New exercises for Entropy, Order Parameters, and Complexity: Statistical Mechanics

James P. Sethna

March 18, 2024

Contents

Chapter 1

Entropy, Order Parameters, and Complexity: New Exercises

Recent exercises, developed after publication, for *Statistical Mechanics: Entropy, Order Parameters, and Complexity, Second Edition*, James P. Sethna, Oxford University Press, 2021.

N1.1 **Entropy and currents.** ⓟ

Consider a locally conserved density $\rho(x,t)$ in an isolated one-dimensional system with a current $J(x,t)$. Imagine that the system has a complicated diffusion constant, with strong density dependence and also dependent on position, so $J(x,t)$ may not be simply related to ρ. Imagine, however, that the internal energy is not important in the system, so entropy $S = -k_B \int \rho \log \rho \, \mathrm{d}x$ must increase with time. Are there conditions on $J(x,t)$ which would guarantee this?

Show that the entropy will increase with time if $J(x,t)$ is in the direction of decreasing ρ. (This is closely related to Exercise 5.10; again, you will need to integrate by parts. You may assume $J(x)$ and $\rho(x)$ go to zero as $x \to \pm\infty$.)

N1.2 **First-digit law and priors.** (Statistics) ⓟ

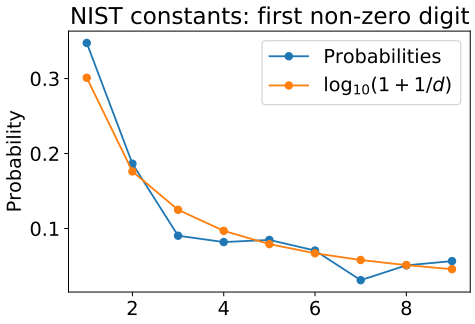


Fig. N1.1 Fraction of first digits for 354 fundamental physical constants. (2019 CODATA internationally recommended values [1]).

Bayesian statistics, like statistical mechanics, incorporates known experimental results into a probabilistic prediction for the behavior of the system in the future (see Exercise N5.3). In statistical mechanics, if we only know the energy of a system then Liouville's theorem tells us that all points in the energy shell are equally likely *a priori*. In Bayesian statistics, they have no theorem like Liouville's, so they need to assume a *prior*. For example, if you want to estimate a time constant τ for a chemical reaction (which can range from nanoseconds to years), you might want a prior $P_\tau(\tau)$ that gives equal weight to each decade: finding τ in the range $(10^{-9}, 10^{-8})$ seconds is equally plausible as finding τ in the range $(10^5, 10^6)$ seconds.

Show that $P_\tau(\tau) \propto 1/\tau$ has this reasonable property. Show that this choice also makes the decay rates $\Gamma = 1/\tau$ have this same nice property: $P_\Gamma(\Gamma) \propto 1/\Gamma$. (Hint: If τ lies in a small range $\Delta\tau$, then Γ will lie in a corresponding small range $\Delta\Gamma$, so $P_\Gamma(\Gamma)|\Delta\Gamma| = P_\tau(\tau)|\Delta\tau|$.) *Show that this distribution predicts that the first non-zero decimal digit d of τ will have probability $\log_{10}(1+1/d)$ (Fig. N5.7).* (Hint: Do it assuming τ lies in one decade first.) *Show your steps.* (Note: Feel free to consult the extensive discussions on the Web.)

Simon Newcomb, using a book of logarithms in 1881[1] discovered this by noticing that the pages in the beginning (1.000001, 1.000002, ...) were dirtier than the ones at the end (9.000001, 9.000002, ...). Frank Benford fleshed this out in 1938, showing that areas of rivers, molecular weights of compounds, and physical constants like the proton mass, Planck length, and Avogadro's constant (Fig. N5.7) also obey this law.

N1.3 **Accelerators vs. ergodicity.**[2] (Accelerator, Mathematics, Computation, Dynamical

[1] Before calculators, people used printed books of logarithms, which allow one to multiply and divide quickly.

[2] Hints for the computations can be found at the book website [31].

systems) ③

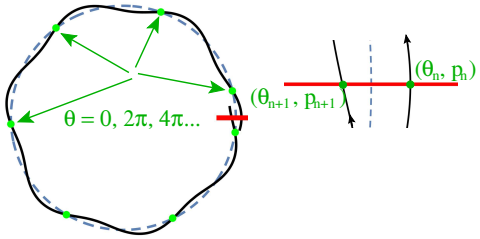


Fig. N1.2 Orbit of particle around synchrotron. Dashed line shows the intended center of the beam; solid line the trajectory of a particular particle. The particle oscillates around the center with a phase θ and an amplitude proportional to $\sqrt{p}$, with anharmonicity determined by K. The standard map is a Poincaré section; it tells one what θ and p will be as the particle next crosses the horizontal axis along the red line.

Synchrotrons push charged particles around circular orbits at near the speed of light, stably over billions of orbits (Fig. **??**). Why haven't the nonlinear fields used to steer the particles kicked them out of the beam? Like the cat map in Exercise 5.8 and the three-body problem of Exercise 4.4, physicists distill the dynamics using a Poincaré section into a two-dimensional area-preserving nonlinear map. The classic example, due to Chirikov, is the so-called *standard map*

$$\begin{aligned} \theta_{n+1} &= \theta_n + p_{n+1} = \theta_n + p_n + K\sin(\theta_n) \\ p_{n+1} &= p_n + K\sin(\theta_n). \end{aligned} \tag{N1.1}$$

Here p is proportional to the squared amplitude of the deviation from center of beam and θ represents the phase of the oscillation around that center (Figure **??**). Large K represents a strong nonlinearity in the fields. For plotting, we shall take θ mod 2π, and occasionally also p mod 2π.

We remember from Exercise 5.8 that area-preserving maps mimic the volume-preservation in phase space guaranteed by Liouville's theorem.

(a) *Show that the standard map (eqn* **??***) preserves area, by showing that its Jacobian has determinant one.*

(b) *Iterate the standard map a few thousand times for* $K = 4$*, starting with* $(\theta_0, p_0) = (0.1, 0.11)$ *(chosen arbitrarily), and plot* p_n *versus* n *for a thousand points* n*.* (For this part, do not take p mod 2π). *Does the amplitude of the particle oscillations around the beam path (as measured by* p*) remain bounded in amplitude as* n *gets large? Would designing an accelerator with this large an anharmonicity be a good idea?*

The situation changes for smaller anharmonicity.

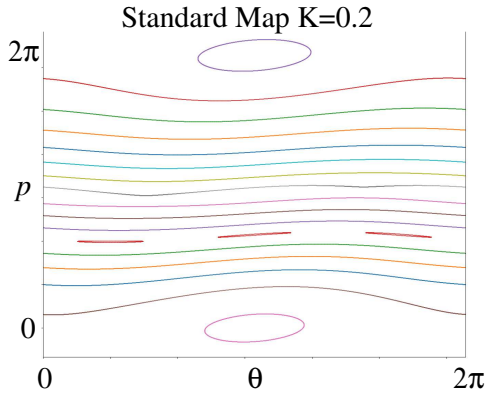


Fig. N1.3 Trajectories at $K = 0.2$ for a variety of initial conditions.

(c) *Implement the standard map graphically, so that you can interactively select a variety of initial conditions and run each for a few thousand steps to explore the trajectories, using* $\mathrm{mod}(2\pi)$ *for* θ *but not for* p*. Check that you get trajectories for* $K = 0.2$ *that are compatible with those shown in Fig.* **??***. Plot the behavior for* $K = 0.7$*, selecting a variety of initial conditions. Be sure to illustrate "oval" KAM*[3] *tori around stable periodic orbits, "horizontal" KAM tori that span from* $\theta = 0$ *to* 2π*, and chaotic regions.* (Remember that your trajectory is a cross section once per orbit around the accelerator. So a closed curve becomes a tube, which mathematically is called a torus.) *Explore initial conditions spanning* $p \in (0, 2\pi)$*. Do the trajectories stay bounded in* p*? Which keeps* p *from growing indefinitely – the oval or horizontal tori?* Keeping the magnets close to the beam is best – until the beam starts hitting the magnets. The coordinate p is related to the squared maximum deviation of the particle from the designed beam center. *Based on your plot, what range* Δp *would you use to set the magnet positions for an accelerator at* $K = 0.7$*?*

The equilibrium state predicted from statistical mechanics would fill all of phase space uniformly. The map at $K = 0.7$ is not *ergodic*.

In Section 4.2 we defined a map (or time evolution) to be ergodic if and only if all the ergodic components of the space either have zero volume or have a volume equal to the space. Here ergodic components were sets R that remain invariant under the map (so if (θ_0, p_0) is in R then (θ_n, p_n) is also in R).

(d) *Identify a few ergodic components in your plot of* $K = 0.7$ *in part (c). Are the KAM tori ergodic components? Do they have non-zero volume? Do they surround ergodic components with non-zero volume? Does the chaotic regions surrounding* $\theta \approx p \approx 0$ *appear to be an ergodic component with non-zero volume?*

(e) *Alter your standard map to also take* p *mod* 2π*, to keep the motion bounded.*[4] *Plot*

[3] KAM stands for Kolmogorov, Arnol'd, and Moser, who proved that tori survive for certain irrational *winding numbers*; see part (e).

[4] Note that your map of $K = 0.7$ in part (c) is periodic in p with period 2π, so this makes sense.

and examine carefully the time evolution at $K = 4$. Do all initial conditions fill out the entire volume? If not, describe ergodic components that neither have zero volume nor fill the square.

We define the *winding number* of a particle trajectory as the average number of oscillations $\theta/2\pi$ per iteration of the map. The KAM theorem tells us that the tori whose winding numbers are difficult to approximate by rationals are the most stable.

(f) *What is the winding number at $K = 0$, for a trajectory starting at initial condition (θ_0, p_0)? Write a routine to approximate the winding number for $K = 0.7$ and $\theta = \pi$, and plot it for $0 < p < 2\pi$. By examining your plot of the different trajectories for $K = 0.7$ from part (c), explain the plateaus you observe at rational winding numbers.*

The last "horizontal" KAM torus for the standard map is destroyed by the nonlinearity at $K_c = 0.971635\ldots$. Its winding number is the inverse Golden Mean, $(\sqrt{5}-1)/2 = 0.618033\ldots$, which is the most irrational number – the number hardest to approximate by ratios of small integers. This transition is associated with self-similar behavior and scaling, and has been studied using the renormalization group methods we study in Chapter 12.

N1.4 **Bosons in two states.** (Quantum) ③

(a) *Consider three noninteracting identical bosons in a system with two single-particle energy eigenstates at energies $E_0 = 0$ and $E_1 = \epsilon$ at temperature T. Give a formula for the partition function Z and the expected number of bosons $\langle n \rangle$ in the excited state E_1, as a sum over states.*

(b) *What would your answer be for $\langle n \rangle$ if there were an* infinite *number of bosons distributed between the two levels? What is the expected number of bosons in the upper state at temperature $k_B T = 10\epsilon$?* (Hint: How does your answer relate to the thermal occupation probabilities for harmonic oscillator eigenstates?)

We treat the excitations of harmonic oscillators as particles (phonons and photons) because they act like noninteracting bosons (see Exercise 7.2). Each harmonic oscillator acts like a different single-particle eigenstate, for a grand canonical ensemble of bosonic photons with chemical potential zero. The n^{th} harmonic excitation is n photons occupying the state. Our lower level here can be viewed as a particle analogy of the heat bath; the heat bath sets the temperature, our lower level sets the chemical potential.

(c) *Now consider a system with* 1000 *distinguishable, noninteracting particles that can each occupy the same two states, $E_0 = 0$ and $E_1 = \epsilon$. What is the expected number of distinguishable particles in the upper state at temperature $k_B T = 10\epsilon$? Discuss how and why the behavior of bosons and distinguishable particles are different in this exercise.*

(d) *Are the bosons in part (b) forming a Bose–Einstein condensate?*

N1.5 Beer and bubble nucleation. ③

A glass of beer illustrates many aspects of statistical mechanics. The head of foam at the top of the beer exhibits rigidity like a solid (Exercise N1.7), and the foam coarsens with time (Exercise N1.6). In this exercise, we shall study the phase transitions triggered when one opens the bottle. You may wish to experiment with a bottle or can of beer or soda.

We shall model beer and soda, for the purposes of this exercise, as a phase consisting of carbon dioxide and water. When one opens the bottle, the absolute pressure drops from around $P_{\text{beer}} \sim 200\,\text{kPa}$ (depending on the beer), to atmospheric pressure $P_{\text{atm}} \approx 100\,\text{kPa}$. The solubility of CO_2 at 0 C in water at atmospheric pressure is about $1.8\,\text{cm}^3$ CO_2 per gram of water [15]. Note that there is a small amount of gas above the beer. Henry's law tells us that the number of dissolved CO_2 molecules is proportional to its partial pressure in the gas.[5] H_2O at 0 C has a partial pressure of 0.6 kPa, so we assume that the pressure in the gas atop the beer bottle is almost completely due to CO_2.

When you pour a glass of beer, much of the volume is filled with foam. Bartenders will pour off the foam to give you a full glass of liquid. Assume the carbonated liquid starts at P_{beer}, and ignore the volume change in the liquid as the pressure drops and the gas separates. Assume the two form a closed system.

(a) *As the pressure drops and the liquid separates from the vapor, what fraction of the volume in the system at the new equilibrium is gas? Does that correspond qualitatively to your observations?* (Note that cold beer may need to be shaken considerably to extract all of the gas.)

Let x_{CO_2} be the fraction of the molecules in the liquid that are CO_2 molecules. Henry's law for our system can be written

$$P_{CO_2} = K x_{CO_2}, \tag{N1.2}$$

with Henry's constant $K \approx 70\,\text{MPa}$ at 0 C (ice-cold beer) [6].

(b) *Compare the solubility determined from Henry's constant to the quoted solubility 1.8 cm^3/gram.*

We shall use X to denote the fraction of CO_2 molecules in the entire system: the total number of CO_2 molecules (both dissolved and in vapor), divided by the total number of molecules including the water (both liquid and vapor). Hence $x = X$ if the CO_2 is entirely dissolved in the liquid.

The vapor pressure of water at $T = 0\,\text{C}$ is $\sim 0.6\,\text{kPa}$, independent of the pressure and roughly independent of the amount of CO_2 dissolved. CO_2 is a gas at all pressures of interest, and both gases are well-approximated as ideal gases of molecules. Let us draw the P–X phase diagram.

[5] The pressure of a gas mixture in the ideal gas approximation is the sum of the *partial pressures* contributed by each constituent.

(c) *Use Henry's law to determine the total pressure (vapor pressure of water plus CO_2 partial pressure) at which the liquid beer begins to coexist with the vapor. Find the pressure at low pressures and high CO_2 molar fraction X at which the liquid disappears (the gas begins to coexist with the liquid). Draw the P–X phase diagram for the H_2O / CO_2 system at $T = 0\,C$, for $P \in (0, 400)\,kPa$ and $X \in (0, 0.01)$, and include a line connecting the beer initial and final pressures.*[6] *Draw the phase diagram again for $P \in (0, 10)\,kPa$ and $X \in (0, 1)$ to illustrate the other side of the coexistence curve. In both diagrams, indicate the liquid phase, the gas phase, and the two-phase coexistence region.*

It is apparent that the phase separation in beer is nucleated through the formation of bubbles of gas. Let us calculate the nucleation rate of a spherical bubble. The surface tension of water in contact with air is $\sigma \sim 7.56\,\mathrm{N/m}$ at $0\,\mathrm{C}$. The prefactor Γ_0 for the Arrhenius rate (called "prefactors" in eqn 11.11) may be roughly estimated [5, 3] as $\Gamma_0 = 4 \times 10^{41}\,\mathrm{m^{-3}\,s^{-1}}$, corresponding to a physically sensible prefactor $\approx 10^{13}/\mathrm{s}$ of the number of attempts to form a critical droplet centered at each water molecule in the cubic meter.

(d) *What is the pressure energy gained by forming a bubble of radius R? What is the surface tension energy cost? What is the radius of the critical bubble? What is the nucleation barrier for generating a bubble in the bulk of the beer (homogeneous nucleation)? Would you plan to wait until a bubble forms when you open a half-liter bottle of beer?*

Water has an unusually large surface tension, and our calculation clearly shows that homogeneous nucleation cannot occur. Even for extremely clean liquids with low surface tension (like liquid helium), homogeneous nucleation almost always is unimportant compared to nucleation on surfaces, dust particles, dislocation tangles, scratches on surfaces, etc. Many of your beer bubbles probably form on small flaws on the glass surface. If your glass of beer has not gone flat yet, see if you notice a chain of bubbles all emitting from the same spot on the glass.

N1.6 **Beer foam and coarsening.**[7] ③

Pour a beer. Note that the bubbles rise to the top under gravity, and form a foam—the "head" of a beer. (If you do not have a beer handy, soap suds or shampoo bubbles will work as well.) As time goes on, the bubbles in the foam will grow larger, as the big bubbles absorb the little ones. This is loosely reminiscent of the coarsening seen after abrupt phase transitions (Section 11.4.1).

There are two major mechanisms for this coarsening: the popping of the walls between bubbles as they grow thin, and the diffusion of gas between bubble walls. Which is more important can be controlled through changing the composition of the fluid. In

[6] Let X be determined by the liquid portion of the beer; assume the small amount of gas at the top escaped when we open the bottle.

[7] Thanks to Douglas Durian for helpful consultations and Fig. N1.2.

this exercise, we shall study a theory of bubble coarsening via diffusion, following a classic analysis in two dimension by von Neumann [33] and Mullins [23] (see [18]).

The bubbles near the top of the beer are *dry*; they have thin walls separating relatively large bubbles. In a "two-dimensional" experiment with thickness much smaller than the bubbles, these walls form a network of curves connecting nodes where three bubbles touch (Fig. N1.2). These walls move to balance the forces on them, on a time scale much faster than the diffusion of gases between bubbles that drive the coarsening we study.

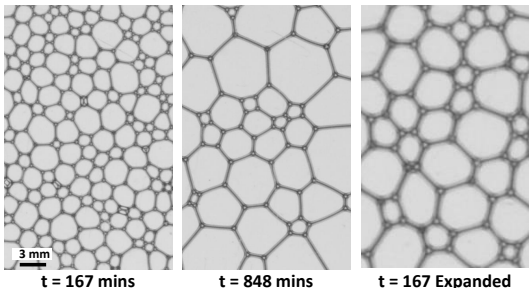


Fig. N1.4 Foam scaling. Two snapshots of a simulated 2D foam, coarsening by diffusion, with elapsed time t in minutes, and a rescaled version of the early time snapshot (graciously provided by Anthony Chieco and Douglas Durian). Note that the patterns look statistically similar, apart from an overall growth in the size of the bubbles.

Let us first examine the structure of the force-balanced configurations. Let the line tension (energy per unit length) of the walls be λ, and assume the energy per node is negligible.

(a) *Draw a free body diagram for the forces on a node at a three-wall intersection.* (Nodes where more than three bubbles meet are *non-generic*. That is, any tiny perturbation will split them into three-fold intersections.) *What force will each wall exert on the node? Show that the three walls must intersect at 120° angles.*

Each bubble will be filled with a gas of a certain pressure. The pressure difference ΔP across a wall determines its radius of curvature $R = \Lambda/\Delta P$.

(b) *Determine the constant Λ in terms of the wall tension λ.*

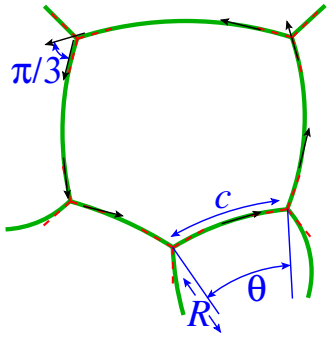


Fig. N1.5 2D Bubble geometry. A thin-walled bubble with five neighbors, whose walls are arcs of circles of angle θ_i, radii R_i, and arc length c_i. The bubble walls meet at nodes, forming angles of $2\pi/3 = 120°$ (red dashed lines). The local tangent vector to the bubble perimeter winds by 2π as one circles the bubble (black arrows), but jumps by $\pi/3 = 60°$ every time it passes a node. It rotates by the opening angle θ during the arc c_i for each bubble.

Hence a thin-walled bubble with n neighbors with pressure changes ΔP_i will be formed by n circular arcs of radii $R_i \propto 1/\Delta P_i$ connecting the nodes where walls meet. As one might imagine, the rate of gas diffusion through the wall per unit length is given by a constant μ times the pressure drop ΔP. So

$$\mathrm{d}A_n/\mathrm{d}t = \mu \sum_{i=1}^{n} \Delta P_i c_i \tag{N1.3}$$

where c_i is the arc length of the corresponding wall (Fig. N1.3).

John von Neumann, known for his work in computer science and game theory, deduced a law for the size evolution of two-dimensional domains [33] like our bubbles. The area A_n of a domain with n neighbors will vary as

$$\mathrm{d}A_n/\mathrm{d}t = K_0(n-6) \tag{N1.4}$$

The bubbles with only a few neighbors vanish. But the average number of neighbors should stay around six (which you can show using Euler's theorem about faces, edges, and nodes).

(c) *Find a video of this coarsening process, such as that in Ref. [16]. Identify a bubble with six neighbors. Does it stay roughly the same size? When it adds / subtracts a neighbor, does it start growing / shrinking as expected? Identify a bubble with more than six neighbors at early times, and follow its evolution. Does it die? (If not, start again with a second many-neighbored bubble. Most bubbles should eventually disappear, since so few survive.) Does the number of neighbors of your bubble fluctuate both up and down?*

What was von Neumann's argument? It was purely geometrical.

(d) *Show that* $\Delta P_i c_i = \Lambda \theta_i$, *where* θ_i *is the angle between the tangents to the* i^{th} *boundary at the nodes on either end as shown in Fig. N1.3 (negative for convex boundaries 'bulging out', positive for concave boundaries). Noticing that the tangent vector jumps by* $\pi/3 = 60^\circ$ *at each node (Fig. N1.3) and that the tangent vector must rotate by* 2π *as it winds around the bubble, derive von Neumann's law (eqn N1.3). What is* K_0 *in terms of* μ *and* Λ*?*

For non-conserved order parameters, we saw in Section 11.4.1 that the typical length scale $L(t) \sim t^\beta$ with $\beta = \frac{1}{2}$. For conserved order parameters, where the particles or spins must diffuse over a length scale $L(t)$ to rearrange the structure, we found $\beta = \frac{1}{3}$.

What is β for 2D dry foams? In Section 11.4.1, we argued qualitatively that if droplets or features of size L *disappear* in a time $t \propto L^{1/\beta}$, then the typical size of the surviving features will *grow* as $L \sim t^\beta$. In the next part, you can give an analogous argument for foams.

(e) *In part (c), you found that the number of neighbors of a droplet changed in a complicated way with time. For simplicity, let us consider a droplet with a constant number of neighbors* n*, which at* $t = 0$ *has area* A_0*. Write a formula for* $A(t)$*. At what time* t_0 *does* $A(t)$ *vanish? What happens to* t_0 *for* $n > 6$*? If we define* $L(t) = \sqrt{A(t)}$*, write* $L(t) = C(n)|t - t_0|^\beta$*. What is* $C(n)$*? Does* β *depend on the number of sides* n*?*

See also Exercise N1.5, which studies the bubble formation in beer, and Exercise N1.7, which studies the transition from liquid beer to rigid foam ...

N1.7 **Beer and rigidity: Jamming.**[8] ③

Beer is a liquid: the bubbles in the beer do not prevent it from pouring. But the head of a beer (the frothy foam on the top of the beer where the bubbles accumulate) is rigid. The foam is a random packing of bubbles that can support shear, just as the glass containing the beer is a random arrangements of atoms bonded together that does not flow under stress.

Fig. N1.6 A head of beer. Beer is liquid on the bottom and forms a rigid network of bubbles in the foam at the top.

[8] This exercise was developed in collaboration with Stephen Thornton.

As in Exercise N1.5, we can view the beer as exploring a slice of the vapor/fluid phase diagram, with gravity pulling the fluid downward leading to polyhedral bubbles separated by thin walls forming the head, gradually becoming rounder and wetter with depth, until they separate into individual spherical bubbles in the bulk of the glass.

Somewhere, as we move upward in the glass, there must be a first point where we go from a fluid phase to a rigid phase.[9] What is the nature of this phase transition?

For large objects like bubbles, powders, sand, and colloids, this transition is a problem in mechanics. When two bubbles touch, they resist moving closer together, which we can think of as an overlap energy. The bubbles can therefore pack together without costing overlap energy only until there are enough contacts that each bubble is is held in place by its neighbors.

This kind of rigidity was first studied by Maxwell[10] in his study of the stability of mechanical trusses. A contact between two spherical bubbles constrains their separation to be the sum of their radii; a truss is a rigid rod that constrains its endpoints to be a fixed distance apart. The Maxwell criterion involves subtracting the number of constraints in a system (bubble contacts) from the the number of motional degrees of freedom of the system (three translations per bubble). Roughly speaking,[11] when the two are equal the system becomes rigid. If there are fewer constraints, the system will be floppy, with each missing constraint replaced by a zero-energy mode of deformation.

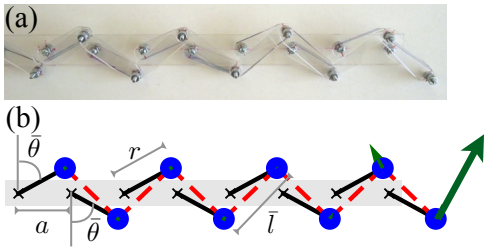


Fig. N1.7 Rigidity and Maxwell constraint counting. An array of nodes (blue dots), connecting rotors (black lines) to springs (red dashed lines). Is the system floppy or stiff? Each rotor and each spring can be viewed as a constraint. The two coordinates of the nodes are each degrees of freedom. The green arrows indicate that there is at least one floppy mode. With permission from Chen et al. [8].

Let us consider an example, a mechanical network [8] inspired by recent work on topological insulators. Figure N1.5 shows the right-hand side of a chain of rotors whose ends are nodes connected by springs.

[9] One can envision a tall beer in weak gravity, which could act to gradually change the chemical potential of the fluid and the gas. On Earth the gradient of chemical potential for a bubble is rather steep.

[10] Yes, it is the same Maxwell from electromagnetism and the Maxwell relations.

[11] The Maxwell criterion has various refinements. Truss networks can have states of *self stress*, where some regions are overconstrained. A network of fibers can become rigid with fewer constraints if they are pulled straight; this can be important in cells. Ellipsoids can become rigid even though they can still rotate. None of these subtleties are significant in bubble networks.

(a) *Check, for an infinite system, that the number of constraints (springs and rotors) per node in this mechanical network would be equal to the number of directions the node can move.* The figure shows that when the chain ends, the missing final spring leads to a floppy mode. The other end of the chain also ends with a rotor with only one spring attached. *Use the Maxwell criterion to argue that there is only one floppy mode: the other dangling end of the chain must be rigid.* (See Ref. [8] for many more fascinating aspects of the problem.)

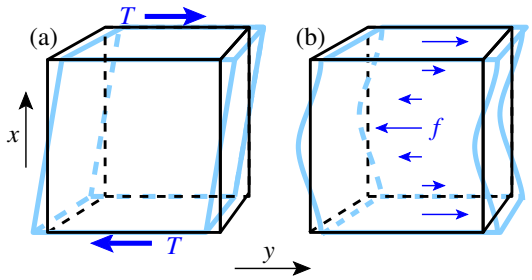


Fig. N1.8 Force densities and tractions. (a) The shear elastic constant of a solid and the viscosity of a liquid are both defined as the response to a traction T (a force per unit area) applied to two ends 0 and L of a sample. (b) The transverse susceptibility $\widetilde{\chi}_T(q,\omega)$ is the response to a force per unit volume $\mathbf{f}(x,y,z,t) \propto \exp(\mathrm{i}qx)\exp(-\mathrm{i}\omega t)\,\widehat{\mathbf{y}}$. We can view the traction in (a) as a force per unit volume $f(x,y,z) = T(\delta(x-L)-\delta(x))$.

Let us prepare to study this rigidity transition by considering the response of normal solids and liquids to external shear. At long lengths and low frequencies, both liquids and isotropic solids respond to a one-dimensional, time-dependent, transverse force density $\mathbf{f}(x,y,z,t) \equiv f(x,t)\widehat{\mathbf{y}}$ by obeying the equation

$$\rho\frac{\partial^2 u}{\partial t^2} = G\frac{\partial^2 u}{\partial x^2} + \eta\frac{\partial^3 u}{\partial x^2 \partial t} + f \tag{N1.5}$$

where $u(x,t)$ is the motion in the $\widehat{\mathbf{y}}$ direction, ρ is the mass density, G is the shear modulus, and η is the dynamic viscosity (see Fig. N1.6(b)).[12]

We can test this equation by considering the case of uniform shear, given by a traction on the top and bottom surfaces of the cube $f(x,t) = T(\delta(x-L)-\delta(x))$.

(b) *If the material is rigid ($G > 0$), show that this force can result in a static deformation satisfying $G\,\mathrm{d}u/\mathrm{d}x = T$, agreeing with the standard definition of the shear modulus.*

(c) *If the material is a liquid ($G = 0$), write an equation for $\dot{u}$, and show that this force produces a steady-state velocity gradient $\eta\,\mathrm{d}\dot{u}/\mathrm{d}x = T$, agreeing with the standard definition of the viscosity.*

[12] Note that the viscosity is in the form of Kelvin damping, the lowest-order form allowed by Galilean invariance (see Exercises 9.6, 9.14, and 10.9).

(d) *Use the equation of motion N1.4 to write the transverse susceptibility at long wavelengths* $\widetilde{\chi}_T(q,\omega) = \tilde{u}/\tilde{f}$. *Show that*

$$G^{-1} = \lim_{q\to 0}\lim_{\omega\to 0} q^2 \widetilde{\chi}_T(q,\omega). \tag{N1.6}$$

Show that

$$\eta = \lim_{\omega\to 0}\lim_{q\to 0}(\rho^2\omega^3/q^2)\chi_T''(q,\omega), \tag{N1.7}$$

where as usual $\chi'' = \mathrm{Im}(\widetilde{\chi})$.

Since our transverse susceptibility describes the typical long-wavelength, low frequency behavior of any isotropic fluid or solid,[13] these last two formulas can be considered the definitions of shear modulus and dynamic viscosity for other systems (like foams near jamming) where the behavior is more interesting at short times and lengths.

We now turn our focus back to beer. Maxwell's constraint on the rigidity of systems also holds for general disordered packings of spheres (like the bubbles in our beer). As we move upwards through the foam, there is a point where the foam can first support shear deformation. This is known as the *jamming* transition. The tunable parameter that induces this transition is $\delta z \equiv z - z_c$, where z is the *coordination number* counting the average number of nearest neighbor contacts each bubble has. The critical coordination number $z_c = 2d$ in d dimensions by Maxwell's constraint-counting argument.

Recent work [17] provides us with an explicit formula for the universal scaling function for the transverse susceptibility near the jamming transition. Near jamming in three spatial dimensions in a viscous medium, the transverse part of the susceptibility has a particular universal scaling form in terms of our distance to jamming δz and intrinsic damping parameter γ:

$$\widetilde{\chi}_T(q,\omega) = \Big[aq^2|\delta z|\Big(\sqrt{1 - c\frac{\mathrm{i}\gamma\omega + d\rho\omega^2}{|\delta z|^2}} \pm 1\Big) - \mathrm{i}\gamma q^2\omega - \rho\omega^2\Big]^{-1} \tag{N1.8}$$

where the $\pm$ refers to the solid "rigid foam" and liquid "wet foam" sides of the transition, respectively.

(e) *Calculate shear the modulus* G *of the material (eqn N1.5) in terms of constants and the distance to jamming* $|\delta z|$, *both on the floppy and the rigid side of the transition. Is the transition from rigid to floppy continuous, in terms of the shear elastic constant? Calculate the dynamic viscosity* η *on the floppy side of the transition (eqn N1.6) in terms of constants and the distance to jamming* $|\delta z|$. (Hint: Expand the square root to first order, and use $1/(a+b\mathrm{i}) = (a-b\mathrm{i})/(a^2+b^2)$. The algebra is a bit messy.) *How does the viscosity depend on the distance to jamming? Explain how we might view this as a continuous transition.*

[13] One might consider this a requirement for a material to be described as a fluid or isotropic solid.

So, unlike the abrupt way that liquids becomes rigid when they freeze into crystals, jammed solids have a continuous transition in their response to external shear. Their elastic response to shear ($1/G$) diverges from the rigid side, and their viscous response to shear ($1/\eta$) from the floppy side smoothly goes to zero. Equation N1.7 explains many other unusual properties exhibited by jammed solids like beer foam.

N1.8 **Where is the antimatter?**[14] (Astrophysics) ③

The Universe experimentally is composed primarily of matter. This is a surprise, since in all other ways matter and antimatter appear the same. Sakharov in 1967 proposed a set of three necessary conditions for the period during which the Universe generated this baryon asymmetry.[15]

1. Reactions that do not conserve net baryon number B (the number of baryons minus antibaryons) must be present.
2. The quantum theory of the Universe must violate both the symmetries C and CP (more details below).
3. The Universe must be out of equilibrium.

In this exercise, we shall give a cartoon discussion of "electroweak baryogenesis" [10], which generates this asymmetry as bubble walls sweep through the Universe during an abrupt phase transition.

First, we need to discuss important symmetries in quantum field theory. There are three symmetries of Schrödinger's equation that we often take for granted. P (parity) is the symmetry inverting all positions through a point, $\mathbf{x} \to -\mathbf{x}$. T (time reversal) is the symmetry inverting the direction of time (e.g., taking momenta $\mathbf{p} \to -\mathbf{p}$). And C (charge conjugation) replaces each particle with its antiparticle. It seems clear that breaking charge conjugation symmetry in Sakharov's second condition is necessary to get more matter than antimatter. Parity is also important to Sakharov because of spin. We shall ignore the spin of the baryons in our model of the bubble wall (see note 17 on page 38).

It was a great surprise when it was discovered in 1956 by Chien–Shiung Wu that the weak interaction violates parity symmetry P. (Her discovery was prompted by the theoretical suggestion that it had never been tested.) Thus Sakharov's second condition is satisfied, even in our current Universe. Later, weak interactions were also shown to violate CP. But there is a famous CPT theorem arguing that any Lorentz-invariant quantum field theory must have CPT invariance. That is, if one simultaneously inverts time, space, and swaps particles with their antiparticles, the laws of the Universe would be unchanged.

[14] This exercise was developed in collaboration with Mitrajyoti Ghosh.

[15] Protons and neutrons are baryons, as are all particles made up of three quarks. Three antiquarks make up antiprotons, antineutrons and other antibaryons.

We can treat C, P, T, and baryon number B as quantum operators, like the Hamiltonian of the Universe $\mathcal{H}$. The CPT theorem tells us that $[CPT, \mathcal{H}] = 0$.

(a) *How do C, P, and T individually affect the baryon number operator? In particular, use $C^2 = T^2 = P^2 = \mathbb{1}$ to compute CBC^{-1}, PBP^{-1}, and TBT^{-1} in terms of B.* (Hint: One approach is to check their behavior applied to an eigenstate of baryon number $B|b\rangle = b|b\rangle$.)

We can now derive Sakharov's third condition. You might think that the energy of a baryon is given by its mass, mc^2, so that in equilibrium the densities of baryons and antibaryons would be the same (since their masses are the same). But baryons interact in complex ways with the particles around them (think nuclear physics and the polarization of local vacuum fluctuations), and these are governed by rules that (for the weak interaction) violate C. Can we get a general proof that the equilibrium densities are the same?

An equilibrium Universe has a density matrix[16] $\rho = \exp(-\beta\mathcal{H})$. The net baryon number is thus given by $\langle B\rangle = \mathrm{Tr}(\rho B)$.

(b) *Use CPT symmetry to show that $\langle B\rangle = 0$.* (Hint: Use $(CPT)(CPT)^{-1} = \mathbb{1}$. Insert $\mathbb{1}$ into the trace, use the fact that CPT commutes with the Hamiltonian, and use the cyclic invariance of the trace.)

Experiments in our current cold Universe have never observed baryon number violations. (Even a very small decay rate of protons to, say, a positron, photon, and antineutrino would cause a huge flood of γ rays emanating from the Earth.)

In the early Universe, we believe there was an *electroweak epoch* when particles were massless (like the photon is now), and the four bosons $W^{\pm}$, Z, and photon were symmetric partners of one another. During this epoch when the electroweak symmetry is unbroken, the theory allows for a *sphaleron*[17] reaction that transforms nine quarks (three baryons) into three antileptons (changing B, as Sakharov required). At the electroweak transition, the Higgs boson was created and many particles became massive, and the sphaleron reaction developed a high energy barrier (explaining why protons do not decay now).

The transition in electroweak baryogenesis is abrupt, and happens via the nucleation of rare bubbles that expand at nearly the speed of light—transforming the equilibrium unbroken phase with $\langle B\rangle = 0$ into a high-temperature version of the current Universe. Calculations show that the sphalerons are turned off almost completely in our current Universe (with breaking of the electroweak symmetry).

[16] Do not confuse the density matrix ρ with the baryon density ρ in later parts of this exercise.

[17] Sphalerons are non-perturbative solutions to the equations of motion of the quantum Hamiltonian of our universe. In particular, sphalerons are a minimum action solution to the field equations with a non-trivial topology. The order parameter space is $SU(2)$, and the field configuration takes the Universe (with topology S^3) into $SU(2)$. So the sphaleron can be viewed as a three-dimensional topological defect, whose strength is measured by the homotopy group $\Pi_3(SU(2))$ (given by non-equivalent mappings of the Universe S^3 into the order parameter space).

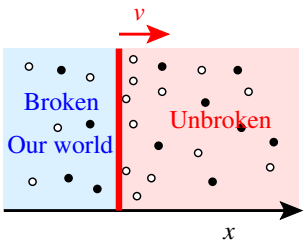


Fig. N1.9 Electroweak transition wall, sweeping at a high velocity v, ending the symmetry-unbroken electroweak epoch and giving our particles masses. The wall interacts with the baryons (black) and antibaryons (white), letting those that impact the wall penetrate with probability $\mathcal{T}$. Deep inside the unbroken phase, sphalerons equilibrate the system and destroy any matter-antimatter asymmetry. Inside our world, sphalerons are inactive and baryon number is conserved.

This bubble wall is wildly out of equilibrium (Fig. N1.7). As it sweeps into the unbroken phase, it encounters baryons and antibaryons, but (because C is not a symmetry) the wall interacts differently with matter and antimatter. In particular, baryons are allowed to enter our universe (the broken-symmetry phase) with probability[18] $\mathcal{T}$, and antibaryons are allowed in with probability $\overline{\mathcal{T}}$.[19] This leads to a pile-up of baryon density $\rho_B(x)$ and antibaryon density $\rho_{\overline{B}}(x)$ ahead of the front. The bubbles are huge, and so the bubble walls are nearly flat. We can thus study this pile-up as a one-dimensional problem, and change to a moving reference frame $y = x - vt$ to find a time-invariant solution to the evolution laws.

We start in parts (c) through (f) by ignoring transitions between baryons and antibaryons, to work out the boundary conditions and the equations of motion in the moving reference frame. This means we can study ρ_B and $\rho_{\overline{B}}$ independently. Let ρ_∞ be the density $\rho_B(\infty)$ deep in the unbroken phase. Since the unbroken phase is presumed to be in equilibrium before the wall sweeps through,[20] $\rho_{\overline{B}}(x)$ will also be equal to ρ_∞ as $x \to +\infty$.

(c) *In steady state, ignoring any transitions changing baryons to antibaryons, can the density $\rho_B(x)$ deep in our universe (as $x \to -\infty$) differ from ρ_∞?* (Hint: What is the current $\widehat{J}(y)$ in the moving reference frame?) *Do the differing transmission probabilities suffice to explain the dearth of antimatter in our world?*

Assume that baryons and antibaryons diffuse with diffusion constant D. Let us begin

[18] Warning: $\mathcal{T}$ is the probability of transmission, not the amplitude transmission constant T used in quantum mechanics and optics.

[19] The spin of a massless particle becomes its *helicity*. So left-handed and right-handed photons carry circular polarization. If our quantum Hamiltonian had CP symmetry, left-handed massless antibaryons impinging from the unbroken symmetry phase would have the same transmission probability as right-handed baryons, leading to no net baryon number in our Universe. This is why Sakharov needs CP violation in addition to C violation.

[20] That is, it is in metastable equilibrium, unstable only to our broken–symmetry universe

by ignoring the sphaleron–induced reactions that change baryon number.

(d) *Write the diffusion equation for baryons, and shift to the moving reference frame y, setting $y = 0$ to be the position of the bubble wall. Derive the evolution equation for $\widehat{\rho}_B(y,t) = \rho_B(y + vt, t)$. Derive an ordinary differential equation for the stationary state, $\widehat{\rho}_B(y)$ in the moving reference frame by setting $\partial\widehat{\rho}_B(y,t)/\partial t = 0$.* Note: $\widehat{\rho}$ evaluated at $y = x - v(t + \mathrm{d}t)$ is

$$\begin{aligned}\widehat{\rho}(y, t+\mathrm{d}t) =& \rho(y + v(t+\mathrm{d}t), t+\mathrm{d}t) \\ =& \rho(y+vt, t) + v\left.\frac{\partial\rho}{\partial x}\right|_{y+vt}\mathrm{d}t \\ &+ \left.\frac{\partial\rho}{\partial t}\right|_{y+vt}\mathrm{d}t.\end{aligned} \tag{N1.9}$$

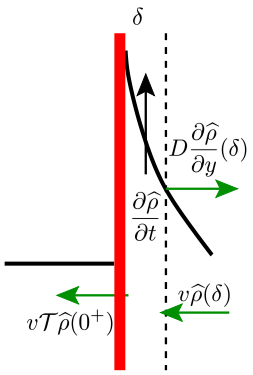


Fig. N1.10 Boundary conditions at electroweak wall. Near the wall, current conservation sets the boundary condition at the wall (eqn N1.9).

We also need to know the boundary condition at the wall $y = 0$. Consider a small region $0 < y < \delta$ on the unbroken symmetry side of the wall (Fig. N1.8). The net number of particles entering this region, divided by δ, gives $\partial\widehat{\rho}/\partial t$ just outside the wall. The current entering the region due to the moving reference frame is given by the velocity times $\widehat{\rho}$. The current transmitted through the wall is the velocity times $\mathcal{T}\widehat{\rho}$ (the current impinging on the wall due to the moving reference frame, times the transmission fraction). Finally, the current exiting the region due to diffusion is D times the slope.

(e) *Why must the net current be zero for a stationary state? In the limit $\delta \to 0$, show that*

$$\left.\frac{\mathrm{d}\widehat{\rho}_B}{\mathrm{d}y}\right|_{0^+} = -(1-\mathcal{T})(v/D)\widehat{\rho}_B(0^+). \tag{N1.10}$$

Finally, argue that this boundary condition must hold even if the density has not reached a stationary state (and hence must hold even for the original evolution of $\rho(x,t)$.)

(f) *Show, in the absence of sphalerons, that the steady-state density of baryons is*

$$\widehat{\rho}_B(y) = \rho_\infty(1 + (1/\mathcal{T} - 1)\mathrm{e}^{-(v/D)y}), \tag{N1.11}$$

and hence $\widehat{\rho}_{\overline{B}}(y) = \rho_\infty(1 + (1/\overline{\mathcal{T}} - 1)\mathrm{e}^{-(v/D)y})$, (This should allow you to check your answers from part (c).)

Now let us add the sphaleron interaction which changes the net baryon number. For simplicity, let us ignore the leptons, and assume a baryon-nonconserving reaction that has an equal rate Γ for changing baryons to antibaryons and for changing antibaryons to baryons.

(g) *Write the coupled differential equations in the moving reference frame for the steady-states* $\widehat{\rho}_B(y)$ *and* $\widehat{\rho}_{\overline{B}}(y)$. (For example, an extra term $\Gamma(\widehat{\rho}_{\overline{B}} - \widehat{\rho}_B)$ is needed for $\widehat{\rho}_B$.) *Add and subtract these equations to find uncoupled equations for* $\overline{\rho}(y) = (\widehat{\rho}_B(y) + \widehat{\rho}_{\overline{B}}(y))/2$ *and* $\delta\rho(y) = (\widehat{\rho}_B(y) - \widehat{\rho}_{\overline{B}}(y))/2$. *Ignoring the boundary conditions at zero, but applying the boundary conditions at* $y = \infty$, *solve for the general solution of these coupled equations (retaining two unknown amplitudes). Show that* $\delta\rho(y) \propto \exp(-\Lambda y)$ *decays exponentially, and determine* Λ.

If you have access and fluency in symbolic manipulation, you could now use the boundary conditions (eqn N1.9) at $y = 0$ for the baryons and the antibaryons to solve for the two unknown amplitudes in your equations. Using the densities at $y = 0^+$, you could then determine the transmitted densities. The algebra is a bit messy, so we are not asking you to do this. You would find in the end that the net baryon density in our Universe is predicted to be

$$\rho_B - \rho_{\overline{B}} = \frac{2(\mathcal{T} - \overline{\mathcal{T}})(\sqrt{v^2 + 8D\Gamma} - v)\rho_\infty}{(\sqrt{v^2 + 8D\Gamma} - v)(\mathcal{T} + \overline{\mathcal{T}}) + 4\mathcal{T}\overline{\mathcal{T}}v}. \tag{N1.12}$$

(h) *Find two choices of constants in the problem for which the baryon asymmetry is zero in our Universe.* (You may assume D stays positive.) *How do each of these relate to Sakharov's conditions? What does the net baryon density become when* $\overline{\mathcal{T}} = 0$ *(perfect exclusion of antibaryons)? Explain physically why this is to be expected.*

At the time of the electroweak transition, there were many more baryons and antibaryons than there are now. Almost all the antibaryons were annihilated by baryons into photons, leaving the small density of baryons we have now. Various estimates for the original ratio suggest that it was less than 10^{-9} away from half and half, perhaps as small as 10^{-11}.

Finally, one should note that the discovery of the Higgs boson produced a huge challenge for the electroweak baryogenesis. The sphaleron reaction rate Γ becomes very small when the Higgs mass becomes too high. As noted in part (g), the predicted net baryon density of our Universe vanishes as $\Gamma \to 0$ (when baryon number is conserved). Physicists used the current baryon number density and this feature of electroweak baryogenesis to predict a maximum value for the Higgs mass of 42 GeV. But the experimental mass turned out to be 125 GeV; the theory was not sufficient to explain the asymmetry. Where the antimatter went remains a mystery.

N1.9 **Chiral waves: Fourier and Green.** ③

Recent research shows that many systems have chiral *edge states* – propagating modes in topological insulators and the quantized Hall effect that flow in only one direction. Such a system might be described by a field $\Xi(x,t)$ which evolves according to

$$\partial\Xi/\partial t = A\,\partial\Xi/\partial x. \tag{N1.13}$$

(see Exercise 9.13).

Our methods for solving the diffusion equation are most useful for equations which are both linear and translation invariant.

(a) *Is the propagation law N1.12 linear? Is it translation invariant?*

This equation can be solved by inspection, for any initial condition, as a traveling wave. In this exercise, we shall reproduce the traveling wave solution using more sophisticated methods. First, what is the traveling wave solution?

(b) *Given an initial condition* $\Xi_0(x) = \Xi(x,0)$*, what will* $\Xi(x,t)$ *be?* (Hint: Use parts (c) and (d) to check your answer.)

We can solve eqn N1.12 with Fourier transforms. Let $\widehat{\Xi}_k(t)$ be the Fourier transform of $\Xi(x,t)$ with respect to x, using the conventions in the Appendix.

(c) *Derive an equation for* $\mathrm{d}\widehat{\Xi}_k/\mathrm{d}t$*. Solve it for* $\widehat{\Xi}_k(t)$ *in terms of* $\widehat{\Xi}_k(0)$*. Perform the inverse transform to evaluate* $\Xi(x,t)$ *in terms of* $\Xi_0(x) = \Xi(x,0)$*. Show your steps.*

We can use the Greens function for eqn N1.12 to solve it.

(d) *Use your general solution of part (b) to guess the Greens function* $G(x,t)$ *for eqn N1.12. Use the Greens function to evaluate* $\Xi(x,t)$ *for the initial condition* $\Xi(x,0) = \Xi_0(x)$*. Show your steps.*

N1.10 **Taste & smell with ensembles.** (Biology) ③

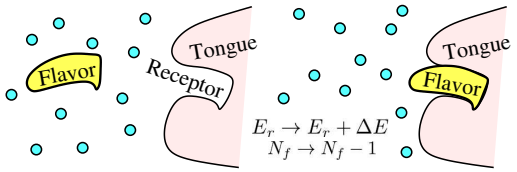


Fig. N1.11 Receptor binding in your nose or mouth leads to a signal that your brain interprets as taste or smell.

In Exercise 3.16, we modeled our sense of taste and smell in the microcanonical ensemble, and derived an expression for the signal involving the chemical potential. Using the grand canonical ensemble, we can now address this directly.

Figure N1.9 illustrates one of the ways our cells measure aspects of their environment – by placing receptors for given chemicals and measuring how often they are bound.

Model the receptor in the grand canonical ensemble, and assume it is in equilibrium with the fluid (air or saliva) surrounding it. Assume the fluid is at temperature T and the chemical potential for the flavor molecules in the fluid is μ. Let the energy of the receptor change by $\Delta E = E^B - E^U < 0$ as the receptor is bound, and let the entropy of the receptor change by $\Delta S = S^B - S^U$. Assume the volume in the system as a whole is unchanged when the flavor molecule binds, and for simplicity set $E^U = S^U = 0$.

Write the grand partition function Ξ as a sum over the two states.[21] *What is the probability that the receptor is bound (and sending a signal to the brain)? Does your answer agree with your microcanonical answer, eqn 3.81?*

N1.11 Entropy of Mastermind^{TM}.[22] ③

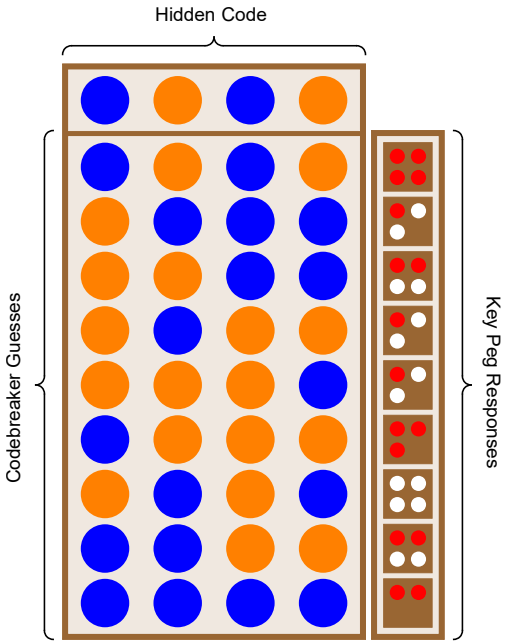


Fig. N1.12 Mastermind™ sample game, with two colors (B&O) and $N = 4$ pegs.

Mastermind is a game where a *maker* secretly picks a *hidden code* of N colored pegs, and a *breaker* learns information about the positions and colors of the pegs through a sequence of guesses. After each guess, the maker responds with the number of guessed pegs which match in position and color (0r, 1r, . . . , 4r), and the number of guessed pegs which share a color but are in the wrong position (0w, 1w, . . . , 4w), as in Fig. N1.10.[23]

[21] To be fussy, we are summing over the bound and unbound Helmholtz free energies, which each sum over e^{S/k_B} states.

[22] This exercise was developed in collaboration with Stephen Thornton

[23] You can view r and w as Right and Wrong, but they also match the red and white colors of the tiny pegs used in the response of the maker (Fig. N1.10).

In this exercise, we shall explore this game for the case of two colors, blue (B) and orange (O). Furthermore, instead of the breaker playing until the hidden code is matched, we shall play only one turn. The breaker's job is to maximize the information about the hidden code after their guess, averaged over many games. The maker will eventually be allowed to choose a strategy optimized to frustrate the breaker, but for now we will assume that the maker chooses each of the possible codes with equal probability.

(a) *If the breaker in a four-peg game guesses all blue and the response is 2r0w (Fig. N1.10), how many of the* $2^4 = 16$ *possible hidden codes are still possible? What is the corresponding probability* q_{20} *that the response will be 2r0w, over many games where the breaker guesses all blue? How many hidden codes are still possible if BBBB yielded a response of 1r0w? 0r0w? 3r0w? 4r0w? What is the probability* q_{rw} *of getting each of these responses?*

The breaker wins big in this case either if they guess correctly (4r0w) or if they guess completely wrong (0r0w) – they know precisely what the peg arrangement is. But how do we measure the information learned by guessing BBBB *on average*? Do we average the number of remaining peg arrangements over the possible responses? Do we do a weighted average of the number of remaining possible arrangements, based on the probability of getting each response?

In Section 5.3.2, we argued that the best way of measuring our ignorance is through entropy – the only measure that satisfies three key properties. (1) *Entropy is maximum for equal probabilities.* (2) *Entropy is unaffected by extra states of zero probability.* (3) *Entropy changes properly for conditional probabilities.*

(b) *Which of the three properties is relevant for calculating the entropy of the breaker before the game starts? Which is relevant for calculating the entropy after they guess BBBB? Which tells us that we can also include legal but missing responses, like 2r1w for the guess BBBB?*

(c) *What is the starting entropy of the breaker, in bits? What is the breaker's entropy* S_{20}^{BBBB} *after guessing BBBB and learning 2r0w? How many bits did the entropy decrease? What would their entropy* S_{rw}^{BBBB} *have been after each of the other possible responses* $rw = 00$*, 10, 30, and 40? What is the average entropy* S^{BBBB} *averaging over all the responses, if all the hidden codes are equally likely?* (Note: This last question is closely related to the discussion in Section 5.3.2 about your expected ignorance after asking your roommate where your keys were last seen.)

Now we must consider other possible initial guesses, to find the optimal choice. We can simplify our calculations by noticing some symmetries.

(d) *Given that the maker picks codes with equal likelihood, can the average entropy after guessing BBBB and OOOO be different? Between BBBO and BOBB? Between BBBB and OBBB?*

There is a permutation symmetry (reordering the pegs) and an inversion symmetry (B⇔O) in this problem; all initial guesses in a class related by these symmetries have the same entropy. This allows us to significantly reduce the number of breaker guesses

we need to calculate. To further reduce effort, we also reduce the number of pegs to $N = 3$.

The following table has two columns for breaker guesses and eight rows for the maker's possible hidden codes (all equally likely).

(e) *The other possibilities for breaker guesses should be related to the two shown. What is the multiplicity of BBB – the number of other breaker guesses in the same symmetry class? What is the multiplicity of BBO?*

	BBB	BBO
BBB		
BBO		
BOB		
OBB		
BOO		2r0w
OBO		
OOB		
OOO		

(f) *Copy the table, and fill it in with the correct responses. Use the table to calculate* q_{rw}^{BBB} *and* S_{rw}^{BBB} *for the responses* rw *in column two, and* q_{rw}^{BBO} *and* S_{rw}^{BBO} *for column three. Calculate the average entropy after a guess in the BBB class and after one in the BBO class, as you did for BBBB in part (c). Which will yield more information?* (Hint: Use Fig. N1.11 to roughly check your answer.)

Now let us allow the maker to develop an optimal strategy to frustrate the breaker. What can we say about the optimal probability distribution choice for the breaker in this game? We assume that both breaker and maker can figure out the other's optimal strategy. By playing a large number of games, any deviations from that optimal strategy can be detected and exploited.

Before the game begins, the maker's choices have the same permutation and inversion symmetries as we saw for the breaker's first guess.

(g) *If the maker insists on a strategy that to break the symmetry – say, choosing BOB more often than OBB, can the breaker exploit knowing this to increase his knowledge after his guess? If the breaker insists on a strategy that guesses BOB more often than OBB, can the maker exploit it? Must the optimal strategies of the two satisfy the inversion and permutation symmetries?*

We shall presume, whatever your answer for part (g), that the maker chooses BBB and OOO each with probability $p \leq \frac{1}{2}$, and all other codes with probability $1/6 - p/3$. Since the breaker know of their strategy and choice of p, the breaker's initial entropy changes.

(h) *Calculate the breaker's initial entropy as a function of p, before they make a guess. If the maker chooses p near* $\frac{1}{2}$ *and the breaker knows this, will the breaker's optimal strategy change from the one you calculated in part (f)?*

Since the maker's strategy maintains the same permutation and inversion symmetries as the equal-weighted strategy, we can use the same table to calculate the breaker's entropy as a function of p.

The table of codes, guesses, and responses now includes a column for the probabilities. The missing entries (unchanged from the earlier table) are filled in for one particular response $r = 2$, $w = 0$ for an initial guess of BBO.

		BBB	BBO
p	BBB		2r0w
$1/6 - p/3$	BBO		
$1/6 - p/3$	BOB		
$1/6 - p/3$	OBB		
$1/6 - p/3$	BOO		2r0w
$1/6 - p/3$	OBO		2r0w
$1/6 - p/3$	OOB		
p	OOO		

Calculating the resulting average entropy after the two guesses is now both laborious and similar to your calculation in part (f). The particular response 2r0w for BBO, however, involves three outcomes with differing probabilities, so is worth investigation.

(i) *What is the net probability for the response* q_{20}^{BBO} *as a function of* p*? What is the entropy* S_{20}^{BBO} *of the breaker after he gets that response to a BBO guess? In terms of these two quantities (without plugging in your formulas), what is the contribution of the 20 response to the average entropy of the breaker after the guess? Plot* S_{20}^{BBO} *versus* $p \in (0, \frac{1}{2})$.

Figure N1.11 shows the uncertainty of the breaker after the two possible classes of guesses. The entropy S_{20}^{BBO} you calculated in part (i) is a necessary step in calculating S^{BBO} in the plot below.

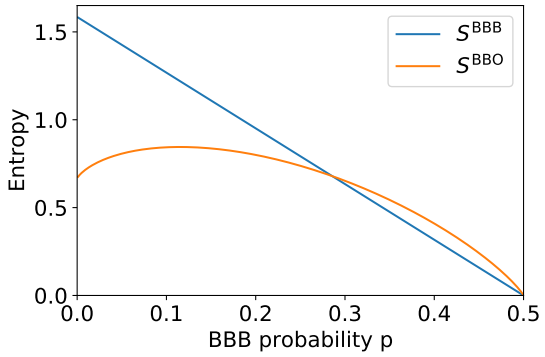


Fig. N1.13 Breaker uncertainty entropies, after one guess in Mastermind$^{\mathrm{TM}}$.

In non-cooperative games, each player tries to choose from an ensemble of strategies that best optimizes their outcome. The breaker's best strategy, in general, depends on the maker's strategy (here the choice of p), and vice versa. Mathematician John Nash[24] won the Nobel Prize in Economics for showing that such games will generally have a *Nash equilibrium*, when neither player has anything to gain by changing their strategy.

(j) *What is the optimal strategy for the breaker, for low p and high p? Use Fig. N1.11 to estimate approximately at what value of p the breaker's strategy changes. Assuming the breaker is using the optimal strategy, estimate approximately at what value of p the maker should use to frustrate the breaker as much as possible.* This combination of strategies is the Nash equilibrium for our game.

N1.12 Interpolation and free energies. ③

Alemi and Fischer [2] have unified a large number of deep neural network models into what they call TherML (Thermodynamics of Machine Learning). They combine four quantities R, D, C, and S that various algorithms attempt to minimize into a single optimization goal

$$R + \delta D + \gamma C + \sigma S, \tag{N1.14}$$

weighing the importance of minimizing R, D, C, S differently depending on the task. This sum is then minimized with respect to millions of parameters in the neural network, to achieve the user's goals. Here C measures the classification error, R the complexity of the representation, S the relative entropy in the parameters of the model, and D the distortion measuring the 'unsupervised' learning performance. This looks remarkably like our free energy formulas like the grand free energy $\Phi = E - TS + PV$ and the energy $E = TS - PV + \mu N$. Note that E weighs the relative 'goals' of S, V, and N by their costs T, $-P$, and μ.

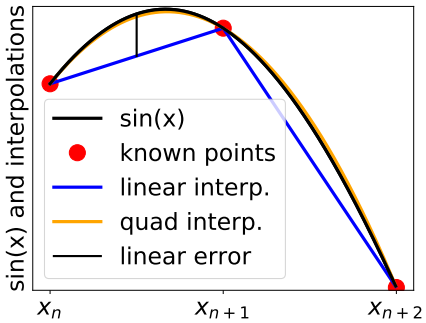


Fig. N1.14 Interpolations for small step sizes usually have their maximum error near midway between fitted points.

[24] This is thee same John Nash that was featured in "A Beautiful Mind".

Without pretending to explain these machine-learning terms, let us pursue a more tangible task with the same form: interpolating a function using polynomials (Fig. N1.12). Our goal is to minimize the error E in the interpolation by varying computer time T and storage N.

Consider approximating the periodic function $\sin(x)$ for x in $(0, 2\pi)$ by linearly interpolating between $N+1$ values at $x_n = n\epsilon$ with $\epsilon = 2\pi/N$ (Fig. N1.12). In each interval, the maximum error will approximately be in the center x between the known values at $x \pm \epsilon/2$, where it is $|\sin(x) - \frac{1}{2}(\sin(x - \epsilon/2) + \sin(x + \epsilon/2))|$.

(a) *Expanding about x to second order in ϵ, at what x is the error biggest? Show that the maximum error E goes as $E = A/N^2$ as $N \to \infty$. For $\sin(x)$, what is A? How many digits of accuracy $D(N) = -\log_{10}(E)$ will linear interpolation provide?*

We will consider polynomial interpolation methods, that fit the nearest M points separated by ϵ with a polynomial of degree $M-1$. For large N (small ϵ), these fits match the first M terms of the Taylor expansion about x, so their errors involve ϵ^M times the M^{th} Taylor coefficient (which depends on the function being interpolated). For simplicity, we shall ignore this dependence, so the error $E \propto N^{-M}$ and the number of digits $D = -\log_{10} E \propto M \log(N)$. (You may check that this agrees with your answer to part (a)).

The time needed to evaluate this polynomial scales as $T \propto M^2$ [27, Section 3.2]. For convenience, we shall assume time $M = \sqrt{T}$ and digits $D = M \log N = \sqrt{T} \log N$ (by choosing suitable units).

We are willing to pay an amount χ per digit of accuracy. Let σ be the cost of a unit of time, and μ be the cost of a unit of memory.

We can view the digits $D(N, T)$ as an analog of the microcanonical ensemble, where the amount of storage N and the computer time per evaluation T are fixed.

(b) *How does the number of digits $D(N, T)$ increase as we increase $N \to N + \mathrm{d}N$? Starting at (N, T), how much are we willing to pay to increase $N \to N + \mathrm{d}N$? How much would we pay for more computation time, $T \to T + \mathrm{d}T$? Evaluate $\mu(N, T)$ and $\sigma(N, T)$, this 'marginal cost' of additional storage and time. Give a general Maxwell relation giving $(\partial\mu(N, T)/\partial T)|_N$ as a derivative of σ, and check it using your formulas for μ and σ.*

So our general goal is to maximize

$$G(\chi, \sigma, \mu) = \chi D - \sigma T - \mu N \tag{N1.15}$$

with respect to T and N, in analogy with TherML's optimization in eqn N1.13. Here D is a function of N and T. G changes variables from D, N, and T just as the Gibbs free energy $G(T, P)$ changes variables from the microcanonical entropy $S(E, V)$. There $1/T$ was analogous to the cost in entropy of buying a unit of energy, and P/T analogous to the cost of a unit of volume. Here σ and μ are literally the cost in dollars for buying computer time and memory.

Minimizing G tells us that we have increased N and T until it costs χ per digit to make further improvements.

(c) *Suppose we start at some non-optimized N and T, and make a small change $(\Delta N, \Delta T)$ that produces a change ΔD in the digits of precision. How much are we willing to pay to benefit by the increased precision? How much does the change cost? Does the place where G is minimized make the total cost equal to the total benefit? Or something else?*

Let us imagine we have solved for G. What could we do with it? Just as derivatives of the Gibbs free energy $\mathrm{d}G = -S\mathrm{d}T + V\mathrm{d}P + \mu\mathrm{d}N$ allow us to write the entropy and the volume as functions of temperature and pressure, so our $G(\chi, \sigma, \mu)$ allows us to budget the computer time and storage in terms of their costs σ and μ and the value per digit χ. So, before plunging into the calculation of G...

(d) *If you were planning a simulation, how would you use G to determine the memory and computer time to budget, and the resulting number of digits you should purchase?*

Now we calculate G.

(e) *Extremize G in eqn N1.14 with respect to N to find a formula for μ. Extremize with respect to T to find a formula for σ.*

You will need a new special function to solve for G. Let $w(A)$ be the solution to the equation $\log(w) = Aw$.[25]

(f) *Solve both formulas from part (e) for $\sqrt{T}$. Equate the solutions to give a self-consistent formula for N. Using the Lambert W function to solve for N. Using one of your two formulas for T to write T as a power of N, and through that as a function of χ, μ, and σ.*

(g) *Let $\sigma = 3$ and $\mu = 0.1$. Plot N, T, and the number of digits for the range $\chi \in (3, 6)$.* (Hint: T should start near four, corresponding to the linear interpolation $M = 2$ in part (a).) *How much does the order M change as we add digits?*

In principle, we could now combine this information into eqn N1.14 to get an explicit (but horrible) formula for G.

N1.13 **Convexity and phase separation.** (Thermodynamics) ③

A piston, initially completely filled with water vapor, compresses the gas until it is completely liquid, moving between the marked positions in Fig. N1.13. The piston is held at temperature $T = 550\mathrm{K}$ at all times.

[25] This can be written in terms the -1 branch of the Lambert W function, $w(A) = -W(-A, k = -1)/A$. See the hints files [31] for implementations in Mathematica and Python.

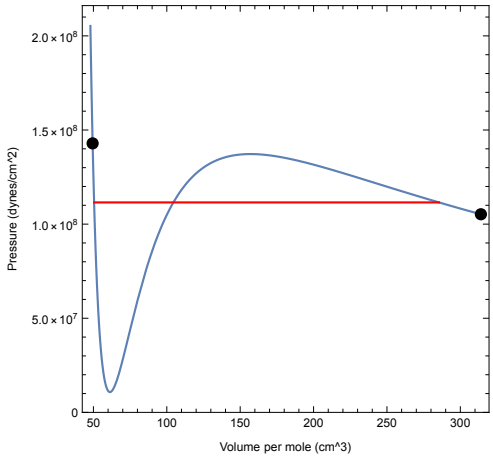


Fig. N1.15 Pressure vs. volume for the van der Waals model applied to one mole of H_2O at $T = 550$K. The red line shows the vapor pressure P_v at this temperature.

(a) *Draw the path on Fig. N1.13 taken if the piston moves slowly enough that the system remains in thermal equilibrium at all times.*

(b) *Sketch a path on Fig. N1.13 taken if the piston moves fast enough so that the pressure rises past the vapor pressure (say, to* 1.2×10^8 *dynes/cm*^{2}*) before a liquid water drop nucleates, but slowly enough so that the subsequent condensation of vapor into the water stays in equilibrium.*

Figure N1.14 shows the chemical potential for the van der Waals model as a function of density at this same temperature. Remember that the Gibbs free energy is μN, so this is also the Gibbs free energy per molecule. Note that the van der Waals solution assumes that the system is filled with molecules at a uniform density ρ, not a mixture of liquid and gas.

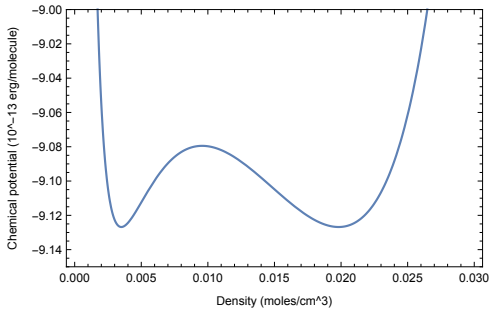


Fig. N1.16 Chemical potential μ vs. density ρ for the van der Waals model for H_2O at $T = 550$K.

(c) *Sketch on a copy of Fig. N1.14 the free energy one would obtain by allowing for the separation of the water into coexisting liquid and gas.* (Ignore the small contribution of

surface tension.)

If a system can be broken up into two weakly interacting subsystems, then the minimum free energy for the system in the limit of infinite size must be convex (see note 4 on page 324).

(d) *In your solution to part (c), what are the two weakly interacting subsystems? Why did we need to take the limit of infinite size to ignore surface tension? Is your answer convex?*

N1.14 Spinodals vs. Nucleation. ③

Here we explore the predicted edges of metastability for the liquid and gas, which are called *spinodals* in the older literature. Spinodals provide some insight into the behavior of materials near abrupt transitions, but should not be taken literally—fluctuations ignored by models like these cause the crossover from nucleation to 'spinodal decomposition' to become blurred.

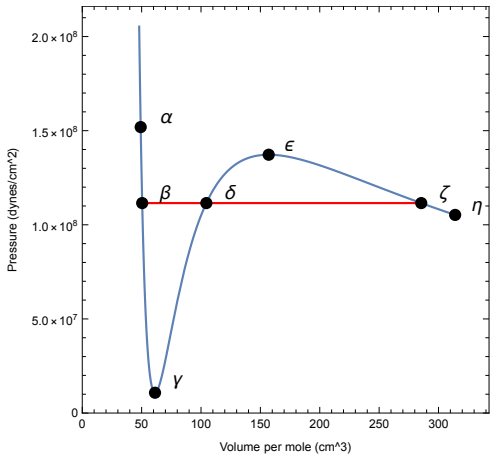


Fig. N1.17 Volume vs. pressure for the van der Waals model applied to one mole of H_2O at $T = 550\,\mathrm{K}$. The red line shows P_v at this temperature.

Let us explore how the Gibbs free energy per particle $G/N = \mu$ (Fig. N1.16) varies as we move between different points on the $P-V$ diagram Fig. N1.15.

(a) *At point α in Fig. N1.15 at the highest pressure, how many other solutions are there with that pressure? Which of the labeled points in the free energy plots of Fig. N1.16 corresponds to the state α?*

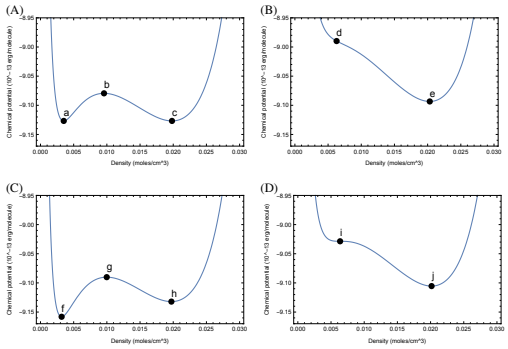


Fig. N1.18 Chemical potential vs. densityat fixed temperature $T = 550\,\mathrm{K}$, at some of the pressures corresponding to the points α–ζ in Fig. N1.15. The labeled points a-j here are solutions to the van der Waals model.

(b) *Is the system at point δ in Fig. N1.15 stable, unstable, or metastable? Which point a–j in Fig. N1.16 corresponds to δ? Explain what about the free energy tells you whether a state is stable, unstable, or metastable.*

We gradually compress a hot water vapor in a vertical piston, keeping the temperature at 550 K. We first study this in the P–V plane and in terms of the chemical potential.

(c) *At which labeled point in Fig. N1.15 does the metastable gas state go unstable?* This is called a spinodal point. *Which point a–j in Fig. N1.16 corresponds to this spinodal point for the gas phase?* In Exercise 11.3, we saw that the surface tension energy for nucleating a bubble is related to the barrier in free energy between the two phases. *At the spinodal point, does the free energy barrier between the two phases disappear?*

If one quickly changes the temperature or volume or other parameter across an abrupt phase transition line, the phase transition qualitatively happens in one of two ways. It can be nucleated by slow-forming bubbles or droplets. Or, if one moves into the unstable region, it can undergo *spinodal decomposition*, spontaneously separating into two phases without a nucleation barrier, with small random density fluctuations growing on many length scales. The boundary between these two is blurry except in mean-field theories like that of van der Waals.

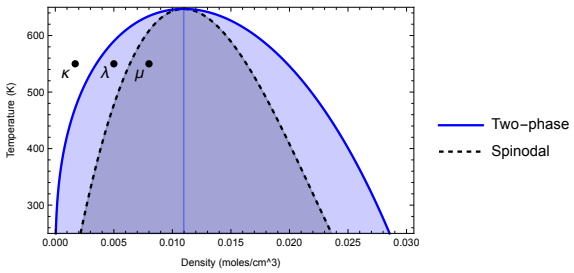


Fig. N1.19 Spinodal and two-phase regions for van der Waals water, as we vary density ρ and temperature T.

The three points κ, λ, and μ in Fig. N1.17 correspond to our compression experiment in part (c). We prepare the state at λ and μ by starting at state κ at $t = 0$ and compressing the piston quickly while cooling to keep the temperature T fixed. The states at λ and μ begin in a uniformly compressed state with only microscopic fluctuations. Assume the system is in very low gravity (so the liquid water, when it forms, does not quickly fall to the bottom). For each, consider the density $\rho_\lambda(h,t)$ and $\rho_\mu(h,t)$ along a one-dimensional line h rising from the bottom of the piston to the top.

(d) *Roughly sketch $\rho_\lambda(h,t)$ as a function of h for three times: $t = 0$ just after compression, a time t_{droplets} when the line passes through a nucleated droplet or two (which have yet to grow to use up the extra water vapor, or to fall to the bottom), and a much later time t_{equilib} when it is fully phase separated.* (Hint: What fraction of the volume will be liquid?) *Label your density axis with the equilibrium densities of the liquid and gas at 550 K, and with the initial density ρ_λ.*

(e) *Roughly sketch $\rho_\mu(h,t)$ as a function of h at a time t_{spinodal} when tiny random thermal density fluctuations have been noticably magnified by the instability, but have yet to approach the equilibrium densities of the two phases. Label your density axis with the equilibrium densities of the liquid and gas at 550 K, and with the initial density ρ_μ.*

After evolving for long times (ignoring gravity), a system which was launched quickly with spinodal decomposition or slowly by nucleation will evolve into rather similar states which coarsen with time.

N1.15 Cell signaling and mutual information.[26] (Biology, Statistics) ③

To survive, living systems must accurately measure and and respond to their environment. For example, *E. coli* bacteria famously execute a process called chemotaxis. By sensing gradients in nutrients, they decide whether to to run (propel themselves forward) or tumble (change direction) using their propellor-like *flagella* (see Exercises 2.19 and 2.22). The environmental signal is transmitted from the receptors on the cell surface to the motors that control the flagella via a cascade of chemical reactions. These

[26] This exercise was developed in collaboration with David Hathcock.

cascades involve kinases—protein enzymes that are activated (phosphorylated) by reactions that are catalyzed by upstream enzymes (the cell membrane receptors or another kinase). The activated kinase has a finite lifetime, with dephosphorylation reactions catalyzed by so-called phosphatase enzymes. Fluctuations in active kinase populations carry the signal through the cascade. Multi-level cascades can amplify signals, enabling sensitive detection of the environment, but also introduce noise due to the stochastic nature of enzymatic reactions. In this problem we will investigate how cells can mitigate this noise, optimizing the information transmitted by the signaling circuit.

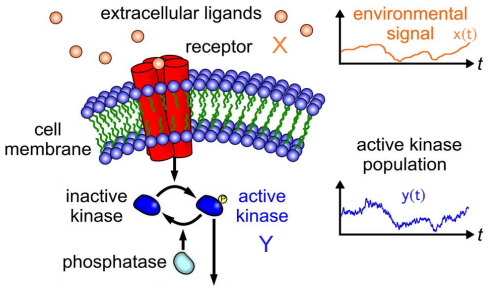


Fig. N1.20 The simplest cellular signaling circuit involves: (i) the environmental signal from extracellular ligands X, (ii) an kinase population Y within the cell, activated by the membrane receptors. This minimal model is a coarse-grained representation of actual chemical circuits that regulate chemotaxis in *E. coli*, olfactory (smell) sensing in mammals, and yeast response to osmotic pressure [30, 14].

Let $X(t)$ be the local concentration of food as the bacteria swims forward. This concentration randomly fluctuates, driven by noise ξ_x due to both to the bacteria's movement and coupling to the environment, but changes relatively slowly, with a long correlation time γ_x^{-1}. The dynamics of X are given by a Langevin equation, $\dot{X} = A_x - \gamma_x X + \xi_x(t)$, where A_x keeps the mean food level $\overline{X} = A_x/\gamma_x$ positive. The cell's measurement of this signal is noisy: small numbers of food molecules randomly bump into the receptors, which in turn causes activation of R kinase within the cell. The dynamics of active kinase population Y are given by, $\dot{Y} = A_y + R(X - \overline{X}) - \gamma_y Y + \xi_y(t)$, where ξ_y is the measurement noise, A_y is the background activation rate, and γ_y is the deactivation rate, determined by phosphatase concentrations. As with X, γ_y^{-1} sets the time-scale for fluctuations in Y. How can the cell tune γ_y to filter out the measurement noise at the receptors and gain the most information about the food in its environment?

For our calculations, we will assume the terms $\xi_a(t)$ with $a = x, y$ are white noise with correlation[27] $\langle \xi_a(t)\xi_b(t')\rangle = 2F_a\delta_{ab}\delta(t-t')$. Information is carried by the fluctuations in X and Y about the mean values, $\overline{X} = A_x/\gamma_x$ and $\overline{Y} = A_y/\gamma_y$; the bacteria wants to swim toward locations where there is more food than average. Subtracting off the

[27] If the number of molecules attaching to the receptor is the cause of the noise (this is called *shot noise*), then $F_x = A_x$ and $F_y = A_y$.

means, $x = X - \overline{X}$ and $y = Y - \overline{Y}$, we are left with,

$$\dot{x} = -\gamma_x x + \xi_x(t) \quad \dot{y} = -\gamma_y y + Rx + \xi_y(t) \tag{N1.16}$$

The calculations in this exercise are somewhat involved, so we shall provide many intermediate results to allow you to bypass parts if you get stuck. *If you use a computer algebra system (Mathematica™, SymPy, ...) please provide your code. If you use paper and pencil, please show your steps and prune your dead ends.*

(a) *Fourier transform eqn (N1.15) and solve for $\widetilde{x}(\omega)$ and $\widetilde{y}(\omega)$. Express your answers in terms of the Fourier transformed noise $\widetilde{\xi}_x(\omega)$, $\widetilde{\xi}_y(\omega)$ as well as γ_x, γ_y, and R.*

If $C_{ab}(\tau) = \langle a(t+\tau)b(t)\rangle$ is the correlation between a and b, then $\widetilde{C}_{ab}(\omega) = \widetilde{a}(\omega)\widetilde{b}^*(\omega)$ (see Appendix A). For example, the noise correlation relation becomes $\widetilde{\xi}_a(\omega)\widetilde{\xi}_b^*(\omega) = 2F_a\delta_{ab}$ and the equal time correlation is $C_{ab}(0) = 1/(2\pi)\int d\omega\, \widetilde{a}(\omega)\widetilde{b}^*(\omega)$. For the next parts, the following integrals are useful:

$$\begin{aligned}
\frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{e^{-i\omega t}d\omega}{\omega^2 + c^2} &= \frac{e^{-c|t|}}{2c} \\
\frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{d\omega}{(\omega^2 + c^2)(i\omega + d)} &= \frac{1}{2c(c+d)} \\
\frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{d\omega}{(\omega^2 + c^2)(\omega^2 + d^2)} &= \frac{1}{2cd(c+d)}
\end{aligned}$$

(b) *Show that $x(t)$ has exponential time correlations, $C_{xx}(\tau) = F_x/\gamma_x \exp(-\gamma_x|\tau|)$.* As we mentioned above the signal has a correlation time scale set by γ_x^{-1}.

(c) *Show that the equal time correlations are $C_{xx}(0) = F_x/\gamma_x$, $C_{xy}(0) = F_x R/(\gamma_x^2 + \gamma_x\gamma_y)$ and $C_{yy}(0) = F_x R^2/(\gamma_x^2\gamma_y + \gamma_x\gamma_y^2) + F_y/\gamma_y$.* To simplify the notation below, we will define $\sigma_x^2 = C_{xx}(0)$, $\sigma_y^2 = C_{yy}(0)$, and $\mathrm{cov}_{xy} = C_{xy}(0)$.

So far, we have understood the fluctuations and correlations in two chemical populations: food molecules in the environment and signaling proteins within the cell. How can we use these to quantify the information gained by the bacteria? Fortunately, Shannon explored this question in his seminal work on information theory. He developed the machinery of information entropy when working for the telephone company.[28] The phones in those days had lots of static. How could one calculate when the static noise made it impossible to hear the words? Shannon introduced the idea of a *mutual information* between the original signal and the noisy final signal.

Suppose the probability distribution of an original message x is $\rho_X(x)$, and the probability distribution of x being turned into a noisy final message y is $\rho_T(y|x)$ (the probability of y given x). This transmitted probabilty distribution can be written in terms of the joint probability distribution $\rho_{X,Y}(x,y)$, the likelihood of a random transmission having

[28] This was in the olden days, when there was only one phone company in the US, Bell Telephone.

original message x and final noisy message y. Note that $\rho_X(x) = \int dy\, \rho_{X,Y}(x,y)$ and similarly for $\rho_Y(y)$.

(d) *Express in words the true statement* $\rho_T(y|x)\rho_X(x) = \rho_{X,Y}(x,y)$ *in a way that would convince a non-scientist.*

The mutual information[29] is defined to be

$$\mathcal{I}(X,Y) = k_S \int dx\, dy\, \rho_{X,Y}(x,y) \log\left(\frac{\rho_{X,Y}(x,y)}{\rho_X(x)\rho_Y(y)}\right), \tag{N1.17}$$

The mutual information is symmetric in X and Y. How can it be used to study how much the noisy final signal y tells us about the original signal x?

(e) *Show that* $\mathcal{I}(X,Y) = S(X) - S(X|Y)$, *the entropy of the probability distribution of general noisy input signals* $\rho_X(x)$ *minus the entropy of the noisy signals resulting in a particular received signal* y, $\rho_T(x|y)$, *averaged over* y (Hint: the identity from part (d) also holds if we swap x and y. Plug this in and separate the log into a difference of two terms). *Explain in words why the first measures your ignorance of the message* X *before getting a telephone call* Y, *and the latter is your average ignorance after a call.*

To compute the mutual information for the cell signaling system, we require the joint distribution $\rho_{X,Y}(x,y)$. In general, for systems driven by white-noise (with delta-function correlations), this is done by converting the Langevin equations into a Fokker-Planck equation (see Exercise 8.22), which can then be solved to obtain the equilibrium distribution. When the equations are linear (as is the case for our model, eqn (N1.15)), one finds that the joint distribution is a bivariate Gaussian[30],

$$\rho_{X,Y}(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-r^2}} \exp\left(-\frac{1}{2\,(1-r^2)}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} - \frac{2rxy}{\sigma_x\sigma_y}\right)\right). \tag{N1.18}$$

Here, the standard deviations $\sigma_x = \langle x^2\rangle^{1/2}$ and $\sigma_y = \langle y^2\rangle^{1/2}$ and correlation $r = \langle xy\rangle/(\sigma_x\sigma_y) = \mathrm{cov}_{xy}/(\sigma_x\sigma_y)$ are precisely those you computed in part (c).

(f) *Argue that the marginal distributions* $\rho_X(x)$ *and* $\rho_Y(y)$ *are also Gaussian, with widths* σ_x *and* σ_y *respectively. Show that the mutual information for the bivariate Gaussian is* $-(k_S/2)\log(1-r^2)$. *Interpret this result for* $r \to 0$ *and* $r \to 1$. *Is the mutual information dependent on the magnitude of the fluctuations?* (Hint: what happens if we scale $x \to \alpha x$).

[29] The mutual information is also the distance in probability space between the joint distribution $\rho_{X,Y}(x,y)$ and the uncorrelated case $\rho_X(x)\rho_Y(y)$, using a distance measure called the Kullback–Liebler divergence (see Exercise N5.18).

[30] This is a generalization of a familiar result from statistical mechanics: the Boltzmann distribution for an over-damped harmonic oscillator (which feels forces linear in the displacement) is Gaussian.

Biologically, the rate γ_y is easiest to tune (e.g. through evolution and natural selection). This rate is controlled by concentrations of phosphatase in the cell, whereas the activation rate R requires energy dissipation.

(g) *Argue that maximizing mutual information is equivalent to maximizing* r^2. *Show that* $r^2 = Z\Lambda/((1+Z)(1+Z+\Lambda))$, *where* $Z = \gamma_y/\gamma_x$ *and* $\Lambda = F_x R^2/(F_y\gamma_x^2)$. *Maximize* r^2 *with respect to* γ_y. *Show that when the information transmission is optimized,* $\gamma_y = \gamma_x\sqrt{1+\Lambda}$ *and* $\mathcal{I} = \frac{1}{2}[k_S \log(1+\sqrt{1+\Lambda}) - 1]$.

With all else fixed, there is an ideal response time: γ_y^{-1} must be faster than γ_x^{-1} by the factor $1/\sqrt{1+\Lambda}$. If γ_y is too small, Y will integrate over the changes in X, filtering out too much of the the signal. On the other hand, if γ_y is too large, Y will fail to filter out enough of the noise introduced at the cell receptors. The mutual information quantifies how much information the cell has about the current state of its environment, obtained by measuring the history of the signal and filtering out receptor noise. The dimensionless parameter Λ serves as a measure of the signal fidelity. When $F_x \gg F_y$, then $\Lambda \gg 1$: it is easy to reproduce the signal if the magnitude of fluctuations in X are large compared to the noise introduced by the receptors. On the other hand if $F_x \approx F_y$, $\Lambda = R^2/\gamma_x^2$, which measures the sensitivity of Y to changes in the X population over the timescale γ_x^{-1}. Real biological signaling circuits, like those in yeast and *E. coli*, tend to lie in the range $\Lambda = 100 - 1000$.

(h) *For the range of* Λ *listed above, how much information, in bits, can the cell learn about its environment from monitoring* $Y(t)$*?*

N1.16 **Emittance and particle beams.**[31] (Accelerator) ③

Particle accelerators take bunches of protons and antiprotons up to near the speed of light, and smash them head-on to see what happens. Electron microscopes take bunches of electrons and focus them to image materials at the atomic scale. X-ray sources accelerate electrons to near the speed of light, and use undulators to wiggle them to create X-rays (synchrotron X-ray sources) or free electron lasers (coherent beams of X-rays). In all of these applications, in addition to the energy per ion and the number of ions, the key property of a good bunch is its *emittance*. In this exercise, we shall explain why emittance is important, relate it to the entropy of the bunch, analyze its quantum limit, and explore the use of electron bunches to cool bunches of protons.

The 3D emittance of a bunch is loosely given[32] by the product of the volume it occupies in position space and in momentum space:

$$\varepsilon = (\Delta q_x \Delta p_x)(\Delta q_y \Delta p_y)(\Delta q_z \Delta p_z). \tag{N1.19}$$

[31] This exercise was developed in collaboration with Michael Kaemingk. We have used real numbers as input for this exam, but our assumptions for the calculations are not reliable. For example, real bunches are not as tidy as our Gaussian model bunches, and their momentum spread is not thermal.

[32] In part (a) you shall derive another more rigorous but less practical definition of the emittance in terms of the entropy per particle.

For our purposes, Δq and Δp will represent the standard deviations of the positions and momenta in the center of mass frame of the bunch. Here z is the 'longitudinal' direction in which the beam is moving, and x and y are 'transverse'. For synchrotrons, y is the 'vertical' direction perpendicular to the plane of the circular orbit and x the 'horizontal' direction in the plane of the circle but perpendicular to the motion. One also speaks of the 2D emittance of the transverse directions $(\Delta q_x \Delta p_x)(\Delta q_y \Delta p_y)$ perpendicular to the velocity of the bunch, or the 1D emittance along one of the axes. We shall see that many beams are strongly anisotropic, with different widths and momentum spreads along these three axes.

Emittance is a limiting parameter in the performance of any accelerator. In electron microscope/diffraction accelerators, the emittance limits the resolution; in colliders, it limits the luminosity; in free electron lasers, it limits the gain length and the minimum wavelength of the laser. Low beam emittance is therefore desirable, and reducing it is a central research goal of the Center for Bright Beams (CBB), a collaboration in which Cornell plays a leading role.

At the Large Hadron Collider (LHC), the bunches have $N = 1.2 \times 10^{11}$ protons, a bunch radius of $\sigma_q = 3.5\,\mu$m, and a transverse emittance of $3.75\,\mu\text{rad} = 1.88 \times 10^{-24}\,\text{kg}\,\text{m}^2/\text{s}$, implying a RMS bunch momentum of $\sigma_p = 5.4 \times 10^{-19}\,\text{kg}\,\text{m/s}$.

The ions in a bunch are often nearly noninteracting and uncorrelated, with all ions having nearly the same probability distribution in phase space. In this case, $\rho_N(\mathbb{P}, \mathbb{Q}) = \prod_{n=1}^{N} \rho(\mathbf{p}_n, \mathbf{q}_n)$. In this exercise, our bunches will mostly have Gaussian distributions. Here we very roughly approximate the LHC bunch as an isotropic, uncorrelated product of spherically symmetric Gaussian distributions

$$\rho_{\text{LHC}}(\mathbf{p}, \mathbf{q}) = \frac{\mathrm{e}^{-\mathbf{q}^2/2\sigma_q^2} \mathrm{e}^{-\mathbf{p}^2/2\sigma_p^2}}{(2\pi\sigma_q^2)^{3/2}(2\pi\sigma_p^2)^{3/2}}. \tag{N1.20}$$

(a) *Write the formula for the 3D emittance of the bunch in eqn N1.20.* Note that the momenta have a thermal distribution. *Write a formula for the temperature of the bunch. Write the entropy*

$$\begin{aligned} S = &- k_B \langle \log \rho \rangle = -k_B \int \rho \log \rho \\ = &- k_B \int \mathrm{d}\mathbb{P}\, \mathrm{d}\mathbb{Q}\, \rho(\mathbb{P}, \mathbb{Q}) \log \rho(\mathbb{P}, \mathbb{Q}) \\ &- 3N k_B \log h. \end{aligned} \tag{N1.21}$$

(Warning: There is an error in the corresponding eqn 5.20 in the print edition of the text.) *Evaluate the temperature and the entropy of the bunch using the LHC parameters above. Write a formula for the emittance in terms of* $\exp(S/Nk_B)$. We may view your last result as a more rigorous definition of the emittance.

Thus, the emittance, like the entropy, can only grow as the bunch passes through accelerating fields and focusing magnets.[33] Making a low emittance bunch, and keeping the emittance low during its acceleration and focusing, is key to all accelerator applications.

Why do we care about the momentum spread Δp in eqn N1.20, if we want a dense beam or a sharp focus? The conservation of emittance forces a tradeoff between a narrow beam and one that stays narrow as it moves.

The angular dispersion of a beam is due to the momentum spread σ_p: the momentum in the transverse direction will make the beam grow in width. The angular spread in a beam is given by the ratio of the transverse momentum spread Δp over the mean momentum of the ions in the forward direction. The latter, mv for nonrelativistic motion (with m the particle mass) becomes $mv/\sqrt{1-v^2/c^2} = \beta\gamma mc$ for a relativistic beam, with $\beta = v/c$ and $\gamma = 1/\sqrt{1-v^2/c^2}$. Thus the angular spread $\Delta\theta = \Delta p/\beta\gamma mc$ radians with respect to the direction of motion.

(b) *An electron microscope has a beam with velocity $v = 0.62c$, a width $\sigma_q^i = 200\,\mu$m and a vertical emittance $\varepsilon = 2.7\times10^{-30}\,\mathrm{kg\,m^2/s}$. What is its vertical momentum spread σ_p? What is its angular spread $\Delta\theta_i$ in radians?* A lens system focuses the beam into a smaller width $40\,\mu$m, without increasing the entropy or changing its velocity. *What is the new angular spread $\Delta\theta_f$, in terms of $\Delta\theta_i$? How far can it propagate before the new spread gives the beam a width larger than the original width σ_q^i?* (Rough estimates are fine.)

One might be surprised that the ions in a bunch can be treated as non-interacting, given the strong Coulomb interactions between particles. Indeed, these 'space-charge' effects are important when the bunch is first formed. But as it reaches near the speed of light, these interactions become much less important. Using an ultrafast electron diffraction apparatus at Cornell as an example, the current state-of-the-art photoemission bunch starts out as a bunch of radius $20\,\mu$m in the transverse directions, and a duration of 200 femtoseconds. Suppose this beam is now placed in an X-ray Free Electron Laser and accelerated to a speed very close to that of light (but leaving its duration fixed). The packet in the laboratory frame is $60\,\mu$m long (roughly spherical). But the packet in the laboratory frame is Lorentz contracted by a factor of $\gamma \approx 34000$, so in the center of mass frame it is a long tube of length $\gamma\times 60\,\mu\mathrm{m} = 1.3\,\mathrm{m}$. The force on the electrons in the beam are mostly due to the charges within a distance along the tube roughly given by the distance of the electron from the center of the tube ($20\,\mu$m), which in turn is roughly $1/\gamma$ of the total. Furthermore, the motion due to these forces in the lab frame takes γ times longer to happen in the lab frame due to time dilation. The two effects combined imply that the Coulomb interactions are suppressed by $1/\gamma^2$, making them basically negligible.

Strong transverse bunch shape anisotropy is also seen in electron beams in synchrotrons. At the Cornell Electron Storage Ring (CESR), $\varepsilon_x \approx 500\,\mathrm{eV\,m/c} = 2.7\times10^{-25}\,\mathrm{kg\,m^2/s}$ along the horizontal (x) direction, and $\varepsilon_y \approx 0.1\,\mathrm{eV\,m/c} = 5.3\times10^{-29}\,\mathrm{kg\,m^2/s}$ along

[33] The beam can lower its emittance by emitting X-rays, which is important in synchrotrons and undulators.

the vertical direction:[34] a factor of 1000 anisotropy! Indeed, one of the ambitious goals in the accelerator community is to get the transverse emittance down to the *quantum limit*.

(c) *What would the vertical emittance $\Delta q_y \Delta p_y$ be at the quantum limit, set by the uncertainty principle? By what factor must CESR shrink their emittance to approach this goal?*

Liouville's theorem implies that reducing the entropy or the emittance cannot be done simply with the standard tools of accelerators (magnetic lenses, focusing solenoids, bunchers, etc.). Instead, the beam must be coupled to another system into which the entropy can be dumped. There are different methods for doing this, such as synchrotron cooling, electron cooling, and stochastic cooling (for which Simon van der Meer received the Nobel Prize in 1984). Here we will consider a simplified model of electron cooling.

Electron cooling is a mechanism by which an electron beam is used to reduce the momentum spread, and therefore the entropy, of a beam of heavier ions. It is being used in the Relativistic Heavy Ion Collider (RHIC), colliding gold ions to create a quark-gluon plasma. The electron and gold ion beams are overlapped with nearly the same velocity. Here again we assume the momenta of the two beams have a Gaussian distribution, characterized by a temperature.

(d) *Suppose the center-of-mass momenta in a gold ion bunch has initial temperature $T_{\rm Au}$, and is put into contact with a co-moving electron bunch of the same radius σ_q, the same number of particles N and a temperature $T_e \ll T_{\rm Au}$. What will the final temperature T_f of the two ion beams be, if they have enough time to exchange kinetic energies and thermalize?* (You may ignore the interactions between the various ions.) *Calculate the ratio of the one-dimensional emittances $\varepsilon_{\rm Au}/\varepsilon_e$ and the beam angular dispersion $\Delta\theta_{\rm Au}/\Delta\theta_e$ of the resulting gold ion and electron bunches.*

As the gold ion beam travels through the electron gas, it will experience two effects. One is a drag force proportional to the velocity with respect to the center of mass. The other is fluctuations due to individual scattering events with the electrons. In Exercise 6.18, you found a fluctuation-response relationship between the noise and the drag coefficient in simulations of Langevin dynamics for one-dimensional motion. The relationship you found in that exercise was

$$k_B T = \langle \Delta p^2 \rangle / 2\eta \Delta t. \tag{N1.22}$$

We express this relation in terms of η, the inverse of the mobility, so the force on a moving gold ion is $F = -\eta v$. (The mobility is called γ in Exercise 6.18.) For us, Δp is a momentum change due to a collision, T is the temperature of the electron beam, and Δt is the time between collisions.

So, to find the drag force for a moving gold ion, we shall calculate the noise on a stationary ion and use this relationship.

[34] The horizontal width is large because of the randomness introduced by the emission of X-ray photons as the beam circles around.

The geometry of collisions between electrons and ions via the Coulomb interaction is complicated. One simplifying concept is that of a scattering cross section—the effective circular disk that would suffer the same forces. For the densities of electrons and gold ions in our two bunches, you may assume that a gold ion near the center of the electron bunch has a cross section $\Sigma = 0.1\,\mathrm{nm}^2 = 10^{-19}\,\mathrm{m}^2$. The momentum exchange in a collision with that disk will solely depend on the momentum p_x of the electron, where x is the axis perpendicular to that disk.

How big a momentum change will a gold ion have during a collision with an electron? (How much momentum is exchanged when a ping-pong ball hits a car?)

(e) *Taking the limit $m_{\mathrm{Au}}/m_e \to \infty$, what is the net change Δp for a stationary gold disk, given a head-on elastic collision with an electron with x momentum p_x?* Assume that the particles are non-relativistic in the center-of-mass frame.

Let us now calculate the other terms in eqn N1.23.

(f) *Give a formula for the average time Δt between collisions from the cross section Σ of the stationary ion disk, the electron number density n, and the momentum distribution of the electrons. Give a formula for the average squared momentum transfer $\langle \Delta p^2 \rangle$ for these collisions. Give the formula for the temperature $k_B T$ of the electrons.* (Rough estimates are fine. You may assume the electrons have a Gaussian momentum distribution with width σ_p.)

At the Relativistic Heavy Ion Collider (RHIC) work is underway to cool gold ions with an electron beam. In one scheme, a gold ion beam circles a ring with circumference $C = 3.9\mathrm{km}$, and each turn passes through a segment of length $0.0078C$ with a co-moving electron gas of density $10^8\,\mathrm{cm}^{-3} = 10^{14}/\mathrm{m}^3$ and momentum spread $1.2 \times 10^{-23}\,\mathrm{kg\,m/s}$. The electrons are refreshed continuously using a Cornell-invented energy recovery Linac (ERL). The bunches are moving fast: their value of $\gamma = 1/\sqrt{1 - v^2/c^2} = 105$. Time dilation makes the interaction time needed larger by a factor of γ.

(g) *Using your answers from part (f) and eqn N1.23 from Exercise 6.18, find a formula for the inverse mobility η. Using $F = -\eta v = \dot{p}$, find a formula for the exponential decay time τ for the gold ion velocity in the center of mass frame. Evaluate it for RHIC's electron cooling scenario.*

(h) *What is the exponential decay time for the gold ion beam in the lab frame of RHIC?*

N1.17 **Nonabelian defects.**[35] (Mathematics) ③

In this problem, we will try to understand the defects that occur in a liquid crystal composed of a planar tri-headed molecule (Figure N1.19). The ordered states of this system have all molecules in the same orientation, but with liquid translational order (randomly placed centers). This exercise will explore subtle questions regarding classification of nonabelian defects. It will explore *braiding* with mathematics related to the problems of anyon statistics and topological quantum computing. And it will provide

[35] This exercise was developed in collaboration with Stephen Thornton.

a complete analysis of the transformation of one defect as it is pulled around another, explored in the case of fingerprints in Exercise 9.17.

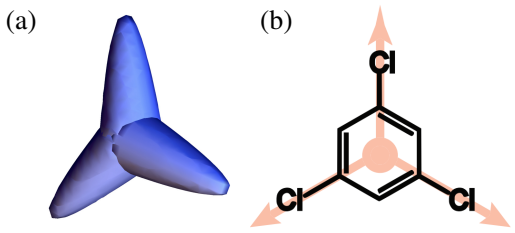


Fig. N1.21 Dihedral molecules. (a) The tri-headed molecule studied in this problem. (b) The molecular structure of 1, 3, 5-Trichlorobenzene, with a representation of a tri-headed molecule superimposed [4]. If these molecules were to form a liquid crystal, the liquid crystal would have the types of defects described in this exercise.

If the molecule 1, 3, 5-Trichlorobenzene (Fig. N1.19b), which has the same symmetry as our cartoon molecule, had a liquid crystalline phase[36] where they oriented parallel to one another, one would find the defects we study here [4]. These systems are generally called *dihedral* liquid crystals for reasons that we will explore shortly.[37]

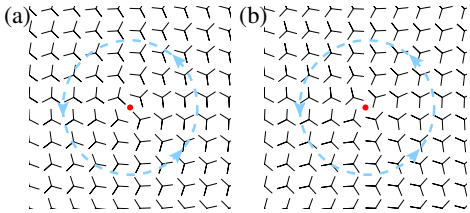


Fig. N1.22 Defects with molecules confined to the plane. Note that the molecules are at random positions; the orientations are shown on a square lattice for convenience.

Figure N1.20 shows two defects (red points) around which the molecules swirl. Recall that the winding number is defined as the fraction of a full rotation $\Delta\phi/2\pi$ an object does as we travel around a defect counterclockwise in real space. $\Delta\phi$ has a sign associated with it. Is the defect labeled by the winding number? In part (a) we find that the answer is yes if the molecular "legs" are confined to the plane, but otherwise no:

(a) *What are the winding numbers of the defects shown in Figure N1.20? If we allow the molecules to rotate in three dimensions, are these two defects topologically equivalent? Why or why not?* (Hint: Try applying a rotation about various axes to all the molecules.)

We saw in part (a) that the molecular orientation, after following the path around a

[36] There is currently no experimental evidence for this specific type of liquid crystal.

[37] Note: Most papers on dihedral liquid crystals confine the molecules to the plane. We are allowing them to rotate in 3 dimensions.

defect, must return to an orientation related by the symmetry group of the molecule. Our molecule has a dihedral symmetry group D_3 (hence the name dihedral liquid crystal). We imagine that the homotopy group (the nonequivalent possible circular paths in the order parameter space) will be related somehow to D_3.

D_3 is a six element group, generated[38] by two elements: a $2\pi/3$ counter-clockwise rotation a about an axis normal to the plane of the molecule (taken to face out of the page for the planar configurations), and a flip x, rotating by π about a particular leg of the molecule (the left-hand portion of the *group presentation* N1.24 below). The multiplication table (Table N1.1) can be derived from the three equality relations on the right side of presentation N1.24:

$$D_3 = \left\langle x, a \;\middle|\; x^2 = 1, a^3 = 1, xax^{-1} = a^{-1} \right\rangle . \tag{N1.23}$$

The group element gh is found in the row labeled by g and the column labeled by h. This is important because D_3 is nonabelian.

$\mathbf{D_3}$	**1**	**a**	$\mathbf{a^2}$	**x**	**xa**	$\mathbf{xa^2}$
1	1	a	a^2	x	xa	xa^2
a	a	a^2	1	xa^2	x	xa
$\mathbf{a^2}$	a^2	1	a	xa	xa^2	x
x	x	xa	xa^2	1	a	a^2
xa	xa	xa^2	x	a^2	1	a
$\mathbf{xa^2}$	xa^2	x	xa	a	a^2	1

Table N1.1: Multiplication table for the dihedral group D_3. Note that the group is nonabelian: gh is not equal to hg. The element gh would be found in the row labeled by g and the column labeled by h.

(b) *What are the symmetry group elements given by the paths in Fig. N1.20 (a) and (b)? Explain in words why $xax^{-1} = a^{-1}$. Use the multiplication rules to calculate $xa^2 \otimes xa$ and $xa \otimes xa^2$. Compare your answers to those in the table.*

We know from part (a) that two different dihedral symmetry group elements can be topologically equivalent—so the defect strength is not just given by the element of D_3. We shall explore this further below. But first, let us see if the same dihedral symmetry group element can correspond to two non-equivalent topological defects.

[38] A group is *generated* by a subset of elements (called *generators*) if every member of the group can be represented by repeatedly multiplying generators and their inverses.

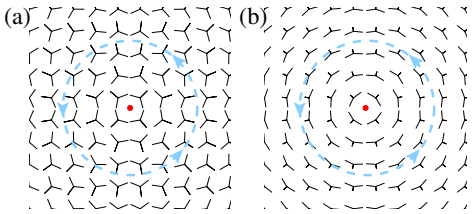


Fig. N1.23 Two equivalent defects, one with the molecules confined to the plane, and the other with one leg rotated to point 'upward'.

(c) *What is the winding number of the defect shown in Fig. N1.21(a)? If D_3 were the homotopy group, and we were to assign this defect an element of D_3, which element would it be? What does this suggest about the defect?* Can we heal this defect by rotating into the third dimension? *Does the rotation into the third dimension shown in Fig. N1.21(b) heal the defect in the order parameter field?*

We need a homotopy group that somehow has a category for a defect that returns the molecule to its original orientation trivially (the identity element in D_3), and somehow also allows a defect to return the molecule differently (nontrivially) to the same labelling of the legs (the element a^3). There is a wonderful treatment of the topological theory of defects by Mermin [21], which we draw upon here.

How can a 2π rotation in three dimensions be different from the identity? For those experienced in quantum physics, rotating a spin $\frac{1}{2}$ electron by 2π changes the sign of the wavefunction. This is because the spin wavefunction is represented not by a vector, but by a 2×2 unitary matrix in SU(2), and there are two SU(2) matrices for every rotation in SO(3). Similarly, in our problem for every rotation g in D_3 there are two rotations $\pm g$ in Dic_3: the 12-element *dicyclic* group, which is the first homotopy group for our dihedral liquid crystal.[39] The multiplication rules for Dic_3 are

$$\mathrm{Dic}_3 = \left\langle x, a \;\middle|\; x^2 = -1, a^3 = -1, xax^{-1} = a^{-1} \right\rangle \tag{N1.24}$$

Here we presume that a corresponds to a 1/3 counter-clockwise turn. (We will not write out the whole multiplication table.)

(d) *View the defect in Fig. N1.21(a) as three* 1/3 *counterclockwise turns. What is the homotopy group element in Dic_3 for the defect? How did lifting to Dic_3 fix the problem we had in part (c)?*

The rotation of the molecular orientation by 2π gives a net minus sign, just as for an electron spin.

[39] Mermin tells us [21] that the order parameter space of our dihedral liquid crystal is $\mathrm{SO}(3)/D_3$ (the continuous broken symmetry modulo the residual symmetry group). The first homotopy group of a simply–connected group modulo a discrete subgroup is the discrete subgroup—but SO(3) is not simply connected. He prescribes using the simply-connected universal cover SU(2), and the corresponding discrete group Dic_3. Thus $\Pi_1\left(SO(3)/D_3\right) \cong \Pi_1\left(SU(2)/\mathrm{Dic}_3\right) \cong \mathrm{Dic}_3$.

Now we still need to understand how the two defects in Fig. N1.20, with different group elements (in both D_3 and Dic_3), can be the same defect. Mermin tells us that defects with non-abelian homotopy groups are not classified by their homotopy group elements, but by the *conjugacy classes* of the group. The conjugacy classes are subsets of the group whose elements are related by conjugation. Two group elements g and h are said to be conjugate if there exists a group element γ such that $h = \gamma g \gamma^{-1}$.

(e) *Show in complete generality that the identity element $g = 1$ of a homotopy group is the only element in its conjugacy class. Interpret this physically: can there be two different ways of having no defect?* One can check that -1 commutes with all other group elements, implying that it too is the only element in its class.

Do the two defects in Fig. N1.20 indeed lie in the same conjugacy class?

(f) *What is the homotopy group element $\beta \in Dic_3$ for the defect in Fig. N1.20(a)?* The two defects in that figure combine to form no defect, so the defect in Fig. N1.20(b) must have homotopy group element $\widehat{\beta} = \beta^{-1}$. *Find a group element γ that shows $\widehat{\beta} = \gamma\beta\gamma^{-1}$.*

Is there a physical reason for this peculiar conjugacy class[40] criterion? Indeed: it is precisely the transformation of the defect β when it *braids* around another defect γ.

Braiding[41] in three dimensions is associated with hair-styling: you take two or more strings of hair or string and weave them over and under one another to form a kind of rope. If we move a point defect with strength β in two dimensions around another defect α (with the necessary continuous readjustments of the order parameter fields), we can view the process as braiding in space-time.

[40] The conjugacy classes of Dic_3 are as follows:

$$C_0 = \{1\}\,,\ \ \overline{C_0} = \{-1\}\,,\ \ C_1 = \{a, -a^2\}\,,\ \ C_2 = \{a^2, -a\}\,,\\ C_{\rm even} = \{x, xa^2, -xa\}\,,\ \ C_{\rm odd} = \{xa, -x, -xa^2\}\,. \tag{N1.25}$$

[41] Braiding is important for the study of anyons. Nonabelian braiding is important for topological quantum computing.

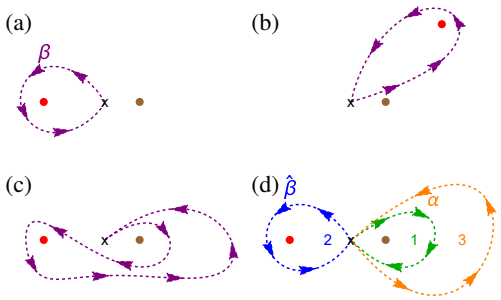


Fig. N1.24 Braiding two defects. (a-d) can represent the motion of a red two-dimensional point defect around a brown defect in time, or a red line defect curling around a brown line defect in three dimensions. (a) The red defect has strength β, measured by the path beginning and ending at **X**. (The connection to **X** makes this a *based* homotopy class, which is important for the argument but not crucial here.) (b-c) The defect is continuously dragged around the brown defect of strength α. The measured strength cannot change, either as the red defect continuously is dragged, or as the path anchored at **X** is continuously modified to surround β without touching α. (d) When the red defect returns to its original position, the original path 2 may measure a different strength $\widehat{\beta}$ than β measured by the deformed path $(1 \to 2 \to 3)$.

Figure N1.22 visually argues that a defect β encircling a defect α indeed changes its homotopy group element.[42] It argues that the original path β deforms into another path (in the same homotopy class) as the defect circumnavigates α. This path becomes the product of three simpler loops.

(g) *In Fig. N1.22(d), what homotopy group elements g_1, g_2, and g_3 correspond to the three segments of the path?* (The labels $\widehat{\beta}$ and α suggest the answers for two of the three.) *What is β, in terms of g_1, g_2, and g_3?*[43] *What is $\widehat{\beta}$ in terms of β and α?*

The schematic of braiding in Figure N1.22 is translated into a real configuration of our order parameter with two defects in Figure N1.23.

[42] We found in Fingerprints (Exercise 9.17) that moving a dislocation around a disclination can change the sign of its Burger's vector. This too is due to a non-abelian homotopy group, discussed by Poenaru et al. [26] in one of the founding papers in this field. They find that the combined order parameter space of rotations and translations relevant for fingerprints is the Klein bottle. Its first homotopy group is nonabelian, which provides an explanation for the dislocation annihilation in Fig. 9.36.

[43] It turns out that there is ambiguity in whether a concatenation of two loops (traveling along 1 and then 2) with homotopy elements g_1 and g_2 corresponds to a single combined loop with homotopy element $g_1 g_2$ or $g_2 g_1$. However, once a convention is chosen, the element assigned to the concatenation of several loops $g_1, \ldots, g_n$ is fixed. You are free to choose this convention.

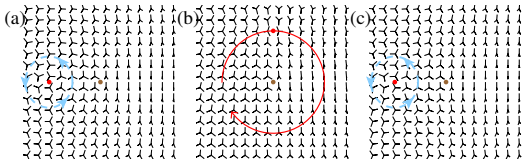


Fig. N1.25 Braiding in practice. (a) The order parameter field in the vicinity of two defects. (b) The order parameter field after one defect has been pulled a quarter-way around the other. (c) The resulting order parameter field after the braiding process. Also, view the evolution of the order parameter in time at the course Web site [31].

(h) *What is the homotopy group element* $\beta \in \mathrm{Dic}_3$ *of the left defect in Fig. N1.23(a)? What is the element* α *for the defect in the center of the figure?*[44] The central brown defect does not change during the braiding process, just as in Figure N1.23. *What is the homotopy group element* $\widehat{\beta}$ *for the left defect in frame (c)? Use the group multiplication rules (eqn N1.25) to verify that your prediction in part (g) for* $\widehat{\beta}$ *is correct.*

N1.18 **Light proton tunneling.** (Dimensional analysis) ③

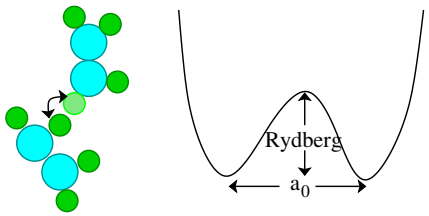


Fig. N1.26 Atom tunneling. A hydrogen atom tunnels a distance a_0, breaking a bond of strength E_{bind} equal to its ionization energy.

In this exercise, we continue to examine a parallel world where the proton and neutron masses are equal to the electron mass, instead of ∼2,000 times larger.

With everything two thousand times lighter, will atomic tunneling become important? Let's make a rough estimate of the tunneling suppression (given by the approximate WKB formula $\exp(-\sqrt{2MV}Q/\hbar)$).

Imagine an atom hopping between two positions, breaking and reforming a chemical bond in the process—an electronic energy barrier, and an electronic-scale distance. The distance will be some fraction of a Bohr radius a_0 and the barrier energy will be some fraction of a Rydberg, but the the atomic mass would be some multiple of the proton mass.

[44] The element assigned to this defect also depends on multiple conventions, including whether the molecules in Fig. N1.23 rotate up or down in the third dimension. Your choice of convention will affect the element α, but not the result of the braiding $\widehat{\beta}$.

In our world, what would the suppression factor be for an hydrogen atom of mass $\sim M_p$ tunneling through a barrier of height V of one Rydberg $= \hbar^2/(2m_e a_0^2)$, and width Q equal to the Bohr radius a_0? How would this change in the parallel world where $M_p \to m_e$? (Simplify your answer as much as possible.) (Use the real-world[45] a_0 and Rydberg for the parallel world, not your answers from a previous exercise. Also please use the simple formula above: don't do the integral. Your answer should involve only two of the fundamental constants.)

N1.19 **Beer.** ③

We shall experimentally explore the statistical mechanics of beer. To explore all aspects, you will need

- Warm and cold non-alcoholic beer, in bottles or cans. It should not be too filtered, so that the head of beer is stable for some time. (Sodas will not work well.)
- Disposable transparent cups, spoons, bottle opener if necessary,
- Salt, sugar, and sand,
- A tub or pail to dispose of liquids.

Avoid shaking the beer before opening. (We will shake it later.) Carefully open a beer, catching any overflow in the cup. Notice there is a pressure release upon opening.

There are a variety of ideas one might explore.

Rigidity transition. *Is beer foam a solid, liquid, or gas? Are beer bubbles in the beer a solid, liquid, or gas? What is the nature of the transition as the bubbles approach the surface, losing the fluid between them?*

Coarsening. *Are the bubbles in the beer changing in size or arrangement before they pop at the top? Can you observe what the mechanisms are for bubble evolution?*

Nucleation. *Are new bubbles formed deep inside the liquid? Or do they form along the edges of the container? Does this change if you shake the container? Do you remember where and how bubbles of boiling water form?*

Try adding salt. Are the effects chemical, or physical? Speculate. Is there a relation to cloud seeding?

If you are feeling quantitative, what should the nucleation rate be, according to theory? (The surface tension of water is $\sigma \sim 7.56$ N/m at 0C. The prefactor for the nucleation rate is about $10^{13}/s$ per water molecule.)

Thermodynamics. *Which is more foamy, a cold bottle of beer or a warm one? What does that mean about the temperature dependence of the chemical potential of CO_2 in water? Do you remember seeing small bubbles form as you heat water, long before it boils?*

[45] The reduced mass effects you found in the earlier exercise will be much less important for larger atoms and molecules, so we shall not include them here.

Examine an unopened, unshaken bottle of beer. Are there any bubbles?

Take a mostly empty container of beer. Holding it shut, shake it to get as much foam as possible. Can you get bubbles with a larger volume than that of the original beer? How much larger? If you can, find some non-alcoholic beer (like Guinness) that uses nitrogen rather than CO_2. Is the behavior different?

At what partial pressure of CO_2 in the beer will bubbles stop forming? At what partial pressure will it reach equilibrium with the air in the room? How does this relate to the boiling and evaporation of water?

Consider a bubble of gaseous CO_2 in equilibrium with the beer. What determines whether molecules of CO_2 will move into or out of the bubble? If you are feeling quantitative, the beer acts as an ideal gas of CO_2 (or N_2) except for a free energy per molecule due to interactions of a dissolved molecule with the surrounding water. The beer foam forms at nearly constant temperature when the pressure changes as the bottle is opened.

N1.20 **Zeros in a byte.** Computer Science ③

Messages are sent in binary code (ones and zeros). The messages are all of length eight (one byte). On average, every incoming message is equally likely.

(a) *Before reading the message, how many different messages could this be? Measure your initial ignorance using the Shannon entropy. How much does your entropy decrease when you read the message?*

A filter keeps track of the number of zeros in the message (so 01010111 would report three). Note for example that there are several possible messages with three 0s, but only one with eight 0s.

(b) *If you learned that the filter returned three, what is your remaining ignorance S_3? What is your average ignorance $\overline{S}$ after knowing the result of the filter, averaged equally over all possible one-byte messages? Is your ignorance after being told there were three zeros larger or smaller than your average ignorance after being told the number of zeros? Measure your ignorance in Shannon entropy.* (See eqn 5.34 in Section 5.3.2 and Exercise N1.11.)

(c) *Assuming again that the incoming possible bytes are all equally likely, what fraction of the information per byte is contained in the number of zeros, as measured by Shannon entropy?*

N1.21 **Pendulum ergodicity.**[46] ③

A Hamiltonian system is never ergodic in phase space if the energy $\mathcal{H}$ varies with position and momentum, since energy is conserved. One discusses instead whether a Hamiltonian system at a certain energy is ergodic. Even here, we must take the thickness $\delta E \to 0$ before discussing ergodicity. Mathematically, the energy shell is thus the surface of constant energy with a measure that weighs regions according to their thickness (according to $1/|\nabla\mathcal{H}|$, see Exercise 3.14).

[46] This problem was developed in collaboration with Stephen Thornton.

Can a Hamiltonian be ergodic at certain energies, and not at other energies? Consider the case of a pendulum.

The energy of a simple frictionless pendulum is given by

$$\mathcal{H}(\theta, p) = \frac{p^2}{2m} + mg\ell\,(1 - \cos\theta) \tag{N1.26}$$

where, as usual, m is a point mass at the end of the pendulum at a length ℓ, g the acceleration due to gravity, and θ the angle away from pointing straight down. The pendulum is hung from a rod that can spin freely around its axis, so that the pendulum, when launched at high energy, can spin round and round without friction. As usual, the small oscillation frequency at the bottom of the well is $\omega = \sqrt{g/\ell}$. Note that the potential energy is in the range $[0, 2mg\ell]$, corresponding to the pendulum pointing down or up.

The experimentalist measures the time average of the total energy $\mathcal{H}$, the angular momentum $p\ell = m\ell^2\dot{\theta}$ around the axis of the rod, and the kinetic energy $p^2/2m$.

(a) The pendulum is started horizontally at rest, with energy $E = mg\ell$. *Describe the motion of the pendulum at later times. What is the time average of the energy? Of the angular momentum? Is the motion ergodic at this energy? Is the time average of the kinetic energy equal to the microcanonical average? Describe the ergodic components of the energy shell to justify your answer.* (Hint: No calculations should be required.)

(b) The pendulum is started vertically pointing down, with kinetic energy $E = 3mg\ell$ with $p > 0$. *Describe the motion of the pendulum at later times. What is the time average of the energy? Is the motion ergodic at this energy? Is the time average of the kinetic energy equal to the microcanonical average? Is the time average of the angular momentum equal to the microcanonical average over the energy shell? Describe the ergodic components of the energy shell to justify your answer.* (Hint: No calculations should be required.)

Clearly our microcanonical pendulum at low energies is not what physicists would call an equilibrium system, even if it satisfies the mathematical definition of ergodic. To address this issue, the mathematicians introduce a stronger condition – that of *mixing*. A system is mixing if the correlation between the system at two different times goes to zero as the time goes to infinity – the memory of the past disappears with time.

(c) *As in part (a), consider the motion of our undamped pendulum with energy $E = mg\ell$. Is it mixing? What would you measure at distant times t and $t + \tau$ that would show you that the system remembers the distant past?*

We now connect the pendulum to a heat bath (say, by running a Langevin simulation), at temperature $k_B T \ll mg\ell$.

(d) *Using your analysis in Exercise 6.11, estimate the rate at which the pendulum will swing over the top, changing θ by either $\pm 2\pi$. Do you expect the time averages will equal the canonical averages for the three measured quantities? Will the two types of averages be equal for the measured quantities if $k_B T \gg mg\ell$?*

N1.22 **Random walks on a lattice.**[47] ③

In Exercise 9.12, we shall explore vacancies diffusing in silicon, and how they flow upward under a gravitational force. In this exercise, we shall treat the random walk of a single vacancy with a simplified square-lattice model.

The silicon vacancy moves whenever one of its four neighbors hops into its position, allowing it to effectively hop to the neighbor position. Assume a random walk with a hop of $a = 2.35$Å, the silicon nearest-neighbor distance. Each vacancy hops every $\Delta t = 6\,\mathrm{ms}$ at random to one of its four neighbors along directions $\pm a\widehat{x}$ and $\pm a\widehat{y}$. We have the option of applying a strong bias field V to the vacancy in the $+y$ direction, such that the probability of hopping in the four directions changes to $p_y^+ = 0.3$, $p_y^- = 0.2$, $p_x^+ = 0.25$, and $p_x^- = 0.25$.

(a) *In the absence of the bias field, what is the RMS distance a vacancy starting at the origin will travel in one minute (ten thousand hops)? What is the diffusion constant D for our square lattice?* (Warning: The relation between the RMS distance and the diffusion constant is dimension dependent.)

Your answer should be within a factor of two of the the real diffusion constant at 1200°C, $D_{\mathrm{Si}} = 3.3 \times 10^{-18} m^2/s$. This is about 200°C below silicon's melting point (see Exercise 9.12).

(b) *In the presence of a bias field, what is the mean displacement it will travel in time Δt? What is the drift velocity v_{model}? Estimate the mobility γ for real silicon at $T = 1200°C = 1473\,K$ using the Einstein relation $D_{\mathrm{Si}}/\gamma = k_B T$ (see Equation 2.22 in Section 2.3). Estimate the drift velocity v_{Si} from the force $F = m_{\mathrm{Si}} g$ due to gravity in real silicon at this temperature. Is our bias field much stronger or weaker than gravity?*

Most external perturbations like electric fields and mechanical stresses are tiny on the scale of atoms. Gravity is an unusually weak perturbation.

(c) *In the presence of the bias field, what is the variance $\overline{x^2}$ in the x direction after a single step from the origin? What is the variance $\overline{(y-\overline{y})^2}$ in the y direction after a single step?* (First write your answers in terms of a^2, and then evaluate in terms of the lattice constant. Keep enough decimal places to distinguish the two.)

The external bias makes the diffusion anisotropic. The new emergent diffusion equation for the probability $\rho(x, y, t)$ in a field should be of the form

$$\frac{\partial \rho}{\partial t} = -\gamma F \frac{\partial \rho}{\partial y} + D_x \frac{\partial^2 \rho}{\partial x^2} + D_y \frac{\partial^2 \rho}{\partial y^2} \tag{N1.27}$$

(d) *Using your answers above, evaluate D_x, D_y, and the combination γF for our model. How does the diffusion parts differ from your answer for part (a)?*

[47] This problem was developed in collaboration with Matthew Dykes.

N1.23 **Averaging over disorder.**[48] ③

A two-state spin takes values $S = \pm 1$. It is in an external field h, so that its Hamiltonian is

$$\mathcal{H} = -hS. \tag{N1.28}$$

It is connected to a heat bath at temperature T.

(a) *Compute its partition function Z, its Helmholtz free energy A, the entropy S, and the specific heat*[49] *c as a function of h and T. What is the entropy at $T = 0, h > 0$ and at $T = \infty$? (The $T \to 0$ limit is tricky: a graphical solution is fine.) Is the difference as expected from our understanding of information entropy?*

To model a system with dirt – a disordered system – one often adds a random term to the Hamiltonian (like a random field for each spin). One then averages the answer over the probability distribution of the disorder to predict the behavior of a large system. This turns out to be trickier than it seems.

Let us calculate the average properties of our spin in a random field h, averaged over a Gaussian probability distribution $\rho(h) = \exp(-h^2/2\sigma^2)/(\sqrt{2\pi}\sigma)$.

(b) *Write in integral form the average of each of the quantities $\overline{Z}$, $\overline{A}$, $\overline{S}$, and $\overline{c}$ over the probability density $\rho(h)$.* All but one of these will be infeasible to evaluate in closed form. *Evaluate the integral for $\overline{Z}$.*

In interacting systems like spin glasses, it is much easier to calculate the average of Z than the average of $\log Z$ or A. But we run into trouble.

(c) *Define $Z_a = \overline{Z}$, and calculate the corresponding quantities A_a and S_a. Show that S_a goes negative at low temperatures.*

The entropy for each disorder you calculated in part (a) never goes negative. So its average cannot be negative! We seem to be stuck with the integrals we cannot do in closed form.

(d) *Define $A_q = \overline{A}$. Argue that S_q, defined as the appropriate derivative of A_q, is equal to $\overline{S}$ from part (b).*

Let us briefly consider a simpler scenario, where h can take only the three values 0 or $\pm h_0$ (with $h_0 > 0$), each with probability 1/3.

(e) *Write A_q and A_a exactly for this case, and evaluate them in the limit $T \to 0$. Using $\overline{A} = \overline{\langle E \rangle} - T\overline{S}$, what value should you expect for the average free energy at $T = 0$? Does A_a appear to be giving unfair weights to disorder configurations with lower-energy states?*

Thus A_a gives an unfairly large weight to members of the disordered ensemble that have unusually low energy configurations. For spin glasses, A_a gives unfair weights to

[48] This problem was developed in collaboration with Stephen Thornton.

[49] Section 6.1 discusses the specific heat at constant volume c_v, but the formulas are the same because here there is no volume to be fixed.

systems like the non-disordered Ising model, where a single spin configuration can make all the bonds happy. This leads to an unphysical ferromagnetic–like transition.

Why the choice of subscripts? When we want to freeze our dirt into a particular configuration, we quench the system quickly to a low temperature. (The blacksmith pounding the red-hot horseshoe, after they get it into shape, quenches it in a bucket of water.) A_q is the *quenched* free energy. We anneal a defective crystal by heating it up to a large temperature T_0 where its defects have enough energy to rearrange and come to equilibrium. $A_a = -k_B T \log(\overline{Z})$ is called the *annealed* free energy. But why does our A_a correspond to an annealed free energy, where the "defects" come to equilibrium?

(f) *Show that $Z_a(T_0)$ from part (c) at a particular temperature T_0 is the true partition function for a Hamiltonian*

$$\mathcal{H}_a = h^2 k_B T_0 / 2\sigma^2 - hS + C, \tag{N1.29}$$

where the constant $C = \frac{1}{2} k_B T_0 \log(2\pi\sigma^2)$. Thus Z_a discusses a system where h and S are both weighted according to the Boltzmann distribution (so the field fluctuates to equilibrate with the spin). In systems like spin glasses, one can calculate annealed averages because they are, in disguise, the correct partition function for an undisordered equilibrium system.

We must end with the *replica trick* that people use to bypass the infeasible integrals we get from trying to average the $\log(Z)$, as in $A_q = -k_B T \overline{\log Z}$. One can often calculate $\overline{Z^n}$, the annealed disorder average of n replicas of a system. (Again, it is feasible because it is in disguise an equilibrium physical system, whose dirt equilibrates with the spins.) We then can find the average $\overline{\log(Z)}$ and hence A_q:

(g) *Show that $\log x = \lim_{n\to 0}(x^n - 1)/n$ by writing $x^n = \exp(n \log x)$.*

We can then take the average of both sides and write $\overline{\log(Z)} = \lim_{n\to 0}(\overline{Z^n} - 1)/n$. Finding the right way of taking the limit $n \to 0$ is harder than we are suggesting. The original researchers used a "replica symmetric" method that works for many systems, and works well in spin glasses for temperatures above the glass transition. Below the glass transition, one must do something more exotic. Georgio Parisi received the Nobel Prize in Physics in 2021 for showing certain disordered systems undergo a "replica symmetry breaking" transition as the temperature is lowered, where certain correlations within the system change dramatically in the spin glass phase. These methods have been shown by Parisi and others to be powerful tools for solving models of ordinary glass, analyzing deep neural network models in machine learning, and providing the fastest algorithms for challenging "NP complete" models in computer science (see Exercise 8.15).

N1.24 Distinguished and undistinguished particles.[50] ③

In quantum physics, all electrons are fundamentally indistinguishable. In classical physics, we often have systems where the particles are in principle distinguishable, but

[50] This exercise was developed in collaboration with Stephen Thornton.

nothing in our experiment depends on which particle is which. (Examples might include colloidal particles, grains of sand, atoms differing only in their isotopes or nuclear spins, etc.) In Section 3.5, we introduced the idea of "undistinguished" particles, and in Section 7.5 we argued that non-interacting undistinguished particles obey Maxwell–Boltzmann statistics. How can we derive the properties for these undistinguished particles, and why is it useful?[51]

In statistical mechanics, when our experiments do not depend on some microscopic details about a system, we create a free energy by "integrating out" the microscopic details. If we promise to ignore the differences between particles that are in principle distinguishable, we must sum over all the possible ways of labeling the particles.

(a) *Suppose we have a system composed of M different types of particles, with N_1 indistinguishable particles of type 1, ..., N_M indistinguishable particles of type M. The system has a total of N particles, in a microcanonical ensemble with total energy E in a volume V, which has an energy shell volume of*

$$\Omega^{\text{dist}}(E, V, N_1, N_2, \ldots, N_M). \tag{N1.30}$$

Assume the energy of the particles is independent of their identities – that all Hamiltonians describing our experiment is the same under a permutation P_n of the positions and their momenta,

$$\begin{aligned}\mathcal{H}(\mathbf{x}_1, \ldots \mathbf{x}_N, \mathbf{p}_1, \ldots, \mathbf{p}_N) = \\ \mathcal{H}(\mathbf{x}_{P_1}, \ldots \mathbf{x}_{P_N}, \mathbf{p}_{P_1}, \ldots, \mathbf{p}_{P_N}).\end{aligned} \tag{N1.31}$$

Calculate $\Omega^{\text{summed}}(E, V, N)$ by summing Ω^{dist} over all permutations P, and write it in terms of factorials and $\Omega^{\text{dist}}(E, V, N, 0, \ldots, 0)$, as if all particles were of the first type. Show that, if all particles are distinct, so $M = N$ and $N_m \equiv 1$, that $\Omega^{\text{summed}}(E, V, N) = N!\, \Omega^{\text{dist}}(E, V, N, 0, \ldots, 0)$.

Given our experiments do not care about the identities, Ω^{summed} will give exactly the same predictions as did Ω^{dist} (as is always true of free energies formed in this way). But is that what we want?

Remember Nernst's law (the "third law of thermodynamics"), that the quantum entropy per particle vanishes at zero temperature? This is true if all of the particles are identical, or if all of the particles are different in ways important to the Hamiltonian (so the ground state is not invariant under permutations). This is because there are usually only a small number of states with the minimum possible energy E_{min} (see footnote 20 on page 151).

(b) *Consider a quantum system with N distinguishable particles, with a Hamiltonian invariant under permutations of the particles, and with a crystalline ground state with energy E_{min}. What is the entropy per per particle $S^{\text{summed}}(E_{\text{min}}, V, N)/N$ at zero temperature? Use Stirling's formula to simplify it. Does it go to zero for large N? Is it at*

[51] See also Cates and Manoharan [7], who discuss several colloidal paradoxes connected to this exercise.

least extensive? (Hint: Assume $\Omega^{\text{dist}}(E_{\text{min}}, V, N, 0, \ldots, 0)$ has only one lowest energy configuration.)

This is technically correct. If the particles are distinct, even at zero temperature they can be swapped around in lots of ways. But if our experiments do not care about this swapping, perhaps we can drop the $N!$ and forget about this annoying permutation entropy?

This drastic step is precisely how we defined Maxwell–Boltzmann statistics in eqn 7.48, $\Omega(N)^{\text{MB}} = \Omega(N)^{\text{dist}}/N!$.

The fact that entropy should be extensive was a consequence of the third key property of Shannon entropy (Section 5.3.2). The entropy S^{summed} is not extensive: the sum $S_1 + S_2$ for two weakly interacting subsystems (and hence the product $\Omega_1\Omega_2$ of their energy shell volumes) is not anywhere near that of the combined system. In particular, if you double the total number of particles, energy, and volume, S^{summed} does not double: the permutation entropy grows much faster than linearly in N. Maxwell–Boltzmann statistics cures this second annoyance.

In section 3.5, we divided the distinguishable energy shell volume Ω_{crude} for the ideal gas (eqn 3.48) by $N!$ to get our final formula for the entropy (eqn 3.57).

(c) *Show that the three terms in the last line of eqn 3.49 for the 'crude' entropy S^{crude} in the limit of large systems scale in the same way with system size, and are not extensive.* (Hint: Use Stirling's formula.) *Show that our final formula for the ideal gas entropy (eqn 3.57) is extensive. In detail, fill in the steps between eqn 3.55 and 3.57 (ignoring Planck's constant) to see how dividing Ω^{crude} in eqn 3.48 by $N!$ fixes this.*

Maxwell–Boltzmann statistics also is needed to talk about separating weakly-interacting subsystems – separating our experiment from the rest of the world. We use this extensively in our model of temperature and chemical potential as connecting our subsystem to an external bath. Here we illustrate this problem, not with a classical phase space, but with a system with discrete energy levels. We shall see that Maxwell–Boltzmann statistics in this case does not have a simple interpretation in terms of counting how to place undistinguished particles into energy levels, but does serve to fix another serious problem with applying Ω^{diff} to systems with varying number of particles.

Figure N1.25 shows a system at fixed energy E and particle number N, with a subsystem with energy E_s and number N_s, and a bath with the remaining energy $E_b = E - E_s$ and $N_b = N - N_s$. All particles in our bath are in a state with energy zero. Particles in the subsystem occupy one of two states, with energies $+\Delta$ or $-\Delta$.

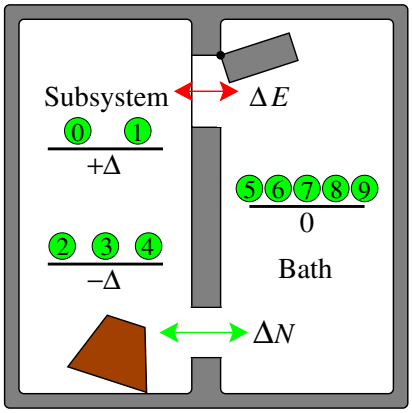


Fig. N1.27 Microcanonical, fixed N system composed of a subsystem potentially exchanging energy and particles with a bath. The current state has system and subsystem energy $E = E_s = -\Delta$, $N_s = 5$ and $N = 10$.

Let us begin with $N = 3$ and $E = 0$. We shall be interested in the volume of the energy shells $\Omega(E, N)$, which here is given by the number of different configurations that the particles can take. Usually we would use something like eqn 3.23, giving the probability $\rho_s(E_s, N_s)$ of the subsystem to have energy E_s, should generalize to

$$\rho_s(E_s, N_s) \stackrel{?}{=} \frac{\Omega_s(E_s, N_s)\Omega_b(E - E_s, N - N_s)}{\Omega(E, N)}; \tag{N1.32}$$

the probability should be the fraction of the energy surface that has the desired energy and number in the subsystem. Here Ω_s and Ω_b are calculated by assuming that they are isolated from one another.

First consider distinguishable particles. Here, to calculate $\Omega_s^{\text{dist}}(E_s, N_0, N_1, N_2)$ for the isolated subsystem we fill it with particles 0 and 1, and the bath with particle 2.

(d) *What is the number of states* $\Omega_s^{\text{dist}}(0, 1, 1, 0)$ *for when the subsystem has zero energy and has particles* 0 *and* 1 *inside? What is* $\Omega_b^{\text{dist}}(0, 0, 0, 1)$ *when particle* 2 *is in the bath? What is the total number of states* Ω^{dist} *with energy zero and three distinguishable particles, assuming the two sides are allowed to exchange energy and particles? Compute the probability* $\rho_s^{\text{frac}}(0, 2)$ *from the fraction of configurations of the whole system that have zero energy and two particles in the subsystem. Compute the probability* $\rho_s^{\text{pred}}(0, 2)$ *predicted by eqn N1.33, assuming* $\Omega_s(E_s, 2)$ *and* $\Omega_b(E - E_s, 1)$ *are given by the uncoupled calculations* $\Omega^{\text{dist}}(E, 1, 1, 0)$ *and* $\Omega_b^{\text{dist}}(E, 0, 0, 1)$. *Do they agree? Which is correct? Why are they different?*

This too is fixed by using Maxwell–Boltzmann statistics.

(e) *What are* $\Omega_s^{\text{MB}}(0, 2)$ *and* $\Omega_b^{\text{MB}}(0, 1)$*? What is* $\Omega^{\text{MB}}(0, 3)$*, including contributions from other particle arrangements of zero energy?* (Hint: Use eqn 7.48 and your answers for Ω^{dist}.) *Which of the three has no natural interpretation as a number of ways of filling*

the states with undistinguished particles? Compute the probability given by eqn N1.33. Does it agree with the correct answer from part (d)?

These $N!$ factors might remind you of the entropy of mixing in Section 5.2.1. There we started with a symmetric split, with the same number of atoms in the subsystem and the bath, and with all the interaction energies between the atoms the same (the Hamiltonian is independent of permutations). If the atoms on the left and right are different (orange and blue), when the partition is lifted the entropy jumped. But if they are the same, the entropy hardly shifted.

Of course, the entropy of mixing is real! If we had a membrane that would let only orange particles through, half of them would diffuse into the blue half leading to a big pressure difference. We could then extract work by moving the wall between the two subsystems. We can still use Maxwell–Boltzmann statistics for the orange atoms and the blue atoms separately. But we should not use Maxwell–Boltzmann statistics ignoring the orange/blue difference, if we plan later to use a membrane whose Hamiltonian distinguishes them.

N1.25 **Localization.**[52] (Quantum, Condensed Matter) ③

In Section 7.4, we discussed how non-interacting electrons provide a useful model for metals, even though the electron-electron interactions are strong. The Fermi liquid of quasiparticles is a kind of adiabatic continuation of the noninteracting electron system, connected by perturbation theory. Here we shall study how a one-dimensional non-interacting metal responds to disorder. We shall discuss how metals with weak disorder are understood by perturbing around the clean state. We shall discover that strong disorder leads to an insulating system whose eigenstates are not extended, but *localized*. We will describe these localized states explicitly by perturbing about a state of isolated atomic states.

Consider a one–dimensional chain of atoms n, each with one noninteracting electron state $|n\rangle$ of energy U_n, that can be occupied by either zero or one spinless electrons.

[52] Hints for the computations can be found at the book web site [31].

Electrons can hop between atoms with matrix element t, leading to the Hamiltonian

$$\begin{aligned}\mathcal{H} &= \sum_{n=1}^{N} U_n |n\rangle\langle n| \\ &\quad - t \sum_{n=1}^{N-1} \left(|n\rangle\langle n+1| + |n+1\rangle\langle n| \right) \\ &= \begin{pmatrix} U_1 & -t & 0 & \dots & 0 & 0 \\ -t & U_2 & -t & \dots & 0 & 0 \\ 0 & -t & U_3 & \dots & 0 & 0 \\ 0 & 0 & -t & \dots & 0 & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots & -t \\ 0 & 0 & 0 & \dots & -t & U_N. \end{pmatrix}\end{aligned} \tag{N1.33}$$

We shall take the random energies U_n as uniformly distributed between $-W$ and W.

Without disorder ($W = 0$), this is a textbook model used to describe energy bands in crystals. Three dimensional analogs of this 'tight-binding' model are quite realistic models of Fermi surfaces and energy bands in real materials.[53]

(a) *Write a function that builds the Hamiltonian matrix of eqn* **??** *with size N, bandwidth $2W$, and hopping matrix element t. Studying zero disorder $W = 0$, find the eigenvectors for $t = 1$, and $N = 100$, sorted by their eigenvalues. Plot the eigenvectors for the four lowest energies. Check numerically that these four are sinusoidal with wavevectors $k_\alpha = \pi\alpha/(N+1)$ appropriate for a box of size N with hard-wall boundary conditions half a grid spacing to either side. Check that their four eigenvalues are the corresponding $E_{k_\alpha} = -2t\cos(k_\alpha)$*

Imagine a 1D metal at zero temperatures with electrons filling the states up to a Fermi surface, here just two points at some $\pm k_{\text{Fermi}}$. Consider a packet of electrons made up of eigenstates near k_{Fermi} traveling to the right. The wavepacket[54] will travel, as usual, at the group velocity $\mathrm{d}E_k/\mathrm{d}k|_{k_{\text{Fermi}}}$, without dissipation.

Now let us explore what happens when we add a weak disorder.

(b) *Build a Hamiltonian with weak disorder $W = W_{\text{weak}} = 0.04$, $t = 1$, and $N = 100$. Plot the lowest four eigenvectors. Are the eigenstates still extensive (reaching from one side of the box to the other)?*

[53] We shall see that even a small disorder changes the metallic behavior of one-dimensional electrons in a qualitative and interesting way. Indeed, one dimensional electrons are unstable in many interesting ways. Adding interactions between electrons, they become *Luttinger liquids*, with emergent scale invariance. Adding interactions with lattice vibrations, they can become topological insulators, with solitons and fractional charges.

[54] Wavepackets are used to connect waves to particle-like motion. In a non-disordered system, one superimposes states with similar momenta to make a spatially localized wavefunction, which then moves with the group velocity of the wave. We discuss wavepackets to motivate the effects of disorder, but no knowledge about them is required to do this exercise.

At this point, we could use perturbation theory to calculate the disordered eigenstates and energy levels. We could then create wavepackets and see how they evolve. In three dimensions, the scattering off of the disorder changes the electron transport qualitatively. Instead of wavepackets moving forever in one direction (ballistic transport, infinite conductivity), one gets diffusive motion of the electron probability through space (disorder providing an elastic scattering length, and a finite conductivity). In three dimensions, this is a good model for metals with impurities or dopants, illustrating how one can understand complex behavior by perturbing around solvable special cases.

Instead, let us examine what happens at large disorder W, or equivalently, small hopping t. (All of our eigenvectors depend only on W/t, and we will perturb in t to study the localized states.)

(c) *Set $t = t_{\rm Weak} = 0.1$, $W = 1$, and $N = 100$, plotting the ten eigenvectors with lowest energies. Also do a log-linear plot of the probability density (absolute square of the wavefunctions) for these eigenvectors. Do the eigenstates still look as if the will be extensive (stretching from one end of a macroscopic wire to the other)? What solvable special limit of $\mathcal{H}$ should we use to capture this new behavior?*

Here we find the eigenstates appear *localized* – fixed in space near individual 'atoms'. The probabilities in these states fall exponentially with distance from their centers. A wavepacket formed from localized states like these cannot transport current: for large disorder, our model describes an insulator.

Just as one can use perturbation theory to describe dirty metals in three dimensions from models like ours, we can use perturbation theory to calculate and understand these localized states. You should remember the use of second-order perturbation theory to describe the energies of a Hamiltonian $\mathcal{H} = \mathcal{H}_0 + V$ for small V. You may not remember that the first step was to use first-order perturbation theory to determine the eigenvectors. If $|\Psi_i^{(0)}\rangle$ has unperturbed eigenvalue $E_i^{(0)}$, then to first order

$$|\Psi_i^{(1)}\rangle = |\Psi_i^{(0)}\rangle + \sum_{i\neq j} \frac{\langle\Psi_j^{(0)}|V|\Psi_i^{(0)}\rangle}{E_i^{(0)} - E_j^{(0)}} |\Psi_j^{(0)}\rangle. \tag{N1.34}$$

If the hopping is small compared to the disorder, let us perturb in t.

(d) *What are the eigenenergies for $\mathcal{H}$ in eqn **??** with $t = 0$? Argue that, to first order in t, the new eigenstates will be confined to three adjacent sites.*

(e) *Write a function, given $\mathcal{H}$, i, t, and N, that gives the perturbed eigenstate to lowest order in t that is centered at site i. Find the site of the ground state with largest probability. Plot the true ground state and your first-order approximation to it.* (Hint: You may be unlucky, and happen to have a neighbor site with a near degenerate energy. Just create a new Hamiltonian and try again.)

What controls whether our model is a metal or an insulator? For a given W/t, are all the states either extended or all localized? Or could there be some mixture?

We can examine this by defining a rough measure of how spread out the wavefunction is, called the *participation ratio*:

$$P(\psi) = \frac{\left(\sum_n |\psi(n)|^2\right)^2}{\sum_n |\psi(n)|^4} = \frac{1}{\sum_n |\psi(n)|^4}. \tag{N1.35}$$

(f) *Show that a state whose probability is spread uniformly among M sites has $P = M$. At zero disorder, what is the participation ratio for the lowest energy state? For the long-wavelength next few states? What is the ratio for a localized state that decays exponentially, $\psi(j) \propto \exp(-|i-j|/\lambda)$, in an infinite chain, with λ much larger than one?*

If the participation ratio $P \ll N$ we can reasonably expect that the eigenstate is localized.

(g) *Calculate the participation ratio for all the eigenstates for intermediate disorder $W_{\rm inter} = 0.5$, $t = 1$, $N = 100$, and plot them against the energy. Is there a systematic variation? Plot the wavefunction for an energy in the middle of the energy band (eigenvalue E near zero), and one at the top and bottom of the band. Which are less localized – the states near the edges of the band, or the states in the center?*

In experiments, one finds a region of localized states at the edges of a band, and extended states in the middle of the band.[55] Between these is a *mobility edge*, where a metal-insulator transition occurs as more electrons are added.

Finally, can we find a mobility edge for our model? One thing to check is if the wavefunctions might have decay lengths larger than our system (so they just look extended at $N = 100$).

(h) *Find the eigenvectors and eigenvalues for the same parameters as in part (g), $W = W_{inter} = 0.5$ and $t = 1$, except for a much larger system ($N = 2000$ if it is feasible on your system). Plot the participation ratio verses energy, and plot eigenstate in the center of the band and at the two edges. Do the states in the middle of the band now appear localized? Are their participation ratios larger than 100 – the system size in part (g)? Does it make sense that they looked extended in the smaller system, but clearly in an infinite system are localized?*

As it happens, disordered electrons in one dimension are always localized, even for tiny disorder. The spinless, noninteracting electrons we study here are also always localized in two dimensions. In two dimensions, they can become extended when interactions, spin orbit scattering, or strong magnetic fields are added. In particular, 2D electrons in a strong magnetic field exhibit the quantum Hall effect (with extended states around the

[55] Each time we add one order to perturbation theory, we get a wavefunction extending outward by one atom. It appears from the first two terms that each order multiplies the terminal amplitude by a factor $|\psi(i+n)/\psi(i+(n-1))| = |t/(U_i - U_{i+n})|$. Roughly speaking, if U_i is at the edge of the band, this factor is twice as small as if U_i is in the center, so there is less localization at the center. Notice, though, that this is useful only when $t > |U_i - U_{i+n}|$. Rare, nearly degenerate states can mix strongly, even at long distances, making the arguments subtle.

edges). Even more interesting, interacting electrons in a strong magnetic field exhibit the fractional quantum Hall effect – our first example of an experimental system with fractional charges and fractional statistics.

N1.26 Correlation matching.[56] ③

Figure N1.26 shows the 2D correlation function C(**r**) and its 2D Fourier transform $\widetilde{C}(\mathbf{k})$ for a periodic stripe pattern (for which an 8x8 block is shown).

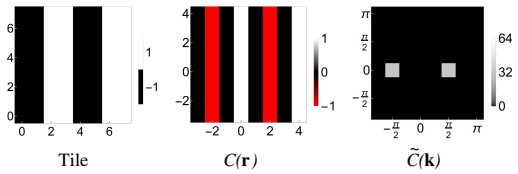


Fig. N1.28 Stripe pattern correlations. Tile: An 8x8 Ising pattern $s(x, y)$ that is repeated to fill space. Black is -1, white is $+1$. C(**r**): The 2D correlation function for the tiled space. $\widetilde{C}(\mathbf{k})$: The Fourier transform of the correlation function $\widetilde{C}(\mathbf{k})$. Note that the zero of both the correlation function and its Fourier transform are not centered; you may imagine invisible rows $-4 \equiv 4$ and $-\pi \equiv \pi$.

Figure N1.27 shows three other 8x8 blocks of spins.

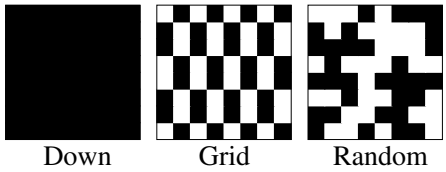


Fig. N1.29 Tiles. Snapshot tiles of an Ising model with period 8 in the x and y directions.

For each of the tiles in Fig. N1.27, identify which of the choices in Fig. N1.28 corresponds to C(**r**)*, and which corresponds to* $\widetilde{C}(\mathbf{k})$*. The color scale and axes for each are given by Fig. N1.26.*

[56] This exercise was developed in collaboration with Stephen Thornton.

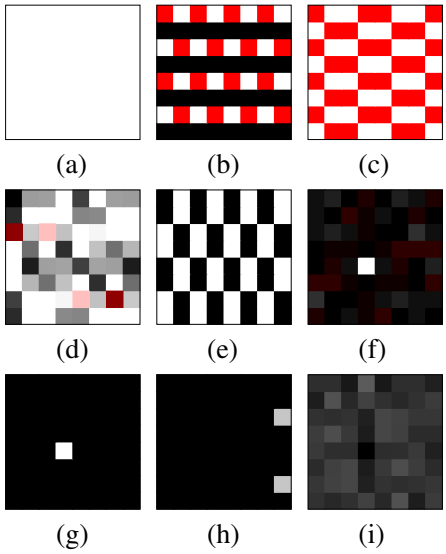


Fig. N1.30 Choices.

N1.27 **Coarsening correlations.**[57] ③

Consider an Ising model on a square lattice, starting in an infinite-temperature random configuration, and shifted to a low temperature. It will coarsen with time, as shown in the two panels in Fig. N1.29.

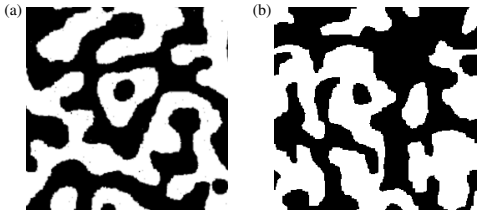


Fig. N1.31 Coarsening snapshots. Two Ising models, quenched from random (infinite temperature) states to temperatures below T_c, each at a snapshot in time after substantial coarsening. One has traditional single-spin flip dynamics, that does not preserve the overall magnetization. The other exchanges the spin of two neighboring sites with a probability that thermalizes the system at constant magnetization (from [32]).

These two pictures and were created with different dynamical rules. In one of them, we use a traditional Metropolis algorithm, flipping one spin at a time. This does not properly describe, say, oil and water separating from one another, because one cannot "flip" an oil molecule into a water molecule – the molecules are locally conserved. The other uses an algorithm invented by Kawasaki, which swaps the spins of neighboring

[57] This exercise was developed in collaboration with Stephen Thornton.

sites (with a probability determined by their six neighbors). This allows up-spin "oil" molecules to diffuse through down-spin "water" regions.

It probably is not obvious from the two snapshots which evolved without conservation and which was conserved. We shall use correlation functions to determine which is which.

What correlation function should we use? Our systems are translation invariant,[58] so the correlation function is expected to depend only on the separation between spins, $\mathbf{r} = \mathbf{x} - \mathbf{x}'$. The system is qualitatively changing in time (as it coarsens), so $\mathrm{C}(\mathbf{r},t) = \langle s(\mathbf{x},t)s(\mathbf{x}+\mathbf{r},t)\rangle$ is a natural choice. The insets of the two panels in Fig. N1.30 show a 2D plot of $\mathrm{C}(\mathbf{r},t)$ for the evolution conditions of the corresponding snapshots in Fig. N1.29, and the main panels plot the angular average[59] $C(r,t) = (1/2\pi)\int_0^{2\pi} \mathrm{d}\theta\, \mathrm{C}\left(r\cos(\theta), r\sin(\theta), t\right)$.

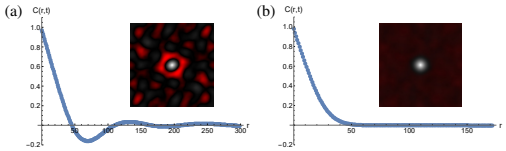


Fig. N1.32 Coarsening correlation functions. The two angular-averaged spin-spin correlation functions, each for the corresponding snapshots in Fig. N1.29(a) & (b). Distances are measured in lattice spacings. (a) The noise in the 2D inset would disappear with an ensemble average over initial conditions, but the oscillations in the $C(r)$ plot are real. (b) The correlation functions are an average of many ensemble snapshots prepared like Fig. N1.29(b).

Thinking about how the two systems grow should allow you to deduce which is which without mathematics.

(a) *Note the oscillations in $C(r,t)$ on the left, that are absent on the right. Decide whether oscillations are more clearly necessary for conserved or non-conserved growth. Explain your reasoning physically, without calculations.* (In particular, your reasoning should be different from the argument below!)

Let us now produce a definitive argument using $C(r,t)$ to settle which figure has non-conserved dynamics. Consider $\mathrm{C}^s(\boldsymbol{\delta},t) = \langle s(\mathbf{x})s(\mathbf{x}+\boldsymbol{\delta})\rangle_{\mathbf{x}}$, the correlation function for a snapshot of a particular coarsening system, and $C^s(r,t)$, its angular average.

(b) *Show, for an N-spin Ising model, that*

$$\int \mathrm{d}\delta_x \mathrm{d}\delta_y\, \mathrm{C}^s(\delta_x,\delta_y,t) = M(t)^2/N, \tag{N1.36}$$

where M is the total magnetization of the snapshot at time t. Go to polar coordinates, and relate this formula to $\int_{r>0} rC^s(r,t)\mathrm{d}r$.

[58] They have periodic boundary conditions.

[59] Here and elsewhere we approximate the discrete lattice of spins as a continuous integral over space, for simplicity. Also, the apparent rotation invariance in these snapshots is only approximate. It is true that the Ising model at T_c has an emergent rotational invariance, but this is not true for low temperatures.

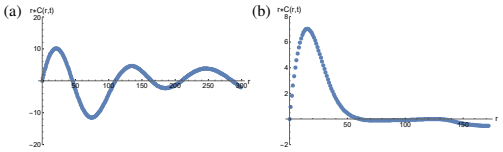


Fig. N1.33 Using $rC(r)$ to study coarsening. The correlation functions shown in Fig. N1.30, multiplied by r. We shall use them to determine which is conserved and which non-conserved, and use one of them to measure the coarsening length scale.

(c) *Presuming both conserved and non-conserved simulations start with zero magnetization at infinite temperature,*[60] *which of the two snapshots in Fig. N1.29 evolved with conserved dynamics? Use Fig. N1.31 and eqn N1.34 to explain why.*

Now let us explore the coarsening dynamics using our understanding of emergent scaling behavior. When the features in the system are large compared to the lattice and small compared to the size of the system, the equal-time correlation function is expected to take a scaling form

$$\mathrm{C}\left(\mathbf{r},t\right) = \mathcal{C}\left(\mathbf{r}/L(t)\right). \tag{N1.37}$$

The system looks statistically similar to itself at previous times, but rescaled by a growing lengthscale $L(t)$. Our scaling assumptions suggest that $L(t) \propto t^{\rho}$, where the critical exponent ρ depends on the type of dynamics (conserved, non-conserved, surface diffusion, ...).

In testing whether universal scaling functions hold, we often use scaling collapses. Usually we would plot $C(r,t)$ versus r/t^{ρ}, especially if we know what power to expect. But in many cases we would like to use our simulations or experiments to measure $L(t)$, see if they obey our scaling assumption (eqn N1.35), and measure ρ. So, we would like a reliable way of measuring $L(t)$.

(d) *Show that the scaling assumption of eqn N1.35 implies that*

$$\int \mathrm{d}^2\mathbf{r}\, \mathrm{C}(\mathbf{r},t) = \alpha\, L(t)^2, \tag{N1.38}$$

where α is a time-independent constant. What is α in terms of the scaling function $\mathcal{C}\left(x,y\right)$?

This suggests that, so long as α is nonzero,[61] a good measure of the length scale is

$$L(t) \propto \sqrt{2\pi \int_0^{\infty} \mathrm{d}r\, r\, C(r,t)}, \tag{N1.39}$$

[60] We could be more fussy. At $T=\infty$ in the finite Ising model, M is the sum of N numbers ± 1, so random walks tell us that $\langle M^2 \rangle = N$, which is precisely the contribution to $\mathrm{C}(\mathbf{r},0)$ at $\mathbf{r}=0$. So the negative regions for must cancel the positive regions for $\mathbf{r}>0$.

[61] Note that we learned from parts (b) and (c) that α is zero for coarsening with a conserved order parameter. All is not lost. We could use a different moment of $C(r,t)$, another feature (like the first zero-crossing), or just rescale the curves until they overlap.

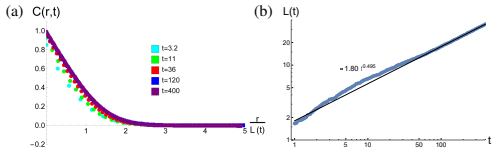


Fig. N1.34 Coarsening correlation collapse. (a) When plotted versus $r/L(t)$, the correlation functions $C(r,t)$ at late times collapse onto a single curve, validating our scaling ansatz in eqn N1.35. (b) The measured $L(t)$ curves are roughly straight on a log-log plot, implying a power law $L(t) \sim t^{\rho}$, as predicted by emergent scale self-similarity. The fit is to the late-time behavior.

To test that our scaling ansatz works, we can measure $L(t)$ for various times t, and plot all the curves $C(r,t)$ versus $r/L(t)$. Shown in Fig. N1.32(b) is a plot of $L(t)$ versus t for the evolution used in panel (b) of the previous figures in this exercise. Fig. N1.32(a) shows $C(r,t)$ plotted against $r/L(t)$ for various times t.

At late times, we expect the 2D scaling form in eqn N1.35 to hold, and we observe in Fig. N1.32(a) that the 1D angular averages $C(r,t)$ curves collapse well when plotted versus $r/L(t)$.

(e) *Argue that the common curve $C(r,t)$ vs. $r/L(t)$ at late times is the angular average of $\mathcal{C}$.* Scaling collapses allow one to measure universal scaling functions with experiments or simulations. *Is the power law found in Fig. N1.32(b) closer to that predicted for conserved, or non-conserved coarsening?*

N1.28 Ising critical correlations.[62] ③

Here we examine numerical simulations of the equilibrium, equal-time, spin-spin correlation function

$$\mathrm{C}(\mathbf{r}) = \langle s(\mathbf{x}) s(\mathbf{x}+\mathbf{r}) \rangle = C(r) \tag{N1.40}$$

near the Ising critical point (where the 2D correlation function becomes isotropic, so it only depends on $r = |\mathbf{r}|$).

Figure N1.33 shows snapshots of the Ising model at $T < T_c$, at $T = T_c$, and at $T > T_c$, together with their equilibrium, equal-time, radially averaged correlation functions $C(r)$.

[62] This exercise was developed in collaboration with Stephen Thornton.

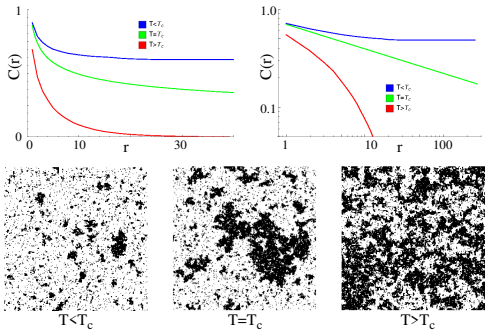


Fig. N1.35 Ising equilibrium correlation functions, on a linear-linear and log-log plot, at the three temperatures shown in the snapshots below. The distance r is measured in lattice spacings, and the snapshots are for a 256×256 lattice. From [9].

It is challenging to estimate the correlation length from snapshots of systems that are approaching critical points.

(a) *Estimate the correlation length using the snapshots for $T < T_c$ and $T > T_c$, and mark the corresponding point on the corresponding curve in the correlation function plots. Discuss how you determined the correlation length in each case. What feature of $C(r)$ in the curves roughly corresponds to your estimates of the correlation length from the snapshots?*

(b) *Use the plots of the correlation function at T_c to estimate the critical exponent η for the 2D Ising model, where $C(r) \sim r^{-\eta}$ at T_c in $d = 2$. Compare with the known exact value.*

(c) *Sketch the connected correlation function $C_{conn}(r) = \langle s(\mathbf{r}+\mathbf{x})s(\mathbf{x})\rangle - \langle s(\mathbf{x})\rangle^2$ for the $T < T_c$ curve in the linear-linear plot.*

N1.29 Rubber band dynamics I: Random walk.[63] ③

Exercise 5.12 introduced an entropic model for a rubber band – N segments of length d pointing forward and backward at random. Here we shall consider the fluctuations of this entropic rubber band, as the individual segments flip back and forth. We shall also examine how it evolves when its endpoint is pulled by an external parabolic potential (Fig. **??**).

How does the length evolve in time, in the absence of a force from the parabolic potential? Consider flipping one of the segments at random. If we choose one of the n_+ segments pointing forward, flipping it will decrease the length L by $2d$. Conversely, flipping one of the $n_- = N - n_+$ segments will increase the length. For convenience, let us set $d = 1$ for the simulation. We also measure time in sweeps (attempting to flip each segment once), so $\Delta t = 1/N$ each time a step in our random walk is taken.

[63] This exercise was developed in collaboration with Stephen Thornton

(a) *What are n_+ and n_- in terms of L and N? Write a routine* `flip(L,N)` *that, with probability n_+/N returns $L-2$, and with probability n_-/N returns $L+2$. Assume our chain starts out with its endpoint at the origin, $L=0$. Plot the evolution of the length with time, for a chain length $N=100$ and for* 10,000 *steps (to time $t=100$). Does the random walk drift away at long times?*

In Exercise 5.12, we calculated the spring constant K for the entropic chain. Examine your solution (or the answer key) for that exercise. At a temperature T, our rubber band should mostly explore only configurations where the free energy $\frac{1}{2}KL^2$ is not much larger than T.

(b) *Use equipartition and K from Exercise 5.12 to derive a formula for the average mean square $\langle L^2 \rangle$ expected for a chain of length N. Compare this with that of your simulated random walk.* (Hint: Your answer should not depend on the temperature! And the equipartition answer should agree with the length of a random walk with stepsize ± 1.)

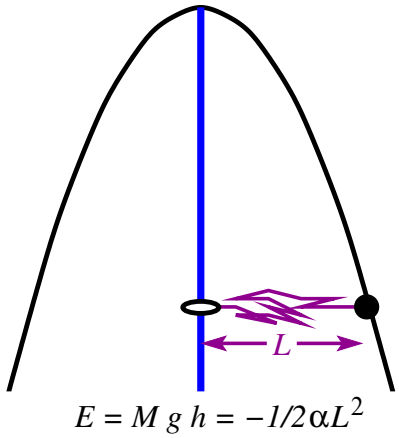


Fig. N1.36 Rubber band stretched by weight on a hill. We place the endpoint of the spring (disk at L) in a parabolic potential $-\frac{1}{2}\alpha L^2$, as suggested by this schematic diagram.

We could now add an external constant force F, and see the spring stretch numerically, as we studied theoretically in Exercises 5.12, 6.16, and 6.17. Instead, let us consider adding a repulsive external quadratic potential $E(L) = -\frac{1}{2}\alpha L^2$ to the endpoint (Fig. **??**).[64] For simplicity, we shall measure energies in units of k_BT, or equivalently we set $k_BT = 1$.

Now, when we flip a segment, we increase or decrease the energy from $E(L)$ to $E(L \pm 2)$. It is natural to do this by equilibrating the two orientations of the segment, the *heat bath* algorithm.[65] Let us focus first on equilibrating a rightward-pointing segment. We want the segment directions after the step to have relative probabilities given by the Boltzmann distribution, which depends on $E(L) - E(L-2)$.

[64] This will be motivated in Exercise N1.31 as the interaction between spins in an infinite-range Ising model.
[65] As it happens, our method in part (a) implements the *Metropolis* algorithm.

(c) *What is the partition function Z for the two states of an initially rightward-pointing segment of a chain of length L? What is the probability that it will shift to point left?*
Our rubber band only has even lengths. Let L be an even integer, and $P^+(L)$ be the probability that a chain of length L will flip one of its leftward-pointing segments to make it shift to a length $L+2$. Similarly, let $P^-(L)$ be the probability per flip that L will shift to $L-2$.
(d) *Show that*

$$P^+(L) = \left(\frac{N-L}{2N}\right) \frac{1}{1+\exp\left(E(L+2)-E(L)\right)}$$
$$P^-(L) = \left(\frac{N+L}{2N}\right) \frac{1}{1+\exp\left(E(L-2)-E(L)\right)}. \tag{N1.41}$$

Show that, for no external force, the heat bath time step does nothing half the time. (The Metropolis algorithm of part (a) is more efficient, but less physical.)
(e) *Adapt your routine to* `flip(L,N,α)`*, that with probability $P^+(L)$ returns $L+2$, with probability $P^-(L)$ returns $L-2$, and otherwise returns L. Check it by running with $\alpha = 0$. Explore different values of α. At what value α_c does the external repulsion balance the entropic spring force? Does the behavior change qualitatively as you go above α_c?*

N1.30 **Rubber band dynamics II: Diffusion.**[66] ③

Exercise N1.29 studied the dynamic fluctuations of an entropic model for a rubber band: N segments of length $d = 1$, fluctuating between pointing forward and backward at random. It studied the random walk of lengths L as the segments hopped, both without and with an external parabolic potential stretching the band (Fig. **??**). Here we shall derive a spatially dependent diffusion equation describing the evolution of the probability distribution of lengths with time, in the limit of large N.

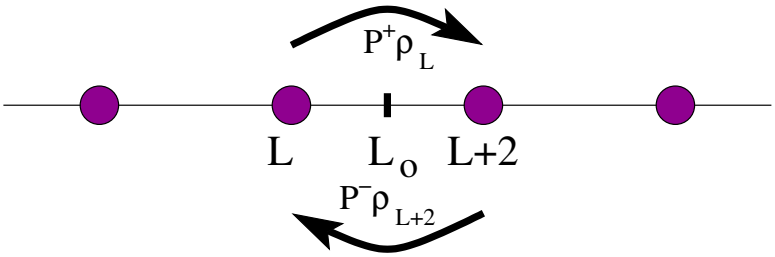


Fig. N1.37 Current is flow forward minus flow backward. The current past the midpoint $L_o = L+1$ between two possible lengths L and $L+2$ is given by the probabilities $\rho(L)$ and $\rho(L+2)$ times the probabilities $P^{\pm}$ of flipping forward and backward. (For this exercise, we shall assume N is even.)

Since the sum of the probabilities[67] of being at length L, $\sum_{L=-N}^{N} p_L = 1$, is constant,

[66] This exercise was developed in collaboration with Stephen Thornton
[67] Note that $p_L = 0$ for odd integers L

and our dynamics only shifts L locally (by $\pm 2 \ll N$), we are advised to write our dynamics in terms of the probability current. Let $J(L_o)$, for odd L_o (midway between possible lengths of the chain) be the net current from $L-1$ to $L+1$ per segment flip (Fig. **??**). In Exercise N1.29, eqn **??** we gave the probability $P^+(L)$ per flip that a chain of length L will grow to $L+2$ (contributing to J at $L_o = L+1$), and $P^-(L)$ that a chain of length L will shrink to $L-2$. (contributing to J at $L_o = L-1$).

(a) *Our rubber band ensemble at time t has probability $p(L)$ of having length L. Argue that the probability current $J(L_o)$ of our rubber band ensemble growing past the (odd) length L_o is*

$$J(L_o)\Delta t = p(L_o-1)P^+(L_o-1) - p(L_o+1)P^-(L_o+1) \tag{N1.42}$$

where Δt is the time for one segment flip.

Here by convention we set $\Delta t = 1/N$, so a sweep that flips N segments takes one unit of time.

In taking the continuum limit as $N \to \infty$ (Fig. **??**, let us keep the total unfolded length fixed. To do so, we use $x = L/N$. Also, the harmonic stretching force $F = \alpha L = \alpha N x$, so we change variables to $a = N\alpha$. Finally, the probability $p(L)$ represents the probability density $\rho(x)$ between $L-1$ and $L+1$ (i.e., $x - 1/N$ and $x + 1/N$), so we substitute $\rho(x)(2/N)$ for $p(L)$.

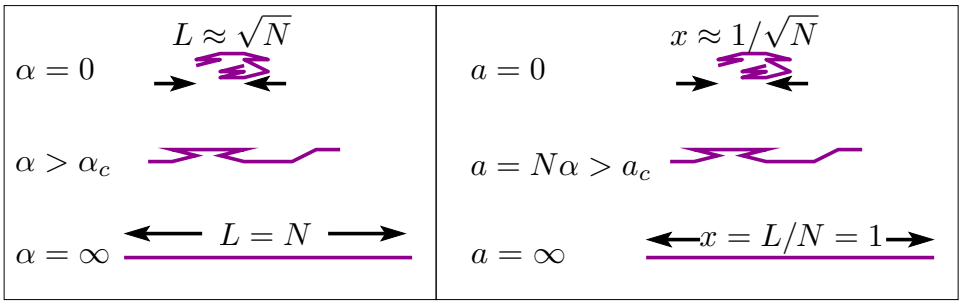


Fig. N1.38 Changing to continuum variables. In going from the microscopic description to the continuum limit, we change all lengths by a factor of N, we change the negative "spring" constant α to $a = N\alpha$, and we change from probabilities $p(L)$ to probability densities $\rho(x) = p(L)N/2$.

(b) *Substituting $L = Nx$, $\alpha = a/N$, and $P^\pm$ from eqn **??** into eqn **??**, show that*

$$J(x) = \frac{\mathrm{e}^{ax}(1 - x + 1/N)\rho(x - 1/N) - \mathrm{e}^{-ax}(1 + x + 1/N)\rho(x + 1/N)}{\mathrm{e}^{ax} + \mathrm{e}^{-ax}} \tag{N1.43}$$

(c) *What is the net current J in the limit $N \to \infty$ in eqn **??** (holding a and x constant)? Use the fact that $(\mathrm{e}^{ax} - \mathrm{e}^{-ax})/(\mathrm{e}^{ax} + \mathrm{e}^{-ax}) = \tanh(ax)$ to remove the exponentials from*

your answer, and show your work. Argue that the end of the rubber band as $N \to \infty$ has a position-dependent velocity

$$v(x) = \tanh(ax) - x. \tag{N1.44}$$

In Exercise N1.29(e), you found that for large a the length quickly moved to a final off-center position. *Find a numerical solution for this final length at $a = 0.75$ and $a = 1.5$ for $N \to \infty$. Derive from $v(x)$ the critical value a_c when the rubber band equilibrium length splits away from the origin.*

(d) *Using your random-walk simulation from Exercise N1.29, make a histogram of lengths explored by one random walk for $N = 100$ at $a = 0.75$ and $N = 1000$ at $a = 1.5$, adding points until you get good histograms. When necessary, drop the transient first part of the trajectory, while the rubber band moves from zero to the new minimum. Are your histograms concentrated near the predicted value you found in part (c)?*

(e) *Starting at $x = 0$, launch a trajectory for $N = 1000$ and $a = 1.5$, and examine how it flows to its final value once it deviates from the local fixed point at zero (say, in the first 20000 segment flips). Compare the flow to that predicted by your equation for the velocity as a function of length in part (c). (That is, numerically solve $\mathrm{d}x/\mathrm{d}t = v(x)$ from part (c), starting from a small positive or negative value of x, and rescale it back from x and a, to L and α.)* The time spent in the vicinity of $L = 0$ in the random walk will depend on the fluctuations. *Adjust the theory curve right and left to make a good comparison.*

The current in the limit $N \to \infty$ is the answer that the continuum limit supplies. Thermodynamics and other continuum theories often ignore the fluctuations in the system. We can study the statistical mechanics of the fluctuations by studying the leading corrections in $1/N$.

(f) *Now find the first correction in $1/N$ to the net current. Write the current to this order in the form*

$$J(x) = (v(x) + v_1(x)/N)\rho(x) - D(x)\partial\rho/\partial x. \tag{N1.45}$$

Calculate $v_1(x)$. Show that

$$D(x) = (1 - x\tanh(ax))/N. \tag{N1.46}$$

The correction $v_1(x)/N$ make tiny corrections to the fixed point of part (d) and the velocity curve of part (e) above – unimportant for large N. But the term proportional to $\partial\rho/\partial x$ (the typical domain of statistical mechanics) dominates many of the properties.

(g) *Write the emergent forced diffusion equation governing our entropic rubber band. (To simplify things, leave it in terms of v, D, ρ, and their derivatives. Ignore the terms involving v_1.)*

(h) *Approximate your differential equation to linear order x about zero (again ignoring v_1), and solve your diffusion equation for the stationary distribution at $a = 0.75$.* (Hint: Solving for the distribution that makes the current equal to zero is easier. Try a Gaussian.) *Compare to a histogram of equilibrated values of your random walk simulation.*

N1.31 **Rubber band dynamics III: Free energy and statics.**[68] ③

Statistical mechanics is a complete theory for the static properties of Hamiltonian systems: the probability of a snapshot of the system having any particular configuration. It constrains the dynamics of the system (entropy cannot decrease, ...) but different microscopic physics or simulation methods can change how a system evolves in time.

Here we study the statics and two kinds of dynamics in the entropic rubber band model, introduced in Exercise 5.12 in the microcanonical ensemble, and analyzed in Exercise 6.16 in the fixed-force ensemble. In Exercise N1.29 we added a parabolic potential energy to the model, and found a transition between a state with one equilibrium length at zero and a state with two stable equilibrium lengths.

We start by analyzing the static properties of the rubber band model in an ensemble fixing the external force on the random chain, and with the external parabolic potential.

We reformulate our model in terms of segment orientations $\mathbf{s}$. Each of the N segments of the rubber band has length one and can point in one of two directions $s_i = \pm 1$, with the rubber band length $L = \sum_{i=1}^{N} s_i$. F is the external force on the tip of the rubber band, and the external potential is $\frac{1}{2}\alpha L^2$, so

$$\begin{aligned}
\mathcal{H}(L) &= -\tfrac{1}{2}\alpha L^2 - FL \\
\mathcal{H}(\mathbf{s}) &= -\tfrac{1}{2}\alpha \left(\sum_{i=1}^{N} s_i \right)^2 - F \sum_{i=1}^{N} s_i \\
&= -\tfrac{1}{2}\alpha \left(\sum_{i=1}^{N} s_i \right) \left(\sum_{j=1}^{N} s_j \right) - F \sum_{i=1}^{N} s_i \\
&= -\alpha \sum_{i=1}^{N} \sum_{j=i+1}^{N} s_i s_j - F \sum_{i=1}^{N} s_i - \tfrac{1}{2}\alpha N,
\end{aligned} \tag{N1.47}$$

where $\mathbf{s} = \{s_1, \ldots, s_{2^N}\}$ runs over all 2^N possible segment orientations. Here the last formula for $\mathcal{H}$ connects our rubber band problem to the well-studied infinite-range Ising model with $J/N = \alpha$ and $F = H$.

Our ensemble fixes the force F and the coupling α. The partition function sums the Boltzmann weight over all possible segment orientation patterns,

$$\begin{aligned}
Z(F, \alpha) &= \sum_{\mathbf{s}} \mathrm{e}^{-\mathcal{H}(\mathbf{s})} \\
&= \sum_{L} Z_{F,\alpha}(L),
\end{aligned} \tag{N1.48}$$

[68] This exercise was developed in collaboration with Stephen Thornton

where we measure energies in units of k_BT and entropy in *nats* ($k_B = 1$).[69] Thus we shall change F and α up and down rather than changing the temperature T down and up to explore the behavior.

What is this last decomposition into $Z(L)$? Since $\mathcal{H}(\mathbf{s})$ depends on the spins only through their sum, we can count the number of segment configurations $\Omega(L) = \binom{N}{(L+N)/2}$ and weigh them by $\exp(-\mathcal{H}(L))$:

$$Z(L) = \Omega(L)\exp(-\mathcal{H}(L)) = \exp(-(\mathcal{H}(L) - S(L))). \tag{N1.49}$$

where $S(L) = \log(\Omega(L))$ is the microcanonical entropy we studied in Exercise 5.12. Instead of using Stirling's formula to approximate the entropy, we will study the exact $Z(L)$ and $\mathcal{F}(L)$ numerically.

The separation $Z = \sum Z(L)$ allows us to find the entire probability distribution of lengths at fixed force. Just as we studied the free energy density for the ideal gas in Section 6.7, we can use $Z(L)$ to define a free energy density $\mathcal{F}(L)$ for the rubber band at fixed force and coupling.

(a) *As in Exercise 6.17, give the formula for $\mathcal{F}(L) = -\log Z(L)$ in terms of L, α, F, and $S(L)$. Write the probability $p(L) = \rho(L)\Delta L = 2\rho(L)$ of the equilibrium rubber band being of length L, in three ways. First, write it in terms of $Z(L)$ and Z. Then write it in terms of $\mathcal{F}(L)$ and Z. And finally, write it in terms of the Boltzmann-like weights $\exp(-\mathcal{F}(L'))$, for all the different lengths L'.*

(b) *Plot $\mathcal{F}(L)$ and $\rho(L)$ for $F = 0$, $N = 100$, and with $\alpha N = a = 0, 0.25, \ldots, 1.5$. (Remember that $\rho(L) = p(L)/\Delta L = p(L)/2$.) Check that the free energy at $\alpha = 0$ and small L agrees well with that given by the spring constant K predicted in Exercise 5.12 as $N \to \infty$. Confirm that the critical value α_c at which $\rho(L)$ splits away from the origin is close to K.*

Note that the free energy near the transition is quantitatively similar to that of the quartic potential $f_0 + \frac{1}{2}aL^2 + gL^4$ as $a(T)$ passes through zero. Exercise 9.5 discusses Landau's approach to the Ising phase transition using this quartic polynomial (eqn 9.18). He posits a quartic free energy density as a function of magnetization at fixed temperature and external field. See also Exercises 12.5 and 12.26 for other mean-field approaches to the Ising model.

(c) *Plot $\mathcal{F}(L)$ and $\rho(L)$ for $a = 1.25$, $N = 100$, and with a few interesting values for the force F.* (Notice that for small values of F there are two stable minima. We call the higher energy minimum metastable.) *At what value F_c does the metastable minimum become unstable?* (A rough answer for F_c is fine. But if you want a precise answer, calculate the spring force $f(L)$ needed for part (d), and see where it last crosses zero as the local minimum of $\mathcal{F}(L, F)$ disappears at F_c.)

In Exercise N1.30, we extracted a prediction for the evolution law of the length from heat bath dynamics. But this is not the only choice. In later chapters, we shall

[69] Hence, $Z(L) = \mathrm{e}^{-(\mathcal{H}(L)-TS(L))/k_BT} = \mathrm{e}^{-\mathcal{H}(L)/\cancel{k_BT}+S(L)/\cancel{k_B}} = \mathrm{e}^{-(\mathcal{H}(L)-S(L))}$.

often assume *gradient* dynamics: that the velocity is a mobility γ times minus the (variational) derivative of the free energy with respect to the "order parameter" (in this case L, see also Section 2.3). Gradient dynamics says that the tip of the rubber band evolves with the law

$$\frac{\mathrm{d}L}{\mathrm{d}t} = v_{\text{gradient}} = \gamma f(L) = -\gamma \frac{\mathrm{d}\mathcal{F}}{\mathrm{d}L}. \tag{N1.50}$$

Here the force $f(L)$ is the force exerted by the spring when it is *not* in its equilibrium position. It is partly due to the external force $\alpha L + F$ and partly due to the entropic spring force.

Let us consider the case where there is no force F from the external world.

(d) *Numerically compute the force $f(L)$ (either by finite differences or by symbolic differentiation) for $F = 0$, $\alpha = 1.25/N$ and $N = 1000$. Does it go to zero at the equilibrium lengths? Compare it to the velocity of the tip of the rubber band given by heat-bath dynamics, $v_{\text{HB}}(L) = N \tanh(\alpha L) - L$, derived in Exercise N1.30, by plotting $\gamma f(L)$ and $v_{\text{HB}}(L)$ on the same graph. Can you find a constant mobility γ that makes these two agree everywhere? (γ is proportional to N.) Can you find a constant mobility that that allows them to agree near the positive and negative equilibrium lengths?* (Focus on matching slopes; a rough estimate is fine. The fixed point shifts quite a bit between $N = 100$ for v_{gradient} and $N = \infty$ for v_{HB}; we reduce this when feasible by using $N = 1000$.)

So, we can match gradient to heat-bath dynamics locally near equilibrium by a suitable choice of the mobility. This is reassuring. But they disagree in general! Is one or the other wrong? Or are they both consistent, possible dynamics that yield the same equilibrium behavior?

The heat-bath algorithm is not an accurate representation of real rubber bands! Had we written a diffusion equation for the (efficient, but somewhat unphysical) Metropolis algorithm (Exercise 8.6), or the (grossly unphysical) Wolff algorithm (Exercise 8.8),[70] we would have yet a different (rather strange) prediction for the velocity.

What do we need to check to see if gradient dynamics and heat-bath dynamics are both OK? Let us add fluctuations to answer this question. Again, we can compare two stochastic dynamics.

The tradition in the field is to extend gradient dynamics to Langevin dynamics by adding noise. They assume a constant γ, and white noise corresponding to a fixed diffusion constant D (see Exercises 6.18, 6.19, and 10.7). By fixing $D/\gamma = k_B T$, they guarantee that the ensemble generated at late times is the equilibrium thermal ensemble given by the Boltzmann distribution.

Feynman [62], at the end of vol. I, sec. 43.5, derives the Einstein relation $D/\gamma = k_B T$. He notes that the current from diffusion must cancel the current from the force

[70] The Wolff algorithm probably cannot be used for our infinite-range Ising model. And, unlike the heat bath algorithm, they likely will give strange, unphysical dynamics even for small perturbations.

due to the free energy in order for the system to be in equilibrium. He then uses the fact that the equilibrium density is given by the Boltzmann distribution. Let us consider a general free energy $\mathcal{G}(x)$ with equilibrium probability density $\rho(x) = \exp(-\mathcal{G}(x)/k_BT)/Z$, diffusion current $-D\rho'(x)$, and force-driven current $\gamma f(x)\rho(x)$.

(e) *Derive the Einstein relation $D/\gamma = k_BT$ by balancing the currents and using the equilibrium probability density.* (It should be easier to do this on the fly than to look up Feynman's argument, but his discussion is worth reading.)

So, do both gradient dynamics and heat-bath dynamics pass the Einstein relation test? Since Langevin dynamics uses the Einstein relation to set the noise from the damping, it certainly passes. But what about heat-bath dynamics?

In Exercise N1.30, in addition to finding $v_{\mathrm{HB}}(L)$, we used the microscopic heat-bath dynamics to derive the spatially dependent diffusion constant $D_{\mathrm{HB}}(L) = N - L\tanh(\alpha L)$. The Einstein relation then implies a spatially dependent mobility $\gamma_{\mathrm{HB}}(L)$.

(f) *Check numerically if $\gamma_{\mathrm{HB}}(L)$ does yield the heat-bath $v_{\mathrm{HB}}(L)$, by plotting the latter along with $\gamma_{\mathrm{HB}}(L)f(L)$ for $\alpha = 1.25/N$ and $N = 100$ (or 1000 if feasible). Is our heat-bath diffusion equation consistent with free energies and the Einstein relation?* (Remember, for us $k_BT = 1$.) *Discuss.*

Chapter 2

Advanced Statistical Mechanics

These exercises cover topics in more advanced statistical mechanics and the renormalization group, algorithms, and computer science.

N2.1 Singular corrections to scaling and the RG. ③

In this exercise, we derive the form of the scaling function (eqn 12.102) for the effects of *irrelevant* operators on the properties of systems near critical points (see Exercise 12.31). Remember that irrelevant directions shrink under coarse-graining. Let χ be the susceptibility of the Ising model, as a function of the reduced temperature $t = T_c - T$ and some irrelevant operator u:

$$\begin{aligned} \mathrm{d}\chi_\ell/\mathrm{d}\ell &= -(\gamma/\nu)\chi_\ell, \\ \mathrm{d}t_\ell/\mathrm{d}\ell &= t_\ell/\nu, \\ \mathrm{d}u_\ell/\mathrm{d}\ell &= -y u_\ell. \end{aligned} \tag{N2.1}$$

How do we derive the universal scaling function $X(z)$ from these renormalization-group flows? Consider the flows illustrated in Fig. 12.8, except now with a third dimension involving the prediction χ. Consider a point (t_0, u_0, χ_0) in the system space, and the invariant curve defined by $z = u_0 t_0^\omega$ (dashed lines). Our renormalization group allows us to calculate $\chi_\ell(t_\ell, u_\ell)$ along these curves—relating the behavior everywhere near the critical manifold (vertical swath flowing toward S^*) to the properties along the outgoing trajectories, which approach closer and closer to the unstable manifold (the horizontal swath flowing away from S^*).

For example, we can define the universal scaling function $X(z)$ (for positive time t) to be the χ_{ℓ^*} where the flow crosses $t_{\ell^*} = 1$.

(a) *Solve eqns N2.1 for u_ℓ and t_ℓ. Setting $t_{\ell^*} = 1$, what is u_{ℓ^*} in terms of your invariant combination z?*

So we label each invariant scaling curve by the value of the vertical position u_{ℓ^*} where it crosses $t_{\ell^*} = 1$.

(b) *Solve eqns N2.1 for $\chi_{\ell^*}(1, u_{\ell^*})$, in terms of z, t_0, and $\chi_0(t_0, u_0)$. Use your solution to solve for the physical behavior $\chi_0(t_0, u_0)$ in terms of t and $X(z)$. Express $X(z)$ in terms of $\chi_{\ell^*}(1, u_{\ell^*})$.* Does your answer agree with the form in eqn 12.102?

N2.2 **Nonlinear RG flows and analytic corrections.**[1] ③

We consider the effects of nonlinear terms in renormalization-group flows. The Ising model in zero field has one relevant variable (the deviation t of the temperature from T_c). To calculate the specific heat, we shall also consider the flow of the free energy per spin f under the renormalization group. Instead of a discrete coarse-graining by a factor b, here we use a continuous coarse-graining measured by ℓ. One can think of one coarse-graining step by $b = (1 + \epsilon)$ as incrementing $\ell \to \ell + \epsilon$; equivalently, coarse-graining to ℓ changes length scales by $\exp(\ell)$.

Consider the particular flow equations[2]

$$\begin{aligned} \mathrm{d}f_\ell/\mathrm{d}\ell &= Df_\ell - at_\ell{}^2 \\ \mathrm{d}t_\ell/\mathrm{d}\ell &= t_\ell/\nu, \end{aligned} \tag{N2.2}$$

where D is the dimension of space and $at_\ell{}^2$ is a nonlinear term that will be important in two dimensions.

We shall call the free energy per spin of the actual system f_0, and the temperature of the actual system t_0. The coarse-grained free energy and temperature are f_ℓ and t_ℓ, after being coarse-grained by a factor $\exp(\ell)$. (Hence at $\ell = 0$ we have not yet coarse-grained, so $f_0 = f$ and $t_0 = t$.) Notice here that the free energy of our system is the *initial condition* $f_0(t_0)$ of this differential equation, and $f_\ell = f(\ell)$ and $t_\ell = t(\ell)$ are the renormalization-group flows of the two variables. To derive the scaling behavior, we shall coarse-grain to ℓ^* where $t_{\ell^*} = 1$, at which point the coarse-grained free energy is f_{ℓ^*}.

Let us start with the linear case $a = 0$.

(a) *Solve for f_ℓ and t_ℓ for $a = 0$. Setting $t_{\ell^*} = 1$, solve for f_0 in terms of f_{ℓ^*}, t_0, D, and ν. Solve for the specific heat per spin $c = T\partial^2 f_0/\partial T^2$, where $t_0 = (T - T_c)/T_c$. Show that the specific heat near T_c has a power-law singularity $c \propto t^{-\alpha}$, with $\alpha = 2 - D\nu$.* (For example, in $D = 3$, $\nu \approx 0.63$, so $\alpha = 2 - D\nu \sim 0.11$; the specific heat diverges at T_c.) *Writing*

$$c = t^{-\alpha}(c_0 + c_1 t + c_2 t^2 + \dots), \tag{N2.3}$$

what is the first correction c_1 to the specific heat near $t = 0$ in the absence of the nonlinear term?

Why is the linear term in the free-energy flow equal to the dimension, $\mathrm{d}f/\mathrm{d}\ell = Df + \dots$, where all other terms are hard-to-compute critical exponents? There is no completely

[1] This exercise was developed in collaboration with Colin Clement

[2] Note that these are total derivatives. So the first equation tells us the total change in f after coarsening by a factor $1 + \mathrm{d}\ell$. $f(t)$ then will coarse-grain to $f_\ell(t_\ell)$ without needing to worry about the chain rule $\mathrm{d}f(t)/\mathrm{d}\ell = \partial f/\partial\ell + \partial f/\partial t \mathrm{d}t/\mathrm{d}\ell$.

general answer to this question (although there are arguments for specific models). Indeed, in other models of disordered systems and glasses, and models above the upper critical dimension, the linear term is not given by the dimension. The relation $2 - \alpha = D\nu$ is called a *hyperscaling* relation (to emphasize they involve the dimension D), and these other models are said to violate hyperscaling. One physical description of hyperscaling involves the singular part of the free energy f inside a correlated region.

(b) *In the case $a = 0$, show that the singular free energy f contained in a correlated volume ξ^D near the critical temperature becomes independent of the distance to the critical point.*

Glassy and disordered systems become extremely sluggish as they are cooled. In at least some cases, this is precisely because the energy barriers needed to continue equilibration diverge as their correlation lengths grow—they are glassy because their RG flows violate the hyperscaling relation.

So much for the power law singularity—what about the correction term c_1 in part (a)? It is an *analytic correction to scaling*.[3] Here it is *subdominant*—near the critical temperature where $t \to 0$, it is less singular than the leading term. One expects in a real physical system that the microscopic bond free energy J between spins will be some analytic function of temperature, and the physical free energy and specific heat will be multiplied by J. Expanding J in a Taylor series about $t = 0$ would also give us terms like those in eqn N2.3.

Does the introduction of the higher-order nonlinear term $at_\ell{}^2$ in eqn N2.2 change the behavior in an important way? Rather than exercising your expertise in analytic solutions of nonlinear differential equations, eqn N2.4 provides not f_ℓ and t_ℓ as functions of ℓ, but the relation between the two:

$$f_\ell(t_\ell) = f_0 \left(\frac{t_\ell}{t_0}\right)^{D\nu} - \frac{a\nu t_\ell{}^2}{(2 - D\nu)}(1 - (t_\ell/t_0)^{-(2-D\nu)}). \tag{N2.4}$$

(c) *Show that $f_\ell(t_\ell)$ in eqn N2.4 satisfies the differential equation given by eqn N2.2, using*

$$\frac{\mathrm{d}f_\ell}{\mathrm{d}t_\ell} = \frac{\mathrm{d}f_\ell}{\mathrm{d}\ell} \Big/ \frac{\mathrm{d}t_\ell}{\mathrm{d}\ell} \,. \tag{N2.5}$$

Show that it has the correct initial conditions at $\ell = 0$. What is f_{ℓ^*} at ℓ^*, where $t_\ell = 1$? Show your method.

Again, it is important to remember that $f_\ell(t_\ell)$ is not the free energy as a function of temperature—it is the coarse-grained free energy as a function of the coarse-grained temperature of a system starting at a free energy f_0 at a temperature t_0. It is $f_0(t_0)$

[3] These are distinct from *singular* corrections to scaling that arise from irrelevant terms under the renormalization group (Exercise 12.31).

that we want to know. Since here we have only one relevant variable (in zero field), all the flows lead to the same[4] final point $f_{\ell^*}(t_{\ell^*} = 1)$

(d) *Solve for f_0 in terms of f_{ℓ^*} and t_0. Solve for the specific heat $c = T\partial^2 f/\partial T^2$, where $t = (T - T_c)/T_c$. Show that it can be written in the form*

$$c = c_{+\text{analytic}}(t) + t^{-\alpha} c_{*\text{analytic}}(t) \tag{N2.6}$$

where the additive correction $c_{+\text{analytic}}(t)$ and the multiplicative correction $c_{*\text{analytic}}(t)$ have a simple Taylor series about $t = 0$. Write these two corrections, in terms of f_{ℓ^*}, T_c, a, ν, and D.

Here the nonlinear term a gives us both a smooth multiplicative term and a smooth additive background to the specific heat. This makes sense physically. An additive term would come from other degrees of freedom like atomic vibrations, or a box holding the magnet. A multiplicative term would arise, for example, from an effective temperature dependence of the coupling between spins (say, due to thermal expansion separating the atoms).

N2.3 Beyond power laws: 2D Ising logs.[5] ③

The two-dimensional Ising model has a logarithmic singularity in the specific heat. The exact result shows that the specific heat per spin is [20, eqn 3.119]

$$\begin{aligned} c(T) =& k_B \frac{2}{\pi} \left(\frac{2J}{k_B T_c}\right)^2 \big[-\log(1 - T/T_c) \\ & + \log(k_B T_c/(2J)) - (1 + \pi/4)\big] \\ =& -\frac{8}{\pi k_B T_c^2} \log\left(\frac{t}{\frac{1}{2} k_B T_c \exp(-(1+\pi/4))}\right) \\ =& -c_0 \log\left(\frac{t}{\tau}\right), \end{aligned} \tag{N2.7}$$

where $t = (T - T_c)/T_c$ and we set $J = 1$. (Remember $\log(t)$ is negative for small t.) Linearizing the flows around the renormalization-group fixed point predicts a power law $c \sim t^{-\alpha}$. When $\alpha \to 0$ one often observes logarithmic corrections, as we see in eqn N2.7. But such corrections are not predicted by the RG flows after linearizing! The key nonlinear term is the term at_ℓ^2 of eqn N2.2 we studied for 3D in Exercise N2.2.

(a) *Is the solution for $f_\ell(t_\ell)$ in eqn N2.4 useful in $D = 2$? Why or why not?* (Hint: The exponent $\nu = 1$ for the two-dimensional Ising model.)

Again, we provide the solution of the nonlinear RG eqns N2.2, now for $D = 2$

$$f_\ell(t_\ell) = f_0(t_\ell/t_0)^2 - a t_\ell^2 \log(t_\ell/t_0). \tag{N2.8}$$

[4] Remember for systems with more than one variable (say t and h), the free energy depends on the invariant curve departing from the fixed point, labeled, say, by $h/t^{\beta\delta} = h_{\ell^*}(t_{\ell^*} = 1)$. We solve for $f_0(t_0, h_0)$ in terms of $f_{\ell^*}(1, h_{\ell^*})$ just as we do in part (b).

[5] This exercise was developed in collaboration with Colin Clement.

(b) *Show that $f_\ell(t_\ell)$ in eqn N2.8 satisfies the differential equation given by eqn N2.2, with the correct initial conditions. Solve for f_0 in terms of f_{ℓ^*} and t_0, where $t_{\ell^*} = 1$. Solve for the specific heat $c = T\partial^2 f/\partial T^2$, where $t = t_0 = (T - T_c)/T_c$ and $f = f_0$. (Remember the chain rule: $\partial f/\partial T = (\partial f/\partial t)(\mathrm{d}t/\mathrm{d}T)$.) Does it agree asymptotically with the exact result in eqn N2.7? What are c_0 and τ, in terms of a and f_{ℓ^*}?*

Thus for the 3D Ising model (Exercise N2.2), nonlinear terms in the renormalization-group flow equations give only analytic corrections to scaling, where in the 2D Ising model they introduce logarithms in the specific heat. Normal form theory (see Exercise 12.4) can be used to determine when one may safely linearize these equations, and to organize the other critical points into universality *families* [28].

N2.4 **Eigenvectors near the renormalization-group fixed point.** ③

The critical exponents in the renormalization group are given by the eigenvalues of the RG transformation linearized near the fixed point. What do the eigenvectors mean?

Consider a two-dimensional Ising model with two parameters, a nearest-neighbor bond[6] $K = J/T$ and a next-neighbor interaction $K_2 = J_2/T$ lying along the diagonal bonds.

$$\begin{aligned}\mathcal{H} = &-K \sum_{i,j} S_{i,j}S_{i+1,j} + S_{i,j}S_{i,j+1} \\ &-K_2 \sum_{i,j} S_{i,j}S_{i+1,j+1} + S_{i,j}S_{i+1,j-1}\end{aligned} \tag{N2.9}$$

If we decimate to the black squares of a checkerboard (say, $i + j$ even), we get a new square-lattice Hamiltonian rotated by 45° coarse-grained by a factor $b = \sqrt{2}$. The next-neighbor bond basically becomes a nearest-neighbor bond—it mostly renormalizes to zero in one step, and contributes its value to the new nearest-neighbor coupling. The deviation of the nearest-neighbor bond from the critical point K^*, we may crudely assume, rescales by a factor $b^{1/\nu}$ under coarse-graining (remember $K \sim J/T$) and then is increased by K_2. So under one coarse-graining step

$$\begin{aligned}K' - K^* &= b^{1/\nu}(K - K^*) + K_2, \\ K_2' &= 0.\end{aligned} \tag{N2.10}$$

(a) *Our crude renormalization-group flow is already linear. What is the fixed point? What is the Jacobian J about the fixed point? What are the eigenvalues λ_0 and λ_1?* (Let λ_1 be the relevant eigenvalue, greater than one.)

Our Jacobian matrix is not symmetric (or Hermitian), so it has two sets of eigenvectors—left eigenvectors $\hat{\boldsymbol{\ell}}_\alpha J = \lambda_\alpha \hat{\boldsymbol{\ell}}_\alpha$, and right eigenvectors $J\hat{\mathbf{r}}_\alpha = \lambda_\alpha \hat{\mathbf{r}}_\alpha$.

(b) *What are the left and right eigenvectors? Are the left eigenvectors orthonormal? Are they normal to the right eigenvector that has a different eigenvalue?*

[6] Instead of thinking of a Hamiltonian space with temperature as an extra parameter, it is convenient to work at fixed temperature, and mimic raising temperature by lowering the overall scale of the energy.

(c) *Draw the flow in the (K, K_2) plane near the fixed point. Indicate the directions of the left eigenvectors and right eigenvectors in different colors. Also draw the boundary between the ferromagnetic and paramagnetic phase. How is this boundary related to the stable manifold of the fixed point? Is it related to any of the eigenvectors?*

Consider a new set of *scaling variables* u_α, given by the dot products of the displacement from the fixed point with the left eigenvectors:

$$u_\alpha = \boldsymbol{\ell}_\alpha \cdot (K - K^*, K_2) \tag{N2.11}$$

(d) *Show that the phase boundary has $u_1 = 0$ (using the convention that λ_1 is the relevant direction). How do the coordinates u_α flow under the renormalization group?*

In general, there is a nonlinear transformation between the parameters $T, H, J_2, \ldots$ in a Hamiltonian and the natural coordinates $t(T, H, J_2), h(T, H, J_2), u(T, H, J_2)$ which flow simply under the renormalization group. This coordinate change is one of the contributors to analytic corrections to scaling.

(e) *Are u_0 and u_1 relevant, irrelevant, or marginal? Which coordinate, u_0 or u_1, is the scaling variable corresponding to the reduced temperature $t(K, K_2)$? If we write a property of our system $X(K, K_2) = X(K(u_0, u_1), K_2(u_0, u_1)) = u_1^x \mathcal{X}(u_0/u_1^y)$, can there be any dependence on u_0, within our crude model? How does X vary near the phase boundary?*

N2.5 **Is the fixed point unique? Period doubling.**[7] (Dynamical systems) ③

Is the fixed point of the renormalization group unique? (It seems unlikely that coarse-graining the Ising model in momentum space gives the same fixed point Hamiltonian as real-space decimation on a square lattice. One would be spherically symmetric, the other has a square symmetry. Naturally, both look the same on long length and time scales, but their short-distance behavior is different.) If not, and there are many alternative fixed points in system space describing a phase transition, can any system at the critical point be a fixed point, for a suitable renormalization group?

We shall answer this question for the particular case of the period-doubling onset of chaos. In particular, we shall investigate what happens to the renormalization-group fixed point as we change coordinates. There is no reason to expect that Nature measures distances x in the same way, though, as we do. We could equally well decimate and rescale in a different coordinate $y = \phi(x)$, where we assume ϕ is smooth, monotone increasing, has a smooth inverse ϕ^{-1}. We will be considering one-humped maps with a maximum at $x = 0$, so we shall assume $\phi(0) = 0$ to keep the maximum at zero.

(a) *Give the formula for the function $\widetilde{g}(y)$ corresponding to the map $g(x)$, in terms of g and ϕ.* Hint: You need to find the x-value from y, then apply g to get the new x, and then find the new y. *Show that the inverse formula is for g in terms of ϕ and $\widetilde{g}$ is*

$$g(x) = \phi^{-1} \widetilde{g}(\phi(x)). \tag{N2.12}$$

[7] There are many exercises exploring this chaotic logistic map (see Index).

If $g(x)$ is at the onset of chaos (showing universal scaling behavior), will $\widetilde{g}(y)$ also be at the onset?

Recall from Exercises 12.9 and 12.29 that the transition to chaotic motion (Fig. 12.17) of one-humped maps $g(x)$ is understood using a renormalization group that decimates in time by a factor of two using $g(g(x))$, and rescaling the coordinate x by a factor of $\alpha \approx -2.5$:

$$T[g](x) = \alpha g(g(\alpha^{-1}x)). \tag{N2.13}$$

Let g^* be the fixed point $T[g^*] = g^*$ of our renormalization group in the space of one-humped maps.

(b) *Write a renormalization-group tranformation $\widetilde{T}$ for which $\widetilde{g}^*$ is a fixed point.* (Hint: Use the formula $T[g^*] = g^*$, and the formula for g^* in terms of $\widetilde{g}^*$ and ϕ from part (a). Change variables to y and solve for $\widetilde{g}(y)$ on the right-hand side.) *Show that $\widetilde{T}$ can be viewed as a decimation plus a nonlinear "stretching" function $\widetilde{\alpha}(y)$, $\widetilde{T} = \widetilde{\alpha}\circ\widetilde{g}\circ\widetilde{g}\circ\widetilde{\alpha}^{-1}(y)$. What is $\widetilde{\alpha}(y)$?*

So there is an infinite family of plausible, nonlinear renormalization-group transformations, with an infinite family of fixed points given taking our original fixed point $g^*(x)$ and changing variables $x \to \phi(x)$.

Can we change variables so that our fixed point is a parabola (as in the logistic map)? Can we make any one-humped map at the onset of chaos a fixed point, by choosing an appropriate change of coordinates?

There is an elegant proof that this is not possible. Our one-humped map has a fixed point $g^*(x^*) = x^*$, and also many periodic orbits[8] $g^{*[2^n]}(x_n) = g^*(g^*(\ldots(x_n)\ldots)) = x_n$. with period $2, 4, 8, \ldots$ (Beware! g^* is a fixed point in function space under the transformation T. x^* is a fixed point on the real line under the function g.)

(c) *Show that* $\mathrm{d}x/\mathrm{d}y = \phi^{-1\prime}(\phi(x)) = 1/(\mathrm{d}y/\mathrm{d}x) = 1/\phi'(x)$. *If x^* is a fixed point of g^*, what is the fixed point y^* of $\widetilde{g}^*$? Find the slope $\mathrm{d}\widetilde{g}^*/\mathrm{d}y|_{y^*}$ at the new fixed point, and show that it equals the slope $g^{*\prime}(x^*)$ at the old fixed point. Similarly, show that* $\mathrm{d}g^{*[2^n]}/\mathrm{d}x|_{x_n} = \mathrm{d}\widetilde{g}^{[2^n]}/\mathrm{d}y|_{y_n}$. (Note that $\mathrm{d}g(g(\ldots(x)\ldots))/\mathrm{d}x = g'(g(\ldots))g'(\ldots)\ldots$, so the product of the derivatives along any period orbit is invariant under coordinate changes.)

So, unless the logistic map happens to have the same derivative as g^* at their respective fixed points, there can be no coordinate transform taking one to the other, and there is no reason to think there is some nonlinear renormalization-group transformation that has the logistic map as a fixed point. (Later we shall find a more direct way to see that the logistic map has different behavior than any fixed point.) Indeed, all the fixed points of all the iterates of a map would have to agree with the iterates of g^* to allow for such a transformation. So our critical surface in function space, where one-humped maps

[8] Remember the period doubling bifurctions, which start with a stable 2^n cycle and end with a stable 2^{n+1} cycle surrounding an unstable 2^n cycle, forming 2^n little pitchforks. All of these unstable 2^n cycles survive to the onset of chaos.

are poised at the onset of chaos, has two subsurfaces—the maps that can be formed under coordinate changes (and could be fixed points), and the maps that cannot be fixed points.

We can gain more insight into these two types of critical points by considering the renormalization-group flow of functions $\widetilde{g}^*$. (It is a fixed point of $\widetilde{T}$ in part (b), but it must flow to g^* under our original renormalization-group transformation T.) Consider maps near to g^*, formed by picking coordinate transformations ϕ that are near to the identity:

$$\phi(x) = x + \epsilon \sum_{p=1}^{\infty} \phi_p x^p. \tag{N2.14}$$

(d) *Write $\phi^{-1}(y)$ to linear order in ϵ, as a similar sum. Calculate the change $\widetilde{g}^*(x) - g^*(x)$ to linear order in ϵ.* (Hint: Start with the equation $\phi(g^*(x)) = \widetilde{g}^*(\phi(x))$.) *Express it as a sum $\epsilon \sum_{p=1}^{\infty} \phi_p \Psi_p(x)$.*

(e) *Show that $g^*(g^*(x/\alpha)) = g^*(x)/\alpha$. By differentiating $T[g^*][x]$, show that $g^{*\prime}(x) = g^{*\prime}(g(x/\alpha))g^{*\prime}(x/\alpha)$. Calculate the change $T[\widetilde{g}^*](x) - g^*(x)$ to linear order in ϵ. Use your first two formulas to show that the change can be written $\epsilon \sum_{p=1}^{\infty} \phi_p \alpha^{1-p} \Psi_p(x)$.*

You have shown that $\Psi_p(x)$ is an eigenfunction of T, with eigenvalue α^{1-p}. You have *also* shown that only these eigenvalues and eigenfunctions can be generated by infinitesimal changes of coordinates. Feigenbaum conjectured in his early paper [12, p.687] that these eigenfunctions were the only ones. Our argument in part (c) suggests that there must be functions near to g^* that cannot be reached by a change of coordinates.

The lowest few eigenvalues λ_n of T, calculated for perturbations in the even subspace, are approximately

$$\{\lambda_1, \lambda_2, \dots\} = \{4.67, 1.00, 0.160, -0.124, -0.0573, 0.0255, -0.0101, \dots\} \tag{N2.15}$$

For your convenience, we provide the powers of $\alpha \approx -2.503$, so

$$\alpha^{1-p} = \{1.00, -0.400, 0.160, -0.0638, 0.0255, -0.0102, 0.00407, \dots\} \tag{N2.16}$$

(f) *Is λ_4 a power of α? Was Feigenbaum correct? Give the next few eigenvalues λ_n that are* not *given by coordinate changes. Which powers p are missing from the list? Do those powers correspond to even, or odd perturbations to g^*?*

So, at least at the period-doubling onset of chaos, there are two kinds of critical points, those that can be reached by a smooth change of coordinates (and could be fixed points of a different renormalization group), and those that cannot. We propose a new nomenclature, based on the analogy with gauge transformations in electromagnetism and field theory. In electromagnetism, a gauge transformation can be viewed as a spatially varying change in the convention of how we measure (gauge) the phase of the wavefunction for charged particles: $\Psi(\mathbf{x}_1, \mathbf{x}_2, \dots) \to \prod_j \mathrm{e}^{\mathrm{i}\chi(\mathbf{x}_j)} \Psi$. Here smooth changes of coordinates acts to change the way we measure position x in our chaotic system; more

generally we might consider also changing the way we measure the control parameter μ, or (for an Ising model) changing the definitions of magnetization M, temperature T, external field H, *etc.* We shall call the first *gauge* critical points and the second (more common) category *singular* critical points. More generally, there are three sources of corrections to scaling. There are *singular corrections* to scaling, which are generated by irrelevant eigenvectors of the renormalization group that are new, irrational numbers not related to the other exponents (perhaps like λ_4 in part (f)). Smooth coordinate transformations in the control variables (here the parameter μ in the logistic map) generate what are traditionally termed analytic corrections to scaling. Finally, there are what are traditionally termed redundant corrections to scaling, which correspond to systems that could have been fixed points of a different renormalization group. We shall call both of these last two *gauge* corrections to scaling: they can give corrections with integer power laws (like the analytic background in the specific heat), or corrections that are combinations of integers and the relevant and gauge-irrelevant critical exponents.

How in practice can we tell the special "gauge" critical points from the "singular" ones?

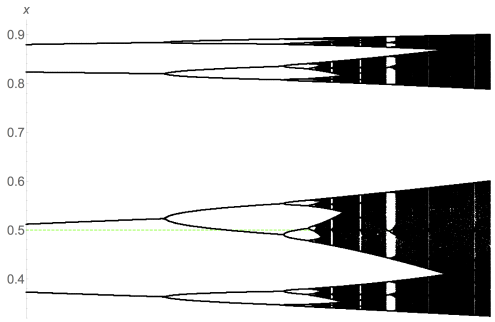


Fig. N2.1 Expanded bifurcation diagram. The attractor at the fixed point g^* has a perfect self-similarity upon flipping and rescaling about the function maximum $x = 0.5$ (green-dashed line). The lower first-tier branch of the bifurcation diagram maps onto the entire diagram when rescaled vertically by α and horizontally by δ; the same happens to the lower branch (whose upper second-tier branch maps onto the entire lower first-tier branch), and so on. Any other map will have this property only asymptotically, near the maximum (green dashed line) and near the critical point.

Suppose we study a map at its critical point, which deviates from g^* primarily along the eigendirection corresponding to $\lambda_4 = -0.124$. This deviation will lead to changes in the spatial patterns seen in the attractor (Fig. N2.1); for example, the ratio α_n of the widths of the n^{th} and $(n+1)^{st}$ tier branch will not be precisely α.

Let the $\psi_n(x)$ be the n^{th} eigenfunction of T.[9]

(g) *Consider an infinitesimal perturbation of the fixed point $g^* + \epsilon\psi_4(x)$ along this direction. How does the leading correction to $\alpha_n - \alpha$ scale with λ_n?* (Hint: How does

[9] We already have a name for the special eigenfunctions, so for example the eigenfunction associated with $\lambda_2 = 1 = \alpha^{1-p}$ with $p = 1$ is $\psi_2(x) = \Psi_1(x) = \psi_2(x)$).

α_n change under one application of the renormalization group operator T?)

Thus the special, gauge critical points have corrections to scaling that are expressable solely in terms of the relevant critical exponents.

Conjectures:

N2.1 The allowed renormalization-group fixed point functions are precisely those accessible by analytic coordinate transformations $y = \phi(x)$, which are precisely those generated by moving along the scaling variables u_n^0 associated with the eigenvalues λ_n not given by powers of α (in a way similar to how infinitesimal Lie algebra symmetries are related to Lie group operations). In particular, a large change of variables will have corrections to scaling at the critical point only involving powers of α.

N2.2 We conjecture that all corrections to scaling for one-humped maps [19] will be given by changing to *normal form* coordinates. So, for the logistic map $f(x) = \mu x(1-x)$, the traditional analysis suggests there are nonlinear scaling coordinates that satisfy $T(u_n) = \lambda u_n$; the curve in function space traced by the logistic map then has coordinates $u_\mu(\mu) = \mu - \mu_\infty + b_1(\mu - \mu_\infty)^2 + \ldots$, $u_2(\mu) = a_2 + b_2(\mu - \mu_\infty) + c_2 \ldots$, ... We are led to conjecture that a further change of coordinates $x \to \phi(x)$ allows us to remove all corrections involving the special eigendirections with $\lambda = \alpha^{1-p}$ for integer $p > 1$. That is, we conjecture that not only is there a gauge submanifold in the surface of critical points, but that the critical points can be *foliated* into surfaces where the singular corrections are the same.

N2.3 We conjecture that the separation between gauge and singular fixed points we define here have analogies in the Ising model and other critical points. (Indeed, we took the name from that literature, where it involves changes in the coordinates describing the fields.) See in particular reference [13], which answers the analogous question as to whether there is a renormalization group that can make any Ising critical point into its fixed point.

N2.4 We note that, in the case of thermodynamic critical points, the distinction between control variables (like $t = T - T_c$ and $h = H/T$) and "results" (like magnetization $M(t,h)$, entropy $S(t,h)$, and energy $E(t,h)$) are rather artificial; under a Legendre transformation we can take a system described by the Gibbs free energy $G(T,H)$ to a system described by the microcanonical entropy $S(E,H)$ or the Helmholtz free energy $A(T,M)$. We conjecture that Legendre transformations will convert the analytic corrections to gauge corrections in our nomenclature. If so, it may be better to treat these on an equal footing, and term them both analytic corrections to scaling.

N2.5 We conjecture that a subset of the irrelevant eigenvalues for other critical points will similarly be given in terms of combinations of relevant eigenvalues. We conjecture that analytic changes in the results variables can remove these corrections to scaling. (Again, Nature does not tell you how you should measure your magnetization, just as it did not tell you how to measure your applied field.)

N2.6 Many of the ideas here were prompted by an analysis of the onset of chaos from quasiperiodic motion [24, 29]. There the coordinate transformation taking the map to a simple rotation was numerically straightforward, central to understanding maps below the onset of chaos, and led to self-similar, nonanalytic maps at the onset of chaos. The coordinate transformations that connected two points on the critical manifold were not analytic, but numerically had a continuous first derivative. We conjecture that the corrections to scaling at the quasiperiodic onset of chaos will again separate into gauge and singular components.

N2.6 **A fair split? Number partitioning.**[10] (Computer science, Mathematics, Computaion, Statistics) ③

A group of N kids want to split up into two teams that are evenly matched. If the skill of each player is measured by an integer, can the kids be split into two groups such that the sum of the skills in each group is the same?

This is the *number partitioning problem* (NPP), a classic and surprisingly difficult problem in computer science. To be specific, it is **NP**$--$*complete*—a category of problems for which no known algorithm can guarantee a resolution in a reasonable time (bounded by a polynomial in their size). If the skill a_j of each kid j is in the range $1 \leq a_J \leq 2^M$, the "size" of the NPP is defined as NM. Even the best algorithms will, for the hardest instances, take computer time that grows faster than any polynomial in MN, getting exponentially large as the system grows.

In this exercise, we shall explore connections between this numerical problem and the statistical mechanics of disordered systems. Number partitioning has been termed "the easiest hard problem". It is genuinely hard numerically; unlike some other **NP** $--$*complete* problems, there are no good heuristics for solving NPP (i.e., that work much better than a random search). On the other hand, the random NPP problem (the ensembles of all possible combinations of skills a_j) has many interesting features that can be understood with relatively straightforward arguments and analogies.

We start with the brute-force numerical approach to solving the problem.

(a) *Write a function* `ExhaustivePartition(S)` *that inputs a list* S *of* N *integers, exhaustively searches through the* 2^N *possible partitions into two subsets, and returns the minimum cost (difference in the sums). Test your routine on the four sets [22]* $S_1 = [10, 13, 23, 6, 20]$, $S_2 = [6, 4, 9, 14, 12, 3, 15, 15]$, $S_3 = [93, 58, 141, 209, 179, 48, 225, 228]$, *and* $S_4 = [2474, 1129, 1388, 3752, 821, 2082, 201, 739]$. Hint: S_1 has a balanced partition, and S_4 has a minumum cost of 48. You may wish to return the signs of the minimum-cost partition as part of the debugging process.

What properties emerge from studying ensembles of large partitioning problems? We find a *phase transition*. If the range of integers (M digits in base two) is large and there are relatively few numbers N to rearrange, it is unlikely that a perfect match

[10] This exercise draws heavily from [22, chapter 7]. Hints for the computations can be found at the book website [31].

can be found. (A random instance with $N = 2$ and $M = 10$ has a one chance in $2^{10} = 1{,}024$ of a perfect match, because the second integer needs to be equal to the first.) If M is small and N is large it should be easy to find a match, because there are so many rearrangements possible and the sums are confined to a relatively small number of possible values. It turns out that it is the ratio $\kappa = M/N$ that is the key; for large random systems with $M/N > \kappa_c$ it becomes extremely unlikely that a perfect partition is possible, while if $M/N < \kappa_c$ a fair split is extremely likely.

(b) *Write a function* `MakeRandomPartitionProblem(N,M)` *that generates N integers randomly chosen from $\{1, \ldots, 2^M\}$, rejecting lists whose sum is odd (and hence cannot have perfect partitions). Write a function* `pPerf(N,M,trials)`*, which generates* `trials` *random lists and calls* `ExhaustivePartition` *on each, returning the fraction $p_{\rm perf}$ that can be partitioned evenly (zero cost). Plot $p_{\rm perf}$ versus $\kappa = M/N$, for $N = 3$, 5, 7 and 9, for all integers M with $0 < \kappa = M/N < 2$, using at least a hundred trials for each case. Does it appear that there is a phase transition for large systems where fair partitions go from probable to unlikely? What value of κ_c would you estimate as the critical point?*

Should we be calling this a phase transition? It emerges for large systems; only in the "thermodynamic limit" where N gets large is the transition sharp. It separates two regions with qualitatively different behavior. The problem is much like a spin glass, with two kinds of random variables: the skill levels of each player a_j are fixed, "quenched" random variables for a given random instance of the problem, and the assignment to teams can be viewed as spins $s_j = \pm 1$ that can be varied ("annealed" random variables)[11] to minimize the cost $C = |\sum_j a_j s_j|$.

(c) *Show that the square of the cost C^2 is of the same form as the Hamiltonian for a spin glass, $H = \sum_{i,j} J_{ij} s_i s_j$. What is J_{ij}?*

The putative phase transition in the optimization problem (part (b)) is precisely a zero-temperature phase transition for this spin-glass Hamiltonian, separating a phase with zero ground-state energy from one with nonzero energy in the thermodynamic limit.

We can understand both the value κ_c of the phase transition and the form of $p_{\rm perf}(N, M)$ by studying the distribution of possible "signed" costs $E_{\mathbf{s}} = \sum_j a_j s_j$. These energies are distributed over a maximum total range of $E_{\rm max} - E_{\rm min} = 2\sum_{j=1}^{N} a_j \leq 2N\,2^M$ (all players playing on the plus team, through all on the minus team). For the bulk of the possible team choices $\{s_j\}$, though, there will be some cancellation in this sum. The probability distribution $P(E)$ of these energies for a particular NPP problem $\{a_j\}$ is not simple, but the average probability distribution $\langle P(E)\rangle$ over the ensemble of NPP problems can be estimated using the central limit theorem. (Remember that the central limit theorem states that the sum of N random variables with mean zero and

[11] *Quenched* random variables are fixed terms in the definition of the system, representing dirt or disorder that was frozen in as the system was formed (say, by quenching the hot liquid material into cold water, freezing it into a disordered configuration). *Annealed* random variables are the degrees of freedom that the system can vary to explore different configurations and minimize its energy or free energy.

standard deviation σ converges rapidly to a normal (Gaussian) distribution of standard deviation $\sqrt{N}\sigma$.)

(d) *Estimate the mean and variance of a single term $s_j a_j$ in the sum, averaging over both the spin configurations s_j and the different NPP problem realizations $a_j \in [1, \ldots, 2^M]$, keeping only the most important term for large M.* (Hint: Approximate the sum as an integral, or use the explicit formula $\sum_1^K k^2 = K^3/3 + K^2/2 + K/6$ and keep only the most important term.) *Using the central limit theorem, what is the ensemble-averaged probability distribution $P(E)$ for a team with N players?* Hint: Here $P(E)$ is nonzero only for *even* integers E, so for large N $P(E) \approx (2/\sqrt{2\pi}\sigma)\exp(-E^2/2\sigma^2)$; the normalization is doubled.

Your answer to part (d) should tell you that the possible energies are mostly distributed among integers in a range of size $\sim 2^M$ around zero, up to a factor that goes as a power of N. The total number of states explored by a given system is 2^N. So, the expected number of zero-energy states should be large if $N \gg M$, and go to zero rapidly if $N \ll M$. Let us make this more precise.

(e) *Assuming that the energies for a specific system are randomly selected from the ensemble average $P(E)$, calculate the expected number of zero-energy states as a function of M and N for large N. What value of $\kappa = M/N$ should form the phase boundary separating likely from unlikely fair partitions? Does that agree well with your numerical estimate from part (b)?*

The assumption we made in part (e) ignores the correlations between the different energies due to the fact that they all share the same step sizes a_j in their random walks. Ignoring these correlations turns out to be a remarkably good approximation.[12] We can use the random-energy approximation to estimate $p_{\rm perf}$ that you plotted in part (b).

(f) *In the random-energy approximation, argue that $p_{\rm perf} = 1 - (1 - P(0))^{2^{N-1}}$. Approximating $(1 - A/L)^L \approx \exp(-A)$ for large L, show that*

$$p_{\rm perf}(\kappa, N) \approx 1 - \exp\left[-\sqrt{\frac{3}{2\pi N}}2^{-N(\kappa-\kappa_c)}\right]. \tag{N2.17}$$

Rather than plotting the theory curve through each of your simulations from part (b), we change variables to $x = N(\kappa - \kappa_c) + (1/2)\log_2 N$, where the theory curve

$$p_{\rm perf}^{\rm scaling}(x) = 1 - \exp\left[-\sqrt{\frac{3}{2\pi}}2^{-x}\right] \tag{N2.18}$$

[12] More precisely, we ignore correlations between the energies of different teams $\mathbf{s} = \{s_i\}$, *except* for swapping the two teams $\mathbf{s} \to -\mathbf{s}$. This leads to the $N - 1$ in the exponent of the exponent for $p_{\rm perf}$ in part (f). Notice that in this approximation, NPP is a form of the random energy model (REM, Exercise 3.19), except that we are interested in states of energy near $E = 0$, rather than minimum energy states.

is independent of N. If the theory is correct, your curves should converge to $p_{\rm perf}^{\rm scaling}(x)$ as N becomes large

(g) *Reusing your simulations from part (b), make a graph with your values of* $p_{\rm perf}(x,N)$ *versus* x *and* $p_{\rm perf}^{\rm scaling}(x)$. *Does the random-energy approximation explain the data well?*

Rigorous results show that this random-energy approximation gives the correct value of κ_c. The entropy of zero-cost states below κ_c, the probability distribution of minimum costs above κ_c (of the Weibull form, Exercise 1.9), and the probability distribution of the k lowest cost states are also correctly predicted by the random-energy approximation. It has also been shown that the correlations between the energies of different partitions vanish in the large (N,M) limit so long as the energies are not far into the tails of the distribution, perhaps explaining the successes of ignoring the correlations.

What does this random-energy approximation imply about the computational difficulty of NPP? If the energies of different spin configurations (arrangements of kids on teams) were completely random and independent, there would be no better way of finding zero-energy states (fair partitions) than an exhaustive search of all states. This perhaps explains why the best algorithms for NPP are not much better than the exhaustive search you implemented in part (a); even among **NP** $-$ $-$*complete* problems, NPP is unusually unyielding to clever methods.[13] It also lends credibility to the conjecture in the computer science community that **P** $\neq$ **NP** $-$ $-$*complete*; any polynomial-time algorithm for NPP would have to ingeneously make use of the seemingly unimportant correlations between energy levels.

N2.7 **Cardiac dynamics.**[14] (Computation, Biology, Complexity) ④

Reading: References [25, 34], Niels Otani, various web pages on cardiac dynamics, http://otani.vet.cornell.edu, and Arthur T. Winfree, "Varieties of spiral wave behavior: An experimentalist's approach to the theory of excitable media", *Chaos*, **1**, 303-334 (1991). See also spiral waves in Dictyostelium by Bodenschatz and Franck, http://newt.ccmr.cornell.edu/Dicty/diEp47A.mov and http://newt.ccmr.cornell.edu/Dicty/diEp47A.avi.

The cardiac muscle is an excitable medium. In each heartbeat, a wave of excitation passes through the heart, compressing first the atria which pushes blood into the ventricles, and then compressing the ventricles pushing blood into the body. In this exercise we will study simplified models of heart tissue, that exhibit *spiral waves* similar to those found in arrhythmias.

An excitable medium is one which, when triggered from a resting state by a small stimulus, responds with a large pulse. After the pulse there is a refractory period during which it is difficult to excite a new pulse, followed by a return to the resting

[13] The computational cost does peak near $\kappa = \kappa_c$. For small $\kappa \ll \kappa_c$ it is relatively easy to find a good solution, but this is mainly because there are so many solutions; even random search only needs to sample until it finds one of them. For $\kappa > \kappa_c$ showing that there is no fair partition becomes slightly easier as κ grows [22, Fig. 7.3].

[14] This exercise and the associated software were developed in collaboration with Christopher Myers.

state. The FitzHugh-Nagumo equations provide a simplified model for the excitable heart tissue:[15]

$$\begin{aligned}\frac{\partial V}{\partial t} &= \nabla^2 V + \frac{1}{\epsilon}(V - V^3/3 - W) \\ \frac{\partial W}{\partial t} &= \epsilon(V - \gamma W + \beta),\end{aligned} \tag{N2.19}$$

where V is the transmembrane potential, W is the recovery variable, and $\epsilon = 0.2$, $\gamma = 0.8$, and $\beta = 0.7$ are parameters. Let us first explore the behavior of these equations ignoring the spatial dependence (dropping the $\nabla^2 V$ term, appropriate for a small piece of tissue). The dynamics can be visualized in the (V, W) plane.

(a) *Find and plot the* nullclines *of the FitzHugh-Nagumo equations: the curves along which* $\mathrm{d}V/\mathrm{d}t$ *and* $\mathrm{d}W/\mathrm{d}t$ *are zero (ignoring* $\nabla^2 V$*). The intersection of these two nullclines represents the resting state* (V^*, W^*) *of the heart tissue. We apply a stimulus to our model by shifting the transmembrane potential to a larger value—running from initial conditions* $(V^* + \Delta, W^*)$*. Simulate the equations for stimuli* Δ *of various sizes; plot* V *and* W *as a function of time* t*, and also plot* $V(t)$ *versus* $W(t)$ *along with the nullclines. How big a stimulus do you need in order to get a pulse?*

Excitable systems are often close to regimes where they develop spontaneous oscillations. Indeed, the FitzHugh-Nagumo equations are equivalent to the van der Pol equation (which arose in the study of vacuum tubes), a standard system for studying periodic motion.

(b) *Try changing to* $\beta = 0.4$*. Does the system oscillate?* The threshold where the resting state becomes unstable is given when the nullcline intersection lies at the minimum of the V nullcline, at $\beta_c = 7/15$.

Each portion of the tissue during a contraction wave down the heart is stimulated by its neighbors to one side, and its pulse stimulates the neighbor to the other side. This triggering in our model is induced by the Laplacian term $\nabla^2 V$. We simulate the heart on a two-dimensional grid $V(x_i, y_j, t)$, $W(x_i, y_j, t)$, and calculate an approximate Laplacian by taking differences between the local value of V and values at neighboring points.

There are two natural choices for this Laplacian. The five-point discrete Laplacian is generalization of the one-dimensional second derivative, $\partial^2 V/\partial x^2 \approx (V(x + \mathrm{d}x) - 2V(x) + V(x - \mathrm{d}x))/\mathrm{d}x^2$:

$$\begin{aligned}\nabla^2_{[5]} V(x_i, y_i) \approx\ & (V(x_i, y_{i+1}) + V(x_i, y_{i-1}) \\ & + V(x_{i+1}, y_i) + V(x_{i-1}, y_i) \\ & - 4V(x_i, y_i))/\mathrm{d}x^2 \\ \leftrightarrow & \frac{1}{\mathrm{d}x^2} \begin{pmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{pmatrix}\end{aligned} \tag{N2.20}$$

[15] Nerve tissue is also an excitable medium, modeled using different *Hodgkin-Huxley* equations.

where $\mathrm{d}x = x_{i+1} - x_i = y_{i+1} - y_i$ is the spacing between grid points and the last expression is the *stencil* by which you multiply the point and its neighbors by to calculate the Laplacian. The nine-point discrete Laplacian has been fine-tuned for improved circularly symmetry, with stencil

$$\nabla^2_{[9]} V(x_i, y_i) \leftrightarrow \frac{1}{\mathrm{d}x^2} \begin{pmatrix} 1/6 & 2/3 & 1/6 \\ 2/3 & -10/3 & 2/3 \\ 1/6 & 2/3 & 1/6 \end{pmatrix}. \tag{N2.21}$$

We will simulate our partial-differential equation (PDE) on a square 100×100 grid with a grid spacing $\mathrm{d}x = 1$.[16] As is often done in PDEs, we will use the crude Euler time-step scheme $V(t + \Delta) \approx V(t) + \Delta \partial V/\partial t$ (see Exercise 3.12): we find $\Delta \approx 0.1$ is the largest time step we can get away with. We will use "no-flow" boundary conditions, which we implement by setting the Laplacian terms on the boundary to zero (the boundaries, uncoupled from the rest of the system, will quickly turn to their resting state). If you are not supplied with example code that does the two-dimensional plots, you may find them at the book website [31].

(c) *Solve eqn N2.21 for an initial condition equal to the fixed point* (V^*, W^*) *except for a* 10×10 *square at the origin, in which you should apply a stimulus* $\Delta = 3.0$. (Hint: Your simulation should show a pulse moving outward from the origin, disappearing as it hits the walls.)

If you like, you can mimic the effects of the sinoatrial (SA) node (your heart's natural pacemaker) by stimulating your heart model periodically (say, with the same 10×10 square). Realistically, your period should be long enough that the old beat finishes before the new one starts.

We can use this simulation to illustrate general properties of solving PDEs.

(d) **Accuracy.** *Compare the five and nine-point Laplacians. Does the latter give better circular symmetry?* **Stability.** *After running for a while, double the time step* Δ. *How does the system go unstable? Repeat this process, reducing* Δ *until just before it goes nuts. Do you see inaccuracies in the simulation that foreshadow the instability?*

This checkerboard instability is typical of PDEs with too high a time step. The maximum time step in this system will go as $\mathrm{d}x^2$, the lattice spacing squared—thus to make $\mathrm{d}x$ smaller by a factor of two and simulate the same area, you need four times as many grid points and four times as many time points—giving us a good reason for making $\mathrm{d}x$ as large as possible (correcting for grid artifacts by using improved Laplacians). Similar but much more sophisticated tricks have been used recently to spectacularly increase the performance of lattice simulations of the interactions between quarks [11].

As mentioned above, heart arrhythmias are due to spiral waves. To generate spiral waves we need to be able to start up more asymmetric states—stimulating several rectangles at different times. Also, when we generate the spirals, we would like to

[16] Smaller grids would lead to less grainy waves, but slow down the simulation a lot.

emulate electroshock therapy by applying a stimulus to a large region of the heart. We can do both by writing code to interactively stimulate a whole rectangle at one time. Again, the code you have obtained from us should have hints for how to do this.

(e) *Add the code for interactively stimulating a general rectangle with an increment to V of size $\Delta = 3$. Play with generating rectangles in different places while other pulses are going by: make some spiral waves. Clear the spirals by giving a stimulus that spans the system.*

There are several possible extensions of this model, several of which involve giving our model spatial structure that mimics the structure of the heart. (One can introduce regions of inactive "dead" tissue. One can introduce the atrium and ventricle compartments to the heart, with the SA node in the atrium and an AV node connecting the two chambers . . .) Niels Otani has an exercise with further explorations of a number of these extensions, which we link to from the Cardiac Dynamics website.

N2.8 **Quantum dissipation from phonons.** (Quantum) ②

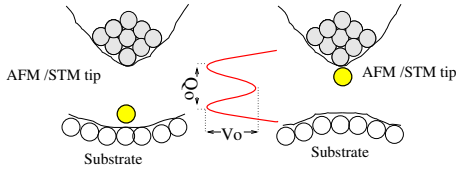


Fig. N2.2 Atomic tunneling from a tip. Any *internal* transition among the atoms in an insulator can only exert a force impulse (if it emits momentum, say into an emitted photon), or a force dipole (if the atomic configuration rearranges); these lead to nonzero phonon overlap integrals only partially suppressing the transition. But a quantum transition that changes the net force between two macroscopic objects (here a surface and a STM tip) can lead to a change in the net force (a force monopole). We ignore here the surface, modeling the force as exerted directly into the center of an insulating elastic medium.[17]See "Atomic Tunneling from a STM/AFM Tip: Dissipative Quantum Effects from Phonons" Ard A. Louis and James P. Sethna, *Phys. Rev. Lett.* **74**, 1363 (1995), and "Dissipative tunneling and orthogonality catastrophe in molecular transistors", S. Braig and K. Flensberg, *Phys. Rev. B* **70**, 085317 (2004).

Electrons cause overlap catastrophes (X-ray edge effects, the Kondo problem, macroscopic quantum tunneling); a quantum transition of a subsystem coupled to an electron bath ordinarily must emit an infinite number of electron-hole excitations because the bath states before and after the transition have zero overlap. This is often called an *infrared* catastrophe (because it is low-energy electrons and holes that cause the zero overlap), or an *orthogonality* catastrophe (even though the two bath states aren't just orthogonal, they are in different Hilbert spaces). Phonons typically do not produce overlap catastrophes (Debye–Waller, Frank–Condon, Mössbauer). This difference is usually attributed to the fact that there are many more low-energy electron-hole pairs

(a constant density of states) than there are low-energy phonons ($\omega_k \sim ck$, where c is the speed of sound and the wave-vector density goes as $(V/2\pi)^3 \mathrm{d}^3k$).

However, the coupling strength to the low energy phonons has to be considered as well. Consider a small system undergoing a quantum transition which exerts a net force at $x = 0$ onto an insulating crystal:

$$\mathcal{H} = \sum_k p_k^2/2m + 1/2\, m\omega_k^2 q_k^2 + F \cdot u_0. \tag{N2.22}$$

Let us imagine a kind of scalar elasticity, to avoid dealing with the three phonon branches (two transverse and one longitudinal); we thus naively write the displacement of the atom at lattice site x_n as $u_n = (1/\sqrt{N}) \sum_k q_k \exp(-\mathrm{i}kx_n)$ (with N the number of atoms), so $q_k = (1/\sqrt{N}) \sum_n u_n \exp(\mathrm{i}kx_n)$.

Substituting for u_0 in the Hamiltonian and completing the square, find the displacement Δ_k of each harmonic oscillator. (Physically, the force F adds a small linear term to the phonon mode with wavevector k, whose minimum becomes displaced by some amount Δ_k.) Let $|F\rangle$ be the ground state of the harmonic oscillators under the force F. Write the formula for the likelihood $\langle F|0\rangle$ that the phonons will all end in their ground states, as a product over k of the phonon overlap integral $\exp(-\Delta_k^2/8\sigma_k^2)$ (with $\sigma_k = \sqrt{\hbar/2m\omega_k}$ the zero-point motion in that mode). Converting the product to the exponential of a sum, and the sum to an integral $\sum_k \sim (V/(2\pi)^3 \int \mathrm{d}^3\mathbf{k}$, do we observe an overlap catastrophe?

Note that you've calculated the probability of a zero-phonon transition—the likelihood that the quantum transition can happen without emitting any phonons is zero. But the same argument shows that there is zero probability of emitting one phonon, or any finite number of phonons. The only allowed transitions emit an infinite number of low-energy phonons. The initial and final ground states are in *different Hilbert spaces*—no finite number of excitations can connect them.

N2.9 **Ising lower critical dimension.** (Dimension dependence) ③

What is the lower critical dimension of the Ising model? If the total energy ΔE needed to destroy long-range order is finite as the system size L goes to infinity, and the associated entropy grows with system size, then surely long-range order is possible only at zero temperature.

(a) **Ising model in D dimensions.** *Consider the Ising model in dimension D on a hypercubic lattice of length L on each side. Estimate the energy*[18] *needed to create a domain wall splitting the system into two equal regions (one spin up, the other spin down). In what dimension will this wall have finite energy as $L \to \infty$? Suggest a bound for the lower critical dimension of the Ising model.*

The scaling at the lower critical dimension is often unusual, with quantities diverging in ways different from power laws as the critical temperature T_c is approached.

[18] Energy, not free energy! Think about $T = 0$.

(b) **Correlation length in 1D Ising model.** *Estimate the number of domain walls at temperature T in the 1D Ising model. How does the correlation length ξ (the distance between domain walls) grow as $T \to T_c = 0$?* (Hint: Change variables to $\eta_i = S_i S_{i+1}$, which is -1 if there is a domain wall between sites i and $i+1$.) *The correlation exponent ν satisfying $\xi \sim (T - T_c)^{-\nu}$ is 1, ~ 0.63, and 1/2 in dimensions $D = 2$, 3, and ≥ 4, respectively. Is there an exponent ν governing this divergence in one dimension? How does ν behave as $D \to 1$?*

N2.10 **XY lower critical dimension and the Mermin-Wagner Theorem.** (Dimension dependence) ③

Consider a model of continuous unit-length spins (e.g., XY or Heisenberg) on a D-dimensional hypercubic lattice of length L. Assume a nearest-neighbor ferromagnetic bond energy

$$- J \, \mathbf{S}_i \cdot \mathbf{S}_j. \tag{N2.23}$$

Estimate the energy needed to twist the spins at one boundary 180° with respect to the other boundary (the energy difference between periodic and antiperiodic boundary conditions along one axis). In what dimension does this energy stay finite in the thermodynamic limit $L \to \infty$? Suggest a bound for the lower critical dimension for the emergence of continuous broken symmetries in models of this type.

Note that your argument produces only one thick domain wall (unlike the Ising model, where the domain wall can be placed in a variety of places). If in the lower critical dimension its energy is fixed as $L \to \infty$ at a value *large* compared to $k_B T$, one could imagine most of the time the order might maintain itself across the system. The actual behavior of the XY model in its lower critical dimension is subtle.

On the one hand, there cannot be long-range order. This can be seen convincingly, but not rigorously, by estimating the effects of fluctuations at finite temperature on the order parameter, within linear response. Pierre Hohenberg, David Mermin and Herbert Wagner proved it rigorously (including nonlinear effects) using an inequality due to Bogoliubov. One should note, though, that the way this theorem is usually quoted ("continuous symmetries cannot be spontaneously broken at finite temperatures in one and two dimensions") is too general. In particular, for two-dimensional crystals one has long-range order in the crystalline *orientations*, although one does not have long-range broken translational order.

On the other hand, the XY model does have a phase transition in its lower critical dimension at a temperature $T_c > 0$. The high-temperature phase is a traditional paramagentic phase, with exponentially decaying correlations between orientations as the distance increases. The low-temperature phase indeed lacks long-range order, but it does have a *stiffness*—twisting the system (as in your calculation above) by 180° costs a free energy that goes to a constant as $L \to \infty$. In this stiff phase the spin–spin correlations die away not exponentially, but as a power law.

The corresponding *Kosterlitz–Thouless phase transition* has subtle, fascinating scaling properties. Interestingly, the defect that destroys the stiffness (a vortex) in the

Kosterlitz–Thouless transition does *not* have finite energy as the system size L gets large. We shall see that its energy grows $\sim \log L$, while its entropy grows $\sim T \log L$, so entropy wins over energy as the temperature rises, even though the latter is infinite.

N2.11 **Long-range Ising.** (Dimension dependence) ③

The one-dimensional Ising model can have a finite-temperature transition if we give each spin an interaction with distant spins.

Long-range forces in the 1d Ising model. *Consider an Ising model in one dimension, with long-range ferromagnetic bonds*

$$\mathcal{H} = \sum_{i>j} \frac{J}{|i-j|^\sigma} S_i S_j. \tag{N2.24}$$

For what values of σ will a domain wall between up- and down-spins have finite energy? Suggest a bound on σ analogous to the lower critical dimension (the maximum power σ, above which a ferromagnetic state is only possible when the temperature is zero). (Hint: Approximate the sum by a double integral. Avoid $i = j$.)

The long-range 1D Ising model at the lower critical power law has a transition that is closely related to the Kosterlitz-Thouless transition. It is in the same universality class as the famous (but obscure) Kondo problem in quantum phase transitions. And it is less complicated to think about and less complicated to calculate with than either of these other two cases.

N2.12 **Equilibrium Crystal Shapes.** (Condensed matter) ③

What is the equilibrium shape of a crystal? There are nice experiments on single crystals of salt, gold, and lead crystals (see http://www.lassp.cornell.edu/sethna/CrystalShapes/Equilibrium_Crystal_Shapes.html). They show beautiful faceted shapes, formed by carefully annealing single crystals to equilibrium at various temperatures. The physics governing the shape involves the anisotropic surface tension $\gamma(\hat{n})$ of the surface, which depends on the orientation $\hat{n}$ of the local surface with respect to the crystalline axes.

We can see how this works by considering the problem for atoms on a 2D square lattice with near-neighbor interations (a lattice gas which we map in the standard way onto a conserved-order parameter Ising model). Here $\gamma(\theta)$ becomes the line tension between the up and down phases—the interfacial free energy per unit length between an up-spin and a down-spin phase. We draw heavily on Craig Rottman and Michael Wortis, *Phys. Rev. B* **24**, 6274 (1981), and on W. K. Burton, N. Cabrera and F. C. Frank, *Phil. Trans. R. Soc. Lond. A* **243** 299-358 (1951).

(a) Interfacial free energy, $T = 0$. *Consider an interface at angle θ between up-spins and down-spins. Show that the energy cost per unit length of the interface is*

$$\gamma_0(\theta) = 2J(\cos(|\theta|) + \sin(|\theta|)), \tag{N2.25}$$

where length is measured in lattice spacings.

(b) Interfacial free energy, low T. *Consider an interface at angle* $\theta = \arctan N/M$, *connecting the origin to the point* (M, N). *At zero temperature, this will be an ensemble of staircases, with* N *steps upward and* M *steps forward. Show that the total number of such staircases is* $(M+N)!/(M!N!)$. Hint: The number of ways of putting k balls into ℓ jars allowing more than one ball per jar (the "number of combinations with repetition") is $(k+\ell-1)!/(k!(\ell-1)!)$. Look up the argument. *Using Stirling's formula, show that the entropy per unit length is*

$$\begin{aligned} s_0(\theta) =& k_B\Big((\cos(|\theta|) + \sin(|\theta|)) \\ & \qquad \log\left(\cos(|\theta|) + \sin(|\theta|)\right) \\ & - \cos(|\theta|)\log(\cos(|\theta|)) \\ & - \sin(|\theta|)\log(\sin(|\theta|))\Big). \end{aligned} \tag{N2.26}$$

How would we generate an equilibrium crystal for the 2D Ising model? (For example, see "The Gibbs-Thomson formula at small island sizes - corrections for high vapour densities" Badrinarayan Krishnamachari, James McLean, Barbara Cooper, and James P. Sethna, *Phys. Rev. B* **54**, 8899 (1996).) Clearly we want a conserved order parameter simulation (otherwise the up-spin "crystal" cluster in a down-spin "vapor" environment would just flip over). The tricky part is that an up-spin cluster in an infinite sea of down-spins will evaporate—it is unstable.[19] The key is to do a simulation below T_c, but with a net (conserved) negative magnetization slightly closer to zero that expected in equilibrium. The extra up-spins will (in equilibrium) mostly find one another, forming a cluster whose time-average will give an equilibrium crystal.

Rottman and Wortis tell us that the equilibrium crystal shape (minimizing the perimeter energy for fixed crystal area) can be found as a parametric curve

$$\begin{aligned} x &= \cos(\theta)\gamma(\theta, T) - \sin(\theta)\frac{\mathrm{d}\gamma}{\mathrm{d}\theta} \\ y &= \sin(\theta)\gamma(\theta, T) + \cos(\theta)\frac{\mathrm{d}\gamma}{\mathrm{d}\theta} \end{aligned}$$

where $\gamma(\theta, T) = \gamma_0(\theta) - Ts(\theta, T)$ is the free energy per unit length of the interface. Deriving this is cool, but somewhat complicated.[20]

(c) Wolff construction. *Using the energy of eqn N2.28 and approximating the entropy at low temperatures with the zero-temperature form eqn N2.29, plot the Wulff shape*

[19] If you put on an external field favoring up-spins, then large clusters will grow and small clusters will shrink. The borderline cluster size is the *critical droplet* (see Entropy, Order Parameters, and Complexity, section 11.3). Indeed, the critical droplet will in general share the equilibrium crystal shape.

[20] See Burton, Cabrera, and Frank, Appendix D. This is their equation D4, with $p \propto \gamma(\theta)$ given by equation D7.

for $T = 0.01$, 0.1, 0.2, 0.4, and 0.8J for one quadrant ($0 < \theta < \pi/2$). Hint: Ignore the parts of the face outside the quadrant; they are artifacts of the low temperature approximation. *Does the shape become circular*[21] *near $T_c = 2J/\log(1+\sqrt{2}) \sim 2.269J$? Why not?* If you're ambitious, Rottman and Wortis's article above gives the exact interfacial free energy for the 2D Ising model, which should fix this problem. *What is the shape at low temperatures, where our approximation is good? Do we ever get true facets?* The approximation of the interface as a staircase is a *solid-on-solid* model, which ignores overhangs. It is a good approximation at low temperatures.

Our model does not describe faceting—one needs three dimensions to have a *roughening transition*, below which there are flat regions on the equilibrium crystal shape.[22]

[21] This is the emergent spherical symmetry at the Ising model critical point.

[22] Flat regions demand that the free energy for adding a step onto the surface become infinite. Steps on the surfaces of three-dimensional crystals are long, and if they have positive energy per unit length the surface is faceted. To get an infinite energy for a step on a two-dimensional surface you need long-range interactions.

Chapter 3

Numerical Methods

These exercises cover numerical methods, and are designed to complement the text *Numerical Recipes*, by William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery.

N3.1 **Condition number and accuracy.**[1] (Numerical) ③

You may think this exercise, with a 2x2 matrix, hardly demands a computer. However, it introduces tools for solving linear equations, condition numbers, singular value decomposition, all while illustrating subtle properties of matrix solutions. Use whatever linear algebra packages are provided in your software environment.

Consider the equation $A\mathbf{x} = \mathbf{b}$, where

$$A = \begin{pmatrix} 0.780 & 0.563 \\ 0.913 & 0.659 \end{pmatrix} \quad \text{and} \quad \mathbf{b} = \begin{pmatrix} 0.217 \\ 0.254 \end{pmatrix}. \tag{N3.1}$$

The exact solution is $\mathbf{x} = (1, -1)$. Consider the two approximate solutions $\mathbf{x}_\alpha = (0.999, -1.001)$ and $\mathbf{x}_\beta = (0.341, -0.087)$.

(a) *Compute the residuals* $\mathbf{r}_\alpha$ *and* $\mathbf{r}_\beta$ *corresponding to the two approximate solutions.* (The residual is $\mathbf{r} = \mathbf{b} - \mathbf{A}\mathbf{x}$.) *Does the more accurate solution have the smaller residual?*

(b) *Compute the condition number*[2] *of* A. *Does it help you understand the result of (a)?* (Hint: V^T maps the errors $\mathbf{x}_\alpha - \mathbf{x}$ and $\mathbf{x}_\beta - \mathbf{x}$ into what combinations of the two singular values?)

(c) *Use a black-box linear solver to solve for* $\mathbf{x}$. *Subtract your answer from the exact one. Do you get within a few times the machine accuracy of* 2.2×10^{-16}*? Is the problem*

[1] Adapted from Saul Teukolsky, 2003.

[2] See section 2.6.2 for a technical definition of the condition number, and how it is related to singular value decomposition. Look on the Web for the more traditional definition(s), and how they are related to the accuracy.

the accuracy of the solution, or rounding errors in calculating A and b? (Hint: try calculating $A\mathbf{x} - \mathbf{b}$.)

N3.2 **Sherman–Morrison formula.** (Numerical) ③

Consider the 5x5 matrices

$$T = \begin{pmatrix} E-t & t & 0 & 0 & 0 \\ t & E & t & 0 & 0 \\ 0 & t & E & t & 0 \\ 0 & 0 & t & E & t \\ 0 & 0 & 0 & t & E-t \end{pmatrix} \tag{N3.2}$$

and

$$C = \begin{pmatrix} E & t & 0 & 0 & t \\ t & E & t & 0 & 0 \\ 0 & t & E & t & 0 \\ 0 & 0 & t & E & t \\ t & 0 & 0 & t & E \end{pmatrix}. \tag{N3.3}$$

These matrices arise in one-dimensional models of crystals.[3] The matrix T is *tridiagonal*: its entries are zero except along the central diagonal and the entries neighboring the diagonal. Tridiagonal matrices are fast to solve; indeed, many routines will start by changing basis to make the array tridiagonal. The matrix C, on the other hand, has a nice periodic structure: each basis element has two neighbors, with the first and last basis elements now connected by t in the upper-right and lower-left corners. This periodic structure allows for analysis using Fourier methods (Bloch waves and $\mathbf{k}$-space).

For matrices like C and T which differ in only a few matrix elements[4] we can find C^{-1} from T^{-1} efficiently using the Sherman-Morrison formula (section 2.7).

Compute the inverse[5] *of T for $E = 3$ and $t = 1$. Compute the inverse of C. Compare the difference $\Delta = T^{-1} - C^{-1}$ with that given by the Sherman-Morrison formula*

$$\Delta = \frac{T^{-1}\mathbf{u} \otimes \mathbf{v}\mathbf{T^{-1}}}{1 + \mathbf{v} \cdot T^{-1} \cdot \mathbf{u}}. \tag{N3.4}$$

[3] As the Hamiltonian for electrons in a one-dimensional chain of atoms, t is the hopping matrix element and E is the on-site energy. As the potential energy for longitudinal vibrations in a one-dimensional chain, $E = -2t = K$ is the spring constant between two neighboring atoms. The tridiagonal matrix T corresponds to a kind of free boundary condition, while C corresponds in both cases to periodic boundary conditions.

[4] More generally, this works whenever the two matrices differ by the outer product $\mathbf{u} \otimes \mathbf{v}$ of two vectors. By taking the two vectors to each have one nonzero component $u_i = u\delta_{ia}, v_j = v\delta_{jb}$, the matrices differ at one matrix element $\Delta_{ab} = uv$; for our matrices $\mathbf{u} = \mathbf{v} = 1, 0, 0, 0, 1$ (see section 2.7.2).

[5] Your software environment should have a *solver* for tridiagonal systems, rapidly giving $\mathbf{u}$ in the equation $T \cdot \mathbf{u} = \mathbf{r}$. It likely will not have a special routine for inverting tridiagonal matrices, but our matrix is so small it is not important.

N3.3 **Methods of interpolation.** (Numerical) ③

We've implemented four different interpolation methods for the function $\sin(x)$ on the interval $(-\pi, \pi)$. On the left, we see the methods using the five points $\pm\pi$, $\pm\pi/2$, and 0; on the right we see the methods using ten points. The graphs show the interpolation, its first derivative, its second derivative, and the error. The four interpolation methods we have implemented are (1) Linear, (2) Polynomial of degree three ($M = 4$), (3) Cubic spline, and (4) Barycentric rational interpolation. *Which set of curves (A, B, C, or D) in Fig. N3.1 corresponds with which method?*

N3.4 **Numerical definite integrals.** (Numerical) ③

In this exercise we will integrate the function you graphed in the first, warmup exercise:

$$F(x) = \exp(-6\sin(x)). \tag{N3.5}$$

As discussed in Numerical Recipes, the word *integration* is used both for the operation that is inverse to differentiation, and more generally for finding solutions to differential equations. The old-fashioned term specific to what we are doing in this exercise is *quadrature*.

(a) *Black-box. Using a professionally written black-box integration routine of your choice, integrate $F(x)$ between zero and π. Compare your answer to the analytic integral*[6] *($\approx 0.34542493760937693$) by subtracting the analytic form from your numerical result. Read the documentation for your black box routine, and describe the combination of algorithms being used.*

(b) *Trapezoidal rule. Implement the trapezoidal rule with your own routine. Use it to calculate the same integral as in part (a). Calculate the estimated integral $Trap(h)$ for $N+1$ points spaced at $h = \pi/N$, with $N = 1, 2, 4, \ldots, 2^{10}$. Plot the estimated integral versus the spacing h. Does it extrapolate smoothly to the true value as $h \to 0$? With what power of h does the error vanish? Replot the data as $Trap(h)$ versus h^2. Does the error now vanish linearly?*

Numerical Recipes tells us that the error is an *even* polynomial in h, so we can extrapolate the results of the trapezoidal rule using polynomial interpolation in powers of h^2.

(c) *Simpson's rule (paper and pencil). Consider a linear fit (i.e., $A+Bh^2$) to two points at $2h_0$ and h_0 on your $Trap(h)$ versus h^2 plot. Notice that the extrapolation to $h \to 0$ is A, and show that A is $4/3\ \mathrm{Trap}(h_0) - 1/3\ \mathrm{Trap}(2h_0)$. What is the net weight associated with the even points and odd points? Is this Simpson's rule?*

(d) *Romberg integration. Apply M-point polynomial interpolation (here extrapolation) to the data points $\{h^2, \mathrm{Trap}(h)\}$ for $h = \pi/2, \ldots, \pi/2^M$, with values of M between two and ten.* (Note that the independent variable is h^2.) *Make a log–log plot of the absolute value of the error versus $N = 2^M$. Does this extrapolation improve convergence?*

[6] $\pi(\mathrm{BesselI}[0,6] - \mathrm{StruveL}[0,6])$, according to Mathematica.

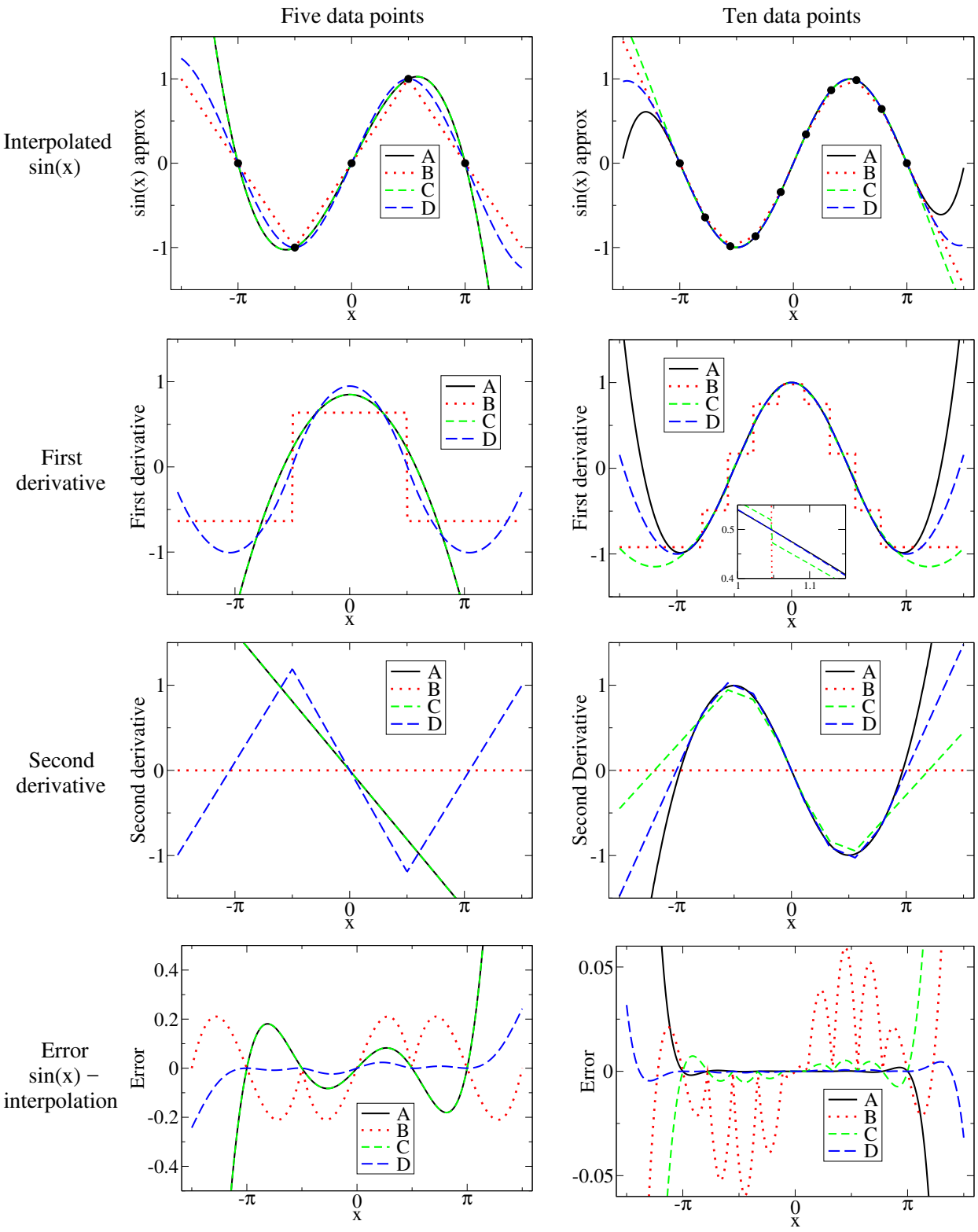


Fig. N3.1 Interpolation methods.

(e) *Gaussian Quadrature. Implement Gaussian quadrature with N points optimally chosen on the interval* $(0, \pi)$*, with* $N = 1, 2, \ldots 5$. (You may find the points and the weights appropriate for integrating functions on the interval $(-1, 1)$ on the course website; you will need to rescale them for use on $(0, \pi)$.) *Make a log–log plot of the absolute value of your error as a function of the number of evaluation points N, along with the corresponding errors from the trapezoidal rule and Romberg integration.*

(f) *Integrals of periodic functions. Apply the trapezoidal rule to integrate* $F(x)$ *from zero to* 2π*, and plot the error on a log plot (log of the absolute value of the error versus N) as a function of the number of points N up to* $N = 20$. (The true value should be around 422.44623805153909946.) *Why does it converge so fast?* (Hint: Don't get distracted by the funny alternation of accuracy between even and odd points.)

The location of the Gauss points depend upon the class of functions one is integrating. In part (e), we were using Gauss-Legendre quadrature, appropriate for functions which are analytic at the endpoints of the range of integration. In part (f), we have a function with *periodic boundary conditions*. For functions with periodic boundary conditions, the end-points are no longer special. What corresponds to Gaussian quadrature for periodic functions is just the trapezoidal rule: equally-weighted points at equally spaced intervals.

N3.5 **Numerical derivatives.** (Rounding errors, Accuracy) ②

Calculate the numerical first derivative of the function $y(x) = \sin(x)$*, using the centered two-point formula* $dy/dx \sim (y(x+h) - y(x-h))/(2h)$*, and plot the error* $y'(x) - \cos(x)$ *in the range* $(-\pi, \pi)$ *at 100 points.* (Do a good job! Use the step-size h described in Numerical Recipes section 5.7 to optimize the sum of the truncation error and the rounding error. Also, make sure that the step-size h is exactly representable on the machine.) *How does your actual error compare to the fractional error estimate given in NR section 5.7? Calculate and plot the numerical second derivative using the formula*

$$\frac{\mathrm{d}^2 y}{\mathrm{d}x^2} \sim \frac{y(x+h) - 2y(x) + y(x-h)}{h^2}, \tag{N3.6}$$

again optimizing h and making it exactly representable. Estimate your error again, and compare to the observed error.

N3.6 **Summing series.** (Efficiency) ②

Write a routine to calculate the sum

$$s_n = \sum_{j=0}^{n} (-1)^j \frac{1}{2j+1}. \tag{N3.7}$$

As $n \to \infty$, $s_\infty = \pi/4$. *About how many terms do you need to sum to get convergence to within* 10^{-7} *of this limit? Now try using Aitken's* Δ^2 *process to accelerate the convergence:*

$$s'_n = s_n - \frac{(s_{n+1} - s_n)^2}{s_{n+2} - 2s_{n+1} + s_n}. \tag{N3.8}$$

About how many terms do you need with Aitken's method to get convergence to within 10^{-7}*?*

N3.7 **Random histograms.** (Random numbers) ②

(a) *Investigate the random number generator for your system of choice. What is its basic algorithm? Its period?*

(b) *Plot a histogram with 100 bins, giving the normalized probability density of 100,000 random numbers sampled from (a) a uniform distribution in the range* $0 < x < 2\pi$*, (b) an exponential distribution* $\rho(x) = 6\exp(-6x)$*, and (c) a normal distribution of mean* $\bar{x} = 3\pi/2$ *and standard deviation* $\sigma = 1/\sqrt{6}$*. Before each plot, set the seed of your random number generator. Do you now get the same plot when you repeat?*

N3.8 **Monte Carlo integration.** (Random numbers, Robust algorithms) ③

How hard can numerical integration be? Suppose the function f is wildly nonanalytic, or has a peculiar or high-dimensional domain of integration? In the worst case, one can always try Monte Carlo integration. The basic idea is to pepper points at random in the integration interval. The integration volume times the average of the function $V\langle f\rangle$ is the estimate of the integral.

As one might expect, the expected error in the integral after N evaluations is given by $1/\sqrt{N-1}$ times the standard deviation of the sampled points (NR equation 7.7.1).

(a) Monte Carlo in one dimensional integrals. *Use Monte Carlo integration to estimate the integral of the function introduced in the preliminary exercises*

$$y(x) = \exp(-6\sin(x)) \tag{N3.9}$$

over $0 \le x < 2\pi$*.* (The correct value of the integral is around 422.446.) *How many points do you need to get 1% accuracy? Answer this last question both by experimenting with different random number seeds, and by calculating the expected number from the standard deviation.* (You may use the fact that $\langle y^2\rangle = (1/2\pi)\int_0^{2\pi} y^2(x)\mathrm{d}x \approx 18{,}948.9$.)

Monte Carlo integration is not the most efficient method for calculating integrals of smooth functions like $y(x)$. Indeed, since $y(x)$ is periodic in the integration interval, equally spaced points weighted equally (the trapezoidal rule) gives exponentially rapid convergence; it takes only nine points to get 1% accuracy. Even for smooth functions, though, Monte Carlo integration is useful in high dimensions.

(b) Monte Carlo in many dimensions (No detailed calculations expected). *For a hypothetical ten-dimensional integral, if we used a regular grid with nine points along each axis, how many function evaluations would we need for equivalent accuracy? Does the number of Monte Carlo points needed depend on the dimension of the space, presuming (perhaps naively) that the variance of the function stays fixed?*

Our function $y(x)$ is quite close to a Gaussian. (Why? Taylor expand $\sin(x)$ about $x = 3\pi/2$.) We can use this to do *importance sampling*. The idea is to evaluate the integral of $h(x)g(x)$ by randomly sampling h with probability g, picking $h(x) = y(x)/g(x)$. The

variance is then $\langle h^2\rangle - \langle h\rangle^2$. In order to properly sample the tail near $x = \pi/2$, we should mix a Gaussian and a uniform distribution:

$$g(x) = \frac{\epsilon}{2\pi} + \frac{1-\epsilon}{\sqrt{\pi/3}} \exp(-6(x - 3\pi/2)^2/2). \tag{N3.10}$$

I found minimum variance around $\epsilon = 0.005$.

(c) Importance Sampling (Optional for 480). *Generate 1,000 random numbers with probability distribution g.*[7] *Use these to estimate the integral of $y(x)$. How accurate is your answer?*

N3.9 **Washboard potential.** (Solving) ②

Consider a washboard potential[8]

$$V(r) = A_1 \cos(r) + A_2 \cos(2r) - Fr \tag{N3.11}$$

with $A_1 = 5$, $A_2 = 1$, and F initially equal to 1.5.

(a) *Plot $V(r)$ over $(-10, 10)$. Numerically find the local maximum of V near zero, and the local minimum of V to the left (negative side) of zero. What is the potential energy barrier for moving from one well to the next in this potential?*

Usually finding the minimum is only a first step—one wants to explore how the minimum moves and disappears...

(b) *Increasing the external tilting field F, graphically roughly locate the field F_c where the barrier disappears, and the location r_c at this field where the potential minimum and maximum merge.* (This is a saddle-node bifurction.) *Give the criterion on the first derivative and the second derivative of $V(r)$ at F_c and r_c. Using these two equations, numerically use a root-finding routine to locate the saddle-node bifurcation F_c and r_c.*

N3.10 **Sloppy minimization.** (Statistics) ③

"With four parameters I can fit an elephant. With five I can make it waggle it's trunk." This statement, attributed to many different sources (from Carl Friedrich Gauss to Fermi), reflects the problems found in fitting multiparameter models to data. One almost universal problem is *sloppiness*—the parameters in the model are poorly constrained by the data.

[7] Take $(1-\epsilon)M$ Gaussian random numbers and ϵM random numbers uniformly distributed on $(0, 2\pi)$. The Gaussian has a small tail that extends beyond the integration range $(0, 2\pi)$, so the normalization of the second term in the definition of g is not quite right. You can fix this by simply throwing away any samples that include points outside the range.

[8] A washboard is what people used to hand-wash clothing. It is held at an angle, and has a series of corrugated ridges; one holds the board at an angle and rubs the wet clothing on it. Washboard potentials arise in the theory of superconducting Josephson junctions, in the motion of defects in crystals, and in many other contexts.

Consider the classic ill-conditioned problem of fitting exponentials to radioactive decay data. If you know that at $t = 0$ there are equal quantities of N radioactive materials with half-lives θ_α, the radioactivity that you would measure is

$$y_{\boldsymbol{\theta}}(t) = \sum_{\alpha=0}^{N-1} \theta_\alpha \exp(-\theta_\alpha t). \tag{N3.12}$$

Now, suppose you do not know the decay rates θ_α. Can you reconstruct them by fitting the data to experimental data $d(t)$?

Start with just two radioactive decay elements $N = 2$. Suppose the actual decay constants for $d(t)$ are $\boldsymbol{\theta}^{[0]} = [1, 2]$ (so the data fall on the curve $d(t) = \exp(-t) + 2\exp(-2t) = y_{\boldsymbol{\theta}^{[0]}}(t)$). For convenience, suppose we have perfect data at all times, with uniform error bars, so the cost is an integral over all times of the square of the error

$$C[\boldsymbol{\theta}] = \int_0^\infty \left(y_{\boldsymbol{\theta}}(t) - d(t)\right)^2 \, \mathrm{d}t. \tag{N3.13}$$

(a) *Draw a contour plot of C in the square $0.5 < \theta_\alpha < 2.5$, with contours at $C = \{2^{-12}, 2^{-11}, \ldots, 2^0\}$. Set the number of grid points per side to 40 to see the two minima.*

One can see from the contour plot that measuring the two rate constants separately would be a challenge. This is because the two exponentials have similar shapes, so increasing one decay rate and decreasing the other can almost perfectly compensate for one another.

(b) *If we assume both elements decay with the same decay rate $\theta = \theta_0 = \theta_1$, minimize the cost to find the optimum choice for θ. Where is this point on the contour plot? Plot $d(t)$ and $y(t)$ with this single-exponent best fit on the same graph, over $0 < t < 2$. Do you agree that it would be difficult to distinguish these two fits?*

This problem can become much more severe in higher dimensions. The banana-shaped ellipses in your contour plot can become needle-like, with aspect ratios of more than a thousand to one (about the same as a human hair). The relative widths of the ellipses are given by the square roots of the eigenvalues of the cost C.

(c) *For our exercise, where the data are perfectly fit by $\boldsymbol{\theta} = \boldsymbol{\theta}^{[0]}$, show that the cost Hessian is a continuous integral*

$$H_{\alpha\beta} = (J^T J)_{\alpha\beta} = J_{t\alpha} J_{t\beta} = \int_0^\infty J(t,\alpha) J(t,\beta) \mathrm{d}t \tag{N3.14}$$

where the Jacobian is now the $\infty \times N$ "matrix" $J(t, \alpha) = \exp(-\theta_\alpha t)(1 - \theta_\alpha t)$.

(d) *Write a routine to calculate $H(\boldsymbol{\theta})$ by doing the indefinite integral in eqn N3.14. Find the eigenvalues and the eigenvectors for the cost Hessian H for your plot in part (b), evaluated at $\boldsymbol{\theta}^{[0]}$, and check them against your contour plot. What is the ratio of the long axis to the short axis, as predicted from your eigenvalues? For a sum of nine exponentials, with $\boldsymbol{\theta}^{[0]} = [1, 2, 3, \ldots, 9]$, construct the Hessian, find its eigenvalues. By what factor does each successive eigenvalue shrink? Are they sloppy (roughly equally spaced in log)?*

N3.11 **Sloppy monomials.**[9] (Statistics) ③

The same function $f(x)$ can be approximated in many ways. Indeed, the same function can be fit in the same interval by the same type of function in several different ways! For example, in the interval $[0, 1]$, the function $\sin(2\pi x)$ can be approximated (badly) by a fifth-order Taylor expansion, a Chebyshev polynomial, or a least-squares (Legendre[10]) fit:

$$\begin{aligned} f(x) &= \sin(2\pi x) \\ f_{\text{Taylor}} &\approx 0.000 + 6.283x + \ \ 0.000x^2 - 41.342x^3 \\ &\quad + \ \ 0.000x^4 + 81.605x^5 \\ f_{\text{Chebyshev}} &\approx 0.0066 + 5.652x + \ \ 9.701x^2 - 95.455x^3 \\ &\quad + 133.48x^4 - 53.39x^5 \\ f_{\text{Legendre}} &\approx 0.016 + 5.410x + 11.304x^2 - 99.637x^3 \\ &\quad + 138.15x^4 - 55.26x^5 \end{aligned}$$

It is not a surprise that the best fit polynomial differs from the Taylor expansion, since the latter is not a good approximation. But it is a surprise that the last two polynomials are so different. The maximum error for Legendre is less than 0.02, and for Chebyshev is less than 0.01, even though the two polynomials differ by

$$\begin{aligned} \text{Chebyshev} - \text{Legendre} = & \\ & - 0.0094 + 0.242x - 1.603x^2 \\ & + 4.182x^3 - 4.67x^4 + 1.87x^5 \end{aligned} \tag{N3.15}$$

a polynomial with coefficients two hundred times larger than the maximum difference!

(a) *Plot $f(x)$, f_{Legendre}, and $f_{\text{Chebyshev}}(x)$ between zero and one on the same graph. Plot $f(x) - f_{\text{Legendre}}$ and $f(x) - f_{\text{Chebyshev}}(x)$ on the same graph with the same range. The first minimizes the squared difference on $[0, 1]$ (eqn* **??***), but it has large errors near the edges. If you were writing a routine to use for calculating $\sin(2\pi x)$ to machine precision in this range, would it be better to use the Legendre or the Chebyshev approximation? Now plot Plot $f_{\text{Chebyshev}}(x) - f_{\text{Legendre}}$ in the range $-1, 2$. Does it indeed get much flatter than you would expect given the coefficients?*

This flexibility in the coefficients of the polynomial expansion is remarkable. We can study it by considering the dependence of the quality of the fit on the parameters. Least-squares (Legendre) fits minimize a cost C, the integral of the squared difference

[9] Thanks to Joshua Waterfall, whose research is described here.

[10] The orthogonal polynomials used for least-squares fits on [-1,1] are the Legendre polynomials, assuming continuous data points. Were we using orthogonal polynomials for this exercise, we would need to shift them for use in [0,1].

between the polynomial and the function:

$$C = (1/2) \int_0^1 (f(x) - y_{\boldsymbol{\theta}}(x))^2 \, \mathrm{d}x,$$
$$y_{\boldsymbol{\theta}}(x) = \sum_{\alpha=0}^{N-1} \theta_\alpha x^\alpha \tag{N3.16}$$

How quickly does this cost increase as we move the N parameters θ_α away from their best-fit values? Varying any one monomial coefficient will of course make the fit bad. But apparently certain coordinated changes of coefficients do not cost much—for example, the difference between least-squares and Chebyshev fits given in eqn **??**.

How should we explore the dependence in arbitrary directions in parameter space? We can use the eigenvalues of the Hessian to see how sensitive the fit is to moves along the various eigenvectors...

(b) *Note that the first derivative of the cost C is zero at the best fit. Analytically (paper and pencil) show that the Hessian second derivative of the cost in eqn* **??** *is*

$$\mathcal{H}_{\alpha\beta} = \frac{\partial^2 C}{\partial\theta_\alpha \partial\theta_\beta} = \frac{1}{\alpha + \beta + 1}. \tag{N3.17}$$

This Hessian is the Hilbert matrix, famous for being ill-conditioned (having a huge range of eigenvalues). Tiny eigenvalues of $\mathcal{H}$ correspond to directions in polynomial space where the fit does not change.

(c) *Numerically calculate the eigenvalues of the 6×6 Hessian for fifth-degree polynomial fits. Do they indeed span a large range? How big is the condition number (the ratio of the largest to the smallest eigenvalue)? Are the ratios all approximately equal (a characteristic of sloppy models)?*

Notice from Eqn **??** that the dependence of the polynomial fit on the monomial coefficients is *independent of the function $f(x)$* being fitted. We can thus vividly illustrate the sloppiness of polynomial fits by considering fits to the *zero* function $f(x) \equiv 0$. A polynomial given by an eigenvector of the Hilbert matrix with small eigenvalue must stay close to zero everywhere in the range $[0, 1]$. Let us check this.

(d) *Calculate the eigenvector corresponding to the smallest eigenvalue of $\mathcal{H}$, checking to make sure its norm is one (so the coefficients are of order one). Note that the elements of this vector are the coefficients of a polynomial pertubation $\delta f(x)$ that changes the cost the smallest amount for a unit vector $\boldsymbol{\theta}$. What is that polynomial? Plot the corresponding polynomial in the range $[0, 1]$: does it stay small everywhere in the interval?*

Especially for larger M, the monomial coefficients of the best fit to a function become sloppy—they can vary over large ranges without damaging the fit, if the other coefficients are allowed to compensate. Only a few combinations of coefficients (those of the largest Hessian eigenvalues) are well determined. This turns out to be a fundamental

property that is shared with many other multiparameter fitting problems. Many different terms are used to describe this property. The fits are called *ill-conditioned*: the parameters θ_n are not well constrained by the data. The *inverse problem* is challenging: one cannot practically extract the parameters from the behavior of the model. Or, as our group describes it, the fit is *sloppy*: only a few directions in parameter space (eigenvectors corresponding to the largest eigenvalues) are constrained by the data, and there is a huge space of models (polynomials) varying along sloppy directions that all serve well in describing the data.

At root, the problem with polynomial fits is that all monomials x^n have similar shapes on $[0, 1]$: they all start flat near zero and bend upward. Thus they can be traded for one another; the coefficient of x^4 can be lowered without changing the fit if the coefficients of x^3 and x^5 are suitably adjusted to compensate.

One should note that, were we change basis from the coefficients θ_n of the monomials x^n to the coefficients ℓ_n of the orthogonal (shifted Legendre) polynomials, the situation completely changes. The Legendre polynomials are designed to be different in shape (orthogonal), and hence cannot be traded for one another. Their coefficients ℓ_n are thus well determined by the data, and indeed the Hessian for the cost C in terms of this new basis is the identity matrix. This puzzled us for some time—is the sloppiness intrinsic, or just a sign of a poor choice of variables. Later work, examining the predictions of nonlinear models using *information geometry*, resolved this question: sloppiness is under rather general conditions expected for the collective predictions of multiparameter nonlinear models.

N3.12 **Conservative differential equations: Accuracy and fidelity.** (Ordinary differential equations) ③

In this exercise, we will solve for the motion of a particle of mass $m = 1$ in the potential

$$V(y) = (1/8)\, y^2(-4A^2 + \log^2(y^2)). \tag{N3.18}$$

That is,

$$\begin{aligned} d^2y/dt^2 &= -dV/dy \\ &= -(1/4)y(-4A^2 + 2\log(y^2) + \log^2(y^2)). \end{aligned} \tag{N3.19}$$

We will start the particle at $y_0 = 1$, $v_0 = dy/dt|0 = -A$, and choose $A = 6$.

(a) *Show that the solution to this differential equation is*[11]

$$F(t) = \exp(-6\sin(t)). \tag{N3.20}$$

Note that the potential energy $V(y)$ is zero at the five points $y = 0$, $y = \pm\exp(\pm A)$.

[11] I worked backward to do this. I set the kinetic energy to $1/2(dF/dt)^2$, set the potential energy to minus the kinetic energy, and then substituted y for t by solving $y = F(t)$.

(b) *Plot the potential energy for* $-3/2\exp(\pm A) < y < 3/2\exp(\pm A)$ *(both zoomed in near* $y = 0$ *and zoomed out).* The correct trajectory should oscillate in the potential well with $y > 0$, turning at two points whose energy is equal to the initial total energy. *What is this initial total energy for our the particle? How much of an error in the energy would be needed, for the particle to pass through the origin when it returns? Compare this error to the maximum kinetic energy of the particle (as it passes the bottom of the well).* Small energy errors in our integration routine can thus cause significant changes in the trajectory.

(c) (Black-box) *Using a professionally written black-box differential equation solver of your choice, solve for* $y(t)$ *between zero and* 4π*, at high precision. Plot your answer along with* $F(t)$ *from part (a). (About half of the solvers will get the answer qualitatively wrong, as expected from part (b).) Expand your plot to examine the region* $0 < y < 1$*; is energy conserved? Finally, read the documentation for your black box routine, and describe the combination of algorithms being used.* You may wish to implement and compare more than one algorithm, if your black-box has that option.

Choosing an error tolerance for your differential equation limits the error in each time step. If small errors in early time steps lead to important changes later, your solution may be quite different from the correct one. Chaotic motion, for example, can never be accurately simulated on a computer. All one can hope for is a *faithful* simulation—one where the motion is qualitatively similar to the real solution. Here we find an important discrepancy—the energy of the numerical solution is drifting upward or downward, where energy should be exactly conserved in the true solution. Here we get a dramatic change in the trajectory for a small quantitative error, but any drift in the energy is qualitatively incorrect.

The Leapfrog algorithm is a primitive looking method for solving for the motion of a particle in a potential. It calculates the next position from the previous two:

$$y(t+h) = 2y(t) - y(t-h) + h^2 f[y(t)] \tag{N3.21}$$

where $f[y] = -dV/dy$. Given an initial position $y(0)$ and an initial velocity $v(0)$, one can initialize and finalize

$$\begin{aligned} y(h) &= y(0) + h(v(0) + (h/2)f[y(0)]) \\ v(T) &= (y(T) - y(T-h))/h + (h/2)f[y(T)] \end{aligned} \tag{N3.22}$$

Leapfrog is one of a variety of *Verlet* algorithms. In a more general context, this is called *Stoermer's rule*, and can be extrapolated to zero step-size h as in the Bulirsch–Stoer algorithms. A completely equivalent algorithm, with more storage but less roundoff error, is given by computing the velocity v at the midpoints of each time step:

$$\begin{aligned} v(h/2) &= v(0) + (h/2)f[y(0)] \\ y(h) &= y(0) + h\,v(h/2) \\ v(t+h/2) &= v(t-h/2) + hf[y(t)] \\ y(t+h) &= y(t) + h\,v(t+h/2) \end{aligned} \tag{N3.23}$$

where we may reconstruct $v(t)$ at integer time steps with $v(t) = v(t-h/2)+(h/2)f[y(t)]$.

(d) (Leapfrog and symplectic methods) *Show that $y(t+h)$ and $y(h)$ from eqns N3.18 and N3.19 converge to the true solutions as as $h \to 0$—that is, the time-step error compared to the solution of $d^2y/dt^2 = f[y]$ vanishes faster than h. To what order in h are they accurate after one time step?*[12] *Implement Leapfrog, and apply it to solving eqn N3.16 in the range $0 < x < 4\pi$, starting with step size $h = 0.01$. How does the accuracy compare to your more sophisticated integrator, for times less than $t = 2$? Zoom in to the range $0 < y < 1$, and compare with your packaged integrator and the true solution. Which has the more accurate period (say, by measuring the time to the first minimum in y)? Which has the smaller energy drift (say, by measuring the change in depth between subsequent minima)?*

Fidelity is often far more important than accuracy in numerical simulations. Having an algorithm that has a small time-step error but gets the behavior qualitatively wrong is less useful than a cruder answer that is faithful to the physics. Leapfrog here is capturing the oscillating behavior far better than vastly more sophisticated algorithms, even though it gets the period wrong.

How does Leapfrog do so well? Systems like particle motion in a potential are Hamiltonian systems. They not only conserve energy, but they also have many other striking properties like conserving phase-space volume (Liouville's theorem, the basis for statistical mechanics). Leapfrog, in disguise, is also a Hamiltonian system. (Eqns N3.20 can be viewed as a composition of two canonical transformations—one advancing the velocities at fixed positions, and one advancing the positions at fixed velocities.) Hence it exactly conserves an approximation to the energy—and thus doesn't suffer from energy drift, satisfies Liouville's theorem, etc. Leapfrog and the related Verlet algorithm are called *symplectic* because they conserve the *symplectic form* that mathematicians use to characterize Hamiltonian systems.

It is often vastly preferable to do an exact simulation (apart from rounding errors) of an approximate system, rather than an approximate analysis of an exact system. That way, one can know that the results will be physically sensible (and, if they are not, that the bug is in your model or implementation, and not a feature of the approximation).

[12] Warning: for subtle reasons, the errors in Leapfrog apparently build up quadratically as one increases the number of time steps, so if your error estimate after one time step is h^n the error after $N = T/h$ time steps can't be assumed to be $Nh^n \sim h^{n-1}$, but is actually $N^2h^n \sim h^{N-2}$.

Chapter 4

Quantum

These exercises were designed for a graduate quantum mechanics course at Cornell.

N4.1 **Quantum notation.** (Notation) ②

For each entry in the first column, match the corresponding entries in the second and third column.

$$
\begin{array}{c}
\textit{Schrödinger} \\
\psi(x) \\
\psi^*(x) \\
|\psi(x)|^2 \\
\int \phi^*(x)\psi(x)\mathrm{d}x \\
|\int \phi^*(x)\psi(x)\mathrm{d}x|^2 \\
\int |\psi(x)|^2\mathrm{d}x \\
(\hbar/2m\mathrm{i})(\psi^*\nabla\psi - \psi\nabla\psi^*)
\end{array}
$$

Dirac
A. $\langle\psi|x\rangle p/2m\langle x|\psi\rangle - \langle x|\psi\rangle p/2m\langle\psi|x\rangle$
B. Ket $|\psi\rangle$ *in position basis, or* $\langle x|\psi\rangle$
C. Bra $\langle\psi|$ *in position basis, or* $\langle\psi|x\rangle$
D. $\langle\psi|x\rangle\langle x|\psi\rangle$
E. Braket $\langle\phi|\psi\rangle$
F. $\langle\psi|\psi\rangle$
G. $\langle\psi|\phi\rangle\langle\phi|\psi\rangle$

Physics
I. The number one
II. Probability density at x
III. Probability ψ is in state ϕ
IV. Amplitude of ψ at x
V. Amplitude of ψ in state ϕ
VI. Current density
VII. None of these.

N4.2 **Light proton atomic size.** (Dimensional analysis) ③

In this exercise, we examine a parallel world where the proton and neutron masses are equal to the electron mass, instead of ∼2,000 times larger.

In solving the hydrogen atom in your undergraduate quantum course, you may have noted that in going from the 6-dimensional electron-proton system into the three-dimensional center-of-mass coordinates, the effective mass gets shifted to a *reduced mass* $m_{\text{red}} = \mu = 1/(1/m_e + 1/m_{\text{nucleus}})$, and is otherwise the hydrogen potential with a fixed (infinite-mass) nucleus. Let us assume that the atomic sizes and the excitation energies are determined solely by this mass shift.

What is the reduced mass for the hydrogen atom in the parallel world of light protons, compared to the electron mass? How much larger will the atom be? How much will the binding energy of the atom change? (You may approximate $M_p \sim \infty$ when appropriate.) (Units hint: $[\hbar] = ML^2/T$, $[ke^2] = Energy * L = ML^3/T^2$, and $[m_e] = M$. Here $k = 1$ in CGS units, and $k = 1/(4\pi\epsilon_0)$ in SI units.)

N4.3 **Light proton tunneling.** (Dimensional analysis) ③

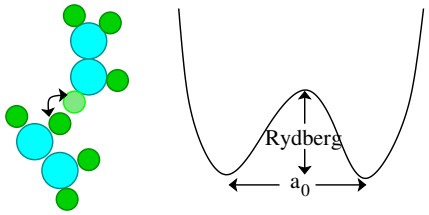


Fig. N4.1 Atom tunneling. A hydrogen atom tunnels a distance a_0, breaking a bond of strength E_{bind} equal to its ionization energy.

In this exercise, we continue to examine a parallel world where the proton and neutron masses are equal to the electron mass, instead of ∼2,000 times larger.

With everything two thousand times lighter, will atomic tunneling become important? Let's make a rough estimate of the tunneling suppression (given by the approximate WKB formula $\exp(-\sqrt{2MVQ}/\hbar)$).

Imagine an atom hopping between two positions, breaking and reforming a chemical bond in the process—an electronic energy barrier, and an electronic-scale distance. The

distance will be some fraction of a Bohr radius a_0 and the barrier energy will be some fraction of a Rydberg, but the the atomic mass would be some multiple of the proton mass.

In our world, what would the suppression factor be for an hydrogen atom of mass $\sim M_p$ *tunneling through a barrier of height* V *of one Rydberg* $= \hbar^2/(2m_e a_0^2)$*, and width* Q *equal to the Bohr radius* a_0*? How would this change in the parallel world where* $M_p \to m_e$*? (Simplify your answer as much as possible.)* (Use the real-world[1] a_0 and Rydberg for the parallel world, not your answers from a previous exercise. Also please use the simple formula above: don't do the integral. Your answer should involve only two of the fundamental constants.)

N4.4 **Light proton superfluid.** (Quantum) ③

In this exercise, we yet again examine a parallel world where the proton and neutron masses are equal to the electron mass, instead of ∼2,000 times larger.

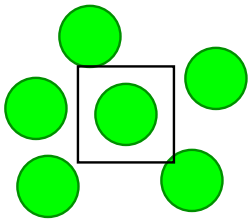


Fig. N4.2 Atoms with imaginary box of size equal to average space per atom

Atoms Bose condense when the number density n is high and the temperature T is low. One can view this condensation point as where temperature and the confinement energy become competitive. Up to a constant of order one, the thermal energy at the transition temperature $k_B T_c$ equals the confinement energy needed to put a helium atom in a box with impenetrable walls of length the mean nearest-neighbor spacing $L = n^{-1/3}$.

Calculate the ratio of the confinement energy $E_{\mathrm{He,ours}}$ *of Helium in our world and water* $E_{\mathrm{H_2O,light}}$ *in the parallel world.* Helium goes superfluid at $\sim 2°K$ in our universe. *From that, estimate the superfluid transition temperature* T_c *for water in the parallel universe. Will it be superfluid at room temperature?* (Assume the box sizes for He in our world and water in the parallel world are the same.[2] Room temperature is about $300°K$. Helium 4, the isotope that goes superfluid, has two neutrons and two protons, with roughly equal masses. Oxygen 16 has eight protons, eight neutrons, and eight electrons.)

[1] The reduced mass effects you found in the earlier exercise will be much less important for larger atoms and molecules, so we shall not include them here.

[2] The number of molecules of water per unit volume is comparable to the number of atoms of liquid helium per unit volume in the real world, and the water molecule will stay about the same size in the parallel world.

N4.5 **Aharonov-Bohm Wire.** (Parallel transport) ③

What happens to the electronic states in a thin metal loop as a magnetic flux Φ_B is threaded through it? This was a big topic in the mid-1980's, with experiments suggesting that the loops would develop a spontaneous current, that depended on the flux Φ_B/Φ_0, with $\Phi_0 = hc/e$ the quantum of flux familiar from the Bohm-Aharnov effect. In particular, Nandini Trivedi worked on the question while she was a graduate student here:

> Nandini Trivedi and Dana Browne, "Mesoscopic ring in a magnetic field: Reactive and dissipative response", *Phys. Rev. B* **38**, 9581-9593 (1988);

she's now a faculty member at Ohio State.

Some of the experiments clearly indicated that the periodicity in the current went as $\Phi_0/2 = hc/2e$—half the period demanded by Bohm and Aharonov from fundamental principles. (This is OK; having a *greater* period would cause one to wonder about fractional charges.) Others found (noisier) periods of Φ_0. Can we do a free-particle-on-a-ring calculation to see if for some reason we get half the period too?

Consider a thin wire of radius R along $x^2+y^2 = R^2$. Let a solenoid containing magnetic flux Φ_B, thin compared to R, lie along the $\hat{z}$ axis. Let ϕ be the angle around the circle with respect to the positive x-axis. (Don't confuse the flux Φ_B with the angle ϕ!) We'll assume the wire confines the electron to lie along the circle, so we're solving a one-dimensional Schrödinger's equation along the coordinate $s = R\phi$ around the circle. Assume the electrons experience a random potential $V(s)$ along the circumference $C = 2\pi R$ of the wire.

(a) Choose a gauge for $\vec{\mathbf{A}}$ so that it points along $\hat{\phi}$. What is the one-dimensional time-independent Schrödinger equation giving the eigenenergies for electrons on this ring? What is the boundary conditions for the electron wavefunction ψ at $s = 0$ and $s = C$? (Hint: the wire is a circle; nothing fancy yet. I'm not asking you to solve the equation—only to write it down.)

Deriving the Bohm-Aharonov effect using Schrödinger's equation is easiest done using a singular gauge transformation.

(b) Consider the gauge transformation $\Lambda(r, \phi) = -\phi\Phi_B/(2\pi)$. Show that $\vec{\mathbf{A}}' = \vec{\mathbf{A}} + \nabla\Lambda$ is zero along the wire for $0 < s < C$, so that we are left with a zero-field Schrödinger equation. What happens at the endpoint C? What is the new boundary condition for the electron wave function ψ' after this gauge transformation? Does the effect vanish for $\Phi_B = n\Phi_0$ for integer n, as the Bohm-Aharonov effect says it should?

Realistically, the electrons in a large, room-temperature wire get scattered by phonons or electron-hole pairs (effectively, a quantum measurement of sorts) long before they propagate around the whole wire, so these effects were only seen experimentally when the wires were cold (to reduce phonons and electron-hole pairs) and mesoscopic (tiny, so the scattering length is comparable to or larger than the circumference).

Finally, let's assume free electrons, so $V(s) = 0$. What's more, to make things simpler, let's imagine that there is only one electron in the wire.

(c) Ignoring the energy needed to confine the electrons into the thin wire, solve the one-dimensional Schrödinger equation to give the ground state of the electron as a function of Φ_B. Plot the current in the wire as a function of Φ_B. Is it periodic with period Φ_0, or periodic with period $\Phi_0/2$?

In the end, it was determined that there were two classes of experiments. Those that measured many rings at once (measuring an average current, an easier experiment) got periodicity of $hc/2e$, while those that attempted the challenge of measuring one mesoscopic ring at a time find hc/e.

N4.6 **Anyons.** (Statistics) ③

Frank Wilczek, "Quantum mechanics of fractional-spin particles", *Phys. Rev. Lett.* **49**, 957 (1982).
Steven Kivelson, Dung-Hai Lee, and Shou-Cheng Zhang, "Electrons in Flatland", Scientific American, March 1996.

In quantum mechanics, identical particles are truly indistinguishable (Fig. N4.3). This means that the wavefunction for these particles must return to itself, up to an overall phase, when the particles are permuted:

$$\Psi(\mathbf{r}_1, \mathbf{r}_2, \cdots) = \exp(\mathrm{i}\chi)\Psi(\mathbf{r}_2, \mathbf{r}_1, \cdots). \tag{N4.1}$$

where $\cdots$ represents potentially many other identical particles.

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. On the 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.[3]

[3]This idea is due to Feynman's thesis advisor, John Archibald Wheeler. As Feynman quotes in his Nobel lecture, *I received a telephone call one day at the graduate college at Princeton from Professor Wheeler, in which he said, "Feynman, I know why all electrons have the same charge and the same mass." "Why?" "Because, they are all the same electron!" And, then he explained on the telephone, "suppose that the world lines which we were ordinarily considering before in time and space - instead of only going up in time were a tremendous knot, and then, when we cut through the knot, by the plane corresponding to a fixed time, we would see many, many world lines and that would represent many electrons, except for one thing. If in one section this is an ordinary electron world line, in the section in which it reversed itself and is coming back from the future we have the wrong sign to the proper time - to the proper four velocities - and that's equivalent to changing the sign of the charge, and, therefore, that part of a path would act like a positron."*

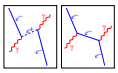


Fig. N4.3 Feynman diagram: identical particles.

In three dimensions, χ must be either zero or π, corresponding to bosons and fermions. In two dimensions, however, χ can be anything: *anyons* are possible! Let's see how this is possible.

In a two-dimensional system, consider changing from coordinates $\mathbf{r}_1, \mathbf{r}_2$ to the center-of-mass vector $\mathbf{R} = (\mathbf{r}_1 + \mathbf{r}_2)/2$, the distance between the particles $r = |\mathbf{r}_2 - \mathbf{r}_1|$, and the angle ϕ of the vector between the particles with respect to the $\hat{x}$ axis. Now consider permuting the two particles counter-clockwise around one another, by increasing ϕ at fixed r. When $\phi = 180° \equiv \pi$, the particles have exchanged positions, leading to a boundary condition on the wavefunction

$$\Psi(\mathbf{R}, r, \phi, \cdots) = \exp(\mathrm{i}\chi)\Psi(\mathbf{R}, r, \phi + \pi, \cdots). \tag{N4.2}$$

Permuting them counter-clockwise (backward along the same path) must then[4] give $\Psi(\mathbf{R}, r, \phi, \cdots) = \exp(-\mathrm{i}\chi)\Psi(\mathbf{R}, r, \phi - \pi, \cdots)$. This in general makes for a many-valued wavefunction (similar to Riemann sheets for complex analytic functions).

Why can't we get a general χ in three dimensions?

(a) *Show, in three dimensions, that* $\exp(\mathrm{i}\chi) = \pm 1$*, by arguing that a counter-clockwise rotation and a clockwise rotation must give the same phase.* (Hint: The phase change between ϕ and $\phi + \pi$ cannot change as we wiggle the path taken to swap the particles, unless the particles hit one another during the path. Try rotating the counter-clockwise path into the third dimension: can you smoothly change it to clockwise? What does that imply about $\exp(\mathrm{i}\chi)$?)

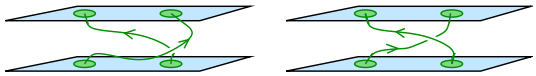


Fig. N4.4 Braiding of paths in two dimensions. In two dimensions, one can distinguish swapping clockwise from counter-clockwise. Particle statistics are determined by representations of the *Braid group*, rather than the permutation group.

Figure N4.4 illustrates how in two dimensions rotations by π and $-\pi$ are distinguishable; the trajectories form *braids* that wrap around one another in different ways. You can't change from a counter-clockwise braid to a clockwise braid without the braids crossing (and hence the particles colliding).

[4]The phase of the wave-function doesn't have to be the same for the swapped particles, but the gradient of the phase of the wavefunction is a physical quantity, so it must be minus for the counter-clockwise path what it was for the clockwise path.

An angular boundary condition multiplying by a phase should seem familiar: it is quite similar to that of the Bohm-Aharonov effect we studied in exercise 2.4. Indeed, we can implement fractional statistics by producing *composite particles*, by threading a magnetic flux tube of strength Φ through the center of each 2D boson, pointing out of the plane.

(b) *Remind yourself of the Bohm-Aharonov phase incurred by a particle of charge e encircling counter-clockwise a tube of magnetic flux* Φ*. If a composite particle of charge e and flux* Φ *encircles another identical composite particle, what will the net Bohm-Aharonov phase be?* (Hint: You can view the moving particle as being in a fixed magnetic field of all the other particles. The moving particle doesn't feel its own flux.)

(c) *Argue that the phase change* $\exp(i\chi)$ *upon swapping two particles is exactly half that found when one particle encircles the other. How much flux is needed to turn a boson into an anyon with phase* $\exp(i\chi)$*?* (Hint: The phase change can't depend upon the precise path, so long as it braids the same way. It's *homotopically invariant*, see Chapter 9 of "Entropy, Order Parameters, and Complexity".)

Anyons are important in the quantum Hall effect. What is the quantum Hall effect? At low temperatures, a two-dimensional electron gas in a perpendicular magnetic field exhibits a Hall conductance that is quantized, when the *filling fraction* ν (electrons per unit flux in units of Φ_0) passes near integer and rational values.

Approximate the quantum Hall system as a bunch of composite particles made up of electrons bound to flux tubes of strength Φ_0/ν. As a perturbation, we can imagine later relaxin the binding and allow the field to spread uniformly.[5]

(d) **Composite bosons and the integer quantum Hall effect.** *At filling fraction* $\nu = 1$ *(the integer quantum Hall state), what are the effective statistics of the composite particle? Does it make sense that the (ordinary) resistance in the quantum Hall state goes to zero?*

- The excitations in the *fractional* quantum Hall effect are anyons with fractional charge. (The $\nu = 1/3$ state has excitations of charge $e/3$, like quarks, and their wavefunctions gain a phase $\exp(i\pi/3)$ when excitations are swapped.)
- It is conjectured that, at some filling fractions, the quasiparticles in the fractional quantum Hall effect have *nonabelian* statistics, which could become useful for quantum computation.
- The composite particle picture is a centeral tool both conceptually and in calculations for this field.

[5]This is not nearly as crazy as modeling metals and semiconductors as noninteracting electrons, and adding the electron interactions later. We do that all the time—electons and holes in solid-state physics, 1s, 2s, 2p electrons in multi-electron atoms, all have obvious meanings only if we ignore the interactions. Both the composite particles and the noninteracting electron model are examples of how we use *adiabatic continuity*—you find a simple model you can solve, that can be related to the true model by turning on an interaction.

N4.7 **Bell.**[6] (Quantum, Qbit) ③

Consider the following cooperative game played by Alice and Bob: Alice receives a bit x and Bob receives a bit y, with both bits uniformly random and independent. The players win if Alice outputs a bit a and Bob outputs a bit b, such that $(a+b = xy) mod 2$. They can agree on a strategy in advance of receiving x and y, but no subsequent communication between them is allowed.

(a) *Give a deterministic strategy by which Alice and Bob can win this game with 3/4 probability.*

(b) *Show that no deterministic strategy lets them win with more than 3/4 probability. (Note that Alice has four possible deterministic strategies* $[0, 1, x, \sim x]$*, and Bob has four* $[0, 1, y, \sim y]$*, so there's a total of 16 possible joint deterministic strategies.)*

(c) *Show that no probabilistic strategy lets them win with more than 3/4 probability. (In a probabilistic strategy, Alice plays her possible strategies with some fixed probabilities* $p_0, p_1, p_x, p_{\sim x}$*, and similarly Bob plays his with probabilities* $q_0, q_1, q_y, q_{\sim y}$*.)*

The upper bound of <= 75% of the time that Alice and Bob can win this game provides, in modern terms, an instance of the Bell inequality, where their prior cooperation encompasses the use of any local hidden variable.

Let's see how they can beat this bound of 3/4, by measuring respective halves of an entangled state, thus quantum mechanically violating the Bell inequality.[7]

Suppose Alice and Bob share the entangled state $\frac{1}{\sqrt{2}}(|\uparrow\rangle_\ell|\uparrow\rangle_r + |\downarrow\rangle_\ell|\downarrow\rangle_r)$, with Alice holding the left Qbit and Bob holding the right Qbit. Suppose they use the following strategy: if $x = 1$, Alice applies the unitary matrix $R_{\pi/6} = \begin{pmatrix} \cos\frac{\pi}{6} & -\sin\frac{\pi}{6} \\ \sin\frac{\pi}{6} & \cos\frac{\pi}{6} \end{pmatrix}$ to her Qbit, otherwise doesn't, then measures in the standard basis and outputs the result as a. If $y = 1$, Bob applies the unitary matrix $R_{-\pi/6} = \begin{pmatrix} \cos\frac{\pi}{6} & \sin\frac{\pi}{6} \\ -\sin\frac{\pi}{6} & \cos\frac{\pi}{6} \end{pmatrix}$ to his Qbit, otherwise doesn't, then measures in the standard basis and outputs the result as b. (Note that if the Qbits were encoded in photon polarization states, this would be equivalent to Alice and Bob rotating measurement devices by $\pi/6$ in inverse directions before measuring.)

(d) *Using this strategy: (i) Show that if* $x = y = 0$*, then Alice and Bob win the game with probability 1.*
(ii) Show that if $x = 1$ *and* $y = 0$ *(or vice versa), then Alice and Bob win with probability 3/4.*

[6] This exercise was developed by Paul Ginsparg, based on an example by Bell '64 with simplifications by Clauser, Horne, Shimony, & Holt ('69).

[7] There's another version for GHZ state, where three people have to get a+b+c mod 2 = x or y or z. Again one can achieve only 75% success classically, but they can win *every* time sharing the right quantum state

(iii) Show that if $x = y = 1$, then Alice and Bob win with probability 3/4.
(iv) Combining parts (i)–(iii), conclude that Alice and Bob win with greater overall probability than would be possible in a classical universe.

This proves an instance of the CHSH/Bell Inequality, establishing that "spooky action at a distance" cannot be removed from quantum mechanics. Alice and Bob's ability to win the above game more than 3/4 of the time using quantum entanglement was experimentally confirmed in the 1980s (A. Aspect et al.).[8]

(e) *(Bonus) Consider a slightly different strategy, in which before measuring her half of the entangled pair Alice does nothing or applies $R_{\pi/4}$, according to whether x is 0 or 1, and Bob applies $R_{\pi/8}$ or $R_{-\pi/8}$, according to whether y is 0 or 1. Show that this strategy does even better than the one analyzed in a–c, with an overall probability of winning equal to $\cos^2 \pi/8 = (1 + \sqrt{1/2})/2 \approx .854$.*
(Extra bonus) Show this latter strategy is optimal within the general class of strategies in which before measuring Alice applies R_{α_0} or R_{α_1}, according to whether x is 0 or 1, and Bob applies R_{β_0} or R_{β_1}, according to whether y is 0 or 1.

This will demonstrate that no local hidden variable theory can reproduce all predictions of quantum mechanics for entangled states of two particles.

N4.8 **Parallel Transport, Frustration, and the Blue Phase.** (Liquid crystals) ③

"Relieving Cholesteric Frustration: The Blue Phase in a Curved Space," J. P. Sethna, D. C. Wright and N. D. Mermin, *Phys. Rev. Lett.* **51**, 467 (1983).

"Frustration, Curvature, and Defect Lines in Metallic Glasses and the Cholesteric Blue Phases," James P. Sethna, *Phys. Rev. B* **31**, 6278 (1985).

"Frustration and Curvature: the Orange Peel Carpet", http://www.lassp.cornell.edu/sethna /FrustrationCurvature/

"The Blue Phases, Frustrated Liquid Crystals and Differential Geometry", http://www.lassp.cornell.edu/sethna /LiquidCrystals/BluePhase/BluePhases.html.

(Optional: for those wanting a challenge.) Both the Aharonov Bohm effect and Berry's phase (later) are generalizations of the idea of *parallel transport*. Parallel transport, from differential geometry, tells one how to drag tangent vectors around on smooth surfaces. Just as we discovered that dragging the phase of a wavefunction around a closed loop in space gave it a net rotation due to the magnetic field (Aharonov-Bohm),

[8] Ordinarily, an illustration of these inequalities would appear in the physics literature not as a game but as a hypothetical experiment. The game formulation is more natural for computer scientists, who like to think about different parties optimizing their performance in various abstract settings. As mentioned, for physicists the notion of a classical strategy is the notion of a hidden variable theory, and the quantum strategy involves setting up an experiment whose statistical results could not be predicted by a hidden variable theory.

the phase can rotate also when the Hamiltonian is changed around a closed curve in Hamiltonian space (Berry's phase). Here we discuss how vectors rotate as one drags them around closed loops, leading us to the *curvature tensor*.

(a) **Parallel transport on the sphere.** *Imagine you're in St. Louis (longitude 90° W, latitude $\sim 40°$ N), pointing north. You walk around a triangle, first to the North Pole, then take a right-angle turn, walk down through Greenwich, England and Madrid, Spain (longitude $\sim 0°$ W) down to the equator, turn right, walk back to longitude 90° W along the equator, turn north, and walk back to St. Louis. All during the walk, you keep pointing your finger in the same direction as far as feasible (i.e., straight ahead on the first leg, off to the left on the second, and so on). What angle does the vector formed by your final pointing finger make with respect to your original finger? If you turned west in Madrid and walked along that latitude ($\sim 40°$ N) to St. Louis (yes, it is just as far north), would the angle be the same? (Hint: what about transport around a tiny quarter circle centered at the north pole?)*

Lightning Intro to Differential Geometry. Parallel transport of vectors on surfaces is described using a *covariant derivative* $(D_i v)^j = \partial_i v^j + \Gamma^j_{ik} v^k$ involving the Christoffel symbol $\Gamma^\mu_{i\nu}$. The amount a vector v^ν changes when taken around a tiny loop $\Delta s \Delta t(-\Delta s)(-\Delta t)$ is given by the Riemannian curvature tensor and the area of the loop

$$v'^i - v^i = R^i_{jk\ell} v^j \Delta s^k \Delta t^\ell. \tag{N4.3}$$

The algebra gets messy on the sphere, though (spherical coordinates are pretty ugly).

Instead, we'll work with a twisty kind of parallel transport, that my advisor and I figured out describe how molecules of the Blue Phase like to locally align. These long, thin molecules locally line up with their axes in a direction $\mathbf{n}(\mathbf{r})$, and are happiest when that axis twists so as to make zero the covariant derivative

$$(D_i n)^j = \partial_i n^j - q\epsilon_{ijk} n_k, \tag{N4.4}$$

where ϵ_{ijk} is the totally antisymmetric tensor[9] with three indices (Fig. N4.5 and N4.6). Since the Blue Phases are in flat space, you don't need to worry about the difference between upper and lower indices.

[9] Hence $\Gamma^j_{ik} = -q\epsilon_{ijk}$. Warning: this connection has *torsion*: it is not symmetric under interchange of the two bottom indices. This is what allows it to have a nonzero curvature even when the liquid crystal lives in flat space. But it will make some of the traditional differential geometry formulas incorrect, which usually presume zero torsion.

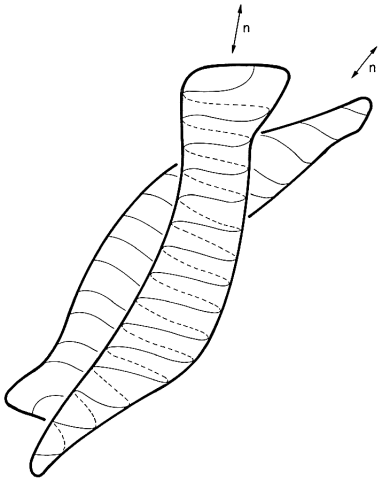


Fig. N4.5 Blue Phase Molecules, long thin molecules aligned along an axis $\hat{\mathbf{n}}(\mathbf{r})$, like to sit at a slight angle with respect to their neighbors. To align the threads and grooves on the touching surface between the molecules demands a slight twist. This happens because, like threaded bolts or screws, these molecules are *chiral*.

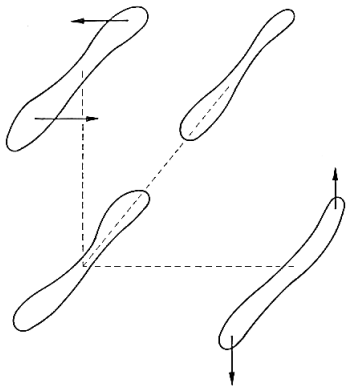


Fig. N4.6 Twisted Parallel Transport. The natural parallel transport, $D_i n_j = \partial_i n_j - q\epsilon_{ijk} n_k$, twists perpendicular to the long axis, but doesn't twist when moving along the axis.

(b) What direction will a molecule with $\mathbf{n} = \hat{x}$ *like to twist as one moves along the x-direction? The y-direction? The z-direction? Is the local low energy structure more like that of Fig. N4.7(a) (the low-temperature state of this system), or that of Fig. N4.7(b)?*

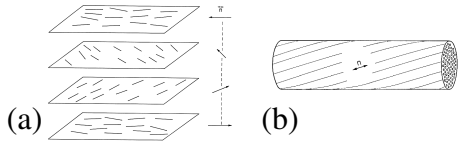


Fig. N4.7 Local structures of cholesteric liquid crystals. (a) Helical. (b) Tube.

So, what's the curvature tensor (eqn N4.14) for the blue phase? I figured out at the time that it came out to be

$$R_{ijk\ell} = q^2(\delta_{i\ell}\delta_{kj} - \delta_{ik}\delta_{\ell j}). \tag{N4.5}$$

This curvature means that the blue phases are *frustrated*; you can't fill space everywhere with material that's "happy", with $D\mathbf{n} = \mathbf{0}$.

(c) If $\mathbf{n}$ *starts at* $\hat{x}$*, and we transport along the tiny loop* $\{\Delta x, \Delta y, -\Delta x, -\Delta y\}$*, calculate the vector* $\mathbf{n}'$ *as it returns to the origin using the curvature tensor (eqns N4.14 and N4.16). according to eqn N4.14. Is the shift in* $\mathbf{n}$ *along the same direction we observe for the larger loop in Fig. N4.8?*

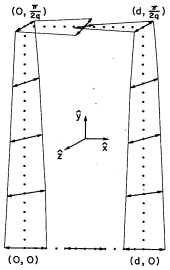


Fig. N4.8 Parallel transport frustration. We can check your answer to part (c) qualitatively by thinking about a larger loop (Fig. N4.8). Consider starting from two places separated $\Delta x = d$ apart along the axis of a double-twisting region (the center of the tube in Fig. N4.7(b)). If we move a distance $\Delta y = \pi/(2q)$ radially outward, the orientation of both will now point along $\hat{z}$. Transporting them each inward by $d/2$ then will cause a further twist.

We can use the curvature tensor of the Blue Phase to calculate a scalar curvature $R_{ijij} = -6q^2$. Thus the blue phases are negatively curved, even though they live in flat space. We looked to see what would happen if we put the blue phases into a positively curved space. Picking the sphere in four dimensions with the correct radius, we could make the curvature (and the frustration) go away, and find the ideal template for the blue phases. We think of the real blue phase as pieces of this ideal template, cut, flattened, and sewn together to fill space, like an orange-peel carpet (Fig. N4.9).

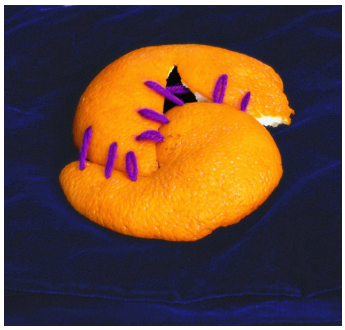


Fig. N4.9 Orange-peel carpet. (Copyright Pamela Davis Kivelson)

N4.9 **Crystal field theory: d-orbitals.** (Group reps) ③

The vector space of functions $f(x, y, z)$ on the unit sphere transforms into itself under rotations $f(\mathbf{x}) \to_R f(R^{-1}\mathbf{x})$. These transformations are linear ($af(\mathbf{x}) + g(\mathbf{x}) \to_R af(R^{-1}\mathbf{x}) + g(R^{-1}\mathbf{x})$), and obey the group composition rule, and thus form a representation of the rotation group.

(a) *Argue that the homogeneous polynomials of degree ℓ,*

$$f(x, y, z) = \sum_{m=0}^{\ell} \sum_{n=0}^{\ell-m} f_{\ell mn} x^m y^n z^{\ell-m-n} \tag{N4.6}$$

form a subspace that is invariant under rotations.

Thus the irreducible representations are contained in these invariant subspaces. Sakurai indeed mentions in his section 3.11 on tensor operators that $Y_1^0 = \sqrt{3/\pi}z/r$ and $Y_1^{\pm 1} = \sqrt{3/2\pi}(x \pm \mathrm{i}y)/r$, and also gives a formula for $Y_2^{\pm 2}$; since $r = 1$ on our unit sphere these are homogeneous polynomials of degree one.

(b) *Look up the $\ell = 2$ spherical harmonics (e.g. in Sakurai's appendix B) and write them as quadratic polynomials in x, y, and z.*

The $\ell = 2$ spherical harmonics are the angular parts of the wavefunctions for electrons in d orbitals (e.g. of transition metal atoms).[10] Electrons in d-orbitals are much more

[10] Here we use the common independent-electron language, where the complex many-body wavefunction of an atom, molecule, or solid is viewed as filling single-electron states, even though the electron-electron repulsion is almost as strong as the electron-nuclear attraction. This idea can be dignified in three rather different ways. First, one can view each electron as feeling an effective potential given by the nucleus plus the average density of electrons. This leads to mean-field Hartree-Fock theory. Second, one can show that the ground state energy can be written as an unknown functional of the electron density (the *Hohenberg-Kohn theorem*, and then calculate the kinetic energy terms as an effective single-body Schrödinger equation in the resulting effective potential due to the net electron density (the *Kohn-Sham equations*). Third, one can start with independent electrons (or Hartree-Fock electrons) and slowly "turn on" the electron-electron repulsion. The independent-electron excited eigenstates develop lifetimes and become *resonances*. For atoms these lifetimes represent Auger decay rates. For crystals these resonances are called *quasiparticles* and the theory is called *Landau Fermi liquid theory*. Landau Fermi-liquid theory is usually derived using Green's functions and Feynman diagrams, but it has recently been re-cast as a renormalization-group flow.

tightly contained near the nucleus than p and s orbitals. In molecules and solids, the s and p orbitals usually hybridize (superimpose) into chemical bonds and broad electron bands, where the original orbitals are strongly distorted. In contrast, d-electrons rarely participate in chemical bonds, and their electron bands are narrow—almost undistorted orbitals with small hopping rates. The energy levels of the five d-orbitals are, however, shifted from one another by their environments. (For crystals, these shifts are called *crystal field splittings.*)

We can use group representation theory to understand how the d-orbitals are affected by their molecular or crystalline environment.

First, we need to calculate the character $\chi(R) = \chi(\hat{n}, \phi)$ of the $\ell = 2$ representation. Remember that the character is the trace of the (here 5×5) matrix corresponding to the rotation. Remember that this trace depends only on the *conjugacy class* of R—that is, if S is some other group element then $\chi(S^{-1}RS) = \chi(R)$. Remember that any two rotations by the same angle ϕ are conjugate to one another.[11]

In class, we found that the character of the $\ell = 1$ representation by using the Cartesian x, y, z basis, where $R_{\hat{z}}(\phi) = \begin{pmatrix} \cos(\phi) & -\sin(\phi) & 0 \\ \sin(\phi) & \cos(\phi) & 0 \\ 0 & 0 & 1 \end{pmatrix}$. Hence $\chi^{(1)}(\phi) = 1 + 2\cos(\phi)$. We can do this same calculation in the m_z basis of the spherical harmonics, where Y_1^0 is unchanged under rotations and $Y_1^{\pm 1} \to \exp(\pm \mathrm{i}\phi) Y_1^{\pm 1}$. Here $R_{\hat{z}}(\phi) = \begin{pmatrix} \exp(\mathrm{i}\phi) & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \exp(-\mathrm{i}\phi) \end{pmatrix}$, and again $\chi^{(1)}(\phi) = 1 + 2\cos(\phi)$.

(c) *Calculate $\chi^{(2)}(\phi)$. Give the characters for rotations C_n by $2\pi/n$, for $n = 1$, 2, 3, and 4 (the important rotations for crystalline symmetry groups.)*

The most common symmetry groups for d-electron atoms in crystals is O, the octahedral group (the symmetry group of a cube). Look up the character tables for the irreducible representations of this finite group. (To simplify the calculation, we'll assume that inversion symmetry is broken; otherwise we should use O_h, which has twice the number of group elements.)

(d) *Use the orthogonality relations of the characters of irreducible representations for O, decompose the $\ell = 2$ representation above into irreducible representations of the octahedral group. How will the energies of the single-particle d-orbitals of a transition metal atom split in an octahedral environment? What will the degeneracies and symmetries (A_1, A_2, E, ...) be of the different levels?* (Hint: If you are doing it right, the dot product of the characters should equal a multiple of the number of octahedral group elements $o(O) = 24$, and the dimensions of the sub-representations should add up to five.)

The five d-orbitals are often denoted d_{xy}, d_{xz}, d_{yz}, d_{z^2} and $d_{x^2-y^2}$. This is a bit of a cheat; really d_{z^2} should be written $d_{2z^2-x^2-y^2}$ or something like that.

(e) *Figure out which of these orbitals are in each of the two representations you found*

[11] If the two rotations have axes $\hat{n}$ and $\hat{n}'$, choose S to rotate $\hat{n}'$ into $\hat{n}$.

in part (d). (Hint: Check how these five orbitals transform under the octahedral symmetries that permute x, y, and z among themselves.)

N4.10 **Entangled Spins.** (Spins) ③

In class, we studied the entanglement of the singlet spin state $|\mathcal{S}\rangle = (1/\sqrt{2})(|\uparrow\rangle_\ell|\downarrow\rangle_r - |\downarrow\rangle_\ell|\uparrow\rangle_r)$ of electrons of a diatomic molecule as the atoms L and R are separated;[12] the spins on the two atoms are in opposite directions, but the system is in a superposition of the two choices. Here we discuss another such superposition, but with a different relative phase for the two choices:

$$|\chi\rangle = (1/\sqrt{2})(|\uparrow\rangle_\ell|\downarrow\rangle_r + |\downarrow\rangle_\ell|\uparrow\rangle_r) \tag{N4.7}$$

You should know from angular momentum addition rules that the space of product wavefunctions of two spin ½ states can be decomposed into a spin 1 and a spin 0 piece: $\frac{1}{2} \otimes \frac{1}{2} = 1 \oplus 0$. So there are two orthogonal eigenstates of S_z with eigenvalue zero: one of total spin zero and one of total spin one.

(a) **Total spin.** *Which state is which?* (If you don't know from previous work, calculate!) *Why do we call $|\mathcal{S}\rangle$ a singlet?*

Now, is the spin wavefunction compatible with what we know about electron wavefunctions?

(b) **Symmetry.** *When the two spins are exchanged, how does $|\chi\rangle$ change? If the total wavefunction $\Psi(x_L, s_L, x_R, s_L)$ is a product of this spin wavefunction $\chi(s_L, s_R)$ and and a two-particle spatial wavefunction $\psi(x_L, x_R)$, what symmetry must ψ have under interchange of electrons?*

We noted in class that two other spin-1 product states, $|\uparrow\rangle_\ell|\uparrow\rangle_r$ and $|\downarrow\rangle_\ell|\downarrow\rangle_r$ do not form entangled states when L and R separate. Is $|\chi\rangle$ like these spin-1 states, or is it entangled like $|\mathcal{S}\rangle$ is?

(c) **Entanglement.** *Give the Schmidt decomposition of $|\chi\rangle$. What are the singular values? What is the entanglement entropy?* (Hint: The steps should be very familiar from class.)

(d) **Singular value decomposition (SVD).** *Let M be the matrix which gives $|\chi\rangle$ as a product of left and right spin states:*

$$|\chi\rangle = \begin{pmatrix} \Uparrow\rangle_\ell & \Downarrow\rangle_\ell \end{pmatrix} M \begin{pmatrix} \Uparrow\rangle_r \\ \Downarrow\rangle_r \end{pmatrix}. \tag{N4.8}$$

What is M? Give an explicit singular value decomposition[13] $M = U\Sigma V^T$ of the matrix M. Explain how the SVD gives us the Schmidt decomposition of part (c).

[12] We assumed that, when separated, one electron is localized basically on each of the two atoms, and the spin kets are labeled based on the primary locale of the corresponding spatial wavefunction for that electron.

[13] Remember that the SVD guarantees that U and V have orthonormal columns, and Σ is a diagonal matrix whose diagonal elements σ_i are all positive and decreasing (so $\sigma_i \geq \sigma_{i+1} \geq 0$). There is some flexibility in the singular vectors (i.e., matched pairs can both have their signs changed), but the singular values are unique and hence a property of the matrix.

N4.11 **Square well ground state.**[14] (Quantum numeric) ②

Consider a particle of mass M in one dimension, confined in a potential that vanishes for $-a \le x \le a$ and becomes infinite at $x = \pm a$, so the wavefunction must vanish at $x = \pm a$.

Find the normalized ground state $\psi_0(x)$. Suppose that the particle is placed in a state ψ with a wave function proportional to $a^2 - x^2$, which vanished properly at the edges. If the energy of the particle is measured, what is the probability that the particle will be found in the state of lowest energy? Use a computational method (symbolic or numerical) to avoid tedious evaluation of integrals. *Evaluate your answer numerically (it should not depend on a). Also, plot the ground state and the trial state, using $a = 3$; they should look rather similar.* This exercise may take much longer than it seems, since you'll likely spend some time setting up your computational environment—installing software, getting plots to work, and so forth. The Hints files should help once the software is set up.

N4.12 **Evolving Schrödinger: Free particles and uncertainty.** (Computation) ③

Several of our computational exercises will involve solving the time-dependent Schrödinger equation for one-dimensional quantum systems. In this first of these exercises, we shall evolve the free particle Hamiltonian

$$\begin{aligned} \mathrm{i}\hbar\frac{\partial\psi}{\partial t} &= -\frac{\hbar^2}{2m}\frac{\partial^2\psi}{\partial x^2}\psi = H_{\text{kin}}\psi \\ \psi(x,t) &= U(t)\psi(0) = \mathrm{e}^{-\mathrm{i}H_{\text{kin}}t/\hbar}\psi(x,0) \end{aligned} \tag{N4.9}$$

In a later exercise, we shall solve for the behavior of a hydrogenic harmonic oscillator with angular frequency $\omega = 10^{12}$ radians/sec. We'll use the ground state of this harmonic oscillator as our initial condition for the free-particle evolution:

$$\psi_0(x) = \left(\frac{m\omega}{\pi\hbar}\right)^{1/4}\mathrm{e}^{-m\omega x^2/2\hbar} = \left(\frac{1}{2\pi a_0^2}\right)^{1/4}\mathrm{e}^{-x^2/4a_0^2}. \tag{N4.10}$$

Here $a_0 = \sqrt{\hbar/2m\omega}$ is the RMS width of the Gaussian probability distribution $|\psi(x)|^2$.

With no potential energy, we can solve for the motion of a free particle, $\widetilde{\psi}(k,t) = U_{\text{kin}}(t)\psi(k,t=0)$ using Fourier transforms.

$$\widetilde{\psi}(k,t) = \mathrm{e}^{-\mathrm{i}(\hbar k^2/2m)t}\widetilde{\psi}(k,t=0) \tag{N4.11}$$

$$\widetilde{\psi}(x,t) = IFFT[\mathrm{e}^{-\mathrm{i}(\hbar k^2/2m)t}FFT[\widetilde{\psi}(x,t=0)]] \tag{N4.12}$$

where FFT takes a Fourier transform of the wavepacket, and IFFT takes an inverse Fourier transform. (FFT stands for Fast Fourier Transform.)

[14] Adapted from Weinberg, *Lectures on Quantum Mechanics*, ex. 1.1. Hints files available on the course website for Mathematica, Python, and Octave/Matlab).

Let's numerically solve for the evolution of $\psi_0(x)$ from eqn N4.26. We'll evaluate it at a discrete set of $N_p = 200$ points spanning a distance $L = 80a_0$. With $\mathrm{d}x = L/N_p$, the points will be at $-L/2, L/2 + \mathrm{d}x, \ldots, L/2 - \mathrm{d}x$. In CGS units, the hydrogen mass is about the proton mass 1.672610^{-24} gm, $\hbar = 1.0545710^{-27}$ erg sec, and we decided $\omega = 10^{12}$ radians/sec.

(a) *Define ψ_0 on this grid. Plot your $|\psi_0|^2$, and check that it roughly has width a_0.*

The Fast Fourier Transform of ψ returns $\widetilde{\psi}(k)$ evaluated at points

$$k = (0, \mathrm{d}k, 2\mathrm{d}k, \ldots, -2\mathrm{d}k, -\mathrm{d}k), \tag{N4.13}$$

with $\mathrm{d}k = 2\pi/L$. (These correspond to the plane waves with period L; FFTs assume periodic boundary conditions.) The maximum value $k_{\max} = \pi/\mathrm{d}x$ happens in the *center* of the FFT. To do our time evolution, we need to define an array k^2 evaluated at these points, which should rise quadratically, come to a cusp, and then fall back to zero.

(b) *Define k^2 on this grid. Plot it.*

Now we can evolve $\psi(x,t) = U_{\mathrm{kin}}(t)\psi_0(x)$.

(c) *Create a routine that calculates $\psi(x,t)$ using eqn N4.28. Plot the real and imaginary parts of $\psi(x,t)$ at $t = P/4$, P, and $2P$ on separate plots, where $P = 2\pi/\omega$ is the period of the harmonic oscillator in the later exercise. What happens to the width of the wave packet?*

The Heisenberg uncertainty principle tells us that $\Delta x \Delta p \geq \hbar/2$. The Gaussian wavepacket we use is the form that minimizes this inequality, so we expect the packet width to grow like vt with v given by the momentum uncertainty $v \sim \Delta p/m$

(d) *Give the formula for v for our packet. Calculate $\Delta x = \sqrt{\langle x^2 \rangle}$ as a function of time, for points $0 \leq t \leq 2P$. Plot Δx versus time and vt versus time on the same plot. Why do they not agree at short times?*

N4.13 **Rotating Fermions.** (Group theory) ③

In this exercise, we'll explore the *geometry* of the space of rotations.

Spin 1/2 fermions transform upon rotations under $SU(2)$, the unitary 2×2 matrices with determinant one. Vectors transform under $SO(3)$, the ordinary 3×3 rotation matrices you know of.

Sakurai argues that a general $SU(2)$ matrix $U = \begin{pmatrix} a & b \\ -b^* & a^* \end{pmatrix}$ with $|a|^2 + |b|^2 = 1$. Viewing $\{\mathrm{Re}(a), \mathrm{Im}(a), \mathrm{Re}(b), \mathrm{Im}(b)\}$ as a vector in four dimensions, $SU(2)$ then geometrically is the unit sphere $\mathbb{S}^3$ in $\mathbb{R}^4$.

Remember that for every matrix R in $SO(3)$, there are two unitary matrices U and $-U$ corresponding to the same physical rotation. The matrix $-U$ has coordinates $(-a, -b)$—it is the antipodal point on $\mathbb{S}^3$, exactly on the opposite side of the sphere. So $SO(3)$ is geometrically the unit sphere with *antipodal points identified*. This is called (for obscure reasons) the *projective plane*, $\mathbb{RP}^3$.

Feynman's plate (in Feynman's plate trick) as it rotates 360° travels in rotation space from one orientation to its antipode. While I'm not sure anyone has figured out whether arms, shoulders, and elbows duplicate the properties of fermions under rotations, the plate motion illustrates the possibility of a minus sign.

But we can calculate this trajectory rather neatly by mapping the rotations not to the unit sphere, but to the space $\mathbb{R}^3$ of three-dimensional vectors. (Just as the 2-sphere $\mathbb{S}^2$ can be projected onto the plane, with the north pole going to infinity, so can the 3-sphere $\mathbb{S}^3$ be projected onto $\mathbb{R}^3$.) Remember the axis-angle variables, where a rotation of angle ϕ about an axis $\hat{\mathbf{n}}$ is given by

$$\exp(-\mathrm{i}\mathbf{S}\cdot\hat{\mathbf{n}}\phi/\hbar) = \exp(-\mathrm{i}\boldsymbol{\sigma}\cdot\hat{\mathbf{n}}\phi/2) = \exp(-\mathrm{i}\mathbf{J}\cdot\hat{\mathbf{n}}\phi/\hbar) \tag{N4.14}$$

where the middle formula works for $SU(2)$ (where $\mathbf{S} = \hbar\sigma/2$, because the particles have spin 1/2) and the last formula is appropriate for $SO(3)$.[15] We figure $\mathbf{n}$ will be the direction of the vector in $\mathbb{R}^3$ corresponding to the rotation, but how will the length depend on ϕ? Since all rotations by 360° are the same, it makes sense to make the length of the vector go to infinity as $\phi \to 360°$. We thus define the *Modified Rodrigues coordinates* for a rotation to be the vector $\mathbf{p} = \hat{\mathbf{n}}\tan(\phi/4)$.

(a) *Fermions, when rotated by 360°, develop a phase change of* $\exp(\mathrm{i}\pi) = -1$ *(as discussed in Sakurai & Napolitano p. 165, and as we illustrated with Feynman's plate trick). Give the trajectory of the modified Rodrigues coordinate for the fermion's rotation as the plate is rotated* 720° *about the axis* $\hat{\mathbf{n}} = \hat{z}$. (We want the continuous trajectory on the sphere $\mathbb{S}^3$, perhaps which passes through the point at ∞. Hint: The trajectory is already defined by the modified Rodrigues coordinate: just describe it.)

(b) *For a general Rodrigues point* $\mathbf{p}$ *parameterizing a rotation in* $SO(3)$*, what antipodal point* $\mathbf{p}'$ *corresponds to the same rotation?* (Hint: A rotation by ϕ and a rotation by $\phi + 2\pi$ should be identified.)

N4.14 **Lithium ground state symmetry.** (Quantum) ③

A simple model for heavier atoms, that's surprisingly useful, is to ignore the interactions between electrons (the *independent electron* approximation).[16]

$$\mathcal{H}^{\mathrm{Z}} = \sum_{i=1}^{Z} p_i^2/2m - k_\varepsilon Z e^2/r_i \tag{N4.15}$$

Remember that the eigenstates of a single electron bound to a nucleus with charge Z are the hydrogen levels ($\psi_n^{\mathrm{Z}} = \psi_{1s}^{\mathrm{Z}}, \psi_{2s}^{\mathrm{Z}}, \psi_{2p}^{\mathrm{Z}}, \dots$), except shrunken and shifted upward

[15] The factor of two is there for $SU(2)$ because the spin is 1/2. For $SO(3)$, infinitesimal generators are antisymmetric matrices, so $\mathbf{S}_{jk}^{(i)} = \mathbf{J}_{jk}^{(i)} = \mathrm{i}\hbar\epsilon_{ijk}$ in the xyz basis; in the usual quantum basis $m_z = (-1, 0, 1)$ the formula will be different.

[16] Here $k_\varepsilon = 1$ in CGS units, and $k_\varepsilon = 1/(4\pi\varepsilon_0)$ in SI units. We are ignoring the slight shift in effective masses due to the motion of the nucleus.

in binding energy (E^{Z} more negative):

$$\begin{aligned} \mathcal{H}^{\mathrm{Z}}\psi_n^{\mathrm{Z}} &= E_n^{\mathrm{Z}}\psi_n \\ \psi_n^{\mathrm{Z}}(\mathbf{r}) &= \psi_n^{\mathrm{H}}(\lambda_r \mathbf{r}) \\ E^{\mathrm{Z}} &= \lambda_E E^{\mathrm{H}} \end{aligned} \tag{N4.16}$$

(a) *By what factor λ_r do the wavefunctions shrink? By what factor λ_E do the energies grow?* (Hint: Dimensional arguments are preferred over looking up the formulas.)

In the independent electron approximation, the many-body electron eigenstates are created from products of single-electron eigenstates. The Pauli exclusion principle (which appears only useful in this independent electron approximation) says that exactly two electrons can fill each of the single-particle states.

(b) *Ignoring identical particle statistics, show that a product wavefunction*

$$\Psi(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \dots) = \psi_{n_1}^{\mathrm{Z}}(\mathbf{r}_1)\psi_{n_2}^{\mathrm{Z}}(\mathbf{r}_2)\psi_{n_3}^{\mathrm{Z}}(\mathbf{r}_3)\dots \tag{N4.17}$$

has energy $E = \sum_i E_{n_i}^{\mathrm{Z}}$.

The effect of the electron-electron repulsion in principle completely destroys this product structure. But for ground-state and excited-state quantum numbers, the language of filling independent electron orbitals is quite useful.[17] However, the energies of these states are strongly corrected by the interactions between the other electrons.

(c) *Consider the 2s and 2p states of an atom with a filled 1s shell (one electron of each spin in 1s states). Which state feels a stronger Coulomb attraction from the nucleus? Argue heuristically that the 2s state will generally have lower (more negative) energy and fill first.*

Let's check something I asserted, somewhat tentatively, in lecture. There I said that, for atoms with little spin-orbit coupling, the ground state wavefunction can be factored into a spatial and a spin piece:

$$\Psi(\mathbf{r}_1, s_1; \mathbf{r}_2, s_2; \mathbf{r}_3, s_3; \dots) \stackrel{?}{=} \psi(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3 \dots)\chi(s_1, s_2, s_3 \dots) \tag{N4.18}$$

We'll check this in the first nontrivial case—the lithium atom ground state, in the independent electron approximation. From part (c), we know that two electrons should occupy the 1s orbital, and one electron should occupy the 2s orbital. The two spins in the 1s orbital must be antiparallel; let us assume the third spin is pointing up $\uparrow_3$:

$$\Psi^0(\mathbf{r}_1, s_1; \mathbf{r}_2, s_2; \mathbf{r}_3, s_3) = \psi_{1s}^{\mathrm{Li}}(\mathbf{r}_1)\psi_{1s}^{\mathrm{Li}}(\mathbf{r}_2)\psi_{2s}^{\mathrm{Li}}(\mathbf{r}_3)\uparrow_1\downarrow_2\uparrow_3 . \tag{N4.19}$$

[17] The excited states of an atom aren't energy eigenstates, they are *resonances*, with a finite lifetime. If you think of starting with the independent electron eigenstates and gradually turning on the Coulomb interaction and the interaction with photons, the true ground state and the resonances are adiabatic continuations of the single-particle product eigenstates—inheriting their quantum numbers.

But this combination is not antisymmetric under permutations of the electrons.

(d) *Antisymmetrize* Ψ^0 *with respect to electrons 1 and 2. Show that the resulting state is a singlet with respect to these two electrons. Antisymmetrize* Ψ^0 *with respect to all three electrons (a sum of six terms). Does it go to zero (in some obvious way)? Can it be written as a product as in eqn N4.34?*

N4.15 **Exponentials of matrices.** (Math) ③

In quantum mechanics, one often takes exponentials of operators. The exponential of a matrix $\exp(M)$ can be computed using several different equivalent relations.

First, one can compute it as a power series:

$$\exp(M) = \sum_{n=0}^{\infty} M^n/n! \tag{N4.20}$$

Let's take the exponential $\exp(-\mathrm{i}\phi\sigma_2/2)$, where $\sigma_2 = \left(\begin{smallmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{smallmatrix}\right)$ is the second Pauli matrix (also known as σ_y). This is the definition of how spin $\frac{1}{2}$ particles transform under rotations.

(a) Note that $\sigma_2^2 = 1$*. Separate the infinite series into even and odd terms, and express* $\exp(-\mathrm{i}\phi\sigma_2/2)$ *as a linear combination of the identity matrix* $\mathbb{1}$ *and the matrix* σ_2*. In your answer, note that a* 360° *rotation is not equal to the identity, but to minus the identity!*

Secondly, one can compute it as an infinite product of infinitesimal transformations:

$$\begin{aligned} \exp(M) &= \lim_{n\to\infty} \exp(M/n)^n \\ &= \lim_{n\to\infty} (\mathbb{1} + M/n)^n. \end{aligned}$$

This will be the basic trick we use to generate the path-integral formulation of quantum mechanics. It is also the way we generate symmetry operations (like rotations) from infinitesimal generators (like angular momentum).[18] For example, in two dimensions the angular momentum operator is $J = \mathrm{i}\hbar \left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)$.

(b) Show that a 2×2 *rotation matrix by an angle* θ/n*, in the limit* $n \to \infty$*, can be written as* $\mathbb{1} + (C\theta/n)J$*. What is the constant* C*? Argue, without calculation, that the product in eqn* **??** *must generate the finite-angle rotations.*

Finally, many matrices which arise in quantum mechanics (symmetric matrices, Hermitian matrices, and the more general category of *normal matrices*) can be diagonalized by a unitary change of basis: $D = \left(\begin{smallmatrix} \lambda_1 & 0 & \cdots \\ 0 & \lambda_2 & \cdots \\ 0 & 0 & \cdots \end{smallmatrix}\right) = U^\dagger M U$, with $U^\dagger = (U^T)^* = U^{-1}$. Thus $M^n = (UDU^\dagger)^n = UD(U^\dagger U)D\cdots DU^\dagger = UD^nU^\dagger$. For these matrices, we can

[18] Continuous groups like the rotations are called *Lie groups*. The corresponding infinitesimal generators, and their commutation relations, are called the *Lie algebra* for the group.

compute the exponential of a matrix by doing a coordinate change to the basis that diagonalizes it:

$$\begin{aligned}\exp(M) &= \sum_{n=0}^{\infty} M^n/n! = \sum_{n=0}^{\infty} UD^nU^\dagger/n! \\ &= U\left(\sum_{n=0}^{\infty} D^n/n!\right)U^\dagger \\ &= U\begin{pmatrix} \mathrm{e}^{\lambda_1} & 0 & \cdots \\ 0 & \mathrm{e}^{\lambda_2} & \cdots \\ 0 & 0 & \cdots \end{pmatrix}U^\dagger\end{aligned} \tag{N4.21}$$

Let's apply this to the time evolution operator $\exp(-\mathrm{i}Ht/\hbar)$ for the Hamiltonian we studied in the Eigen exercise (3.1): $H = \left(\begin{smallmatrix} 0 & -4 \\ -4 & 6 \end{smallmatrix}\right)$.

(c) Apply the relation eqn N4.38 to calculate the 2×2 *time evolution operator* $\exp(-\mathrm{i}Ht/\hbar)$ *for our Hamiltonian. Apply the resulting time evolution operator to the state* $\psi(0) = \left(\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right)$ *to calculate* $\psi(t)$*. Also write the time evolved state as* $\sum_n \exp(-\mathrm{i}E_nt/\hbar)|n\rangle\langle n|\psi\rangle$*, where* $|n\rangle$ *are the eigenstates of* H*. Do the two methods agree?*

N4.16 **Baker-Campbell-Hausdorff identity.** (Expanded upon from Gottfried & Yan exercise 2.13.) ③

For operators A and B, we know $\mathrm{e}^{A+B} \neq \mathrm{e}^A\mathrm{e}^B$ unless A and B commute, so $C = [A, B] = 0$. The BCH theorem tells us that, for any two linear operators A and B, that

$$\mathrm{e}^A\mathrm{e}^B = \exp(A + B + \tfrac{1}{2}[A, B] + \tfrac{1}{12}([A, [A, B] - [B, [A, B]]) \dots) \tag{N4.22}$$

where ... alludes to multiple commutators of even higher order.

We start with the special case where A and B do not commute with each other but which both commute with $[A, B]$. In that case, they satisfy

$$\mathrm{e}^{A+B} = \mathrm{e}^A\mathrm{e}^B\mathrm{e}^{-\frac{1}{2}[A,B]} \tag{N4.23}$$

(a) *Examine the power series of both sides. Show that they agree at low orders.*

(b) *To prove this, first show that* $[B, \mathrm{e}^{xA}] = \mathrm{e}^{xA}[B, A]x$*, where* x *is a scalar (a number, which commutes with everything). Next, define* $G(x) = \mathrm{e}^{xA}\mathrm{e}^{xB}$ *and show that*

$$\frac{\mathrm{d}G}{\mathrm{d}x} = (A + B + [A, B]x)G. \tag{N4.24}$$

Integrate this to obtain the desired result.

(c) *Let* K *and* P *be two noncommuting operators (e.g., the kinetic and potential energy parts of the Hamiltonian* $H = K + P$*). Let* ϵ *be small (e.g.,* $\mathrm{i}\mathrm{d}t/\hbar$ *in a small time-step* $U(\mathrm{d}t) = \exp(-\mathrm{i}H\mathrm{d}t/\hbar)$*. Show that*

$$\mathrm{e}^{\epsilon(K+P)} = \mathrm{e}^{\epsilon K}\mathrm{e}^{\epsilon P} + O(\epsilon^2) \tag{N4.25}$$

by explicitly expanding out the power series. What order is the error in the approximation

$$\mathrm{e}^{\epsilon(K+P)} \approx \mathrm{e}^{\epsilon K/2}\mathrm{e}^{\epsilon P}\mathrm{e}^{\epsilon K/2}? \tag{N4.26}$$

(Hint: We have a good reason to do the extra work of adding the third step.)

N4.17 **Harmonic oscillators and symbolic manipulation.** (Computation) ③

In this exercise, we shall use *symbolic manipulation* environments (*Mathematica* or *SymPy*) to explore the raising and lowering operators a and $a^\dagger$. We'll use them to generate the position-space eigenstates $\psi_n(x)$, along with their associated Hermite polynomials. We'll distinguish between *analytical* calculations (paper and pencil) and *symbolic* calculations (using the symbolic manipulation package on the computer).

Remember that the Hamiltonian for a simple harmonic oscillator[19] is

$$H = p^2/2m + \tfrac{1}{2}m\omega^2 x^2. \tag{N4.27}$$

with

$$p = -\mathrm{i}\hbar\frac{\partial}{\partial x} \tag{N4.28}$$

The ground state probability distribution is a Gaussian of width $a_0 = \sqrt{\hbar/2m\omega}$, so

$$\psi_0(x) = (m\omega/\pi\hbar)^{1/4}\mathrm{e}^{-m\omega x^2/(2\hbar)}. \tag{N4.29}$$

For plots, we'll take constants from McEuen's bouncing buckyballs (Park et al., "Nanomechanical oscillations in a single-C_{60} transistor", *Nature* **407**, 57 (2000)). Thus $m \approx 60 * 12$amu, with an amu = 1.66054e-24 gm, $\hbar \approx 1.0545716 \times 10^{-27}$ erg sec, and $\omega = 1.2$THz.

(a) *Do a symbolic integration to check if ψ_0 is properly normalized. Plot ψ_0 from $-4a_0$ to $4a_0$ using McEuen's constants. How do the zero-point fluctuations for McEuen's buckyball compare to the size of an atom?*

(b) *Define an operator $H(\psi)$ that symbolically takes a function $\psi(x) = |\psi\rangle$ and returns another function $H|\psi\rangle$. Symbolically calculate the ground state energy $E_0 = H(\psi_0)/\psi_0$.* (Hint: it should be independent of x.)

Remember that the ladder operators are written in terms of the position and momentum:

$$a = \sqrt{m\omega/2\hbar}(x + \mathrm{i}p/m\omega) \tag{N4.30}$$

$$a^\dagger = \sqrt{m\omega/2\hbar}(x - \mathrm{i}p/m\omega) \tag{N4.31}$$

Remember the number operator $N = a^\dagger a$.

(c) *Using the commutation relation $[x, p] = \mathrm{i}\hbar$, analytically (paper and pencil) show that $(N + \tfrac{1}{2})\hbar\omega = H$, $[a, a^\dagger] = 1$, $[N, a^\dagger] = a^\dagger$, and $[N, a] = -a$.*

(d) *Using the commutation relations from above, analytically show that $H(a^\dagger)^n\psi_0 = (n + \tfrac{1}{2})\hbar\omega(a^\dagger)^n\psi_0$, so $\psi_n \propto (a^\dagger)^n\psi_0$.*

(e) *As in parts (b), symbolically define the operator p from eqn N4.45. Using it, define the operator $a^\dagger(\psi)$. Is it normalized? Symbolically calculate $E_1 = H\left(a^\dagger\psi_0\right)/a^\dagger\psi_0$.*

[19] The spring constant $K = m\omega^2$; this gives $\omega = \sqrt{K/m}$, which may be more familiar.

(f) *Symbolically, is $a^\dagger a^\dagger \psi_0$ normalized? Calculate symbolically the norm of higher powers of $(a^\dagger)^n \psi_0$ until you figure out what we need to divide it by to normalize it.*

(g) *Analytically calculate the norm of $(a^\dagger)^n \psi_0$, using the commutation relations of part (c). Does it agree with your conclusion of part (f)?*

(h) *Symbolically define ψ_n recursively in terms of ψ_{n-1}, using the proper normalization from parts (f) and (g). Evaluate it symbolically for $n = 1, 2, 3, 4$.*

(i) *Using McEuen's buckyball constants, plot $\psi_n(x)$ for $n = 1, 2, 3, 4$ for $-5a_0 < x < 5a_0$.*

N4.18 **Matrices, wavefunctions, and group representations.** (Group reps) ③

In this exercise, we shall explore the *tensor product* of two vector spaces, and how they transform under rotations. We'll draw analogies between two examples: vectors → matrices and single-particle-states → two-particle-wavefunctions.

The tensor product between two vectors is $(\mathbf{v} \otimes \mathbf{w})_{ij} = v_i w_j$. The tensor product between two single-particle wavefunctions $\zeta(x)$ for particle A and $\phi(y)$ for particle B is the product wavefunction $\Psi(x, y) = \zeta(x)\phi(y)$. If $H^{(A)}$ and $H^{(B)}$ are the Hilbert spaces for particles A and B, the tensor product space $H^{(AB)} = H^{(A)} \otimes H^{(B)}$ is the space of all linear combinations of tensor products $\Psi(x, y) = \zeta(x)\phi(y)$ of states in $H^{(A)}$ and $H^{(B)}$. Two-particle wavefunctions live in $H^{(AB)}$.

Let $\{\hat{\mathbf{e}}_i\} = \{\hat{x}, \hat{y}, \hat{z}\}$ be an orthonormal basis for $\mathbb{R}^3$, and let $\{\zeta_i\}$ and $\{\phi_j\}$ be orthonormal bases for the Hilbert spaces $H^{(A)}$ and $H^{(B)}$ for particles A and B.

(a) *Show that the tensor products $\Psi_{ij}(x, y) = \zeta_i(x)\phi_j(y)$ are orthonormal.* (The dot product is the usual $\int \mathrm{d}x\mathrm{d}y\Psi^*\Psi$.) With some more work, it is possible to show that they are also complete, forming an orthonormal basis of $H^{(AB)}$.)

Suppose the two particles are both in states with total angular momentum $L_A = L_B = 1$, and are then coupled with a small interaction. Angular momentum addition rules then say that the two-particle state can have angular momentum $L_{(AB)}$ equal to 2 or 1 or 0: $1 \otimes 1 = 2 \oplus 1 \oplus 0$. In group representation theory, this decomposition corresponds to finding three subspaces that are invariant under rotations.

The tensor product space $\mathbb{R}^3 \otimes \mathbb{R}^3$ are normally written as 3×3 matrices M_{ij}, where $M = \sum_{i=1}^3 \sum_{j=1}^3 M_{ij}\hat{\mathbf{e}}_i\hat{\mathbf{e}}_j$. Vectors transform under $L = 1$, so we would expect matrices to decompose into $L = 2$, 1, and 0.

(b) *Show that the antisymmetric matrices, the multiples of the identity matrix, and the traceless symmetric matrices are all invariant under rotation (i.e., $R^{-1}MR$ is in the same subspace as M for any rotation R). Which subspace corresponds to which angular momentum?*

(c) *Consider the $L = 1$ subspace of matrices M. Provide the (standard) formula taking this space into vectors in $\mathbb{R}^3$. Why are these called pseudovectors?*

I always found torque $\boldsymbol{\tau} = \mathbf{r} \times \mathbf{F}$ quite mysterious. (Its direction depends on whether you are right- or left-handed!) Fundamentally, we see now that this is because torque is not a vector—it is an antisymmetric 3×3 matrix.

How does this relate back to quantum wavefunctions? Suppose our two $L = 1$ particles are identical, with spins in the same state.

(d) *Which angular momentum states are allowed for spin-aligned fermions? For spin-aligned or spinless bosons?*

Many physical properties are described by symmetric matrices: the dielectric constant in electromagnetism, the stress and strain tensors in elastic theory, and so on.

N4.19 **Molecular rotations.** (Quantum) ③

In class, we estimated the frequency of atomic vibrations, by generating a simple model of an atom of mass AM_P in a harmonic potential whose length and energy scales were set by electron physics (a Bohr radius and a fraction of a Rydberg). In the end, we distilled the answer that atomic vibrations were lower in frequency than those of electrons by a factor $\sqrt{M_P/m_e}$, times constants of order one.

Here we consider the frequencies of molecular *rotations*.

(a) *By a similar argument, derive the dependence of molecular rotation energy splittings on the mass ratio M_P/m_e.*

(b) *Find some molecular rotation energy splittings in the literature. Are they in the range you expect from your estimates of part (a)?*

N4.20 **Propagators to path integrals.** (PathIntegrals) ③

In class, we calculated the propagator for free particles, which Sakurai also calculates (eqn 2.6.16):

$$K(x',t;x_0,t_0) = \sqrt{\frac{m}{2\pi \mathrm{i}\hbar(t-t_0)}}\exp\left[\frac{\mathrm{i}m(x'-x_0)^2}{2\hbar(t-t_0)}\right]. \tag{N4.32}$$

Sakurai also gives the propagator for the simple harmonic oscillator (eqn 2.6.18):

$$\begin{aligned} K(x',t;x_0,t_0) = &\sqrt{\frac{m\omega}{2\pi \mathrm{i}\hbar\sin[\omega(t-t_0)]}} \\ &\times \exp\Big[\left\{\frac{\mathrm{i}m\omega}{2\hbar\sin[\omega(t-t_0)]}\right\} \\ &\left[(x'^2+x_0^2)\cos[\omega(t-t_0)] - 2x'x_0\right]\Big]. \end{aligned} \tag{N4.33}$$

In deriving the path integral, Feynman approximates the short-time propagator in a potential $V(x)$ using the trapezoidal rule:

$$\begin{aligned} &K(x_0{+}\Delta x, t_0+\Delta t; x_0, t_0) \\ &\quad = N_{\Delta t}\exp\left[\frac{\mathrm{i}\Delta t}{\hbar}\left\{\tfrac{1}{2}m(\Delta x/\Delta t)^2 - V(x_0)\right\}\right], \end{aligned} \tag{N4.34}$$

where the expression in the curly brackets is the straight-line approximation to the Lagrangian $\frac{1}{2}m\dot{x}^2 - V(x)$. Check Feynman's approximation: is it correct to first order

in Δt for the free particle and the simple harmonic osillator? For simplicity, let's ignore the prefactors (coming from the normalizations), and focus on the terms inside the exponentials.

Taking $t = t_0 + \Delta t$ and $x' = x_0 + \dot{x}\Delta t$, expand to first order in Δt the terms in the exponential for the free particle propagator (eqn N4.50) and the simple harmonic oscillator (eqn N4.51). Do they agree with Feynman's formula? (Hint: For the simple harmonic oscillator, the first term is proportional to $1/\Delta t$, so you'll need to keep the second term to second order in Δt.)

N4.21 **Three particles in a box.** (Quantum) ③

(Adapted from Sakurai, p. 4.1)

Consider free, noninteracting particles of mass m in a one-dimensional box of length L with infinitely high walls.

(a) *What are the lowest three energies of the single-particle energy eigenstates?*

If the particles are assumed noninteracting, the quantum eigenstates can be written as suitably symmetrized or antisymmetrized single-particle eigenstates. One can use a level diagram, such as in Fig. N4.10, to denote the fillings of the single particle states for each many-electron eigenstate.

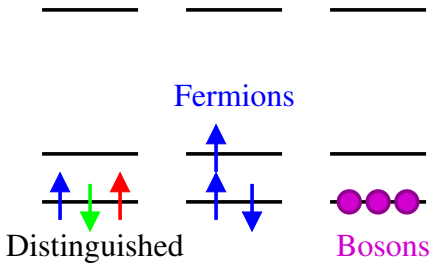


Fig. N4.10 Level diagram, showing one of the ground states for each of the three cases.

(b) *If three distinguishable spin-1/2 particles of the same mass are added to the box, what is the energy of the three-particle ground state? What is the degeneracy of the ground state? What is the first three-particle energy eigenvalue above the ground state? Its degeneracy? The degeneracy and energy of the second excited state? Draw a level diagram for one of the first excited states, and one of the second excited states (the ground state being shown on the left in Fig. N4.10).*

(c) *The same as part (b), but for three identical spin-1/2 fermions.*

(d) *The same as part (b), but for three identical spin-zero bosons.*

N4.22 **Rotation matrices.** (Mathematics) ②

A rotation matrix R takes an orthonormal basis $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$, $\hat{\mathbf{z}}$ into another orthonormal triad $\hat{\mathbf{u}}$, $\hat{\mathbf{v}}$, $\hat{\mathbf{w}}$, with $\hat{\mathbf{u}} = R\hat{\mathbf{x}}$, $\hat{\mathbf{v}} = R\hat{\mathbf{y}}$, and $\hat{\mathbf{w}} = R\hat{\mathbf{z}}$.

(a) Which is another way to write the matrix R?

I. $R = \begin{pmatrix} u_1v_1 + v_1w_1 + w_1u_1 & \cdots \\ \cdots & \end{pmatrix}$

II. $R = \begin{pmatrix} (& \hat{\mathbf{u}} &) \\ (& \hat{\mathbf{v}} &) \\ (& \hat{\mathbf{w}} &) \end{pmatrix}$;

III. $R = \begin{pmatrix} \begin{pmatrix}\hat{\mathbf{u}}\end{pmatrix} & \begin{pmatrix}\hat{\mathbf{v}}\end{pmatrix} & \begin{pmatrix}\hat{\mathbf{w}}\end{pmatrix} \end{pmatrix}$;

IV. $R = \hat{\mathbf{u}} \otimes \hat{\mathbf{v}} + \hat{\mathbf{v}} \otimes \hat{\mathbf{w}} + \hat{\mathbf{w}} \otimes \hat{\mathbf{u}}$

Rotation matrices are to real vectors what unitary transformations (common in quantum mechanics) are to complex vectors. A unitary transformation satisfies $U^\dagger U = \mathbb{1}$, where the dagger gives the complex conjugate of the transpose, $U^\dagger = (U^T)^*$. Since R is real, $R^\dagger = R^T$.

(b) Argue that $R^T R = \mathbb{1}$.

Thus R is an *orthogonal* matrix, with transpose equal to its inverse.

(c) In addition to (b), what other condition do we need to know that R is a proper rotation (i.e., in $SO(3)$), and not a rotation-and-reflection with determinant -1?

(I) $\hat{\mathbf{u}}$, $\hat{\mathbf{v}}$, *and* $\hat{\mathbf{w}}$ *must form a right-handed triad (presuming as usual that* $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$, *and* $\hat{\mathbf{z}}$ *are right-handed),*

(II) $\hat{\mathbf{u}} \cdot \hat{\mathbf{v}} \times \hat{\mathbf{w}} = 1$

(III) $\hat{\mathbf{w}} \cdot \hat{\mathbf{u}} \times \hat{\mathbf{v}} = 1$

(IV) All of the above

One of the most useful tricks in quantum mechanics is multiplying by one. The operator $|k\rangle\langle k|$ can be viewed as a projection operator: $|k\rangle\langle k|\psi\rangle$ is the part of $|\psi\rangle$ that lies along direction $|k\rangle$. If k labels a complete set of orthogonal states (say, the eigenstates of the Hamiltonian), then the original state can be reconstructed by adding up the components along the different directions: $|\psi\rangle = \sum_k |k\rangle\langle k|\psi\rangle$. Hence the identity operator $\mathbb{1} = \sum_k |k\rangle\langle k|$. We'll use this to derive the path-integral formulation of quantum mechanics, for example. Let's use it here to derive the standard formula for rotating matrices.

Under a change of basis R, a matrix A transforms to $R^T AR$. We are changing from the basis $\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2, \hat{\mathbf{x}}_3 = |x_i\rangle$ to the basis $|u_j\rangle$, with $|u_n\rangle = R|x_n\rangle$. Since $|u_j\rangle = R|x_j\rangle$, we know $\langle x_i|u_j\rangle = \langle x_i|R|x_j\rangle = R_{ij}$, and similarly $\langle u_i|x_j\rangle = R^T_{ij}$. Let the original components of the operator $\mathbb{A}$ be $A_{k\ell} = \langle x_k|A|x_\ell\rangle$ and the new coordinates be $A'_{ij} = \langle u_i|A|u_j\rangle$.

(d) Multipling by one twice into the formula for A': $A'_{ij} = \langle u_i|\mathbb{1}A\mathbb{1}|u_j\rangle$ and expanding the first and second identities in terms of x_k and x_ℓ, derive the matrix transformation formula $A'_{ij} = R^T_{ik} A_{k\ell} R_{\ell j} = R^T AR$, where we use the Einstein summation convention over repeated indices.

N4.23 **Trace.** (Mathematics) ②

The trace of a matrix A is $Tr(A) = \sum_i A_{ii} = A_{ii}$ where the last form makes use of the *Einstein summation convention*.

(a) Show the trace has a cyclic invariance: $Tr(ABC) = Tr(BCA)$. (Hint: write it out as a sum over components. Matrices don't commute, but products of components of matrices are just numbers, and do commute.) *Is* $Tr(ABC) = Tr(ACB)$ *in general?*

Remember from exercise N4.22(b) that a rotation matrix R has its inverse equal to its transpose, so $R^T R = \mathbb{1}$, and that a matrix A transforms into $R^T A R$ under rotations.

(b) Using the cyclic invariance of the trace, show that the trace is invariant under rotations.

Rotation invariance is the primary reason that the trace is so important in mathematics and physics.

N4.24 **Complex exponentials.** (Mathematics) ②

You can prove the double angle formulas using complex exponentials: just take the real and imaginary parts of the identity $\cos(A+B) + \mathrm{i}\sin(A+B) = \mathrm{e}^{\mathrm{i}(A+B)} = \mathrm{e}^{\mathrm{i}A}\mathrm{e}^{\mathrm{i}B} = (\cos A + \mathrm{i}\sin A)(\cos B + \mathrm{i}\sin B) = \cos A\cos B - \sin A\sin B + \mathrm{i}(\cos A\sin B + \sin A\cos B)$.

In a similar way, use complex exponentials and $(x+y)^3 = x^3 + 3x^2y + 3xy^2 + y^3$ *to derive the triple angle formulas. Which is true?*
(A) $\cos(3\theta) = \cos^3(\theta) - 3\cos(\theta)\sin^2(\theta)$, $\sin(3\theta) = \sin^3(\theta) - 3\cos^2(\theta)\sin(\theta)$;
(B) $\sin(3\theta) = \cos^3(\theta) - 3\cos(\theta)\sin^2(\theta)$, $\cos(3\theta) = 3\cos^2(\theta)\sin(\theta) - \sin^3(\theta)$;
(C) $\cos(3\theta) = \cos^3(\theta) - 3\cos(\theta)\sin^2(\theta)$, $\sin(3\theta) = 3\cos^2(\theta)\sin(\theta) - \sin^3(\theta)$;
(D) $\sin(3\theta) = 3\cos^2(\theta)\sin(\theta) + \sin^3(\theta)$; $\cos(3\theta) = \cos^3(\theta) + 3\cos(\theta)\sin^2(\theta)$,
(E) $\cos(3\theta) = \cos^3(\theta) - 3\mathrm{i}\cos(\theta)\sin^2(\theta)$, $\sin(3\theta) = 3\mathrm{i}\cos^2(\theta)\sin(\theta) - \sin^3(\theta)$.

N4.25 **Dirac δ-functions.** (Mathematics) ③

Quantum bound-state wavefunctions are unit vectors in a complex Hilbert space. If there are N particles in 3 dimensions, the Hilbert space is the space of complex-valued functions $\psi(\mathbf{x})$ with $\mathbf{x} \in \mathbb{R}^{3N}$ whose absolute squares are integrable: $\int \mathrm{d}\mathbf{x}|\psi(\mathbf{x})|^2 < \infty$.

But what about unbound states? For example, the propagating plane-wave states $\psi(x) = |k\rangle \propto \exp(-\mathrm{i}kx)$ for a free particle in one dimension? Because unbound states are spread out over an infinite volume, their probability density at any given point is zero—but we surely don't want to normalize $|k\rangle$ by multiplying it by zero.

Mathematicians incorporate continuum states by extending the space into a *rigged Hilbert space*. The trick is that the unbound states form a *continuum*, rather than a discrete spectrum—so instead of summing over states to decompose wavefunctions $|\phi\rangle = \mathbb{1}\phi = \sum_n |n\rangle\langle n|\phi\rangle$ we integrate over states $\phi = \mathbb{1}\phi = \int \mathrm{d}k\,|k\rangle\langle k|\phi\rangle$. This tells us how we must normalize our continuum wavefunctions: instead of the Kronecker-δ function $\langle m|n\rangle = \delta_{mn}$ enforcing orthonormal states, we demand $|k'\rangle = \mathbb{1}|k'\rangle = \int \mathrm{d}k\,|k\rangle\langle k|k'\rangle = |k'\rangle = \int \mathrm{d}k\,|k\rangle\delta(k-k')$ telling us that $\langle k|k'\rangle = \delta(k-k')$ is needed to ensure the useful decomposition $\mathbb{1} = \int \mathrm{d}k\,|k\rangle\langle k|$.

Let's work out how this works as physicists, by starting with the particles in a box, and then taking the box size to infinity. For convenience, let us work in a one dimensional box $0 \le x < L$, and use *periodic boundary conditions*, so $\psi(0) = \psi(L)$ and $\psi'(0) = \psi'(L)$. This choice allows us to continue to work with plane-wave states $|n\rangle \propto \exp(-\mathrm{i}k_n x)$ in the box. (We could have used a square well with infinite sides, but then we'd need to fiddle with wave-functions $\propto \sin(kx)$.)

(a) What values of k_n are allowed by the periodic boundary conditions? What is the separation Δk between successive wavevectors? Does it go to zero as $L \to \infty$, leading to a continuum of states?

To figure out how to normalize our continuum wavefunctions, we now start with the relation $\langle m|n\rangle = \delta_{mn}$ and take the continuum limit. We want the normalization N_k of the continuum wavefunctions to give $\int_{-\infty}^{\infty} N_k N_{k'} \exp(-\mathrm{i}(k'-k)x)\mathrm{d}x = \delta(k'-k)$.

(b) What is the normalization $\langle x|n\rangle = N_n \exp(\mathrm{i}k_n x)$ for the discrete wave-functions in the periodic box, to make $\langle n|n'\rangle = \int_0^L N_n N_m \exp(-\mathrm{i}(k'_n - k_n)x)\mathrm{d}x = \delta_{mn}$? Write $\mathbb{1} = \sum_n |n\rangle\langle n|$, and change the sum to an integral in the usual way ($\int \mathrm{d}k\, |k\rangle\langle k| \approx \sum_n |n\rangle\langle n|\Delta k$). Show that the normalization of the continuum wavefunctions must be $N_k = 1/\sqrt{2\pi}$, so $\psi_k(x) = \langle x|k\rangle = \exp(\mathrm{i}kx)/\sqrt{2\pi}$. (Hint: If working with operators is confusing, ensure that $\langle x|\mathbb{1}|x'\rangle$ for $0 < x, x' < L$ is the same for $\mathbb{1} = \sum_n |n\rangle\langle n|$ (valid in the periodic box) and for $\mathbb{1} = \int \mathrm{d}k|k\rangle\langle k|$ (valid for all x).

Notice some interesting ramifications:

I. The fact that our continuum plane waves have normalization $1/\sqrt{2\pi}$ incidentally tells us one form of the δ function:

$$\delta(k'-k) = \langle k|k'\rangle = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-\mathrm{i}(k'-k)x}\mathrm{d}x. \tag{N4.35}$$

Also, $\delta(x'-x) = 1/(2\pi)\int_{-\infty}^{\infty} \mathrm{d}k \exp(\mathrm{i}k(x'-x))$.

II. The same normalization is used for position eigenstates $|x\rangle$, so $\langle x'|x\rangle = \delta(x'-x)$ and $\mathbb{1} = \int \mathrm{d}x|x\rangle\langle x|$.

III. The Fourier transform can be viewed as a change of variables from the basis $|x\rangle$ to the basis $|k\rangle$:

$$\begin{aligned}\widetilde{\phi}(k) &= \langle k|\phi\rangle = \langle k|\mathbb{1}|\phi\rangle \\ &= \int \mathrm{d}x \langle k|x\rangle\langle x|\phi\rangle \\ &= 1/\sqrt{2\pi}\int \mathrm{d}x \exp(-\mathrm{i}kx)\phi(x)\end{aligned} \tag{N4.36}$$

Note that this definition is different from that I used in the appendix of my book (Statistical Mechanics: Entropy, Order Parameters, and Complexity, http://pages.physics.cornell.edu/~sethna/Stat Mech/EntropyOrderParametersComplexity.pdf); there the 2π

is placed entirely on the inverse Fourier transform, which here it is split symmetrically between the two, so the inverse Fourier transform is

$$\phi(x) = \langle x|\phi\rangle = \langle x|\mathbb{1}|\phi\rangle \tag{N4.37}$$

$$= \int \mathrm{d}k \langle x|k\rangle\langle k|\phi\rangle \tag{N4.38}$$

$$= 1/\sqrt{2\pi} \int \mathrm{d}k \exp(\mathrm{i}kx)\widetilde{\phi}(k).$$

IV. The Dirac δ-function can be written in many different ways. It is basically the limit[20] as $\epsilon \to 0$ of sharply-peaked, integral-one functions of width ϵ and height $1/\epsilon$ centered at zero. Let's use this to derive the useful relation:

$$\lim_{\epsilon\to 0+} \frac{1}{x - \mathrm{i}\epsilon} = p.v.\frac{1}{x} + \mathrm{i}\pi\delta(x). \tag{N4.39}$$

Here all these expressions are meant to be inside integrals, and p.v. is the Cauchy principal value of the integral:[21]

$$p.v. \int_{-\infty}^{\infty} = \lim_{\epsilon\to 0+} \int_{-\infty}^{-\epsilon} + \int_{\epsilon}^{\infty}. \tag{N4.40}$$

(c) Note that $\epsilon/(x^2+\epsilon^2)$ has half-width ϵ at half-maximum, height ϵ, and integrates to π, so basically (i.e., in the weak limit) $\lim_{\epsilon\to 0} \epsilon/(x^2+\epsilon^2) = \pi\delta(x)$. Argue that $\lim_{\epsilon\to 0} \int f(x)/(x-\mathrm{i}\epsilon)\mathrm{d}x = p.v. \int f(x)/x\,\mathrm{d}x + \mathrm{i}\pi f(0)$. (Hints: The integral of $1/(1+y^2)$ is $\arctan(y)$, which becomes $\pm\pi/2$ at $y = \pm\infty$. Multiply numerator and denominator by $x + \mathrm{i}\epsilon$.)

N4.26 **Eigen Stuff.** (Mathematics) ②

Consider an operator for a two-state system $O = \begin{pmatrix} 0 & -4 \\ -4 & 6 \end{pmatrix}$ Its eigenvectors are $|e_1\rangle = \frac{1}{\sqrt{5}}\begin{pmatrix} 2 \\ 1 \end{pmatrix}$ and $|e_2\rangle = \frac{1}{\sqrt{5}}\begin{pmatrix} -1 \\ 2 \end{pmatrix}$

(a) What are the associated eigenvalues o_1 and o_2?

(b) Use $|e_1\rangle$ and $|e_2\rangle$ to construct a rotation matrix R that diagonalizes O, so $R^T O R = \begin{pmatrix} o_1 & 0 \\ 0 & o_2 \end{pmatrix}$. (Hint: See problem N4.22(a). We want R to rotate the axes into $\hat{\mathbf{u}} = |e_1\rangle$ and $\hat{\mathbf{v}} = |e_2\rangle$.) *What angle does R rotate by?*

(c) Assume that the system is in a state $|L\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ Decompose $|L\rangle$ into the eigenvectors of O. (Hint: As in Exercise N4.22(d), multiplying $|L\rangle$ by one is useful.) *If the observable corresponding to the operator O is measured for state $|L\rangle$, what is the probability of finding the value o_1? Does the probability of finding either o_1 or o_2 sum to one?*

[20] Clearly this is not a limit in the ordinary sense: the difference between functions does not go to zero as ϵ goes to zero, but rather (within ϵ of the origin) has large positive and negative values that cancel. It is a *weak limit*—when integrated against any smooth functions, the differences go to zero.

[21] If $f(x)$ is positive at zero, $\int f(x)/x\,\mathrm{d}x$ is the sum of minus infinity for $x < 0$ and positive infinity for $x > 0$; taking the principal value tells us to find the sum of these two canceling infinities by chopping them symmetrically about zero.

N4.27 **Fine and hyperfine structure: Hydrogen and angular momentum addition.** (Angular Momentum) ③

Symmetries have powerful implications for energy eigenstates of composite systems. They are ordinarily the only cause for degenerate states, for example. Here we use rotational symmetry, and the corresponding angular-momentum addition laws, to derive the degeneracies of the hydrogen $n = 2$ states.

Including the spin ½ of the electron and the spin ½ of the proton, and the four $n = 2$ states of hydrogen, there are sixteen degenerate energy eigenstates in Schrödinger's solution for hydrogen with $n = 2$. In this exercise, we shall follow how these energy eigenstates split up when we include the *fine splitting* and *hyperfine splitting*. We shall not need to do any calculations with Hamiltonians; we shall just use the rotational symmetry of the Hamiltonian and angular momentum addition rules.

(a) *What is the energy of the $n = 2$ state of hydrogen, ignoring spin, relativity, and the nuclear spin?* (Include the fact that the proton and electron have spin ½ in the degeneracy calculation, but ignore their effects on the energy for now.)

The 2s and 2p states in hydrogen both have $n = 2$, and are degenerate to this order. This degeneracy is not due to a straightforward symmetry of the Hamiltonian.[22] It is split by terms of order α^2, where $\alpha = e^2/\hbar c \approx 1/137$ is the fine structure constant, representing the importance of relativity.

The relativistic correction to the kinetic energy splits the 2s and 2p states, but does not couple to the electron or proton spin.

(b) *Including these kinetic energy terms, how do the sixteen original states split up in energy?*

The spin-orbit coupling, also of order α^2, is proportional to $\mathbf{L}\cdot\mathbf{S}$, where $\mathbf{L}$ is the angular momentum of the electron and $\mathbf{S}$ is the spin of the electron. Because it is a dot product, it maintains rotational symmetry.

(c) *Using angular momentum addition rules, discuss what happens to the twelve 2p orbitals after incorporating the spin-orbit coupling. What values of j are are allowed, where $J = L + S$? What are the degeneracies of the coupled states?* (Hint: The different energy eigenstates with the same J are related by rotations. You should not need the form of the interaction to solve this part or the next.)

The splitting due to the spin-orbit interaction is called *fine structure*, and also arises in heavier atoms. For example, the yellow light from sodium vapor lamps is comprised of two nearby spectral lines, split by the spin-orbit interaction.[23]

For states with $L > 0$ the coupling to the nuclear spin I is approximately given by $\hat{A}\mathbf{I}\cdot\mathbf{J}$. This is called the *hyperfine* splitting; it is smaller than the fine structure

[22] It's peculiar to the $1/\mathbf{r}$ potential energy law, and an associated conserved Lenz's vector. The hydrogen problem can be mapped in an obscure way to the four-dimensional harmonic oscillator: see S&N sect. 4.1.

[23] Wikipedia also calls the 2s-2p splitting in hydrogen a fine structure effect, but I'm not sure that's standard. In heavier atoms, the energies of these orbitals (quasiparticle resonance energies, not eigenstates) are shifted primarily not due to relativity, but due to the effects of the other electrons.

splittings because the nucleus is heavy compared to the electron. Again, this interaction maintains rotational symmetry (as it must).

(d) *For each of your degenerate families of 2p states in part (c) ignoring the hyperfine splitting, what are the allowed values of $F = I + J$? What degeneracies in the final eigenvalues do you expect?*

N4.28 **Mystery: Properties of the group character table.** (Group Reps) ③

In the week following this assignment, we shall learn about representations of finite groups. Group representation theory involves new conceptual ideas, new mathematical theorems, and some new calculational methods. Even knowing the ideas and the theorems, I find the calculational methods seem mysterious, almost magical. Let's try to introduce these tools first, to motivate the lectures to come. I am not pretending to introduce why we do these manipulations—this is an experiment, giving you the mechanics of the calculation before we explain the context in order to motivate your interest.

Consider the following table. It is an expanded version of the character table for the group representations of C_3v, the symmetry group of a triangle. But just treat it as a list of row vectors A_1, A_2, and E, along the six 'directions' labeled by the six symmetry group elements $g = e$, r, r^2, v, rv, and r^2v in the group G.

C_{3v}	e	r	r^2	v	r v	r^2 v
A_1	1	1	1	1	1	1
A_2	1	1	1	-1	-1	-1
E	2	-1	-1	0	0	0

Table N4.1: Expanded character table for C_{3v}. The group elements $g = e$, r, ... label the columns; the representations $R = A_1$, A_2, ... label the rows, and the entries are the characters $\chi_R(g)$.

(a) Orthogonality *Show that the three character row vectors are orthogonal to one another. Show that the naive "dot product" of a row vector with itself is equal to the number of group elements (called $O(G)$.*

Thus the three representations A_1, A_2, and E are orthonormal using the *inner product* given by the naive dot product divided by the order of the group:

$$\chi_1 * \chi_2 = (1/O(g)) \sum_{g \in G} \chi_1(g)\chi_2(g). \tag{N4.41}$$

Group representations give one matrix $R(g)$ for each abstract symmetry operation g. So rotation matrices form a representation of the rotation group. (Mathematicians carefully distinguish between the abstract multiplication table G for a group, and the implementation $R(g)$ of that group in matrix form.) The *characters* of the group $\chi(g)$

are the traces of these matrices. (Much more about groups and characters will in lecture.)

For example, we can write a representation of the triangle symmetry group C_{3v} by thinking of how each symmetry operation permutes the three vertices of the triangle. Label the three vertices of the triangle by the three unit vectors. Let $R(g)_{ij}$ be one if vertex j shifts to vertex i under the symmetry operation g.

What triangle symmetry corresponds to the six group elements e, r, ...? We always use e to represent the 'do-nothing' symmetry, so $R(e) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$. The matrix r rotates the triangle 180°, so $R(r) = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$. The matrix v flips the triangle around the first vertex, so $R(v) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$.

(b) *What are the characters $\chi(e)$, $\chi(r)$, and $\chi(v)$?*

We define the product of two symmetries (say rv) as performing the symmetry operations from left to right (so, flipping by v and then rotating by r).

(c) *What is the matrix $R(rv)$? Do this two ways. Figure out how the symmetry rv permutes the vertices. Or use the property $R(rv) = R(r)R(v)$; the matrices have the same multiplication table as the group. What is $\chi(rv)$? Is it the same as that of e, r, or v?*

Notice in Table N4.1 that the column vectors labeled by r and r^2 are the same, while v, rv, and r^2v also share the same characters. This is generally true: the characters of two elements in the same *conjugacy class* are always the same. *Use this to check your character for rv in section (c).* [We will explain why this is true in lecture.] We put the two rotations r and r^2 into the conjugacy class C_3, and we put the three reflections v, rv, and r^2v into the conjugacy class σ_v; the identity e is put into the one-element class E. This allows us to make a more efficient character table (Table N4.2), where the number of elements in multiply occupied conjugacy classes is included in the column heading (hence $3\sigma_v$, because there are three σ_v rotations). Now, to find the character of a representation, you only need to compute the trace of one element of each conjugacy class, and to take the inner product of two characters (eqn N4.62) one can sum over conjugacy classes but multiply by the multiplicity (number of elements in the class).

C_{3v}	E	$2C_3$	$3\sigma_v$
A_1	1	1	1
A_2	1	1	-1
E	2	-1	0

Table N4.2: Traditional character table for C_{3v}

One of the main uses for character tables is for finding *decompositions* of representations into *irreducible* representations. This turns out to be related to Fourier transforms, to angular momentum addition rules, and to many other standard problems in mathematics and quantum mechanics. We shall leave what this means mysterious until

lecture, but let us perform a decomposition of the representation R we have described in parts (b) and (c).

(d) *What would the character row for our representation R look like in Table N4.2? Show that the inner product (eqn N4.62) of the representation with itself is an integer, but not one.* Irreducible representations have norm one. *Take the inner product of χ_R with the three irreducible representations, and show that they are integers.* Any reducible representation can be decomposed into integer numbers of the irreducible representations.

I always find it surprising when my naive dot products work out to be multiples of the size of the group. In more complicated cases, it seems magical.

N4.29 **F-electrons and graphene.** (Quantum) ③

In this exercise, we shall explore how seven degenerate f-electron states of an atom split under a weak perturbation which breaks the rotational symmetry.

Atoms often sit atop surfaces with weak interactions without strong bonding; we describe them as *adsorbed*. Consider a light atom[24] in an electronic f-state (i.e., with $\ell = 3$), adsorbed on a monolayer of graphene (Fig. N4.11). Assume the atom is positioned above a point of hexagonal symmetry, so the symmetry group for the atom is broken from $SO(3)$ to C_{6v}.

How do we know this? Why is the symmetry group not just C_6? Why is our system not symmetric under D_{6h}, the symmetry group of graphene?

(a) *What symmetry is exhibited by our adsorbed atom that is not in C_6? What symmetry in D_{6h} is not a symmetry of our adsorbed atom?*

The character of a spin-ℓ representation for SO(3) for a rotation by angle θ is $\chi^{(\ell)}(\theta) = \sin[(\ell + \frac{1}{2})\theta]/\sin[\frac{1}{2}\theta]$. (Check this for the $\ell = 1$ representation, where you know $\chi^{(1)}$ in terms of $\cos[\theta]$. You'll need to use L'Hôpital's rule to evaluate $\chi^{(\ell)}(0)$.)

Six of the symmetry operations in C_{6v} (conjugacy classes σ_v and σ'_v) are reflections—in $O(3)$ but not in $SO(3)$. The characters for representations of $O(3)$ are not so commonly studied. Let's figure them out for the special case of reflections.

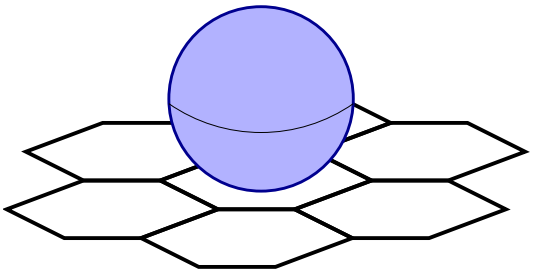


Fig. N4.11 ***Atom adsorbed on graphene.***

[24] The atom is light so that we may ignore the spins of the electrons. A heavy atom would have significant spin-orbit interactions.

C_{6v}	E	C_2	$2C_3$	$2C_6$	$3\sigma_v$	$3\sigma'_v$
A_1	1	1	1	1	1	1
A_2	1	1	1	1	-1	-1
B_2	1	-1	1	-1	1	-1
B_1	1	-1	1	-1	-1	1
E_2	2	2	-1	-1	0	0
E_1	2	-2	-1	1	0	0

Table N4.3: Character table for C_{6v}

Every reflection $\Sigma(\hat{n})$ in $O(3)$ takes the mirror plane into itself, and the perpendicular $\hat{n}$ of the mirror plane to $-\hat{n}$. Thus $\Sigma(\hat{y})$ is a reflection in the $x-z$ mirror plane. Let $R_{\hat{n}}$ be a rotation that takes the coordinate axis $\hat{y}$ to $\hat{n}$.

(b) *Using $R_{\hat{n}}$, show that all reflections in $O(3)$ are conjugate to $\Sigma(\hat{y})$.*

Since the trace is invariant under rotations, and conjugacy in $SO(3)$ is a rotation, and the character is a trace, this means that all reflections will have the same character under representations of $O(3)$. Consider the angular momentum ℓ representation of $O(3)$ generated by the rotations of the spherical harmonics $Y_\ell^m(\theta,\phi)$. Remember that θ is the angle from the $\hat{z}$ axis, and ϕ is measured from the $\hat{x}$ axis.

(c) *How does Y_ℓ^m transform under the reflection $\Sigma(\hat{y})$ in the $x-z$ plane? In the $(2\ell+1)$-dimensional space of Y_ℓ^m for fixed ℓ, what are the elements of the $(2\ell+1)\times(2\ell+1)$ matrix $D_{mm'}$ representing $\Sigma(\hat{y})$? Show that the trace $\chi^{(\ell)}(\Sigma(\hat{y}))=1$, and hence that the character for all reflections is one in all (integer) representations of $O(3)$, independent of ℓ.*

Table N4.3 gives the character table for C_{6v}.

(d) *When the f-electron eigenstates are split by the hexagonal crystal field from the graphene, what irreducible representations and degeneracies will be represented?* (Hint: Use the orthogonality of the representations to decompose the $\ell=3$ representation. Also, check that the total number of states equals the number of f-states.) For example, your answer might be "Two nondegenerate eigenstates with reps A_1 and B_2, and three doublet eigenstates, two with reps E_2 and one with rep E_1.")

N4.30 **Juggling buckyballs.** (Path Integrals) ③

Paul McEuen in Physics and Jiwoong Park in Chemistry here discovered in 2000 that buckyballs (C_{60} molecules) bounce inside their transistors.[25] Here use path integrals to discuss how buckyballs evolve under juggling. (We'll focus on juggling one buckyball, by throwing it straight up into the air and waiting for it to fall down.) The Lagrangian for the buckyball is

$$\mathcal{L} = \tfrac{1}{2}m\dot{y}^2 - mgy. \tag{N4.42}$$

[25] See "Nanomechanical oscillations in a single-C_60 transistor", by Hongkun Park, Jiwoong Park, Andrew K.L. Lim, Erik H. Anderson, A. Paul Alivisatos, and Paul L. McEuen, *Nature* **407**, 57-60 (2000).

(a) *In classical mechanics, if the buckyball starts and ends at $y = 0$ and travels for a time $2\Delta t$, how high y_{peak} must its trajectory reach at the midpoint?* (Hint: Nothing tricky yet.)

Feynman tells us that the propagator for a particle starting at $(y = y_i, t = t_i)$ and ending at $(y = y_f, t = t_f)$ is a path integral over all trajectories $y(t)$:

$$\begin{aligned}\langle y_f|U(t_f - t_i)|y_i\rangle &= \iiint_{y_i,t_i}^{y_f,t_f} \mathcal{D}[y(t)] \exp\left(\mathrm{i}/\hbar S[y(t)]\right) \\ &= \iiint_{y_i,t_i}^{y_f,t_f} \mathcal{D}[y(t)] \exp\left(\mathrm{i}/\hbar \int \mathcal{L}\mathrm{d}t\right)\end{aligned} \tag{N4.43}$$

where the three integral signs represent a suitably normalized integral over all paths $y(t)$. We, like Feynman, will make a discrete "trapezoidal rule" approximation to the propagator. As a rough example, we'll do two segments and only one intermediate point y_2:

$$\begin{aligned}S[y(t)] \approx \Big[&\tfrac{1}{2}m\left(\frac{y_3 - y_2}{\Delta t}\right)^2 - \tfrac{1}{2}mg(y_1 + y_2) \\ &- \tfrac{1}{2}mg(y_2 + y_3) + \tfrac{1}{2}m\left(\frac{y_2 - y_1}{\Delta t}\right)^2\Big]\Delta t.\end{aligned} \tag{N4.44}$$

(b) *What intermediate point y_2^* minimizes the trapezoidal action (eqn N4.65), for general y_1 and y_3? For the symmetric path $y_1 = y_3 = 0$, how does this compare to the peak of the trajectory in part (a)? What is the action $S^* = S[y_2^*]$ for this symmetric minimum action trajectory?* (Note: we're doing an approximation; the heights need not be the same. Hint: Check units of S^*. Also, does it have the right sign?)

(c) *What is our one-point trapezoidal approximation to the propagator*

$$\langle y = 0|U(2\Delta t)|y = 0\rangle? \tag{N4.45}$$

(Request: Please write your answer factoring out the contribution from the minimum action part S^*. Hints: Don't forget the "weight factor" from Sakurai. You can check that you've included the right number of weight factors by checking the units of your propagator: at $t_f = t_i$, for example, $\langle y_f|U(0)|y_i\rangle = \delta(y_f - y_i)$ has units of inverse length. Also, $\int_{-\infty}^{\infty} \mathrm{d}x \exp(-\mathrm{i}Ax^2) = \sqrt{\pi/\mathrm{i}A}$.)

N4.31 **Solving Schrödinger: Nuclear decays and resonances.** (Computation) ③

(a)

N4.32 **Resonances: α-decay.** (Quantum) ③

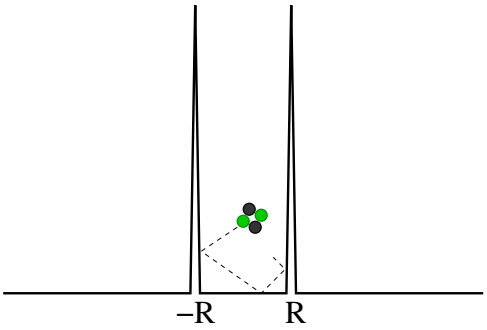


Fig. N4.12 ***One-dimensional nuclear potential.***

In this exercise, we solve a one-dimensional model of radioactive α-decay, where a nucleus ejects a particle formed by two protons and two neutrons (a Helium-4 nucleus).

We assume that the strong force minus the Coulomb repulsion provides a constant potential for the α particle inside a nucleus of radius R, which for simplicity we shall assume is zero. At the edge of the nucleus in the real world, the (short-range) strong interaction drops rapidly to zero, but the Coulomb repulsion decays slowly with distance, leading to a tunneling barrier. We model this barrier with a δ-function of strength $U > 0$[26] (see Fig. N4.12). Both inside and outside the nucleus, the potential is zero:

$$V(x) = U\delta(x \pm R)$$

(The attractive case $U < 0$ is a model for the hydrogen molecule, and is discussed for example in Wikipedia's *Double Delta Potential* article, http://en.wikipedia.org/wiki/Delta_potential#Double_Delta_Potential.)

Parts (a)-(c) of this exercise solve analytically for the energy eigenstates, but getting them correct is important for the later parts.[27]

Our Hamiltonian has a symmetry which allows us to choose energy eigenstates that are even (ψ_E) or odd (ϕ_E).

(a) *What symmetry of the Hamiltonian is this? Given an energy eigenstate $\zeta_E(x)$ with mixed symmetry (in particular, ζ_E is not odd), construct an even eigenstate of the same energy (ignoring the overall normalization).*

In this exercise, we will be interested in the family of even eigenstates states ψ_E which can be nonzero at $x = 0$, and for which $\psi'_E(0) = \partial\psi_E/\partial x|_{x=0} = 0$. To solve for these even energy eigenstates, there are three steps.

[26] In one-dimensional quantum mechanics, the first derivative of the wave-function jumps where the potential has a δ-function. Find details in a textbook or on the Web.

[27] Feel free to check your answers by solving Schrödinger's equation numerically, approximating $\delta(x-R) = (1/\sqrt{2\pi\sigma^2})\exp(-x^2/(2\sigma^2))$ for σ as small as is numerically convenient.

First, we deduce the form of the wavefunction. Note that, away from the δ-function, the wavefunction has wave-vector $k(E) = \sqrt{2mE}/\hbar$; it is convenient to label the wavefunctions by $k(E)$ instead of E. Using the boundary condition at zero, we write the wavefunction for $|x| < R$ as $\psi_k^{\text{nuc}} = A_k \cos(kx)$, with an overall amplitude A_k. For $x > R$, we write the wavefunction as a standing sine wave[28] $\psi_k^{\text{out}} = B \sin(kx + \Delta_k)$. Note that there is a continuum of ψ_k eigenstates, so it is proper for us to use the δ-function normalization $\langle \psi_k | \psi_{k'} \rangle = \delta(k - k')$.

(b) *Show that $B = 1/\sqrt{\pi}$ for our continuum wavefunction to be properly normalized.* (Hints: Since we're studying only even eigenstates, $k \geq 0$. Also, because the region $|x| < R$ is finite, we can ignore it for the normalization in an infinite box.)

Second, we impose the conditions induced by the δ-potential at the edge of the nucleus.

(c) *Write the condition on A_k and Δ_k given by imposing continuity of $\psi_k(x)$ at $x = R$. Write the conditions on A_k and Δ_k given by the discontinuity of $\psi_k'(x)$ imposed by the δ-function potential (see note 26).* For convenience, write your answers from here on in terms of the unitless ratio $\widetilde{U} = 2mRU/\hbar^2$.

Third, we solve for the eigenstates of our Hamiltonian that are nonzero at $x = 0$.

(d) *Use the conditions of part (c), solve for A_k^2.* (Trick: Arrange the two equations of part (c) to be $\sin(kR + \Delta_k) = \cdots$ and $\cos(kR + \Delta_k) = \cdots$, where $\cdots$ is independent of Δ_k. Sum the squares of the right-hand sides: what must the sum be equal to?)

We now consider the decay of an α-particle injected into this potential at $x = 0$. That is, consider an initial wavefunction $\Psi(x) = \delta(x)$.[29]

(e) *What is the probability*[30] *$P(k)$ of being in eigenstate ψ_k?* (Write your answer abstractly in terms of $\psi_k(x)$. This you can do without solving parts (a-d).)

(f) *Plot the probabilities $P(k)$ versus kR with $\widetilde{U} = 30$ and for $0 < kR < 10$.*

In the limit $U \to \infty$, the nucleus should approximate a particle in a box of size $2R$. In that limit, the injection of an α-particle can only occur at certain discrete energies—the nuclear eigenstates $\mathcal{E}_m^\infty$ of a free particle in a box of size $2R$.

(g) *Compare the peaks you found in part (f) to the wavevectors for the particle-in-a-box states. Why are you missing half of the peaks?*

(h) *(Extra credit) Change variables from $P(k)$ to $P(E)$ by using $\mathrm{d}E/\mathrm{d}k$.*[31] *Using the FWHM of the peaks in $P(E)$, estimate the lifetimes of the first three even resonances*

[28] For $x < -R$, we use the even symmetry of ψ_E to set $\psi_k = \psi_k^{\text{out}}(-x) = B \sin(-kx + \Delta_k)$. Note that we are solving for standing waves in this problem. For other purposes, scattering waves or outgoing waves might be preferable.

[29] This is a nuclear version of tunneling from an STM tip; $P(E) = P(k(E))\,(\mathrm{d}k/\mathrm{d}E)$ measures the local density of states for the α particle at the center of the nucleus.

[30] The position eigenstate $\Psi(x) = |x = 0\rangle$ is δ-function normalized, with $\langle x | x' \rangle = \delta(x - x')$. Hence the "probability" $P(k)$ integrates to infinity, and not to one. You can alternatively think of this calculation as the first step in evaluating the Green's function $G(x', t'; 0, 0)$ from $x = t = 0$, which evolves an initial packet $\delta(x)$ from the origin.

[31] The overall normalization of these densities of states may be off by a factor with dimensions of length.

of our nucleus (either numerically or analytically). Calculate the integrated probability for being in each of these three resonances. Do they go to the particle-in-a-box values as $U \to \infty$?

N4.33 **Solving Schrödinger: Alpha decay, Green's functions, and resonances.** (Computation) ③

(a)

N4.34 **Harmonic Fermi sea.** (Quantum, fermions) ③

N identical spin ½ Fermions are subject to a three-dimensional simple harmonic-oscillator potential. Ignore any mutual interactions between the particles.

(a) *Show that the change in the ground state energy when adding an additional particle jumps at certain "magic" N. These are analogous to (but not the same as) the atomic numbers of the noble gases in atomic physics.*[32] *Give the first four of these magic numbers.*

The effective potential for nuclear matter is smooth near the center of the nucleus, so the magic numbers for the number of protons Z and neutrons N in the nucleus are similar to those found in the harmonic oscillator potential.

(b) *Show, if N is large, that the Fermi energy is approximately $E_F = \hbar\omega\sqrt[3]{3N}$, and the ground state energy is $E_B = (3/4)\hbar\omega\sqrt[3]{3N^4}$.*

The number of neutrons in heavy nuclei is larger than the number of protons, because it costs extra Coulomb energy to push protons into the nucleus. (Protons can turn into neutrons by emitting positrons in a β^+ decay.) But when there are more neutrons, they become more costly because their Fermi energy is higher (the Pauli exclusion principle forces them into a high-energy, unoccupied harmonic oscillator state). This is incorporated, in the high-N limit (ignoring the shell structure) into a "Pauli term" in the semi-empirical mass formula for estimating nuclear binding energies.

(c) *If there are $Z = A/2 - \Delta/2$ protons and $N = A/2 + \Delta/2$ neutrons in a 3D harmonic oscillator of frequency $\hbar\omega$, what is the change in the ground state energy $E_B(Z,N) - E_B(A/2, A/2)$ for small Δ, to second order in Δ?*

This does not quite give the form used in the semi-empirical mass formula (Pauli energy $= -a_A(A-2Z)^2/A$).[33] It is usual to use the free Fermi gas in a confining sphere instead of a 3D harmonic oscillator to model the nuclear potential; the latter gives a different denominator.

At root, this is because I started with a δ-function wave packet, whose squared norm is infinity rather than one. I should have put the system on a lattice and taken the lattice to zero, or used Green's functions. Feel free to proceed.

[32] Noble gases are particularly stable; the binding energy of electrons filling the closed shell is unusually high, and the binding energy for the electron for the next alkali metal is unusually low. This corresponds to a jump in the energy per particle at the atomic number of the noble gas.

[33] Note that the semi-empirical mass formula gives the binding energy, which is a constant minus the ground state energy—hence the minus sign.

N4.35 **Periodic Table.** (Atomic) ③

In this exercise, we examine the periodic table to gain insight into the amazing but mysterious utility of thinking of electrons in atoms as filling independent orbitals.

The electron-electron repulsion in atoms, molecules, and solids is almost as large as the electron-nuclear attraction. However, in physics and chemistry we discuss many-electron systems in the language of noninteracting electrons. Thus metals and semiconductors have electron and hole excitations, atoms have 1s electrons near the nucleus and 3d electrons in the transition metals, and diamond has tetrahedral coordination because of sp^3 hybridization of the 2s/2p and 3s/3p orbitals. All of these labels would be valid if the electrons did not interact—but the true electron wavefunctions for an N-electron atom is a complex function of $3N$ variables that in no way factors into orbitals.

First, let us pretend that electrons did not interact with one another: the Coulomb repulsion between electrons is set to zero. Remember that the first four angular momentum states $L = 0$, 1, 2, 3 are called s, p, d, and f for obscure historical reasons.[34] Remember also that the hydrogen spectrum has energy levels $E_n = -(1/n^2)13.6\text{eV}$, with the n^{th} energy level including states with $L = 0, \ldots, n-1$.

(a) *How many noninteracting electrons can fit into the s, p, d, and f states? How many total electrons can fit into each level E_n? If we assume the elements in this noninteracting world kept the same names for each atomic number, which elements would be noble gases?*

These atomic levels are not just a fiction. One can use a high-energy electron or X-ray to eject an electron from an atom (creating a 'core hole' in, say, the 1s state). An electron from the 2s state may then transition into the hole, simultaneously ejecting a third 2p electron from the nucleus. This emitted particle is called an *Auger electron*. The kinetic energies of the Auger electrons will be given, to a good approximation, by the energy difference $-(E_{1s} - E_{2s} - E_{2p})$, just as one would expect if the electrons did not interact.[35] Auger transitions are often used to identify chemical species.

(b) *In our noninteracting world, we can have multi-electron atoms with a hydrogen nucleus. What would the energy be of the emitted Auger electron in the above transition?*

The problem here, however, is if the electrons did not interact the 1s core hole would be an eigenstate that would not decay.[36] This last point is the key to the puzzle. The eigenstates of the noninteracting electrons become *resonances* when interactions are turned on. Their energies are complex, with imaginary parts that correspond to their

[34] There is yet a different notation for these states in the X-ray community (K, L_1, ..., M_5), involving also the total angular momentum $J = L + S$ of the electrons.

[35] Corrections for the interaction energy between the 2s and 2p holes in the final state, and corrections for the electronic screening energies, improve the accuracy of this rough estimate.

[36] The 2s-1s transition could happen by two-photon emission, but would not eject the other electron. Of course, allowing photon emission also makes the electronic excited states into resonances, not eigenstates. Photon emission is small because the fine structure constant $\alpha = e^2/\hbar c \sim 1/137$ is small. The electron-electron interaction is not small, which makes the question more subtle.

lifetimes—here due to their interaction with other electrons. Electrons and holes in metals and semiconductors become *quasiparticles*—quasielectrons and quasiholes that carry around a screening cloud or atomic polarization cloud, and decay eventually into lower-energy excitations.

The resonance energy levels of a 1s core excitation of an atom with nuclear charge Z should be roughly given by the 1s ionization energy of the corresponding helium-like two-electron ion; the outer electrons are mostly farther away from the nucleus and thus will not lower the interaction energy. But the 2s electron resonance energy will roughly be given by the 2s state for an atom with charge $Z-2$, since the 1s electrons will partially screen the nucleus.[37] The different orbitals (1s, 2s, 2p, 3s, 3p, 3d, ...) will get shifted in energy away from their hydrogenic values because of this screening. Various approximate quantum methods for incorporating this screening energy can be developed (e.g., Hartree-Fock).

We can gain some understanding of the power of this picture, without getting buried in arcane Auger tables, by examining the periodic table. Find a periodic table that conveniently shows the fillings of the different subshells (1s, 2s, 2p, 3s, ...).

(c) *Find an ordering of the energies of the subshells that mostly explains the ground state level filling of the different atoms, up to Radon (i.e., ignoring the late radioactive ones). What are the exceptions to your rule?*

(d) *Are the noble metals closed-shell like the noble gases?*

N4.36 **Nuclear Shell Model.** (nuclear) ③

Nuclear physics is challenging. Unlike atomic physics, where the interaction of electromagnetism with matter is weak (of order $\alpha = 1/137$), the interaction between quarks (and hence nucleons) is strong—so we cannot use perturbation theory. Unlike condensed matter physics, where we can assume many particles and hence describe liquids and crystals with continuum theories, the number of protons and neutrons in a nucleus is relatively small.

Nuclear physics is thus a field where creative use of simple models is widespread. We have explored earlier the use of *random matrix theory* to describe excitations of nuclei. In this exercise, we shall introduce both the *nuclear shell model* and the nuclear *semi-empirical mass formula*.

We shall use real data, downloaded directly from the Web. There is a table of atomic masses of various isotopes at the Atomic Mass Data Center, http://amdc.in2p3.fr/masstables/Ame2011int/mass.mas114.

[37] Actually, in electronic structure calculations one often treats the core electrons with a *pseudopotential*. Instead of treating the 2s excitation of a Z atom as a 2s excitation of a $Z-2$ atom, one treats it as a 1s excitation of a $Z-2$ atom, but with a potential that smoothly blurs out the nucleus inside the radius of the 1s shell. They choose this smooth potential to match the scattering amplitudes of the original core-electron ion.

(a) *Download the mass table directly into your computational environment. After reading them, drop the first 39 lines of header. For each N and Z in the table with experimental data*[38] *make a table of the nuclear names (e.g., 56Fe) and of the "mass excess" column (converted to floats). The mass excess is the atomic mass minus one amu (atomic mass unit) per nucleon, where an amu = 931.494061 MeV is one-twelfth the energy of* $^{1}2C$*. The semi-empirical mass formula estimates the nuclear mass, which means we need to add back $A\times$amu and subtract the Z electron masses m_e. Store the binding energy indexed by Z and N:*

$$\text{nuclear mass} = \text{mass excess} + A\,\text{amu} - Zm_e. \tag{N4.46}$$

$$\text{binding energy} = Zm_p + Nm_n - \text{nuclear mass}. \tag{N4.47}$$

Create a matrix with these entries (zero where no experimental data), and make a plot of the nonzero entries.[39]

(Hint: the main feature of this plot will be that larger A have larger nuclear masses.)

The semi-empirical mass formula treats the nucleus primarily as a drop of liquid, with a "condensation energy" $a_V A$, where $A = N + Z$ is the number of nucleons, and a surface tension energy $a_S A^{2/3}$. (If the nucleus is a liquid of nucleons of roughly constant density, then its radius $R \sim A^{1/3}$ and hence the surface area $\sim R^2 \sim V^{2/3}$.) Packing Z protons into the nucleus costs a Coulomb energy (as in $(Zq)^2/R$) of $-a_C Z^2/A^{1/3}$; this Coulomb energy is why there are more neutrons than protons in heavy nuclei.

In addition, there are two quantum terms. The first is the Pauli term, which is related to Exercise N4.34; if the number of neutrons N is different than the number of protons Z, the neutron Fermi energy will be different than the proton Fermi energy, and there will be a energy cost that grows as the difference. Since either positive or negative differences will reduce the binding energy,[40] it is natural to approximate this difference with the square $(N - Z)^2$. It turns out that the magnitude of this effect grows weaker as A grows, so using $A = N + Z$ we approximate the Pauli energy $-a_A(N - Z)^2/A = -a_A(A - 2Z)^2/A$.

The second quantum term is a pairing energy. If the number of protons (or neutrons) is even, the energy is lower than if it is odd, by a pairing energy that goes roughly as $\delta = a_P/A^{1/2}$. (Some sources say this is due to the fact that the second nucleon can "go into the same orbital" as the first; others say this even-odd term is due to spin-orbit interactions; others blame it on the formation of superconducting Cooper pairs of protons or neutrons.) The pairing energy is considered zero for even-odd nuclei, positive (binding) for even-even nuclei, and negative (destabilizing) for odd-odd nuclei. (It is interesting to note that there are only four stable nuclei with an odd number of protons and an odd number of neutrons.)

[38] Warning: data with # instead of decimal points are theoretical extrapolations. Drop those points.
[39] You can use a *masked array* to do this in Python; in Mathematica you can set ColorRules → 0.0→White.
[40] Usually there will be more neutrons

We use Rohlf's values (Rohlf: Modern Physics from a to Z0, James William Rohlf, Wiley, 1994, section 11.3, quoted from Wikipedia), who give $a_V = 15.75\,\text{MeV}$, $a_S = 17.8\,\text{MeV}$, $a_C = 0.711\,\text{MeV}$, $a_A = 23.7\,\text{MeV}$, and $a_P = 11.18\,\text{MeV}$.

(b) *Create a function SemiEmpiricalMassFormula(Z,N) that evaluates this formula. Check against the actual value for 56Fe from your calculation in part (a).*[41] (Hint: Iron has $Z = 26$ and hence ^{5}6Fe has $N = 30$. Your answer should be within less than a percent of the experimental value. This can also help debug part (a). For assistance in checking for typos, I got volumeTerm=882.0 MeV, surfaceTerm=-260.54 MeV, coulombTerm=-125.628 MeV, PauliTerm=-6.77 MeV, δ=1.49 MeV.)

One major piece of physics that the semi-empirical mass formula misses are the *magic numbers*. These are explained by assuming that, like atoms, the nucleons approximately fill orbitals that have shells (like the 1s, 2s, 2p, ... for atoms). Just as noble gases arise when a shell fills and the next orbital has a big jump in energy, so nuclei have especially stable states. Notice, though, that there is only one type of electron, while there are two types of nucleons. So we expect especially stable nuclei when either the proton or the neutron number is magic: horizontal or vertical bands in a color plot of mass versus Z and N.

(c) *Make a 2D color plot of the difference between the semi-empirical mass formula and the actual binding energy, as a grid of Z versus N. Don't plot colors (or plot white) where experimental data is not available.*[42] *Note the horizontal and vertical bands where the theory underestimates the binding energy. Estimate these 'magic numbers'. Do your estimates agree with the Harmonic Fermi sea estimates of an earlier exercise? Do they agree with Wikipedia's list 2, 8, 20, 28, 50, 82, 126?*

N4.37 **Mirror path integrals.** (Path Integrals) ③

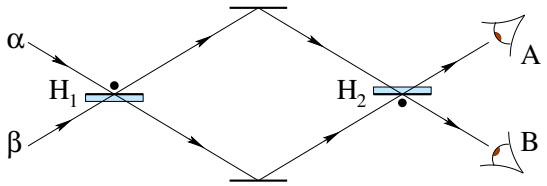


Fig. N4.13 Qbit weirdness. A photon, passing through a pair of half-silvered mirrors H_1 and H_2, undergoes quantum interference between different paths.

One of the most compelling examples of Qbits and their weirdness is provided by the example of photons and half-silvered mirrors. Fig. N4.13 shows a photon[43] coming

[41] Iron 56 is one of the most stable nuclei. It dominates the endpoint of fusion reactions in stars. Lower mass nuclei tend to fuse; higher masses tend to fission. I hear that ^{6}2Ni is even more stable, but is not accessible easily in nuclear reactions.

[42] You can use a *masked array* to do this in Python; in Mathematica you can set ColorRules → 0.0→White.

[43] We assume, as before, that the polarization of the photon lies perpendicular to the plane of the paper, so that it remains unchanged and hence unimportant to the interference.

from the left in a superposition $\binom{\alpha}{\beta}$, through a set of mirrors, to two detectors named Alice (A) and Bob (B). We work in the basis $|1\rangle = \binom{1}{0}$ representing the upper of the two beams at a given position, and $|0\rangle = \binom{0}{1}$ representing the lower of two beams. As discussed in Schumacher & Westmoreland, the half-silvered mirror H_2 approximately acts as a unitary transformation $H_2 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$. The first mirror H_1, with its mirrored side on the top, changes the sign of the beam reflecting from its top, hence $\frac{1}{\sqrt{2}}\begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix}$. The product $G = H_2 H_1$ is analogous to the propagator, or Green's function, for this system.[44]

(a) *What is G? Given the impinging wave $\binom{\alpha}{\beta}$ from the left, what are the probabilities P_A and P_B that Alice and Bob will see the photon?* (Hint: Remember Bob did not see anything when the initial photon came from below.)

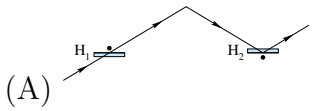
We can develop a kind of discrete path integral representation of the propagator G by writing

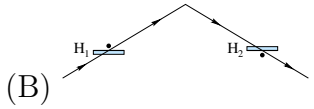
$$G = \mathbb{1} H_2 \mathbb{1} H_1 \mathbb{1} \tag{N4.48}$$

$$= \sum_{x_i=0}^{1} \sum_{x_m=0}^{1} \sum_{x_f=0}^{1} |x_f\rangle\langle x_f| H_2 |x_m\rangle\langle x_m| H_1 |x_i\rangle\langle x_i|. \tag{N4.49}$$

Here i, m, f representing the initial states, the states in the middle, and the final (detected) states, and $\mathbb{1} = |1\rangle\langle 1| + |0\rangle\langle 0|$. If we assume the initial photon is coming from the bottom left ($x_i = 0$), there are four remaining paths in this sum.

(b) *Give the four amplitudes contributed by these four paths. Which ones contribute to $\langle 1|G|0\rangle = G_{10}$? Which ones contribute to $\langle 0|G|0\rangle = G_{00}$? Which go to Bob? Do the sum of the amplitudes going to Bob equal zero (as they should)?*





[44] Note two confusing things in our notation. First, the photon moves from left to right (hitting H_1, then H_2), but the matrices describing the evolution propagate from right to left ($H_2 H_1$). Second, $|1\rangle = \binom{1}{0}$ is a vector whose first element is one [zeroth in Python/C], and $|0\rangle = \binom{0}{1}$ is a vector whose *second* element is one [first in Python/C].

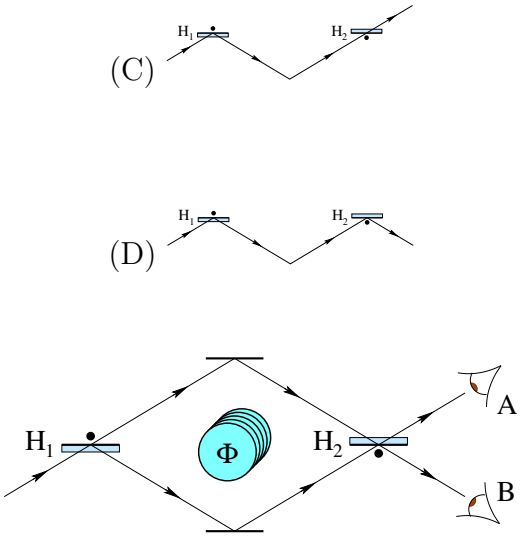


Fig. N4.14 Bohm-Aharonov and mirrors.

Imagine an electron traversing an electron-mirror array, impinging from below. The mirrors H_1 and H_2 have the same effect on the amplitudes as the former half-silvered ones did for the photon. Here, though, we thread a solenoid between the upper and lower paths in the middle region, enclosing a net magnetic flux Φ pointing upward out of the page. The field is zero outside the solenoid, and you may ignore the electron's spin.

(c) *As a function of Φ, what is the probability that an initial electron will be seen by Alice? What values of Φ, in multiples of the elementary flux quantum $\Phi_0 = hc/e = 2\pi\hbar c/e$, prevent Alice from seeing any electrons?* (Remember, the initial electron comes from below. Hints: $\oint_C \mathbf{A}\cdot \mathrm{d}\ell = \Phi$ if the path C encircles the solenoid counterclockwise once. The path-integrand amplitude for $\mathbf{x}(t)$ in a field $\mathbf{A}$ gains a phase $\zeta = \int (q/c)\mathbf{A}(x)\cdot \mathrm{d}\mathbf{x}/\hbar$. The charge on an electron is $q = -e$.)

N4.38 **Harmonic oscillator spectrum: The propagator.** (Path integrals) ③

(a) *Show that the trace of the propagator can be written in terms of the energy eigenvalues:*

$$\int_{-\infty}^{\infty} K(x,t_2;x,t_1)\mathrm{d}x = \sum_n \exp(-\mathrm{i}E_n(t_2-t_1)/\hbar). \tag{N4.50}$$

(Hint: write

$$\begin{aligned} K &= \langle x_2|U(t_2-t_1)|x_1\rangle = \langle x_2|\mathrm{e}^{-\mathrm{i}H(t_2-t_1)/\hbar}|x_1\rangle \\ &= \langle x_2|\mathbb{1}\mathrm{e}^{-\mathrm{i}H(t_2-t_1)/\hbar}|x_1\rangle \end{aligned}$$

and insert a complete set of energy eigenstates for $\mathbb{1}$.)

(b) *Sum the geometrical series in eqn N4.86 for a one-dimensional harmonic oscillator of frequency ω.*

The propagator for the harmonic oscillator is

$$K_{HO}(x_2,t_2;x_1,t_1) = \sqrt{\frac{m\omega}{2\pi \mathrm{i}\hbar \sin(\omega(t_2-t_1))}} \exp\left[\frac{\mathrm{i}m\omega}{2\hbar\sin(\omega(t_2-t_1))}\{(x_2^2+x_1^2)\cos(\omega(t_2-t_1))-2x_2x_1)\right].$$

(There is a typo in Sakurai's formula 2.6.18).

(c) *Write $K_{HO}(x,t_2,x,t_1) = f(t)\exp(-\mathrm{i}A(t)x^2)$. Evaluate the trace. Show that you get the same answer as part (b).* (Hint: Use the Gaussian integral formula $\int_{-\infty}^{\infty}\exp(-\mathrm{i}A(t)x^2) = \sqrt{\pi/(\mathrm{i}A)}$. You may want to use the half-angle formula $\sin(a/2) = \sqrt{\frac{1}{2}(1-\cos(a)}$.)

N4.39 **Evolving Schrödinger: Coherent states.** (Computation) ③

In this exercise, we shall build upon the numerical work of exercise 3.6 (free particle evolution), exercise 4.2 (Baker-Campbell Hausdorff idenity), and exercise 4.3 (Coherent States). We shall solve the time-dependent Schrödinger equation for a harmonic oscillator in its ground state, and after the ground state is translated to the side by a distance x_0.

The time-dependent Schrödinger equation for our one-dimensional quantum system is:

$$\begin{aligned} \mathrm{i}\hbar\frac{\partial\psi}{\partial t} &= -\frac{\hbar^2}{2m}\frac{\partial^2\psi}{\partial x^2} + V(x)\psi = H\psi = H_{\mathrm{kin}}\psi + H_{\mathrm{pot}}\psi \\ \psi(t) &= U(t)\psi(0) = \mathrm{e}^{-\mathrm{i}Ht/\hbar}\psi(0) = \mathrm{e}^{-\mathrm{i}H_{\mathrm{kin}}t/\hbar - \mathrm{i}H_{\mathrm{pot}}t/\hbar}\psi(0) \end{aligned} \tag{N4.51}$$

As in the last exercise, we use the constants for McEuen's bouncing buckyballs, with $m = 60m_C \sim 12*60m_p$ and the frequency to $\omega = 10^{12}$ radians/sec, and will evaluate it at $N_p = 200$ points spanning $L = 30a_0$,

In the free particle example (no potential energy), we advanced time by dt by multiplying the Fourier transform by $U_{\mathrm{kin}}(k,dt) = \exp(\mathrm{i}(\hbar^2k^2/2m)dt/\hbar)$. If, on the other hand, there were no *kinetic* energy (infinite mass), we could solve for the time evolution $\psi(x,t) = U_{\mathrm{pot}}(t)\psi(x,t=0)$ by multiplying $\psi(x)$ in real space by a time-dependent phase depending on position:

$$\begin{aligned} \psi(x,t+dt) &= U_{\mathrm{pot}}(dt)\psi(x,t) = \mathrm{e}^{-\mathrm{i}H_{\mathrm{pot}}dt/\hbar}\psi(x,t) \\ &= \mathrm{e}^{-\mathrm{i}V(x)dt/\hbar}\psi(x,t). \end{aligned} \tag{N4.52}$$

To approximately solve Schrödinger's time evolution, we alternate advancing the wave function in real space and Fourier space, using the Baker–Campbell–Hausdorff formula of exercise 4.2(b):

$$\psi(t+dt) = \mathrm{e}^{-\mathrm{i}H_{\mathrm{kin}}t/\hbar - \mathrm{i}H_{\mathrm{pot}}t/\hbar}\psi(0) \quad \approx \mathrm{e}^{-\mathrm{i}H_{\mathrm{pot}}t/2\hbar}\mathrm{e}^{-\mathrm{i}H_{\mathrm{kin}}t/\hbar}\mathrm{e}^{-\mathrm{i}H_{\mathrm{pot}}t/2\hbar}\psi(0) = U_{\mathrm{pot}} \tag{N4.53}$$

(a) *Define the two arrays* UkinTildeDt *and* UpotDtOver2. *Define the initial wavefunction* $\psi[0](x)$. (Hint: If your implementation stores $\psi[n][x]$ as a two-dimensional complex array, you may want to allocate it and initialize $\psi[0][x]$ as part of that array.)

(b) *Evolve the wavefunction to a time equal to twice the period* P *of the oscillator, in steps of* $dt = P/100$, *storing your answer after each step. Plot* $\psi(x, P/5)$, *showing the real part, the imaginary part, and the absolute value all on the same graph.* (Why don't we plot $|\psi^2(x)|$ on this graph?) *If possible, animate these three curves; otherwise, plot several snapshots until you see the evolution. What happens to the probability density? Why? What happens to the real and imaginary parts? Why?*

(c) *Now shift the wavefunction* $\psi(x, t = 0) = \psi_0(x - x0)$, *with* $x_0 = 10a_0$, *where* a_0 *is the RMS width of the ground state wavefuction (see exercise 3.6). Time evolve as in part (b). How does the evolution compare to a classical particle in the harmonic well?*

(d) *Using your answer to exercise 4.3(d), write the initial wavefunction for part (c) in terms of a coherent state. What is* λ*?*

For photons and phonons and other harmonic systems, the coherent states evolve just as classical particles would.

N4.40 **Coherent State Evolution.** (Operator algebra) ③

Consider the annihilation operator a for a simple harmonic oscillator, transformed into the time-dependent Heisenberg-representation operator $\mathbf{a}_{\mathrm{H}}(t)$:

$$\mathbf{a}_{\mathrm{H}}(t) = \mathrm{e}^{\mathrm{i}\mathcal{H}t/\hbar} a \mathrm{e}^{-\mathrm{i}\mathcal{H}t/\hbar} = U^\dagger(t) a U(t). \tag{N4.54}$$

The time evolution for an operator in the Heisenberg representation is given by the commutator with the Hamiltonian, so

$$\frac{\mathrm{d}\mathbf{a}_{\mathrm{H}}}{\mathrm{d}t} = \frac{\mathrm{i}\mathcal{H}}{\hbar}\mathbf{a}_{\mathrm{H}} - \mathbf{a}_{\mathrm{H}}\frac{\mathrm{i}\mathcal{H}}{\hbar} = -\mathrm{i}/\hbar[\mathbf{a}_{\mathrm{H}}, \mathcal{H}]. \tag{N4.55}$$

You may use the fact that the Hamiltonian for the harmonic oscillator in the Schrödinger representation is $\mathcal{H} = \hbar\omega(a^\dagger a + \frac{1}{2})$, and that $[a, a^\dagger] = 1$.

(a) *Calculate* $[\mathbf{a}_{\mathrm{H}}, \mathcal{H}]$, *and write it in terms of* $\mathbf{a}_{\mathrm{H}}$. *What is* $\mathrm{d}\mathbf{a}_{\mathrm{H}}/\mathrm{d}t$*?* (Simplify your answers until they only involve $\mathbf{a}_{\mathrm{H}}$ and constants, not $\mathcal{H}$ or a.)

(b) *Show that* $\mathbf{a}_{\mathrm{H}}(t) = \exp(-\mathrm{i}\omega t)\mathbf{a}_{\mathrm{H}}(0) = \exp(-\mathrm{i}\omega t)a$ *is the solution to the time evolution you found in part (a).* (Hint: This can also be a check for part (a).)

We discovered in a computational exercise that the probability density for a displaced harmonic oscillator ground state oscillates like a classical particle with the oscillator frequency ω. In another exercise, we showed that a displaced harmonic oscillator ground state is one example of a *coherent state*, an eigenstate of the annihilation operator a:

$$a|\lambda\rangle = \lambda|\lambda\rangle, \tag{N4.56}$$

which is also normalized $\langle\lambda|\lambda\rangle = 1$. Here $\lambda \in \mathbb{C}$ can be any complex number.

(c) *In the Schrödinger representation*[45] *show that a coherent state* $|\lambda\rangle$ *evolves after a time* t *to a state* $|\xi\rangle = U(t)|\lambda\rangle$ *which is also an eigenstate of the annihilation operator* a*. What is its eigenvalue* $\tilde{\lambda}$*?* (Hints: Multiply $a|\xi\rangle = aU(t)|\lambda\rangle$ on the left by $\mathbb{1} = U(t)U(-t)$ and use part (b). You don't need to *compute* $U(t)|\lambda\rangle$, you just need to show it is an eigenstate of a.)

Since there is only one coherent state with eigenvalue $\tilde{\lambda}$, our evolved state $U(t)|\lambda\rangle = C|\tilde{\lambda}\rangle$ for some constant C. Since time evolution conserves probability (and hence $U(t)$ is unitary), $\langle\lambda|U^\dagger(t)U(t)|\lambda\rangle = |C|^2 = 1$, so C is a pure phase.

The phase C depends on time and λ. Let us solve for the case of the oscillator in the ground state.

(d) *Calculate* $C(t)$ *for the special case* $\lambda = 0$*.* (Hint: the coherent state with $\lambda = 0$ is the ground state of the harmonic oscillator. You don't need to know the solutions of previous sections to solve this.)

N4.41 **Decoherence.**[46] (Density Matrices) ③

In this exercise, we will explore the effects of decoherence on a quantum system using density matrices and the Bloch sphere. We will study the dynamics of spin-1/2 particles in a magnetic field, with and without decoherence. We will work in the z-spin basis, and denote the spins pointing parallel to and anti-parallel to the z-direction by $|\uparrow_z\rangle$ and $|\downarrow_z\rangle$. The spins are subjected to a magnetic field $\vec{B} = B\hat{x}$ in the x-direction. Convince yourself that the Hamiltonian modeling this is $\mathcal{H} = -\mu_0 B(|\uparrow_x\rangle\langle\uparrow_x| - |\downarrow_x\rangle\langle\downarrow_x|)$.

(a) *Write this Hamiltonian in the z-spin basis.*

(b) *Suppose the initial wavefunction is* $|\psi(t=0)\rangle = |\uparrow_z\rangle$*. Solve the Schrödinger equation to find* $|\psi(t)\rangle$*. Do you observe that the spin oscillates between* $|\uparrow_z\rangle$ *and* $|\downarrow_z\rangle$*? What is the frequency* ω *of the oscillation?*

Recall from problem 11.1 that any 2×2 density matrix can be written as $\rho = \frac{1}{2}(\mathbb{1}+\vec{n}\cdot\vec{\sigma})$. The vector $\vec{n}$ is called the *Bloch vector*, and always has norm $|\vec{n}| \le 1$, forming the solid *Bloch sphere*. (Remember $\sigma_x = \left(\begin{smallmatrix}0&1\\1&0\end{smallmatrix}\right)$, $\sigma_y = \left(\begin{smallmatrix}0&-\mathrm{i}\\\mathrm{i}&0\end{smallmatrix}\right)$, and $\sigma_z = \left(\begin{smallmatrix}1&0\\0&-1\end{smallmatrix}\right)$.)

(c) *Calculate* $\rho(t)$ *for* $|\psi(t)\rangle$ *from your calculation in part (b), in terms of* ω*. Calculate* $\vec{n}(t)$*, and use the double angle formulas to simplify your answer. Geometrically, what is the trajectory of* $\vec{n}(t)$*? Show that this agrees with your solution*[47] *to exercise 11.1,* $\mathrm{d}\vec{n}/\mathrm{d}t \propto -\vec{B}\times\vec{n}$*.*

(d) *Show that the eigenvalues of a general 2×2 density matrix* $\rho = \frac{1}{2}(\mathbb{1} + \vec{n}\cdot\vec{\sigma})$ *are* $\frac{1}{2}(1\pm|\vec{n}|)$*. What is the entropy* $S = -k_B\mathrm{Tr}(\rho\log\rho)$ *of a general density matrix in terms of* $\vec{n}$*?* (Hint: use the basis in which ρ is diagonal, and your eigenvalues.) *Show that the zero-entropy pure states are those on the surface* $|\vec{n}| = 1$ *of the Bloch sphere.* (Hint: $x\log x$ is negative for $0 < x < 1$, and equal to 0 at the end points $x = 0$ and $x = 1$.) *Does your solution* $\vec{n}(t)$ *from part (c) stay a zero entropy pure state, as it should?*

[45] As opposed to the Heisenberg representation of part (b).

[46] Developed in collaboration with Bhuvanesh Sundar.

[47] That exercise had funny units, but the form of the equation and the sign should agree.

Decoherence arises in a system due to interaction with a large environment. Essentially, the universe is constantly looking at our system, and as a result of interaction with the rest of the universe, our spins get entangled with the universe. Since we observe only the spins and do not observe the infinitely many degrees of freedom in the rest of the universe, it appears to us that the spins lose information about any coherences they may have developed.

(e) *How does a general density matrix* $\rho = \left(\begin{smallmatrix}\rho_{11} & \rho_{12}\\ \rho_{21} & \rho_{22}\end{smallmatrix}\right)$ *written in the* z *basis change when its* s_z *component is measured? Show that the effect of a measurement in the* z*-basis is to project* $\vec{n}$ *onto the* z*-axis.*

For the remainder of the exercise, we consider the evolution under the specific Hamiltonian $\mathcal{H}$ you studied in parts (a) through (c). We shall model decoherence as a measurement being done on the spins with small probability Γ per unit time. In a small time interval δt, the spin is measured in the z-basis with a probability $\Gamma\delta t$, and not measured with a probability $1-\Gamma\delta t$, and then the system evolves for a time δt.

(f) *What is* $\vec{n}(t+\delta t)$ *in terms of* $\vec{n}(t)$*, including first the possibility of observation and then the time evolution from* $\mathcal{H}$*? Write a differential equation for the components of* $\vec{n}(t)$ *by taking* $\delta t \to 0$*.*

(g) *Show that* n_z *obeys the second-order differential equation for a damped harmonic oscillator,* $\mathrm{d}^2 n_z/\mathrm{d}t^2 + \eta \mathrm{d}n_z/\mathrm{d}t + \omega_0^2 n_z = 0$*. What are* η *and* ω_0 *in terms of* Γ *and* ω*?*

(h) *What is the long-time limit for* $\vec{n}$*? For* ρ*? For the entropy?*

N4.42 **Quantum Algorithms.**[48] (Quantum Information Processing) ④

Are quantum computers faster than our standard classical computers?

Clearly, we need to define our terms here—since factoring 143 (the current quantum-computing world record) doesn't take long on a classical computer. The key question is how the computer time would scale for large problems. Factoring M-digit numbers on a quantum computer takes no more than $O(M^3)$ time (that is, some constant times M^3, using Shor's algorithm), while the most efficient known method for factoring on classical computers takes $O(e^{1.9M^{1/3}(\log M)^{2/3}})$. For large numbers of digits, quantum computers win (if they can be built). There are a few other problems where classical computers are known to be much slower than quantum computers: Grover's algorithm for searching an unsorted database, Simon's algorithm, ...

Here we explore a somewhat artificial problem, solved in the quantum case using the Deutsch-Josza algorithm.[49] This will also introduce the *reversibility* of quantum computing, and the use of *gates*—unitary operators that transform Qbits to execute the quantum computer program.

The Deutsch-Josza algorithm considers functions that map n bits to one bit. Let us denote the n bits as $x_0, x_1, \cdots, x_{n-1}$, where $x_0, x_1, \cdots, x_{n-1}$ are all 0 or 1. Let

[48] Developed in collaboration with Bhuvanesh Sundar, based on an exercise by Paul Ginsparg.

[49] There are many discussions of the Deutsch-Josza algorithm in the literature—feel free to consult them. If you find one that is particularly pertinent, reference it properly in your writeup.

$\mathbf{x} = \sum_{m=0}^{n-1} 2^m x_m$ denote the integer represented by the bit sequence $x_0, x_1, \cdots, x_{n-1}$; x_0 is the least significant bit and x_{n-1} is the most significant bit. We'll also denote $|x_0\rangle|x_1\rangle\cdots|x_{n-1}\rangle$ as $|\mathbf{x}\rangle$.

Let f be a function that maps the n bits $x_0, x_1, \cdots, x_{n-1}$ to one bit (that is, either True or False, one or zero). For example, f could be a function Prime that returns True if the integer $\mathbf{x}$ is prime, or Even that returns True if the integer is divisible by two, or Big that returns True if the integer is greater than or equal to 2^{n-1}. Our algorithm is not concerned with implementing $f(\mathbf{x})$, but with testing properties of an unknown f by sampling its output. For example, testing whether f is a *constant function* (either True for all possible $\mathbf{x}$, or False for all arguments) is a challenge for classical computers. (An experiment may find a thousand Trues in a row, but to be sure that the function always returns True one must test all 2^n choices of $\mathbf{x}$.) We define a *balanced* function to be one which returns True for exactly half of the possible inputs. Thus Even and Big above are balanced, but Prime is neither balanced nor constant.

(a) *Write the four possible functions $f(x_0)$ for $n = 1$ (two possible inputs, two possible outputs). Which are constant? Which are balanced?* For larger n, most possible functions are neither balanced nor constant.

Deutsch and Josza considered the artificial problem of distinguishing between balanced and constant functions. Let us define DJ functions to be those functions guaranteed to be either balanced or constant. Given that f is a DJ function, can a quantum computer probing f distinguish between the two cases faster than a classical computer? Let us first consider how a classical computer would solve this.

(b) *Argue that in the worst case, the n-bit DJ function f would have to be called $2^{n-1}+1$ times in order to determine for certain whether it is balanced or constant.*

Our challenge is to use a quantum computer program to do this calculation with *one* operation of the operator f. How do we set this up?

A quantum computer performs unitary operations on Qbits to execute the program. Unitary operations are reversible;[50] indeed, the only irreversible step in a perfect quantum computer is the macroscopic observer reading the answer. This means that no quantum computer can perform the classical AND operation, for example—since $AND(x_0, x_1)$ is False for three different values of x_0 and x_1, it would throw out information that could not be retrieved. The workaround is to encode the answer in a final Qbit y. So if $n = 2$ and $f(x_0, x_1) = AND(x_0, x_1)$ (a function that is neither balanced nor constant), we could implement f on a quantum computer by writing a program that took $|x_0\rangle|x_1\rangle|y\rangle$ and returned $U_f(|x_0\rangle|x_1\rangle|y\rangle) = |x_0\rangle|x_1\rangle|y \oplus AND(x_0, x_1)\rangle$ where $\oplus$ is addition modulo 2. If you input $|x_0\rangle|x_1\rangle|y = 0\rangle$, the output value of $|y\rangle$ gives $f(\mathbf{x})$. If you input $|x_0\rangle|x_1\rangle|y = 1\rangle$, the output value of $|y\rangle$ gives $1 \oplus f(\mathbf{x}) = NOT(f(\mathbf{x}))$—this feature is important to keep U_f reversible. U_f is also linear: for example, $U_f((\alpha|0\rangle + \beta|1\rangle)|0\rangle|y\rangle) = \alpha|0\rangle|0\rangle|y \oplus f(0,0)\rangle + \beta|1\rangle|0\rangle|y \oplus f(1,0)\rangle$.

[50] The reverse operation is $U^\dagger = U^{-1}$.

(c) *Show that U_f is reversible for the case where $f = AND$ by giving an explicit method for reconstructing x_0, x_1, and y from x_0, x_1, and $y \oplus AND(x_0, x_1)$. Then show in general that U_f is its own inverse for any n-bit function f.*

We are now given a quantum computer operation that evaluates an unknown DJ function f: $U_f(|x_0\rangle|x_1\rangle\cdots|y\rangle) = |x_0\rangle|x_1\rangle\cdots|y \oplus f(\mathbf{x})\rangle)$.

Just as a classical computer can be made of AND gates, NOT gates, OR gates, etc., so a quantum computer is composed of gates that manipulate one or two Qbits by application of unitary operators. The single-Qbit gates are thus 2×2 unitary matrices.

(d) *In the basis*[51] $|0\rangle = \left(\begin{smallmatrix}1\\0\end{smallmatrix}\right)$ *and* $|1\rangle = \left(\begin{smallmatrix}0\\1\end{smallmatrix}\right)$, *write the single-Qbit gate NOT as a* 2×2 *matrix. Show that the Hadamard gate, written* $H = 1/\sqrt{2}\left(\begin{smallmatrix}1 & 1\\1 & -1\end{smallmatrix}\right)$, *is unitary.* We can implement both the NOT gate and the H gate on electron Qbits, for example, by exposing them in a magnetic field with a suitable direction and orientation.

Our strategy will be to apply U_f not to a Qbit product that corresponds to a classical bit sequence $|x_0\rangle|x_1\rangle\cdots|x_{n-1}\rangle$, but rather to a Qbit string that represents a quantum superposition of all possible classical bit sequences. Let us first consider[52] the case $n = 1$. Our strategy is to use the Hadamard gate to create a superposition of bit sequences and then apply U_f, and then re-apply the Hadamard gate to find out whether our function is constant or balanced. We shall abuse notation to use $H^{n+1}|\mathbf{x}\rangle|y\rangle$ to mean $(H|x_0\rangle)(H|x_1\rangle)\cdots(H|x_{n-1}\rangle)(H|y\rangle)$.

(e) *Starting with the case of $n = 1$ Qbit plus $|y\rangle$, initialize our two Qbits to $|\Psi_0\rangle = |x_0\rangle|y\rangle = |0\rangle|1\rangle$. Apply the Hadamard operation on both Qbits (exposing them both to the same magnetic field). What is the resulting superposition? Apply U_f for the four cases of f you found in part (a), and then apply the Hadamard transformation on both the Qbits again. What is the measured final value of x_0 for the constant functions with $n = 1$? For the balanced functions?*

You should have found that you could conclusively say if f were constant or balanced with just one call to U_f.

Now that you have worked out the $n = 1$ case, let us generalize to arbitrary n. The algorithm proceeds in the same way. We initialize each of the n Qbits $|x_0\rangle, |x_1\rangle, \cdots, |x_{n-1}\rangle$ to $|0\rangle$, and $|y\rangle$ to $|1\rangle$, so $|\Psi_0\rangle = |0\rangle^n|1\rangle$. We perform the Hadamard operation on all the Qbits, so $|\Psi_1\rangle = H^{n+1}|\Psi_0\rangle = (H|0\rangle)^n(H|1\rangle)$. We pass them through U_f, so $|\Psi_2\rangle = U_f|\Psi_1\rangle$. We perform another Hadamard operation on all the Qbits, $|\Psi_3\rangle = H^{n+1}|\Psi_2\rangle$. Finally, we measure the overlap with the initial state, $|\langle\Psi_0|\Psi_3\rangle|^2$, measuring the probability that $x_0 = 0$, $x_1 = 0$, ..., $x_{n-1} = 0$, $y = 1$.

Let us do this step by step. The initial state of the Qbits is $|\Psi_0\rangle = |0\rangle^n|1\rangle$. A Hadamard operation is then applied on all of them. The state of the Qbits after this operation is $|\Psi_1\rangle = H^{n+1}|\Psi_0\rangle = (H|0\rangle)^n(H|1\rangle)$.

[51] I apologize for the shift in notation: I used the in Exercise N4.43, $|1\rangle = \left(\begin{smallmatrix}1\\0\end{smallmatrix}\right)$ and $|0\rangle = \left(\begin{smallmatrix}0\\1\end{smallmatrix}\right)$.

[52] The special case $n = 1$ of the Deutsch-Josza algorithm is called the Deutsch's algorithm.

(f) *Write* $|\Psi_1\rangle$ *as a superposition of* $|\mathbf{x}\rangle|y\rangle$ *for all possible* n*-bit binary numbers* $\mathbf{x}$ *and both values of* y*. Show that the probabilities of being in these states are all equal (but the amplitudes may have different signs).*

Now the Qbits are passed through U_f. The state of the Qbits after passing through U_f is $|\Psi_2\rangle = U_f|\Psi_1\rangle$. When f is a *constant* function, U_f changes the Qbit y in the same way for all arguments $\mathbf{x}$.

(g) *If* f *is a constant function, show that* $|\Psi_2\rangle$ *is a constant times* $|\Psi_1\rangle$*. What is this constant if* $f \equiv 0$*? If* $f \equiv 1$*? Show that the measured values of* $x_0, x_1, \cdots, x_{n-1}$ *in* $|\Psi_3\rangle$ *after the final Hadamard operation are all* 0*, so in particular that* $|\langle\Psi_0|\Psi_3\rangle|^2 = 1$*.* (Hint: Do not try to apply H^{n+1} on $|\Psi_2\rangle$ written as a superposition of several terms. Instead, decompose $|\Psi_2\rangle$ as a product of a state for each Qbit ($|\Psi_2\rangle = \Pi_{0\le i<n}|\phi_i\rangle$ where $|\phi_i\rangle$ is a superposition of $|0\rangle$ and $|1\rangle$), and use the fact that $H^2 = \mathbb{1}$).

Hence for a constant function, the result of our quantum computation always has unit probability of returning the state with all $x_i = 0$ and $y = 1$.

When f is a balanced function, the value in the Qbit y is changed differently for different arguments $\mathbf{x}$; for half of those 2^n arguments, y is left unchanged, and for half of those arguments, y is flipped (from 0 to 1 or vice versa). Let us illustrate this with an example: let us consider the function Even, which returns True if the integer $\mathbf{x}$ is divisible by two.

(h) *What is the least significant bit of an integer if it was even? If the integer was odd? Argue that whether the Qbit* y *is flipped by* U_f *or not is determined solely by the least significant bit* x_0 *in* $|\Psi_1\rangle$*.* We know that the Qbits were in a product state (a product of single Qbits) $|\Psi_1\rangle = (H|0\rangle)^n(H|1\rangle)$ before passing through U_f. *Show that the Qbits are in a product state after passing through* U_f *as well (i.e.* $|\Psi_2\rangle = \Pi_{0\le i<n}|\phi_i\rangle$*), and write this product explicitly. Is* $|\Psi_2\rangle$ *different from* $|\Psi_1\rangle$*? What are the measured values of the Qbits* $x_0, \cdots, x_{n-1}$ *after the final Hadamard operation?* (Hint: Perform the Hadamard operation on each term in the product above, and use the fact that $H^2 = \mathbb{1}$.)

You should have found that $x_0, \cdots, x_{n-1}$ are measured to be something other than all zeros. The Deutsch-Josza algorithm states that for any balanced function f, the probability of measuring $x_0, \cdots, x_{n-1}$ to be all zeros is 0. We'll prove this in the following way.

(i) *Show that for an arbitrary balanced function* f*,*

$$|\Psi_2\rangle = U_f|\Psi_1\rangle = \frac{1}{2^{\frac{n+1}{2}}} \sum_{0\le\mathbf{x}<2^n} (-1)^{f(\mathbf{x})}|\mathbf{x}\rangle(|0\rangle - |1\rangle) \tag{N4.57}$$

$$= \frac{1}{2^{\frac{n+1}{2}}} \left(\sum_{\mathbf{x}:f(\mathbf{x})=0} |\mathbf{x}\rangle - \sum_{\mathbf{x}:f(\mathbf{x})=1} |\mathbf{x}\rangle \right) (|0\rangle - |1\rangle)\,.$$

After the final Hadamard operation, show that the probability of measuring $x_0 = 0$*,* $x_1 = 0$*, ...,* $x_{n-1} = 0$*,* $y = 1$ *is zero, i.e.* $\langle\Psi_0|(H^{n+1}|\Psi_2\rangle) = 0$*.* (Hint: Rather than cal-

culating $\langle\Psi_0|\Psi_3\rangle$, calculate the same quantity in the form $\langle\Psi_1|\Psi_2\rangle = (\langle\Psi_0|H^{n+1})|\Psi_2\rangle = \langle\Psi_0|(H^{n+1}|\Psi_2\rangle) = \langle\Psi_0|\Psi_3\rangle$, and use eqn N4.97.)

Hence with one application of the function f, with 100% certainty a constant function returns the initial state and a balanced function with 100% certainty will never return the initial state.

It is amazing that we could determine whether a DJ function f is constant or balanced in just *one* evaluation of U_f. The Deutsch-Josza algorithm achieves an *exponential* speedup over its classical counterpart. The problems considered in the above (Deutsch and Deutsch-Josza) algorithms may seem far removed from applications to real world problems, but these algorithms paved the way for more complicated and powerful algorithms.

Why are we still factoring 143? The great challenge in building quantum computer is *decoherence*, the tendency of Qbits to interact with the environment and go from quantum superpositions into mixtures.

N4.43 **Supersymmetric harmonic oscillator.**[53] (Quantum, Supersymmetry) ③

One of the main predictions of supersymmetry[54] is that each particle comes with a supersymmetric partner with the same mass but with opposite statistics.[55] For example, the fermionic electron is paired with the bosonic selectron. Supersymmetry is also a potential symmetry of nature, with an unusual connection to the translational symmetries in space and time (the Poincaré group). Finally, supersymmetry allows one to calculate remarkable things about certain Hamiltonians. In this exercise, we shall explore a "zero-dimensional"[56] example of a supersymmetric Hamiltonian, and try to illustrate each of these features of supersymmetry.[57]

Remember the commutation relations for creation and annihilation operators suitable for bosons

$$[a, a^\dagger] = 1 \qquad [a, a] = [a^\dagger, a^\dagger] = 0, \tag{N4.58}$$

and fermions

$$\{b, b^\dagger\} = 1 \qquad \{b, b\} = \{b^\dagger, b^\dagger\} = 0. \tag{N4.59}$$

[53] Developed in collaboration with John Stout, Fall 2013.

[54] The footnotes in this problem are meant as inspiration—tying it to fundamental ideas in theoretical physics. *None of the footnotes are necessary or useful for solving the problem*—ignore them if you wish.

[55] Supersymmetric partners have the same mass as long as supersymmetry is unbroken. We expect supersymmetry to be *spontaneously broken* at low energy scales, given that we have not yet detected any supersymmetric partners of the Standard Model particles.

[56] We often talk about quantum field theories in d spatial dimensions and one time dimension as d+1-dimensional field theories: our space-time is thus 3+1 dimensional. We can view nonrelativistic quantum mechanics as a $d = 0$ quantum field theory, and it is in this regard that we consider the supersymmetric Hamiltonians described here as "zero-dimensional" or 0+1-dimensional.

[57] There are a number of discussions of the supersymmetric harmonic oscillator and zero-dimensional supersymmetry in the literature and on the Web. Feel free to consult these. If you find one particularly useful, reference it properly in your writeup.

where $[A, B] = AB - BA$ is the commutator and $\{A, B\} = AB + BA$ is the anticommutator.[58] For this simple example, we take our bosons and fermions to be noninteracting, so their creation and annihilation operators commute,

$$[a, b] = [a, b^\dagger] = [a^\dagger, b] = [a^\dagger, b^\dagger] = 0. \tag{N4.60}$$

In one dimension, the Hamiltonian of the simple harmonic oscillator of frequency ω can be written either in terms of x and p:

$$\mathcal{H}_B = p^2/2m + \tfrac{1}{2}m\omega^2 x^2 \tag{N4.61}$$

or in terms of the creation and annihilation operators

$$\mathcal{H}_B = \hbar\omega(a^\dagger a + \tfrac{1}{2}). \tag{N4.62}$$

Here $\frac{1}{2}\hbar\omega$ is the ground state energy of the harmonic oscillator—the zero-dimensional analogue of the *vacuum energy* in field theory.

The harmonic oscillator Hamiltonian can be written in a more symmetric way by using the anticommutator.

(a) *Show that* $\mathcal{H}_B = \frac{1}{2}\hbar\omega\{a^\dagger, a\}$. *Is the vacuum energy still* $\frac{1}{2}\hbar\omega$*?*

Note that we're now calling the ladder operators a and $a^\dagger$ creation and annihilation operators. In this new language, the n^{th} excited state of the harmonic oscillator can be viewed as a state with n bosons.

Define a "fermionic harmonic oscillator" in analogy to the bosonic one, $\mathcal{H}_F = \frac{1}{2}\hbar\omega[b^\dagger, b]$. Again, we can view the n^{th} excited state as a state of n fermions.

(b) *What is the ground state energy of* $\mathcal{H}_F$*? How many fermions are in the ground state, in this new language? What is the energy of the state with one fermion?*

(c) *If we write the zero-fermion state as*[59] $|0\rangle = \left(\begin{smallmatrix}0\\1\end{smallmatrix}\right)$ *and the one-fermion state as* $|1\rangle = \left(\begin{smallmatrix}1\\0\end{smallmatrix}\right)$*, then write* b*,* $b^\dagger$*, and* $\mathcal{H}_F$ *in terms of the three Pauli matrices* σ_x*,* σ_y*, and* σ_z*. Check that your form for* b *and* $b^\dagger$ *satisfy the anticommutation relations of eqn N4.99*
(Remember $\sigma_x = \left(\begin{smallmatrix}0&1\\1&0\end{smallmatrix}\right)$, $\sigma_y = \left(\begin{smallmatrix}0&-\mathrm{i}\\\mathrm{i}&0\end{smallmatrix}\right)$, and $\sigma_z = \left(\begin{smallmatrix}1&0\\0&-1\end{smallmatrix}\right)$.)

We can write our first supersymmetric Hamiltonian by adding the boson and fermion harmonic oscillators:

$$\mathcal{H}_S = \mathcal{H}_B + \mathcal{H}_F = \tfrac{1}{2}\hbar\omega\left(\{a^\dagger, a\} + [b^\dagger, b]\right). \tag{N4.63}$$

[58] Be sure to avoid getting confused by our multiple uses of the terms boson and fermion in this exercise. There are really three different ways we use the terms, each extremely useful and compelling. They are:

N4.1 the objects which vibrate or have spins, that produce harmonic oscillators or two-state systems,

N4.2 the 'primitive' bosons (and fermions) which are excitations within a harmonic oscillator (e.g., N bosons = Nth excited state inside the vibrating object)

N4.3 the composite objects inside the supersymmetric Hamiltonian that merge zero or more 'primitive' bosons and fermions.

[59] Here we use the convention $|1\rangle = \left(\begin{smallmatrix}0\\1\end{smallmatrix}\right)$ and $|0\rangle = \left(\begin{smallmatrix}1\\0\end{smallmatrix}\right)$, instead of the notation used in quantum computing.

Note that the ground state energy for this Hamiltonian is zero.[60]

This supersymmetric Hamiltonian is not particularly difficult to solve. Because there is no interaction between the bosonic and fermionic parts of the Hamiltonian, the solution separates and the eigenstates are just products $\psi(x)\chi(s)$, and the energy of the eigenstate is the sum of the Fermi and Bose energies.

Remember that a composite particle with an odd number of primitive fermions is a fermion—so half of our eigenstates represent composite bosons, and half represent composite fermions.

(d) *Solve for the energies for the eigenstates of $\mathcal{H}_S$. Which eigenstates represent composite fermions? Which composite bosons? Draw the level diagram for $\mathcal{H}_S$, with the first few composite boson eigenenergies as a column of horizontal lines on the left, and the first few composite fermion eigenenergies on the right. On each line, write the number of primitive bosons and fermions making up the composite. Is there a composite fermion state for each composite boson state? What state is the exception?* We shall hitherto drop the "composite" label. If we interpret the energy of a state as the mass of a particle[61] , supersymmetry gives us for every fermion a boson with the same mass.

The fact that our Hamiltonian has (almost) one fermion state for each boson state is a result of an unusual symmetry of the Hamiltonian. To see this, let's define an operator, called the *supercharge*,

$$Q = b\left(\frac{p}{\sqrt{m}} + \mathrm{i}\sqrt{m}\omega x\right) = \mathrm{i}\sqrt{2\hbar\omega}ba^\dagger. \tag{N4.64}$$

(Remember that $x = \sqrt{\hbar/2m\omega}(a^\dagger + a)$ and $p = \mathrm{i}\sqrt{m\omega\hbar/2}(a^\dagger - a)$.)

(e) *Show that $[\mathcal{H}_S, Q] = 0$.* (Hence Q is a symmetry of the Hamiltonian.) *Show that Q acting on a fermion state gives a constant times a boson state of the same energy, and that $Q^\dagger$ acting on a boson state almost always gives a constant times a fermion state of the same energy. Which of the ground states is the exception to this rule? Show that this ground state is an eigenfunction of Q and $Q^\dagger$ with eigenvalue zero.*[62]

Supersymmetry has been shown (by Haag, Lopuszanski, and Sohnius[63]) to be the only way to consistently extend the symmetries of spacetime. Spacetime has a spatial

[60] That is a hint for part (b).

[61] We can motivate this by remembering that we are dealing with a theory with zero spatial dimensions, and so the usual relativistic energy of a particle (which should correspond to a eigenstate of our Hamiltonian) $E = \sqrt{\mathbf{p}^2c^2 + m^2c^4}$ reduces to $E = mc^2$. We often interpret the mass of a particle as being the energy required to create a "copy" of the particle at rest, and it is analogous to the band gap energy in semiconductors.

[62] $Q\Psi = 0$ gives us a first-order differential equation which can be directly integrated to obtain this ground state wave function! This trick extends to field theory applications too—yet another way in which supersymmetry simplifies theorists lives.

[63] The story starts with the Coleman-Mandula no-go theorem in 1967. (According to n-Lab, a no-go theorem is "any theorem...that shows that an idea is not possible even though it may appear as if it should be." Thus Bell's theorem is a no-go theorem dictating the impossibility of local, hidden variable theories that reproduce the predictions of quantum mechanics.) The Coleman-Mandula theorem tells us that in a realistic quantum field theory, space-time symmetries (like the Lorentz group) can only be combined with internal

translational symmetry (with an associated conserved momentum), a time-translational symmetry (associated with the conserved energy, with the Hamiltonian giving the infinitesimal time-translation operator), and other symmetries (rotations and relativistic boosts). Combining these symmetries gives us the Poincaré group.

In our "zero-dimensional" harmonic oscillator, only the time-translational symmetry remains from the Poincaré group. How does supersymmetry extend time-translation invariance? Can we somehow create a time translation by supersymmetrically transforming it?

(f) *Show that* $\mathcal{H}_S = \frac{1}{2}\{Q, Q^\dagger\}$. We see that a combination of two supercharges generates a time translation!

The supersymmetric harmonic oscillator we looked at above may seem pretty trivial: how hard is it to get degenerate states when all states have the same energy splitting?

However, we can generate lots of interacting supersymmetric Hamiltonians by specifying a supercharge

$$Q_W = b\left(\frac{p}{\sqrt{m}} + \mathrm{i}\sqrt{m}W'(x)\right) \tag{N4.65}$$

where $W'(x) = \mathrm{d}W/\mathrm{d}x$, and requiring that $\mathcal{H}_W = \frac{1}{2}\{Q_W, Q_W^\dagger\}$, where the real function $W(x)$ is called the *superpotential*.

Our Hamiltonian $\mathcal{H}_S$ can be viewed as the special case of $W(x) = \frac{1}{2}\omega x^2$. Note that our superpotential need not have units of energy.

(g) *Show that* Q_W *and* $\mathcal{H}_W$ *are 2×2 matrices*

$$\begin{aligned} \mathcal{H}_W &= \begin{pmatrix} \mathcal{H}_1 & 0 \\ 0 & \mathcal{H}_2 \end{pmatrix} \\ Q_W &= \begin{pmatrix} 0 & 0 \\ A & 0 \end{pmatrix}. \end{aligned} \tag{N4.66}$$

where the elements of the matrices are functions of p *and* x. (Hint: Remember $p = -\mathrm{i}\hbar\partial/\partial x$. You might check this against the Web, which has different units.)

---

symmetries (like the SU(3) of the strong interaction) in a trivial way (so that the total symmetry group is (space-time symmetry)×(internal symmetry group).

How did the no-go theorem go? You may remember, according to Noether's theorem, that all continuous symmetries are associated with conserved quantities: thus momentum and energy are the conserved quantities related to translations in space and time, and conversely $\mathbf{p}$ and $\mathcal{H}$ (or P^j and P^0 in four-vector notation) generate infinitesimal space and time translations. Coleman and Mandula showed that spacetime symmetry generators had to commute with generators of any new internal symmetries represented by commutation relations.

Haag, Lopuszanski, and Sohnius were able to skirt the Coleman-Mandula theorem by avoiding the hidden assumption that the new symmetry had to obey commutation relations: the new supersymmetries involve *anticommutation* relations. In fact, they were able to show that this is the *only* way of extending the Poincaré group for consistent, interacting quantum field theories with massive particles.

There is a lovely relationship between the eigenvalues and eigenfunctions of $\mathcal{H}_1$ and $\mathcal{H}_2$, two seemingly different Hamiltonians. Let $\Psi_n^{(1)}(x)$ and $\Psi_m^{(2)}(x)$ be the n-th and m-th eigenfunctions of $\mathcal{H}_1$ and $\mathcal{H}_2$, respectively.

(h) *Using the fact that* $[\mathcal{H}_W, Q_W] = [\mathcal{H}_W, Q_W^\dagger] = 0$*, show that* $A^\dagger\Psi_m^{(2)}(x)$ *is an eigenstate of* $\mathcal{H}_1$ *and* $A\Psi_n^{(1)}(x)$ *is an eigenstate of* $\mathcal{H}_2$*.* (Thus, if we know the eigenfunctions and eigenenergies of one of the Hamiltonians, we know them for the other.)

Let us work out a specific example. Consider $W'(x) = (\pi\hbar/mL)\cot(\pi x/L)$.

(i) *Show that* $\mathcal{H}_1$ *is the particle-in-a-box Hamiltonian (Fig. N4.16) shifted by a constant to set its ground state energy to zero. Show that* $\mathcal{H}_2$ *is a Hamiltonian with potential*[64]

$$V(x) = \frac{\pi^2\hbar^2}{2mL^2}\left(2\csc^2\left(\frac{\pi x}{L}\right) - 1\right). \tag{N4.67}$$

Using the first excited state $\Psi_2^{(1)}(x) = \sqrt{2/L}\sin(2\pi x/L)$ *of* $\mathcal{H}_1$ *and the operator* A*, generate the ground state of* $\mathcal{H}_2$ *and show that it is proportional to* $\sin^2(\pi x/L)$*. Explicitly show (taking the derivatives) that* $A\Psi_2^{(1)}(x)$ *is an eigenfunction of* $\mathcal{H}_2$ *and thus verify that its energy is the same as that of* $\Psi_2^{(1)}$*.*

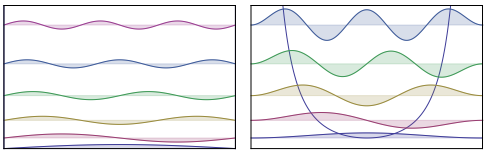


Fig. N4.15 Supersymmetric eigenenergies and eigenstates. (Left) Eigenstates for $\mathcal{H}_1$, the square well potential, displaced vertically by their eigenenergies. (Right) Eigenstates for $\mathcal{H}_2$, the $\csc^2 x$ potential, which is the supersymmetric pair for the square well.

While supersymmetry may not exist in nature, it has proved to be an excellent tool for gaining insight into the way theories with gauge symmetry behave. (For example, we have no proof that the strong interaction confines quarks, but Seiberg and Witten were able to demonstrate confinement in certain supersymmetric theories.) It also has allowed physicists to prove theorems in pure mathematics. Ed Witten, high-energy theorist at the Institute for Advanced Study, was awarded the Fields Medal (the Nobel equivalent in math) for his use of supersymmetry to figure out topological properties of a manifold (such as the Euler characteristic, related to the number of holes or handles a manifold has) by using the difference in the number of zero-energy fermion and boson wavefunctions on it.[65]

[64] Note that the $\Psi = 0$ boundary conditions for the two Hamiltonians are the same for both $\mathcal{H}_1$ and $\mathcal{H}_2$.
[65] E. Witten, *Supersymmetry and Morse Theory*, J. Diff Geom. **17**, 661–692 (1982).

N4.44 **Fourier series and group representations.** (Math) ③

In class, we focused on finite-dimensional group representations for finite groups. In quantum mechanics, the most useful symmetries are often continuous, and Hilbert space is infinite dimensional. With some small modifications, all of our results can go through to the continuous case.

Here we apply group representation theory to the continuous rotations in the plane, $SO(2)$. Let $g_\phi \in SO(2)$ be the rotation by angle ϕ.[66]

(a) *Show that every different g_ϕ is in its own conjugacy class.* (This is true for any commutative group.)

Thus we may label the conjugacy classes by the angle ϕ.

Consider the action of g_ϕ on a function $f(\theta)$:

$$R(g_\phi) : f(\theta) \to f(\theta - \phi). \tag{N4.68}$$

Here θ represents a point on a circle, the complex function $f(\theta)$ is a vector in the Hilbert space of complex functions[67] on the circle, and $R(g_\phi)$ is a linear mapping of that function space into itself.[68]

(b) *Show that, for any nonnegative integer m, that the two-dimensional space spanned by the basis $\{\cos(m\theta), \sin(m\theta)\}$ is an invariant subspace under $SO(2)$. Give the explicit 2×2 matrix for $R(g_\phi)$ acting on this subspace in this basis. What is the character $\chi(\phi)$ of this representation?* (Hint: Use the angle addition formulas. Check that the character of the identity is the dimension of the representation.)

In the space of complex functions on the circle, this two-dimensional representation is not irreducible. It can be decomposed into two invariant subspaces, with bases $\{e^{im\theta}\}$ and $\{e^{-im\theta}\}$.

(c) *What is the character of the m-representation given by the one-dimensional invariant subspace of multiples of $\{e^{im\theta}\}$?*

Thus the character table for $SO(2)$ would have an infinite number of rows (one for each integer $\pm m$) and a continuous infinity of columns (one for each angle ϕ).

For finite groups, we decomposed representations into irreducible pieces using the *little orthogonality theorem*: for any two irreducible representations α and β, the sum over group elements $\sum_{g\in G} \chi^{(\alpha)}(g)\chi^{(\beta)}(g)^* = o(G)\delta_{\alpha\beta}$, where $o(G)$ is the number of elements of the group. For continuous groups, the sum must be replaced by an integral over

[66] This exercise is mostly about understanding the definitions. If you find resources on the Web or elsewhere that are helpful, just properly acknowledge them. In particular, I found http://www.cmth.ph.ic.ac.uk/people/d.vvedensky/groups/Chapter8.pdf which discusses the application of group reps to $SO(2)$. No guarantees that my conventions agree with those in the literature, though.

[67] Particularly, L^2 functions on the circle.

[68] In the past, we viewed group representations as mappings of the group into spaces of matrices that preserve multiplication. But matrices are just linear transformations of vectors; here we are using infinite dimensional vectors instead. Thus $R(g)$ is a linear map taking a function to another function.

the group,[69] and the number of elements of the group replaced by the "volume" of the group. For two-dimensional rotations, we find

$$\int_0^{2\pi} \mathrm{d}\phi\, \chi^{(\alpha)}(\phi)\chi^{(\beta)}(\phi)^* = 2\pi\delta_{\alpha\beta}. \tag{N4.69}$$

(d) *Show that the characters of your irreducible representations from part (c) satisfy the orthogonality relation N4.125. Is the character of your reducible representation in part (b) orthogonal to all the irreducible representations? Use the little orthogonality relation explicitly to decompose this reducible representation into its irreducible components.*

For finite-dimensional representations of finite groups, we knew that any representation could be decomposed into irreducible representations: that is, any general vector could be written as a sum of vectors in the different invariant subspaces. For example, in Alemi's analysis of vibrations in a triangular molecule, he found the normal modes by using a projection operator

$$P^{(\alpha)} = (f^{(\alpha)}/o(G)) \sum_{g\in G} \chi^{(\alpha)}(g)^* R(g). \tag{N4.70}$$

Here[70] $f^{(\alpha)}$ is the dimension of the representation α.

For example, any random deformation of the molecule, when averaged over the group, gave a uniform dialation of the triangle. This dialation is invariant under triangular symmetries—so it transforms under the representation A_1. Since $\chi^{(A_1)}(g) \equiv 1$, this is just what eqn N4.126 suggests. When Alex multiplied by the characters of the two-dimensional representation E (using $P^{(E)}$ in eqn N4.126), though, he discovered a different normal mode that was doubly degenerate.

Let us return now to our infinite-dimensional space of complex functions on the circle, to connect our irreducible representation decomposition with the theory of Fourier series. For our continuous group $SO(2)$, the corresponding projection operator is

$$P^{(\alpha)} = (1/2\pi) \int_0^{2\pi} \mathrm{d}\phi \chi^{(\alpha)}(\phi)^* R(g_\phi). \tag{N4.71}$$

Let $f(\theta)$ be a particular complex function on the circle. Let $R(g_\phi)$ be defined on the function as in eqn N4.124.

(e) *Show that the projection operator in eqn N4.127, using the m representation of part (c), takes $f(\theta)$ into a coefficient times the basis vector for that representation. How is the coefficient related to the Fourier series coefficient*[71] *$\tilde{f}_m$ for f?*

[69] For $SO(2)$, this is just an integral over ϕ. More generally, and in particular for $SO(3)$, you have an extra factor in the integral (the *Haar measure*).

[70] Alex didn't bother with the factor $(f^{(\alpha)}/o(G))$, since he just wanted a vector in the subspace. We want to make the sum over representations α equal to the original function. Alex also, I think, missed the complex conjugate (but all his characters were real).

[71] There are many different conventions for Fourier series. Clearly state which one you are using.

If we sum the projections of a vector into all the invariant subspaces, we should get the vector back again. The inverse Fourier series takes the Fourier coefficients and resums them back into the original function.

(f) *How is the inverse Fourier series related to the sum of the projections of $f(\theta)$ over all the irreducible representations of $SO(2)$?*

N4.45 **Solving Schrödinger: WKB, instantons, and the double well.** (Computation) ③

We study the problem of quantum tunneling in a symmetric double well potential. If the barrier between the two wells is large, the ground state and first excited state are well approximated as symmetric and antisymmetric superpositions of the ground states in the two separated wells. Indeed, the low-energy physics of the double-well system can be approximated by a two-level system (TLS), another example of a Qbit. In the symmetric case, in a basis where $\left(\begin{smallmatrix}1\\0\end{smallmatrix}\right)$ is the state localized in the left well and and $\left(\begin{smallmatrix}0\\1\end{smallmatrix}\right)$ is in the right well, the Hamiltonian is

$$H_{\rm TLS} = \begin{pmatrix} 0 & -\Delta \\ -\Delta & 0 \end{pmatrix}, \tag{N4.72}$$

where Δ is called the tunneling matrix element.

(a) *What is the energy splitting for $H_{\rm TLS}$? Calculate the time evolution operator $U(t) = \exp(-{\rm i}H_{\rm TLS}t/\hbar)$. If we start in the left well at $t = 0$, what is the probability being in the left well after time t?*

It is traditional and useful to study the double well given by the quartic polynomial

$$V(y) = V_0 \left(\frac{y}{Q_0} - 1\right)^2 \left(\frac{y}{Q_0} + 1\right)^2. \tag{N4.73}$$

In our notation, this potential has minima at $\pm Q_0$ separated by a barrier of height V_0; a particle of mass m has small oscillation frequency ω near the minima in each of the two wells.

(b) *Calculate the barrier height V_0 algebraically in terms of m, Q_0, and ω, that sets the small oscillation frequency near $y = Q_0$ equal to ω.*

(c) *Calculate the classical instanton action (or, equivalently, the WKB exponent) $S_0 = \int_{-Q_0}^{Q_0} \sqrt{2mV(y)}{\rm d}y$ for our potential. Let $Q_0 = na_0$ with $a_0 = \sqrt{\hbar/2m\omega}$ the width of the ground state in the harmonic approximation. What is $S_0/\hbar$ in terms of n? What numerical value does it have for $n = 5$? How big is $\exp(-S_0/\hbar)$, the Euclidean action's suppression to the contribution of the instanton path?* (Hint: $S_0/\hbar$ is dimensionless. As suggested by the latter part of the exercise, for our potential it should depend only on the dimensionless number n, given that V_0 is given as in part (b).)

We shall assume our particle has the mass of a hydrogen atom, the small oscillation frequency in the left and right wells is 1THz, and $Q_0 = 5a_0$. Generate a grid of length

$L = 20a_0$ with $N_p = 200$ points. Let the initial wavefunction $\psi[0] = \psi_0(x + Q_0)$ be the harmonic ground state in the left well. As in the coherent-state exercise, generate the appropriate arrays $U_{\rm pot}({\rm d}t/2)$ and $U_{\rm kin}({\rm d}t)$ to evolve ψ using your BCH formula, with ${\rm d}t = P/20$ where P is the period of the oscillator.

(d) *Solve for the time evolution over 1,000 periods. Animate the probability density* $|\psi(x,t)|^2$ *versus time. Does it travel between the two wells periodically, as should be predicted by your answer to part (a)?*

(e) *Plot the probability* $\int_{-\infty}^{0} |\psi(x,t)|^2$ *that the particle is found in the left well.* (The high-frequency wiggles are due to the small components of higher eigenstates of our initial state in the left well.) *By eye, estimate the period of oscillation, and also* Δ*, within 10%.*[72]

(f) *Use a nonlinear least-squares method to fit your prediction from part (a) to the probability you calculated in part (e). Include your fit in the plot for part (e). What is* $\Delta_{\rm Fit}$*?* (If your answer from part (e) is off by much more than 30%, see if you can track down the problem.) *What is the ratio of your estimated energy splitting (using the formula from part (a)), to the energy splitting between states in one of the two wells (in the harmonic approximation)?* Quantum tunneling is one of the most important sources of low-energy, long-time behavior in physics.

The instanton formula for the tunnel splitting is closely related to the WKB formula.[73] Both are of the form $\Delta = \hbar\omega_0 \exp(-S_0/\hbar)$. In the WKB formula, the prefactor $\hbar\omega_0$ is given by a matching calculation; in the instanton method it is given by a path integral incorporating small oscillations about the instanton path.[74] Gildener and Patrascioiu (*Phys. Rev. D* **16**, *423, 1977*, referred to by Coleman) did the calculation explicitly for the quartic well using instanton methods, and got

$$\Delta \sim \hbar\omega \sqrt{\frac{6S_0}{\pi\hbar}} \exp(-S_0/\hbar) \tag{N4.74}$$

so in our notation $\omega_0/\omega = \sqrt{6S_0/\pi\hbar}$.

(g) *Calculate your numerical estimate of* ω_0/ω*, and compare with Gildener and Patrascioiu's estimate.*

There is an interesting story about communication between high-energy and condensed-matter physics. Sidney Coleman (high-energy Harvard theorist) and Jim Langer (condensed matter theorist on sabbatical at Harvard) were both working on barrier crossing at the same time. Coleman was stuck for months, because his instanton calculation was a factor of two higher than the WKB estimate (the 'double well done doubly well'

[72] Nonlinear least-squares routines typically get lost if you don't start near the best fit.

[73] The two turning points for the classical path in the harmonic well are not far enough apart for the traditional WKB formula to be accurate; a suitable generalization accounting for that proximity does agree with the instanton formula.

[74] Coleman tells us the answer from the instanton calculation is given in terms of *Wronskians*.

appendix to his Les Houches lectures, calculating quantum fluctuations causing transitions through barriers). Unbeknownst to him, Jim Langer in a nearby office had solved the problem in the statistical mechanics context (critical droplet theory, calculating thermal fluctuations causing transitions over barriers). Eventually, Coleman realized as Langer had earlier that only half the fluctuations that reach the barrier actually cross it.

N4.46 **GHZ and σ matrices.** (Weirdness) ③

The GHZ state is the state of three spins held in spatial isolation from one another by Alice, Bob, and Carol: $|GHZ\rangle = (1/\sqrt{2})(|\uparrow_z,\uparrow_z,\uparrow_z\rangle - |\downarrow_z,\downarrow_z,\downarrow_z\rangle)$, where the first label in the ket represents the state of the A-spin, the second the B-spin, etc. In class, we showed that $XXX = X_AX_BX_C$ acting on $|GHZ\rangle$ always gave -1, but gave $+1$ for $XYY = X_AY_BY_C$ or YXY or YYX, each time they are measured. Multiplying the results of measurements of the latter three, $(X_AY_BY_C)(Y_AX_BY_C)(Y_AY_BX_C)$ gave plus one, and not minus one (as one would presume if Bell was right and hidden variables stored "true" states of X and Y).

Instead of considering products of observations, consider the operator product

$$\Omega = (X_AY_BY_C)(Y_AX_BY_C)(Y_AY_BX_C) \tag{N4.75}$$

In this exercise you shall show that Ω is equal to $-X_AX_BX_C$ as an operator (independent of the GHZ state). There are two steps in the proof.

First, since Alice, Bob, and Carol are isolated from one another, Alice's measurements do not affect Bob's, etc. This means that any operator (X, Y, or Z) at A should commute with any operator at B or C, and that operators at B and C should commute. Also, since all the operators have eigenvalues ±1, their squares are the identity (e.g., $Y_B^2 = \mathbb{1}$).

(a) *Using only this fact, rewrite Ω of eqn N4.131 as the product of five operators.*

The operator Z at a site has eigenvalue $+1$ if the spin is quantized up along the z axis, and -1 if the spin is down. Hence Z acting on a spin at that site[75] gives $Z = \left(\begin{smallmatrix}1&0\\0&-1\end{smallmatrix}\right)$. Similarly, in the z basis $X = \left(\begin{smallmatrix}0&1\\1&0\end{smallmatrix}\right)$ and $Y = \left(\begin{smallmatrix}0&-\mathrm{i}\\\mathrm{i}&0\end{smallmatrix}\right)$.

(b) *Use this and your answer to part (a) to show that $\Omega = -X_AX_BX_C$.*

The GHZ state is unusual in that it is an eigenstate of XXX, XYY, YXY, and YYX, so that measurements of $|GHZ\rangle$ by these operators leave it unchanged (up to a sign). This allowed us to view the conundrum that separate measurements of the three operators were incompatible with Bell. But we have shown that any state ψ will share the property of the GHZ state that $\Omega|\psi\rangle = -XXX|\psi\rangle$.

[75] Thus the spin operator $\sigma_z = (\hbar/2)Z$. Also, X, Y, or Z acting on a spin at another site is the identity.

N4.47 **No cloning theorem.** (Quantum Computing) ③

Can one make an exact copy of an arbitrary, unknown quantum state $|\psi\rangle$, without changing the original state? Sadly, this is impossible.[76]

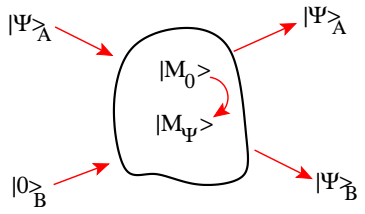


Fig. N4.16 Hypothetical cloning machine. Our machine takes an unentangled state $|\psi\rangle_A$, and copies it using a blank state $|0\rangle_B$. In the process, the machine changes state in some fashion that can depend on ψ.

Pretend it was possible. In particular, let us start with three subsystems (Fig. N4.17): A (to be cloned, with coordinates $\mathbf{x}$), B (to store the clone, coordinates $\mathbf{y}$), and M (the quantum cloning machine, with coordinates $\mathbf{Z}$). Suppose subsystem A is in state $|\psi\rangle_A = \psi(\mathbf{x})$; system B starts in a blank system $|0\rangle_B = \beta(\mathbf{y})$, and our cloning machine starts in a resting state $|M_0\rangle_M = \mathcal{M}_\prime(\mathbf{Z})$. Before and after the machine operates, A, B, and M are independent, uncoupled systems. The machine should act as a unitary operator U_c, taking $\Psi(\mathbf{x},\mathbf{y},\mathbf{Z}) = \psi(\mathbf{x})\beta(\mathbf{y})\mathcal{M}_0(\mathbf{Z})$ into

$$U_c|\Psi\rangle = \psi(\mathbf{x})\psi(\mathbf{y})\mathcal{M}_\psi(\mathbf{Z}) \tag{N4.76}$$

where $\mathcal{M}_\psi(\mathbf{Z})$ is the final state of the machine after it processes the state Ψ. Our machine need only copy states $\psi(\mathbf{x})$ which are unentangled with the rest of the universe, and thus we expect the final state $\psi(\mathbf{y})$ should be also unentangled by our hypothetical machine.

(a) *Viewing the state in eqn N4.132 as a system B in an environment made up of A and M, is the quantum state entangled, or not? If not entangled, write it as a product; if entangled, give its Schmidt decomposition.* (Note: In writing the Schmidt decomposition $\sum_k \sigma_k u_k(\mathbf{y})v_k(\mathbf{x},\mathbf{Z})$, you must show $\sigma_k > 0$, $\sum_k \sigma_k^2 = 1$, the $|u_k\rangle_B$ are orthonormal, and that the $v_k(\mathbf{y},\mathbf{Z})$ are orthonormal. Hint: You can probably do this part by inspection.) *If you do a Schmidt decomposition, explicitly check the orthogonality of the AM states.*

Suppose our initial cloning machine acts on two *orthonormal* input states $\psi(\mathbf{x})$ and $\phi(\mathbf{x})$, so in addition to eqn N4.132, we have

$$U_c|\Phi\rangle = \phi(\mathbf{x})\phi(\mathbf{y})\mathcal{M}_\phi(\mathbf{Z}) \tag{N4.77}$$

[76] As with many impossible things, this would be very useful! Copying Qbits being used for secure quantum communication would allow one to listen in. Copying Mr. Spock while beaming up would have allowed exciting new episodes in the 1970's ...

(b) *Let U_c act on the superposition $(1/\sqrt{2})(|\Psi\rangle + |\Phi\rangle)$; expand it using linearity and eqns N4.132 and N4.133. Is the final state entangled, or not? Again, write it as a product, or give its Schmidt decomposition.* (Remember $\psi(\mathbf{x})$ and $\phi(\mathbf{x})$ are orthogonal. You may assume $|M_\phi\rangle$ and $|M_\psi\rangle$ are normalized after the unitary transformation, but they *need not be orthogonal*.) *If you do a Schmidt decomposition, explicitly check the orthogonality of the AM states.*

Thus our cloning machine must necessarily be unsuccessful in generating an unentangled clone of ψ, ϕ, and their superposition. No general-purpose cloning operator is possible over a Hilbert space with more than one dimension.

N4.48 **Aharonov-Bohm wire: straightforward approach.** (Parallel Transport) ③

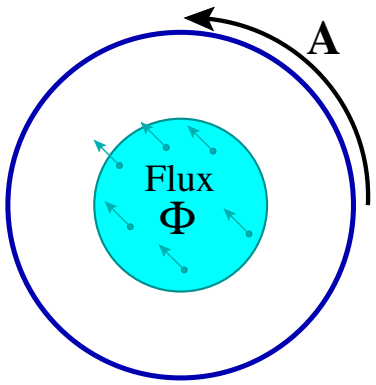


Fig. N4.17 Aharonov-Bohm wire loop. A circular wire of radius R encloses a magnetic field of total flux Φ.

An electron is confined to a one-dimensional wire loop of radius R (Fig. N4.18). A uniform magnetic field of strength B perpendicular to the loop has a net magnetic flux $\Phi = \pi R^2 B$. We pick a gauge in which the electromagnetic potential $\mathbf{A} = A\hat{\theta} = \Phi\hat{\theta}/(2\pi R)$ at the radius of the wire, where $\hat{\theta}$ is the unit vector tangent to the wire. Let s measure distance around the wire. Schrödinger's time-independent equation then becomes

$$\begin{aligned} E\psi(s) &= -\frac{\hbar^2}{2m}\left(\frac{\partial}{\partial s} - \frac{\mathrm{i}eA}{\hbar c}\right)^2 \psi(s) \\ &= -\hbar^2/(2m)\left(\frac{\partial}{\partial s} - \frac{\mathrm{i}e\Phi}{2\pi R\hbar c}\right)^2 \psi(s) \end{aligned}$$

where $e < 0$ is the charge on the electron. The energy eigenstates are of the form $\psi_n(s) \propto \exp(\mathrm{i}k_n s)$. In this problem, we are *not* changing to a singular gauge.

(a) *What values of k_n are allowed?* (Hints: The wavefunction must be periodic in arclength s with period $2\pi R$. The answer does not depend on Φ.)

(b) *Show that ψ_n is an eigenstate of $\big(\partial/\partial s - \mathrm{i}eA/\hbar c\big)$. Using that, find the eigenenergies E_n. Are the eigenenergies of the system the same for $\Phi = 0$ and for $\Phi = \Phi_0 = hc/e = 2\pi\hbar c/e$, as we showed in lecture?* (Hint: The label n of the energies may change, but is the *spectrum* of eigenenergies unchanged?)

The flux through the loop is set to exactly half a flux quantum, $\Phi = \Phi_0/2 = hc/2e = \pi\hbar c/e$.

(c) *What eigenstate or eigenstates have lowest energy? Are all the eigenstates doubly degenerate?* (Show your work.)

N4.49 **Number, Phase, and Josephson.** (Quantum commutation) ③

In this exercise, we use the Josephson effect in superconductors to explore how the phase of the superconducting wavefunction is related to the indeterminacy in the number of particles.

A Josephson junction is a thin insulating layer or a weak link joining two regions of superconducting material or superfluid, see Fig. N4.19(a). Let the number of bosons (charge $2e$ Cooper pairs) in the two pieces be N_1 and N_2, and the superconducting phases be ϕ_1 and ϕ_2. There is a term in the energy $U(\phi_2 - \phi_1)$ that grows with the phase difference between the two sides. Naturally, this energy cost must be periodic in $\delta\phi = \phi_2 - \phi_1$. For weak links, this energy cost is very nearly some constant U_0 times $1 - \cos(\delta\phi)$,

$$U(\delta\phi) = U_0(1 - \cos(\delta\phi)), \tag{N4.78}$$

as plotted in Fig. N4.19(b).

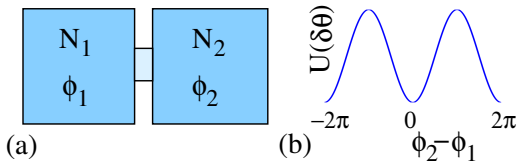


Fig. N4.18 Josephson junction schematic. (a) A Josephson junction is made from a weak link (light small box) connecting two pieces of superconductor or superfluid. (b) The energy of the superconducting weak link as a function of the phase difference across it.

In class, we remarked that number and phase for superconductors are conjugate variables, $[\phi, N] = \mathrm{i}$, similar to how position and momentum are conjugate $[x, p] = \mathrm{i}\hbar$. We said this could be used to explain the behavior of Josephson junctions. Here we will work backward. We shall start with the experimental behavior of the junction (as predicted by Josephson from a more microscopic calculation), and deduce the commutation relation.

First, a Josephson junction experimentally has a supercurrent $I/(2e) = \mathrm{d}N_2/\mathrm{d}t = -\mathrm{d}N_1/\mathrm{d}t$ which depends on $\delta\phi$:

$$I = I_c \sin(\delta\phi). \tag{N4.79}$$

This is the *DC Josephson effect*; electron pairs can tunnel across the insulating gap, but only up to a *critical current* I_c. Let $\delta N = (N_2 - N_1)/2$, the net number of $2e$ bosons that have crossed the junction. Note from eqn N4.137 that $\mathrm{d}(\delta N)/\mathrm{d}t = I\,(\delta\phi)/(2e) \propto -\mathrm{d}U/\mathrm{d}\delta\phi$, just as the rate of change of momentum $\dot{p}$ is given by the force (the negative derivative of the potential energy). We can use this to deduce the commutation of $\delta\phi$ and δN up to an overall constant x.

(a) *Use the general Heisenberg relation* $\mathrm{d}O/\mathrm{d}t = (1/\mathrm{i}\hbar)[O, \mathcal{H}]$*, and assume that* $\mathcal{H}$ *is some (so far) unknown function of* δN *(the "kinetic energy") plus* $U(\delta\phi)$ *(the "potential energy"). Show that* $\delta N = (x/\mathrm{i})\partial/\partial(\delta\phi)$ *gives a DC Josephson relation for any value of* x*. Use this to write* $[\delta\phi, \delta N]$ *in terms of* x*. Write* x *in terms of* I_c *and* U_0*.*

(b) *Using your commutator* $[\delta\phi, \delta N]$ *from part (a), deduce that* $\delta\phi = (y/\mathrm{i})\partial/\partial\delta N$ *by showing that* $[\delta\phi, \delta N^m] = (y/\mathrm{i})m\delta N^{m-1}$ *for all positive integers* m*.* (Hint: work by induction.) *Write* y *in terms of* x*.*

Second, with a fixed voltage V across a Josephson junction, we measure an AC current $I(t) = I_c \sin((2e/\hbar)Vt)$. This is the *AC Josephson effect*, and has been used in the past to define the volt. Using the DC Josephson relation (eqn N4.138), this implies $\mathrm{d}(\delta\phi)/\mathrm{d}t = 2eV/\hbar$. Note that $2eV$ is the energy to move one Cooper pair from the left to the right. For most junctions, a change of $\delta N = 1$ Cooper pair can be viewed as an infinitesimal change in δN. We can use this to deduce the constant x.

(c) *Use* $\mathrm{d}(\delta\phi)/\mathrm{d}t = (1/\mathrm{i}\hbar)[\delta\phi, \mathcal{H}]$ *and the form of* $\delta\phi$ *from part (b), deduce* y*,* x*, and the relation between* U_0 *and* I_c*. In particular, verify that the commutator* $[\delta\phi, \delta N] = \mathrm{i}$*, as suggested in class.* (Hint: How is the voltage related to the energy, and thus the Hamiltonian?)

While the superconducting wavefunction is a macroscopic manifestation of quantum mechanics, it usually is not subject to large quantum fluctuations. Instead, it acts like a magnetization or a crystalline orientation—a macroscopic property of a material, that has a well-defined value. When will our Hamiltonian for the Josephson junction, which describes the *classical* evolution of the current and phase, have important *quantum* fluctuations?

Just as $[x, p] = \mathrm{i}\hbar$ implies $\sigma_x\sigma_p \geq \hbar/2$, $[\delta\phi, \delta N] = \mathrm{i}$ implies $\sigma_{\delta\phi}, \sigma_{\delta N} \geq \frac{1}{2}$. Under normal circumstances, the number of Cooper pairs δN exchanged between the superconductors is enormous. In many ways, the superconducting condensate acts like a classical field. But for a tiny junction, the total number of Cooper pairs can be small—leading to a noticeable quantum fluctuation in the superconducting phase. A group at IBM tested ideas of Tony Leggett by building such a Josephson junction (Washburn et al., "Effects of dissipation and temperature on macroscopic quantum tunneling", PRL 54, 25 (1985)).

Let the thin insulating layer in Fig. N4.19(a) be of thickness $t = 10\mathrm{nm}$, and width and height L. Treat the Josephson junction as a capacitor, with a charge $Q = 2e\delta N$ transferred from one side to another. This gives us a model for the "kinetic energy" in $\mathcal{H}$. For many purposes, the complicated physics of a Josephson junction can be

summarized by the critical current I_c, the capacitance C, and a resistance (not discussed here).

(d) *Including the capacitor energy as the kinetic energy, give the total Hamiltonian $\mathcal{H}$ for our junction, in terms of the junction capacitance C and the critical current I_c. Take the harmonic approximation $U(\delta\phi) \approx \frac{1}{2}U_0\delta\phi^2$, and derive the capacitance necessary to make $\langle\delta\phi^2\rangle = 1$ in the ground state, in terms of the critical current I_c.* Washburn et al. used a junction with a capacitance of $C = 0.02 \pm 0.005$pF, and a critical current of $I_c = 55\mu$A. *How big is $\delta\phi$ for their junction?*

Many experiments and theoretical calculations have been done on these quantum fluctuations. A superconducting loop with a tiny Josephson junction weak link, when put in a magnetic field, will allow flux to quantum tunnel out of the junction. This quantum tunneling can be viewed as quantum fluctuations allowing the phase $\delta\phi$ to tunnel from one minimum of $U(\delta\phi)$ to the next.

N4.50 **Localization.**[77] (Quantum, Condensed Matter) ③

In Section 7.4, we discussed how non-interacting electrons provide a useful model for metals, even though the electron-electron interactions are strong. The Fermi liquid of quasiparticles is a kind of adiabatic continuation of the noninteracting electron system, connected by perturbation theory. Here we shall study how a one-dimensional noninteracting metal responds to disorder. We shall discuss how metals with weak disorder are understood by perturbing around the clean state. We shall discover that strong disorder leads to an insulating system whose eigenstates are not extended, but *localized*. We will describe these localized states explicitly by perturbing about a state of isolated atomic states.

Consider a one–dimensional chain of atoms n, each with one noninteracting electron state $|n\rangle$ of energy U_n, that can be occupied by either zero or one spinless electrons. Electrons can hop between atoms with matrix element t, leading to the Hamiltonian

$$\begin{aligned}\mathcal{H} &= \sum_{n=1}^{N} U_n |n\rangle\langle n| \\ &\quad - t\sum_{n=1}^{N-1} \left(|n\rangle\langle n+1| + |n+1\rangle\langle n|\right) \\ &= \begin{pmatrix} U_1 & -t & 0 & \dots & 0 & 0 \\ -t & U_2 & -t & \dots & 0 & 0 \\ 0 & -t & U_3 & \dots & 0 & 0 \\ 0 & 0 & -t & \dots & 0 & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots & -t \\ 0 & 0 & 0 & \dots & -t & U_N. \end{pmatrix}\end{aligned} \tag{N4.80}$$

[77] Hints for the computations can be found at the book web site [31].

We shall take the random energies U_n as uniformly distributed between $-W$ and W.

Without disorder ($W = 0$), this is a textbook model used to describe energy bands in crystals. Three dimensional analogs of this 'tight-binding' model are quite realistic models of Fermi surfaces and energy bands in real materials.[78]

(a) *Write a function that builds the Hamiltonian matrix of eqn* **??** *with size N, bandwidth $2W$, and hopping matrix element t. Studying zero disorder $W = 0$, find the eigenvectors for $t = 1$, and $N = 100$, sorted by their eigenvalues. Plot the eigenvectors for the four lowest energies. Check numerically that these four are sinusoidal with wavevectors $k_\alpha = \pi\alpha/(N+1)$ appropriate for a box of size N with hard-wall boundary conditions half a grid spacing to either side. Check that their four eigenvalues are the corresponding $E_{k_\alpha} = -2t\cos(k_\alpha)$*

Imagine a 1D metal at zero temperatures with electrons filling the states up to a Fermi surface, here just two points at some $\pm k_{\text{Fermi}}$. Consider a packet of electrons made up of eigenstates near k_{Fermi} traveling to the right. The wavepacket[79] will travel, as usual, at the group velocity $\mathrm{d}E_k/\mathrm{d}k|_{k_{\text{Fermi}}}$, without dissipation.

Now let us explore what happens when we add a weak disorder.

(b) *Build a Hamiltonian with weak disorder $W = W_{\text{weak}} = 0.04$, $t = 1$, and $N = 100$. Plot the lowest four eigenvectors. Are the eigenstates still extensive (reaching from one side of the box to the other)?*

At this point, we could use perturbation theory to calculate the disordered eigenstates and energy levels. We could then create wavepackets and see how they evolve. In three dimensions, the scattering off of the disorder changes the electron transport qualitatively. Instead of wavepackets moving forever in one direction (ballistic transport, infinite conductivity), one gets diffusive motion of the electron probability through space (disorder providing an elastic scattering length, and a finite conductivity). In three dimensions, this is a good model for metals with impurities or dopants, illustrating how one can understand complex behavior by perturbing around solvable special cases.

Instead, let us examine what happens at large disorder W, or equivalently, small hopping t. (All of our eigenvectors depend only on W/t, and we will perturb in t to study the localized states.)

(c) *Set $t = t_{\text{Weak}} = 0.1$, $W = 1$, and $N = 100$, plotting the ten eigenvectors with lowest energies. Also do a log-linear plot of the probability density (absolute square of*

[78] We shall see that even a small disorder changes the metallic behavior of one-dimensional electrons in a qualitative and interesting way. Indeed, one dimensional electrons are unstable in many interesting ways. Adding interactions between electrons, they become *Luttinger liquids*, with emergent scale invariance. Adding interactions with lattice vibrations, they can become topological insulators, with solitons and fractional charges.

[79] Wavepackets are used to connect waves to particle-like motion. In a non-disordered system, one superimposes states with similar momenta to make a spatially localized wavefunction, which then moves with the group velocity of the wave. We discuss wavepackets to motivate the effects of disorder, but no knowledge about them is required to do this exercise.

the wavefunctions) for these eigenvectors. Do the eigenstates still look as if the will be extensive (stretching from one end of a macroscopic wire to the other)? What solvable special limit of $\mathcal{H}$ should we use to capture this new behavior?

Here we find the eigenstates appear *localized* – fixed in space near individual 'atoms'. The probabilities in these states fall exponentially with distance from their centers. A wavepacket formed from localized states like these cannot transport current: for large disorder, our model describes an insulator.

Just as one can use perturbation theory to describe dirty metals in three dimensions from models like ours, we can use perturbation theory to calculate and understand these localized states. You should remember the use of second-order perturbation theory to describe the energies of a Hamiltonian $\mathcal{H} = \mathcal{H}_0 + V$ for small V. You may not remember that the first step was to use first-order perturbation theory to determine the eigenvectors. If $|\Psi_i^{(0)}\rangle$ has unperturbed eigenvalue $E_i^{(0)}$, then to first order

$$|\Psi_i^{(1)}\rangle = |\Psi_i^{(0)}\rangle + \sum_{i \neq j} \frac{\langle \Psi_j^{(0)} | V | \Psi_i^{(0)} \rangle}{E_i^{(0)} - E_j^{(0)}} |\Psi_j^{(0)}\rangle. \tag{N4.81}$$

If the hopping is small compared to the disorder, let us perturb in t.

(d) *What are the eigenenergies for $\mathcal{H}$ in eqn* **??** *with $t = 0$? Argue that, to first order in t, the new eigenstates will be confined to three adjacent sites.*

(e) *Write a function, given $\mathcal{H}$, i, t, and N, that gives the perturbed eigenstate to lowest order in t that is centered at site i. Find the site of the ground state with largest probability. Plot the true ground state and your first-order approximation to it.* (Hint: You may be unlucky, and happen to have a neighbor site with a near degenerate energy. Just create a new Hamiltonian and try again.)

What controls whether our model is a metal or an insulator? For a given W/t, are all the states either extended or all localized? Or could there be some mixture?

We can examine this by defining a rough measure of how spread out the wavefunction is, called the *participation ratio*:

$$P(\psi) = \frac{\left(\sum_n |\psi(n)|^2\right)^2}{\sum_n |\psi(n)|^4} = \frac{1}{\sum_n |\psi(n)|^4}. \tag{N4.82}$$

(f) *Show that a state whose probability is spread uniformly among M sites has $P = M$. At zero disorder, what is the participation ratio for the lowest energy state? For the long-wavelength next few states? What is the ratio for a localized state that decays exponentially, $\psi(j) \propto \exp(-|i-j|/\lambda)$, in an infinite chain, with λ much larger than one?*

If the participation ratio $P \ll N$ we can reasonably expect that the eigenstate is localized.

(g) *Calculate the participation ratio for all the eigenstates for intermediate disorder* $W_{\rm inter} = 0.5$, $t = 1$, $N = 100$, *and plot them against the energy. Is there a systematic variation? Plot the wavefunction for an energy in the middle of the energy band (eigenvalue* E *near zero), and one at the top and bottom of the band. Which are less localized – the states near the edges of the band, or the states in the center?*

In experiments, one finds a region of localized states at the edges of a band, and extended states in the middle of the band.[80] Between these is a *mobility edge*, where a metal-insulator transition occurs as more electrons are added.

Finally, can we find a mobility edge for our model? One thing to check is if the wavefunctions might have decay lengths larger than our system (so they just look extended at $N = 100$).

(h) *Find the eigenvectors and eigenvalues for the same parameters as in part (g),* $W = W_{inter} = 0.5$ *and* $t = 1$, *except for a much larger system (*$N = 2000$ *if it is feasible on your system). Plot the participation ratio verses energy, and plot eigenstate in the center of the band and at the two edges. Do the states in the middle of the band now appear localized? Are their participation ratios larger than 100 – the system size in part (g)? Does it make sense that they looked extended in the smaller system, but clearly in an infinite system are localized?*

As it happens, disordered electrons in one dimension are always localized, even for tiny disorder. The spinless, noninteracting electrons we study here are also always localized in two dimensions. In two dimensions, they can become extended when interactions, spin orbit scattering, or strong magnetic fields are added. In particular, 2D electrons in a strong magnetic field exhibit the quantum Hall effect (with extended states around the edges). Even more interesting, interacting electrons in a strong magnetic field exhibit the fractional quantum Hall effect – our first example of an experimental system with fractional charges and fractional statistics.

[80] Each time we add one order to perturbation theory, we get a wavefunction extending outward by one atom. It appears from the first two terms that each order multiplies the terminal amplitude by a factor $|\psi(i+n)/\psi(i+(n-1))| = |t/(U_i - U_{i+n})|$. Roughly speaking, if U_i is at the edge of the band, this factor is twice as small as if U_i is in the center, so there is less localization at the center. Notice, though, that this is useful only when $t > |U_i - U_{i+n}|$. Rare, nearly degenerate states can mix strongly, even at long distances, making the arguments subtle.

N4.51 **Eight-Fold Way.** (Particles, Group reps) ③

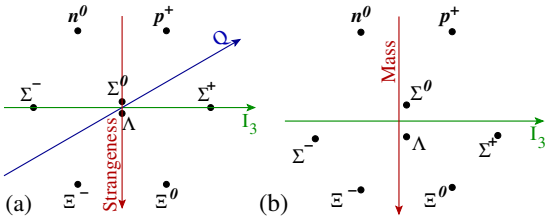


Fig. N4.19 Baryon octet. (a) Shown are a family of eight baryons, headed by the neutron and proton, organized by their electrical charge Q, their *strangeness* S, and their isospin I_3.[81] Thus for example the neutron has isospin -½, and Σ^- has isospin -1. These baryons transform under an irreducible representation **8** of the group $SU(3)$. The charge Q is a linear combination of I_3 and strangeness. (b) The same baryons, plotting mass versus isospin I_3.

Years before the realization that the strong interaction has an exact "color" gauge symmetry group $SU(3)$, it was noticed that it had an approximate "flavor" SU(3) symmetry. This was noticed because the baryons and mesons naturally organized themselves into multiplets which mimicked representations of $SU(3)$ (Fig. N4.20(a)). Figure N4.20(b) shows the masses of one family of baryons, indicating the original experimental evidence for the model.

In this exercise,[82] we will explore the geometry shown in Fig. N4.20(a), exploring the rows, the angles, and their relation to the symmetry groups.

The story began with the neutron and the proton. They are amazingly similar, apart from their charge. (For example, the proton has a mass 0.9986 times that of the neutron—only about 2½ electron masses different. A little lighter, and the neutron could not decay.) Werner Heisenberg made an analogy with spin-½ particles—the proton has isospin $I_3 = +½$ and the neutron isospin $I_3 = -½$.

In a world with only the strong interaction, and where the up and down quark had the same mass, the strong interaction would have an SU(2) total isospin symmetry as well as an SU(2) rotational symmetry—thus an overall symmetry $SU(2) \times SU(2)$. The nucleon ground state, with spin ½ and isospin ½, thus would have four degenerate states ($n^\uparrow$, $n^\downarrow$, $p^\uparrow$, and $p^\downarrow$).[83]

(a) *Draw a figure corresponding to Fig. N4.20(a) for these four nucleon states, plotting I_3 horizontally but plotting spin as the vertical axis.*

Consider now the deuteron—the bound state of a proton and a neutron—in a world with perfect isospin symmetry. In such a world, we can think of the bound states of

[82] Reading: I found K. Schulten, "Notes on Quantum Mechanics", ch. 12 to be quite useful in developing this exercise (http://www.ks.uiuc.edu/Services/Class/PHYS480/qm_PDF/chp12.pdf).

[83] It is traditional to ignore the rotational $SU(2)$ symmetry when discussing gauge field theories, since it is an *external* symmetry inherited from the symmetry of space-time, rather than an *internal* symmetry of the particle fields. We break with tradition in part (a).

two identical *nucleons*, which can have two (ordinary) rotational spin states $\pm\frac{1}{2}$ and two isospin states $I_3 = \pm\frac{1}{2}$ (proton or neutron). We may assume that the ground state spatial wavefunction for the deuteron is symmetric under interchange of the two nucleons, with orbital angular momentum zero.[84] Two spin $\frac{1}{2}$ nucleons may thus form rotational spin either zero or one ($\frac{1}{2} \otimes \frac{1}{2} = 1 \oplus 0$). Their total isospins combine in the same way. That is, isospin has the same symmetry group structure, so we know that the eigenstates of our isospin-symmetric deuteron will also be eigenstates of isospin with total $I = 1$ or $I = 0$.

(b) *First, write the three total isospin triplet states in terms of $|p\rangle$ and $|n\rangle$. Are there bound states with $I_3 = \pm 1$, to your knowledge?* (See section 12.2 of Schulten if you get stuck.) *Second, how does the total isospin singlet state transform under interchange of the two nucleons? Treating the two nucleons as identical fermions (differing by their isospin I_3), use the antisymmetry of the wavefunction to argue that the rotational spin of the deuteron must be equal to one.*

Isospin is indeed so close to a symmetry of the strong interaction[85] that one can use isospin Clebsch-Gordon coefficients to predict branching ratios in particle collisions.

We now turn to a much more strongly broken symmetry. The strange quark is notably heavier than the up and down quarks. But if we extend the SU(2) total isospin symmetry to SU(3), incorporating the strange quark, we can rationalize the elementary particles into families organized by their isospin, strangeness, and charge. Figure N4.20(a) shows eight particles which transform under the irreducible representation of SU(3) called **8**.

You remember that angular momentum has J_x, J_y, and J_z, and that the *ladder* operators $J_\pm = J_x \pm \mathrm{i}J_y$ were useful in deducing facts about the different spin states. In particular, the fact that $[J_z, J_\pm] = \pm\hbar J_\pm]$ told us that, given an eigenstate $|j_z\rangle$ of J_z, we can find other eigenstates $|j_z \pm 1\rangle = J_\pm|j_z\rangle$ (raising and lowering the rung of the j_z ladder).[86] Other calculations told us how the ladders ended, and gave us the rules that the states with total angular momentum $J^2 = j(j+1)$ came as multiplets of degeneracy $2j+1$, with $j_z = (-j, \ldots, j)$. We will focus here just on generalizing these ladder operators to SU(3).

The three quarks

$$|u\rangle = \begin{pmatrix}1\\0\\0\end{pmatrix}, \quad |d\rangle = \begin{pmatrix}0\\1\\0\end{pmatrix}, \quad |s\rangle = \begin{pmatrix}0\\0\\1\end{pmatrix} \tag{N4.83}$$

form a three-dimensional representation of $SU(3)$ by multiplying by the corresponding matrix. (Here the language gets confusing, since $SU(3)$ denotes both the abstract

[84] At high energies, wavefunctions become less quantitatively useful. The internal state of a nucleus is actually a mixture of states with different numbers of particles. Also, for the deuteron, spin-orbit interactions lead to a small mixing of states with orbital angular momentum $L = 2$. We ignore such subtle effects.

[85] It is of course not a symmetry of the electromagnetic interaction: the proton and neutron have different charges! It is also not a symmetry of the weak interaction: neutrons weakly decay into protons emitting isospin-free electrons and neutrinos.

[86] As a quick review, this works because $J_z|j_z \pm 1\rangle = J_z J_\pm|j_z\rangle = (J_\pm J_z + [J_z, J_\pm])|j_z\rangle = J_\pm \hbar j_z|j_z\rangle \pm \hbar J_\pm|j_z\rangle = \hbar(j_z \pm 1)J_\pm|j_z\rangle = \hbar(j_z \pm 1)|j_z \pm 1\rangle$.

group and its traditional representation as a set of 3×3 complex unitary matrices that all have determinant one.) We call this representation **3**, just as we call the octet representation **8**.

The total isospin symmetry $SU(2)$ is the subgroup of $SU(3)$ which just mixes the first two components u and d. In the **3** representation, it has three infinitesimal generators which can be turned into the isospin operator $I_3 = \begin{pmatrix} ½ & 0 & 0 \\ 0 & -½ & 0 \\ 0 & 0 & 0 \end{pmatrix}$ and the two ladder operators $I_+ = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ and $I_- = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$. Just as for spin, these ladder operators exist in all representations. For example, in the octet representation **8** of Fig. N4.20(a), $I_\pm$ shifts one right and left, so $I_+|\Xi^-\rangle \propto |\Xi^0\rangle$ and $I_-|p^+\rangle \propto |n^0\rangle$. Similarly, we can use the generators of the $SU(2)$ subgroup that mixes the up and strange quarks to form ladder operators $V_+ = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ and $V_- = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}$, shifting up and down along the axis in Fig. N4.20(a) up-and-to-the-right at roughly[87] 60°. The subgroup mixing down and strange forms a third pair of ladders $U_\pm$ moving at roughly $-60°$.

(c) *Show that* $[I_3, V_+] = ½V_+$, *using the representation* ***3*** *for the generators above.*

The three-fold symmetry of the raising and lowering operators gives us the three-fold symmetry of the baryon octet of Fig. N4.20(a).

(d) *If* $V_+|\Sigma^-\rangle = \alpha|n^0\rangle$ *for some constant* α, *calculate* $I_3V_+|\Sigma^-\rangle$ *and* $V_+I_3|\Sigma^-\rangle$. (Note that the isospin operator in $I_3V_+|\Sigma^-\rangle$ evaluates the isospin of a neutron, while in $V_+I_3|\Sigma^-\rangle$ it evaluates the isospin of the Σ^-.) *Check against your answer from part (c).*

Let us finally consider the fact that this $SU(3)$ flavor symmetry is not exact. In particular, the strange quark is quite heavy: note that the masses in Fig. N4.20(b) are mostly proportional to strangeness. Adding a big mass to the strange quark breaks the $SU(3)$ symmetry, but retains the isospin symmetry $SU(2) \subset SU(3)$. By decomposing the various $SU(3)$ representations into irreducible representations of $SU(2)$, we can understand another feature of the octets.

(e) *In the three-dimensional representation* ***3*** *used for the three quarks above, what are the two invariant subspaces*[88] *under isospin* $SU(2)$, *written in terms of the three quark directions in eqn N4.139?* (Hint: Which quarks are mixed by the isospin symmetry group? This is not subtle.) *Show that*[89] *the SU(3) representation* ***3*** *decomposes into*

[87] To make the diagram in Fig. N4.20(a) reflect the permutation symmetry between the three quarks of SU(3), we really should squash the vertical axis a bit, so that the horizontal distance between neutron and proton along the isospin direction I_3 equals the distance between Σ^- and n^0 up-and-to-the-right. Because we plot strangeness S along the vertical axis rather than the more symmmetrical $\sqrt{3}S/2$, moving along V_+ actually is at 63.43°.

[88] Reminder: a subspace of a 3D vector space is a line or plane through the origin—spanned by one or two vectors. Fundamental to group representation theory is the decomposition of representations into irreducible representations. Decomposing the representation involves finding subspaces which are invariant—mapped to themselves—under the group.

[89] (Do not be confused by the notation. It is traditional in SU(3) to label the representation by the number of dimensions of the corresponding vector space in **boldface**—hence **8** has eight dimensions. It is traditional

isospin SU(2) representations $0 \oplus \frac{1}{2}$. (Hint: you don't need character tables—the different isospin representations all have different dimensions.)

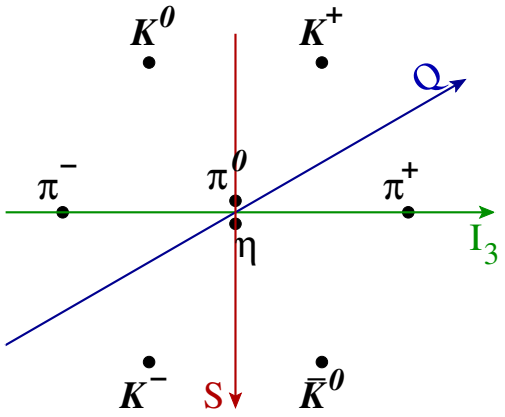


Fig. N4.20 Meson octet. The light mesons also share the same group representation **8** of SU(3). (Note that the strangeness here is zero for the pions. Note that the masses for these mesons is not linear in strangeness—it grows for both positive and negative strangeness away from zero. Note finally that the strangeness arrow is reversed both here and in Fig. N4.20(a) from the standard convention, which would have the neutron and proton have highest (least negative) strangeness of their family of baryons.)

We now want to build particles out of quarks. Adding three quarks to make a baryon leads to several representations. Let us instead add a quark and an antiquark to make a meson (Fig. N4.21). The antiquarks transform under a related representation $\bar{\mathbf{3}}$, which also decomposes into $0 \oplus \frac{1}{2}$ when SU(3) is broken to SU(2). A quark and an antiquark together form a nine-dimensional representation $\mathbf{3} \otimes \bar{\mathbf{3}} = \mathbf{8} \oplus \mathbf{1}$, where **1** is a one-dimensional representation corresponding in this case to the η' meson.

We need not refer to character orthogonality relations to find the isospin decomposition of **8**. Given that we know how the two three dimensional SU(3) reps **3** and $\bar{\mathbf{3}}$ decompose when broken to SU(2), we can deduce the decomposition of **8**. The direct sum of the decomposition of the nine-dimensional product rep $\mathbf{3} \otimes \bar{\mathbf{3}} = \mathbf{8} \oplus \mathbf{1}$. Since the one-dimensional rep **1** must lead to a one-dimensional isospin-zero piece, the isospin subspaces of **8** then can be deduced by elimination.[90]

in SU(2) to label a representation of dimension D by the maximum value of the spin $j = (D-1)/2$ without boldface. Hence **3** is a three-dimensional rep of SU(3), while 3 is a seven dimensional rep of SU(2).)

[90] Again, the notation is confusing: we use the same symbols for direct sums and products of vectors, vector spaces, and group representations. Let us give a review and summary, in case it is more helpful than it is distracting. The direct product of two vectors a_α and c_γ of dimensions d_A and d_C is the $d_A d_C$-dimensional vector with components $(\mathbf{a} \otimes \mathbf{c})_{\alpha\gamma} = a_\alpha c_\gamma$; thus a spin-up electron in a p_x state $\chi(\mathbf{x})$ and a spin-up muon in a d_{xy} state $\phi(\mathbf{y})$ make for a direct product wavefunction $\Psi(\mathbf{x}, \mathbf{y}) = \chi(\mathbf{x})\phi(\mathbf{y})$. The direct product of two vector spaces A and C is the space $A \times C$ spanned by all linear combinations of vectors in the two spaces. Thus a spin-up electron with $L = 1$ and a spin-up muon with $L = 2$ will have a Hilbert space of fifteen dimensions. The direct sum of **a** with **b** of dimension d_B is a vector $\mathbf{a} \oplus \mathbf{b}$ of dimension $d_A + d_B$ with first d_A components

To do this, we need to know that "direct product" is distributive over "direct sum", for group representations. Let A, B, and C be vector spaces, with elements $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$. Let R_A, R_B, and R_C be representations of the same group G, so for each $g \in G$ $R_A[g]$ is a matrix acting on vectors in A, etc. We want to show that $(R_A \oplus R_B) \otimes R_C$ is basically the same as (isomorphic to) $(R_A \otimes R_C) \oplus (R_B \otimes R_C)$.

The key is to label the vectors in the spaces $(A \oplus B) \otimes C$ and $(A \otimes C) \oplus (B \otimes C)$ in ways that make the correspondence clear. Let us write a vector in the spaces $(A \oplus B) \otimes C$ as $(\mathbf{a} \oplus \mathbf{b}) \otimes \mathbf{c})_{\omega,\gamma} = \mathbf{a}_\omega \mathbf{c}_\gamma$ if $\omega \leq d_A$ and $(\mathbf{a} \oplus \mathbf{b}) \otimes \mathbf{c})_{\omega,\gamma} = \mathbf{b}_{\omega - d_A} \mathbf{c}_\gamma$ if $\omega > d_A$. (The direct product is often written with two indices; indeed, matrices are direct products of vectors. The direct sum is usually implemented with one vector appended to the end of the other.) Let us write the corresponding vector in $(A \otimes C) \oplus (B \otimes C)$ so that the labels of the components are the same: $((\mathbf{a} \otimes \mathbf{c}) \oplus (\mathbf{b} \otimes \mathbf{c}))_{\omega,\gamma} = \mathbf{a}_\omega \mathbf{c}_\gamma$ if $\omega \leq d_A$ and $((\mathbf{a} \otimes \mathbf{c}) \oplus (\mathbf{b} \otimes \mathbf{c}))_{\omega,\gamma} = \mathbf{b}_{\omega - d_A} \mathbf{c}_\gamma$ if $\omega > d_A$.

Now we can argue that $(R_A \oplus R_B) \otimes R_C = (R_A \otimes R_C) \oplus (R_B \otimes R_C)$, and hence that direct products are distributive over direct sums for group representations too. We do this by showing that corresponding vectors (as defined above) rotate to corresponding vectors, i.e. that $(R_A[g]\mathbf{a} \oplus R_B[g]\mathbf{b}) \otimes R_c[g]\mathbf{c}$ corresponds to $(R_A[g]\mathbf{a} \otimes R_c[g]\mathbf{c}) \oplus (R_B[g]\mathbf{b} \otimes R_C[g]\mathbf{c})$.[91]

$$
\begin{aligned}
((R_A[g]\mathbf{a} \oplus R_B[g]\mathbf{b}) \otimes R_c[g]\mathbf{c})_{\omega,\gamma} &= (R_A[g]\mathbf{a})_\omega (R_c[g]\mathbf{c})_\gamma && \text{if } \omega \leq d_A, \text{and} \\
((R_A[g]\mathbf{a} \oplus R_B[g]\mathbf{b}) \otimes R_c[g]\mathbf{c})_{\omega,\gamma} &= (R_B[g]\mathbf{b})_{\omega - d_A} (R_c[g]\mathbf{c})_\gamma && \text{if } \omega > d_A.
\end{aligned}
$$

Similarly, (N4.84)

$$
\begin{aligned}
((R_A[g]\mathbf{a} \otimes R_c[g]\mathbf{c}) \oplus (R_B[g]\mathbf{b} \otimes R_C[g]\mathbf{c}))_{\omega,\gamma} &= (R_A[g]\mathbf{a})_\omega (R_c[g]\mathbf{c})_\gamma && \text{if } \omega \leq d_A, \text{and} \\
((R_A[g]\mathbf{a} \otimes R_c[g]\mathbf{c}) \oplus (R_B[g]\mathbf{b} \otimes R_C[g]\mathbf{c}))_{\omega,\gamma} &= (R_B[g]\mathbf{b})_{\omega - d_A} (R_c[g]\mathbf{c})_\gamma && \text{if } \omega > d_A.
\end{aligned}
$$

Hence the group acts on the two vector spaces in the same way.

(f) *Substitute the isospin decompositions for* $\mathbf{3}$ *and* $\bar{\mathbf{3}}$ *into the direct product* $\mathbf{3} \otimes \bar{\mathbf{3}}$ *representing the quark-antiquark bound states. Use the above distributive property to*

---

a_α followed by d_B components b_β. The direct sum of vector spaces A and B is the vector space $A \oplus B$ of these vectors. Thus if our electron could be in either a p state or an f state, it has a Hilbert space of dimension $3+7 = 10$ of possible states. The group representation $R_A \oplus R_B$ is block diagonal and acts on the vector space $A \oplus B$; $(R_A \oplus R_B)(g) \cdot (\mathbf{a} \oplus \mathbf{b}) = (R_A(g) \cdot \mathbf{a}) \oplus (R_B \cdot \mathbf{b})$; if we rotate about z by 90°, the p_x components of our electron wavefunction will rotate to p_y and the f_{xzz} components of our wavefunction will rotate to f_{yzz}. The group representation $R_A \otimes R_C$ acts on the vector space $A \otimes C$, $(R_A \otimes R_B)(g) \cdot (\sum \mathbf{a} \otimes \mathbf{b}) = \sum (R_A(g) \cdot \mathbf{a}) \otimes (R_B \cdot \mathbf{b})$; rotating our electron-muon wavefunction $\Psi(\mathbf{x}, \mathbf{y}) \to \Psi(R\mathbf{x}, R\mathbf{y})$ rotates single-particle product wavefunctions $\chi(\mathbf{x})\phi(\mathbf{y}) \to \chi(R\mathbf{x})\phi(R\mathbf{y})$; general wavefunctions are sums of these product wavefunctions. In part (f), we are thus showing that $(A \oplus B) \otimes C = (A \otimes C) \oplus (B \otimes C)$; an electron in either a p or an f state, together with a muon in a d state will be in a fifty dimensional Hilbert space. In part (g), we are decomposing a direct product of direct sums into a direct sum of invariant spaces. In our electron-muon system this would correspond to $(1 \oplus 3) \otimes 2 = (1 \otimes 2) \oplus (3 \otimes 2) = (3 \oplus 2 \oplus 1) \oplus (5 \oplus 4 \oplus 3 \oplus 2 \oplus 1)$; two reps of $L = 1$, 2, and 3, plus one rep each with $L = 4$ and 5. Do not forget—the SU(2) representation S has dimension $2S + 1$, so $1 \otimes 2$ is of dimension $3 \times 5 = 15$.

[91] If we are being fussy, we should show this is true also for a sum over triples of $\mathbf{a}_n$, $\mathbf{b}_n$, and $\mathbf{c}_n$, since a general vector in the direct product of two spaces is a linear combination of direct products of vectors in the two spaces.

decompose $\mathbf{3}\otimes\bar{\mathbf{3}}$ *into irreducible representations of the isospin subgroup* $SU(2)$. (Hint: There should be five invariant subspaces of this nine-dimensional space.)

(g) *How many of the isospin subspaces from part (f) should be part of the octet in Fig. N4.21? Give the total isospin and* I_3 *for each meson, grouping them into isospin multiplets (the mesons forming the different representations).*

N4.52 **Quantum dissipation from phonons.** (Quantum) ②

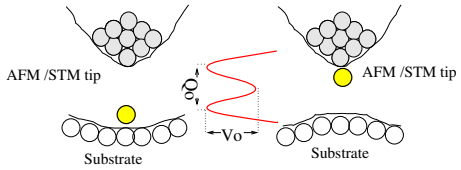


Fig. N4.21 Atomic tunneling from a tip. Any *internal* transition among the atoms in an insulator can only exert a force impulse (if it emits momentum, say into an emitted photon), or a force dipole (if the atomic configuration rearranges); these lead to nonzero phonon overlap integrals only partially suppressing the transition. But a quantum transition that changes the net force between two macroscopic objects (here a surface and a STM tip) can lead to a change in the net force (a force monopole). We ignore here the surface, modeling the force as exerted directly into the center of an insulating elastic medium.[92]See "Atomic Tunneling from a STM/AFM Tip: Dissipative Quantum Effects from Phonons" Ard A. Louis and James P. Sethna, *Phys. Rev. Lett.* **74**, 1363 (1995), and "Dissipative tunneling and orthogonality catastrophe in molecular transistors", S. Braig and K. Flensberg, *Phys. Rev. B* **70**, 085317 (2004).

Electrons cause overlap catastrophes (X-ray edge effects, the Kondo problem, macroscopic quantum tunneling); a quantum transition of a subsystem coupled to an electron bath ordinarily must emit an infinite number of electron-hole excitations because the bath states before and after the transition have zero overlap. This is often called an *infrared* catastrophe (because it is low-energy electrons and holes that cause the zero overlap), or an *orthogonality* catastrophe (even though the two bath states aren't just orthogonal, they are in different Hilbert spaces). Phonons typically do not produce overlap catastrophes (Debye–Waller, Frank–Condon, Mössbauer). This difference is usually attributed to the fact that there are many more low-energy electron-hole pairs (a constant density of states) than there are low-energy phonons ($\omega_k \sim ck$, where c is the speed of sound and the wave-vector density goes as $(V/2\pi)^3 \mathrm{d}^3k$).

However, the coupling strength to the low energy phonons has to be considered as well. Consider a small system undergoing a quantum transition which exerts a net force at $x = 0$ onto an insulating crystal:

$$\mathcal{H} = \sum_k p_k^2/2m + 1/2\, m\omega_k^2 q_k^2 + F \cdot u_0. \tag{N4.85}$$

Let us imagine a kind of scalar elasticity, to avoid dealing with the three phonon branches (two transverse and one longitudinal); we thus naively write the displacement of the atom at lattice site x_n as $u_n = (1/\sqrt{N}) \sum_k q_k \exp(-\mathrm{i}kx_n)$ (with N the number of atoms), so $q_k = (1/\sqrt{N}) \sum_n u_n \exp(\mathrm{i}kx_n)$.

Substituting for u_0 in the Hamiltonian and completing the square, find the displacement Δ_k of each harmonic oscillator. (Physically, the force F adds a small linear term to the phonon mode with wavevector k, whose minimum becomes displaced by some amount Δ_k.) Let $|F\rangle$ be the ground state of the harmonic oscillators under the force F. Write the formula for the likelihood $\langle F|0\rangle$ that the phonons will all end in their ground states, as a product over k of the phonon overlap integral $\exp(-\Delta_k^2/8\sigma_k^2)$ (with $\sigma_k = \sqrt{\hbar/2m\omega_k}$ the zero-point motion in that mode). Converting the product to the exponential of a sum, and the sum to an integral $\sum_k \sim (V/(2\pi)^3 \int \mathrm{d}^3\mathbf{k}$, do we observe an overlap catastrophe?

Note that you've calculated the probability of a zero-phonon transition—the likelihood that the quantum transition can happen without emitting any phonons is zero. But the same argument shows that there is zero probability of emitting one phonon, or any finite number of phonons. The only allowed transitions emit an infinite number of low-energy phonons. The initial and final ground states are in *different Hilbert spaces*—no finite number of excitations can connect them.

Chapter 5

Sloppy Models, Information Geometry, and Emergent Simplicity

Exercises for tiny bits of a draft textbook on information geometry, available at https://sethna.lassp.cornell.edu/Teaching/BasicTraining/SloppyBook.pdf.

N5.1 **Emergent vs. fundamental.** ⓟ

Statistical mechanics is central to condensed matter physics. It is our window into the behavior of materials—how complicated interactions between large numbers of atoms lead to physical laws (Fig. N5.1). For example, the theory of sound emerges from the complex interaction between many air molecules governed by Schrödinger's equation. More is different [10].

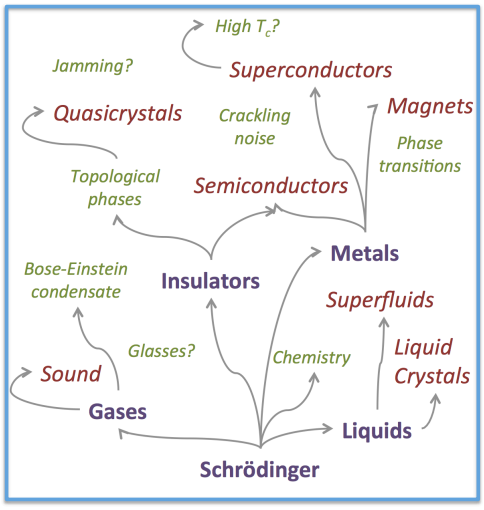


Fig. N5.1 Emergent. New laws describing macroscopic materials emerge from complicated microscopic behavior [178].

For example, if you inhale helium, your voice gets squeaky like Mickey Mouse. The dynamics of air molecules change when helium is introduced—the same law of motion, but with different constants.

(a) *Look up the wave equation for sound in gases. How many constants are needed? Do the details of the interactions between air molecules matter for sound waves in air?*

Statistical mechanics is tied also to particle physics and astrophysics. It is directly important in, e.g., the entropy of black holes (Exercise 7.16), the microwave background radiation (Exercises 7.15 and 10.1), and broken symmetry and phase transitions in the early Universe (Chapters 9, 11, and 12). Where statistical mechanics focuses on the *emergence* of comprehensible behavior at low energies, particle physics searches for the *fundamental* underpinnings at high energies (Fig. N5.2). Our different approaches reflect the complicated science at the atomic scale of chemistry and nuclear physics. At higher energies, atoms are described by elegant field theories (the *standard model* combining electroweak theory for electrons, photons, and neutrinos with QCD for quarks and gluons); at lower energies effective laws emerge for gases, solids, liquids, superconductors, ...

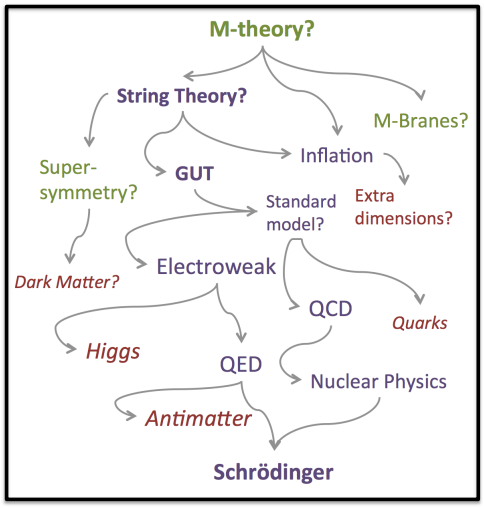


Fig. N5.2 Fundamental. Laws describing physics at lower energy emerge from more fundamental laws at higher energy [178].

The laws of physics involve parameters—real numbers that one must calculate or measure, like the speed of sound for a each gas at a given density and pressure. Together with the initial conditions (e.g., the density and its rate of change for a gas), the laws of physics allow us to predict how our system behaves.

Schrödinger's equation describes the Coulomb interactions between electrons and nuclei, and their interactions with electromagnetic field. It can in principle be solved to describe almost all of materials physics, biology, and engineering, apart from radioactive decay and gravity, using a Hamiltonian involving only the parameters $\hbar$, e, c, m_e,

and the the masses of the nuclei.[1] Nuclear physics and QCD in principle determine the nuclear masses; the values of the electron mass and the fine structure constant $\alpha = e^2/\hbar c$ could eventually be explained by even more fundamental theories.

(b) *About how many parameters would one need as input to Schrödinger's equation to describe materials and biology and such?* Hint: There are 253 stable nuclear isotopes.

(c) *Look up the Standard Model—our theory of electrons and light, quarks and gluons, that also in principle can be solved to describe our Universe (apart from gravity). About how many parameters are required for the Standard Model?*

In high-energy physics, fewer constants are usually needed to describe the fundamental theory than the low-energy, effective emergent theory—the fundamental theory is more elegant and beautiful. In condensed matter theory, the fundamental theory is usually less elegant and messier; the emergent theory has a kind of parameter compression, with only a few combinations of microscopic parameters giving the governing parameters (temperature, elastic constant, diffusion constant) for the emergent theory.

Note that this is partly because in condensed matter theory we confine our attention to one particular material at a time (crystals, liquids, superfluids). To describe all materials in our world, and their interactions, would demand many parameters.

My high-energy friends sometimes view this from a different perspective. They note that the methods we use to understand a new superfluid, or a topological insulator, are quite similar to the ones they use to study the Universe. They admit a bit of envy—that we get a new universe to study every time an experimentalist discovers another material.

N5.2 **Width of the height distribution.**[2] (Statistics) ③

In this exercise we shall explore statistical methods of fitting models to data, in the context of fitting a Gaussian to a distribution of measurements. We shall find that *maximum likelihood* methods can be *biased*. We shall find that all sensible methods converge as the number of measurements N gets large (just as thermodynamics can ignore fluctuations for large numbers of particles), but a careful treatment of fluctuations and probability distributions becomes important for small N (just as different ensembles become distinguishable for small numbers of particles).

The Gaussian distribution, known in statistics as the *normal* distribution

$$\mathcal{N}(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-(x-\mu)^2/(2\sigma^2)} \tag{N5.1}$$

is a remarkably good approximation for many properties. The heights of men or women in a given country, or the grades on an exam in a large class, will often have a his-

[1] The gyromagnetic ratio for each nucleus is also needed in a few situations where its coupling to magnetic fields are important.

[2] This exercise was developed in collaboration with Colin Clement.

togram that is well described by a normal distribution.[3] If we know the heights x_n of a sample with N people, we can write the likelihood that they were drawn from a normal distribution with mean μ and variance σ^2 as the product

$$P(\{x_n\}|\mu,\sigma) = \prod_{n=1}^{N} \mathcal{N}(x_n|\mu,\sigma^2). \tag{N5.2}$$

We first introduce the concept of *sufficient statistics*. Our likelihood (eqn N5.2) does not depend independently on each of the N heights x_n. What do we need to know about the sample to predict the likelihood?

(a) *Write $P(\{x_n\}|\mu,\sigma)$ in eqn N5.2 as a formula depending on the data $\{x_n\}$ only through N, $\overline{x} = (1/N)\sum_n x_n$ and $S = \sum_n (x_n - \overline{x})^2$.*

Given the model of independent normal distributions, its likelihood is a formula depending only on[4] $\overline{x}$ and S, the sufficient statistics for our Gaussian model.

Now, suppose we have a small sample and wish to estimate the mean and the standard deviation of the normal distribution.[5] *Maximum likelihood* is a common method for estimating model parameters; the estimates $(\mu_{\mathrm{ML}}, \sigma_{\mathrm{ML}})$ are given by the peak of the probability distribution P.

(b) *Show that $P(\{x_n\}|\mu_{\mathrm{ML}},\sigma_{\mathrm{ML}})$ takes its maximum value at*

$$\begin{aligned} \mu_{\mathrm{ML}} &= \frac{\sum_n x_n}{N} = \overline{x} \\ \sigma_{\mathrm{ML}} &= \sqrt{\sum_n (x_n - \overline{x})^2/N} = \sqrt{S/N}. \end{aligned} \tag{N5.3}$$

(Hint: It is easier to maximize the log likelihood; $P(\boldsymbol{\theta})$ and $\log(P(\boldsymbol{\theta}))$ are maximized at the same point $\boldsymbol{\theta}_{\mathrm{ML}}$.)

If we draw samples of size N from a distribution of known mean μ_0 and standard deviation σ_0, how do the maximum likelihood estimates differ from the actual values? For the limiting case $N = 1$, the various maximum likelihood estimates for the heights vary from sample to sample (with probability distribution $\mathcal{N}(x|\mu,\sigma^2)$, since the best estimate of the height is the sampled one). Because the average value $\langle \mu_{\mathrm{ML}} \rangle_{\mathrm{samp}}$ over many samples gives the correct mean, we say that μ_{ML} is *unbiased*. The maximum

[3] This is likely because one's height is determined by the additive effects of many roughly uncorrelated genes and life experiences; the central limit theorem would then imply a Gaussian distribution (Chapter 2 and Exercise 12.11).

[4] In this exercise we shall use $\overline{X}$ denote a quantity averaged over a single sample of N people, and $\langle X \rangle_{\mathrm{samp}}$ denote a quantity also averaged over many samples.

[5] In physics, we usually estimate measurement errors separately from fitting our observations to theoretical models, so each experimental data point d_i comes with its error σ_i. In statistics, the estimation of the measurement error is often part of the modeling process, as in this exercise.

likelihood estimate for $\sigma^2_{\rm ML}$, however, is biased. Again, for the extreme example $N = 1$, $\sigma^2_{\rm ML} = 0$ for every sample!

(c) *Assume the entire population is drawn from some (perhaps non-Gaussian) distribution of variance* $\langle x^2 \rangle_{\rm samp} = \sigma_0^2$. *For simplicity, let the mean of the population be zero. Show that*

$$\left\langle \sigma^2_{\rm ML} \right\rangle_{\rm samp} = (1/N) \left\langle \sum_{n=1}^{N} (x_n - \overline{x})^2 \right\rangle_{\rm samp} = \frac{N-1}{N} \sigma_0^2. \tag{N5.4}$$

that the variance for a group of N *people is on average smaller than the variance of the population distribution by a factor* $(N-1)/N$. (Hint: $\overline{x} = (1/N)\sum_n x_n$ is not necessarily zero. Expand it out and use the fact that x_m and x_n are uncorrelated.)

The maximum likelihood estimate for the variance is biased on average toward smaller values. Thus we are taught, when estimating the standard deviation of a distribution[6] from N measurements, to divide by $\sqrt{N-1}$:

$$\sigma^2_{\rm N-1} \approx \frac{\sum_n (x_n - \overline{x})^2}{N-1}. \tag{N5.5}$$

This correction $N \to N-1$ is generalized to more complicated problems by considering the number of independent degrees of freedom (here $N-1$ degrees of freedom in the vector $x_n - \overline{x}$ of deviations from the mean). Alternatively, it is interesting that the bias disappears if one does not estimate both σ^2 and μ by maximizing the joint likelihood, but integrating (or *marginalizing*) over μ and then finding the maximum likelihood for σ^2.

N5.3 **Statistical mechanics and statistics.**[7] (Statistics) ③

Consider the problem of fitting a theoretical model to experimentally determined data. Let our model predict a time-dependent function $y^\theta(t)$, where $\boldsymbol{\theta}$ are the model parameters. Let there be N experimentally determined data points d_i at times t_i with errors of standard deviation σ. We assume that the experimental errors for the data points are independent and Gaussian distributed, so that the probability that a given model produced the observed data points (the probability $P(D|\boldsymbol{\theta})$ of the data given the model) is

$$P(D|\boldsymbol{\theta}) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-\left(y^\theta(t_i) - d_i\right)^2 / 2\sigma^2}. \tag{N5.6}$$

(a) *True or false: This probability density corresponds to a Boltzmann distribution with energy* H *and temperature* T, *with* $H = \sum_{i=1}^{N} (y^\theta(t_i) - d_i)^2/2$ *and* $k_B T = \sigma^2$.

[6] Do not confuse this with the estimate of the error in the *mean* $\overline{x}$.

[7] This exercise was developed with the help of Robert Weiss.

There are two approaches to statistics. Among a family of models, the *frequentists* will pick the parameters $\boldsymbol{\theta}$ with the largest value of $P(D|\boldsymbol{\theta})$ (the *maximum likelihood estimate*); the ensemble of best-fit models is then deduced from the range of likely input data (deduced from the error bars σ). The *Bayesians* take a different point of view. They argue that there is no reason to believe a priori that all models have the same probability. (In model parameter space, there is no analogue of Liouville's theorem, Section 4.1.) Suppose the probability of the model (the *prior*) is $P(\boldsymbol{\theta})$. They use the theorem

$$P(\boldsymbol{\theta}|D) = P(D|\boldsymbol{\theta})P(\boldsymbol{\theta})/P(D). \tag{N5.7}$$

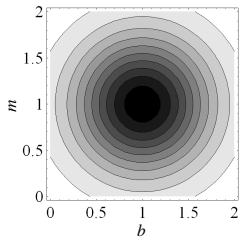
(b) *Prove Bayes' theorem (eqn N5.7) using the fact that* $P(A \text{ and } B) = P(A|B)P(B)$ *(see note 39 on p. 113).*

The Bayesians will often pick the maximum of $P(\boldsymbol{\theta}|D)$ as their model for the experimental data. But, given their perspective, it is even more natural to consider the entire *ensemble* of models, weighted by $P(\boldsymbol{\theta}|D)$, as the best description of the data. This ensemble average then naturally provides error bars for the parameters as well as for the predictions of various quantities.

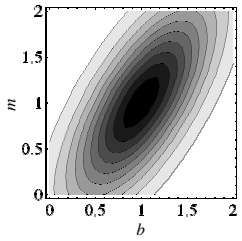
Consider the problem of fitting a line to two data points. Suppose the experimental data points are at $t_1 = 0$, $d_1 = 1$ and $t_2 = 1$, $d_2 = 2$, where both y-values have uncorrelated Gaussian errors with standard deviation $\sigma = 1/2$, as assumed in eqn N5.6 above. Our model $M(m, b)$, with parameters $\boldsymbol{\theta} = (m, b)$, is $y(t) = mt + b$. Our Bayesian statistician has prior knowledge 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 $0 < b < 2$.

(c) *Which of the contour plots shown accurately represent the probability distribution* $P(\boldsymbol{\theta}|D)$ *for the model, given the observed data? (The spacing between the contour lines is arbitrary.)*

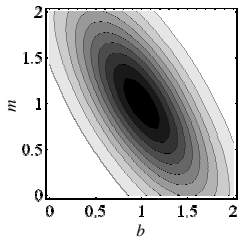
(A)



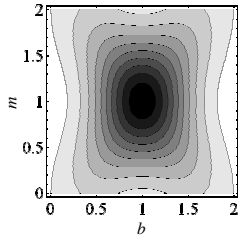
(B)



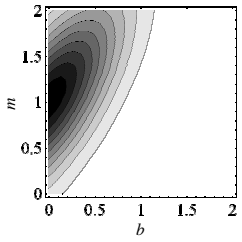
(C)



(D)



(E)



N5.4 **Sloppy exponentials.**[8] (Information geometry, Statistics) ②

The problem of extracting the decay rates from a sum of exponential decays is a famously difficult inverse problem, from the early days of radioactivity to modern simulations of lattice quantum chromodynamics [**?**]. In a series of exercises, we shall use our information geometry ideas to study the simplest version of this problem: the sum

[8] Hints for the computations can be found at the book website [31].

of N exponential decays:

$$y_{\boldsymbol{\theta}}(t) = (1/N)\sum_{\alpha=1}^{N} \exp(-\theta_\alpha t). \tag{N5.8}$$

We anticipate that it will be challenging to disentangle decay rates θ which are close to one another, unless one has high-precision data over large ranges of time. All the decay curves are smoothly monotonically decreasing, and one could imagine modeling a sum of two decays with a single intermediate decay rate. You shall find in these exercises that this simple model illustrates the behavior we have found widespread in multiparameter models in physics, engineering, biology, and other fields.

In this first exercise, we presume we have perfect experimental data for the decay $d(t)$ at M points t_i equally spread for t between 0 and 10, with separation $\Delta t = 10/M$. We shall be considering how well this data can be represented by other values of the parameters $\boldsymbol{\theta}$, so our cost (eqn **??**) is:

$$C(\boldsymbol{\theta}, \boldsymbol{\theta}^{[0]}) = \sum_{i=1}^{M} \left(y_{\boldsymbol{\theta}}(t_i) - y_{\boldsymbol{\theta}^{[0]}}(t_i)\right)^2 / 2\sigma^2 \;\approx \int_0^\infty \tfrac{1}{2} \left(y_{\boldsymbol{\theta}}(t) - y_{\boldsymbol{\theta}^{[0]}}(t)\right)^2 \, \mathrm{d}t. \tag{N5.9}$$

where for convenience (since our data is perfect) we set $\sigma^2 = 1/\Delta t$. We shall use the continuum approximation to evaluate the Hessian at the best fit.

To start, suppose $d(t)$ has two decay rates $\boldsymbol{\theta}^{[0]} = [1, 2]$, so the data $d(t) = \frac{1}{2}(\exp(-t) + \exp(-2t))$.

(a) *Write a function that returns $y_{\boldsymbol{\theta}}(t)$, and a function that computes the cost for $\Delta t = 0.01$. Draw a contour plot of C in the square $0.5 < \theta_\alpha < 2.5$, with contours at $C = \{2^{-12}, 2^{-11}, \ldots, 2^0\}$. Set the number of grid points per side to 40 (so $\Delta\theta = 0.02$) to see the two minima.*

The diagonal in this plot gives single exponential decays. How well does a single exponential capture the behavior at $\boldsymbol{\theta}^{[0]}$?

(b) *Constraining $\theta_1 = \theta_2$, find the point of minimum cost $\theta_{\min}$. Where is the point on the contour plot? Compare the two curves $y_{\theta^{[0]}}(t)$ and $y_{\theta_{\min}}(t)$, and also plot their difference.*

One can see from the contour plot that measuring the two rate constants separately would be a challenge. This is because the two exponentials have similar shapes, so increasing one decay rate and decreasing the other can almost perfectly compensate for one another.

This clearly is not a deep truth for two exponentials. But the effect is hugely magnified when we have many parameters. We can see this by computing the eigenvalues of the cost Hessian.

(c) *Analytically calculate the Jacobian $J_{t\alpha} = \partial y_{\boldsymbol{\theta}}(t)/\partial\theta_\alpha$ in the continuum approximation (eqn **??**). Using the Jacobian, show that the Hessian for the cost evaluated at the*

best fit is

$$\mathcal{H}_{\alpha\beta} = \left.\frac{\partial^2 C(\boldsymbol{\theta}, \boldsymbol{\theta}_0)}{\partial\theta_\alpha \partial\theta_\beta}\right|_{\boldsymbol{\theta}^{[0]}} = \frac{2}{N^2}\frac{1}{(\theta_\alpha + \theta_\beta)^3}. \tag{N5.10}$$

(Hint: See the discussion below eqn **??**.)

(d) *Using your answer from part (c), write a routine to calculate the entire array $H(\boldsymbol{\theta})$. Check it by examining the eigenvectors and eigenvalues for the $N = 2$ case of part (b). What do you predict the ratio R = (long axis/short axis) to be, in terms of the two eigenvalues λ_{stiffer} and $\lambda_{\text{sloppier}}$? Are the directions roughly in line with the eigenvectors?*

(e) *For a sum of seven exponentials, with $\boldsymbol{\theta}^{[0]} = [1, 2, 3, \ldots, 7]$, construct the Hessian, and find its eigenvalues. Are they sloppy (roughly equally spaced in log)? By roughly what factor does each successive eigenvalue shrink?*

This sloppiness makes it strikingly difficult to extract the parameter values from the data.

(f) *Argue that the number of measurements n_{measure} needed to estimate a parameter scales inversely with its variance ($n_{\text{measure}} \sim 1/\sigma^2$). Given that the eigenvalues of the Hessian give the variance along the various eigendirections, by what factor $n_{\text{sloppy}}/n_{\text{stiff}}$ is it harder to measure the parameters along the sloppy directions, for your sum of seven exponentials?*

(g) *Given that the diagonal elements of the inverse cost Hessian, $(\mathcal{H}^{-1})_{\alpha\alpha}$ are proportional to the variance in parameter α for one sampling of the Gaussian given by the cost, what are the variances in the seven parameters $\theta_\alpha^{[0]}$?*

N5.5 **Sloppy monomials.**[9] (Statistics) ③

The same function $f(x)$ can be approximated in many ways. Indeed, the same function can be fit in the same interval by the same type of function in several different ways! For example, in the interval $[0, 1]$, the function $\sin(2\pi x)$ can be approximated (badly) by a fifth-order Taylor expansion, a Chebyshev polynomial, or a least-squares (Legendre[10]) fit:

$$\begin{aligned} f(x) &= \sin(2\pi x) \\ f_{\text{Taylor}} &\approx 0.000 + 6.283x + \ 0.000x^2 - 41.342x^3 \\ &\quad + \ 0.000x^4 + 81.605x^5 \\ f_{\text{Chebyshev}} &\approx 0.0066 + 5.652x + \ 9.701x^2 - 95.455x^3 \\ &\quad + 133.48x^4 - 53.39x^5 \\ f_{\text{Legendre}} &\approx 0.016 + 5.410x + 11.304x^2 - 99.637x^3 \\ &\quad + 138.15x^4 - 55.26x^5 \end{aligned}$$

[9] Thanks to Joshua Waterfall, whose research is described here.

[10] The orthogonal polynomials used for least-squares fits on [-1,1] are the Legendre polynomials, assuming continuous data points. Were we using orthogonal polynomials for this exercise, we would need to shift them for use in [0,1].

It is not a surprise that the best fit polynomial differs from the Taylor expansion, since the latter is not a good approximation. But it is a surprise that the last two polynomials are so different. The maximum error for Legendre is less than 0.02, and for Chebyshev is less than 0.01, even though the two polynomials differ by

$$\begin{aligned}\text{Chebyshev} - \text{Legendre} = \\ &- 0.0094 + 0.242x - 1.603x^2 \\ &+ 4.182x^3 - 4.67x^4 + 1.87x^5\end{aligned} \tag{N5.11}$$

a polynomial with coefficients two hundred times larger than the maximum difference!

(a) *Plot* $f(x)$*,* f_{Legendre}*, and* $f_{\text{Chebyshev}}(x)$ *between zero and one on the same graph. Plot* $f(x) - f_{\text{Legendre}}$ *and* $f(x) - f_{\text{Chebyshev}}(x)$ *on the same graph with the same range. The first minimizes the squared difference on* $[0, 1]$ *(eqn* **??***), but it has large errors near the edges. If you were writing a routine to use for calculating* $\sin(2\pi x)$ *to machine precision in this range, would it be better to use the Legendre or the Chebyshev approximation? Now plot Plot* $f_{\text{Chebyshev}}(x) - f_{\text{Legendre}}$ *in the range* $-1, 2$*. Does it indeed get much flatter than you would expect given the coefficients?*

This flexibility in the coefficients of the polynomial expansion is remarkable. We can study it by considering the dependence of the quality of the fit on the parameters. Least-squares (Legendre) fits minimize a cost C, the integral of the squared difference between the polynomial and the function:

$$\begin{aligned}C &= (1/2) \int_0^1 (f(x) - y_{\boldsymbol{\theta}}(x))^2 \, \mathrm{d}x, \\ y_{\boldsymbol{\theta}}(x) &= \sum_{\alpha=0}^{N-1} \theta_\alpha x^\alpha\end{aligned} \tag{N5.12}$$

How quickly does this cost increase as we move the N parameters θ_α away from their best-fit values? Varying any one monomial coefficient will of course make the fit bad. But apparently certain coordinated changes of coefficients do not cost much—for example, the difference between least-squares and Chebyshev fits given in eqn **??**.

How should we explore the dependence in arbitrary directions in parameter space? We can use the eigenvalues of the Hessian to see how sensitive the fit is to moves along the various eigenvectors...

(b) *Note that the first derivative of the cost* C *is zero at the best fit. Analytically (paper and pencil) show that the Hessian second derivative of the cost in eqn* **??** *is*

$$\mathcal{H}_{\alpha\beta} = \frac{\partial^2 C}{\partial\theta_\alpha \partial\theta_\beta} = \frac{1}{\alpha + \beta + 1}. \tag{N5.13}$$

This Hessian is the Hilbert matrix, famous for being ill-conditioned (having a huge range of eigenvalues). Tiny eigenvalues of $\mathcal{H}$ correspond to directions in polynomial space where the fit does not change.

(c) *Numerically calculate the eigenvalues of the* 6×6 *Hessian for fifth-degree polynomial fits. Do they indeed span a large range? How big is the condition number (the ratio of the largest to the smallest eigenvalue)? Are the ratios all approximately equal (a characteristic of sloppy models)?*

Notice from Eqn **??** that the dependence of the polynomial fit on the monomial coefficients is *independent of the function* $f(x)$ being fitted. We can thus vividly illustrate the sloppiness of polynomial fits by considering fits to the *zero* function $f(x) \equiv 0$. A polynomial given by an eigenvector of the Hilbert matrix with small eigenvalue must stay close to zero everywhere in the range $[0, 1]$. Let us check this.

(d) *Calculate the eigenvector corresponding to the smallest eigenvalue of* $\mathcal{H}$*, checking to make sure its norm is one (so the coefficients are of order one). Note that the elements of this vector are the coefficients of a polynomial pertubation* $\delta f(x)$ *that changes the cost the smallest amount for a unit vector* $\boldsymbol{\theta}$*. What is that polynomial? Plot the corresponding polynomial in the range* $[0, 1]$*: does it stay small everywhere in the interval?*

Especially for larger M, the monomial coefficients of the best fit to a function become sloppy—they can vary over large ranges without damaging the fit, if the other coefficients are allowed to compensate. Only a few combinations of coefficients (those of the largest Hessian eigenvalues) are well determined. This turns out to be a fundamental property that is shared with many other multiparameter fitting problems. Many different terms are used to describe this property. The fits are called *ill-conditioned*: the parameters θ_n are not well constrained by the data. The *inverse problem* is challenging: one cannot practically extract the parameters from the behavior of the model. Or, as our group describes it, the fit is *sloppy*: only a few directions in parameter space (eigenvectors corresponding to the largest eigenvalues) are constrained by the data, and there is a huge space of models (polynomials) varying along sloppy directions that all serve well in describing the data.

At root, the problem with polynomial fits is that all monomials x^n have similar shapes on $[0, 1]$: they all start flat near zero and bend upward. Thus they can be traded for one another; the coefficient of x^4 can be lowered without changing the fit if the coefficients of x^3 and x^5 are suitably adjusted to compensate.

One should note that, were we change basis from the coefficients θ_n of the monomials x^n to the coefficients ℓ_n of the orthogonal (shifted Legendre) polynomials, the situation completely changes. The Legendre polynomials are designed to be different in shape (orthogonal), and hence cannot be traded for one another. Their coefficients ℓ_n are thus well determined by the data, and indeed the Hessian for the cost C in terms of this new basis is the identity matrix. This puzzled us for some time—is the sloppiness intrinsic, or just a sign of a poor choice of variables. Later work, examining the predictions of nonlinear models using *information geometry*, resolved this question: sloppiness is under rather general conditions expected for the collective predictions of multiparameter nonlinear models.

N5.6 **Nonlinear fits.** (Statistics, Information geometry) ③

In this exercise, we briefly introduce some geometrical features of nonlinear model fits to data. These fits involve unknown parameters θ_α, control parameters t_i describing different experimental conditions, experimental data d_i taken under these different conditions, and a nonlinear function $y_i(\boldsymbol{\theta})$ that makes a prediction for the data given values for the parameters. As an example, we might fit a sum of two decaying exponentials to (say) the decay of radiation from a mixture of radioactive elements with unknown decay rates (see [**?**, **?**] and Exercise N5.3.) Our model is

$$\mathbf{y}_{\boldsymbol{\theta}}(t) = \exp(-\theta_1 t) + \exp(-\theta_2 t). \tag{N5.14}$$

Here the parameters $\boldsymbol{\theta} = \{\theta_1, \theta_2\}$ are the decay rates, the control parameter is t the time elapsed, and the data $\mathbf{d} = \{d_i\}$ are the counts from a Geiger counter. We shall assume that the experimental data points $\{d_i \pm \sigma_i\}$ have independent measurement errors with Gaussian distributions of standard deviation σ_i.

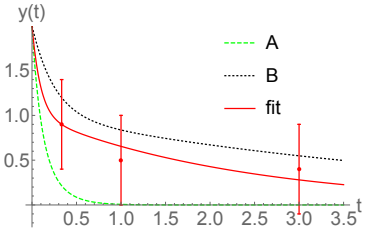


Fig. N5.3 Fitting a nonlinear function to data, here a sum of two exponentials to three data points $y(1/3) = 0.9 \pm 0.5$, $y(1) = 0.5 \pm 0.5$, and $y(2) = 0.4 \pm 0.5$. Fit A decays too quickly and fit B too slowly, although both are within statistical errors.

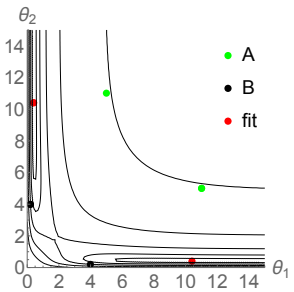


Fig. N5.4 Contours of constant cost $C = \chi^2/2$ in parameter space. Notice the symmetry reflecting around $\theta_1 = \theta_2$. Notice also the narrow canyons—one can fit the data well with $\theta_2 = \infty$ (a single exponential decaying from $y(0) = 1$), a point on the edge of the model manifold.

A nonlinear least-squares fit varies the parameters to minimize a cost

$$C(\boldsymbol{\theta}) = \chi^2/2 = \sum_i (y_i(\theta) - d_i)^2/2\sigma_i^2. \tag{N5.15}$$

The cost is half of what the statisticians call χ^2 (pronounced "chi squared").

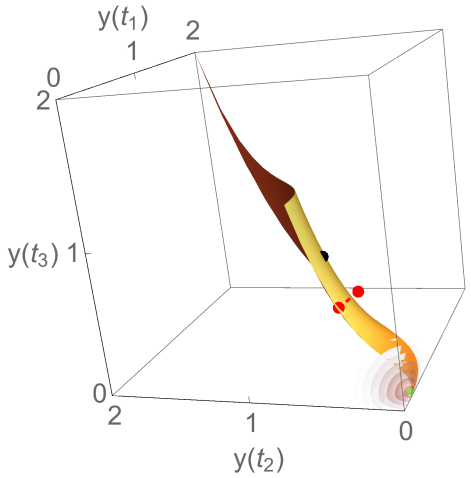


Fig. N5.5 Nonlinear model predictions in data space. The curved surface represents the *model manifold*—the surface of predictions in data space formed by varying the parameters of our nonlinear model. (We rescale the axes by the associated error bars.) The upper dot represents fit B. The dot at the lower right represents fit A, with the fuzzy sphere representing the range of experimental predictions around the fit. The two other dots represent the data and the best fit (the nearest point to the model manifold in data space). Note that the best fit is nearly at an edge of the model manifold.

First, let us provide a few interpretations of the cost.

(a) *[i] Interpret C as half the squared distance in a* data space *(Fig. N5.5) which has one coordinate per experimental measurement. What is the metric tensor g_{ij} in data space, in terms of the error bars σ_i?*[11] *[ii] Suppose the experimental data points d_i have errors that are distributed as independent Gaussians of standard deviation σ_i. How is our cost related to the log-likelihood that the data would have arisen from our model? [iii] View C as a Hamiltonian, and the likelihood $P(\mathbf{d}|\boldsymbol{\theta})$ giving the probability of observing data $\mathbf{d}$ in data space as being a Boltzmann distribution. What is the temperature T?*

Statistical mechanics focuses on predicting the behavior (probability distribution in phase space) for a Hamiltonian $\mathcal{H}(T, P, V, N)$ with known parameters. At fixed temperature, the probability density is proportional to $\exp(-\mathcal{H}/k_BT)$, where Liouville's theorem tells us how to measure phase-space volume. Statistics predicts the distribution of data points for a model $\mathbf{y}(\boldsymbol{\theta})$ with parameters $\boldsymbol{\theta}$. The probability distribution is proportional to $\exp(-C)$ per unit volume in data space, where the distance between two points in data space is determined by the error bars on each data point.

Finding the distribution of data points for a given set of parameters in statistics is not a challenge.

[11] The metric tensor g_{ij} on a Riemannian manifold gives the distance between nearby points. If the two points have coordinates $\mathbf{x}$ and $\mathbf{x} + \boldsymbol{\Delta}$ and $\boldsymbol{\Delta}$ is small, then the squared distance is $\sum_{ij} g_{ij}\Delta_i\Delta_j$.

(b) *Argue, for equal Gaussian measurement errors, that the predicted distribution of data points for a given set of parameters* $\boldsymbol{\theta}$ *is just a blurred, Gaussian sphere in data space (as in the lower right corner of Fig. N5.5). For general* σ_i*, make an analogy with the momentum distribution of classical particles with different masses to describe the probability distribution.*

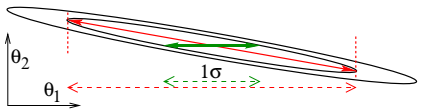


Fig. N5.6 Error estimates for fit parameters. Contours of constant cost C in parameter space $\boldsymbol{\theta}$ near the best fit, ignoring anharmonicity. The ellipse axes are $\mathbf{e}_v = (1/6, 1)$ and $\mathbf{e}_h = (1, -1/6)$. The 1σ range of θ_1 keeping θ_2 fixed is the solid arrow. The total uncertainty σ_1 for θ_1 includes fluctuations of θ_2 (solid diagonal arrow), $\sigma_1^2 = \Sigma_{11} = (H^{-1})_{11}$ (long dashed range).

Our job in nonlinear fitting is to estimate the probabilities of different choices of parameters given the experimental data. Surely we expect the true parameters to have a large probability $P(\mathbf{d}|\boldsymbol{\theta})$ of generating the experimental data—the true $\boldsymbol{\theta}$ will be somewhere near the best fit $\boldsymbol{\theta}^{\text{best}}$ that minimizes the cost $C = \chi^2$. Let us assume that the probability $P(\boldsymbol{\theta}|\mathbf{d})$ of finding a set of parameters given the data is proportional to the probability $P(\mathbf{d}|\boldsymbol{\theta})$ that the model would have generated that data. (See Exercise N5.2 for a discussion of *priors* in Bayesian statistics.) Let us also assume that we are estimating the parameters well enough that we may approximate the cost by a Taylor expansion up to second order about the maximum likelihood. If $\boldsymbol{\theta} = \boldsymbol{\theta}^{\text{best}} + \boldsymbol{\Delta}$ for small $\boldsymbol{\Delta}$,

$$C(\boldsymbol{\theta}) \approx C^{\text{best}} + \tfrac{1}{2} H_{\alpha\beta}\Delta_\alpha\Delta_\beta, \tag{N5.16}$$

where we shall call

$$H_{\alpha\beta} = \frac{\partial^2 C}{\partial\theta_\alpha \partial\theta_\beta} \tag{N5.17}$$

the cost Hessian.

(c) *Which of the eigenvectors* $\mathbf{e}_v$ *or* $\mathbf{e}_h$ *in Fig. N5.6 corresponds to a stiff direction (larger eigenvalue of* H*)? Which is sloppier? Verify that the probability distribution of* θ_1 *holding all other variables fixed is a normal distribution with variance* H_{11} *(short horizontal dashed range).*

It is more of a challenge to calculate the error in our estimate of θ_1 *allowing the other variables* to vary freely (long horizontal dashed range). The variance of the estimate of a variable is given by the corresponding diagonal element of the *covariance matrix* $\Sigma = H^{-1}$, the inverse of the Hessian.

If $P(\boldsymbol{\theta})$ is approximately a multidimensional Gaussian, then the variance in θ_1 is given

by

$$\begin{aligned}\langle\Delta_1^2\rangle &= \int \Delta_1^2 P(\boldsymbol{\Delta})\mathrm{d}\boldsymbol{\Delta} \\ &= \int \frac{\Delta_1^2}{Z}\mathrm{e}^{-\frac{1}{2}\sum_{\alpha\beta}\Delta_\alpha H_{\alpha\beta}\Delta_\beta}\mathrm{d}\boldsymbol{\Delta},\end{aligned} \tag{N5.18}$$

where

$$Z = \int \exp\left(-\tfrac{1}{2}\sum_{\alpha\beta}\Delta_\alpha H_{\alpha\beta}\Delta_\beta\right)\mathrm{d}\boldsymbol{\Delta}. \tag{N5.19}$$

is the normalization factor.

In statistical mechanics, a key method for calculating expectation values $\langle X^n\rangle$ in a Boltzmann distributions is to add a *source term* λX to the Hamiltonian, shifting the partition function to $Z(\lambda) = \sum \exp(-(H+\lambda X)/k_BT)$, with free energy $F(\lambda) = -k_BT\log Z$. Then

$$\begin{aligned}\left.\frac{\mathrm{d}F}{\mathrm{d}\lambda}\right|_{\lambda=0} &= \frac{-k_BT}{Z}\left.\frac{\mathrm{d}Z}{\mathrm{d}\lambda}\right|_{\lambda=0} \\ &= \frac{\cancel{-k_BT}}{Z}\int \frac{X}{\cancel{-k_BT}}\mathrm{e}^{-H/k_BT} \\ &= \langle X\rangle\end{aligned} \tag{N5.20}$$

and

$$\begin{aligned}\left.\frac{\mathrm{d}^2F}{\mathrm{d}\lambda^2}\right|_{\lambda=0} &= \left.\frac{k_BT}{Z^2}\left(\frac{\mathrm{d}Z}{\mathrm{d}\lambda}\right)^2\right|_{\lambda=0} - \left.\frac{k_BT}{Z}\frac{\mathrm{d}^2Z}{\mathrm{d}\lambda^2}\right|_{\lambda=0} \\ &= \frac{\langle X\rangle^2}{k_BT} - \frac{k_BT}{Z}\int \frac{X^2}{(k_BT)^2}\mathrm{e}^{-H/k_BT} \\ &= \frac{\langle X^2\rangle - \langle X\rangle^2}{-k_BT} \\ &= \frac{\left\langle (X-\langle X\rangle)^2\right\rangle}{-k_BT}\end{aligned} \tag{N5.21}$$

We can use this method to calculate $\langle\Delta_1^2\rangle$.

(d) *Add the source term* $\boldsymbol{\lambda}\cdot\boldsymbol{\Delta} = \lambda\Delta_1$ *to our cost, where* $\boldsymbol{\lambda}^T = (\lambda, 0, 0, \dots)$ *is* λ *times a unit vector in the shared* θ_1 *and* Δ_1 *direction, so*

$$Z(\lambda) = \int \mathrm{e}^{-\frac{1}{2}\boldsymbol{\Delta}^T H\boldsymbol{\Delta} - \boldsymbol{\lambda}\cdot\boldsymbol{\Delta}}\mathrm{d}\boldsymbol{\Delta}. \tag{N5.22}$$

Complete the square, and show that $Z(\lambda) = \exp(\frac{1}{2}\lambda^2\Sigma_{11})Z(0)$. *Use eqn N5.15 to show that* $\langle\Delta_1^2\rangle = \Sigma_{11}$.

There is a commonly used approximation to the cost Hessian that has important geometrical significance.

(e) *[i] Write the cost Hessian in eqn N5.9 in terms of first and second derivatives of* $y_i(\boldsymbol{\theta})$. *[ii] If we take the cost Hessian at a point where* $\mathbf{d} = \mathbf{y}(\boldsymbol{\theta})$ *on the model manifold, show that* $H_{\alpha\beta} = \sum_i (\partial y_i/\partial\theta_\alpha)(\partial y_i/\partial\theta_\beta) = (J^T J)_{\alpha\beta}$, *where* $J_{i\alpha} = (1/\sigma_i)\partial y_i/\partial\theta_\alpha$ *is the Jacobian. [iii] Show that the squared distance in data space between two model predictions* $\mathbf{y}(\boldsymbol{\theta})$ *and* $\mathbf{y}(\boldsymbol{\theta}+\boldsymbol{\Delta})$ *is given for small* $\boldsymbol{\Delta}$ *by the metric tensor* $g_{\alpha\beta} = (J^T J)_{\alpha\beta}$.

$g_{\alpha\beta} = (J^T J)_{\alpha\beta} = J_{i\alpha} J_{i\beta}$ is the *induced* metric on the model manifold, inherited from the embedding data space metric g_{ij} of part (a). $g = J^T J$ is called the *Fisher information matrix* in the statistics community.

N5.7 **Fisher information and Cramér–Rao.**[12] (Statistics, Mathematics, Information geometry) ④

Here we explore the geometry of the space of probability distributions. When one changes the external conditions of a system a small amount, how much does the ensemble of predicted states change? What is the *metric* in probability space? Can we predict how easy it is to detect a change in external parameters by doing experiments on the resulting distribution of states? The metric we find will be the *Fisher information matrix* (FIM). The *Cramér–Rao bound* will use the FIM to provide a rigorous limit on the precision of any (unbiased) measurement of parameter values.

In both statistical mechanics and statistics, our models generate probability distributions $P(\mathbf{x}|\boldsymbol{\theta})$ for behaviors $\mathbf{x}$ given parameters $\boldsymbol{\theta}$.

- A crooked gambler's loaded die, where the state space is comprised of discrete rolls $\mathbf{x} \in \{1, 2, \ldots, 6\}$ with probabilities $\boldsymbol{\theta} = \{p_1, \ldots, p_5\}$, with $p_6 = 1 - \sum_{j=1}^{5} \theta_j$.
- The probability density that a system with a Hamiltonian $\mathcal{H}(\boldsymbol{\theta})$ with $\boldsymbol{\theta} = (T, P, N)$ giving the temperature, pressure, and number of particles, will have a probability density $P(\mathbf{x}|\boldsymbol{\theta}) = \exp(-\mathcal{H}/k_B T)/Z$ in phase space (Chapter 3, Exercise N5.10).
- The height of women in the US, $\mathbf{x} = \{h\}$ has a probability distribution well described by a normal (or Gaussian) distribution $P(\mathbf{x}|\boldsymbol{\theta}) = 1/\sqrt{2\pi\sigma^2}\exp(-(x-\mu)^2/2\sigma^2)$ with mean and standard deviation $\boldsymbol{\theta} = (\mu, \sigma)$ (Exercise N5.2).
- A least squares model $y_i(\boldsymbol{\theta})$ for N data points $d_i \pm \sigma$ with independent, normally distributed measurement errors predicts a likelihood for finding a value $\mathbf{x} = \{x_i\}$ of the data $\{d_i\}$ given by

$$P(\mathbf{x}|\boldsymbol{\theta}) = \frac{\mathrm{e}^{-\sum_i (y_i(\theta)-x_i)^2/2\sigma^2}}{(2\pi\sigma^2)^{N/2}}. \tag{N5.23}$$

(Think of the theory curves you fit to data in many experimental labs courses.)

[12] This exercise was developed in collaboration with Colin Clement and Katherine Quinn.

How "distant" is a loaded die is from a fair one? How "far apart" are the probability distributions of particles in phase space for two small system at different temperatures and pressures? How hard would it be to distinguish a group of US women from a group of Pakistani women, if you only knew their heights?

We start with the least-squares model.

(a) *How big is the probability density that a least-squares model with true parameters* $\boldsymbol{\theta}$ *would give experimental results implying a different set of parameters* $\boldsymbol{\phi}$*? Show that it depends only on the distance between the vectors* $|\mathbf{y}(\boldsymbol{\theta}) - \mathbf{y}(\boldsymbol{\phi})|$ *in the space of predictions.* Thus the predictions of least-squares models form a natural manifold in a behavior space, with a coordinate system given by the parameters. The point on the manifold corresponding to parameters $\boldsymbol{\theta}$ is $\mathbf{y}(\boldsymbol{\theta})/\sigma$ given by model predictions rescaled by their error bars, $\mathbf{y}(\boldsymbol{\theta})/\sigma$.

Remember that the metric tensor $g_{\alpha\beta}$ gives the distance on the manifold between two nearby points. The squared distance between points with coordinates $\boldsymbol{\theta}$ and $\boldsymbol{\theta} + \epsilon\boldsymbol{\Delta}$ is $\epsilon^2 \sum_{\alpha\beta} g_{\alpha\beta}\Delta_\alpha\Delta_\beta$.

(b) *Show that the least-squares metric is* $g_{\alpha\beta} = (J^T J)_{\alpha\beta}/\sigma^2$*, where the Jacobian* $J_{i\alpha} = \partial y_i/\partial\theta_\alpha$.

For general probability distributions, the natural metric describing the distance between two nearby distributions $P(\mathbf{x}|\boldsymbol{\theta})$ and $Q = P(\mathbf{x}|\boldsymbol{\theta} + \epsilon\boldsymbol{\Delta})$ is given by the FIM:

$$g_{\alpha\beta}(\boldsymbol{\theta}) = -\left\langle \frac{\partial^2 \log P(\mathbf{x}|\boldsymbol{\theta})}{\partial\theta_\alpha \partial\theta_\beta} \right\rangle_{\mathbf{x}} \tag{N5.24}$$

Are the distances between least-squares models we intuited in parts (a) and (b) compatible with the the FIM?

(c) *Show for a least-squares model that eqn N5.18 is the same as the metric we derived in part (b).* (Hint: For a Gaussian distribution $\exp((x-\mu)^2/(2\sigma^2))/\sqrt{2\pi\sigma^2}$, $\langle x\rangle = \mu$.)

If we have experimental data with errors, how well can we estimate the parameters in our theoretical model, given a fit? As in part (a), now for general probabilistic models, how big is the probability density that an experiment with true parameters $\boldsymbol{\theta}$ would give results perfectly corresponding to a nearby set of parameters $\boldsymbol{\theta} + \epsilon\boldsymbol{\Delta}$?

(d) *Take the Taylor series of* $\log P(\boldsymbol{\theta} + \epsilon\boldsymbol{\Delta})$ *to second order in* ϵ*. Exponentiate this to estimate how much the probability of measuring values corresponding to the predictions at* $\boldsymbol{\theta} + \epsilon\boldsymbol{\Delta}$ *fall off compared to* $P(\boldsymbol{\theta})$. Thus to linear order the FIM $g_{\alpha\beta}$ estimates the range of likely measured parameters around the true parameters of the model.

The *Cramér–Rao bound* shows that this estimate is related to a rigorous bound. In particular, errors in a multiparameter fit are usually described by a *covariance matrix* Σ, where the variance of the likely values of parameter θ_α is given by $\Sigma_{\alpha\alpha}$, and where $\Sigma_{\alpha\beta}$ gives the correlations between two parameters θ_α and θ_β. One can show within our quadratic approximation of part (d) that the covariance matrix is the inverse of the FIM $\Sigma_{\alpha\beta} = (g^{-1})_{\alpha\beta}$. The *Cramér–Rao bound* roughly tells us that no experiment

can do better than this at estimating parameters. In particular, it tells us that the error range of the individual parameters from a sampling of a probability distribution is bounded below by the corresponding element of the inverse of the FIM

$$\Sigma_{\alpha\alpha} \geq (g^{-1})_{\alpha\alpha}. \tag{N5.25}$$

(if the estimator is *unbiased*, see Exercise N5.2). This is another justification for using the FIM as our natural distance metric in probability space.

In Exercise N5.18, we shall examine *global* measures of distance or distinguishability between potentially quite different probability distributions. There we shall show that these measures all reduce to the FIM to lowest order in the change in parameters. In Exercises N5.19, N5.9, and N5.10, we shall show that the FIM for a Gibbs ensemble as a function of temperature and pressure can be written in terms of thermodynamic quantities like compressibility and specific heat. There we use the FIM to estimate the *path length in probability space*, in order to estimate the entropy cost of controlling systems like the Carnot cycle.

N5.8 **Gibbs for pistons.** (Thermodynamics) ④

The degrees of freedom in a piston are $\mathbf{X} = \{\mathbb{P}, \mathbb{Q}, V\}$, where $\mathbb{P}$ and $\mathbb{Q}$ are the $3N$ positions and momenta of the particles, and V is the current volume of the piston. The Gibbs ensemble for a piston is the probability density

$$\rho = (1/\Gamma) \exp(-\beta \mathcal{H}(\mathbb{P}, \mathbb{Q}) - \beta P V). \tag{N5.26}$$

Here Γ is the partition function for the Gibbs ensemble, normalizing the distribution to one.

Let our piston be filled with an ideal gas of particles of mass m. What is the partition function $Z(V, \beta)$ for the canonical ensemble? (Be sure to include the Gibbs factor $N!$; the quantum phase-space refinements are optional.) Show that the partition function for the Gibbs ensemble is

$$\Gamma(P, \beta) = (2\pi m/\beta)^{3N/2} (\beta P)^{-(N+1)}, \tag{N5.27}$$

Show that the joint probability density for finding the N particles with $3N$ dimensional momenta $\mathbb{P}$, the piston with volume V, and the $3N$ dimensional positions $\mathbb{Q}$ inside V (eqn N5.20), is

$$\rho_{\text{Gibbs}}(\mathbb{P}, \mathbb{Q}, V | P, \beta) = (1/\Gamma(P, \beta)) \mathrm{e}^{-\beta \mathbb{P}^2/2m - \beta P V}. \tag{N5.28}$$

N5.9 **Pistons in probability space.**[13] (Mathematics, Information geometry) ④

Fig. 5.3 shows the Carnot cycle as a path in the P–V space of pressure and volume—parameters varied from the outside. One could draw a similar diagram in the space of pressure and temperature, or volume and temperature. Here we shall explore how to describe the path in the space of *probability distributions*. In the process, we shall compute the *model manifold* of the ideal gas, and show that it is a two-dimensional plane.

As discussed in Exercise N5.7, there is a natural distance, or metric, in the space of probability distributions:

$$g_{\mu\nu} = -\left\langle \frac{\partial^2 \log(\rho)}{\partial\theta_\mu \partial\theta_\nu} \right\rangle, \tag{N5.29}$$

the *Fisher information metric*. So, a system in the Gibbs ensemble is described in terms of two parameters, usually P and T. We shall instead use the "natural" parameters $\theta_1 = p = \beta P$ and $\theta_2 = \beta$, where $\beta = 1/k_B T$ (see Exercise N5.10). The squared distance in probability space between two systems with tiny changes in pressure and temperature is then

$$d^2(\rho(\mathbf{X}|\boldsymbol{\theta}), \rho(\mathbf{X}|\boldsymbol{\theta} + \mathrm{d}\boldsymbol{\theta})) = g_{\mu\nu}\mathrm{d}\theta_\mu \mathrm{d}\theta_\nu. \tag{N5.30}$$

(a) *Compute* $g^{(\mathrm{p},\beta)}_{\mu\nu} = -\langle \partial^2 \log(\rho)/\partial\theta_\mu\partial\theta_\nu \rangle$ *using eqn N5.22 from Exercise N5.8.*

The metric tensor $g^{(\mathrm{p},\beta)}$ for the Gibbs ensemble of the piston tells us the distance in probability space between neighboring pressures and temperatures. What kind of surface (the *model manifold*) is formed by this two-parameter family of probability distributions? Does it have an intrinsic curvature?

(b) *Show that one can turn the metric tensor into the identity* $g^{(\mathrm{x},\mathrm{y})}_{\mu\nu} = \delta_{\mu\nu}$ *by a coordinate transformation* $(p, \beta) \to (x = A\log(p), y = B\log(\beta))$. *What are the necessary scale factors* A *and* B*?*

Hence the model manifold of the piston in the Gibbs ensemble is a plane! We can draw our control paths in the (x, y) plane. We label the four steps of the Carnot cycle as in Fig. 5.3.

(c) *Draw the Carnot cycle path in as a parameterized curve in* (x, y)*, with* $P_a = 1$*,* $P_b = 0.5$*,* $T_1 = 1$ *and* $T_2 = 0.8$*, for* $N = 1$. (Hint: eqn 5.8 will be helpful in finding the adiabatic parts of the path $p(\beta)$.) *Is the length of the expansion at fixed pressure the same as you calculated in Exercise N5.19?*

[13] This exercise was developed in collaboration with Ben Machta, Archishman Raju, Colin Clement, and Katherine Quinn

N5.10 **FIM for Gibbs.**[14] (Mathematics, Thermodynamics, Information geometry) ④

In this exercise, we study the geometry in the space of probability distributions defined by the Gibbs ensemble[15] of a general equilibrium system. We compute the Fisher Information Metric (FIM, Exercises N5.7 and N5.9)

$$g_{\mu\nu} = -\left\langle \frac{\partial^2 \log(\rho)}{\partial\theta_\mu \partial\theta_\nu} \right\rangle, \tag{N5.31}$$

of the Gibbs phase space ensemble $\rho(\mathbb{P},\mathbb{Q})$ in terms of thermodynamic properties of the system.

In Exercise N5.9 we calculated $g_{\mu\nu}$ for the ideal gas, using the "natural" variables $\theta_1 = p = \beta P$ and $\theta_2 = \beta$, rather than P and T. Why are these coordinates special? The log of the Gibbs probability distribution for an arbitrary interacting collection of particles with Hamiltonian $\mathcal{H}$ (eqn N5.20) is

$$\begin{aligned} \log(\rho) &= -\beta\mathcal{H}(\mathbb{P},\mathbb{Q}) - \beta PV - \log\Gamma \\ &= -\beta\mathcal{H}(\mathbb{P},\mathbb{Q}) - pV - \log\Gamma. \end{aligned} \tag{N5.32}$$

This is the logarithm of the partition function Γ plus terms *linear* in $p = \beta P$ and β.[16] So the second derivatives with respect to p and β only involve $\log(\Gamma)$. We know that the Gibbs free energy $G(p,\beta) = -k_B T \log(\Gamma) = -(1/\beta)\log(\Gamma(p,\beta))$, so $\log(\Gamma) = -\beta G(p,\beta)$. The first derivatives of the Gibbs free energy $\mathrm{d}G = -S\mathrm{d}T + V\mathrm{d}P + \mu\mathrm{d}N$ are related to things like volume and entropy and chemical potential; our metric is given by the second derivatives (compressibility, specific heat, ...)

(a) *For a collection of particles interacting with Hamiltonian $\mathcal{H}$, relate the four terms $g_{\mu\nu}^{(\mathrm{p},\beta)}$ in terms of physical quantities given by the second derivatives of G. Write your answer in terms of N, p, β, the particle density $\rho = N/\langle V\rangle$, the isothermal compressibility $\kappa = -(1/\langle V\rangle)(\partial\langle V\rangle/\partial P)|_T$, the thermal expansion coefficient $\alpha = (1/\langle V\rangle)(\partial\langle V\rangle/\partial T)|_P$, and the specific heat per particle at constant pressure, $c_\mathrm{P} = (T/N)(\partial S/\partial T)|_P$.* (Hint: $G(P,T) = G(p/\beta, 1/\beta)$. Your answer will be a bit less complicated if you pull out an overall factor of $N/(\rho\beta^2)$.)

The metric tensor for a general Hamiltonian is a bit simpler in the more usual coordinates (P,β) or (P,T).

(b) *Show that*

$$g^{(\mathrm{P},\beta)} = N \begin{pmatrix} \beta\kappa/\rho & \alpha/\beta\rho \\ \alpha/\beta\rho & c_\mathrm{P}/\beta^2 \end{pmatrix}$$

[14] This exercise was developed in collaboration with Ben Machta, Archishman Raju, Colin Clement, and Katherine Quinn

[15] The Fisher information distance is badly defined except for changes in *intensive* quantities. In a microcanonical ensemble, for example, the energy E is constant and so the derivative $\partial\rho/\partial E$ would be the derivative of a δ function. So we study pistons varying P and $\beta = 1/k_B T$, rather than at fixed volume or energy.

[16] In statistics, log probability distributions which depend on parameters in this linear fashion are called *exponential families*. Many common distributions, including lots of statistical mechanical models like ours, are exponential families.

and

$$g^{(\mathrm{P,T})} = N \begin{pmatrix} \kappa/\rho T & -\alpha/\rho T \\ -\alpha/\rho T & c_{\mathrm{P}}/T^2 \end{pmatrix}.$$

(c) *Calculate $g^{(\mathrm{p},\beta)}$ for the ideal gas using your answer from part (a). Compare with your results calculating $g^{(\mathrm{p},\beta)}$ directly from the probability distribution in Exercise N5.9. Is the difference significant for macroscopic systems?* (Hint: If you use $G = A + PV$ directly from eqn 6.24, remember that the thermal de Broglie wavelength λ depends on temperature.)

The standard formulas for an ideal gas do not include the piston wall as a degree of freedom, so part (c) has one fewer positional degree of freedom than in Exercise N5.9. That is, the macroscopic calculation neglects the entropic contribution of the fluctuations in volume (the position of the piston inside the cylinder).

N5.11 **Plotting the model manifold.**[17] (Information geometry, Statistics) ③

In this exercise, we shall use our N parameter model of decaying exponentials explore ways of visualizing the resulting behavior. Remember eqn **??** from Exercise N5.4, $y_{\boldsymbol{\theta}}(t) = (1/N)\sum_{\alpha=1}^{N} \exp(-\theta_\alpha t)$.

One way of visualizing the behavior space is to pick two or three quantities of interest, and explore how they vary with one another. This is a *projection* of the model manifold onto the three coordinate axes of interest.

(a) *Taking $N = 2$ different exponents, draw the projection of the the model manifold onto the axes corresponding to $t = \{1/3, 1, 3\}$.* (That is, do a 3D parametric plot of $\{y_{\boldsymbol{\theta}}(1/3), y_{\boldsymbol{\theta}}(1), y_{\boldsymbol{\theta}}(3)\}$, varying $0 < \theta_1 < \infty$ and $0 < \theta_2 < \infty$.) *You'll want some of the θs to be small and some large.* (I had a range $\sim (10^{-2}$–10^{2}, with equally spaced points in log if on a grid). *Identify the two values of θ for the three pointy endpoints. What simpler model (with only one parameter) is associated with the three edges of the manifold?*

What happens when we take N exponentials? The model manifold will no longer be a surface – it will fill an N-dimensional manifold, and its projection into 3D will also fill a volume. We cannot expect to do this with a grid of points: 10 points in each direction for $N = 7$ θs would be 10^7 curves. Let us choose random vectors to get an idea of the shape.

(b) *Select a large number of random vectors in the space of parameters $0 < \theta_\alpha < \infty$. Starting with $N = 2$, reproduce the model manifold you found in part (a). You'll want some of the θs to be small and some large to get to the edges: choose values with probability $\rho(\theta) \propto 1/\theta$ in the same range you used in part (a). Try $N = 7$. Rotate the 3D plot to see how "thick" the model manifold is.*

You should find a thin sheet with what appears to be pointy-tipped scallops along one "edge". For $N = 1$, there were three points on the manifold, including the two at the ends of the edge.

[17] Hints for the computations can be found at the book website [31].

(c) *How many points do you observe for $N = 7$? From part (a), argue that the three cuspy points for $N = 2$ correspond to values where $y(t)$ is a constant except perhaps at $t = 0$. Is the same (likely) true for $N = 7$?*

The cusps in the model manifold are simpler models with no adjustable parameters! We shall find in general that the edges, corners, and hyper-edges of the model manifold form emergent, simpler models. In Exercise **??**, we shall use noise to sample the edges of the model manifold.

Using three predictions is not an exhaustive study for a complex model. Can we create a 3D view of the entire behavior? In Fig. **??**, we used principal component analysis to rotate our 5000 dimensional stock price information so that the most important few directions could be separated out and viewed. Let us apply principal component analysis to our data. There are packaged routines you can use for this.

(d) *Test your implementation of PCA. Generate random trajectories $y(t)$ for pairs of θs as in part (a), but now for 20 timepoints $y(t)$ evenly spaced with t between zero and ten. Find the first three principal components from these trajectories. You should get a similar manifold (perhaps flipped), except rotated so that the longest axis is along the first component and the narrowest axis is along the third component.*

(e) *Now generate a random set of trajectories with $N = 7$ for $t \in (0, 10)$, and plot the first three principal components. Do they appear to be thinning by roughly a constant factor for each new component? Plot the next three components. Does the manifold continue to get thinner?*

(e) *Now generate a random set of trajectories with $N = 7$ for $t \in (0, 10)$, and plot the first three principal components. Do they appear to be thinning by roughly a constant factor for each new component? Plot the next three components. Does the manifold continue to get thinner?*

The surface swept out by $y_{\boldsymbol{\theta}}(t)$ in the space of trajectories is the *model manifold*. You have found that it forms a *hyperribbon* – a geometrical object that is longer than it is wide, wider than it is thick, and so on for as many perpendicular directions as there are parameters. In practice, most multiparameter models share this behavior [**?**, **?**, **?**], and for NLLS models with certain smoothness conditions this hyperribbon behavior can be rigorously proven [**?**, **?**]. And, just as the edges and corners of our exponential decay model correspond to models with fewer exponentials, so too the hyper-edges of the hyperribbons for models in these other fields give rapidly converging approximate models for systems with complex microscopic laws.

N5.12 **Monomial hyperribbons.**[18] (Statistics) ③

We saw in Exercise N5.5 that the monomial coefficients for polynomial fits to data are ill-determined, and have sloppy eigenvalues for their Hessian. While linear fits with unconstrained coefficients have an unbounded model manifold (an infinite hyperplane, so not a hyperribbon), they allow arbitrarily large gradients in the resulting fit, which

[18] This exercise embodies the results of Quinn, Wilber, et al. [**?**].

are not usually expected in practice and often suppressed by nonlinearities in realistic models (where parameters can often go to infinity in ways that keep the predictions bounded).

Here we consider the model manifold for polynomials $y_{\boldsymbol{\theta}}(x) = \sum_{\alpha=0}^{N-1} \theta_\alpha x^\alpha$ with bounds on the parameters θ_α. The Jacobian

$$J_{m\alpha} = \left.\frac{\partial y_{\boldsymbol{\theta}}}{\partial \theta_\alpha}\right|_{x_m} \tag{N5.33}$$

can be viewed as mapping small vectors $\boldsymbol{\delta}$ in parameter space onto vectors $\Delta y(\mathbf{x})$ in prediction space, where $\mathbf{x} = \{x_1, \ldots, x_M\}$ are the locations of the data points being fit. That is, $y_{\boldsymbol{\theta}+\boldsymbol{\delta}}(x_m) = y_{\boldsymbol{\theta}}(x_m) + \sum_\alpha J_{m\alpha}\delta_\alpha$ (see Fig. **??**).

(a) *Show (or note) that* $J_{m\alpha} = x_m^\alpha$.

Thus

$$J = \begin{pmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^{N-1} \\ 1 & x_2 & x_2^2 & \cdots & x_2^{N-1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_M & x_M^2 & \cdots & x_M^{N-1} \end{pmatrix} \tag{N5.34}$$

is the famous Vandermonde matrix. If $M = N$, this matrix is square, and its determinant is the ratio of the volume of the allowed parameters $\boldsymbol{\theta}$ and the volume of the resulting model manifold. The famous result is that this determinant is given by $\det(J) = \prod_{1\le i<j\le N}(x_j - x_i)$. This can be seen by observing that $\det J$ obeys the rule that swapping rows of J swaps the sign of the determinant, that the determinant has the correct net degree in the x_i (the same degree $N(N-1)/2$ as the product of the diagonal entries), and then checking for the overall multiplicative constant.

Let us consider the constrained system where the monomial coefficients lie in a sphere

$$\sum_\alpha \theta_\alpha^2 < N, \tag{N5.35}$$

and consider fits over the unit interval $x \in (0, 1)$. Each θ on average then has a variance of order one, and we are considering the behavior over a distance of order one. This bound corresponds to controlling the derivatives of the function at zero, since

$$\theta_\alpha = y_{\boldsymbol{\theta}}^{(\alpha)}(0)/\alpha! = (1/\alpha!)\, \partial^\alpha y_{\boldsymbol{\theta}}/\partial x^\alpha|_{x=0} = R^\alpha \tag{N5.36}$$

where R would correspond to a radius of convergence of the function according to a ratio test. (Our assumption in eqn N5.30 that the monomial coefficients lie in a sphere can be made to work by rescaling the length until $R = 1$.) Our general theorem for nonlinear least-squares models [**?**] demands a stronger constraint on the functions $f(x)$ – that the sum of the squares of the mth derivatives be less than N *for every* x in the interval. They also use the fact that polynomials have the biggest range of predictions given the Taylor series bounds, so your calculation is reproducing much of the qualitative physics of the rigorous proof.

If the typical distance between the points x_i is Δx, then the determinant $\det J$ is $\sim (\Delta x)^{N(N-1)/2}$, which becomes really, really tiny as N gets large and the minimal spacing L/M gets small. As soon as the number of data points per radius of convergence becomes larger than two, the volume of the model manifold gets progressively smaller.

(b) *Taking $M = 6$ equally spaced points on $[0, 1]$ and $N = 6$ parameters, numerically check that the determinant of our Vandermonde matrix J is tiny.*

Is this volume small because the the predictions are squeezed into a hyperribbon? If the widths along the nth direction scale as $w_n = (\Delta x)^n$, then this would work, since $\prod_{n=1}^{N} w_n = (\Delta x)^{\sum_{n=1}^{N} n} = (\Delta x)^{N(N-1)/2}$. But usually the number of predictions M is larger than the number of parameters N, so J isn't a square matrix. What mathematical operation gives us the shape of the image of the unit sphere? Singular value decompositions is a powerful generalization of eigenvector decomposition, and precisely serves this purpose.

Singular value decomposition is not well studied in physics, where we usually care about square matrices that are symmetric or Hermitian. See the excellent Wikipedia article [**?**] on SVD. The theory says that any matrix of real numbers can be decomposed into a product of three matrices:

$$\begin{aligned} J &= U\Sigma V^T \\ J_{i\alpha} &= U_{ij}\Sigma_{j\beta}(V^T)_{\beta\alpha}. \end{aligned} \tag{N5.37}$$

Here U is $M \times M$, V is $N \times N$, and Σ is $M \times N$ and diagonal (until the diagonal hits one of the far sides of the rectangle). Here the columns of U and V (and hence the rows of V^T) are an orthonomal basis for the behavior space and the parameter space, and are called the left-singular vectors and the right-singular vectors of J, respectively. (This makes $U^T U$ and $V^T V$ the $M \times M$ and $N \times N$ identity matrices: they are unitary.) Assuming $M > N$, the first N basis vectors of U span the tangent to the model manifold. The αth right singular vectors of V maps onto the αth left singular vector of U after being stretched an amount given by $\sigma_{\alpha\alpha}$.

(c) *Taking $M = 11$ equally spaced points and $N = 6$ parameters, dig up the appropriate SVD routine and find U, Σ, and V for our our Vandermonde Jacobian J.* The left singular vectors of U are unit vectors on our sphere in parameter space. In behavior space, our sphere becomes an ellipsoid, with axes along the right singular vectors. *By how much are the unit axes of our sphere of parameters being squashed in behavior space? Is our hyperellipsoid model manifold roughly thinner by a constant factor for each new parameter?*

Finally, let us connect the skewness of J and its singular values $\Sigma_{\alpha\alpha}$ with the Hessian $\mathcal{H}_{\alpha\beta} = g_{\alpha\beta} = (J^T J)_{\alpha\beta}$.

(d) *Show analytically using the singular value decomposition that the eigenvalues of $\mathcal{H}$ are the squares of the singular values of J. What are the eigenvectors, in terms of the left and right singular vectors?*

N5.13 **First-digit law and priors.** (Statistics) ⓟ

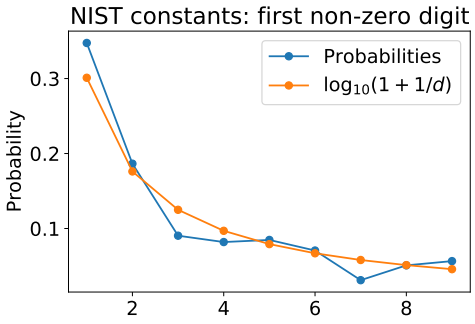


Fig. N5.7 Fraction of first digits for 354 fundamental physical constants. (2019 CODATA internationally recommended values [1]).

Bayesian statistics, like statistical mechanics, incorporates known experimental results into a probabilistic prediction for the behavior of the system in the future (see Exercise N5.3). In statistical mechanics, if we only know the energy of a system then Liouville's theorem tells us that all points in the energy shell are equally likely *a priori*. In Bayesian statistics, they have no theorem like Liouville's, so they need to assume a *prior*. For example, if you want to estimate a time constant τ for a chemical reaction (which can range from nanoseconds to years), you might want a prior $P_\tau(\tau)$ that gives equal weight to each decade: finding τ in the range $(10^{-9}, 10^{-8})$ seconds is equally plausible as finding τ in the range $(10^5, 10^6)$ seconds.

Show that $P_\tau(\tau) \propto 1/\tau$ *has this reasonable property. Show that this choice also makes the decay rates* $\Gamma = 1/\tau$ *have this same nice property:* $P_\Gamma(\Gamma) \propto 1/\Gamma$. (Hint: If τ lies in a small range $\Delta\tau$, then Γ will lie in a corresponding small range $\Delta\Gamma$, so $P_\Gamma(\Gamma)|\Delta\Gamma| = P_\tau(\tau)|\Delta\tau|$.) *Show that this distribution predicts that the first non-zero decimal digit* d *of* τ *will have probability* $\log_{10}(1+1/d)$ *(Fig. N5.7).* (Hint: Do it assuming τ lies in one decade first.) *Show your steps.* (Note: Feel free to consult the extensive discussions on the Web.)

Simon Newcomb, using a book of logarithms in 1881[19] discovered this by noticing that the pages in the beginning (1.000001, 1.000002, ...) were dirtier than the ones at the end (9.000001, 9.000002, ...). Frank Benford fleshed this out in 1938, showing that areas of rivers, molecular weights of compounds, and physical constants like the proton mass, Planck length, and Avogadro's constant (Fig. N5.7) also obey this law.

N5.14 **Bayesian priors.**[20] (Statistics) ③

In this exercise, we shall explore an analogy between statistical mechanics and Bayesian statistics. As in Exercise N5.2, we consider the problem of fitting a Gaussian probability

[19] Before calculators, people used printed books of logarithms, which allow one to multiply and divide quickly.

[20] This exercise was developed in collaboration with Colin Clement.

distribution to a collection of measurements. (See also Exercise N5.7 for an information-geometry analysis of this same problem.)

Consider the population the heights of women in the United States. Several websites quote a mean height of $\mu_0 = 162$ cm for US women, but neglect to mention the variance. We will assume $\mu = \mu_0$ is known, and we would like to estimate the probability distribution of the unknown standard deviation σ, given a single uncorrelated sample of N women. We know

$$\begin{aligned} P(\{x_n\}|\sigma) &= \prod_{n=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-(x_n-\mu_0)^2/(2\sigma^2)} \\ &= \frac{\mathrm{e}^{-\sum_{n=1}^{N}(x_n-\mu_0)^2/2\sigma^2}}{\sqrt{2\pi\sigma^2}^N} \\ &= (2\pi\sigma^2)^{-N/2} \exp\left(-S_N/2\sigma^2\right), \end{aligned} \tag{N5.38}$$

where the value $S_N = \sum_{n=1}^{N}(x_n - \mu_0)^2$ provides sufficient statistics for estimating σ.

In statistical mechanics, we care not only about the average behavior, but the *distribution* of behaviors. If our sample size is small, we should care not only about being correct on average, but also what the distribution will be of the true answers given the data we have. In our case, given the model $P(\{x_n\}|\sigma)$ with unknown parameter $(\boldsymbol{\theta} = (\sigma))$, and knowing only one sample of N data points $\{x_n\}$, what is the probability that the standard deviation of the unknown distribution is in the range $(\sigma, \sigma + \Delta)$?

In Bayesian statistics, we estimate the probability of a given set of parameters $\boldsymbol{\theta}$ given data $\mathbf{d}$ by using Bayes' theorem (see Exercise N5.3):

$$P(\boldsymbol{\theta}|\mathbf{d}) = P(\mathbf{d}|\boldsymbol{\theta})P(\boldsymbol{\theta})/P(\mathbf{d}). \tag{N5.39}$$

with $P(\{x_n\}|\sigma)$ from eqn N5.33. Here the probability of the data, $P(\mathbf{d})$, is independent of the parameters and basically acts to normalize $P(\boldsymbol{\theta}|\mathbf{d})$ to one. The probability density $P(\boldsymbol{\theta})$ is called the *prior*.

There is a close relationship between Bayesian statistics and statistical mechanics. The unknown parameters are analogous to the degrees of freedom in a physical system (say, momenta and positions of the particles). The probability density $P(\boldsymbol{\theta}|\mathbf{d})$ is analogous to the Boltzmann factor $\exp(-\mathcal{H}/k_BT)/Z$ of Chapter 6. The data $\mathbf{d}$ is analogous to the known external conditions (energy, volume, pressure, ...). Statisticians do Monte Carlo in parameter space (stochastic Bayesian analysis [**?**]) using the same techniques we discuss in Chapter 8.

But what is the prior $P(\boldsymbol{\theta})$? It represents knowledge you had about the parameters before the data is taken, or perhaps about how parameters *should* be distributed, if no measurements have yet been taken. In the statistical mechanics of classical particles (Chapter 3), our presumption about the relative probability of different positions and momenta is given by Liouville's theorem—a uniform prior, weighting all regions of

phase space equally. (It is only after we know the temperature or the energy that high momenta become less probable than low momenta.)

Uniform priors in Bayesian statistics *seem* unbiased. We shall compare several priors of the form $P_\alpha(\sigma) \sim \sigma^\alpha$.

There are three values for α of particular interest.

- $\alpha = 0$, the *uniform prior* for σ where every interval $(\sigma, \sigma + \Delta\sigma)$ is equally likely.
- A value for α, where every interval $(v, v + \Delta v)$ in the variance σ^2 is equally likely (uniform prior for σ^2).
- Jeffrey's prior $P(\sigma) = 1/\sigma$, where every *fractional* change $(\sigma, (1 + \Delta)\sigma)$ is equally likely.

Suppose three competing investigators took each took a single sample of women, with $N = 4$, $N = 40$, and $N = 400$, from a population with known mean μ_0. Suppose for simplicity that in each case their sample gave the population average[21] $S_N = N\sigma^2_{\rm pop}$.

(a) *Plot* $P_0(\sigma|S_N)$ *versus* $\sigma/\sigma_{\rm pop}$ *for these three samples* $N = 4$, 40, *and* 400, *assuming uniform prior* $\alpha = 0$ *for* σ *and using* $S_N = N\sigma^2_{\rm pop}$ The normalization ($P(\mathbf{d})$ in eqn N5.34) can be computed either numerically, or analytically in terms of $\Gamma(z) = \int_0^\infty x^{z-1}\exp(-x){\rm d}x$. *How does the maximum likelihood* $\sigma_{\rm ML}$, *where* $P_0(\sigma_{\rm ML})$ *is maximum, vary with* N*? Is it biased, compared to the naive estimate* $\sigma_{\rm pop}$*? Finally, explain why the curve appears so asymmetric for small* N*. Is the average* σ *for this probability distribution biased? In what direction?* (Hint: Is it more likely for a narrow Gaussian to give a widely distributed sample of four points, or for a wide Gaussian to happen to give a tight cluster of four points?)

So the bias in statistical estimates depends on whether you are interested in the mean (average) or the mode (maximum likelihood). From a Bayesian perspective, choosing any single number to represent the probability distribution of the quantity of interest is perhaps misguided.

Note that the bias $\langle X\rangle_{\rm samp}$ we found in Exercise N5.2, eqn N5.4 is quite different than the bias $\langle X\rangle_{\rm BayesAv}$ we discuss here. There we compared height variations over repeated samples of N women; here we use a single sample of N heights and average over the underlying true distributions that could have produced the data.

As we mentioned earlier, uniform priors *seem* unbiased. But a prior uniform in the standard deviation σ is not uniform in the variance σ^2!

(b) *Consider a uniform prior in the variance* $v = \sigma^2$, *so* $P(v)$ *is constant. What is* $P(\sigma)$*?* (Hint: The probability of being in corresponding intervals must agree, so $P(\sigma){\rm d}\sigma = P(v){\rm d}v$.) *What is* α *for a uniform prior on the variance? Calculate the mode and the average for our three samples* $S_N = N\sigma^2_{\rm pop}$, *(*$N = 4$, 40, *and* 400, $\mu = \mu_0$*) for a uniform prior on the variance. Compare with the naive estimate.*

[21] The standard deviation of women's heights in the US turns out to be about $\sigma_{\rm pop} = 6.9$ cm.

Unmeasured rates of biochemical reactions are examples of parameters that are often uncertain over many orders of magnitude. Surely our prior expectation that the rate is in the range $\Gamma \in (10^{-3}, 10^{-4})$ should not be a million times smaller than the rate is in the range $\Gamma \in (10^{3}, 10^{4})$ (as a uniform prior would suggest). Now consider using instead the time-scale $\tau = 1/\Gamma$ as the parameter—a uniform prior in τ would weight the two intervals differently by a factor of a million *in the opposite direction*. Jeffrey's prior, uniform in the logarithm, fixes this problem.

(c) *Consider a uniform prior in the log of the width* $\log(\sigma)$. *Show that* $P(\sigma) \propto 1/\sigma$, *so* $\alpha = -1$. *Check that the prior* $P_\alpha(\sigma)$ *integrated over a range* $(\sigma_0, 10\sigma_0)$ *is indeed constant, independent of* σ_0. *Calculate the mode and the average for our three samples* $S_N = N\sigma^2_{\rm pop}$ *(*$N = 4$, *40, and* 400, $\mu = \mu_0$*) for Jeffrey's prior.*

N5.15 **Hellinger and the FIM.** (Information geometry, Statistics) @

What is the shape of the space of probability distributions?

First, the space of probability distributions is often very high dimensional. The prediction is sometimes discrete: if there are 100 spins in an 10×10 Ising model, there are 2^100 probabilities $\rho(\mathbf{s})$ which sum to one. It is often continuous: if one fits a Gaussian $\rho_{\overline{x},\sigma}(x)$ to data, the model manifold is a two-dimensional surface in an infinite-dimensional space of possible probability functions.

Second, even if the prediction is discrete, it is not natural to treat the prediction as a vector, because the natural distance between two predictions is not the sum of squares of the differences between the individual probabilities. We can see this by considering how difficult it is to measure a small change in probability of one of the predictions.

(a) *Consider flipping a coin to measure the small probability* ρ *that it lands on its edge. After* $F \gg 1/\rho$ *flips, what is the error in your estimate of* ρ, *to lowest order in* ρ? *How many flips do you need to estimate* ρ *to an accuracy* ϵ? *Is it equally easy to measure rare events and common events to an absolute error of* ϵ? *Which is harder?*

The probability $\rho(\mathbf{x})$ is normalized, $\sum_{\mathbf{x}} \rho(\mathbf{x}) = 1$. Taking the square root of each entry gives a point on the unit hypersphere (in the positive "octant", with all components greater than zero). Let us check that it is uniformly challenging to estimate the square roots of the components with differing probabilities.

(b) *In the above experiment, what is the error in your estimate of* $\sqrt{\rho}$, *to lowest order in* ρ? *Is it equally easy to measure the square root of the probability of rare events and common events?*

So, we conjecture that a natural measure of distance on the sphere of $\sqrt{\rho}$ is the Euclidean distance in the embedding space

$$|\rho_1 - \rho_2|_{\rm sphere} = \sqrt{(\sqrt{\rho_1} - \sqrt{\rho_2})^2} = \sqrt{\sum_{\mathbf{x}} (\sqrt{\rho_1(\mathbf{x})} - \sqrt{\rho_2(\mathbf{x})})^2}. \tag{N5.40}$$

This is, up to a constant, the *Hellinger distance*, sometimes called the Hellinger divergence (because all the other measures for separations between probabilities are not proper distances.)

Suppose we now have a model $\rho_{\boldsymbol{\theta}}(\mathbf{x})$ depending on parameters $\boldsymbol{\theta}$. The metric tensor $g_{\alpha\beta}$ is defined to be the dependence of the distance on small changes in parameters, $\boldsymbol{\theta}' = \boldsymbol{\theta} + \epsilon\boldsymbol{\delta}$. The squared distance to quadratic order should be

$$|\rho_{\boldsymbol{\theta}+\epsilon\boldsymbol{\delta}} - \rho_{\boldsymbol{\theta}}|^2 = \epsilon^2 g_{\alpha\beta}\delta_\alpha\delta_\beta. \tag{N5.41}$$

(c) *Show that the distance on the sphere implies that*

$$g^{\text{sphere}}_{\alpha\beta} = \int \mathrm{d}\mathbf{x}\tfrac{1}{4}\rho(\partial_\alpha \log\rho)(\partial_\beta\log\rho) = \tfrac{1}{4}\langle(\partial_\alpha\log\rho)(\partial_\beta\log\rho)\rangle_{\mathbf{x}}. \tag{N5.42}$$

But we have been told that the natural distance in probability space is given by the Fisher Information Matrix (FIM). Equation N5.18 tells us

$$\begin{aligned} g^{\text{FIM}}_{\alpha\beta} &= -\left\langle \frac{\partial^2 \log\rho(\mathbf{x})}{\partial\theta_\alpha\partial\theta_\beta}\right\rangle \\ &= -\int \mathrm{d}\mathbf{x}\rho(\mathbf{x})\frac{\partial^2\log\rho(\mathbf{x})}{\partial\theta_\alpha\partial\theta_\beta} \end{aligned} \tag{N5.43}$$

Are these different? The version g^{FIM} in eqn N5.18 is more convenient for us, but it equals $4g^{\text{sphere}}$, which is also commonly used.

(d) *Show this.*

The Hellinger distance perhaps should have been defined to be twice as big (the distance on a sphere of radius two), so the squared distance for nearby points would agree with the FIM. Or, even better, the FIM should have been defined to be a factor of four smaller. Instead, the Hellinger distance is sometimes the sphere distance, and sometimes the squared distance divided by square-root of two (so with a metric tensor one eighth that of the FIM).

N5.16 **Bhattacharyya and the inPCA embedding.**[22] (Information geometry, Statistics) @

In Exercise N5.15, we saw that the space of probability distributions is naturally viewed as a hypersphere, where $\sqrt{\rho(\mathbf{x})}$ is viewed as a vector with components labeled by $\mathbf{x}$. We also saw that the distances between points on this sphere gave the natural Fisher information metric for local distances between probability distributions. A probabilistic model like the Ising model then has a natural, isometric embedding as a surface on this sphere. Unfortunately, we also saw that this embedding is intrinsically high dimensional

[22] This exercise is based on Quinn et al. [**?**].

for interesting models: there is no way to visualize a large Ising model by projecting it into a small dimensional space.

This is in sharp contrast with the nonlinear least squares models we have studied in the rest of this book! There the predictions lie on hyperribbons in the behavior space, naturally forming an isometric, low-dimensional embedding that we can visualize using, for example, principal component analysis (as in Exercise N5.11). Can we find a way to do this for probabilistic models?

The key problem is that systems like large Ising models, or experiments measuring the cosmic microwave background radiation, have too much precise data. Quinn [156] pointed out that, once parameters have shifted even a tiny amount, the behavior is obviously distinguishable with simple measurements: the probability of a snapshot of one Ising model (or one Universe's cosmic microwave background radiation) being reproduced at the other temperature and pressure (or Hubble constant and baryon density) becomes near zero.

Consider the sphere distance (eqn N5.35), whose square

$$d^2_{\text{sphere}}(\rho_1,\rho_2) = \sum_{\mathbf{x}} \left(\sqrt{\rho_1(\mathbf{x})} - \sqrt{\rho_2(\mathbf{x})}\right)^2 = 2(1 - \sqrt{\rho_1}\cdot\sqrt{\rho_2}), \tag{N5.44}$$

can be written in terms of a dot product between points on the sqrt probability sphere. (The Hellinger distance and the natural FIM local distances agree with the sphere distance up to constants.)

(a) *Note that the two forms of* d^2_{sphere} *in eqn N5.39 are equal, because the densities are normalized. Suppose there is no overlap between two Ising model ensembles. That is, for all spin configurations* $\mathbf{s}$ *of an Ising model with significant probability* $\rho_{T,H}(\mathbf{s}) > 0$*, the probability* $\rho_{T',H'} = 0$*. What is the sphere distance between the two?*

(b) *Suppose there are many ensembles at different temperatures and fields, with mutually orthogonal probability distributions. What geometrical figure will they form on the hypersphere? Can this be viewed as a hyperribbon?*

Quinn et al. used the *replica trick* to take the limit of zero data. Suppose we take n snapshots of our Ising model, $\{\mathbf{s}_1,\ldots,\mathbf{s}_n\}$. We can view this as one snapshot of n uncoupled replicas of the Ising model, all with the same parameters, with a probability distribution

$$\rho^{[n]}_{T,H}(\mathbf{s}_1,\ldots,\mathbf{s}_n) = \rho_{T,H}(\mathbf{s}_1)\rho_{T,H}(\mathbf{s}_2)\ldots\rho_{T,H}(\mathbf{s}_n). \tag{N5.45}$$

(c) *Note that the replicated dot product* $\sqrt{\rho_1^{[n]}}\cdot\sqrt{\rho_2^{[n]}} = \sum_{\mathbf{s}_1}\cdots\sum_{\mathbf{s}_n}\prod_{i=1}^n \sqrt{\rho_1(\mathbf{s}_i)}\sqrt{\rho_2(\mathbf{s}_i)}$*. Show that it can be written as* $\left(\sum_{\mathbf{s}}\sqrt{\rho_1(\mathbf{s})}\sqrt{\rho_2(\mathbf{s})}\right)^n = (\sqrt{\rho_1}\cdot\sqrt{\rho_2})^n$*. Have we made the orthogonality problem better or worse?*

Thus you have shown that the distance per replica

$$(d^{[n]}_{\text{replicated}})^2(\rho_1,\rho_2) = d^2_{\text{replicated}}(\rho_1^{[n]},\rho_2^{[n]})/n = 2(1-(\sqrt{\rho_1}\cdot\sqrt{\rho_2})^n)/n. \tag{N5.46}$$

The replica trick is to boldly use the formula $\lim_{n\to 0}(x^n-1)/n = \log(x)$. Especially regarding the replicated partition function of disordered systems like spin glasses, using this formula is mathematically dubious but physically extremely useful. Here it yields the Bhattacharyya divergence between the two probability distributions.

(d) *Show that* $\lim_{n\to 0} d^{[n]}(\rho_1,\rho_2)^2 = -2\log(\sqrt{\rho_1}\cdot\sqrt{\rho_2})$. Up to the multiplicative factor separating our sphere distance from that of the FIM, you can check on your favorite Web search or AI that this yields the Bhattacharyya divergence.

Using this metric, along with a version of principal component analysis called MDS[23] we succeed in averting the curse of dimensionality, as shown in Figs **??** and **??**.

N5.17 **Kullback–Leibler and isKLe.**[24] (Information geometry, Statistics) @

To visualize the model manifold in information geometry, we want to preserve some measure of the distance between the predictions. For models whose output is a probability defined over a large number of states, it is usually not possible to directly calculate that distance. Of course, a direct measure of the magnetization of an N-spin Ising model would also demand an infeasible 2^N computations of the energy: perhaps one could develop a Monte-Carlo method for calculating the Bhattacharyya divergence of inPCA (Exercise N5.16) or the Hellinger distance (Exercise N5.15).

In Exercises N5.9 and N5.10 we found a deep relation between the Fisher Information Metric (measuring local distances between probability distributions) and second derivatives of the free energy like the specific heat and thermal expansion coefficient. Here we discover a deep relation between a global measure of distance between distributions and first derivatives of the free energy. You shall show that the Kullback–Leibler divergence between two distributions, once symmetrized, can be written in terms of the magnetization and the energy for Ising models (uncovered by Teoh et al. [**?**], see Fig. **??**).

The Ising model is a particular example of what probabilists call an *exponential family*. A probability distribution $\rho_{\boldsymbol{\theta}}(\mathbf{x})$ is an exponential family if it can be written in the form

$$\rho_{\boldsymbol{\theta}}(\mathbf{x}) = f(\mathbf{x})g(\boldsymbol{\theta})\exp\left(\sum_\gamma \eta_\gamma(\boldsymbol{\theta})T_\gamma(\mathbf{x})\right). \tag{N5.47}$$

The T_γ are called the *sufficient statistics* for the distribution; they hold all the information about the configuration $\mathbf{x}$ needed to determine the probability; The η_γ are called the *natural parameters* (see Exercise N5.2).

Note that eqn N5.42 has a form similar to a Boltzmann distribution: for the Ising model $\rho_{T,H}(\mathbf{s}) = \exp\left(-(J\sum_{<ij>} s_i s_j - H\sum_i s_i)/k_B T\right)/Z(T,H)$.

[23] We rederived multidimensional scaling by taking the zero-replica limit of principal component analysis. It can be used for any distance measure, not just the Bhattacharyya divergence. So, for example, our isKLe embeddings [**?**] use MDS with the symmetrized Kullback–Leibler embedding (see Exercise N5.17 and Figs **??**, **??**, and **??**).

[24] This exercise is based on Teoh et al. [**?**].

(a) *What are the two natural parameters and two sufficient statistics for our Ising model? What is $g(\boldsymbol{\theta})$? Show that $f(\mathbf{x}) = 1$.* It will be convenient to work in these natural parameters (as we did in Exercises N5.9 and N5.10). Let us call $\boldsymbol{\eta} = \{-\beta, h\}$ and $\mathbf{T}(\mathbf{s}) = \{e(\mathbf{s}), m(\mathbf{s})\}$.

The Kullback–Leibler divergence between two probability distributions $\rho(\mathbf{x})$ and $\sigma(\mathbf{x})$ is

$$KL(\rho||\sigma) = \sum_{\mathbf{x}} \rho(\mathbf{x}) \log\left(\rho(\mathbf{x})/\sigma(\mathbf{x})\right). \tag{N5.48}$$

Notice that the formula reminds us of the formula for the Shannon entropy $S = -k\sum(\rho\log(\rho))$. It has many physical interpretations and uses in statistics (expected surprise from using the wrong model, extra bits using the wrong coding algorithm, and so on.)

(b) *Calculate the divergence $KL(\rho_1(\beta_1, h_1)||\rho_2(\beta_2, h_2))$ in terms of Z_1, Z_2, the four natural parameters, $e_1 = \langle e(\mathbf{s})\rangle_{\rho_1}$ and $m_1 = \langle m(\mathbf{s})\rangle_{\rho_1}$.*

The KL divergence is not symmetric (as a distance should be). But we can use the symmetrized KL divergence (sometimes called the Jeffrey's divergence)

$$sKL(\rho, \sigma) = KL(\rho||\sigma) + KL(\sigma||\rho) \tag{N5.49}$$

as a kind of distance.[25]

(c) *Show that the sKL divergence is $-(\beta_1 - \beta_2)(e_1 - e_2) + (h_1 - h_2)(m_1 - m_2)$.*

This is a known result, for a general exponential family. But what was not realized (to our knowledge) is that this can be used to find coordinates for the two points! Although, as Exercise N5.16 and more generally for multidimensional scaling embeddings, some of the coordinates are "timelike", with negative squared contributions to the distance. The space-like coordinates are the averages[26] of the natural parameters and their statistics, $(-\beta + e)/2$ and $(h + m)/2$, and the time-like coordinates are half the differences $(-\beta - e)/2$ and $(h - m)/2$,

$$\begin{aligned} T_\beta^\pm &= \tfrac{1}{2}(-\beta \pm e) \\ T_h^\pm &= \tfrac{1}{2}(h \pm m) \end{aligned} \tag{N5.50}$$

(d) *Show that the sKL divergence between ρ_1 and ρ_2 indeed is given by the sum of the squares of the space-like coordinate differences minus the sum of the squares of the time-like coordinate differences.*

Note the relation between our model manifold embedding in a Minkowski space and the *model graph* formed by plotting $(e(\beta, h), m(\beta, h))$ in four dimensions. Rotating each

[25] It is zero if $\rho = \sigma$ and symmetric, and one can check that it agrees with the FIM when the deviation between ρ and σ goes to zero. It does not satisfy the triangle inequality, but this is precisely why it is valuable to us, since the triangle inequality dooms low dimensional embeddings (see Exercise N5.15).

[26] Note that the natural parameter for the energy is $-\beta$. It may seem weird to add and subtract fields and magnetizations! One can find other coordinate sets (e.g., $(\lambda h \pm 1/\lambda m)/2$) which can be used to fix the units. These correspond to Lorentz boosts in the Minkowski-like embedding space.

conjugate pair (β, e) and (h, m) of the model graph by 45 degrees generates the isometric embedding. Note that space-time rotations are *not* isometric, however. Indeed, the fully magetized states in Fig. **??** at low temperatures and large fields are at zero distance from one another (they are the same state), but arrange in a 45 degree diagonal with light-like separations.

The explicit formulas make generating model manifolds in most statistical mechanics problems completely straightforward. Fig. **??**, showing the two-dimensional Ising model, uses a standard Monte-Carlo evaluation of the field and temperature-dependent energy and magnetization. (The Wolff algorithm can be generalized to work in an external field [98], making simulations like these fast even near the critical point.) One imagines using simulations like these to

Figs **??** and **??** show other illustrations of applications of the isKLe embedding.

N5.18 **Distances in probability space.**[27] (Statistics, Mathematics, Information geometry) ③

In statistical mechanics we usually study the behavior expected given the experimental parameters. Statistics is often concerned with estimating how well one can deduce the parameters (like temperature and pressure, or the increased risk of death from smoking) given a sample of the ensemble. Here we shall explore ways of measuring distance or distinguishability between distant probability distributions.

Exercise N5.7 introduces four problems (loaded dice, statistical mechanics, the height distribution of women, and least-squares fits to data), each of which have parameters $\boldsymbol{\theta}$ which predict an ensemble probability distribution $P(\mathbf{x}|\boldsymbol{\theta})$ for data $\mathbf{x}$ (die rolls, particle positions and momenta, heights, ...). In the case of least-squares models (eqn N5.17) where the probability is given by a vector $x_i = y_i(\boldsymbol{\theta}) \pm \sigma$, we found that the distance between the predictions of two parameter sets $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ was naturally given by $|\mathbf{y}(\boldsymbol{\theta})/\sigma - \mathbf{y}(\boldsymbol{\phi})/\sigma|$. We want to generalize this formula—to find ways of measuring distances between probability distributions given by arbitrary kinds of models.

Exercise N5.7 also introduced the Fisher information metric (FIM) in eqn N5.18:

$$g_{\mu\nu}(\boldsymbol{\theta}) = -\left\langle \frac{\partial^2 \log(P(\mathbf{x}))}{\partial\theta_\alpha \partial\theta_\beta} \right\rangle_{\mathbf{x}} \tag{N5.51}$$

which gives the distance between probability distributions for nearby sets of parameters

$$d^2(P(\boldsymbol{\theta}), P(\boldsymbol{\theta} + \epsilon\boldsymbol{\Delta})) = \epsilon^2 \sum_{\mu\nu} \Delta_\mu g_{\mu\nu} \Delta_\nu. \tag{N5.52}$$

Finally, it argued that the distance defined by the FIM is related to how distinguishable the two nearby ensembles are—how well we can deduce the parameters. Indeed, we found that to linear order the FIM is the inverse of the covariance matrix describing

[27] This exercise was developed in collaboration with Katherine Quinn.

the fluctuations in estimated parameters, and that the Cramér–Rao bound shows that this relationship between the FIM and distinguishability works even beyond the linear regime.

There are several measures in common use, of which we will describe three—the Hellinger distance, the Bhattacharyya "distance", and the Kullback–Liebler divergence. Each has its uses. The Hellinger distance becomes less and less useful as the amount of information about the parameters becomes large. The Kullback–Liebler divergence is not symmetric, but one can symmetrize it by averaging. It and the Bhattacharyya distance nicely generalize the least-squares metric to arbitrary models, but they violate the triangle inequality and embed the manifold of predictions into a space with Minkowski-style time-like directions [156].

Let us review the properties that we ordinarily demand from a distance between points P and Q.

- We expect it to be positive, $d(P,Q) \geq 0$, with $d(P,Q) = 0$ only if $P = Q$.
- We expect it to be symmetric, so $d(P,Q) = d(Q,P)$.
- We expect it to satisfy the *triangle inequality*, $d(P,Q) \leq d(P,R) + d(R,Q)$—the two short sides of a triangle must extend at total distance enough to reach the third side.
- We want it to become large when the points P and Q are extremely different.

All of these properties are satisfied by the least-squares distance of Exercise N5.7, because the distances between points on the surface of model predictions is the Euclidean distance between the predictions in data space.

Our first measure, the Hellinger distance at first seems ideal. It defines a *dot product* between probability distributions P and Q. Consider the discrete gambler's distribution, giving the probabilities $\mathbf{P} = \{P_j\}$ for die roll j. The normalization $\sum P_j = 1$ makes $\{\sqrt{P_j}\}$ a unit vector in six dimensions, so we define a dot product $P \cdot Q = \sum_{j=1}^{6} \sqrt{P_j}\sqrt{Q_j} = \int \mathrm{d}\mathbf{x} \sqrt{P(\mathbf{x})}\sqrt{Q(\mathbf{x})}$. The Hellinger distance is then given by the squared distance between points on the unit sphere:[28]

$$\begin{aligned} d^2_{\text{Hel}}(P,Q) &= (P-Q)^2 = 2 - 2P \cdot Q \\ &= \int \mathrm{d}\mathbf{x} \left(\sqrt{P(\mathbf{x})} - \sqrt{Q(\mathbf{x})}\right)^2. \end{aligned} \tag{N5.53}$$

(a) *Argue, from the last geometrical characterization, that the Hellinger distance must be a valid distance function. Show that the Hellinger distance does reduce to the FIM for nearby distributions, up to a constant factor. Show that the Hellinger distance never gets larger than* $\sqrt{2}$. *What is the Hellinger distance between a fair die* $P_j \equiv 1/6$ *and a loaded die* $Q_j = \{1/10, 1/10, \ldots, 1/2\}$ *that favors rolling 6?*

[28] Sometimes it is given by *half* the distance between points on the unit sphere, presumably so that the maximum distance between two probability distributions becomes one, rather than $\sqrt{2}$.

The Hellinger distance is peculiar in that, as the statistical mechanics system gets large, or as one adds more experimental data to the statistics model, all pairs approach the maximum distance $\sqrt{2}$.

(b) *Our gambler keeps using the loaded die. Can the casino catch him? Let* $P_N(\mathbf{j})$ *be the probability that rolling the die* N *times gives the sequence* $\mathbf{j} = \{j_1, \ldots, j_N\}$. *Show that*

$$P_N \cdot Q_N = (P \cdot Q)^N, \tag{N5.54}$$

and hence

$$d^2_{\mathrm{Hel}}(P_N, Q_N) = 1 - (P \cdot Q)^N \tag{N5.55}$$

After $N = 100$ *rolls, how close is the Hellinger distance from its maximum value?*

From the casino's point of view, the certainty that the gambler is cheating is becoming squeezed into a tiny range of distances. (P_N and Q_N becoming increasingly orthogonal does not lead to larger and larger Hellinger distances.) In an Ising model, or a system with N particles, or a cosmic microwave background experiment with N measured areas of the sky, even tiny changes in parameters lead to orthogonal probability distributions, and hence Hellinger distances near its maximum value of one.[29]

The Hellinger overlap $(P \cdot Q)^N = \exp(N \log(P \cdot Q))$ keeps getting smaller as we take N to infinity; it is like the exponential of an extensive quantity.

Our second measure, the Bhattacharyya distance, can be derived from a limit of the Hellinger distance as the number of data points N goes to zero:

$$\begin{aligned} d^2_{\mathrm{Bhatt}}(P, Q) &= \lim_{N \to 0} \tfrac{1}{2} d^2_{\mathrm{Hel}}(P_N, Q_N)/N \\ &= -\log(P \cdot Q) \\ &= -\log\left(\sum_{\mathbf{x}} \sqrt{P(\mathbf{x})} \sqrt{Q(\mathbf{x})} \right). \end{aligned} \tag{N5.56}$$

We sometimes say that we calculate the behavior of N *replicas* of the system, and then take $N \to 0$. Replica theory is useful, for example, in disordered systems, where we can average $F = -k_B T \log(Z)$ over disorder (difficult) by finding the average of Z^N over disorder (not so hard) and then taking $N \to 0$.

(d) *Derive eqn N5.51.* (Hint: $Z^N \approx \exp(N \log Z) \approx 1 + N \log Z$ for small N.)

The third distance-like measure we introduce is the *Kullback–Leibler divergence* from Q to P.

$$d_{\mathbf{KL}}(Q|P) = -\int \mathrm{d}\mathbf{x} P(\mathbf{x}) \log(Q(\mathbf{x})/P(\mathbf{x})). \tag{N5.57}$$

[29] The problem is that the manifold of predictions is being curled up onto a sphere, where the short-cut distance between two models becomes quite different from the geodesic distance within the model manifold.

(c) *Show that the Kullback–Liebler divergence is positive, zero only if $P = Q$, but is not symmetric. Show that, to quadratic order in ϵ in eqn N5.47, that the Kullback–Liebler divergence does lead to the FIM.*

The Kullback–Liebler divergence is sometimes symmetrized:

$$\begin{aligned} d_{\mathbf{sKL}}(Q,P) &= \tfrac{1}{2}(d_{\mathbf{KL}}(Q|P) + d_{\mathbf{KL}}(P|Q)) \\ &= \int \mathrm{d}\mathbf{x}(P(\mathbf{x}) - Q(\mathbf{x}))\log(P(\mathbf{x})/Q(\mathbf{x})). \end{aligned} \tag{N5.58}$$

The Bhattacharyya distance and the symmetrized Kullback–Liebler divergence share several features, both good and bad.

(d) *Show that they are* intensive *[156]—that the distance grows linearly with repeated measurements*[30] *(as for repeated rolls in part (b)). Show that they do not satisfy the triangle inequality. Show that they does satisfy the other conditions for a distance. Show, for the nonlinear least-squares model of eqn N5.17, that they equal the distance in data space between the two predictions.*

N5.19 **Can we burn information?**[31] (Mathematics, Thermodynamics, Information geometry) ④

The use of entropy to measure information content has been remarkably fruitful in computer science, communications, and even in studying the efficiency of signaling and sensing in biological systems. The Szilard engine (Exercise 5.2) was a key argument that thermodynamic entropy and information entropy could be exchanged for one another—that one could burn information. We ask here—can they be exchanged? Are information and entropy fungible?

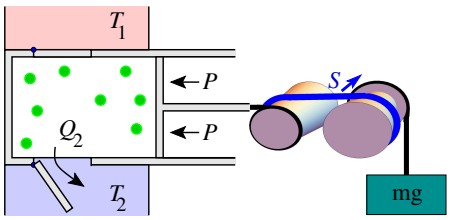


Fig. N5.8 Piston control. Machta [121] studies a piston plus a control system to extract work during expansion. To change the pressure, a continuously variable transmission, controlled by a gradient of the entropy S, connects the piston to a mass under a gravitational force. Minimizing the control cost plus the entropy cost due to fluctuations in the gear ratio lead to a minimum entropy cost for control.

[30] This also makes these measures behave nicely for large systems as in statistical mechanics, where small parameter changes lead to nearly orthogonal probability distributions.

[31] This exercise was developed in collaboration with Ben Machta, Archishman Raju, Colin Clement, and Katherine Quinn

Szilard stores a bit of information as an atom on one side of a piston, and extracts $P\mathrm{d}V$ work $k_BT\log 2$ as the piston expands—the same work needed to store a bit. Machta [121] argues that there is a fundamental bound on the entropy cost for extracting this work. He considers a system consisting of the piston plus a control mechanism to slowly decrease the pressure and extract the work, Fig. N5.8. (See Feynman's *Ratchet and pawl* discussion [62, I.46], discussing fluctuations in a similar system.)

Machta argues that this cost is given by a *path length in parameter space*. To be specific, Machta argues that to guide a system through a change in pressure from P_i to P_f should cost an entropy[32]

$$\langle \Delta S_{\mathrm{control}} \rangle = 2 \int_{P_i}^{P_f} \sqrt{g_{PP}} |\mathrm{d}P|. \tag{N5.59}$$

The metric in this space, as discussed in Exercise N5.7, is

$$g_{PP} = -\langle \partial^2 \log(\rho)/\partial P^2 \rangle, \tag{N5.60}$$

the *Fisher information metric*, giving the natural distance between two nearby probability distributions.

For example, in a Gibbs ensemble at constant pressure $P = \theta_1$, the squared distance between two nearby pressures at the same temperature is

$$d^2(\rho(\mathbf{X}|P), \rho(\mathbf{X}|P + \mathrm{d}P)) = g_{PP}(\mathrm{d}P)^2, \tag{N5.61}$$

leading directly to eqn N5.54.

Let us compute the metric g_{PP} in the coordinates for the ideal gas in a piston (Exercise N5.8), and then analyze the cost for thermodynamic control for Szilard's burning information engine in Exercise 5.2.

(a) *Using eqns N5.55 and N5.22, show that* $g_{PP} = (1 + N)/P^2$.

(b) *What is the entropy cost to expand a piston containing a single atom at constant temperature by a factor of two? What is the work done by the piston? How does this affect Szilard's argument about burning information in Exercise 5.2?*

Machta's result thus challenges Szilard's argument that information entropy and thermodynamic entropy can be exchanged. It also gives a (subextensive) cost for the Carnot cycle (see Exercise N5.9).

N5.20 Averaging over disorder.[33] ③

A two-state spin takes values $S = \pm 1$. It is in an external field h, so that its Hamiltonian is

$$\mathcal{H} = -hS. \tag{N5.62}$$

It is connected to a heat bath at temperature T.

[32] We shall follow Machta and set $k_B = 1$ in this exercise, writing it explicitly only when convenient.

[33] This problem was developed in collaboration with Stephen Thornton.

(a) *Compute its partition function Z, its Helmholtz free energy A, the entropy S, and the specific heat*[34] *c as a function of h and T. What is the entropy at $T = 0, h > 0$ and at $T = \infty$?* (The $T \to 0$ limit is tricky: a graphical solution is fine.) *Is the difference as expected from our understanding of information entropy?*

To model a system with dirt – a disordered system – one often adds a random term to the Hamiltonian (like a random field for each spin). One then averages the answer over the probability distribution of the disorder to predict the behavior of a large system. This turns out to be trickier than it seems.

Let us calculate the average properties of our spin in a random field h, averaged over a Gaussian probability distribution $\rho(h) = \exp(-h^2/2\sigma^2)/(\sqrt{2\pi}\sigma)$.

(b) *Write in integral form the average of each of the quantities $\overline{Z}$, $\overline{A}$, $\overline{S}$, and $\overline{c}$ over the probability density $\rho(h)$.* All but one of these will be infeasible to evaluate in closed form. *Evaluate the integral for $\overline{Z}$.*

In interacting systems like spin glasses, it is much easier to calculate the average of Z than the average of $\log Z$ or A. But we run into trouble.

(c) *Define $Z_a = \overline{Z}$, and calculate the corresponding quantities A_a and S_a. Show that S_a goes negative at low temperatures.*

The entropy for each disorder you calculated in part (a) never goes negative. So its average cannot be negative! We seem to be stuck with the integrals we cannot do in closed form.

(d) *Define $A_q = \overline{A}$. Argue that S_q, defined as the appropriate derivative of A_q, is equal to $\overline{S}$ from part (b).*

Let us briefly consider a simpler scenario, where h can take only the three values 0 or $\pm h_0$ (with $h_0 > 0$), each with probability 1/3.

(e) *Write A_q and A_a exactly for this case, and evaluate them in the limit $T \to 0$. Using $\overline{A} = \overline{\langle E \rangle} - T\overline{S}$, what value should you expect for the average free energy at $T = 0$? Does A_a appear to be giving unfair weights to disorder configurations with lower-energy states?*

Thus A_a gives an unfairly large weight to members of the disordered ensemble that have unusually low energy configurations. For spin glasses, A_a gives unfair weights to systems like the non-disordered Ising model, where a single spin configuration can make all the bonds happy. This leads to an unphysical ferromagnetic–like transition.

Why the choice of subscripts? When we want to freeze our dirt into a particular configuration, we quench the system quickly to a low temperature. (The blacksmith pounding the red-hot horseshoe, after they get it into shape, quenches it in a bucket of water.) A_q is the *quenched* free energy. We anneal a defective crystal by heating it up to a large temperature T_0 where its defects have enough energy to rearrange and come

[34] Section 6.1 discusses the specific heat at constant volume c_v, but the formulas are the same because here there is no volume to be fixed.

to equilibrium. $A_a = -k_B T \log(\overline{Z})$ is called the *annealed* free energy. But why does our A_a correspond to an annealed free energy, where the "defects" come to equilibrium?

(f) *Show that $Z_a(T_0)$ from part (c) at a particular temperature T_0 is the true partition function for a Hamiltonian*

$$\mathcal{H}_a = h^2 k_B T_0 / 2\sigma^2 - hS + C, \tag{N5.63}$$

where the constant $C = \frac{1}{2} k_B T_0 \log(2\pi\sigma^2)$. Thus Z_a discusses a system where h and S are both weighted according to the Boltzmann distribution (so the field fluctuates to equilibrate with the spin). In systems like spin glasses, one can calculate annealed averages because they are, in disguise, the correct partition function for an undisordered equilibrium system.

We must end with the *replica trick* that people use to bypass the infeasible integrals we get from trying to average the $\log(Z)$, as in $A_q = -k_B T \overline{\log Z}$. One can often calculate $\overline{Z^n}$, the annealed disorder average of n replicas of a system. (Again, it is feasible because it is in disguise an equilibrium physical system, whose dirt equilibrates with the spins.) We then can find the average $\overline{\log(Z)}$ and hence A_q:

(g) *Show that $\log x = \lim_{n\to 0}(x^n - 1)/n$ by writing $x^n = \exp(n \log x)$.*

We can then take the average of both sides and write $\overline{\log(Z)} = \lim_{n\to 0}(\overline{Z^n} - 1)/n$. Finding the right way of taking the limit $n \to 0$ is harder than we are suggesting. The original researchers used a "replica symmetric" method that works for many systems, and works well in spin glasses for temperatures above the glass transition. Below the glass transition, one must do something more exotic. Georgio Parisi received the Nobel Prize in Physics in 2021 for showing certain disordered systems undergo a "replica symmetry breaking" transition as the temperature is lowered, where certain correlations within the system change dramatically in the spin glass phase. These methods have been shown by Parisi and others to be powerful tools for solving models of ordinary glass, analyzing deep neural network models in machine learning, and providing the fastest algorithms for challenging "NP complete" models in computer science (see Exercise 8.15).

Bibliography

[1] (2019). The NIST reference on constants, units, and uncertainty, from CODATA internationally recommended 2018 values of the fundamental physical constants. http://physics.nist.gov/cuu/Constants/.

[2] Alemi, Alexander A. and Fischer, Ian (2018). Therml: Thermodynamics of machine learning. cite arxiv:1807.04162Comment: Presented at the ICML 2018 workshop on Theoretical Foundations and Applications of Deep Generative Models.

[3] Blander, Milton and Katz, Joseph L. (1975). Bubble nucleation in liquids. *AIChE Journal*, **21**(5), 833–848.

[4] Bowick, Mark J., Manyuhina, O. V., and Serafin, F. (2016). Shapes and singularities in triatic liquid crystal vesicles. *EPL (Europhysics Letters)*, **117(2):26001**.

[5] Brennen, C.E. (2014). *Cavitation and Bubble Dynamics*. Cavitation and Bubble Dynamics. Cambridge University Press.

[6] Carroll, John J. and Mather, Alan E. (1992). The system carbon dioxide-water and the Krichevsky–Kasarnovsky equation. *Journal of Solution Chemistry*, **21**, 607–21.

[7] Cates, Michael E. and Manoharan, Vinothan N. (2015). Celebrating soft matter's 10th anniversary: Testing the foundations of classical entropy: colloid experiments. *Soft Matter*, **11**, 6538–6546.

[8] Chen, Bryan Gin-ge, Upadhyaya, Nitin, and Vitelli, Vincenzo (2014). Nonlinear conduction via solitons in a topological mechanical insulator. *Proceedings of the National Academy of Sciences*, **111**(36), 13004–13009.

[9] Chen, Yan-Jiun, Paquette, Natalie, Machta, Benjamin B., and Sethna, James P. (2013). Universal scaling function for the two-dimensional Ising model in an external field: A pragmatic approach. http://arxiv.org/abs/1307.6899.

[10] Cline, James M. (2006, 9). Baryogenesis. In *Les Houches Summer School - Session 86: Particle Physics and Cosmology: The Fabric of Spacetime*.

[11] Davies, C. T. H., Follana, E., Gray, A., Lepage, G. P., and *et al.* (2004). High-precision lattice qcd confronts experiment. *Physical Review Letters*, **92**, 022001.

[12] Feigenbaum, M. (1979). The universal metric properties of nonlinear transformations. *Journal of Statistical Physics*, **21**, 669–706.

[13] Fisher, M. E. and Randeria, M. (1986). Location of renormalization-group fixed points. *Physical Review Letters*, **56**, 2332.

[14] Hathcock, David, Sheehy, James, Weisenberger, Casey, Ilker, Efe, and Hinczewski, Michael (2016). Noise filtering and prediction in biological signaling networks. *IEEE Transactions on Molecular, Biological and Multi-Scale Communications*, **2**(1), 16–30.

[15] III, Robert A. Kinzie. Carbon in lakes. http://www2.hawaii.edu/~kinzie/documents/470/CO2solubility.doc.

[16] Larsen, Bue Vedel (2015). 2D dry foam: Coarsening example using parallel methods. https://www.youtube.com/watch?v=DN23diqRAGY.

[17] Liarte, Danilo B., Thornton, Stephen J., Schwen, Eric, Cohen, Itai, Chowdhury, Debanjan, and Sethna, James P. (2021, March). Scaling of dynamical susceptibility at the onset of rigidity for disordered viscoelastic matter. *arXiv e-prints*, arXiv:2103.07474.

[18] MacPherson, Robert D. and Srolovitz, David J. (2007). The von Neumann relation generalized to coarsening of three-dimensional microstructures. *Nature*, **446**, 1053–5.

[19] Mao, J. and Hu, B. (1987). Corrections to scaling for period doubling. *Journal of Statistical Physics*, **46**, 111.

[20] McCoy, B. M. and Wu, T. T. (2014). *The two-dimensional Ising model,* 2nd Edition. Dover Publications.

[21] Mermin, N. D. (1979). The topological theory of defects in ordered media. *Reviews of Modern Physics*, **51**, 591–648.

[22] Mézard, M. and Montanari, A. (2009). *Information, physics, and computation.* Oxford University Press, New York.

[23] Mullins, W. W. (1963). In *Metal Surfaces: Structure, Energetics, and Kinetics*, pp. 17–66. American Society for Metals, Ohio.

[24] Ostlund, S., Rand, D., Sethna, J. P., and Siggia, E. D. (1983). Universal properties of the transition from quasi-periodicity to chaos in dissipative systems. *Physica D*, **8**, 303–42.

[25] Otani, N. (2004). Home page. http://otani.vet.cornell.edu.

[26] Poenaru, Valentin and Toulouse, Gérard (1977). The crossing of defects in ordered media and the topology of 3-manifolds. *Journal de Physique*, **38**(8), 887–895.

[27] Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (2002). *Numerical recipes in C++ [C, Fortran, ...], the art of scientific computing,* 2nd edition. Cambridge University Press, Cambridge.

[28] Raju, A., Clement, C. B., Hayden, L. X., Kent-Dobias, J. P., Liarte, D. B., Rocklin, D. Z., and Sethna, J. P. (2019). Normal form for renormalization groups. *Physical Review X*, **9**, 021014.

[29] Rand, D., Ostlund, S., Sethna, J. P., and Siggia, E. D. (1982). Universal transition from quasiperiodicity to chaos in dissipative systems. *Physical Review Letters*, **49**, 132–5.

[30] Sartori, Pablo and Tu, Yuhai (2011). Noise filtering strategies in adaptive biochemical signaling networks. *Journal of Statistical Physics*, **142**(6), 1206–1217.

[31] Sethna, J. P. and Myers, C. R. (2020). *Entropy, order parameters, and complexity*, computer exercises and course materials. http://sethna.lassp.cornell.edu/StatMech/EOPCHintsAndMaterials.html.

[32] Sicilia, Alberto, Sarrazin, Yoann, Arenzon, Jeferson J, Bray, Alan J, and Cugliandolo, Leticia F (2009). Geometry of phase separation. *Physical Review E*, **80**(3), 031121.

[33] von Neumann, J. (1952). In *Metal Interfaces*, pp. 108–10. American Society for Metals, Cleveland.

[34] Winfree, A. T. (1991). Varieties of spiral wave behavior: An experimentalist's approach to the theory of excitable media. *Chaos*, **1**, 303–34.