left we see the classic problem faced by mapmakers: the peel of an orange (or the crust of the earth) cannot be mapped smoothly onto a flat space without stretching and tearing it. On the left we see the analogous problem faced in many metallic–glass forming liquids whose atoms locally prefer to form nice compact tetrahedra: twenty tetrahedra cannot be glued together to form an icosahedron. Just as the orange peel can be nicely fit together on the sphere, the metallic glasses are unfrustrated in curved space [95].

spins form a loop with an odd number of antiferromagnetic bonds (figure 12.20) we have *frustration*: one of the bonds will have to be left in an unhappy state, since there must be an even number of spin inversions around the loop (figure 12.20).

Real configurational glasses are not as fully understood as systems with frozen–in disorder. It is believed in many cases, however, that frustration is also important for configurational glasses (figure 12.21).

The statistical mechanical study of glasses and glassy systems remains a developing and controversial field. In spin glasses, the mathematical and computational sophistication is high, but basic conceptual questions remain: it is acknowledged that there is a transition to a true spin–glass state, but we still don't know whether the ground state of the spin–glass in an infinite system should be viewed as being a unique spin configuration (up to inversion of all spins) or many competing configurations: cluster and replica methods based on these two different assumptions remain in competition. For configurational glasses, it remains an important and open question whether there is an underlying thermodynamic glass transition [111] or whether the observed transition is a kind of jamming transition in the dynamics.

What is clear both theoretically and experimentally is that these systems in practice do fall out of equilibrium, and a given cooling schedule will result in a large ensemble of low–temperature metastable states



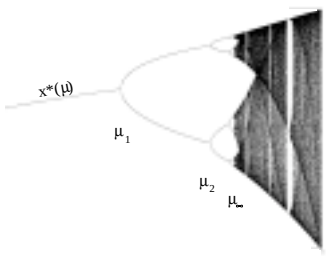



**Fig. 12.22 Bifurcation diagram for period doubling cascade.** The attractor as a function of $\mu$ for the Feigenbaum logistic map $f(x) = 4\mu x(1-x)$. For small $\mu < \mu_1$, repeatedly iterating $f(f(f(\ldots(f(x)))))\ldots$ converges to a stable fixed point $x^*(\mu)$. In a series of bifurcations, this stable fixed point shifts into a stable two–cycle at $\mu_1$, a four–cycle at $\mu_2$, an eight–cycle at $\mu_3$ ..., where the $2^n$-cycle passes through all $2^n$ points in a little dance before repeating. This series of *period doubling* bifurcations come more and more closely together, until at $\mu_\infty$ the system goes into a chaotic state. At the onset of chaos, this attractor shows a self–similar scaling: the bifurcations converge geometrically so that

$$\mu_\infty - \mu_n \propto \delta^n \tag{12.18}$$

where $\delta = 4.669201609102990\ldots$ is a universal constant, shared by for a

(with a considerable residual entropy, see exercise 5.8). There is one particular disordered magnetic system, the random–field Ising model, where it's been proven (in three dimensions, for weak disorder) that the ground state is ferromagnetic, and it is understood in detail how the system nonetheless gets stuck in metastable states, as the coarsening of a quenched unmagnetized state develops diverging energy barriers as it is cooled. Our model for hysteresis and avalanches described earlier (and in exercise 8.12) describes the transitions between these metastable states in the random–field Ising model, except that our transitions are not driven by temperature, but by ramping up an external field.

### 12.3.4 Dynamical Systems and the Onset of Chaos.

Much of statistical mechanics focuses on systems with large numbers of particles, or systems connected to a large external environment. Continuous transitions also arise in isolated or simply driven systems with few relevant degrees of freedom, where they are called *bifurcations*. A bifurcation is a qualitative change in behavior which arises when a parameter in a set of differential equations passes through a critical value. The study of these bifurcations is the theory of *normal forms* (exercise 12.3). Bifurcation theory contains analogies to universality classes, critical exponents, and analytic corrections to scaling.

Dynamical systems, even when they contain only a few degrees of freedom, can exhibit immensely complex, chaotic behavior (exercise 4.2).[23] The mathematical trajectories formed by chaotic systems at late times – the *attractors* – are often fractal in structure, and many concepts and methods from statistical mechanics are useful in studying these sets.[24]

The place where renormalization–group methods have had a spectacular impact, however, is in the study of the onset of chaos. Figure 12.22 shows a simple dynamical system undergoing a series of bifurcations leading to a chaotic state. Feigenbaum (exercise 12.8) analyzed the series using a renormalization group, coarse graining not in space but in *time*. Again, this behavior is universal – exactly the same series of bifurcations (up to smooth coordinate changes) arise in other maps and in real physical systems. Other renormalization–group calculations have been important for the study of the transition to chaos from quasiperiodic motion, and for the chaotic breakdown of the last KAM torus in Hamiltonian systems (see exercise 4.2).

[23] Indeed, it is likely that chaos in large systems is the fundamental underpinning both for the increase of entropy 5.4 and of the validity of the ergodic hypothesis that we can average properties over the energy surface (section 4.2).

## Exercises

[24] For example, statistical mechanical ensembles become *invariant measures*, and the attractors are characterized by various kinds of entropy.

(12.1) **Scaling: Critical Points and Coarsening.**

Near critical points, the self-similarity under rescaling leads to characteristic power-law singularities. These dependences may be disguised, however, by less-singular corrections to scaling. You may find Yeoman's tables 2.3, 2.4, and 3.1 helpful in defining exponents.

*(a) An experiment measures the susceptibility $\chi(T)$ in a magnet for temperatures $T$ slightly above the ferromagnetic transition temperature $T_c$. They find their data is fit well by the form*

$$\chi(T) = A(T-T_c)^{-1.25} + B + C(T-T_c) + D(T-T_c)^{1.77}. \tag{12.19}$$

*Assuming this is the correct dependence near $T_c$, what is the critical exponent $\gamma$?*

When measuring functions of two variables near critical points, one finds universal scaling functions. The whole function is a prediction of the theory!

*(b) The pair correlation function $C(r,T) = \langle S(x)S(x+r)\rangle$ is measured in another, three-dimensional system just above $T_c$. It is found to be spherically symmetric, and of the form*

$$C(r,T) = r^{-1.026} f(r(T-T_c)^{0.59}), \tag{12.20}$$

*where the function $f(x)$ is found to be roughly* $\exp(-x)$. *What is the critical exponent $\nu$? The exponent $\eta$?*

During coarsening, we found that the system changed with time, with a length scale that grows as a power of time: $L(t) \sim t^{1/2}$ for a non-conserved order parameter, and $L(t) \sim t^{1/3}$ for a conserved order parameter. These exponents, unlike critical exponents, are simple rational numbers that can be derived from arguments akin to dimensional analysis. Associated with these diverging length scales there are scaling functions. Coarsening doesn't lead to a system which is self-similar to itself at equal times, but it does lead to a system which at two different times looks the same – apart from a shift of length scales.

An Ising model with non-conserved magnetization is quenched to a temperature $T$ well below $T_c$. After a long time $t_0$, the correlation function looks like $C(\mathbf{r}, T, t_0) = c(\mathbf{r})$.

*(c) Assume that the correlation function at short distances $C(0,T,t)$ will be time independent, and that the correlation function at later times will have the same functional form apart from a rescaling of the length. Write $C(\mathbf{r}, T, 2t_0)$ in terms of $c(\mathbf{r})$. Write a scaling form*

$$C(\mathbf{r}, T, t) = t^{-\omega} \mathcal{C}(\mathbf{r}/t^{\rho}, T). \tag{12.21}$$

*Use the time-independence of $C(0,T,t)$ and the fact that the order parameter is not conserved to predict the numerical values of the exponents $\omega$ and $\rho$.*

It was only recently made clear that the scaling function $\mathcal{C}$ for coarsening *does* depend on temperature (and is, in particular, anisotropic for low temperature, with domain walls lining up with lattice planes). Low-temperature coarsening isn't as "universal" as continuous phase transitions are: even in one model, different temperatures can have different scaling functions.

(12.2) **RG Trajectories and Scaling.**

An Ising model near its critical temperature $T_c$ is described by two variables: the distance to the critical temperature $t = (T-T_c)/T_c$, and the external field $h = H/J$. Under a coarse-graining of length $x' = (1-\epsilon)\,x$, the system is observed to be similar to itself [25] at a shifted temperature $t' = (1+a\epsilon)\,t$ and a shifted external field $h' = (1+b\epsilon)\,h$, with $b > a > 0$ (so there are two relevant eigendirections, with the external field more relevant than the temperature).

*(a) Which diagram from figure 12.23 has curves consistent with this flow, for $b > a > 0$?*

The magnetization $M(t,h)$ is observed to rescale under this same coarse-graining operation to $M' = (1+c\epsilon)\,M$, so $M\left((1+a\epsilon)\,t, (1+b\epsilon)\,h\right) = (1+c\epsilon)\,M(t,h)$.

*(b) Suppose $M(t,h)$ is known at $t = t_1$, the line of filled circles in the various figures in part (a). Give a formula for $M(2t_1, h)$ (open circles) in terms of $M(t_1, h')$.* (Hint: Find the scaling variable in terms of $t$ and $h$ which is constant along the renormalization-group trajectories shown in (a). Write a scaling form for $M(t,h)$ in terms of this scaling variable, and find the critical exponents in terms of $a$, $b$, and $c$. From there calculating $M(t,h)$ at $t = 2t_1$ should be possible, given the values at $t = t_1$.)

(12.3) **Bifurcation Theory and Phase Transitions.** (Mathematics)

[25] By emphasizing self-similarity, we can view these flows in the two-dimensional space of the original model – but the self-similarity is only approximate, valid sufficiently near to the critical point. The renormalization-group flows are in the infinite-dimensional space of Hamiltonians, and the irrelevant directions perpendicular to $t$ and $h$ cause corrections far from $t = (T-T_c)/T_c = h = 0$. It's equivalent to view the flows in part (a) as the flows on the unstable manifold of the renormalization-group fixed point, if you prefer.

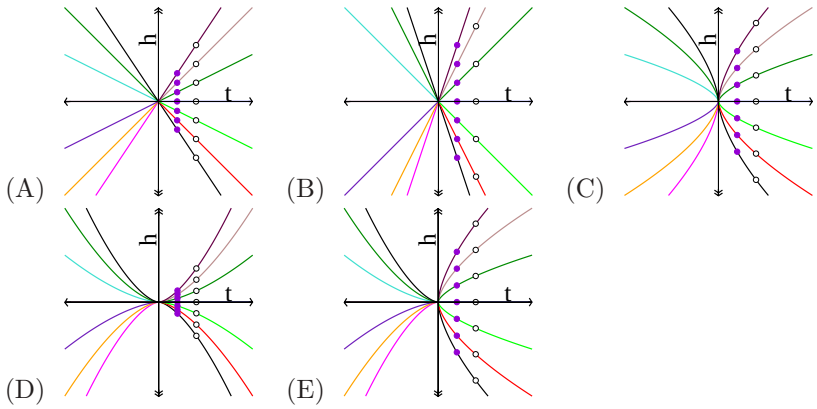



**Fig. 12.23**

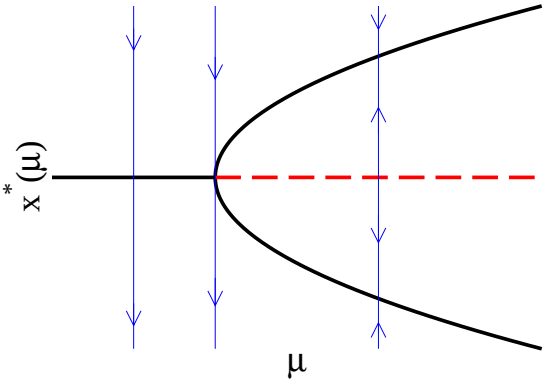



**Fig. 12.24 Pitchfork Bifurcation diagram.** The flow diagram for the pitchfork bifurcation (equation 12.23). The dashed line represents unstable fixed points, the solid thick lines stable fixed points. The thin lines and arrows represent the dynamical evolution directions. It's called a pitchfork because of the three tines on the right emerging from the handle on the left.

Dynamical systems theory is the study of the time evolution given by systems of differential equations. Let $\mathbf{x}(t)$ be a vector of variables evolving in time $t$, let $\boldsymbol{\lambda}$ be a vector of parameters governing the differential equation, and let $\mathbf{F}_{\boldsymbol{\lambda}}(\mathbf{x})$ be the differential equations[26]

$$\dot{\mathbf{x}} = \mathbf{F}_{\boldsymbol{\lambda}}(\mathbf{x}). \tag{12.22}$$

The typical focus of the theory is not to solve the differential equations for general initial conditions, but to study the qualitative behavior. In general, they focus on *bifurcations* – special values of the parameters $\boldsymbol{\lambda}$ where the behavior of the system changes qualitatively.

*(a)* **Pitchfork.** *Consider the differential equation in one variable $x(t)$ with one parameter $\mu$*

$$\dot{x} = \mu x - x^3. \tag{12.23}$$

*Show that there is a bifurcation at $\mu_c = 0$, by showing that an initial condition with large $x(0)$ will evolve qualitatively differently at late times for $\mu > 0$ versus for $\mu < 0$.* Hint: Although you can solve this differential equation explicitly, we recommend instead that you argue this qualitatively from the bifurcation diagram in figure 12.24: a few words should suffice.

Dynamical systems theory has much in common with equilibrium statistical mechanics of phases and phase transitions. The liquid–gas transition is characterized by external parameters $\boldsymbol{\lambda} = (P, T, N)$, and has a current state described by $\mathbf{x} = (V, E, \mu)$. Equilibrium phases correspond to fixed points ($x^*(\mu)$ with $\dot{x}^* = 0$) in the dynamics, and phase transitions correspond to bifurcations.[27] For example, the power laws we find near continuous phase transitions have simpler analogues in the dynamical systems.

*(b) Find the critical exponent $\beta$ giving the size of the non–zero fixed points as a function of $\mu$, $x^*(\mu) \propto (\mu - \mu_c)^\beta$.*

[26] As before, we use $\dot{x} = \frac{dx}{dt}$. This form of the differential equation can also be used to encode second–order differential equations like $\ddot{x} = -\omega^2 x$: one doubles the number of variables to $\dot{x} = v$, $\dot{v} = -\omega^2 x$.

[27] In section 8.1, we noted that inside a phase all properties are analytic in the parameters. Similarly, bifurcations are values of $\lambda$ where nonanalyticities in the long–time dynamics are observed.

Just as there are universality classes of continuous phase transitions, where quite different systems scale in the same way near the critical points, quite different dynamical systems can have the same behaviors near bifurcation points.

*(c) At what value $\lambda_c$ does the differential equation*

$$\dot{m} = m - \tanh(\lambda m). \tag{12.24}$$

*have a bifurcation? Does the fixed–point value $m^*(\lambda)$ behave as a power–law $m^* \sim |\lambda - \lambda_c|^\beta$ near $\lambda_c$ (up to corrections with higher powers of $\lambda - \lambda_c$)? Does the value of $\beta$ agree with that of the pitchfork bifurcation?*

There is a common approximate method for studying phase transitions called *mean–field theory*, where the fluctuating environment of each degree of freedom is replaced by the average behavior. In the Ising model on a square lattice, this amounts to assuming each spin $s_j = \pm 1$ has four neighbors which are magnetized with the average magnetization $m = \langle s_j \rangle$. This leads to a one–spin mean-field Hamiltonian

$$\mathcal{H} = -4Jms_j. \tag{12.25}$$

*(d) At temperature $k_BT$, what is the value for $\langle s_j \rangle$ in equation 12.25, given $m$? At what temperature $T_c$ is the phase transition, in mean field theory?* (Hint: at what temperature is a non–zero $m$ self–consistent?) *Argue from part (c) that $m \propto (T_c - T)^\beta$. Is this value for the critical exponent $\beta$ correct for the Ising model in either two dimensions ($\beta = 1/8$) or three dimensions ($\beta \approx 1/3$)?*[28]

Just as there are different universality classes for continuous phase transitions with different renormalization–group fixed points, there are different classes of bifurcations each with its own *normal form*. Some of the other important normal forms include the saddle–node bifurcation,

$$\dot{x} = \mu - x^2, \tag{12.26}$$

transcritical exchange of stability,

$$\dot{x} = \mu x - x^2, \tag{12.27}$$

and the Hopf bifurcation,

$$\begin{aligned} \dot{x} &= (\mu - (x^2 + y^2))x - y \\ \dot{y} &= (\mu - (x^2 + y^2))y + x. \end{aligned} \tag{12.28}$$

(12.4) **Onset of Lasing as a Critical Point.** (Quantum, Mathematics) (Thanks Gaeta, Sievers. [34])

Lasers represent a stationary, condensed state. It is different from a phase of matter not only because it's made up out of energy, but also because it's intrinsically a non–equilibrium state. In a laser entropy is not maximized, free energies are not minimized – and yet the state has a robustness and integrity reminiscent of phases in equilibrium systems.

In this exercise, we'll study a system of excited atoms coupled to a photon mode just before it begins to lase. We'll see that it exhibits the diverging fluctuations and scaling that we've studied near critical points.

Let's consider a system of atoms weakly coupled to a photon mode. We assume that $N_1$ atoms are in a state with energy $E_1$, $N_2$ atoms are in a higher energy $E_2$, and that these atoms are strongly coupled to some environment that keeps these populations fixed.[29] Below the onset of lasing, the probability $\rho_n(t)$ that the photon mode is occupied by $n$ photons obeys

$$\begin{aligned} \frac{d\rho_n}{dt} = a\big(&n\rho_{n-1}N_2 - n\rho_n N_1 - (n+1)\rho_n N_2 \\ &+ (n+1)\rho_{n+1}N_1\big). \end{aligned} \tag{12.29}$$

The first term on the right–hand side represents the rate at which one of the $N_2$ excited atoms atom experiencing $n-1$ photons will emit a photon; the second term represents the rate at which one of the $N_1$ lower–energy atoms will absorb one of $n$ photons; the third term represents emission in an environment with $n$ photons, and the last represents absorption with $n+1$ photons. The fact that absorption in the presence of $m$ photons is proportional

[28] Note that $m(T)$ is a power law only to lowest order near $T_c$: in phase transitions there are *corrections to scaling*. In normal form theory, one can show that the dynamics for a general system exhibiting a pitchfork bifurcation as $\mu$ is varied can be transformed by a change of coordinates $\mathbf{y}(\mathbf{x}, \mu)$ into the normal form given by equation 12.23. If there are many coordinates $\mathbf{x}$ (equation 12.22), the extra coordinates all exponentially shrink to zero as time elapses. This smooth coordinate transform introduces the extra subdominant power laws found in part (c). In phase transitions, these smooth coordinate transformations give *analytic* corrections to scaling; there are also *singular* corrections to scaling that arise from irrelevant flows in the renormalization group that do not have analogues in bifurcation theory.

[29] That is, we assume that the atoms are being pumped into state $N_2$ to compensate for both decays into our photon mode and decays into other channels. This usually involves exciting atoms into additional atomic levels.

to $m$ and emission is proportional to $m+1$ is a property of bosons (exercises 7.7 and 7.8(c)). The constant $a > 0$ depends on the lifetime of the transition, and is related to the Einstein A coefficient (exercise 7.8).

*(a) Find a simple expression for* $\frac{d\langle n\rangle}{dt}$*, where* $\langle n\rangle = \sum_{m=0}^{\infty} m\rho_m$ *is the mean number of photons in the mode.* (Hint: Collect all terms involving $\rho_m$.) *Show that for* $N_2 > N_1$ *that this mean number grows indefinitely with time, leading to a macroscopic occupation of photons into this single state – a laser.*[30]

**Diverging Correlation Energy and Critical Slowing Down.** Now, let us consider our system just before it begins to lase. Let $\epsilon = (N_2 - N_1)/N_1$ be our measure of how close we are to the lasing instability. We might expect the value of $\langle n\rangle$ to diverge as $\epsilon \to 0$ like $\epsilon^{-\nu}$ for small $\epsilon$. Near a phase transition, one also normally observes *critical slowing down*: to equilibrate, the phase must communicate information over large distances of order the correlation length, which takes a time which diverges as the correlation length diverges. Let us define a critical exponent $\zeta$ for our lasing system, so that the typical relaxation time is proportional to $|\epsilon|^{-\zeta}$ as $\epsilon \to 0$.

*(b) For* $\epsilon < 0$*, below the instability, solve your equation from part (a) for the long–time stationary value of* $\langle n\rangle$*. What is* $\nu$ *for our system? For a general initial condition for the mean number of photons, solve for the time evolution. It should decay to the long–time value exponentially. Does the relaxation time diverge as* $\epsilon \to 0$*? What is* $\zeta$*?*

*(c) Solve for the stationary state* $\boldsymbol{\rho}^*$ *for* $N_2 < N_1$*.* (Your formula for $\boldsymbol{\rho}^*_n$ should not involve $\rho^*$.) *If* $N_2/N_1$ *is given by a Boltzmann probability at temperature* $T$*, is* $\boldsymbol{\rho}^*$ *the thermal equilibrium distribution for the quantum harmonic oscillator at that temperature?* Warning: The number of bosons in a phonon mode is given by the Bose–Einstein distribution, but the probability of different occupations in a quantum harmonic oscillator is given by the Boltzmann distribution (see section 7.2 and exercise 7.9).

We might expect that near the instability the probability of getting $n$ photons might have a scaling form

$$\rho^*_n(\epsilon) \sim n^{-\tau}\mathcal{D}(n|\epsilon|^{\nu}). \tag{12.30}$$

*(d) Show, for small* $\epsilon$*, that there is a scaling form for* $\boldsymbol{\rho}^*$*, with corrections that go to zero as* $\epsilon \to 0$*, using your answer to part (c). What is* $\tau$*? What is the function* $\mathcal{D}(x)$*?* (Hint: In deriving the form of $\mathcal{D}$, $\epsilon$ is small, but $n\epsilon^{\nu}$ is of order one. If you were an experimentalist doing scaling collapses, you'd plot $n^{\tau}\rho_n$ versus $x = n|\epsilon|^{-\nu}$; try changing variables in $n^{\tau}\rho_n$ to replace $\epsilon$ by $x$, and choose $\tau$ to eliminate $n$ for small $\epsilon$.)

(12.5) **Superconductivity and the Renormalization Group.**

This exercise touches upon many advanced topics. There's a lot of background discussion, which is largely irrelevant to doing the exercise. It's OK to skip to the questions if you find the motivations confusing.

Ordinary, low–temperature superconductivity is due to the interaction of electrons with phonons. The phonons produce an effective attractive interaction between certain pairs of electrons which in the end leads to superconductivity. These phonon–mediated interactions are confined to those electrons within a region $\hbar\omega_D$ in energy around the Fermi energy $\epsilon_F$; here $\omega_D$ is a measure of the highest phonon frequency.

Ordinary superconductivity happens at a rather low temperature: in contrast to to phonon energies (hundreds of degrees Kelvin times $k_B$) or electronic energies (tens of thousands of degrees Kelvin), phonon–mediated superconductivity in most materials happens below a few Kelvin. This is largely explained by the BCS theory of superconductivity, which predicts that the transition temperature for weakly coupled superconductors is

$$T_c = 1.764\hbar\omega_D e^{-\frac{1}{Vg(\epsilon_F)}} \tag{12.31}$$

where $V$ is an attraction between electron pairs mediated by the phonons, and $g(\epsilon_F)$ is the density of states (DOS) of the electron gas (equation 7.74) at the Fermi energy. If $V$ is small, $e^{-\frac{1}{Vg(\epsilon_F)}}$ can be exponentially small, explaining why materials often have to be so cold to go superconducting.

Superconductivity was discovered decades before it was explained. Many looked for explanations which would involve interactions with phonons, but there was a serious obstacle. People had studied the interactions of phonons with electrons, and had shown that the system stays metallic (no superconductivity) *to all orders in perturbation theory.*

*(a) Taylor expand* $T_c$ *(equation 12.31) about* $V = 0^+$ *(about infinitesimal positive* $V$*). Guess or show the value of all the terms in the Taylor series. Can we expect to explain superconductivity at positive temperatures by perturbing in powers of* $V$*?*

There are two messages here.

[30] The number of photons will eventually stop growing when they begin to pull energy out of the $N_2$ excited atoms faster than the pumping can replace them – invalidating our equations.

- Proving something to all orders in perturbation theory doesn't make it true.
- Since phases are regions in which perturbation theory converges (see section 8.1), the theorem is not a surprise. It's a condition for a metallic phase with a Fermi surface to exist at all.

In recent times, people have developed a renormalization–group description of the Fermi liquid state and its instabilities.[31] Discussing Fermi liquid theory, the BCS theory of superconductivity, or this renormalization–group description would take us far into rather technical subjects. However, we can illustrate all three by analyzing a rather unusual renormalization–group flow.

Roughly speaking, the renormalization–group treatment of Fermi liquids says that the Fermi surface is a fixed point of a coarse–graining in *energy*. That is, they start with a system space consisting of a partially–filled band of electrons with an energy width $W$, including all kinds of possible electron–electron repulsions and attractions. They coarse–grain by perturbatively eliminating (integrating out) the electronic states near the edges of the band,

$$W' = (1 - \delta)W. \tag{12.32}$$

incorporating their interactions and effects into altered interaction strengths among the remaining electrons. These altered interactions give the renormalization–group flow in the system space. The equation for $W$ gives the change under one iteration ($n = 1$): we can pretend $n$ is a continuous variable and take $\delta n \to 0$, so $\frac{W'-W}{\delta} \to \frac{dW}{dn}$, and hence

$$dW/dn = -W. \tag{12.33}$$

When they do this calculation, they find

- The non–interacting Fermi gas we studied in section 7.7 is a *fixed–point of the renormalization group*. All interactions are zero at this fixed point. Let $V$ represent one of these interactions.[32]
- The fixed point is unstable to an attractive interaction $V > 0$, but is stable to a repulsive interaction $V < 0$.
- Attractive forces between electrons grow under coarse–graining and lead to new phases, but repulsive forces shrink under coarse–graining, leading back to the metallic free Fermi gas.

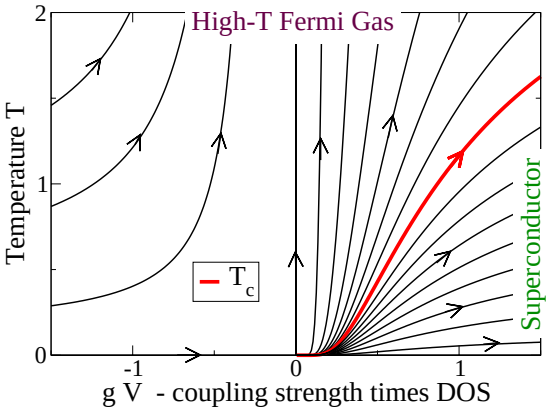



**Fig. 12.25** The renormalization flows defined by equations 12.34 and 12.35. The temperature $T$ is relevant at the free Fermi gas fixed point; the coupling $V$ is marginal. The distinguished curve represents a phase transition boundary $T_c(V)$. Below $T_c$, for example, the system is superconducting; above $T_c$ it is a (finite–temperature) metal.

This is quite different from our renormalization group treatment of phase transitions, where *relevant* directions like the temperature and field were unstable under coarse graining, whether shifted up or down from the fixed point, and other directions were *irrelevant* and stable (figure 4 in the Nature paper). For example, the temperature of our Fermi gas is a relevant variable, which rescales under coarse–graining like

$$\begin{aligned} T' &= (1 + a\delta)T \\ dT/dn &= aT. \end{aligned} \tag{12.34}$$

Here $a > 0$, so the effective temperature becomes larger as the system is coarse–grained. How can they get a variable $V$ which grows for $V > 0$ and shrinks for $V < 0$?

- When they do the coarse–graining, they find that the interaction $V$ is *marginal*: to linear order it neither increases nor decreases.

The next allowed term in the Taylor series near the fixed point gives us the coarse–grained equation for the interaction

$$\begin{aligned} V' &= (1 + b\delta V)V \\ dV/dn &= bV^2. \end{aligned} \tag{12.35}$$

[31] There are also other instabilities of Fermi liquids. Charge–density waves, for example, form due to electron–phonon coupling, and they too have the characteristic $e^{-\frac{1}{aV}}$ dependence on the coupling $V$.

[32] $V$ will be the pairing between opposite–spin electrons near the Fermi surface for superconductors.

- They find $b > 0$.

*(b) True or false? (See figure 12.25.)*

*(T) (F) For $V > 0$ (attractive interactions), the interactions get stronger with coarse–graining.*

*(T) (F) For $V < 0$ (repulsive interactions), coarse–graining leads us back to the free Fermi gas, explaining why the Fermi gas describes metals (section 7.7).*

*(T) (F) Temperature is an irrelevant variable, but dangerous.*

*(T) (F) The scaling variable*

$$x = TV^{-\beta\delta} \tag{12.36}$$

*is unchanged by the coarse–graining (second equations in 12.34 and 12.35), where $\beta$ and $\delta$ are universal critical exponents:*[33] *hence $x$ labels the progress along the curves in figure 12.25 (increasing in the direction of the arrows).*

*(T) (F) The scaling variable*

$$y = Te^{\frac{a}{bV}} \tag{12.37}$$

*is unchanged by the coarse–graining, so each curve in figure 12.25 has a fixed value for $y$.*

Now, without knowing anything about superconductivity, let's presume that our system goes superconducting at some temperature $T_c(V)$ when the interactions are attractive. When we coarse–grain a system that is at the superconducting transition temperature, we must get another system that is at its superconducting transition temperature.

*(c) What value for $a/b$ must they calculate in order to get the BCS transition temperature (equation 12.31) from this renormalization group? What is the value of the scaling variable (whichever you found in part (b)) along $T_c(V)$?*

Thus the form of the BCS transition temperature at small $V$, equation 12.31, can be explained by studying the Fermi gas *without reference to the superconducting phase*!

(12.6) **RG and the Central Limit Theorem: Short.** (Math) [34]

If you're familiar with the renormalization group and Fourier transforms, this problem can be stated very quickly. If not, you're probably better off doing the long version (following page).

*Write a renormalization-group transformation $T$ taking the space of probability distributions into itself, that takes two random variables, adds them, and rescales the width by the square root of two. Show that the Gaussian of width $\sigma$ is a fixed point. Find the eigenfunctions $f_n$ and eigenvectors $\lambda_n$ of the linearization of $T$ at the fixed point. (Hint: it's easier in Fourier space.) Describe physically what the relevant and marginal eigenfunctions represent. By subtracting the fixed-point distribution from a binomial distribution, find the leading correction to scaling, as a function of $x$. Which eigenfunction does it represent? Why is the leading irrelevant eigenvalue not dominant here?*

(12.7) **RG and the Central Limit Theorem: Long.** (Math) [35]

In this exercise, we will develop a renormalization group in *function space*. We'll be using maps (like our renormalization transformation $T$) that take a function $\rho$ of $x$ into another function of $x$; we'll write $T[\rho]$ as the new function, and $T[\rho](x)$ as the function evaluated at $x$. We'll also make use of the Fourier transform (equation A.5)

$$\mathcal{F}[\rho](k) = \int_{-\infty}^{\infty} e^{ikx}\rho(x)\,dx; \tag{12.38}$$

$\mathcal{F}$ maps functions of $x$ into functions of $k$. When convenient, we'll also use the tilde notation: $\widetilde{\rho} = \mathcal{F}[\rho]$, so for example (equation A.6)

$$\rho(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-ikx}\widetilde{\rho}(k)\,dk; \tag{12.39}$$

The central limit theorem states that the sum of many independent random variables tends to a Gaussian, whatever the original distribution might have looked like. That is, the Gaussian distribution is the *fixed point function* for large sums. When summing many random numbers, the details of the distributions of the individual random variables becomes unimportant: simple behavior emerges. We'll study this using the renormalization group, giving an example where we can explicitly implement the coarse–graining transformation. Here our system space is the space of probability distributions $\rho(x)$. There are four steps in the procedure:

**1. Coarse grain.** Remove some fraction (usually half) of the degrees of freedom. Here, we will add pairs of random variables: the probability distribution for sums of $N$ independent random variables of distribution $f$ is the same as the distribution for sums of $N/2$ random variables of distribution $f * f$, where $*$ denotes convolution.

*(a) Argue that if $\rho(x)$ is the probability that a random variable has value $x$, that the probability distribution of*

[33] Thus $\delta$ is not the infinitesimal change in parameter.
[34] See reference [21].
[35] See reference [21].

*the sum of two random variables drawn from this distribution is the convolution*

$$C[\rho](x) = (\rho * \rho)(x) = \int_{-\infty}^{\infty} \rho(x-y)\rho(y)dy. \quad (12.40)$$

Remember (equation A.22) the Fourier transform of the convolution is the product of the Fourier transforms, so

$$\mathcal{F}[C[\rho]](k) = \left(\widetilde{\rho}(k)\right)^2. \quad (12.41)$$

**2. Rescale.** The behavior at larger lengths will typically be similar to that of smaller lengths, but some of the constants will shift (or *renormalize*). Here the mean and width of the distributions will increase as we coarse–grain. We confine our main attention to distributions of zero mean. Remember that the width (standard deviation) of the sum of two random variables drawn from $\rho$ will be $\sqrt{2}$ times the width of one variable drawn from $\rho$, and that the overall height will have to shrink by $\sqrt{2}$ to compensate. We define a rescaling operator $S_{\sqrt{2}}$ which reverses this spreading of the probability distribution:

$$S_{\sqrt{2}}[\rho](x) = \sqrt{2}\rho(\sqrt{2}x). \quad (12.42)$$

*(b) Show that if $\rho$ is normalized (integrates to one), so is $S_{\sqrt{2}}[\rho]$. Show that the Fourier transform*

$$\mathcal{F}[S_{\sqrt{2}}[\rho]](k) = \widetilde{\rho}(k/\sqrt{2}). \quad (12.43)$$

Our renormalization-group transformation is the composition of these two operations,

$$T[\rho](x) = S_{\sqrt{2}}[C[\rho]](x) = \sqrt{2}\int_{-\infty}^{\infty} \rho(\sqrt{2}x - y)\rho(y)\,dy. \quad (12.44)$$

Adding two Gaussian random variables (convolving their distributions) and rescaling the width back should give the original Gaussian distribution: the Gaussian should be a *fixed point*.

*(c) Show that the Gaussian distribution*

$$\rho^*(x) = (1/\sqrt{2\pi}\sigma)\exp(-x^2/2\sigma^2) \quad (12.45)$$

*is indeed a fixed point in function space under the operation $T$. You can do this either by direct integration, or by using the known properties of the Gaussian under convolution.*

*(d) Use equations 12.41 and 12.43 to show that*

$$\mathcal{F}[T[\rho]](k) = \widetilde{T}[\widetilde{\rho}](k) = \widetilde{\rho}(k/\sqrt{2})^2. \quad (12.46)$$

*Calculate the Fourier transform of the fixed point $\widetilde{\rho^*}(k)$ (or see exercise A.5). Using equation 12.46, show that $\widetilde{\rho^*}(k)$ is a fixed point in Fourier space under our coarse-graining operator $\widetilde{T}$.*[36]

These properties of $T$ and $\rho^*$ should allow you to do most of the rest of the exercise without any messy integrals.

The central limit theorem tells us that sums of random variables have probability distributions that approach Gaussians. In our renormalization group framework, to prove this we need to show that our Gaussian fixed point is *attracting*: that all nearby probability distributions will flow under iterations of $T$ to $\rho^*$.[37]

**3. Linearize about the Fixed Point.** Consider a function near the fixed point: $\rho(x) = \rho^*(x) + \epsilon f(x)$. In Fourier space, $\widetilde{\rho}(k) = \widetilde{\rho^*}(k) + \epsilon\widetilde{f}(k)$. We want to find the eigenvalues $\lambda_n$ and eigenfunctions $f_n$ of the derivative of the mapping $T$. That is, they must satisfy

$$\begin{aligned} T[\rho^* + \epsilon f_n] &= \rho^* + \lambda_n\epsilon f_n + O(\epsilon^2). \quad (12.47)\\ \widetilde{T}[\widetilde{\rho^*} + \epsilon\widetilde{f_n}] &= \widetilde{\rho^*} + \lambda_n\epsilon\widetilde{f_n} + O(\epsilon^2). \end{aligned}$$

*(e) Show using equations 12.46 and 12.47 that the transforms of the eigenfunctions satisfy*

$$\widetilde{f_n}(k) = (2/\lambda_n)\widetilde{\rho^*}(k/\sqrt{2})\widetilde{f_n}(k/\sqrt{2}). \quad (12.48)$$

**4. Find the Eigenvalues and Calculate the Universal Critical Exponents.**

*(f) Show that $\widetilde{f_n}(k) = (ik)^n\widetilde{\rho^*}(k)$ is the Fourier transform of an eigenfunction (i.e., that it satisfies 12.48.) What is the eigenvalue $\lambda_n$?*

The directions with eigenvalues greater than one are called *relevant*: they are dangerous, corresponding to deviations from our fixed point that grow under coarse-graining. The directions with eigenvalues equal to one are called *marginal*: they don't get smaller (to linear order) and are thus also potentially dangerous. When you find relevant and marginal operators, you always need to understand each of them on physical grounds.

*(g) The eigenfunction $f_0(x)$ with the biggest eigenvalue corresponds to an unphysical perturbation: why? (Hint: probability distributions must be normalized to one.) The next two eigenfunctions $f_1$ and $f_2$ have important physical interpretations. Show that $\rho^* + \epsilon f_1$ to lowest order is*

[36] To be explicit, the operator $\widetilde{T} = \mathcal{F} \circ T \circ \mathcal{F}^{-1}$ is a renormalization-group transformation that maps Fourier space into itself.

[37] This won't quite be true: there will be three relevant directions in function space that we'll need to consider by hand: the width, the mean, and the normalization.

*equivalent to a shift in the mean of $\rho$, and $\rho^* + \epsilon f_2$ is a shift in the standard deviation $\sigma$ of $\rho^*$.*

All other eigenfunctions should have eigenvalues $\lambda_n$ less than one. This means that a perturbation in that direction will shrink under the renormalization–group transformation: non-Gaussian wiggles $f_n(x)$ in the distribution will die out:

$$T^N(\rho^* + \epsilon f_n) - \rho^* \sim \lambda_n^N \epsilon f_n. \tag{12.49}$$

The first two *irrelevant* eigenfunctions are easier to write in Fourier space: they are

$$\begin{aligned} f_3(x) &\propto \rho^*(x)(3x/\sigma - x^3/\sigma^3) \\ f_4(x) &\propto \rho^*(x)(3 - 6x^2/\sigma^2 + x^4/\sigma^4) \end{aligned} \tag{12.50}$$

For second order phase transitions, temperature is typically a relevant direction, with eigenvalue greater than one. This implies that the deviation of the temperature from the critical point grows under coarse–graining: on longer and longer length scales the system looks farther and farther from the critical point (figure 12.10). Specifically, if the temperature is just above the phase transition, the system appears "critical" on length scales smaller than the correlation length, but on larger length scales the effective temperature has moved far above the transition temperature and the system looks fully disordered.

In our problem, the relevant directions are comprehensible: they change the width and the means of the Gaussian. In a formal sense, we have a line of fixed points and an unstable direction: the renormalization group doesn't tell us that by subtracting the mean and rescaling the width all of the distributions would converge to the same Gaussian.

**Corrections to Scaling and Coin Flips.** Does anything really new come from all this analysis? One nice thing that comes out is the *leading corrections to scaling*. The fixed point of the renormalization group explains the Gaussian shape of the distribution of $N$ coin flips in the limit $N \to \infty$, but the linearization about the fixed point gives a systematic understanding of the corrections to the Gaussian distribution for large but not infinite $N$.

Usually, the largest eigenvalues are the ones which dominate. In our problem, consider adding a small perturbation to the fixed point $f^*$ along the two leading irrelevant directions $f_3$ and $f_4$:

$$\rho(x) = \rho^*(x) + \epsilon_3 f_3(x) + \epsilon_4 f_4(x). \tag{12.51}$$

What happens when we add $2^\ell$ of our random variables to one another (corresponding to $\ell$ applications of our renormalization group transformation $T$)? The new distribution should be given by

$$T^\ell(\rho)(x) \sim \rho^*(x) + \lambda_3^\ell \epsilon_3 f_3(x) + \lambda_4^\ell \epsilon_4 f_4(x). \tag{12.52}$$

Since $1 > \lambda_3 > \lambda_4$, the leading correction should be dominated by the perturbation with the largest eigenvalue.

*(h) Plot the difference between the binomial distribution of $N$ coin flips and a Gaussian of the same mean and width, for $N = 10$ and $N = 20$. (The Gaussian has mean of $N/2$ and standard deviation $\sqrt{N}/2$, as you can see from the case $N = 1$.) Does it approach one of the eigenfunctions $f_3$ or $f_4$ (equations 12.50)?*

*(i) Why didn't a perturbation along $f_3(x)$ dominate the asymptotics? What symmetry forced $\epsilon_3 = 0$? Should flips of a biased coin break this symmetry?*

We should mention that there are other fixed points for sums of many random variables. If the variance of the original probability distribution is infinite, one can get so-called *Levy* distributions.

(12.8) **Period Doubling.** (Math, Complexity) (With Myers. [75])

Chaos is often associated with dynamics which stretch and fold: when a batch of taffy is being pulled, the motion of a speck in the taffy depends sensitively on the initial conditions. A simple representation of this physics is provided by the map[38]

$$f(x) = 4\mu x(1 - x) \tag{12.53}$$

restricted to the domain $(0, 1)$. It takes $f(0) = f(1) = 0$, and $f(\frac{1}{2}) = \mu$. Thus, for $\mu = 1$ it precisely folds the unit interval in half, and stretches it to cover the original domain.

The study of dynamical systems (*e.g.*, differential equations and maps like equation 12.53) often focuses on the behavior after long times, where the trajectory moves along the *attractor*.[39] We can study the onset and behavior of chaos in our system by observing the evolution of the attractor as we change $\mu$. For small enough $\mu$, all

[38] We also study this map in exercises 4.3, 5.11, and 5.13; parts (a) and (b) below overlap somewhat with exercise 4.3.

[39] In statistical mechanics, we also focus on the behavior at long times, which we call the equilibrium state. Microscopically our systems do not settle down onto attractors: Liouville's theorem 4.1 guarantees that no points of phase space attract others. Of course we have attractors for the macroscopic variables in statistical mechanics.

points shrink to the origin: the origin is a stable fixed point which attracts the entire interval $x \in (0,1)$. For larger $\mu$, we first get a stable fixed point inside the interval, and then period doubling.

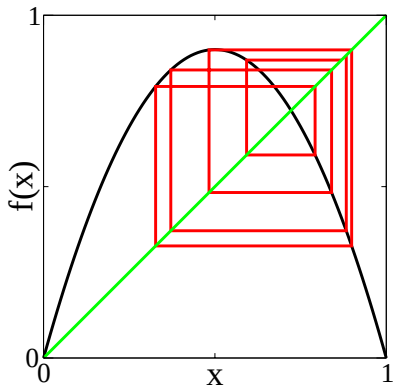



**Fig. 12.26 Period–eight cycle.** Iterating around the attractor of the Feigenbaum map at $\mu = 0.9$.

*(a)* **Iteration:** *Set $\mu = 0.2$; iterate $f$ for some initial points $x_0$ of your choosing, and convince yourself that they all are attracted to zero. Plot $f$ and the diagonal $y = x$ on the same plot. Are there any fixed points other than $x = 0$? Repeat for $\mu = 0.3$, $\mu = 0.7$, and 0.8. What happens?*

*On the same graph, plot $f$, the diagonal $y = x$, and the segments $\{x_0, x_0\}$, $\{x_0, f(x_0)\}$, $\{f(x_0), f(x_0)\}$, $\{f(x_0), f(f(x_0))\}$, ... (representing the convergence of the trajectory to the attractor: see figure 12.26). See how $\mu = 0.7$ and 0.8 differ. Try other values of $\mu$.*

*By iterating the map many times, find a point $a_0$ on the attractor. As above, then plot the successive iterates of $a_0$ for $\mu = 0.7$, 0.8, 0.88, 0.89, 0.9, and 1.0.*

You can see at higher $\mu$ that the system no longer settles into a stationary state at long times. The fixed point where $f(x) = x$ exists for all $\mu > \frac{1}{4}$ – but for larger $\mu$ it is no longer *stable*. If $x^*$ is a fixed point (so $f(x^*) = x^*$) we can add a small perturbation $f(x^* + \epsilon) \approx f(x^*) + f'(x^*)\epsilon = x^* + f'(x^*)\epsilon$; the fixed point is stable (perturbations die away) if $|f'(x^*)| < 1$.[40]

In this particular case, once the fixed point goes unstable the motion after many iterations becomes periodic, repeating itself after *two* iterations of the map – so $f(f(x))$ has two fixed points. Notice that by the chain rule $\frac{df(f(x))}{dx} = f'(x)f'(f(x))$, and indeed

$$\frac{df^{[n]}}{dx} = \frac{df(f(\dots f(x)\dots))}{dx} \tag{12.54}$$

$$= f'(x)f'(f(x))\dots f'(f(\dots f(x)\dots)) \tag{12.55}$$

so the stability of a period $N$ orbit is determined by the product of the derivatives of $f$ at each point along the orbit.

*(b)* **Analytics:** *Find the fixed point $x^*(\mu)$ of the map 12.53, and show that it exists and is stable for $1/4 < \mu < 3/4$. If you're ambitious or have a computer algebra program, show that there is a stable period–two cycle for $3/4 < \mu < (1+\sqrt{6})/4$.*

*(c)* **Bifurcation Diagram:** *Plot the attractor as a function of $\mu$, for $0 < \mu < 1$: compare with figure 12.22. (Pick regularly spaced $\delta\mu$, run $n_{\rm transient}$ steps, record $n_{\rm cycles}$ steps, and plot. After the routine is working, you should be able to push $n_{\rm transient}$ and $n_{\rm cycles}$ both larger than 100, and $\delta\mu < 0.01$.) Also plot the attractor for another one–humped map*

$$f_{\sin}(x) = B\sin(\pi x), \tag{12.56}$$

*for $0 < B < 1$. Do the bifurcation diagrams appear similar to one another?*

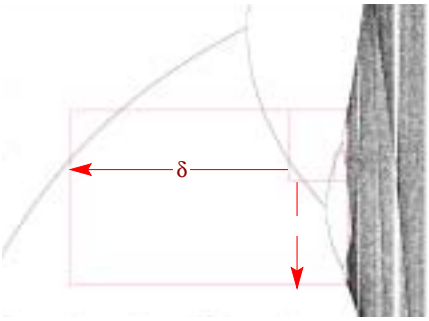



**Fig. 12.27 Self–similarity in period doubling bifurcations.** The period doublings occur at geometrically spaced values of the control parameter $\mu_\infty - \mu_n \propto \delta^n$, and the attractor at the $n^{\rm th}$ period doubling is similar to one half of the attractor at the $(n+1)^{\rm th}$ period doubling, except inverted and larger, rescaled by a factor of $\alpha$. The boxes shown in the diagram illustrate this self-similarity: each box looks like the next, except expanded by $\delta$ along the horizontal $\mu$ axis and flipped and expanded by $\alpha$ along the vertical axis.

Notice the complex, structured, chaotic region for large $\mu$ (which we study in exercise 4.3). How do we get from a

[40] In a continuous evolution, perturbations die away if the Jacobian of the derivative at the fixed point has all negative eigenvalues. For mappings, perturbations die away if all eigenvalues of the Jacobian have magnitude less than one.

stable fixed point $\mu < 3/4$ to chaos? The onset of chaos in this system occurs through a cascade of *period doublings*. There is the sequence of bifurcations as $\mu$ increases – the period two cycle starting at $\mu_1 = 3/4$, followed by a period four cycle starting at $\mu_2$, period eight at $\mu_3$ – a whole period doubling cascade. The convergence appears geometrical, to a fixed point $\mu_\infty$:

$$\mu_n \approx \mu_\infty - A\delta^n \tag{12.57}$$

so

$$\delta = \lim(\mu_{n-1} - \mu_{n-2})/(\mu_n - \mu_{n-1}) \tag{12.58}$$

and there is a similar geometrical self-similarity along the $x$ axis, with a (negative) scale factor $\alpha$ relating each generation of the tree (figure 12.27).

In exercise 4.3, we explained the boundaries in the chaotic region as images of $x = 1/2$. These special points are also convenient for studying period doubling. Since $x = 1/2$ is the maximum in the curve, $f'(1/2) = 0$. If it were a fixed point (as it is for $\mu = 1/2$), it would not only be stable, but unusually so: a shift by $\epsilon$ away from the fixed point converges after one step of the map to a distance $\epsilon f'(1/2) + \epsilon^2/2 f''(1/2) = O(\epsilon^2)$. We say that such a fixed point is *superstable*. If we have a period $N$ orbit that passes through $x = 1/2$, so that the $N^{\text{th}}$ iterate $f^N(1/2) \equiv f(\dots f(1/2)\dots) = 1/2$, then the orbit is also superstable, since (by equation 12.54) the derivative of the iterated map is the product of the derivatives along the orbit, and hence is also zero.[41]

These superstable points happen roughly halfway between the period–doubling bifurcations, and are easier to locate, since we know that $x = 1/2$ is on the orbit. Let's use them to investigate the geometrical convergence and self–similarity of the period doubling bifurcation diagram from part (d). For this part and part (h), you'll need a routine that finds the roots $G(y) = 0$ for functions $G$ of one variable $y$.

*(d)* **The Feigenbaum Numbers and Universality:** *Numerically, find the values of $\mu_n^s$ at which the $2^n$–cycle is superstable, for the first few values of $n$.* (Hint: define a function $G(\mu) = f_\mu^{[2^n]}(1/2) - 1/2$, and find the root as a function of $\mu$. In searching for $\mu_n^s$, you'll want to search in a range $(\mu_{n-1}^s + \epsilon, \mu_n^s + (\mu_n^s - \mu_{n-1}^s)/A)$ where $A \sim 3$ works pretty well. Calculate $\mu_0$ and $\mu_1$ by hand.) *Calculate the ratios $\frac{\mu_{n-1}^s - \mu_{n-2}^s}{\mu_n^s - \mu_{n-1}^s}$: do they appear to converge to the Feigenbaum number $\delta = 4.6692016091029909\dots$? Extrapolate the series to $\mu_\infty$ by using your last two reliable values of $\mu_n^s$ and equation 12.58. In the superstable orbit with $2^n$ points, the nearest point to $x = 1/2$ is $f^{[2^{n-1}]}(1/2)$.*[42] *Calculate the ratios of the amplitudes $f^{[2^{n-1}]}(1/2) - 1/2$ at successive values of $n$; do they appear to converge to the universal value $\alpha = -2.5029078750958928\dots$? Calculate the same ratios for the map $f_2(x) = B\sin(\pi x)$: do $\alpha$ and $\delta$ appear to be universal (independent of the mapping)?*

The limits $\alpha$ and $\delta$ are independent of the map, so long as it folds (one hump) with a quadratic maximum. They are the same, also, for experimental systems with many degrees of freedom which undergo the period-doubling cascade. This self–similarity and universality suggests that we should look for a renormalization–group explanation.

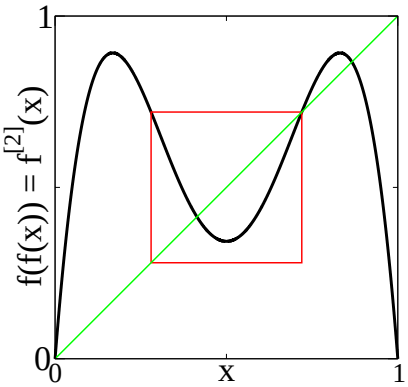



**Fig. 12.28 Renormalization–group transformation.** The renormalization–group transformation takes $g(g(x))$ in the small window with upper corner $x^*$ and inverts and stretches it to fill the whole initial domain and range $(0,1) \times (0,1)$.

[41] This is also why there is a cusp in the density at the boundaries in the chaotic region: the derivative of the function is zero, so points near $x = 1/2$ become compressed into a small region to one side of $f(1/2)$.

[42] This is true because at the previous superstable orbit, $2^{n-1}$ iterates returned us to the original point $x = 1/2$.

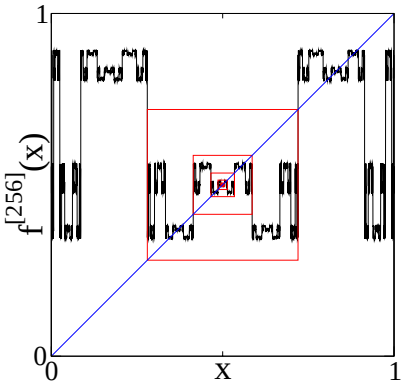



**Fig. 12.29 Self-similar iterated function in period doubling.** The function $\lim_{n\to\infty} f^{[2^n]}(x)$ is self-similar at the period–doubling fixed point $\mu_\infty$.

*(e)* **Coarse–graining in time.** *Plot $f(f(x))$ vs. $x$ for $\mu = 0.8$, together with the line $y = x$ (or see figure 12.28). Notice that the period–two cycle of $f$, naturally, become a pair of stable fixed points for $f^{[2]}$. (We're coarse–graining in time – removing every other point in the time series, by studying $f(f(x))$ rather than $f$.) Compare the plot with that for $f(x)$ vs. $x$ for $\mu = 0.5$. Notice that the region around the stable fixed points for $f^{[2]} = f(f(x))$ looks quite a bit like that around the fixed point for $f$ at the smaller value of $\mu$. Plot $f^{[4]}(x)$ at $\mu = 0.875$; notice again the small one–humped map near $x = \frac{1}{2}$.*

The fact that the one–humped map reappears in smaller form just after the period–doubling bifurcation is the basic reason that succeeding bifurcations so often follow one another. The fact that many things are universal is due to the fact that the little one–humped maps have a shape which becomes *independent of the original map* after several period doublings.

Let's define this renormalization–group transformation $T$, taking function space into itself. Roughly speaking, $T$ will take the small upside–down hump in $f(f(x))$ (figure 12.28), invert it, and stretch it to cover the interval from $(0, 1)$. Notice in your graphs for part (g) that the line $y = x$ crosses the plot $f(f(x))$ not only at the two points on the period–two attractor, but also (of course) at the old fixed point $x^*[f]$ for $f(x)$. This unstable fixed point plays the role for $f^{[2]}$ that the origin played for $f$: our renormalization–group rescaling must map $(x^*[f], f(x^*)) = (x^*, x^*)$ to the origin. The corner of the window that maps to $(1, 0)$ is conveniently located at $1 - x^*$, since our map happens to be symmetric[43] about $x = \frac{1}{2}$. For a general one–humped map $g(x)$ with fixed point $x^*[g]$ the side of the window is thus of length $2(x^*[g] - \frac{1}{2})$. To invert and stretch, we must thus rescale by a factor $\alpha[g] = -1/(2(x^*[g] - \frac{1}{2}))$. Our renormalization–group transformation is thus a mapping $T[g]$ taking function space into itself, where

$$T[g](x) = \alpha[g]\,(g\,(g(x/\alpha[g] + x^*[g])) - x^*[g])\,. \quad (12.59)$$

(This is just rescaling $x$ to squeeze into the window, applying $g$ twice, shifting the corner of the window to the origin, and then rescaling by $\alpha$ to fill the original range $(0, 1) \times (0, 1)$.)

*(f)* **Scaling and the Renormalization Group:** *Write routines that calculate $x^*[g]$ and $\alpha[g]$, and define the renormalization–group transformation $T[g]$. Plot $T[f]$, $T[T[f]]$, ... and compare them. Are we approaching a fixed point $f^*$ in function space?*

This explains the self–similarity: in particular, the value of $\alpha[g]$ as $g$ iterates to $f^*$ becomes the Feigenbaum number $\alpha = -2.5029\ldots$

*(g)* **Universality and the Renormalization Group:** *Using the sine function of equation 12.56, plot $T[T[f_{\sin}]]$ and compare with $T[T[f]]$. Are they approaching the same fixed point?*

By using this rapid convergence in function space, one can prove both that there will (often) be an infinite geometrical series of period doubling bifurcations leading to chaos, and that this series will share universal features (exponents $\alpha$ and $\delta$ and features) that are independent of the original dynamics.

(12.9) **Percolation and Universality.** (Complexity) (With Myers. [75])

**Cluster Size Distribution: Power laws at $p_c$.** A system at its percolation threshold $p_c$ is self–similar. When looked at on a longer length scale (say, with a ruler with notches spaced $1 + \epsilon$ farther apart, for infinitesimal $\epsilon$), the statistical behavior of the large percolation clusters should be unchanged, if we simultaneously rescale various measured properties according to certain rules. Let $x$ be the length and $S$ be the size (number of nodes) in a percolation cluster, and let $n(S)$ be the probability that a given cluster will be of size $S$ at $p_c$.[44] The cluster measured with the new ruler will have a length $x' = x/(1 - \epsilon)$, a size $S' = S/(1 + c\epsilon)$, and will occur with probability $n' = (1 + a\epsilon)\,n$.

[43] For asymmetric maps, we would need to locate this other corner $f(f(x_c)) = x^*$ numerically. As it happens, this asymmetry is irrelevant at the fixed point.

[44] Hence the probability that a given node is in a cluster of size $S$ is proportional to $Sn(S)$.

*(a) In precise analogy to our analysis of the avalanche size distribution (equations 12.3 through 12.6), show that the probability is a power law, $n(S) \propto S^{-\tau}$. What is $\tau$, in terms of $a$ and $c$?*

In two dimensions, there are exact results known for many properties of percolation. In particular, it is known that[45] $\tau = 187/91$. You can test this numerically, either with the code you developed for exercise 8.11, or by using the software at our Web site [108].

*(b) Calculate the cluster size distribution $n(S)$, both for bond percolation on the square lattice and for site percolation on the triangular lattice, for a large system size (perhaps $L \times L$ with $L = 400$) at $p = p_c$.*[46] *At some moderate size $S$ you will begin occasionally to not have any avalanches: plot $\log(n(S))$ versus $\log(S)$ for both bond and site percolation, together with the power law $n(S) \propto S^{-187/91}$ predicted by the exact result. To make better use of the data, one should bin the avalanches into larger groups, especially for larger sizes where the data is sparse. It's a bit tricky to do this nicely, and you can get software to do this at our Web site [108]. Do the plots again, now with all the data included, using bins that start at size ranges $1 \leq S < 2$ and grow by a factor of 1.2 for each bin.* You should see clear evidence that the distribution of clusters does look like a power law (a straight line on your log–log plot), and fairly convincing evidence that the power law is converging to the exact result at large $S$ and large system sizes.

**The Size of the Infinite Cluster: Power laws near $p_c$.** Much of the physics of percolation above $p_c$ revolves around the connected piece left after the small clusters fall out, often called the *percolation cluster*. For $p > p_c$ this largest cluster occupies a fraction of the whole system, often called $P(p)$.[47] The fraction of nodes in this largest cluster for $p > p_c$ is closely analogous to the $T < T_c$ magnetization $M(T)$ in the Ising model (figure 12.8) and the density difference $\rho_l(T) - \rho_g(T)$ near the liquid–gas critical point (figure 12.8). Indeed, the value $P(p)$ goes to zero continuously as $p \to p_c$.

Systems that are not at $p_c$ are not self–similar. However, there is a scaling relation between systems at differing values of $p - p_c$: a system coarsened by a factor $(1 + \epsilon)$ will be similar to one farther from $p_c$ by a factor $(1 + \epsilon/\nu)$, except that the percolation cluster fraction $P$ must be rescaled upward by $(1 + \beta\epsilon/\nu)$.[48] This last rescaling reflects the fact that the percolation cluster becomes more dense as you coarse grain, filling in or blurring away the smaller holes. You may check, just as for the magnetization (equation 12.7), that

$$P(p) \sim (p_c - p)^{\beta}. \tag{12.60}$$

In two dimensions, $\beta = 5/36$ and $\nu = 4/3$.

*(c) Calculate the fraction of nodes $P(p)$ in the largest cluster, for both bond and site percolation, at a series of points $p = p_c + 2^{-n}$ for as large a percolation lattice as is convenient, and a good range of $n$. (Once you get your method debugged, $n = 10$ on a $L \times L$ lattice with $L = 200$ should be numerically feasible.) Do a log–log plot of $P(p)$ versus $p - p_c$, and compare along with the theory prediction, equation 12.60 with $\beta = 5/36$.*

You should find that the numerics in part (c) are not compelling, even for rather large system sizes. The two curves look a bit like power laws, but the slopes $\beta_{\text{eff}}$ on the log-log plot don't agree with one another or with the theory. Worse, as you get close to $p_c$ the curves, although noisy, clearly are not going to zero. Of course this is true: there will always be a largest cluster, and it is only as the system size $L \to \infty$ that the largest cluster can vanish as a fraction of the system size.

**Finite Sized Scaling** (advanced). We can extract better values for $\beta$ from small simulations by explicitly including the length $L$ into our analysis. Let $P(p, L)$ be the mean fraction of nodes[49] in the largest cluster for a system of size $L$.

*(d) On a single graph, plot $P(p, L)$ versus $p$ for bond percolation $L = 5$, 10, 20, 50, and 100, focusing on the region around $p = p_c$ where they differ from one another. (At $L = 10$ you'll want $p$ to range from 0.25 to 0.75; for $L = 50$ the range should be from 0.45 to 0.55 or so.) Five or ten points will be fine. You'll discover that the sample–to–sample variations are large (another finite–size effect), so average each curve over perhaps ten or twenty realizations.*

[45] A non-obvious result!

[46] Conveniently, the critical probability $p_c = \frac{1}{2}$ for both these systems, see exercise 8.11, part(c). This enormously simplifies the scaling analysis, since we don't need to estimate $p_c$ as well as the critical exponents.

[47] For $p < p_c$, there will of course still be a largest cluster, but it will not grow much bigger as the system size grows and the fraction $P(p) \to 0$ for $p < p_c$ as the system length $L \to \infty$.

[49] You can take a microcanonical–style ensemble over all systems with exactly $p\,L^2$ sites or $2p\,L^2$ bonds, but it's simpler just to do an ensemble average over random number seeds.

Each curve $P(p, L)$ is rounded near $p_c$, as the characteristic cluster lengths reach the system box length $L$. Thus this rounding is itself a symptom of the universal long–distance behavior, and we can study the dependence of the rounding on $L$ to extract better values of the critical exponent $\beta$. We'll do this using a *scaling collapse*, rescaling the horizontal and vertical axes so as to make all the curves fall onto a single scaling function.

First, we must derive the scaling function for $P(p, L)$. We know that

$$L' = L/(1+\epsilon) \qquad (12.61)$$
$$(p_c - p)' = (1 + \epsilon/\nu)(p_c - p)$$

since the system box length $L$ rescales like any other length. It is convenient to change variables from $p$ to $X = (p_c - p)L^{1/\nu}$: let $P(p, L) = \bar{P}(L, (p_c - p)L^{1/\nu})$.

*(e) Show that $X$ is unchanged under coarse–graining (equation 12.61). (You can either show $X' = X$ up to terms of order $\epsilon^2$, or you can show $dX/d\epsilon = 0$.)*

The combination $X = (p_c - p)L^{1/\nu}$ is another *scaling variable*. The combination $\xi = |p - p_c|^{-\nu}$ is the way in which lengths diverge at the critical point, and is called the *correlation length*. Two systems of different lengths and different values of $p$ should be similar if the lengths are the same when measured in units of $\xi$. $L$ in units of $\xi$ is $L/\xi = X^\nu$, so different systems with the same value of the scaling variable $X$ are statistically similar. We can turn this verbal assertion into a mathematical scaling form by studying how $\bar{P}(L, X)$ coarse–grains.

*(f) Using equations 12.61 and the fact that $P$ rescales upward by $(1 + \beta\epsilon/\nu)$ under coarse–graining, write the similarity relationship for $\bar{P}$ corresponding to equation 12.12 for $\bar{D}(S, R)$. Following our derivation of the scaling form for the avalanche size distribution (through equation 12.15), show that $\bar{P}(L, X) = L^{-\beta/\nu}\mathcal{P}(X)$ for some function $\mathcal{P}(X)$, and hence*

$$P(p, L) \propto L^{-\beta/\nu}\mathcal{P}((p - p_c)L^{1/\nu}) \qquad (12.62)$$

*Presuming that $\mathcal{P}(X)$ goes to a finite value as $X \to 0$, derive the power law giving the percolation cluster size $L^2 P(p_c, L)$ as a function of $L$. Derive the power law variation of $\mathcal{P}(X)$ as $X \to \infty$ using the fact that $P(p, \infty) \propto (p - p_c)^\beta$.*

Now, we can use equation 12.62 to deduce how to rescale our data. Clearly we can find the finite–sized scaling function $\mathcal{P}$ by plotting $L^{\beta/\nu}P(p, L)$ versus $X = (p - p_c)L^{1/\nu}$, again with $\nu = 4/3$ and $\beta = 5/36$.

*(g) Plot $L^{\beta/\nu}P(p, L)$ versus $X$ for $X = -0.8, +0.8$, plotting perhaps five points for each curve, for both site percolation and bond percolation. Use system sizes $L = 5$, 10, 20, and 50. Average over many clusters for the smaller sizes (perhaps 400 for $L = 5$), and over at least ten even for the largest.*

Your curves should collapse onto two scaling curves, one for bond percolation and one for site percolation.[50] Notice here that the finite–sized scaling curves collapse well for small $L$, where we would need to go to much larger $L$ to see good power laws in $P(p)$ directly (part (c)). Notice also that both site percolation and bond percolation collapse for the same value of $\beta$, even though the rough power laws from part (c) seemed to differ. In an experiment (or a less thoroughly understood theory), one would use these collapses to estimate $p_c$, $\beta$, and $\nu$.

(12.10) **Hysteresis Model: Scaling and Exponent Equalities.** (Complexity)

Find a Windows machine. Download Matt Kuntz' hysteresis simulation from our course Web site [55, 54].

Run it with the default parameters (two dimensions, $R = 0.9$, 1000×1000) You can use the center buttons in the upper right of the subwindow and main window to make them expand to fill the screen.

The simulation is a simplified model of magnetic hysteresis, described in [109]; see also [110]. The spins $s_i$ begin all pointing down, and flip upward as the external field $H$ grows from minus infinity, depending on the spins of their neighbors and a local random field $h_i$. The flipped spins are colored as they flip, with spins in the same *avalanche* sharing the same color. An avalanche is a collection of spins all triggered from the same original spin. In the parameter box, the *Disorder* is the ratio $R$ of the root-mean-square width $\sqrt{\langle h_i^2 \rangle}$ to the ferromagnetic coupling $J$ between spins:

$$R = \sqrt{\langle h^2 \rangle}/J. \qquad (12.63)$$

Examine the $M(H)$ curve for our model (the fourth button, marked with the $S$ curve) and the $dM/dH$ curve (the fifth button, marked with a spiky curve). The individual avalanches should be visible on the first graph as jumps, and the second graph as spikes. This kind of time series (a set of spikes or pulses with a broad range of

[50] These two curves would also collapse onto one another, given a suitable rescaling of the horizontal and vertical axes, if we did the triangular lattice in a square box instead of the rectangular box we get from shearing an $L \times L$ lattice. The finite–size scaling function will in general depend on the boundary condition, and in particular on the shape of the box.

sizes) we hear as *crackling noise*. You can go to our site http://simscience.org/crackling/ to hear the noise resulting from our model, as well as crackling noise we've assembled from crumpling paper, from fires and Rice Krispies©, and from the earth (earthquakes in 1995, sped up to audio frequencies).

Examine the avalanche size distribution (button *A*). The (unlabeled) vertical axis on the log-log plot gives the number of avalanches $D(S,R)$; the horizontal axis gives the size $S$ (with $S = 1$ on the left-hand side). Equivalently, $D(S,R)$ is the probability distribution that a given avalanche during the simulation will have size $S$. The graph is created as a histogram, and the curve changes color after the first bin with zero entries (after which the data becomes much less useful, and should be ignored).

Examine also the spin-spin correlation function $C(x,R)$ (button *C*). It shows a log-log plot of the probability (vertical-axis) that an avalanche initiated at a point $\mathbf{x}_0$ will extend to include a spin $\mathbf{x}_1$ a distance $x = \sqrt{(\mathbf{x}_1 - \mathbf{x}_0)^2}$ away.

Two dimensions is fun to watch, but the scaling behavior is rather confusing. In three dimensions we have good evidence for scaling and criticality at a phase transition in the dynamical evolution. Well below $R_c \sim 2.16$ (for the cubic lattice), one large avalanche flips most of the spins. Well above $R_c$ all avalanches are fairly small: at very high disorder each spin flips individually. We measure the critical disorder as the point, as $L \to \infty$, where one first finds *spanning avalanches*, which extend from one side of the simulation to the other.

Simulate a 3D system with $L = 100$ (one million spins) at $R = R_c = 2.16$ (or larger, if you have a fast machine). It'll be faster if you use the *sorted list* algorithm (which takes much more memory, though, and will pause for a while before starting). The display will show a 100×100 cross-section of the 3D avalanches. Notice that there are many tiny avalanches, and a few blobs. Usually you will find one large colored region spanning most of the system. Look at the $M(H)$ curve (the bottom half of the hysteresis loop, given by the button with the slanted S-like curve). It has many small vertical jumps (avalanches), and one large one (corresponding to the spanning avalanche).

*(a) What fraction of the system is flipped by the one largest avalanche, in your simulation? Compare this with the hysteresis curve at $R = 2.4 > R_c$. Does it have a similar big jump, or is it continuous?*

Below $R_c$ we get a big jump; above $R_c$ all avalanches are small compared to the system size. If the system size were large enough, we believe the fraction of spins flipped by the spanning avalanche at $R_c$ would go to zero. The largest avalanche would nonetheless span the system – just like the percolation cluster at $p_c$ spans the system but occupies volume zero in the limit of large systems.

The other avalanches form a nice power-law size distribution: let's measure it carefully. Do a set of 10 runs (*# Runs 10*) at $L = 100$ and $R = R_c = 2.16$. (If your machine is slow, do five or even one run. If your machine is fast, or you need to eat dinner anyhow, do 10 runs at $L = 200$. Make sure you don't run out of RAM, though: if you do, shift to the *bits* algorithm.)

Watch the avalanches. Notice that sometimes the second-largest avalanche in the view (the largest being the "background color") is sometimes pretty small: this is often because the cross section we view missed it. Look at the avalanche size distribution. (You can watch it as it averages over simulations.) Print it out when the simulations finish by selecting the graph and selecting *Print...* from the *Simulation* menu. Notice that at $R_c$ you find a pretty good power law distribution (a straight line on the log-log plot). This critical exponent is usually denoted by the exponent combination $\tau + \sigma\beta\delta$:

$$D(S, R_c) \sim S^{-(\tau+\sigma\beta\delta)}. \tag{12.64}$$

*(b) From your plot, measure this exponent combination from your simulation. It should be close to two. Is your estimate larger or smaller than two?*

This power-law distribution is to magnets what the Gutenberg-Richter law is to earthquakes. The power law stems naturally from the *self-similarity*

We want to explore how the avalanche size distribution changes as we move above $R_c$. We'll do a series of three or four runs at different values of $R$, and then graph the avalanche size distributions after various transformations. To store the data files, Matt's program assumes that you have a certain set of folders. In the directory where the program HysteresisWin.exe resides, add a folder named *data* and in it put a directory named *average*. The program will save data files into those directories (but won't create them for you). Check *Output Data Files* on the dialog box before starting the runs away from $R_c$.

Do a run at $R = 6$ and $R = 4$ with $L = 100$, and make sure you have files aval_histo_D3_LX in your data directory. Do runs at $R = 3$, $R = 2.5$, and $R = 2.16$ at $L = 200$, using the *bits* algorithm if you start running out of RAM (and your disk starts swapping, making a lot of noise). *Bits* will start faster than *sorted list*, but it will take a long time searching for the last few spins: be patient if you don't get the output files for a while after it looks like it's finished.

*(c) Copy and edit your avalanche size distribution files, removing the data after the first bin with zero avalanches in it. Start up a graphics program, and plot the curves on*

*a log-log plot: they should look like power laws for small $S$ and cut off at larger $S$. Enclose a copy of your plot.*

We expect the avalanche size distribution to have the scaling form

$$D(S,R) = S^{-(\tau+\sigma\beta\delta)}\mathcal{D}(S(R-R_c)^{1/\sigma}). \tag{12.65}$$

sufficiently close to $R_c$. This form is "self universal": a system at $2(R-R_c)$ has the same distribution as a system at $R-R_c$ except for an overall change $A$ in probability and $B$ in the size scale of the avalanches, so $D(S,R-R_c) \approx AD(BS, 2(R-R_c))$.

*(d) What are $A$ and $B$ in this equation for the scaling form given by equation 12.65?*

Of course, at $R = 4$ and 6 we expect substantial corrections! Let's see how well the collapse works anyhow.

*(e) Multiply the vertical axis of each curve by $S^{\tau+\sigma\beta\delta}$. This then should give four curves $\mathcal{D}(S(R-R_c)^{1/\sigma})$ which are (on a log-log plot) roughly the same shape, just shifted sideways horizontally (rescaled in $S$ by the typical largest avalanche size, proportional to $1/(R-R_c)^{1/\sigma}$). Measure the peak of each curve. Make a table with columns $R$, $S_{\rm peak}$, and $R-R_c$ (with $R_c \sim 2.16$). Do a log-log plot of $R-R_c$ versus $S_{\rm peak}$, and estimate $\sigma$ in the expected power law $S_{\rm peak} \sim (R-R_c)^{-1/\sigma}$.*

*(f) Do a scaling collapse: plot $S^{\tau+\sigma\beta\delta}D(S,R)$ versus $(R-R_c)^{1/\sigma}\,S$ for the avalanche size distributions with $R > R_c$. How well do they collapse onto a single curve?*

This problem gives nice answers only near $R_c$, where you need very large systems to get good curves! Olga Perkovic worked really hard on extracting critical exponents some years back for her Ph.D. thesis.

# Appendix: Fourier Methods

# A

Why are Fourier methods important? Why is it so common and useful for us to transform functions of time and space $y(\mathbf{x}, t)$ into functions of frequency and wave-vector $\widetilde{y}(\mathbf{k}, \omega)$?

- *Humans hear frequencies.* The human ear analyzes pressure variations in the air into different frequencies. Large frequencies $\omega$ are perceived as high pitches; small frequencies are low pitches. The ear, very roughly, does a Fourier transform of the pressure $P(t)$ and transmits $|\widetilde{P}(\omega)|^2$ to the brain.[1]
- *Scattering experiments measure Fourier components.* Many experimental methods scatter waves (light, X-rays, electrons, or neutrons) off of materials (section 10.2). These experiments typically probe the absolute square of the Fourier amplitude of whatever is scattering the incoming beam.
- *Common mathematical operations become simpler in Fourier space.* Derivatives, correlation functions, and convolutions can be written as simple products when the functions are Fourier transformed. This has been important to us when calculating correlation functions (equation 10.3), summing random variables (exercise 1.2 and exercise 12.7) and calculating susceptibilities (equations 10.28, 10.37, and 10.51). In each case, we turn a calculus calculation into algebra.
- *Linear equations in translationally-invariant systems have solutions in Fourier space.* We have used Fourier methods for solving the diffusion equation (section 2.4.1 and equation 10.20), and more broadly in our study of correlation functions and susceptibilities (chapter 10).

[1] Actually, this is how the ear *seems* to work, but not how it *does* work. First, the signal to the brain is time dependent, with the tonal information changing as a word or tune progresses: it's more like a wavelet transform, giving the frequency content in various time slices. Second, the phase information in $\widetilde{P}$ is not completely lost: power and pitch is the primary signal, but the relative phases of different pitches is also perceptible. Third, experiments have shown that the human ear is very nonlinear in its mechanical response.

In section A.1 we introduce the conventions typically used in physics for the Fourier series, Fourier transform, and fast Fourier Transform. In section A.2 we derive their integral and differential properties. In section A.3,we interpret the Fourier transform as an orthonormal change-of-basis in function space. And finally, in section A.4 we explain why Fourier methods are so useful for solving differential equations by exploring their connection to translational symmetry.

## A.1 Fourier Conventions

Here we define the Fourier series, the Fourier transform, and the fast Fourier transform, as they will be used in this text.

The Fourier series for periodic functions of period $T$ is

$$\widetilde{y}_m = \frac{1}{T}\int_0^T y(t)\exp(i\omega_m t)dt, \tag{A.1}$$

where $\omega_m = 2\pi m/T$, with integer $m$. The Fourier series can be resummed to retrieve the original function using the *inverse Fourier series*:

$$y(t) = \sum_{m=-\infty}^{\infty} \widetilde{y}_m \exp(-i\omega_m t). \tag{A.2}$$

In a three–dimensional box of volume $V = L \times L \times L$ with periodic boundary conditions, these formulas become

$$\widetilde{y}_{\mathbf{k}} = \frac{1}{V}\int y(\mathbf{x})\exp(i\mathbf{k}\cdot\mathbf{x})dV, \tag{A.3}$$

and

$$y(\mathbf{x}) = \sum_{\mathbf{k}} \widetilde{y}_{\mathbf{k}} \exp(-i\mathbf{k}\cdot\mathbf{x}). \tag{A.4}$$

where the $\mathbf{k}$ run over a lattice of wave vectors $\mathbf{k}_{(m,n,o)} = [\frac{2\pi m}{L}, \frac{2\pi n}{L}, \frac{2\pi o}{L}]$ in the box.

The Fourier transform is defined for functions on the entire infinite line:

$$\widetilde{y}(\omega) = \int_{-\infty}^{\infty} y(t)\exp(i\omega t)\,dt \tag{A.5}$$

where now $\omega$ takes on all values.[2] We regain the original function by doing the inverse Fourier transform.

$$y(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \widetilde{y}(\omega)\exp(-i\omega t)\,d\omega. \tag{A.6}$$

This is related to the inverse Fourier series by a continuum limit (figure A.1:

$$\frac{1}{2\pi}\int d\omega \approx \frac{1}{2\pi}\sum_{\omega}\Delta\omega = \frac{1}{2\pi}\sum_{\omega}\frac{2\pi}{T} = \frac{1}{T}\sum_{\omega} \tag{A.7}$$

where the $1/T$ here compensates for the factor of $T$ in the definitions of the forward Fourier series. In three dimensions the Fourier transform formula A.5 is largely unchanged,

$$\widetilde{y}(\mathbf{k}) = \int y(\mathbf{x})\exp(i\mathbf{k}\cdot\mathbf{x})dV \tag{A.8}$$

while the inverse Fourier transform gets the cube of the prefactor.

$$y(\mathbf{x}) = \frac{1}{(2\pi)^3}\int_{-\infty}^{\infty} \widetilde{y}(\mathbf{k})\exp(-i\mathbf{k}\cdot\mathbf{x})\,d\mathbf{k}. \tag{A.9}$$

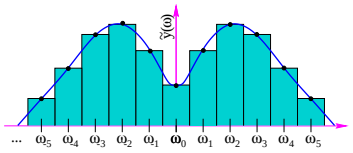



**Fig. A.1 Approximating the integral as a sum.** By approximating the integral $\int \tilde{y}(\omega)\exp(i\omega t)\,d\omega$ as a sum over the equally spaced points $\omega_m$, $\sum_m \tilde{y}(\omega)\exp(i\omega_m t)\Delta\omega$, we can connect the formula for the Fourier transform to the formula for the Fourier series, explaining the factor $1/2\pi$ in equation A.6.

[2] Why do we divide by $T$ or $L$ for the series and not for the transform? Imagine a system in an extremely large box. Fourier series are used for functions which extend over the entire box: hence we divide by the box size to keep them finite as $L \to \infty$). Fourier transforms are usually used for functions which vanish quickly, so they remain finite as the box size gets large.

The Fast Fourier transform (FFT) starts with $N$ equally spaced data points $y_\ell$, and returns a new set of complex numbers $\widetilde{y}_m^{FFT}$:

$$\widetilde{y}_m^{FFT} = \sum_{\ell=0}^{N-1} y_\ell \exp(i2\pi m\ell/N), \tag{A.10}$$

with $m = 0, \ \ldots N-1$. The inverse of the FFT is given by

$$y_\ell = (1/N) \sum_{m=0}^{N-1} \widetilde{y}_m^{FFT} \exp(-i2\pi m\ell/N). \tag{A.11}$$

The FFT essentially samples the function $y(t)$ at equally spaced points points $t_\ell = \ell T/N$ for $\ell = 0, \ldots N-1$.

$$\widetilde{y}_m^{FFT} = \sum_{\ell=0}^{N-1} y_\ell \exp(i\omega_m t_\ell). \tag{A.12}$$

It is clear from equation A.10 that $\widetilde{y}_{m+N}^{FFT} = \widetilde{y}_m^{FFT}$, so the fast Fourier transform is periodic with period $\omega_N = 2\pi N/T$. The inverse transform can also be written

$$y_\ell = (1/N) \sum_{m=-N/2+1}^{N/2} \widetilde{y}_m^{FFT} \exp(-i\omega_m t_\ell) \tag{A.13}$$

where we have centered[3] the sum $\omega_m$ at $\omega = 0$ by using the periodicity.[4]

Often the values $y(t)$ (or the data points $y_\ell$) are real. In this case, equations A.1 and A.5 show that the negative Fourier amplitudes are the complex conjugates of the positive ones: $\widetilde{y}(\omega) = \widetilde{y}^*(-\omega)$. Hence for real functions the real part of the Fourier amplitude will be even and the imaginary part will be odd.[5]

The reader may wonder why there are so many versions of roughly the same Fourier operation.

(1) The function $y(t)$ can be defined on a finite interval with periodic boundary conditions on $(0, T)$ (series, FFT) or defined in all space (transform). In the periodic case, the Fourier coefficients are defined only at discrete wavevectors $\omega_m = 2\pi m/T$ consistent with the periodicity of the function: in the infinite system the coefficients are defined at all $\omega$.
(2) The function $y(t)$ can be defined at a discrete set of $N$ points $t_n = n\Delta t = nT/N$ (FFT), or at all points $t$ in the range (series, transform). If the function is defined only at discrete points, the Fourier coefficients are periodic with period $\omega_N = \frac{2\pi}{\Delta t} = 2\pi N/T$.[6]

[3] If $N$ is odd, to shift the Fourier series to place zero in the center of a plot the sum should be taken over $-(N-1)/2 \leq m \leq (N-1)/2$.

[4] Notice that the FFT returns the negative $\omega$ Fourier coefficients as the last half of the vector, $m = N/2+1, N/2+2, \ldots$. (This works because $-N/2+j$ and $N/2+j$ differ by $N$, the periodicity of the FFT). One must be careful about this when using Fourier transforms to solve calculus problems numerically. For example, to solve the diffusion equation (section 2.4.1) one must multiply the first half of the vector by $e^{-Dk_m^2 t} = e^{-D(m\frac{2\pi}{L})^2 t}$ but multiply the second half by $e^{-D(K-k_m)^2 t} = e^{-D\left((N-m)\frac{2\pi}{L}\right)^2 t}$.

[5] This allows one to write slightly faster FFTs specialized for real functions. One pays for the higher speed by an extra step in unpacking the resulting Fourier spectrum.

[6] There is one more logical possibility: a discrete set of points that fill all space: the atomic displacements in an infinite crystal, for example. In Fourier space, such a system has continuous $k$ with periodic boundary conditions at the edges of the *Brillouin zone* at $\pm K/2 = \pm\pi/a$.

The reader may wonder also why our formulas may be different from those found in other texts. These conventions differ substantially between fields.

- Some use the notation $j = \sqrt{-1}$ instead of $i$.
- More substantively, some use the complex conjugate of our formulas, substituting $-i$ for $i$ in all the formulas.[7] This alternative convention of course makes no change for any real quantity.
- Some use a $1/\sqrt{T}$ and $1/\sqrt{2\pi}$ factor symmetrically on the Fourier and inverse Fourier operations.
- Some use frequency and wavelength ($f = 2\pi\omega$ and $\lambda = 2\pi/k$) instead of angular frequency $\omega$ and wavevector $k$. This makes the transform and inverse transform more symmetric, and avoids some of the prefactors.

[7] Changing $i \to -i$ in the time series formulas, for example, would make $\chi''(\omega) = -Im[\chi(\omega)]$ in equation 10.29 and would make $\chi$ analytic in the *lower* half-plane in figure 10.11.

Our Fourier conventions are those most commonly used in physics, at least for time–frequency transformations. For spatial Fourier transforms, physicists sometimes use the complex conjugate of our formulas.[8]

[8] This inconsistent convention can be convenient for wave problems, because a single Fourier mode then has the form $\exp(i(\mathbf{k}\cdot\mathbf{x} - \omega t))$, which (for $\omega > 0$ is a wave propagating in the direction $\mathbf{k}$: our more consistent convention unintuitively produces the wave $\exp(-i(\mathbf{k}\cdot\mathbf{x} + \omega t))$ which unintuitively propagates in the direction $-\mathbf{k}$. In particular, if we use a single Fourier component with our conventions $\exp(-i\omega t - i\mathbf{k}\cdot\mathbf{x}$ for Schrödinger's equation for a free particle, then the energy $E = i\hbar\frac{\partial}{\partial t} = \hbar\omega$ and the momentum $\mathbf{p} = -i\hbar\frac{\partial}{\partial x} = -\hbar\mathbf{k}$.

## A.2 Derivatives, Convolutions, and Correlations

The important differential and integral operations become multiplications in Fourier space. A calculus problem in $t$ or $x$ thus becomes an algebra exercise in $\omega$ or $k$.

**Integrals and derivatives.** Because $\frac{d}{dt}e^{-i\omega t} = -i\omega e^{i\omega t}$, the Fourier coefficient for the derivative of a function $y(t)$ is $-i\omega$ times the Fourier coefficient of the function:

$$dy/dt = \sum \widetilde{y}_m \left(-i\omega_m \exp(-i\omega_m t)\right) = \sum (-i\omega_m \widetilde{y}_m) \exp(-i\omega_m t) \tag{A.14}$$

so

$$\left.\widetilde{\frac{dy}{dt}}\right|_\omega = -i\omega\widetilde{y}_\omega. \tag{A.15}$$

This holds also for the Fourier transform and the fast Fourier transform. Since the derivative of the integral gives back the original function, the Fourier transform for the indefinite integral of a function $y$ is thus given by dividing by $-i\omega$

$$\widetilde{\int y(t)dt} = \frac{\widetilde{y}_\omega}{-i\omega} = i\frac{\widetilde{y}_\omega}{\omega} \tag{A.16}$$

except at $\omega = 0$.[9]

[9] If the mean of the function is zero, then $\widetilde{y}(\omega)/\omega = 0/0$ is undefined at $\omega = 0$. This makes sense: the indefinite integral has an arbitrary integration constant, which gives its Fourier series an arbitrary value at $\omega = 0$. If the mean of the function $\bar{y}$ is not zero, then the integral of the function will have a term $\bar{y}(t - t_0)$. Hence it is not periodic (for Fourier series) or square integrable (for Fourier transforms).

These relations are invaluable in the solution of many linear partial differential equations. For example, we saw in section 2.4.1 that the diffusion equation

$$\frac{\partial\rho}{\partial t} = D\frac{\partial^2\rho}{\partial x^2} \tag{A.17}$$

becomes trivial when we Fourier transform $x$ to $k$:

$$\frac{\partial \widetilde{\rho}_k}{\partial t} = -Dk^2 \widetilde{\rho}, \tag{A.18}$$

$$\widetilde{\rho}_k(t) = \rho_k(0) \exp(-Dk^2 t). \tag{A.19}$$

**Convolutions and correlation functions.** The absolute square of the Fourier transform[10] $|\widetilde{y}(\omega)|^2$ is given by the Fourier transform of the correlation function $C(\tau) = \langle y(t)y(t+\tau)\rangle$ (see equation 10.3):

$$\begin{aligned}
|\widetilde{y}(\omega)|^2 &= \widetilde{y}(\omega)^* \widetilde{y}(\omega) = \int dt'\, e^{-i\omega t'} y(t') \int dt\, e^{i\omega t} y(t) \\
&= \int dt dt'\, e^{i\omega(t-t')} y(t') y(t) = \int d\tau\, e^{i\omega\tau} \int dt'\, y(t') y(t'+\tau) \\
&= \int d\tau\, e^{i\omega\tau} T \langle y(t)y(t+\tau)\rangle = T \int d\tau\, e^{i\omega\tau} C(\tau) \\
&= T\widetilde{C}(\omega),
\end{aligned} \tag{A.20}$$

where $T$ is the total time $t$ during which the Fourier spectrum is being measured. Thus, as noted in section 10.2, scattering experiments, by measuring the square of the Fourier transform, give us the spatial correlation function for the system.

The convolution[11] $h(z)$ of two functions $f(x)$ and $g(y)$ is defined as

$$h(z) = \int f(x) g(z-x)\, dx. \tag{A.21}$$

The Fourier transform of the convolution is the product of the Fourier transforms. In three dimensions,[12]

$$\begin{aligned}
\widetilde{f}(\mathbf{k})\widetilde{g}(\mathbf{k}) &= \int e^{i\mathbf{k}\cdot\mathbf{x}} f(\mathbf{x}) d\mathbf{x} \int e^{i\mathbf{k}\cdot\mathbf{y}} g(\mathbf{y}) d\mathbf{y} \\
&= \int e^{i\mathbf{k}\cdot(\mathbf{x}+\mathbf{y})} f(\mathbf{x}) g(\mathbf{y}) d\mathbf{x} d\mathbf{y} = \int e^{i\mathbf{k}\cdot\mathbf{z}} d\mathbf{z} \int f(\mathbf{x}) g(\mathbf{z}-\mathbf{x}) d\mathbf{x} \\
&= \int e^{i\mathbf{k}\cdot\mathbf{z}} h(\mathbf{z}) d\mathbf{z} = \widetilde{h}(\mathbf{k}).
\end{aligned} \tag{A.22}$$

## A.3 Fourier Methods and Function Space

.

There is a nice analogy between the space of vectors $\mathbf{r}$ in three dimensions and the space of functions $y(t)$ periodic with period $T$, which provides a simple way of thinking about Fourier series. It is natural to define our function space to including all complex functions $y(t)$. (After all, we want the complex Fourier plane-waves $e^{-i\omega_m t}$ to be in our space.) Let us list the features in common:

- **Vector Space.** A vector $\mathbf{r} = (r_1, r_2, r_3)$ in $\mathbb{R}^3$ can be thought of as a real-valued function on the set $\{1, 2, 3\}$. Conversely, the function $y(t)$ can be thought of as a vector with one component for each $t \in [0, T)$.

[10] The absolute square of the Fourier transform of a time signal is called the *power spectrum*.

[11] Convolutions show up in sums and Greens functions. The sum $\mathbf{z} = \mathbf{x} + \mathbf{y}$ of two random vector quantities with probability distributions $f(\mathbf{x})$ and $g(\mathbf{y})$ has a probability distribution given by the convolution of $f$ and $g$ (exercise 1.2): (The value $\mathbf{x}$ is chosen for the first random variable with probability density $f(\mathbf{x})$: to get a sum of $\mathbf{z}$ the second variable $\mathbf{y}$ must have the value $\mathbf{z}-\mathbf{x}$, with probability density $g(\mathbf{z}-\mathbf{x})$.) An initial condition $f(\mathbf{x}, t_0)$ propagated in time to $t_0 + \tau$ is given by convolving with a Greens function $g(\mathbf{y}, \tau)$ (section 2.4.2).

[12] Clearly the convolution and correlation theorems are closely related: we do convolutions in time and correlations in space to illustrate both the one-dimensional and vector versions of the calculation.

Mathematically, this is an evil analogy. Most functions which have independent random values for each point $t$ are undefinable, unintegrable, and generally pathological. The space becomes well defined if we confine ourselves to functions $y(t)$ whose absolute squares $|y(t)|^2$ can be integrated. This vector space of functions is called $\mathbb{L}^2$.

- **Inner Product.** The analogy to the dot product of two 3D vectors $\mathbf{r} \cdot \mathbf{s} = r_1 s_1 + r_2 s_2 + r_3 s_3$ is an inner product between two functions $y$ and $z$:

$$y \cdot z = \frac{1}{T} \int_0^T y(t) z^*(t)\, dt. \tag{A.23}$$

  You can think of this inner product as adding up all the products $y_t z_t^*$ over all points $t$, except that we weight each point by $dt/T$.
- **Norm.** The distance between two 3D vectors $\mathbf{r}$ and $\mathbf{s}$ is given by the *norm* of the difference $|\mathbf{r} - \mathbf{s}|$. The norm of a vector is the square root of the dot product of the vector with itself, so $|\mathbf{r} - \mathbf{s}| = \sqrt{(\mathbf{r} - \mathbf{s}) \cdot (\mathbf{r} - \mathbf{s})}$. To make this inner product norm work in function space, we need to know that the inner product of a function with itself is never negative. This is why, in our definition A.23, we took the complex conjugate of $z(t)$. This norm on function space is called the $L^2$ norm,

$$||y||_2 = \sqrt{\frac{1}{T} \int_0^T |y(t)|^2}. \tag{A.24}$$

  Thus our restriction to square-integrable functions makes the norm of all functions in our space finite.[13]
- **Basis.** A natural basis for $\mathbb{R}^3$ is given by the three unit vectors $\hat{\mathbf{x}}_1$, $\hat{\mathbf{x}}_2$, $\hat{\mathbf{x}}_3$. A natural basis for our space of functions is given by the functions $\widehat{f}_m = e^{-i\omega_m t}$, with $\omega_m = 2\pi m/T$ to keep them periodic with period $T$.
- **Orthonormality.** The basis in $\mathbb{R}^3$ is orthonormal, with $\hat{\mathbf{x}}_i \cdot \hat{\mathbf{x}}_j$ equaling one if $i = j$ and zero otherwise. Is this also true of the vectors in our basis of plane waves? They are normalized

$$||\widehat{f}_m||_2^2 = \frac{1}{T} \int_0^T |e^{-i\omega_m t}|^2\, dt = 1. \tag{A.25}$$

  They are also orthogonal, with

$$\begin{aligned} \widehat{f}_m \cdot \widehat{f}_n &= \frac{1}{T} \int_0^T e^{-i\omega_m t} e^{i\omega_n t}\, dt = \frac{1}{T} \int_0^T e^{-i(\omega_m - \omega_n)t}\, dt \\ &= \frac{1}{-i(\omega_m - \omega_n)T}\, e^{-i(\omega_m - \omega_n)t} \Big|_0^T = 0 \end{aligned} \tag{A.26}$$

  unless $m = n$, since $e^{-i(\omega_m - \omega_n)T} = e^{-i2\pi(m-n)} = 1$.
- **Coefficients.** The coefficients of a 3D vector are given by taking dot products with the basis vectors: $r_n = \mathbf{r} \cdot \hat{\mathbf{x}}_n$. The analogy in

[13] Another important property is that the only vector whose norm is zero is the zero vector. There are many functions whose absolute squares have integral zero, like the function which is zero except at $T/2$, where it is one, and the function which is zero on irrationals and one on rationals. Mathematicians finesse this difficulty by defining the vectors in $\mathbb{L}^2$ not to be functions, but rather to be *equivalence classes* of functions whose relative distance is zero. Hence the zero vector in $\mathbb{L}^2$ includes all functions with norm zero.

function space gives us the definition of the Fourier coefficients, equation A.1:

$$\widetilde{y}_m = y \cdot \widehat{f}_m = \frac{1}{T}\int_0^T y(t)\exp(i\omega_m t)dt. \tag{A.27}$$

- **Completeness.** We can write an arbitrary 3D vector $\mathbf{r}$ by summing the basis vectors weighted by the coefficients: $\mathbf{r} = \sum r_n \hat{\mathbf{x}}_n$. The analogy in function space gives us the formula A.2 for the inverse Fourier series:

$$\begin{aligned} y &= \sum_{m=-\infty}^{\infty} \widetilde{y}_m \widehat{f}_m \\ y(t) &= \sum_{m=-\infty}^{\infty} \widetilde{y}_m \exp(-i\omega_m t). \end{aligned} \tag{A.28}$$

  One says that a basis is *complete* if any vector can be expanded in that basis. Our functions $\widehat{f}_m$ are complete in $\mathbb{L}^2$.[14]

[14] You can imagine that proving they are complete would involve showing that there are no functions in $\mathbb{L}^2$ which are "perpendicular" to all the Fourier modes. This is the type of tough question that motivates the mathematical field of real analysis.

Our coefficient equation A.27 follows from our completeness equation A.28 and orthonormality:

$$\begin{aligned} \widetilde{y}_\ell \stackrel{?}{=} y \cdot \widehat{f}_\ell &= \left(\sum_m \widetilde{y}_m \widehat{f}_m\right) \cdot \widehat{f}_\ell \\ &= \sum_m \widetilde{y}_m \left(\widehat{f}_m \cdot \widehat{f}_\ell\right) = \widetilde{y}_\ell \end{aligned} \tag{A.29}$$

or, writing things out

$$\begin{aligned} \widetilde{y}_\ell \stackrel{?}{=} &\frac{1}{T}\int_0^T y(t) e^{i\omega_\ell t}\, dt \\ =&\frac{1}{T}\int_0^T \left(\sum_m \widetilde{y}_m e^{-i\omega_m t}\right) e^{i\omega_\ell t}\, dt \\ =&\sum_m \widetilde{y}_m \left(\frac{1}{T}\int_0^T e^{-i\omega_m t} e^{i\omega_\ell t}\, dt\right) = \widetilde{y}_\ell. \end{aligned} \tag{A.30}$$

Our function space, together with our inner product (equation A.23), is a *Hilbert space* (a complete inner product space).

## A.4 Fourier and Translational Symmetry

Why are Fourier methods so useful? In particular, why are the solutions to linear differential equations so often given by plane waves: sines and cosines and $e^{ikx}$?[15] Most of our basic equations are derived for systems with a *translational symmetry*. Time translational invariance holds for any system without an explicit external time–dependent force; invariance under spatial translations holds for all homogeneous systems.

[15] It's true, we're making a big deal about what's usually called the separation of variables method. But why does separation of variables so often work, and why does it always end up with sines?

Why are plane waves special for systems with translational invariance? *Plane waves are the eigenfunctions of the translation operator.* Define $\mathcal{T}_\Delta$, an operator which takes function space into itself, and acts to shift the function a distance $\Delta$ to the right:[16]

$$\mathcal{T}_\Delta\{f\}(x) = f(x-\Delta). \tag{A.31}$$

Any solution $f(x,t)$ to a translation–invariant equation will be mapped by $\mathcal{T}_\Delta$ onto another solution. Moreover, $\mathcal{T}_\Delta$ is a linear operator (translating the sum is the sum of the translated functions). If we think of the translation operator as a big matrix acting on function space, we can ask for its eigenvalues and eigenvectors (now called eigenfunctions) $f_k$.[17]

$$\mathcal{T}_\Delta\{f_k\}(x) = f_k(x-\Delta) = \lambda_k f_k(x). \tag{A.32}$$

This equation is of course solved by our complex plane waves solutions $f_k(x) = e^{-ikx}$ with $\lambda_k = e^{ik\Delta}$.[18]

Why are these eigenfunctions useful? The time evolution of an eigenfunction must have the same eigenvalue $\lambda$! The argument is something of a tongue–twister: translating the time evolved eigenfunction gives the same answer as time evolving the translated eigenfunction, which is time evolving $\lambda$ times the eigenfunction, which is $\lambda$ times the time evolved eigenfunction.[19]

The fact that the different eigenvalues don't mix under time evolution is precisely what made our calculation work: time evolving $A_0 e^{-ikx}$ had to give a multiple $A(t)e^{-ikx}$ since there is only one eigenfunction of translations with the given eigenvalue. Once we've reduced the partial differential equation to a differential equation for a few eigenstate amplitudes, the calculation becomes analytically simple enough to do.

Quantum physicists will recognize the tongue–twister above as a statement about simultaneously diagonalizing commuting operators: since translations commute with time evolution, one can find a complete set of translation eigenstates which are also time evolution solutions. Mathematicians will recognize it from group representation theory: the solutions to a translation–invariant linear differential equation form a representation of the translation group, and hence they can be decomposed into irreducible representations of that group. These approaches are basically equivalent, and very powerful. One can also use these approaches for systems with other symmetries. For example, just as the invariance

[16] That is, if $g = \mathcal{T}_\Delta\{f\}$, then $g(x) = f(x-\Delta)$, so $g$ is $f$ shifted to the right by $\Delta$.

[17] You are familiar with eigenvectors of $3\times 3$ symmetric matrices $M$, which go to multiples of themselves when multiplied by $M$, $M\cdot\vec{e}_n = \lambda_n\vec{e}_n$. The translation $\mathcal{T}_\Delta$ is a linear operator on function space just as $M$ is a linear operators on $\mathbb{R}^3$.

[18] The real exponential $e^{Ax}$ is also an eigenstate, with eigenvalue $e^{-A\Delta}$. Indeed, the analogous eigenstate for time translations is the solution for the time–translation–invariant diffusion equation, which we've seen has solutions which decay in time as $e^{-\omega_k t}$, with $\omega_k = Dk^2$. Exponentially decaying solutions in space also arise in some translation–invariant problems, such as quantum tunneling into barriers and the penetration of electromagnetic radiation (light) into metals.

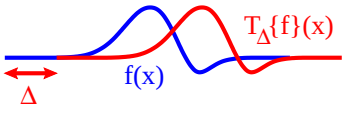



**Fig. A.2** The mapping $\mathcal{T}_\Delta$ takes function space into function space, shifting the function to the right by a distance $\Delta$. For a physical system that is translation invariant, a solution translated to the right is still a solution.

[19] Written out in equations, this simple idea is even more obscure. Let $\mathcal{U}_t$ be the time–evolution operator for a translationally invariant equation (like the diffusion equation of section 2.2). That is, $\mathcal{U}_t\{\rho\}$ evolves the function $\rho(x,\tau)$ into $\rho(x,\tau+t)$. ($\mathcal{U}_t$ is not *translation* in time, but *evolution* in time.) Because our system is translation invariant, translated solutions are also solutions for translated initial conditions: $\mathcal{T}_\Delta\{\mathcal{U}_t\{\rho\}\} = \mathcal{U}_t\{\mathcal{T}_\Delta\{\rho\}\}$. Now, if $\rho_k(x,0)$ is an eigenstate of $\mathcal{T}_\Delta$ with eigenvalue $\lambda_k$, is $\rho_k(x,t) = \mathcal{U}_t\{\rho_k\}(x)$ an eigenstate with the same eigenvalue? Yes indeed:

$$\begin{aligned}\mathcal{T}_\Delta\{\rho_k(x,t)\} &= \mathcal{T}_\Delta\{\mathcal{U}_t\{\rho_k(x,0)\}\} = \mathcal{U}_t\{\mathcal{T}_\Delta\{\rho_k(x,0)\}\} \\ &= \mathcal{U}_t\{\lambda_k\rho_k(x,0)\} = \lambda_k\mathcal{U}_t\{\rho_k(x,0)\} = \lambda_k\rho_k(x,t).\end{aligned} \tag{A.33}$$

because the evolution law $\mathcal{U}_t$ is linear.

of homogeneous systems under translations leads to plane-wave solutions with definite wave vector $k$,

- The invariance of isotropic systems (like the hydrogen atom) under the rotation group leads naturally to solutions involving spherical harmonics with definite angular momenta $\ell$ and $m$,
- The invariance of the strong interaction under $SU(3)$ leads naturally to the '8-fold way' families of mesons and baryons, and
- The invariance of our universe under the Poincaré group of space–time symmetries (translations, rotations, and Lorentz boosts) leads naturally to particles with definite mass and spin!

## Exercises

(A.1) **Fourier for a Waveform.** (Basic,Math)
A musical instrument playing a note of frequency $\omega_1$ generates a pressure wave $P(t)$ periodic with period $2\pi/\omega_1$: $P(t) = P(t + 2\pi/\omega_1)$. The complex Fourier series of this wave (equation A.2) is zero except for $m = \pm 1$ and $\pm 2$, corresponding to the fundamental $\omega_1$ and the first overtone. At $m = 1$, the Fourier amplitude is $2-i$, at $m = -1$ it is $2 + i$, and at $m = \pm 2$ it is 3. What is the pressure $P(t)$?

(A) $\exp((2 + i)\omega_1 t) + 2\exp(3\omega_1 t)$
(B) $\exp((2\omega_1 t))\exp(i(\omega_1 t)) * 2\exp(3\omega_1 t)$
(C) $\cos 2\omega_1 t - \sin\omega_1 t + 2\cos 3\omega_1 t$
(D) $4\cos\omega_1 t - 2\sin\omega_1 t + 6\cos 2\omega_1 t$
(E) $4\cos\omega_1 t + 2\sin\omega_1 t + 6\cos 2\omega_1 t$

(A.2) **Relations between the Fouriers.** (Math)
In this exercise, we explore the relationships between the Fourier series and the fast Fourier transform. The first is continuous and periodic in real space, and discrete and unbounded in Fourier space; the second is discrete and periodic both in real and in Fourier space. Thus, we must again convert integrals into sums (as in figure A.1).

As we take the number of points $N$ in our FFT to $\infty$ the spacing between the points gets smaller and smaller, and the approximation of the integral as a sum gets better and better.

*Let $y_\ell = y(t_\ell)$ where $t_\ell = \ell(T/N) = \ell(\Delta t)$. Approximate the Fourier series integral A.1 above as sum over $y_\ell$, $(1/T)\sum_{\ell=0}^{N-1} y(t_\ell)\exp(-i\omega_m t_\ell)\Delta t$. For small positive $m$, give the constant relating $\tilde{y}_m^{FFT}$ to the Fourier series coefficient $\tilde{y}_m$.*

(A.3) **Fourier Series: Computation.** (Math,Basic)
In this exercise, we will use the computer to illustrate features of Fourier series and discrete fast Fourier transforms using sinusoidal waves. Download the program *Fourier* from the course Web page [108], or the hints file for the programming language of your choice.

First, we will take the Fourier series of periodic functions $y(x) = y(x + L)$ with $L = 20$. We will sample the function at $N = 32$ points, and use a FFT to approximate the Fourier series. The Fourier series will be plotted as functions of $k$, at $-k_{N/2}$, ..., $k_{N/2-2}$, $k_{N/2-1}$. (Remember that the negative $m$ points are given by the last half of the FFT.)

*(a) Analytically derive the Fourier series $\tilde{y}_m$ in this interval for $\cos(k_1 x)$ and $\sin(k_1 x)$. (Hint: they are zero except at two values of $m = \pm 1$.)*

*(b) What spacing $\delta k$ between $k$-points $k_m$ do you expect to find? What is $k_{N/2}$? Evaluate each both as a formula involving $L$ and numerically for $L = 20$.*

Start the program *Fourier*. You should have a graph of a cosine wave $A\cos(k(x - x_0))$, evaluated at 32 points from $x = 0$ to 20 as described above, with $k = k_1 = 2\pi/L$, $A = 1$, and $x_0 = 0$. You should also have a graph of the Fourier series of the cosine: black is the real part, red is the imaginary part. Play with the program, trying various combinations of the real-space, Fourier-space, and parameter options.[20]

[20] You can always return to the initial configuration by quitting the program and starting again.

*(c) Check your predictions from part (a) for the Fourier series for* $\cos(k_1 x)$ *and* $\sin(k_1 x)$*. Check your predictions from part (b) for* $\delta k$ *and for* $k_{N/2}$*. Zoom in on the plots to get better accuracy for your measurements.* (Hint: you can find $k_{N/2}$ by observing the range on the Fourier series plot, which should go from $k_{-N/2}$ to $k_{N/2-1}$ for even $N$.) Increase the number of waves, keeping the number of data points fixed. Notice that the Fourier series looks fine, but that the real-space curves quickly begin to vary in amplitude, much like the patterns formed by beating (superimposing two waves of different frequencies). By increasing the number of data points, you can see that the beating effect is due to the small number of points we sample. Of course, even for large numbers of sampled points $N$, at very small wavelengths (when we get close to $k_{N/2}$) beating will still happen. Try various numbers of waves $m$ up to and past $m = N/2$.

The Fourier series $\tilde{y}_m$ runs over all integers $m$. The fast Fourier transform runs only over $0 \le m < N$. There are three ways to understand this difference: function space dimension, wavelengths, and aliasing.

**Function space dimension.** The space of periodic functions $y(x)$ on $0 \le x < L$ is infinite, but we are sampling them only at $N = 32$ points. The space of possible fast Fourier series must also have $N$ dimensions. Now, each coefficient of the FFT is complex (two dimensions), but the negative frequencies are complex conjugate to their positive partners (giving two net dimensions for the two wavevectors $k_m$ and $k_{-m} \equiv k_{N-m}$). If you're fussy, $\tilde{y}_0$ has no partner, but is real (only one dimension), and if $N$ is even $\tilde{y}_{-N/2}$ also is partnerless, but is real. So $N$ k-points are generated by $N$ real points.

**Wavelength.** The points that we sample the function are spaced $\delta x = L/N$ apart. It makes sense that the fast Fourier transform would stop when the wavelength becomes close to $\delta x$: how can we resolve wiggles shorter than our sample spacing?

*(d) Give a formula for* $y_\ell$ *for a cosine wave at* $k_N$*, the first wavelength* not *calculated by our FFT. It should simplify to a constant. Give the simplified formula for* $y_\ell$ *at* $k_{N/2}$*, the first missing wavevector after we've shifted the large* $m$*'s to* $N - m$ *to get the negative frequencies. Check your prediction for what* $y_\ell$ *looks like for* $\cos(k_N x)$ *and* $\cos(k_{N/2} x)$ *using the program* Fourier*. The Fourier series for the latter should have only one spike, at the edge of the plot.*

So, the FFT returns Fourier components only until there is one point per bump (half-period) in the cosine wave.

**Aliasing.** Suppose our function really does have wiggles with shorter distances than our sampling distance $\delta x$. Then its fast Fourier transform will have contributions to the long-wavelength coefficients $\tilde{y}_m^{FFT}$ from these shorter wavelength wiggles: specifically $\tilde{y}_{m\pm N}$, $\tilde{y}_{m\pm 2N}$, etc.

*(e) Let's work out a particular case of this: a short-wavelength cosine wave. On our sampled points* $x_\ell$*, show that* $\exp(ik_{m\pm N} x_\ell) = \exp(ik_m x_\ell)$*. Show that the short wavelength wave* $\cos(k_{m+N} x_\ell) = \cos(k_m x_\ell)$*, and hence that its fast Fourier transform will be a bogus peak at the long wavelength* $k_m$*. Check your prediction for the transforms of* $\cos(kx)$ *for* $k > k_{N/2}$ *using the program.*

If you sample a function at $N$ points with Fourier components beyond $k_{N/2}$, their contributions get added to Fourier components at smaller wave-vectors. This is called *aliasing*, and is an important source of error in Fourier methods. We always strive to sample enough points to avoid it.

You should see at least once how aliasing affects the FFT of functions that are not sines and cosines. Go back to the default values for the program, and shift to the function *Packet*. Change the width $\sigma$ of the packet to make it thinner. Notice that when the packet begins to look ratty (as thin as the spacing between the sampled points $x_\ell$) the Fourier series hits the edges and overlaps: high frequency components are "folded over" or *aliased* into the lower frequencies.

Make sure you understand the Fourier series for cosines and sines with various numbers of wavelengths: how $x_0$ changes cosines to sines, how the wave-vector $k$ is related to $L$ and the number of wavelengths, how the real and imaginary parts vary with the phase of the wave, and why the imaginary parts of the Fourier series for $\sin(kx)$ have the signs that they do.

**Windowing.** One often needs to take Fourier series of functions which are not periodic in the interval. Set the number of data points $N$ to 256 (powers of two are faster) and compare $m = 20$ with an "illegal" non-integer value $m = 20.5$. Notice that the real-space function $y(x)$ is not periodic in the interval $[0, L)$ for $m = 20.5$. Notice that the Fourier series looks pretty complicated. Each of the two peaks has broadened into a whole staircase. Try looking at the power spectrum (which is proportional to $|\tilde{y}|^2$), and again compare $m = 20$ with $m = 20.5$. This is a numerical problem known as *windowing*, and there are various schemes to minimize its effects as well.

(A.4) **Fourier Series of a Sinusoid.** (Basic,Math)

Which picture in figure A.3 represents the Fourier series (equation A.2) associated with the function $f(x) = 3\sin(x) + \cos(2x)$? (The solid line is the real part, the dashed line is the imaginary part.)

(A.5) **Fourier Transforms and Gaussians: Computation.** (Math,Basic)

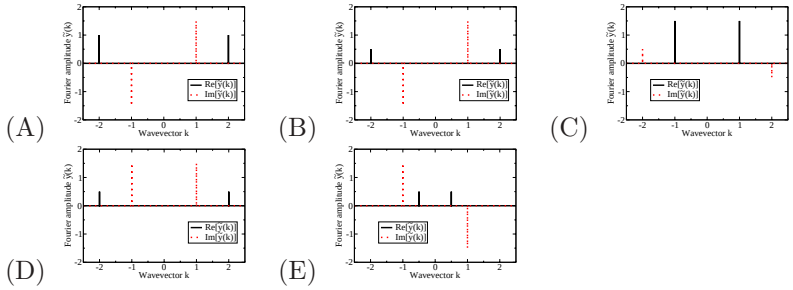



**Fig. A.3**

In this exercise, we will use the computer to illustrate features of Fourier transforms, focusing on the particular case of Gaussian functions, but illustrating general properties. Download the program *Fourier* from the course Web page [108], or the hints file for the programming language of your choice.

The Gaussian distribution (also known as the normal distribution) has the form

$$G(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-(x-x_0)^2/2\sigma^2\right). \tag{A.34}$$

where $\sigma$ is the standard deviation and $x_0$ is the center. Let

$$G_0(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-x^2/2\right) \tag{A.35}$$

be the Gaussian of mean zero and $\sigma = 1$. The Fourier transform of $G_0$ is another Gaussian, of standard deviation one, but no normalization factor:[21]

$$\widetilde{G}_0(k) = \exp(-k^2/2). \tag{A.37}$$

In this exercise, we study how the Fourier transform of $G(x)$ varies as we change $\sigma$ and $x_0$.

**Widths.** As we make the Gaussian narrower (smaller $\sigma$), it becomes more pointy. Shorter lengths mean higher wave-vectors, so we expect that its Fourier transform will get wider.

*(a) Starting with the Gaussian with $\sigma = 1$, zoom in to measure the width of its Fourier transform at some convenient height. (The full width at half maximum, FWHM, is a sensible choice.) Change $\sigma$ to 2 and to* 0.1*, and measure the widths, to verify that the Fourier space with goes inversely with the real width.*

*(b) Show that this rule is true in general. Change variables in equation A.5 to show that if $z(x) = y(Ax)$, that $\widetilde{z}(k) = \widetilde{y}(k/A)/A$. Using equation A.35 and this general rule, write a formula for the Fourier transform of a Gaussian centered at zero with arbitrary width $\sigma$.*

*(c) Compute the product $\Delta x\,\Delta k$ of the FWHM of the Gaussians in real and Fourier space. (Your answer should be independent of the width $\sigma$.)* This is related to the Heisenberg uncertainty principle, $\Delta x\,\Delta p \sim \hbar$, which you learn about in quantum mechanics.

**Translations.** Notice that a narrow Gaussian centered at some large distance $x_0$ is a reasonable approximation

[21] There is a misleadingly elementary-looking derivation of this formula, which is useful to see even though it's only correct for an advanced reason. We complete the square inside the exponent, and change variables from $x$ to $y = x - ik$:

$$\begin{aligned}\widetilde{G}_0(k) &= \int_{-\infty}^{\infty} e^{ikx} G_0(x)\,dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{ikx} \exp(-x^2/2)\,dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp(-(x-ik)^2/2)\,dx\, \exp((-ik)^2/2) \\ &= \left[\int_{-\infty-ik}^{\infty-ik} \frac{1}{\sqrt{2\pi}} \exp(-y^2/2)\,dy\right] \exp(-k^2/2). \end{aligned} \tag{A.36}$$

If we are sloppy about the integration limits, this would give us the answer: the term inside the brackets is one (the integral of a normalized Gaussian), leaving us with $\exp(-k^2/2)$. However, the change-of-variables here *changes the path in the complex plane*: the integration contour is along the line $Im[y] = -k$. To justify shifting this path, we need advanced methods from complex analysis: Cauchy's theorem, the fact that the exponential has no singularities in the complex plane, and the fact that $\exp(-y^2)$ gets very small along the two closing segments at large $Re[y] = \pm X$.

to a $\delta$-function. We thus expect that its Fourier transform will be similar to the plane wave $\widetilde{G}(k) \sim \exp(ikx_0)$ we would get from $\delta(x - x_0)$.

*(d) Change the center of the Gaussian. How does the Fourier transform change? Convince yourself that it is being multiplied by the factor* $\exp(ikx_0)$. *How does the power spectrum* $|\widetilde{G}(\omega)|^2$ *change as we change* $x_0$*?*

*(e) Show that this rule is also true in general. Change variables in equation A.5 to show that if* $z(x) = y(x - x_0)$ *then* $\widetilde{z}(k) = \exp(ikx_0)\widetilde{y}(k)$. *Using this general rule, extend your answer from part (b) to write the formula for the Fourier transform of a Gaussian of width* $\sigma$ *and center* $x_0$.

(A.6) **Uncertainty.** (Basic,Math)

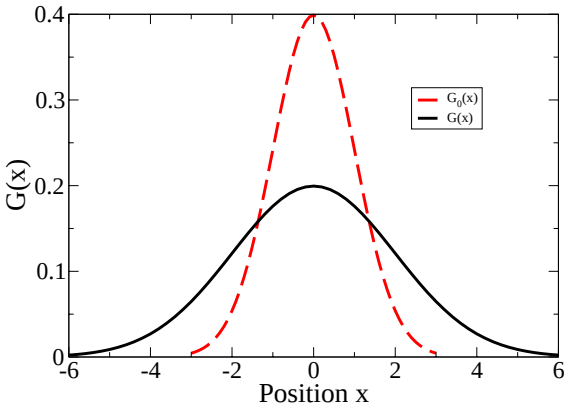



**Fig. A.4** Real Space Gaussians

The dashed line in figure A.4 shows

$$G_0(x) = 1/\sqrt{2\pi}\,\exp(-x^2/2). \tag{A.38}$$

The dark line shows another function $G(x)$. The areas under the two curves $G(x)$ and $G_0(x)$ are the same. The dashed lines in the choices of figure A.5 represent the Fourier transform $\tilde{G}_0(k) = \exp(-k^2/2)$. Which part in figure A.5 has a solid curve that represents the Fourier transform of $G$?

(A.7) **White Noise.** (Math)

White light is a mixture of light of all frequencies. White noise is a mixture of sound of all frequencies, with constant average power per unit frequency. The hissing noise you hear on radio and TV between stations is approximately white noise: there are a lot more high frequencies than low ones, so it sounds high-pitched.

Download the program *Fourier* from the course Web page [108], or the hints file for the programming language of your choice.

What kind of time signal would generate white noise? Start *Fourier*, and select the function *White Noise*. You should see a jagged, random function: each $y_\ell = y(\ell L/N)$ is chosen independently as a random number.[22] Change the number of data points to, say, 1024.

You should also see the Fourier transform of the noise signal. The Fourier transform of the white noise looks amazingly similar to the original signal. It is different, however, in two important ways. First, it is complex: there is a real part (black) and an imaginary part (red). The second is for you to discover.

*Zoom in near* $k = 0$ *on the Fourier plot, and describe how the Fourier transforms of the noisy signal are different from random. In particular, what symmetry do the real and imaginary parts have? Can you show that this is true for any real function* $y(x)$*?*

Now select the *Power Spectrum* for the right graph. The average power at a certain frequency in a time signal $f(t)$ is proportional to $|\tilde{f}(k)|^2$, which is what we plot on the right side. Check that the power is noisy, but on average is crudely independent of frequency. (You can check this best by varying the random number seed.) White noise is usually due to random, uncorrelated fluctuations in time.

(A.8) **Fourier Matching.** (Basic,Math)

*Which of the first six parts (1-6) of figure A.6 is the Fourier representation of each of the last three parts (A-C)? (Dark line is real part, lighter dotted line is imaginary part.)* (This exercise should be fairly straightforward after doing exercises A.3, A.5 and A.7.)

(A.9) **Fourier Series and Gibbs Phenomenon.** (Math)

[22] We choose the numbers with probability given by the Gaussian distribution $G_0(y)$, but it would look about the same if we took numbers with a uniform probability in, say, the range $(-1, 1)$.

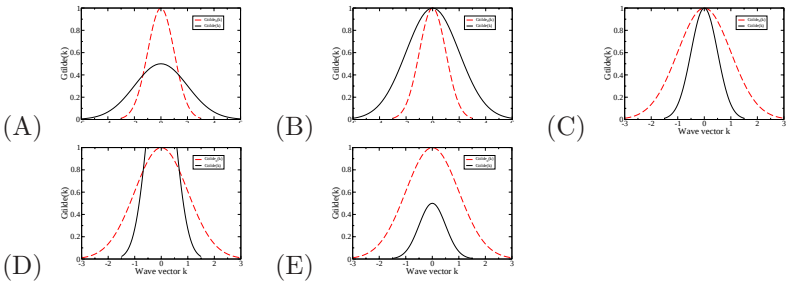



**Fig. A.5**

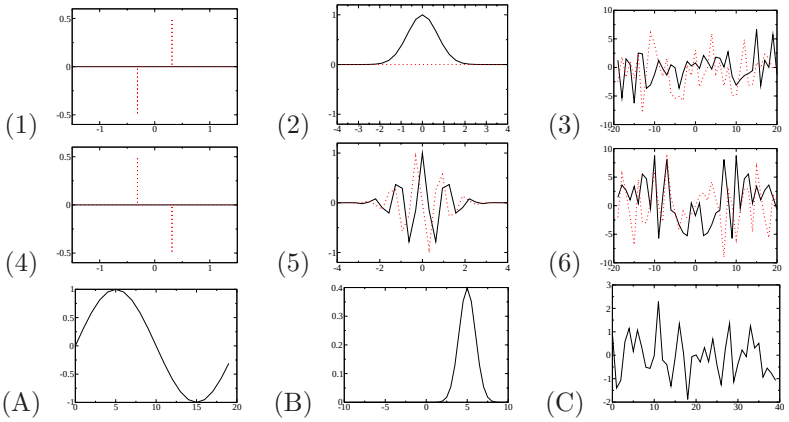



**Fig. A.6**

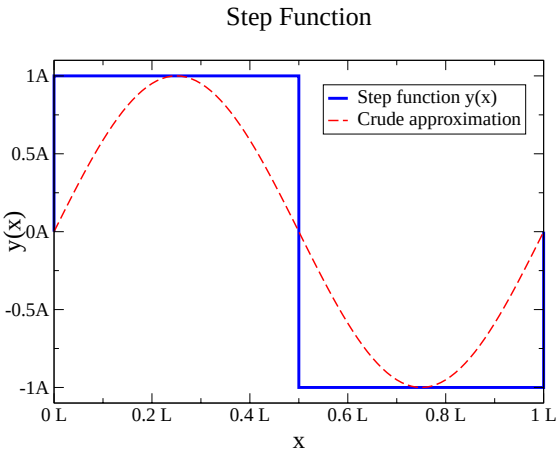



**Fig. A.7** Step Function

In this exercise, we'll look at the Fourier series for a couple of functions, the step function (above) and the triangle function.

Consider a function $y(x)$ which is $A$ in the range $0 < x < L/2$ and minus $A$ in the range $L/2 < x < L$ (shown above). It's a kind of step function, since it takes a step downward at $L/2$.[23]

*(a) As a crude approximation, the step function looks a bit like a chunky version of a sine wave, $A\sin(2\pi x/L)$. In this crude approximation, what would the complex Fourier series be (equation A.2)?*

*(b) Show that the odd coefficients for the complex Fourier series of the step function are $\tilde{y}_m = 2Ai/(m\pi)$ ($m$ odd). What are the even ones? Check that the coefficients $\tilde{y}_m$ with $m = \pm 1$ are close to those you guessed in part (a).*

*(c) Setting $A = 2$ and $L = 10$, plot the partial sum of the Fourier series (equation A.1) for $m = -n, -n+1, \ldots, n$ with $n = 1$, 3, and 5. (You'll likely need to combine the coefficients $\tilde{y}_m$ and $\tilde{y}_{-m}$ into sines or cosines, unless your plotting package knows about complex exponentials.) Does it converge to the step function? If it is not too inconvenient, plot the partial sum up to $n = 100$, and concentrate especially on the overshoot near the jumps in the function at 0, $L/2$, and $L$. This overshoot is called the Gibbs phenomenon, and occurs when you try to approximate functions $y(x)$ with discontinuities.*

One of the great features of the Fourier series is that it makes taking derivatives and integrals easier. What does the integral of our step function look like? Let's sum the Fourier series for it!

*(d) Calculate the Fourier series of the integral of the step function, using your complex Fourier series from part (b) and the formula A.16 for the Fourier series of the integral. Plot your results, doing partial sums up to $\pm m = n$, with $n = 1$, 3, and 5, again with $A = 2$ and $L = 10$. Would the derivative of this function look like the step function? If it's convenient, do $n = 100$, and notice there are no overshoots.*

[23] It can be written in terms of the standard Heaviside step function $\Theta(x) = 0$ for $x < 0$ and $\Theta(x) = 1$ for $x > 0$, as $y(x) = A\,(1 - 2\Theta(x - L/2))$.

# Bibliography


[1] Guenter Ahlers. Critical phenomena at low temperatures. *Rev. Mod. Phys.*, 52:489–503, 1980.

[2] Vinay Ambegaokar, B. I. Halperin, David Nelson, and Eric Siggia. Dissipation in two–dimensional superfluids. *Physical Review Letters*, 40:783–6, 1978.

[3] Vinay Ambegaokar, B. I. Halperin, David Nelson, and Eric Siggia. Dynamics of superfluid films. *Physical Review B*, 21:1806–26, 1980.

[4] M. H. Anderson, J.R. Ensher, M.R. Matthews, C.E. Wieman, and E.A. Cornell. Observation of bose-einstein condensation in a dilute atomic vapor. *Science*, 269, 1995. jilawww.colorado.edu/bec/ . Thanks to the Physics 2000 team for permission to reprint this figure.

[5] P. W. Anderson. Coherent matter field phenomena in superfluids. In *Some Recent Definitions in the Basic Sciences*, volume 2. Belfer Graduate School of Science, Yeshiva University, New York, 1965. Reprinted in [7, p. 229].

[6] P. W. Anderson. Considerations on the flow of superfluid helium. *Rev. Mod. Phys.*, 38, 1966. Reprinted in [7, p. 249].

[7] P. W. Anderson. *Basic Notions of Condensed Matter Physics*. Benjamin–Cummings, Menlo Park, CA, 1984.

[8] Neil W. Ashcroft and N. David Mermin. *Solid State Physics*. Hold, Rinehart, and Wilson, Philadelphia, 1976.

[9] Carl M. Bender and Steven A. Orszag. *Advanced Mathematical Methods for Scientists and Engineers*. McGraw–Hill, New York, 1978.

[10] Howard Berg. Motile behavior of bacteria. *Physics Today*, 53:24–29, 2000. www.aip.org/pt/jan00/berg.htm .

[11] N. O. Birge and S. R. Nagel. Specific heat spectroscopy of the glass transition. *Physical Review Letters*, 54:2674, 1985.

[12] Eberhard Bodenschatz, Brian Utter, and Rolf Ragnarsson. Directional solidification of solvent/polymer alloys. www.milou.ccmr.cornell.edu /solidification.html, 2002.

[13] A.B. Bortz, M.H. Kalos, and J.L. Lebowitz. A new algorithm for monte-carlo simulation of ising spin systems. *J. Comput. Phys.*, 17:10–18, 1975.

[14] Piet Brouwer. This exercise was written in collaboration with Piet Brouwer.

[15] Sarah Buchan. Thanks to Sarah Buchan for suggesting this problem.

[16] Alex Buchel and James P. Sethna. Elastic theory has zero radius of convergence. *Physical Review Letters*, 77:1520, 1996.

[17] Alex Buchel and James P. Sethna. Statistical mechanics of cracks: Thermodynamic limit, fluctuations, breakdown, and asymptotics of elastic theory. *Physical Review E*, 55:7669, 1997.

[18] John Cardy. *Scaling and Renormalization in Statistical Physics*. Cambridge University Press, Cambridge, 1996.

[19] Jean M. Carlson and James A. Langer. Properties of earthquakes generated by fault dynamics. *Physical Review Letters*, 62:2632, 1989.

[20] David Chandler. *Introduction to Modern Statistical Mechanics*. Oxford University Press, Oxford, 1987.

[21] J. T. Chayes, L. Chayes, J. P. Sethna, and D. J. Thouless. A mean-field spin glass with short-range interactions. *Commun. Math. Phys.*, 106:41, 1986.

[22] T.-S. Choy, J. Naset, J. Chen, S. Hershfield, and C. Stanton. The fermi surface database. www.phys.ufl.edu/fermisurface/, 2000.

[23] Chunhwa Chu. *Hysteresis and microstructures: a study of biaxial loading on compound twins of copper-aluminum-nickel single crystals*. PhD thesis, Aerospace Engineering, University of Minnesota, 1993.

[24] Sydney Coleman. Secret symmetry: an introduction to spontaneous symmetry breakdown and gauge fields. In *Aspects of Symmetry, Selected Erice Lectures*. Cambridge University Press, Cambridge, 1985.

[25] Barbara H. Cooper and Tatjana Curcic. Stm images of nanoscale pit decay. www.lassp.cornell.edu/cooper_nanoscale /nanofeatures.html, 1995.

[26] P. G. deGennes and J. Prost. *The Physics of Liquid Crystals, Second Edition*. Clarendon Press, Oxford, 1993.

[27] C. Domb and A. J. Guttman. Low-temperature series for the ising model. *J. Phys. C*, 8:1652–60, 1970.

[28] Freeman J. Dyson. Time without end: Physics and biology in an open universe. *Reviews of Modern Physics*, 51:447, 1979.

[29] Michael B. Elowitz and Stanislas Leibler. A synthetic oscillatory network of transcriptional regulators. *Nature*, 403:335–338, 2000.

[30] E. Fermi, J. Pasta, and S. Ulam. Studies of nonlinear problems. i. In *E. Fermi, Collected Papers, Vol. II*, page 978. Univ. of Chicago Press, Chicago, 1965. (Reprinted from Los Alamos report LA-1940, 1955.).

[31] Richard P. Feynman. *Statistical Mechanics, A set of lectures*. Addison-Wesley, Menlo Park, CA, 1972.

[32] Dieter Forster. *Hydrodynamic Fluctuations, Broken Symmetry, and Correlation Functions*. Benjamin/Cummings, Reading, Massachusetts, 1975.

[33] D. Frenkel and A. A. Louis. Phase separation in a binary hard-core mixture. an exact result. *Phys. Rev. Lett.*, 68:3363, 1992.

[34] Alex Gaeta and Al Sievers. Thanks to Alex Gaeta and Al Sievers for helpful suggestions in writing this exercise.
[35] D.T. Gillespie. Exact simulation of coupled chemical reactions. *J. Comput. Phys.*, 22:403–434, 1976.
[36] Paul Ginsparg. Private communication.
[37] Michelle Girvan and Mark E. J. Newman. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, 12:7821–7826, 2002.
[38] Raymond E. Goldstein and Neil W. Ashcroft. Origin of the singular diameter in coexistence curve of a metal. *Phys. Rev. Lett.*, 55:2164–2167, 1985.
[39] Peter J. E. Goss and Jean Peccoud. Quantitative modeling of stochastic systems in molecular biology by using stochastic petri nets. *Proceedings of the National Academy of Sciences*, 95:6750–6755, 1998.
[40] Dennis S. Greywall and Guenter Ahlers. Second–sound velocity and superfluid density in $^4$he under pressure near $t_\lambda$. *Phys. Rev. A*, 7:2145–2162, 1973.
[41] John Guckenheimer. Private communication.
[42] E. A. Guggenheim. The principle of corresponding states. *J. Chem. Phys.*, 13:253–261, 1945.
[43] B. I. Halperin. Adapted from final exam, Harvard University, 1977.
[44] Peter Hänggi, Peter Talkner, and Michal Borkovec. Reaction–rate theory: fifty years after kramers. *Reviews of Modern Physics*, 62:251, 1990.
[45] P. Heller and G. B. Benedek. Nuclear magnetic resonance in $mnf_2$ near the critical point. *Phys. Rev. Lett.*, 8:428–432, 1962.
[46] Paul A. Houle and James P. Sethna. Acoustic emission from crumpling paper. *Physical Review E*, 54:278, 1996.
[47] Joachim Jacobsen, Karsten W. Jacobsen, and James P. Sethna. Rate theory for correlated processes: Double-jumps in adatom diffusion. *Physical Review Letters*, 79:2843, 1997.
[48] Richard D. James. Private communication.
[49] Roderick V. Jensen and Christopher R. Myers. Images of the critical points of nonlinear maps. *Physical Review A*, 32:1222–1224, 1985.
[50] F. Jülicher, A. Ajdari, and J. Prost. Modeling molecular motors. *Reviews of Modern Physics*, 69:1269–1281, 1997.
[51] A. I. Khinchin. *Mathematical foundations of information theory.* Dover, New York, 1957.
[52] Pamela Davis Kivelson. Neur-on. www.neur-on.com , 2002.
[53] Matthew C. Kuntz, Paul Houle, and James P. Sethna. Crackling noise? simscience.org/ , 1998.
[54] Matthew C. Kuntz and James P. Sethna. Hysteresis, avalanches, and noise: Numerical methods. www.lassp.cornell.edu /sethna/hysteresis/code/, 1998.
[55] Matthew C. Kuntz and James P. Sethna. Hysteresiswin.exe. www.physics.cornell.edu/sethna/StatMech/HysteresisWin.exe, 2004.

[56] L. D. Landau and E. M. Lifshitz. *Quantum Mechanics, Non-relativistic Theory, Second Edition*. Pergamon Press, Oxford, 1965.

[57] L. D. Landau and E. M. Lifshitz. *Statistical Physics, Course of Theoretical Physics Volume 5*. Butterworth Heinemann, Oxford, 1980.

[58] James A. Langer. Statistical theory of the decay of metastable states. *Ann. Phys. (NY)*, 54:258–75, 1969.

[59] Stephen A. Langer, Eric R. Grannan, and James P. Sethna. Nonequilibrium entropy and entropy distributions. *Physical Review B*, 41:2261, 1990.

[60] Stephen A. Langer and James P. Sethna. Textures in a chiral smectic liquid crystal film. *Physical Review A*, 34:5305, 1986.

[61] Stephen A. Langer and James P. Sethna. Entropy of glasses. *Physical Review Letters*, 61:570, 1988. (M. Goldstein noticed the bounds earlier).

[62] B. J. Last and D. J. Thouless. Percolation theory and electrical conductivity. *Phys. Rev. Lett.*, 27:1719–1721, 1971.

[63] J. L. Lebowitz and O. Penrose. Modern ergodic theory. *Physics Today*, pages 23–29, February 1973.

[64] J. Liphardt, B. Onoa, S. B. Smith, I. Tinoco, and C. Bustamante. Reversible unfolding of single rna molecules by mechanical force. *Science*, 292 (5517):733–737, 2001.

[65] Robert S. Maier and D. L. Stein. Escape problem for irreversible systems. *Physical Review E*, 48:931–938, 1993.

[66] O. Malcai, D. A. Lidar, O. Biham, and D. Avnir. Scaling range and cutoffs in empirical fractals. *Physical Review E*, 56:2817–2828, 1997.

[67] Burton G. Malkiel. *The Random Walk Guide to Investing: Ten Rules for Financial Success*. W. W. Norton and Company, New York, 2003.

[68] Burton G. Malkiel. *A Random Walk Down Wall Street*. W. W. Norton and Company, New York, 2004.

[69] N. Marković, C. Christiansen, and A. M. Goldman. Thickness–magnetic field phase diagram at the superconductor–insulator transition in 2d. *Physical Review Letters*, 81:5217–5220, 1998.

[70] Paul C. Martin. Problème à n corps (many-body physics). In *Measurements and Correlation Functions*, pages 37–136. Gordon and Breach, New York, 1968.

[71] Jon Mathews and Robert L. Walker. *Mathematical Methods of Physics*. Addison-Wesley, Redwood City, California, 1964.

[72] Gary McGath and Sergey Buldyrev. The self–avoiding random walk. polymer.bu.edu/java/java/saw/saw.html. BU Center for Polymer Studies.

[73] N. D. Mermin. Lattice gas with short–range pair interactions and a singular coexistence–curve diameter. *Physical Review Letters*, 26:957–959, 1971.

[74] Jehoshua Bruck Michael Gibson. Efficient exact stochastic simulation of chemical systems with many species and many channels. *J. Phys. Chem.*, 104:1876–1899, 2000.

[75] Christopher R. Myers. This exercise, and the associated software, was written in collaboration with Christopher R. Myers.

[76] Mark E. J. Newman. Models of the small world. *J. Stat. Phys.*, 101:819–841, 2000.

[77] Mark E. J. Newman. Scientific collaboration networks. ii. shortest paths, weighted networks, and centrality. *Physical Review E*, 64:016132, 2001.

[78] Mark E. J. Newman and G. T. Barkema. *Monte Carlo Methods in Statistical Physics*. Oxford University Press, Oxford, 1999.

[79] Mark E. J. Newman and D. J. Watts. Renormalization–group analysis of the small–world network model. *Physics Letters A*, 263:341–346, 1999.

[80] National Institute of Health. National human genome research institute – talking glossary: Rna. www.genome.gov /Pages/Hyperion//DIR/VIP/Glossary/Illustration/rna.shtml.

[81] L. Onsager. Motion of ions – principles and concepts. *Science*, 166 (3911):1359, 1969. 1968 Nobel Prize lecture.

[82] Giorgio Parisi. *Statistical Field Theory*. Perseus Book Group, 1988.

[83] Leon Poon. Cat map. University of Maryland Chaos Group, www-chaos.umd.edu/misc/catmap.html.

[84] William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery. *Numerical Recipes in C++ [C, Fortran, ...], The Art of Scientific Computing, 2nd ed.* Cambridge University Press, Cambridge, 2002.

[85] K. Rajagopal and F. Wilczek. Enforced electrical neutrality of the color–flavor locked phase. *Physical Review Letters*, 86:3492–3495, 2001.

[86] Kristen S. Ralls and Robert A. Buhrman. Microscopic study of 1/f noise in metal nanobridges. *Phys. Rev. B*, 44:5800–17, 1991.

[87] Kristen S. Ralls, Daniel C. Ralph, and Robert A. Buhrman. Individual-defect electromigration in metal nanobridges. *Phys. Rev. B*, 40:11561–70, 1989.

[88] K. O. Rasmussen, T. Cretegny, P. G. Kevrekidis, and N. Gronbech-Jensen. Statistical mechanics of a discrete nonlinear system. *Phys. Rev. Lett.*, 84:3740, 2000.

[89] Matthew Sands Richard P. Feynman, Robert B. Leighton. *The Feynman Lectures on Physics*. Addison-Wesley, Menlo Park, CA, 1963.

[90] Craig Rottman, Michael Wortis, J. C. Heyraud, and J. J. Métois. Equilibrium shapes of small lead crystals: Observation of pokrovsky-talapov critical behavior. *Physical Review Letters*, 52:1009–1012, 1984.

[91] Andrew D. Rutenberg and Ben P. Vollmayr-Lee. Anisotropic coarsening: Grain shapes and nonuniversal persistence exponents. *Phys. Rev. Lett.*, 83:3772–3775, 1999.

[92] Evelyn Sander. Marlowe the cat starring in the arnol'd cat map movie. math.gmu.edu/~sander/movies/arnold.html.

[93] Daniel V. Schroeder. *Thermal Physics*. Addison-Wesley Longman, San Francisco, 2000.

[94] U. T. Schwarz, L. Q. English, and A. J. Sievers. Experimental generation and observation of intrinsic localized spin wave modes in an antiferromagnet. *Phys. Rev. Lett.*, 83:223, 1999.

[95] James P. Sethna. Frustration, curvature, and defect lines in metallic glasses and cholesteric blue phases. *Physical Review B*, 31:6278, 1985.

[96] James P. Sethna. Order parameters, broken symmetry, and topology. In *1991 Lectures in Complex Systems: The Proceedings of the 1991 Complex Systems Summer School, Santa Fe, New Mexico*, volume XV. Addison Wesley, 1992. www.lassp.cornell.edu/sethna/OrderParameters/Intro.html.

[97] James P. Sethna. The blue phases. www.lassp.cornell.edu /sethna/LiquidCrystals/BluePhase, 1995.

[98] James P. Sethna. Equilibrium crystal shapes. www.lassp.cornell.edu sethna/CrystalShapes, 1995.

[99] James P. Sethna. Jupiter! the three body problem. www.physics.cornell.edu/sethna/teaching/sss/jupiter/jupiter.htm, 1996.

[100] James P. Sethna. Cracks and elastic theory. www.lassp.cornell.edu /sethna/Cracks/Zero_Radius_of_Convergence.html, 1997.

[101] James P. Sethna. Quantum electrodynamics has zero radius of convergence. www.lassp.cornell.edu/sethna/Cracks/QED.html, 1997.

[102] James P. Sethna. Stirling's formula for n! www.lassp.cornell.edu /sethna/Cracks/Stirling.html, 1997.

[103] James P. Sethna. What is coarsening? www.lassp.cornell.edu /sethna/Coarsening/What_Is_Coarsening.html, 1997.

[104] James P. Sethna. What is the radius of convergence? www.lassp .cornell.edu/sethna/Cracks/What_Is_Radius_of_Convergence.html, 1997.

[105] James P. Sethna. The ising lab. www.physics.cornell.edu/sethna /teaching/sss/ising/ising.htm, 1998.

[106] James P. Sethna. Wolff algorithm ising model. www.physics .cornell.edu/sethna/StatMech/LmcSolution.exe, 2000.

[107] James P. Sethna. Frustration and curvature. www.lassp.cornell.edu /sethna/FrustrationCurvature, 2002.

[108] James P. Sethna. Entropy, order parameters, and emergent properties web site. www.physics.cornell.edu/sethna/StatMech/, 2004.

[109] James P. Sethna, Karin A. Dahmen, and Christopher R. Myers. Crackling noise. *Nature*, 410:242, 2001.

[110] James P. Sethna and Matthew C. Kuntz James P. Sethna. Hysteresis and avalanches. www.lassp.cornell.edu /sethna/hysteresis/, 1996.

[111] James P. Sethna, Joel D. Shore, and Ming Huang. Scaling theory for the glass transition. *Physical Review B*, 44:4943, 1991. (Based on discussions with Daniel Fisher, see erratum *Phys. Rev. B* **47**, 14661 (1993)).

[112] Claude E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27:379–423, 1948.

[113] Joel D. Shore, Mark Holzer, and James P. Sethna. Logarithmically slow domain growth in nonrandom frustrated systems: Ising models with competing interactions. *Phys. Rev. B*, 46:11376–11404, 1992.

[114] Eric D. Siggia. Late stages of spinodal decomposition in binary mixtures. *Phys. Rev. A*, 20:595, 1979.

[115] Public Broadcasting System. The formula that shook the world. www.pbs.org/wgbh/nova/stockmarket/formula.html , 2000.

[116] Lei-Han Tang and Qing-Hu Chen. Finite-size and boundary effects on the i-v characteristics of two-dimensional superconducting networks. *Physical Review B*, 67:024508, 2003.

[117] S. B. Thomas and G. S. Parks. *J. Phys. Chem.*, 35:2091, 1931.

[118] Chikashi Toyoshima, Masayoshi Nakasako, Hiromi Nomura, and Haruo Ogawa. Crystal structure of the calcium pump of sarcoplasmic reticulum at 2.6å resolution. *Nature*, 405:647–655, 2000.

[119] Princeton University. Adapted from author's graduate preliminary exam, fall 1977.

[120] Richard Vawter. Van der waals equation of state. www.ac.wwu.edu /~vawter/PhysicsNet/Topics/Thermal/vdWaalEquatOfState.html.

[121] Duncan J. Watts and Stephen H. Strogatz. Collective dynamics of "small–world" networks. *Nature*, 393:440–442, 1998.

[122] H. G. Wells. *The Time Machine*. Online., several places, 1895.

[123] H. Yin, M. D. Wang, K. Svoboda, R. Landick, S. M. Block, and J. Gelles. Modeling molecular motors. *Science*, 270:1653–1657, 1995.

[124] J. Zinn-Justin. *Quantum field theory and critical phenomena (3rd edition)*. Oxford University Press, Oxford, 1996.

[125] A. Zunger, L. G. Wang, G. L. W. Hart, and M. Sanatai. Obtaining ising-like expansions for binary alloys from first principles. *Modelling Simul. Mater. Sci. Eng.*, 10:685–706, 2002.