

## Unpublished Exercises

# Statistical Mechanics: Entropy, Order Parameters, and Complexity

James P. Sethna

Laboratory of Atomic and Solid State Physics, Cornell University, Ithaca, NY  
14853-2501

The author provides this version of this manuscript with the primary intention of making the text accessible electronically—through web searches and for browsing and study on computers. Oxford University Press retains ownership of the copyright. Hard-copy printing, in particular, is subject to the same copyright rules as they would be for a printed book.

CLARENDON PRESS • OXFORD  
2018



# Contents

<b>Contents</b>	<b>v</b>
<b>List of figures</b>	<b>xi</b>
Chapter 1: What is statistical mechanics?	6
Exercises	6
1.1 Quantum dice and coins	6
1.3 Waiting time paradox	7
1.4 Stirling's formula	8
1.9 First to fail: Weibull	9
1.10 Emergence	9
1.11 Emergent vs. fundamental	9
1.12 The birthday problem	10
1.13 Width of the height distribution	11
1.14 Nonlinear fits	12
1.15 Fisher information and Cramér–Rao	15
1.16 Distances in probability space	16
Chapter 2: Random walks and emergent properties.	19
Exercises	19
2.1 Random walks in grade space	19
2.2 Photon diffusion in the Sun	19
2.14 Modified random walk	19
2.15 Modified diffusion	19
2.16 Density dependent diffusion	20
2.17 Local conservation	20
2.18 Absorbing boundary conditions	20
2.19 Run and tumble	20
2.20 Continuous time walks: Ballistic to diffusive	21
2.21 Random walks and generating functions	22
Chapter 3: Temperature and equilibrium	23
Exercises	23
3.2 Large and very large numbers	23
3.5 Hard sphere gas	23
3.10 Triple product relation	24
3.11 Maxwell relations	24
3.13 Weirdness in high dimensions	25
3.14 Entropy maximum and temperature	25
3.15 Taste and smell and chemical potential	25
3.16 Undistinguished particles	25
3.17 Random energy model	26

<b>Chapter 4: Phase-space dynamics and ergodicity</b>	28
<b>Exercises</b>	28
4.2    Liouville vs. the damped pendulum	28
4.5    No attractors in Hamiltonian systems	28
4.6    Perverse initial conditions	28
4.7    Bayesian priors	28
4.8    Crooks	30
4.9    Jarzynski	31
4.10   2D turbulence and Jupiter's great red spot	33
<b>Chapter 5: Entropy</b>	36
<b>Exercises</b>	36
5.8    The Arnol'd cat map	36
5.12   Rubber band	38
5.13   How many shuffles?	39
5.24   The Dyson sphere	40
5.18   Undistinguished particles and the Gibbs factor	41
5.19   Entropy of socks	41
5.20   Loaded dice and sticky spheres	42
5.21   Gravity and entropy	42
5.23   Aging, entropy, and DNA	42
5.22   Entropy of the galaxy	43
5.25   Nucleosynthesis and the arrow of time	43
<b>Chapter 6: Free energies</b>	46
<b>Exercises</b>	46
6.4    Molecular motors and free energies	46
6.6    Lagrange	47
6.13   Pollen and hard squares	48
6.15   Entropic forces: Gas vs. rubber band	49
6.16   Rubber band free energy	49
6.17   Rubber band: Illustrating the formalism	49
6.18   Langevin dynamics	50
6.19   Newton's theory of sound	50
6.20   Word frequency: Zipf's law	50
6.21   Epidemics and zombies	53
6.22   Nucleosynthesis as a chemical reaction	54
6.23   Entropy cost for controlling engines	56
<b>Chapter 7: Quantum statistical mechanics</b>	59
<b>Exercises</b>	59
7.4    Does entropy increase in quantum systems?	59
7.9    Bosons are gregarious: superfluids and lasers	59
7.17   Eigenstate thermalization	60
7.18   Is sound a quasiparticle?	60
7.19   Drawing wavefunctions	60
7.20   Universe of light baryons	61
7.21   Many-fermion wavefunction nodes	61
7.22   The greenhouse effect and cooling coffee	61
<b>Chapter 8: Calculation and computation</b>	63
<b>Exercises</b>	63

8.5	Detailed balance	63
8.17	Solving the Ising model with parallel updates	63
8.18	Low temperature expansion of the Ising model	64
8.19	Cluster expansions: the 2D Ising model	64
8.20	Fruit flies, Markov matrices, and free energies	64
8.21	Kinetic proofreading in cells	67
<b>Chapter 9: Order parameters, broken symmetry, and topology</b>		71
Exercises		71
9.9	Nematic order parameter half space	71
9.10	Ising order parameter	71
9.11	Pentagonal order parameter	71
9.12	Rigidity of crystals	73
9.13	Chiral wave equation	74
9.15	Superfluid second sound	74
9.16	Can't lasso a basketball	75
<b>Chapter 10: Correlations, response, and dissipation</b>		76
Exercises		76
10.9	Quasiparticle poles and Goldstone's theorem	76
10.10	Human correlations	76
10.11	Onsager regression hypothesis	77
10.12	Liquid free energy	77
10.13	Harmonic susceptibility and dissipation	78
10.14	Liquid susceptibility and dissipation	78
10.15	Harmonic fluctuation dissipation	78
10.16	Liquid static susceptibility and fluctuation–dissipation theorem	79
10.17	Harmonic Kramers–Krönig	79
10.18	Liquid Kramers Krönig	79
10.19	Susceptibilities and correlations	79
10.20	Subway bench correlations	80
<b>Chapter 11: Abrupt phase transitions</b>		82
Exercises		82
11.1	Maxwell and van der Waals	82
11.6	Coarsening in the Ising model	82
11.10	What is it unstable to?	83
11.11	Droplet Nucleation in two dimensions	83
11.12	Linear stability of a growing interface	83
11.13	Gibbs free energy barrier	83
11.14	Elastic theory has zero radius of convergence	84
<b>Chapter 12: Continuous phase transitions</b>		88
Exercises		88
12.7	Renormalization-group trajectories	88
12.5	Mean-field theory	89
12.14	Crackling noises	90
12.15	Hearing chaos	90
12.16	The Gutenberg Richter law	91
12.17	Random walks and critical exponents	91

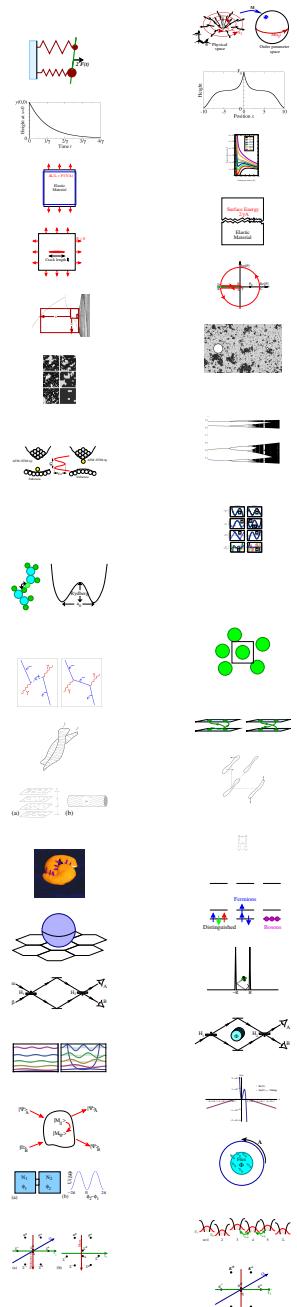
12.18	Hysteresis and Barkhausen Noise	92
12.19	Biggest of bunch: Gumbel	92
12.20	Extreme value statistics: Gumbel, Weibull, and Fréchet	93
12.21	Diffusion equation and universal scaling functions	95
12.22	Period doubling and the onset of chaos	95
12.23	Ising mean field derivation	96
12.24	Mean-field bound for free energy	96
12.26	The onset of chaos: Lowest order renormalization-group for period doubling	97
12.27	The onset of chaos: Full renormalization-group calculation	97
12.28	Singular corrections to scaling	98
12.29	Singular corrections to scaling and the renormalization group	99
12.30	Nonlinear flows, analytic corrections, and hyper-scaling	100
12.31	Beyond power laws: Nonlinear flows and logarithms in the 2D Ising model	101
12.25	Avalanche size distribution	102
12.32	Conformal invariance	103
<b>3</b>	<b>Advanced Statistical Mechanics</b>	<b>107</b>
	Exercises	107
3.1	Eigenvectors near the renormalization-group fixed point	107
3.2	Is the fixed point unique? Period doubling	108
3.3	A fair split? Number partitioning	111
3.4	Cardiac dynamics	113
3.5	Quantum dissipation from phonons	115
3.6	Ising lower critical dimension	116
3.7	XY lower critical dimension and the Mermin-Wagner Theorem	117
3.8	Long-range Ising	117
3.9	Equilibrium Crystal Shapes	118
<b>4</b>	<b>Numerical Methods</b>	<b>121</b>
	Exercises	121
4.1	Condition number and accuracy	121
4.2	Sherman–Morrison formula	121
4.3	Methods of interpolation	122
4.4	Numerical definite integrals	122
4.5	Numerical derivatives	124
4.6	Summing series	124
4.7	Random histograms	124
4.8	Monte Carlo integration	124
4.9	Washboard potential	125
4.10	Sloppy minimization	125

4.11 Sloppy monomials	126
4.12 Conservative differential equations: Accuracy and fidelity	127
<b>5 Quantum</b>	<b>131</b>
Exercises	131
5.1 Quantum notation	131
5.2 Light proton atomic size	131
5.3 Light proton tunneling	131
5.4 Light proton superfluid	132
5.5 Aharonov-Bohm Wire	132
5.6 Anyons	133
5.7 Bell	135
5.8 Parallel Transport, Frustration, and the Blue Phase	136
5.9 Crystal field theory: $d$ -orbitals	139
5.10 Entangled Spins	140
5.11 Rotating Fermions	141
5.12 Lithium ground state symmetry	141
5.13 Matrices, wavefunctions, and group representations	142
5.14 Molecular rotations	143
5.15 Propagators to path integrals	143
5.16 Three particles in a box	143
5.17 Rotation matrices	144
5.18 Trace	144
5.19 Complex exponentials	145
5.20 Dirac $\delta$ -functions	145
5.21 Eigen Stuff	146
5.22 F-electrons and graphene	146
5.23 Juggling buckyballs	147
5.24 Resonances: $\alpha$ -decay	148
5.25 Mirror path integrals	149
5.26 Coherent State Evolution	150
5.27 Decoherence	151
5.28 Quantum Algorithms	152
5.29 Supersymmetric harmonic oscillator	154
5.30 Fourier series and group representations	157
5.31 Solving Schrödinger: WKB, resonances, and lifetimes	159
5.32 GHZ and $\sigma$ matrices	160
5.33 No cloning theorem	161
5.34 Aharonov-Bohm wire: straightforward approach	162
5.35 Number, Phase, and Josephson	162
5.36 Localization	163
5.37 Eight-Fold Way	166
<b>References</b>	<b>171</b>



# List of figures

1.1	Quantum dice	6	
1.2	Emergent	10	
1.3	Fundamental	10	
1.4	Fitting a nonlinear function to data	13	
1.5	Contours of constant cost $C = \chi^2/2$	13	
1.6	Nonlinear model predictions in data space	13	
1.7	Error estimates for fit parameters	14	
3.8	Hard sphere gas	23	
3.9	Excluded area around a hard disk	23	
4.10	Total derivatives	28	
4.11	Crooks fluctuation theorem: evolution in time	30	
4.12	Crooks fluctuation theorem: evolution backward in time	31	
4.13	One particle in a piston	32	
4.14	Collision phase diagram	32	
4.15	Jupiter's Red Spot	33	
4.16	Circles	35	
5.17	Arnol'd cat map	37	
5.18	Evolution of an initially concentrated phase-space density	37	
5.19	Rubber band	39	
5.20	The Universe as a piston	44	
5.21	PV diagram for the universe	44	
6.22	RNA polymerase molecular motor	46	
6.23	Hairpins in RNA	47	
6.24	Telegraph noise in RNA unzipping	47	
6.25	Square pollen	48	
6.26	Piston with rubber band	49	
6.27	Zipf's law for word frequencies	50	
8.28	Fruit fly behavior map	66	
8.29	DNA polymerase	67	
8.30	DNA proofreading	67	
8.31	Naive DNA replication	68	
8.32	Equilibrium DNA replication	68	
8.33	Kinetic proofreading	69	
8.34	Impossible stair	69	
9.35	Projective Plane	71	
9.36	Frustrated soccer ball	72	
9.37	Cutting and sewing: disclinations	72	
9.38	Crystal rigidity	73	
9.39	Sphere	75	



9.40 Escape in the third dimension	75
10.41 Two springs	77
10.42 Pulling	79
10.43 Hitting	79
11.44 $P$ - $V$ plot: van der Waals	82
11.45 Stretched block	85
11.46 Fractured block	85
11.47 Critical crack	85
11.48 Contour integral in complex pressure	87
12.49 Scaling in the period doubling bifurcation diagram	95
12.50 Two proteins in a critical membrane	103
12.51 Ising model under conformal rescaling	104
3.1 Expanded bifurcation diagram	110
3.2 Atomic tunneling from a tip	116
4.1 Interpolation methods	123
5.1 Atom tunneling	131
5.2 Atoms with imaginary box of size equal to average space per atom	132
5.3 Feynman diagram: identical particles	134
5.4 Braiding of paths in two dimensions	134
5.5 Blue Phase Molecules	137
5.6 Twisted Parallel Transport	137
5.7 Local structures	138
5.8 Parallel transport frustration	138
5.9 Orange-peel carpet	139
5.10 Level diagram	144
5.11 <i>Atom adsorbed on graphene</i>	147
5.12 <i>One-dimensional nuclear potential</i>	148
5.13 Qbit weirdness	149
5.14 Bohm-Aharonov and mirrors	150
5.15 Supersymmetric eigenenergies and eigenstates	157
5.16 Real and imaginary part of $V(x)$	160
5.17 Hypothetical cloning machine	161
5.18 Aharonov-Bohm wire loop	162
5.19 Josephson junction schematic	162
5.20 Anderson model of localization	164
5.21 Baryon octet	166
5.22 Meson octet	169

# Preface to second edition

I have received much positive feedback for the exercises included in the first edition of *Statistical Mechanics: Entropy, Order Parameters, and Complexity*. Indeed, my motivation and design of the text was to glue together exercises covering the broad sweep of statistical mechanics with chapters explaining the core of the subject. Building on strength, the primary new material in the second edition are new exercises.

The new exercises come in three categories. First, there are exercises that introduce exciting developments in statistical mechanics: *e.g.* exact results in nonequilibrium statistical mechanics (new Exercises 4.8 and 4.9), Jupiter's red spot and two-dimensional turbulence 4.10, fruit flies and Markov chains 8.20, and nucleosynthesis and the arrow of time 5.25. (Our current low entropy state is best explained as a chemical reaction in an expanding piston.) Second, I have made my computer exercises much more systematic, providing Python and Mathematica notebook hints files to guide the students through the process. New computer exercises include conformal invariance 12.32 and human correlations on subway cars 10.20. Third, I have added many short, targeted exercises. These reflect my recent experience in ‘flipping the classroom’; I now ask the students to read the material at home, and reserve class time for activities. Exercises marked with a  $\circled{P}$  are pre-class questions designed to stimulate thought about topics introduced in specific sections, such as weirdness in high dimensions 3.13, and sound as a quasiparticle 7.18. Exercises marked with a  $\circled{I}$  are used as in-class activities; a nice example is new Exercise 9.11 where I bring soccer balls to class and we roll them around to study frustration and parallel transport. The second edition also refines and revises many of the older exercises.

I have also made substantive changes to the derivation of power laws and scaling at continuous phase transitions (Section 12.2), and minor edits throughout the text.

Thanks to Pietro Torta and Varun Gandhi for stamping out many errors and obscurities.

James P. Sethna  
Ithaca, NY  
2017

## Chapter 1: What is statistical mechanics?

### Exercises

**1.1 Quantum dice and coins.**<sup>1</sup> (Quantum) ①

You are given two unusual ‘three-sided’ dice which, when rolled, show either one, two, or three spots. There are three games played with these dice: *Distinguishable*, *Bosons*, and *Fermions*. In each turn in these games, the player rolls the two dice, starting over if required by the rules, until a legal combination occurs. In *Distinguishable*, all rolls are legal. In *Bosons*, a roll is legal only if the second of the two dice shows a number that is larger or equal to that of the first of the two dice. In *Fermions*, a roll is legal only if the second number is strictly larger than the preceding number. See Fig. 1.1 for a table of possibilities after rolling two dice.

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

	1	2	3
Roll #2	3	4	5
2	3	4	5
1	2	3	4

**Fig. 1.1 Quantum dice.** Rolling two dice. In *Bosons*, one accepts only the rolls in the shaded squares, with equal probability  $1/6$ . In *Fermions*, one accepts only the rolls in the darkly-shaded squares (not including the diagonal from lower left to upper right), with probability  $1/3$ .

(a) *Presume the dice are fair: each of the three numbers of dots shows up  $1/3$  of the time. For a legal turn rolling a die twice in the three games (*Distinguishable*, *Bosons*, and *Fermions*, what is*

*the probability  $\rho(5)$  of rolling a 5?*

(b) *For a legal turn in the three games, what is the probability of rolling a double? (Hint: There is a Pauli exclusion principle: when playing *Fermions*, no two dice can have the same number of dots showing.) Electrons are fermions; no two non-interacting electrons can be in the same quantum state. Bosons are gregarious (Exercise 7.9); non-interacting bosons have a larger likelihood of being in the same state.*

Let us decrease the number of sides on our dice to  $N = 2$ , making them quantum coins, with a head  $H$  and a tail  $T$ . Let us increase the total number of coins to a large number  $M$ ; we flip a line of  $M$  coins all at the same time, repeating until a legal sequence occurs. In the rules for legal flips of quantum coins, let us make  $T < H$ . A legal Boson sequence, for example, is then a pattern  $TTTT \cdots HHHH \cdots$  of length  $M$ ; all legal sequences have the same probability.

(c) *What is the probability that all the  $M$  flips of our quantum coin give the same value ( $HHHH \cdots$  or  $TTTT \cdots$ ), in the three games? Discuss: how is the probability of having all flips be in the same state change with the three games? (Hint: How many legal sequences are there for the three games? How many of these are all the same value?)*

Let us now consider a biased coin, with probability  $p = 1/3$  of landing  $H$  and thus  $1 - p = 2/3$  of landing  $T$ . Note that if two sequences are legal in both *Bosons* and *Distinguishable*, their relative probability is the same in both games.<sup>2</sup>

(d) *What is the probability that all the  $M$  flips of our biased quantum coin give tails, in *Distinguishable*? What is the relative probability in *Bosons* of getting exactly one head versus getting all  $M$  tails? What is the relative probability of getting*

<sup>1</sup>This exercise was developed in collaboration with Sarah Shandera.

<sup>2</sup>The probability of finding a particular legal sequence in *Bosons* is larger by a constant factor due to discarding the illegal sequences. This factor is just one over the probability of a given toss of the coins being legal,  $Z = \sum_{\alpha} p_{\alpha}$  over legal Boson sequences  $\alpha$ . This normalization constant is called the *partition function*, and is surprisingly central to statistical mechanics.

*exactly  $m$  heads compared to all tails in Bosons?*  
 This should give a familiar series. As  $M$  gets large, what is the probability in Bosons that all coins flip tails?

We can view our quantum dice and coins as non-interacting particles, with the biased coin having a lower energy for  $T$  than for  $H$  (Section 7.4). Having a non-zero probability of having all the bosons in the single-particle ground state  $T$  is Bose condensation (Section 7.6), closely related to superfluidity and lasers (Exercise 7.9).

### 1.3 Waiting time paradox.<sup>3</sup> ⓘ

Here we examine the *waiting time paradox*: for events happening at random times, the average time until the next event equals the average time between events. If the average waiting time until the next event is  $\tau$ , then the average time since the last event is also  $\tau$ . Is the mean total gap between two events then  $2\tau$ ? Or is it  $\tau$ , the average time to wait starting from the previous event? Working this exercise introduces the importance of different *ensembles*.

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

A pedestrian repeatedly approaches a bus stop at random times  $t$ , and notes how long it takes before the first bus passes, and before the first car passes.

(a) Draw the probability density for the ensemble of waiting times  $\delta^{\text{Bus}}$  to the next bus observed by the pedestrian. Draw the density for the corresponding ensemble of times  $\delta^{\text{Car}}$ . What is the mean waiting time for a bus  $\langle \delta^{\text{Bus}} \rangle_t$ ? The mean time  $\langle \delta^{\text{Car}} \rangle_t$  for a car? (Hint: You may be familiar with the car probability density from the study of radioactive decay. If not, you can derive it by writing a differential equation for  $S(t)$ , the ‘survival probability’ that a car has not yet arrived at time  $t$ . What is the probability density  $P(\delta^{\text{Car}})$  for the arrival of the *first* car, in terms of  $S(t)$ ?)

In statistical mechanics, we shall describe specific physical systems (a bottle of  $N$  atoms with energy  $E$ ) by considering *ensembles* of systems. Sometimes we shall use two different ensembles to describe the same system (all bottles of  $N$  atoms with energy  $E$ , or all bottles of  $N$  atoms at that temperature  $T$  where the mean energy is  $E$ ). We have been looking at the time-averaged ensemble (the ensemble  $\langle \dots \rangle_t$  over random times  $t$ ). There is also in this problem an ensemble average over the gaps between vehicles ( $\langle \dots \rangle_{\text{gap}}$  over random time intervals); these two give different averages for the same quantity.

A traffic engineer sits at the bus stop, and measures an ensemble of time gaps  $\Delta^{\text{Bus}}$  between neighboring buses, and an ensemble of gaps  $\Delta^{\text{Car}}$  between neighboring cars.

(b) Draw the probability density of gaps she observes between buses. Draw the probability density of gaps between cars. (Hint: Is it different from the ensemble of car waiting times you found in part (a)? Why not?) What is the mean gap time  $\langle \Delta^{\text{Bus}} \rangle_{\text{gap}}$  for the buses? What is the mean gap time  $\langle \Delta^{\text{Car}} \rangle_{\text{gap}}$  for the cars? (Hint: One of these probability distributions cannot be drawn completely. Ignoring measurement error and imperfectly punctual public transportation, the density involves the Dirac  $\delta$ -function.<sup>4</sup>)

You should find that the mean waiting time for a bus in part (a) is half the mean bus gap time in (b), which seems sensible – the gap seen by the pedestrian is the sum of the  $\delta_+^{\text{Bus}} + \delta_-^{\text{Bus}}$  of the waiting time and the time since the last bus. However, you should also find the mean waiting time for a car *equals* the mean car gap time. The equation  $\Delta^{\text{Car}} = \delta_+^{\text{Car}} + \delta_-^{\text{Car}}$  would seem to imply that the average gap seen by the pedestrian is twice the mean waiting time.

(c) How can the average gap between cars measured by the pedestrian be different from that measured by the traffic engineer? Discuss.

(d) Consider a short experiment, with three cars passing at times  $t = 0, 2$ , and  $8$  (so there are two gaps, of length 2 and 6). What is  $\langle \Delta^{\text{Car}} \rangle_{\text{gap}}$ ? What is  $\langle \Delta^{\text{Car}} \rangle_t$ ? Explain why they are different.

When we change between ensembles in statistical mechanics, we will need to check carefully that the

<sup>3</sup>The original form of this exercise was developed in collaboration with Piet Brouwer.

<sup>4</sup>The  $\delta$ -function  $\delta(x - x_0)$  is a probability density which has 100% probability of being in any interval containing  $x_0$ ; thus  $\delta(x - x_0)$  is zero unless  $x = x_0$ , and  $\int f(x)\delta(x - x_0) dx = f(x_0)$  so long as the domain of integration includes  $x_0$ . Mathematically, this is not a function, but rather a distribution or a measure.

properties we calculate are the same in the different ensembles, as we take the number of particles to be large.

**1.4 Stirling's formula.** (Mathematics) ④

Stirling's approximation,  $n! \sim \sqrt{2\pi n}(n/e)^n$ , is remarkably useful in statistical mechanics; it gives an excellent approximation for large  $n$ . In statistical mechanics the number of particles is so large that we usually care not about  $n!$ , but about its logarithm, so  $\log(n!) \sim n \log n - n + \frac{1}{2} \log(2\pi n)$ . Finally,  $n$  is often so large that the final term is a tiny fractional correction to the others, giving the simple formula  $\log(n!) \sim n \log n - n$ .

(a) Calculate  $\log(n!)$  and these two approximations to it for  $n = 2, 4, 1024$ , and one mole ( $6.03 \cdot 10^{23}$ ). Discuss the percentage accuracy of these two approximations for small and large  $n$ .

Note that  $\log(n!) = \log(1 \times 2 \times 3 \times \dots \times n) = \log(1) + \log(2) + \dots + \log(n) = \sum_{m=1}^n \log(m)$ .

(b) Convert the sum to an integral,  $\sum_{m=1}^n \approx \int_0^n dm$ . Derive the simple form of Stirling's formula.

(c) Draw a plot of  $\log(m)$  and a bar chart showing  $\log(\text{ceiling}(n))$ . (Here  $\text{ceiling}(x)$  represents the smallest integer larger than  $x$ .) Argue that the integral under the bar chart is  $\log(n!)$ .

The difference between the sum and the integral in part (c) should look approximately like a collection of triangles, except for the region between zero and one. The sum of the areas equals the error in the simple form for Stirling's formula.

(d) Imagine doubling these triangles into rectangles on your drawing from part (c), and sliding them sideways (ignoring the error for  $m$  between zero and one). Explain how this relates to the term  $\frac{1}{2} \log n$  in Stirling's formula  $\log(n!) - (n \log n - n) \approx \frac{1}{2} \log(2\pi n) = \frac{1}{2} \log(2) + \frac{1}{2} \log(\pi) + \frac{1}{2} \log(n)$ .

**1.9 First to fail: Weibull.**<sup>5</sup> (Mathematics, Statistics, Engineering) ③

Suppose you have a brand-new supercomputer with  $N = 1000$  processors. Your parallelized code, which uses all the processors, cannot be restarted in mid-stream. How long a time  $t$  can you expect to run your code before the first processor fails?

This is example of *extreme value statistics* (see also exercises 12.19 and 12.20), where here we are looking for the smallest value of  $N$  random variables that are all bounded below by zero. For large  $N$  the probability distribution  $\rho(t)$  and survival probability  $S(t) = \int_t^\infty \rho(t') dt'$  are often given by the *Weibull distribution*

$$\begin{aligned} S(t) &= e^{-(t/\alpha)^\gamma}, \\ \rho(t) &= -\frac{dS}{dt} = \frac{\gamma}{\alpha} \left(\frac{t}{\alpha}\right)^{\gamma-1} e^{-(t/\alpha)^\gamma}. \end{aligned} \quad (1.1)$$

Let us begin by assuming that the processors have a constant rate  $\Gamma$  of failure, so the probability density of a single processor failing at time  $t$  is  $\rho_1(t) = \Gamma \exp(-\Gamma t)$  as  $t \rightarrow 0$ , and the survival probability for a single processor  $S_1(t) = 1 - \int_0^t \rho_1(t') dt' \approx 1 - \Gamma t$  for short times.

(a) Using  $(1 - \epsilon) \approx \exp(-\epsilon)$  for small  $\epsilon$ , show that the the probability  $S_N(t)$  at time  $t$  that all  $N$  processors are still running is of the Weibull form (eqn 1.1). What are  $\alpha$  and  $\gamma$ ?

Often the probability of failure per unit time goes to zero or infinity at short times, rather than to a constant. Suppose the probability of failure for one of our processors

$$\rho_1(t) \sim Bt^k \quad (1.2)$$

with  $k > -1$ . (So,  $k < 0$  might reflect a breaking-in period, where survival for the first few minutes increases the probability for later survival, and  $k > 0$  would presume a dominant failure mechanism that gets worse as the processors wear out.)

(b) Show the survival probability for  $N$  identical processors each with a power-law failure rate (eqn 1.2) is of the Weibull form for large  $N$ , and give  $\alpha$  and  $\gamma$  as a function of  $B$  and  $k$ .

The parameter  $\alpha$  in the Weibull distribution just sets the scale or units for the variable  $t$ ; only the exponent  $\gamma$  really changes the shape of the distribution. Thus the form of the failure distribution

at large  $N$  only depends upon the power law  $k$  for the failure of the individual components at short times, not on the behavior of  $\rho_1(t)$  at longer times. This is a type of *universality*,<sup>6</sup> which here has a physical interpretation; at large  $N$  the system will break down soon, so only early times matter.

The Weibull distribution, we must mention, is often used in contexts not involving extremal statistics. Wind speeds, for example, are naturally always positive, and are conveniently fit by Weibull distributions.

**1.10 Emergence.** ②

We begin with the broad statement *Statistical mechanics explains the simple behavior of complex systems*. New laws emerge from bewildering interactions of constituents.

*Discuss which of these emergent behaviors is probably not studied using statistical mechanics.*

(a) *The emergence of the wave equation from the collisions of atmospheric molecules,*

(b) *The emergence of Newtonian gravity from Einstein's general theory,*

(c) *The emergence of random stock price fluctuations from the behavior of traders,*

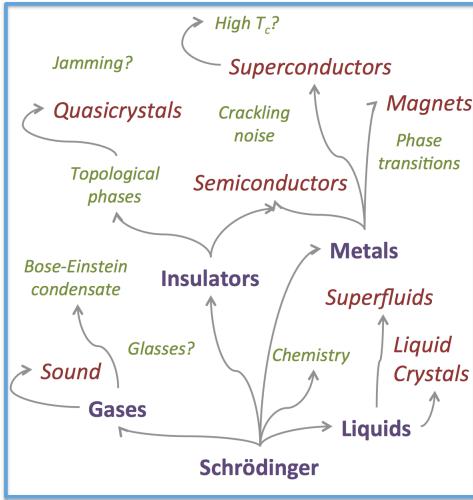
(d) *The emergence of a power-law distribution of earthquake sizes from the response of rubble in earthquake faults to external stresses.*

**1.11 Emergent vs. fundamental.** ②

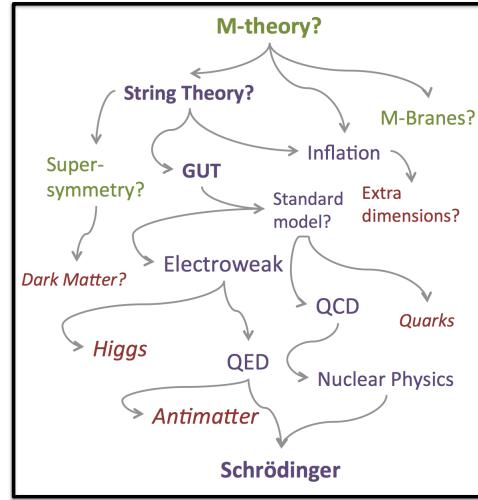
Statistical mechanics is central to condensed-matter physics. It is our window into the behavior of materials – how complicated interactions between large numbers of atoms lead to physical laws (Fig. 1.2). For example, the theory of sound emerges from the complex interaction between many air molecules governed by Schrödinger's equation. More is different [2].

<sup>5</sup>Developed with the assistance of Paul (Wash) Wawrzynek

<sup>6</sup>The Weibull distribution forms a one-parameter family of universality classes; see chapter 12.



**Fig. 1.2 Emergent.** New laws describing macroscopic materials emerge from complicated microscopic behavior [47].



**Fig. 1.3 Fundamental.** Laws describing physics at lower energy emerge from more fundamental laws at higher energy [47].

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. 1.3). 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, . . .

(b) Schrödinger’s equation can in principle be solved to describe almost all of materials physics, biology, . . . , apart from radioactive decay or gravity. How many parameters would one need as input to Schrödinger’s equation to describe materials and biology and such? 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?

### 1.12 The birthday problem. ②

Remember birthday parties in your elementary school? Remember those years when two kids had the same birthday? How unlikely!

How many kids would you need in class to get, more than half of the time, at least two with the same birthday?

(a) Numerical. Write `BirthdayCoincidences(K, C)`, a routine that returns the fraction among  $C$  classes for which at least two kids (among  $K$  kids per class) have the same birthday. (Hint: By sorting a random list of integers, common birthdays will be adjacent.) Plot this probability versus  $K$  for a reasonably large value of  $C$ . Is it a surprise that your classes had overlapping birthdays when you were young?

One can intuitively understand this, by remembering that to avoid a coincidence there are

$K(K - 1)/2$  pairs of kids, all of whom must have different birthdays (probability  $364/365 = 1 - 1/D$ , with  $D$  days per year).

$$P(K, D) \approx (1 - 1/D)^{K(K-1)/2} \quad (1.3)$$

This is clearly a crude approximation – it doesn’t vanish if  $K > D!$  Ignoring subtle correlations, though, it gives us a net probability

$$\begin{aligned} P(K, D) &\approx \exp(-1/D)^{K(K-1)/2} \\ &\approx \exp(-K^2/(2D)) \end{aligned} \quad (1.4)$$

Here we’ve used the fact that  $1 - e \approx \exp(-e)$ , and assumed that  $K/D$  is small.

(b) Analytical. Write the exact formula giving the probability, for  $K$  random integers among  $D$  choices, that no two kids have the same birthday. (Hint: What is the probability that the second kid has a different birthday from the first? The third kid has a different birthday from the first two?) Show that your formula does give zero if  $K > D$ . Converting the terms in your product to exponentials as we did above, show that your answer is consistent with the simple formula above, if  $K \ll D$ . Inverting eqn 1.4, give a formula for the number of kids needed to have a 50% chance of a shared birthday.

Some years ago, we were doing a large simulation, involving sorting a lattice of  $1000^3$  random fields (roughly, to figure out which site on the lattice would trigger first). If we want to make sure that our code is unbiased, we want different random fields on each lattice site – a giant birthday problem.

Old-style random number generators generated a random integer ( $2^{32}$  ‘days in the year’) and then divided by the maximum possible integer to get a random number between zero and one. Modern random number generators generate all  $2^{52}$  possible double precision numbers between zero and one.

(c) If there are  $2^{32}$  distinct four-byte unsigned integers, how many random numbers would one have to generate before one would expect coincidences half the time? Generate lists of that length, and check your assertion. (Hints: It is faster to use array operations, especially in interpreted languages. I generated a random array with  $N$  entries, sorted it, subtracted the first  $N - 1$  entries

from the last  $N - 1$ , and then called `min` on the array.) Will we have to worry about coincidences with an old-style random number generator? How large a lattice  $L \times L \times L$  of random double precision numbers can one generate with modern generators before having a 50% chance of a coincidence? If you have a fast machine with a large memory, you might test this too.

### 1.13 Width of the height distribution.<sup>7</sup> (Statistics) ③

In this exercise and Exercise 4.7, 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

$$\mathcal{N}(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/(2\sigma^2)} \quad (1.5)$$

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 histogram that is well described by a normal distribution.<sup>8</sup> 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  $\mu$  and variance  $\sigma^2$  as the product

$$P(\{x_n\}|\mu, \sigma) = \prod_{n=1}^N \mathcal{N}(x_n|\mu, \sigma^2). \quad (1.6)$$

We first introduce the concept of *sufficient statistics*. Our likelihood (eqn 1.6) 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 1.6 as a formula depending on the data  $\{x_n\}$  only through  $N$ ,  $\bar{x} = (1/N) \sum_n x_n$  and  $S = \sum_n (x_n - \bar{x})^2$ .

<sup>7</sup>This exercise was developed in collaboration with Colin Clement.

<sup>8</sup>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).

Given the model of independent normal distributions, its likelihood is a formula depending only on<sup>9</sup>  $\bar{x}$  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.<sup>10</sup> Maximum likelihood is a common method for estimating model parameters; the estimates  $(\mu_{\text{ML}}, \sigma_{\text{ML}})$  are given by the peak of the probability distribution  $P$ .

(b) Show that  $P(\{x_n\}|\mu_{\text{ML}}, \sigma_{\text{ML}})$  takes its maximum value at

$$\begin{aligned}\mu_{\text{ML}} &= \frac{\sum_n x_n}{N} = \bar{x} \\ \sigma_{\text{ML}} &= \sqrt{\sum_n (x_n - \bar{x})^2 / N} = \sqrt{S/N}.\end{aligned}\quad (1.7)$$

(Hint: It is easier to maximize the log likelihood;  $P(\theta)$  and  $\log(P(\theta))$  are maximized at the same point  $\theta_{\text{ML}}$ .)

If we draw samples of size  $N$  from a distribution of known mean  $\mu_0$  and standard deviation  $\sigma_0$ , 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  $\langle \mu_{\text{ML}} \rangle_{\text{samp}}$  over many samples gives the correct mean, we say that  $\mu_{\text{ML}}$  is *unbiased*. The maximum likelihood estimate for  $\sigma_{\text{ML}}^2$ , however, is biased. Again, for the extreme example  $N = 1$ ,  $\sigma_{\text{ML}}^2 = 0$  for every sample!

(c) Assume the entire population is drawn from some (perhaps non-Gaussian) distribution of variance  $\langle x^2 \rangle_{\text{samp}} = \sigma_0^2$ . For simplicity, let the mean of the population be zero. Show that

$$\begin{aligned}\langle \sigma_{\text{ML}}^2 \rangle_{\text{samp}} &= (1/N) \left\langle \sum_{n=1}^N (x_n - \bar{x})^2 \right\rangle_{\text{samp}} \\ &= \frac{N-1}{N} \sigma_0^2.\end{aligned}\quad (1.8)$$

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:

$\bar{x} = (1/N) \sum_n x_n$  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<sup>11</sup> from  $N$  measurements, to divide by  $\sqrt{N-1}$ :

$$\sigma_{N-1}^2 \approx \frac{\sum_n (x_n - \bar{x})^2}{N-1}. \quad (1.9)$$

This correction  $N \rightarrow N-1$  is generalized to more complicated problems by considering the number of independent degrees of freedom (here  $N-1$  degrees of freedom in the vector  $x_n - \bar{x}$  of deviations from the mean). Alternatively, it is interesting that the bias disappears if one does not estimate both  $\sigma^2$  and  $\mu$  by maximizing the joint likelihood, but integrating (or *marginalizing*) over  $\mu$  and then finding the maximum likelihood for  $\sigma^2$ .

#### 1.14 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  $\theta_\alpha$ , control parameters  $t_i$  describing different experimental conditions, experimental data  $d_i$  taken under these different conditions, and a nonlinear function  $y_i(\theta)$  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 [53, 54] and Exercise 6.14.) Our model is

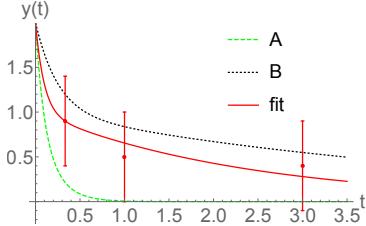
$$\mathbf{y}_\theta(t) = \exp(-\theta_1 t) + \exp(-\theta_2 t). \quad (1.10)$$

Here the parameters  $\theta = \{\theta_1, \theta_2\}$  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  $\sigma_i$ .

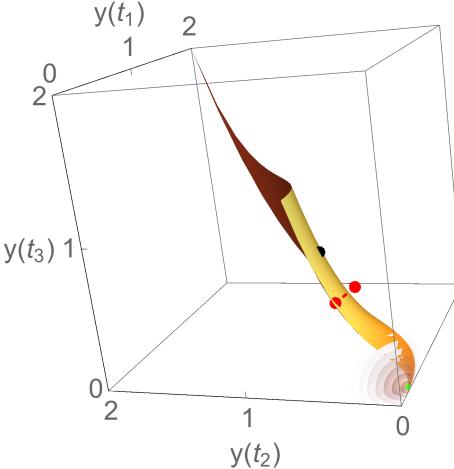
<sup>9</sup>In this exercise we shall use  $\bar{X}$  denote a quantity averaged over a single sample of  $N$  people, and  $\langle X \rangle_{\text{samp}}$  denote a quantity also averaged over many samples.

<sup>10</sup>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  $\sigma_i$ . In statistics, the estimation of the measurement error is often part of the modeling process, as in this exercise.

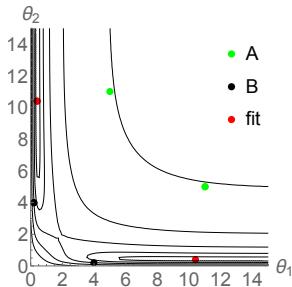
<sup>11</sup>Do not confuse this with the estimate of the error in the *mean*  $\bar{x}$ .



**Fig. 1.4 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. 1.6 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.



**Fig. 1.5 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

$$C(\boldsymbol{\theta}) = \chi^2/2 = \sum_i (y_i(\boldsymbol{\theta}) - d_i)^2 / 2\sigma_i^2. \quad (1.11)$$

The cost is half of what the statisticians call  $\chi^2$  (pronounced ‘chi squared’).

<sup>12</sup>The metric tensor  $g_{ij}$  on a Riemannian manifold gives the distance between nearby points. If the two points have coordinates  $\mathbf{x}$  and  $\mathbf{x} + \Delta$  and  $\Delta$  is small, then the squared distance is  $\sum_{ij} g_{ij} \Delta_i \Delta_j$ .

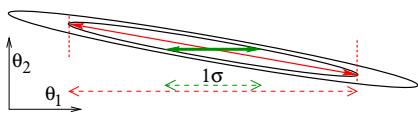
First, let us provide a few interpretations of the cost.

- (a) [i] Interpret  $C$  as half the squared distance in a data space (Fig. 1.6) which has one coordinate per experimental measurement. What is the metric tensor  $g_{ij}$  in data space, in terms of the error bars  $\sigma_i$ ?<sup>12</sup> [ii] Suppose the experimental data points  $d_i$  have errors that are distributed as independent Gaussians of standard deviation  $\sigma_i$ . 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  $\exp(-\mathcal{H}/k_B T)$ , 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.

(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. 1.6). For general  $\sigma_i$ , make an analogy with the momentum distribution of classical particles with different masses to describe the probability distribution.



**Fig. 1.7 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  $1\sigma$  range of  $\theta_1$  keeping  $\theta_2$  fixed is the solid arrow. The total uncertainty  $\sigma_1$  for  $\theta_1$  includes fluctuations of  $\theta_2$  (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 1.13 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  $\boldsymbol{\Delta}$ ,

$$C(\boldsymbol{\theta}) \approx C^{\text{best}} + \frac{1}{2} H_{\alpha\beta} \Delta_\alpha \Delta_\beta, \quad (1.12)$$

where we shall call

$$H_{\alpha\beta} = \frac{\partial^2 C}{\partial \theta_\alpha \partial \theta_\beta} \quad (1.13)$$

the cost Hessian.

(c) Which of the eigenvectors  $\mathbf{e}_v$  or  $\mathbf{e}_h$  in Fig. 1.7 corresponds to a stiff direction (larger eigenvalue of  $H$ )? Which is sloppier? Verify that the probability distribution of  $\theta_1$  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  $\theta_1$  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  $\theta_1$  is given by

$$\begin{aligned} \langle \Delta_1^2 \rangle &= \int \Delta_1^2 P(\boldsymbol{\Delta}) d\boldsymbol{\Delta} \\ &= \int \frac{\Delta_1^2}{Z} e^{-\frac{1}{2} \sum_{\alpha\beta} \Delta_\alpha H_{\alpha\beta} \Delta_\beta} d\boldsymbol{\Delta}, \end{aligned} \quad (1.14)$$

where

$$Z = \int \exp \left( -\frac{1}{2} \sum_{\alpha\beta} \Delta_\alpha H_{\alpha\beta} \Delta_\beta \right) d\boldsymbol{\Delta}. \quad (1.15)$$

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*  $\lambda X$  to the Hamiltonian, shifting the partition function to  $Z(\lambda) = \sum \exp(-(H + \lambda X)/k_B T)$ , with free energy  $F(\lambda) = -k_B T \log Z$ . Then

$$\begin{aligned} \frac{dF}{d\lambda} \Big|_{\lambda=0} &= \frac{-k_B T}{Z} \frac{dZ}{d\lambda} \Big|_{\lambda=0} \\ &= \frac{k_B T}{Z} \int \frac{X}{k_B T} e^{-H/k_B T} \quad (1.16) \\ &= \langle X \rangle \end{aligned}$$

and

$$\begin{aligned} \frac{d^2 F}{d\lambda^2} \Big|_{\lambda=0} &= \frac{k_B T}{Z^2} \left( \frac{dZ}{d\lambda} \right)^2 \Big|_{\lambda=0} - \frac{k_B T}{Z} \frac{d^2 Z}{d\lambda^2} \Big|_{\lambda=0} \\ &= \frac{\langle X \rangle^2}{k_B T} - \frac{k_B T}{Z} \int \frac{X^2}{(k_B T)^2} e^{-H/k_B T} \\ &= \frac{\langle X^2 \rangle - \langle X \rangle^2}{-k_B T} \\ &= \frac{\langle (X - \langle X \rangle)^2 \rangle}{-k_B T} \quad (1.17) \end{aligned}$$

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  $\lambda$  times a unit vector in the shared  $\theta_1$  and  $\Delta_1$  direction, so

$$Z(\lambda) = \int e^{-\frac{1}{2} \boldsymbol{\Delta}^T H \boldsymbol{\Delta} - \boldsymbol{\lambda} \cdot \boldsymbol{\Delta}} d\boldsymbol{\Delta}. \quad (1.18)$$

Complete the square, and show that  $Z(\lambda) = \exp(\frac{1}{2}\lambda^2 \Sigma_{11}) Z(0)$ . Use eqn 1.17 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 1.11 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.

### 1.15 Fisher information and Cramér–Rao.<sup>13</sup>

(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

$\mathbf{x} \in \{1, 2, \dots, 6\}$  with probabilities  $\boldsymbol{\theta} = \{p_1, \dots, p_5\}$ , 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 6.23).
- 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 1.13).
- A least squares model  $y_i(\boldsymbol{\theta})$  for  $N$  data points  $d_i \pm \sigma$  with independent, normally distributed measurement errors (Exercise 1.14) predicts a likelihood for finding a value  $\mathbf{x} = \{x_i\}$  of the data  $\{d_i\}$  given by  $P(\mathbf{x}|\boldsymbol{\theta}) = e^{-\sum_i (y_i(\boldsymbol{\theta}) - x_i)^2 / 2\sigma^2} / \sqrt{2\pi\sigma^2}^N$ .

First we introduce the Fisher information matrix, which gives the metric describing the distance between two nearby probability distributions (see also Exercises 1.14 and 1.16). How ‘distant’ is a loaded die 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 a group of US women from a group of Pakistani women, if you only knew their heights?

The natural distance between two *nearby* probability distributions  $P(\mathbf{x}|\boldsymbol{\theta})$  and  $Q = P(\mathbf{x}|\boldsymbol{\theta} + \epsilon\boldsymbol{\Delta})$  is

$$\begin{aligned} & d^2(P(\mathbf{x}|\boldsymbol{\theta}), P(\mathbf{x}|\boldsymbol{\theta} + \epsilon\boldsymbol{\Delta})) \\ &= \epsilon^2 \sum_{\alpha\beta} \Delta_\alpha g_{\alpha\beta} \Delta_\beta \end{aligned} \quad (1.19)$$

where  $g_{\alpha\beta}$  is the Fisher information matrix:

$$\begin{aligned} g_{\alpha\beta}(\boldsymbol{\theta}) &= \left\langle \frac{\partial \log P(\mathbf{x}|\boldsymbol{\theta})}{\partial \theta_\alpha} \frac{\partial \log P(\mathbf{x}|\boldsymbol{\theta})}{\partial \theta_\beta} \right\rangle_{\mathbf{x}} \quad (1.20) \\ &= \int P(\mathbf{x}|\boldsymbol{\theta}) \frac{\partial \log P(\mathbf{x}|\boldsymbol{\theta})}{\partial \theta_\alpha} \frac{\partial \log P(\mathbf{x}|\boldsymbol{\theta})}{\partial \theta_\beta} d\mathbf{x} \end{aligned}$$

The FIM can also be calculated as the second derivative

$$g_{\alpha\beta}(\boldsymbol{\theta}) = - \left\langle \frac{\partial^2 \log P(\mathbf{x}|\boldsymbol{\theta})}{\partial \theta_\alpha \partial \theta_\beta} \right\rangle_{\mathbf{x}} \quad (1.21)$$

<sup>13</sup>This exercise was developed in collaboration with Colin Clement and Katherine Quinn.

<sup>14</sup>That is, if one can exchange the derivatives with the integral. That this is not always allowed is shown by the distribution  $P(\mathbf{x}|L)$  of ideal gas particles in a piston, varying the length  $L = \theta$ .

in cases where<sup>14</sup>  $\int d\mathbf{x} \partial^2 P(\mathbf{x}|\boldsymbol{\theta})/\partial\theta_\alpha\partial\theta_\beta = \partial^2(\int d\mathbf{x}P(\mathbf{x}|\boldsymbol{\theta}))/\partial\theta_\alpha\partial\theta_\beta = \partial^2(1)/\partial\theta_\alpha\partial\theta_\beta = 0$ .

(a) Show that eqn 1.21 equals eqn 1.20, given that one can exchange the derivatives with the integral as above. (Hint: Expand eqn 1.21 into derivatives of  $P$ . Which term is eqn 1.20? Is the other term zero?)

Consider a set of  $N$  measurements  $x_n$  of a quantity that is expected to have a Gaussian distribution

$$P(\{x_n\}|\mu, \sigma) = \prod_{n=1}^N \frac{1}{\sqrt{2\pi}\sigma} e^{-(x_n - \mu)^2/(2\sigma^2)}, \quad (1.22)$$

with unknown mean  $\mu$  and standard deviation  $\sigma$ . These could be heights for  $N$  women (Exercise 1.13), or the volume enclosed by a piston at constant pressure (Exercise 6.23), ...

(b) Show that the  $2 \times 2$  FIM is<sup>15</sup>

$$g = \begin{pmatrix} g_{\mu\mu} & g_{\mu\sigma} \\ g_{\sigma\mu} & g_{\sigma\sigma} \end{pmatrix} = \begin{pmatrix} N/\sigma^2 & 0 \\ 0 & 2N/\sigma^2 \end{pmatrix} \quad (1.23)$$

for the distribution given by eqn 1.22. (Hint: Use the second form for the FIM, eqn 1.21. Remember to take the expectation value.)

The FIM is the natural metric tensor on the manifold of probability distributions – the natural distance between nearby distributions as the parameters are varied (eqn 1.19). Why do we say this?

In Exercise 1.14, we introduced the FIM in the context of least-squares fits to data.<sup>16</sup> There we showed, to linear order, that the covariance matrix  $\Sigma$  giving the uncertainty of our estimates of parameters is given by the matrix inverse of the FIM,  $\Sigma = g^{-1}$ . In particular, the variance (standard deviation squared)  $\sigma_\alpha^2$  of the uncertainty in parameter  $\theta_\alpha$  is  $\Sigma_{\alpha\alpha}$ .

(c) Suppose we have a model with only one parameter  $\theta_0$ , so  $g$  and  $\Sigma$  are  $1 \times 1$  matrices. Use eqn 1.19 to estimate (to linear order) how much of a change  $\epsilon\Delta_0$  one would need to move a distance  $\epsilon$  in the space of probability distributions. Show that  $\Delta_0$  is equal to  $\sqrt{\Sigma}$ .

Thus, to linear order, the distance one must move in probability space to be distinguishable (the uncertainty in that direction) is one. Be warned that the calculation is more complex for multiple parameters: see Exercise 1.14 part (d).

Our linearized calculations are backed up by a rigorous theorem, the *Cramér–Rao bound*. This states that the elements of the covariance matrix  $\Sigma$  of parameter uncertainties estimated from a sampling of a probability distribution is bounded below by the corresponding element of the inverse FIM

$$\Sigma_{\alpha\alpha} \geq (g^{-1})_{\alpha\alpha}. \quad (1.24)$$

if the estimator is *unbiased* (see Exercise 1.13).

(d) Use the metric tensor of part (b) to calculate the the Cramér–Rao bound on the precision one can measure the mean and standard deviation of the heights of US women from a single (uncorrelated) sample of  $N = 40$  women.

In Exercise 1.16, we shall examine *global* measures of distance or distinguishability between potentially quite different probability distributions. There we shall show that three of these measures (the Kullback–Liebler divergence, the Hellinger distance, and the Bhattacharyya divergence) all reduce to the FIM to lowest order in the change in parameters. In Exercise 6.23 we shall write 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.

### 1.16 Distances in probability space.<sup>17</sup> (Statistics, Mathematics, Information Geometry) ③

Statistical mechanics studies the fluctuations in behavior expected given the external conditions. Fluctuations in energy are related to the temperature (eqn 6.14); fluctuations in volume are related to pressure (Chapter 10). Statistics is concerned with describing 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. Exercise 1.15 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 predicts an ensemble probability distribution  $P(\mathbf{x}|\boldsymbol{\theta})$  for data  $\mathbf{x}$  (die rolls, particle positions and momenta, heights, ...). It introduced

<sup>15</sup>This metric turns out to have a *constant negative curvature*.

<sup>16</sup>In Exercise 1.16 you will demonstrate that these definitions agree.

<sup>17</sup>This exercise was developed in collaboration with Katherine Quinn.

the Fisher information metric (FIM, eqn 1.21)

$$g_{\mu\nu}(\boldsymbol{\theta}) = - \left\langle \frac{\partial^2 \log(P(\mathbf{x}))}{\partial \theta_\alpha \partial \theta_\beta} \right\rangle_{\mathbf{x}} \quad (1.25)$$

which gives the distance between probability distributions for nearby sets of parameters  $d^2(P(\boldsymbol{\theta}), P(\boldsymbol{\theta} + \epsilon \Delta)) = \epsilon^2 \sum_{\mu\nu} \Delta_\mu g_{\mu\nu} \Delta_\nu$ . 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 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.

Here we shall explore ways of measuring distance or distinguishability between distant probability distributions. There are several measures in common use, of which we will describe four – the least-squares metric in data space, the Kullback–Liebler divergence, the Hellinger distance, and the Bhattacharyya ‘distance’. Each has its uses. The least-squares metric is only useful for multivariate Gaussian models. The Kullback–Liebler divergence is not symmetric. The Hellinger distance becomes less and less useful as the amount of information about the parameters becomes large. The Bhattacharyya ‘distance’ nicely generalizes the least-squares metric to arbitrary models, but violates the triangle inequality and embeds the manifold of predictions into a space with Minkowski-style time-like directions [42].

The first distance measure we discuss is used for models which make predictions  $\mathbf{y}(\boldsymbol{\theta}) = \{y_i(\boldsymbol{\theta})\}$  for data points data points  $\mathbf{x} = \{x_i\}$  (For example, as in Exercise 1.14,  $x_i$  could be the number of Geiger counter clicks for a collection of radioactive elements near time  $t_i$ , and  $y_i$  be a sum of exponential decays depending on decay rates and amplitudes given by parameters  $\theta_\alpha$ .) Let us assume there are  $N$  such data points with independent Gaussian experimental uncertainties  $\sigma$ . Thus the probability that  $\mathbf{y}(\boldsymbol{\theta})$  would yield experimental data  $\mathbf{x}$  is

$$P_{\text{LS}}(\mathbf{x}|\boldsymbol{\theta}) = \frac{e^{-\sum_i (y_i(\boldsymbol{\theta}) - x_i)^2 / 2\sigma^2}}{(2\pi\sigma^2)^{N/2}}, \quad (1.26)$$

a function of the *squared distance* between prediction and experiment in data space.

(a) Use the second definition (eqn 1.21) to calculate the FIM  $g_{\alpha\beta}$  for the least-squares probability  $P_{\text{LS}}$ . Show that it is  $J^T J$ , where  $J_{i\alpha} =$

$\partial y_i / \partial \theta_\alpha / \sigma$  (see Exercise 1.14, part d). Show that the infinitesimal squared distance  $ds^2 = g_{\alpha\beta} d\theta_\alpha d\theta_\beta = \sum_i dy_i^2 / \sigma^2$ , proportional to the infinitesimal squared distance between the predictions in data space. Thus one can choose the distance  $d_{\text{LS}}(P(\boldsymbol{\theta}_1), P(\boldsymbol{\theta}_2))$  between two least-squares probability distributions to be proportional to the Euclidean distance between their predictions in data space.

Now let us turn to the more general, but problematic measures of distinguishability. 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 part (a), because the distances between points on the surface of model predictions is the Euclidean distance between the predictions in data space.

The second distance-like measure we discuss is the *Kullback–Leibler divergence* from  $Q$  to  $P$ .

$$d_{\text{KL}}(Q|P) = - \int d\mathbf{x} P(\mathbf{x}) \log(Q(\mathbf{x})/P(\mathbf{x})). \quad (1.27)$$

(b) Show that the Kullback–Liebler divergence is positive, zero only if  $P = Q$ , but is not symmetric. Show that, to quadratic order in  $\epsilon$  in eqn 1.19, that the Kullback–Liebler divergence does lead to the Fisher information metric.

The third 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*  $P \cdot Q = \sum_{j=1}^6 \sqrt{P_j} \sqrt{Q_j} = \int d\mathbf{x} \sqrt{P(\mathbf{x})} \sqrt{Q(\mathbf{x})}$ . The Hellinger distance is then given by the squared

distance between points on the unit sphere:<sup>18</sup>

$$\begin{aligned} d_{\text{Hel}}^2(P, Q) &= (P - Q)^2 = 2 - 2P \cdot Q \\ &= \int d\mathbf{x} \left( \sqrt{P(\mathbf{x})} - \sqrt{Q(\mathbf{x})} \right)^2 \end{aligned} \quad (1.28)$$

(c) 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 Fisher information metric 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  $Q_j = \{1/10, 1/10, \dots, 1/2\}$  that favors rolling 6? 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  $\sqrt{2}$ .

(d) 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  $\mathbf{j} = \{j_1, \dots, j_N\}$ . Show that

$$P_N \cdot Q_N = (P \cdot Q)^N, \quad (1.29)$$

and hence

$$d_{\text{Hel}}^2(P_N, Q_N) = 1 - (P \cdot Q)^N \quad (1.30)$$

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.<sup>19</sup>

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 final measure, Bhattacharyya's, turns out to be an *intensive* measure of distance [42]; it does not change with system size (particles, rolls, or measurements).

The Bhattacharyya distance can be derived from a limit of the Hellinger distance as the number of data points  $N$  goes to zero:

$$\begin{aligned} d_{\text{Bhatt}}^2(P, Q) &= \lim_{N \rightarrow 0} \frac{1}{2} d_{\text{Hel}}^2(P_N, Q_N)/N \\ &= -\log(P \cdot Q) \\ &= -\log \left( \sum_{\mathbf{x}} \sqrt{P(\mathbf{x})} \sqrt{Q(\mathbf{x})} \right). \end{aligned} \quad (1.31)$$

We sometimes say that we calculate the behavior of  $N$  replicas of the system, and then take  $N \rightarrow 0$ . 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  $N \rightarrow 0$ .

(e) Derive eqn 1.31. (Hint:  $Z^N \approx \exp(N \log Z) \approx 1 + N \log Z$  for small  $N$ .) Show that the Bhattacharyya distance does not satisfy the triangle inequality. (Hint: Find two distributions of loaded dice which do not overlap (so  $P \cdot Q = 0$ ), but both overlap with a third.) Show that it does satisfy the other conditions for a distance. Show, for the nonlinear least-squares model of part (a), that the Bhattacharyya distance equals the distance in data space between the two predictions.

Thus the Bhattacharyya distance is the natural generalization of the ‘distance in data space’ for least-squares models.

<sup>18</sup>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  $\sqrt{2}$ .

<sup>19</sup>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.

## Chapter 2: Random walks and emergent properties.

### Exercises

#### 2.1 Random walks in grade space. (Statistics)

Many students complain about multiple-choice exams, saying that it is easy to get a bad grade just by being unlucky in what questions get asked. While easy and unambiguous to grade, are they good measures of knowledge and skill?

A course is graded using multiple choice exams, with ten points for each problem. A particular student has a probability 0.7 of getting each question correct.

(a) Generalize the coin-flip random walk discussion (near eqn 2.3), to calculate the student's mean score  $\langle s_N \rangle$ , the mean square score  $\langle s_N^2 \rangle$ , and the standard deviation  $\sigma_N = \sqrt{\langle (s_N - \langle s_N \rangle)^2 \rangle}$  for a test with  $N$  questions. (Note that  $\langle s_{N-1} \ell_n \rangle \neq 0$  in this case.)

Measuring instruments in physics are often characterized by the signal-to-noise ratio. One can view the random choice of questions as a source of noise, and your calculation in part (a) as an estimate of that noise.

(b) An introductory engineering physics exam with ten ten-point multiple choice questions has a class mean of 70 and a standard deviation of 15. How much of the standard deviation is attributable to random chance (noise), as you calculate in part (a)? If the remainder of the variation between people is considered the signal, is it a good test, comparing the noise to the signal?

#### 2.2 Photon diffusion in the Sun. (Astrophysics)

If fusion in the Sun turned off today, how long would it take for us to notice? This question became urgent some time back when the search for solar neutrinos failed.<sup>20</sup> Neutrinos, created in the same fusion reaction that creates heat in the Solar core, pass through the Sun at near the speed of light without scattering – giving us a current status report. The rest of the energy takes longer to get out.

Most of the fusion energy generated by the Sun is produced near its center. The Sun is  $7 \cdot 10^5$  km

in radius. Convection probably dominates heat transport in approximately the outer third of the Sun, but it is believed that energy is transported through the inner portions (say to a radius  $R = 5 \cdot 10^8$  m) through a random walk of X-ray photons. (A photon is a quantized package of energy; you may view it as a particle which always moves at the speed of light  $c$ . Ignore for this exercise the index of refraction of the Sun.)

There are a range of estimates for the mean free path for a photon in the Sun. For our purposes, assume photons travel at the speed of light, but bounce in random directions (without pausing) with a step size of  $\ell = 0.1\text{cm} = 10^{-3}\text{m}$ .

About how many random steps  $N$  will the photon take of length  $\ell$  to get to the radius  $R$  where convection becomes important? About how many years  $\Delta t$  will it take for the photon to get there? Related formulæ:  $c = 3 \cdot 10^8 \text{ m/s}$ ;  $\langle x^2 \rangle \sim 2Dt$ ;  $\langle s_n^2 \rangle = n\sigma^2 = n\langle s_1^2 \rangle$ . There are  $31\,556\,925.9747 \sim \pi \cdot 10^7 \sim 3 \cdot 10^7$  s in a year.

#### 2.14 Modified random walk.

Random walks with a constant drift term will have a net correlation between steps. This problem can be reduced to the problem without drift by shifting to a ‘moving reference frame’. In particular, suppose we have a random walk with steps independently drawn from a uniform density  $\rho(\ell)$  on  $[0,1]$ , but with a non-zero mean  $\langle \ell \rangle = \bar{\ell} \neq 0$ .

Argue that the sums  $s'_N = \sum_n^N (\ell_n - \bar{\ell})$  describe random walks in a moving reference frame, with zero mean. Argue that the variance of these random walks (the squared standard deviation) is the same as the variance  $\langle (s_N - \bar{s}_N)^2 \rangle$  of the original random walks.

#### 2.15 Modified diffusion.

Photons diffusing in clouds are occasionally absorbed by the water droplets. Neutrons diffusing in a reactor, or algae diffusing in the sea, may multiply as they move.

<sup>20</sup>The missing electron neutrinos, as it happened, ‘oscillated’ into other types of neutrinos.

*How would you modify the derivation of the diffusion equation in eqns (2.9–2.12) to allow for particle non-conservation? Which equation in 2.9 should change? What would the new term in the diffusion equation look like?*

**2.16 Density dependent diffusion.** ⓘ

The diffusion constant can be density dependent; for example, proteins diffusing in a cell membrane are so crowded they can get in the way of one another.

*What should the diffusion equation be for a conserved particle density  $\rho$  diffusing with diffusion constant  $D(\rho)$ ? (Hint: See footnote 20, p. 31.)*

**2.17 Local conservation.** Ⓜ

Tin is deposited on a surface of niobium. At high temperatures, the niobium atoms invade into the tin layer. *Is the number of niobium atoms  $\rho_{Nb}(\mathbf{x})$  a locally conserved quantity?* Politicians in ‘red states’ (primarily Republican) are concerned about migration of democrats into the Sun Belt (the southwest). *Is the number of democrats  $\rho_{Dem}(\mathbf{x})$  locally conserved?*

**2.18 Absorbing boundary conditions.** Ⓜ

A particle starting at  $x'$  diffuses on the positive  $x$  axis for a time  $t$ , except that whenever it hits the origin it is absorbed. The resulting probability density gives the Green’s function  $\rho(x, t) = G(x|x', t) = \int G(x|x', t)\rho(x', 0)dx'$ .

*Solve for  $G$ . (Hint: Use the ‘method of images’: add a negative  $\delta$  function at  $-x'$ .)*

**2.19 Run and tumble.** (Active matter, Biology) Ⓝ

Purcell, in an essay *Life at low Reynolds number* [41], describes the strange physical world at the scale of bacteria.<sup>21</sup>

The bacterium E. coli swims using roughly five corkscrew-shaped propellers called *flagella*, which spin at 100 revolutions per second. These propellers mesh nicely into a bundle when they rotate counterclockwise, causing the bacterium to *run* forward. But when they rotate clockwise, the bundle flies apart and the bacterium *tumbles*. Assume that during a tumble the bacterium does not change position, and after a tumble it is pointed in a random direction. Pretend the runs are of fixed duration  $T \approx 1\text{sec}$  and speed  $V \approx 20\mu\text{m/sec}$ , and they alternate with tumbles

of duration  $\tau \approx 0.1\text{sec}$ . (Real cells shift from runs to tumbles with a continuous distribution of run times, and do not completely scramble their orientation after a tumble, Exercise 2.20)

(a) *What is the mean-square distance  $\langle \mathbf{r}^2(t) \rangle$  moved by our bacterium after a time  $t = N(T+\tau)$ , in terms of  $V$ ,  $T$ ,  $t$ , and  $\tau$ ? What is the formula for the diffusion constant?* (Hint: Be careful; your formula for the diffusion constant should depend on the fact that the diffusion is in three dimensions.)

Purcell tells us that the cell does not need to swim to get to new food after it has exhausted the local supply. Instead, it can just wait for food molecules to diffuse to it, with a rate he says is  $4\pi aND$  food molecules per second. Here  $a$  is the radius of the cell,  $N$  is the food concentration at infinity, and  $D \approx 10^{-9}\text{m}^2/\text{s}$  is the food diffusion constant.

(b) *Assume the food is eaten by the bacterium with perfect efficiency at the sphere of radius  $a$ . Solve the diffusion equation for the density of food molecules, and confirm Purcell’s formula for the rate at which food is eaten.* (Hint: You may want to use the Laplacian in spherical coordinates. For a function  $\rho(\mathbf{r}) = \rho(r)$  with spherical symmetry,  $\nabla^2\rho = (1/r^2)\partial(r^2\partial\rho/\partial r)/\partial r$ .)

The cell lives in an environment which varies in space. It swims to move toward regions with higher concentrations of food, and lower concentrations of poisons (a behavior called *chemotaxis*). Bacteria are too small to sense the concentration gradient from one side of the cell to the other. The run-and-tumble strategy is designed to move them far enough to tell if things are getting better. In particular, the cells run for longer times when things are getting better (but not shorter when things are getting worse).

(c) *Model chemotaxis with a one-dimensional run-and-tumble model along a coordinate  $x$ . The velocity  $\pm V$  is chosen with equal probability at each tumble, with the same velocity and tumble time  $\tau$  as above. But now the duration  $T_+$  of runs in the positive  $x$  direction is larger than the duration  $T$  of runs in the negative direction. Compare the run speed  $V$  to the average velocity  $\langle dx/dt \rangle$  of the bacterium toward a better life.*

<sup>21</sup>He also analyzes swimming as *motion generated by a periodic change in shape*. For small, slow things like bacteria where inertia is unimportant, motions like kicking your feet that trace forward and backward along the same path in shape space (‘reciprocal motions’) return you back where you started; spinning a corkscrew works. The net motion is an integral of a *curvature* over the loop in shape space [57].

**2.20 Continuous time walks: Ballistic to diffusive.**<sup>22</sup> (3)

Random walks at long times diffuse, with  $\langle x^2 \rangle \propto Dt$ , with a proportionality constant that depends on dimension. But what happens at short times? In many cases, the motion at short times is *ballistic* – straight line motion with constant speed. Perfume molecules move in straight lines between collisions with molecules in the air. E-coli and other bacteria alternate between runs and tumbles (Exercise 2.19). Electrons<sup>23</sup> in semiconductors can have long mean-free paths between collisions, and the ballistic to diffusive crossover we study here has practical implications for the behavior of electronic devices.

Let us consider the trajectory of a single particle moving through vacuum, subject to an external force  $F$  and to collisions at random times. The collision at time  $t_\alpha$  resets the velocity to a value  $v_\alpha$ . Under a constant external force  $F$ , the velocity  $v(t) = v_\Omega + (F/m)(t - t_\Omega)$ , where  $t_\Omega$  is the time of the collision immediately preceding the current time  $t$ . We can write this as an equation:

$$v(t') = \sum_{\alpha} (v_\alpha + (F/m)(t' - t_\alpha)) \Theta(t' - t_\alpha) \Theta(t_{\alpha+1} - t'). \quad (2.32)$$

Here the Heaviside step function  $\Theta$  (zero for negative argument and one for positive argument), is used to select the time interval between collisions. We assume the velocities after each collision have mean zero and are uncorrelated with one another, with mean-square velocity  $\sigma_v^2$ , so

$$\langle v_\alpha \rangle = 0, \quad (2.33)$$

$$\langle v_\alpha v_\beta \rangle = \sigma_v^2 \delta_{\alpha\beta}. \quad (2.34)$$

We assume that the scattering events happen at random with an average scattering time  $\Delta t$ , so the intervals  $\delta_\alpha = t_{\alpha+1} - t_\alpha$  have an exponential probability distribution  $\rho(\delta) = \exp(-\delta/\Delta t)/\Delta t$ . (Remember that for random events with rate  $1/\Delta t$ , the gap-averaged probability of the interval length is  $\rho(\delta)$ , the waiting time  $\delta$  to the next collision averaged over the current time  $t$  is  $\rho(\delta)$ , and the time  $\delta$  since the last collision is also distributed by  $\rho(\delta)$ ; see Exercise 1.3).

<sup>22</sup>Feynman has a nice discussion of continuous-time diffusion [19, I.43].

<sup>23</sup>More precisely, it is the electron and hole *quasiparticles* which have large mean-free paths (see footnote 23 on page 180). The bare electrons scatter rapidly with their screening clouds, but the electrons plus their clouds can move long distances ballistically.

Let us first calculate the mobility  $\gamma$ . Again, let the collision immediately before the current time  $t$  be labeled by  $\Omega$ .

(a) Show that the expectation value of the velocity  $\langle v(t) \rangle$  is given by  $F/m$  times the mean time  $\langle t - t_\Omega \rangle$  since the last collision. Your answer should use eqns 2.32 and 2.33, and the probability distribution  $\rho(\delta)$ . Calculate the mobility in terms of  $\Delta t$  and  $m$ . Why is your answer different from the calculation for fixed intervals  $\Delta t$  considered in Section 2.3? (See Exercise 1.3 and Feynman's discussion in [19, I.43].)

Now let us explore the crossover from ballistic to diffusive motion, in the absence of an external force. Let  $r(t)$  be the distance moved by our random walk since time zero. At times much less than  $\Delta t$ , the particle will not yet have scattered, and the motion is in a straight line. At long times we expect diffusive motion.

How does the crossover happen between these two limits? In the next part we calculate  $\langle dr^2/dt \rangle$  as a function of time  $t$  for a continuous-time random walk starting at  $t = 0$ , in the absence of an external force, giving us the crossover and the diffusion constant  $D$ . Assume the random walk has been traveling since  $t = -\infty$ , but that we measure the distance traveled since  $t = 0$ .

(b) Calculate  $\langle dr^2/dt \rangle$  as a function of time, where  $r(t) = x(t) - x(0)$  is the distance moved since  $t = 0$ . To avoid getting buried in algebra, we ask you to do this in a particular sequence. (1) Write  $r$  as an abstract integral over  $v$ . Square it, and differentiate with respect to time. (2) Substitute eqn 2.32 with  $F = 0$ , and use eqn 2.34 to ensemble average over the collision velocities. (Hint: Your formula at this point should only involve the last collision time  $t_\Omega$ , and should be an integral from 0 to  $t$  involving a  $\Theta$  function.) (3) Then take the ensemble average over the collision times to get  $\langle dr^2/dt \rangle$ . What is the diffusion constant  $D$ ? Does it satisfy the relation  $D/\gamma = m\bar{v}^2$  of footnote 20 on page 22? (Hint: You may end up with a double integral over  $t_\Omega$  and  $t'$ . Examine the integration region in the  $(t_\Omega, t')$  plane, and exchange orders of integration.) The crossover between ballistic motion and diffusion is easily found in the literature, and is a

decaying exponential:

$$\langle dr^2/dt \rangle = 2D(1 - \exp(-t/\Delta t)). \quad (2.35)$$

(c) Use this to check your answer in part (b). Integrate, and provide a log-log plot of  $\sqrt{\langle r^2 \rangle}/D\Delta t$  versus  $t/\Delta t$ , over the range  $0.1 < t/\Delta t < 100$ . Solve for the asymptotic power-law dependence at short and long times, and plot them as straight lines on your log-log plot. Label your axes. (Are the power laws tangent to your curves as  $t \rightarrow 0$  and  $t \rightarrow \infty$ ?)

### 2.21 Random walks and generating functions.<sup>24</sup> (Mathematics) ③

Consider a one-dimensional random walk with step-size  $\pm 1$  starting at the origin. What is the probability  $f_t$  that it *first* returns to the origin at  $t$  steps?<sup>25</sup>

(a) Argue that the probability is zero unless  $t = 2m$  is even. How many total paths are there of length  $2m$ ? Calculate the probability for  $f_{2m}$  for up to eight-step hops ( $m < 5$ ) by drawing the different paths that touch the origin only at their endpoints. (Hint: You can save paper by drawing the paths starting to the right, and multiplying by two. Check your answer by comparing to the results for general  $m$  below.)

This *first return* problem is well-studied, and is usually solved using a *generating function*. Gen-

erating functions extend a series (here  $f_{2m}$ ) into a function. The generating function for our one-dimensional random walk first return problem, it turns out, is

$$\begin{aligned} F(x) &= \sum_{m=0}^{\infty} f_{2m} x^m = 1 - \sqrt{1-x} \\ &= \sum_m \frac{2^{-2m}}{2m-1} \binom{2m}{m} x^m. \end{aligned}$$

(Look up the derivation if you want an example of how it is done.)

(b) Evaluate the probability of returning to the origin for the first time at  $t = 20$  steps ( $m = 10$ ). Use Stirling's formula  $n! \sim \sqrt{2\pi n}(n/e)^n$  to give the probability  $f_{2m}$  of first returning at time  $t = 2m$  for large  $t$ . (Note that the rate per unit time  $P_{\text{hop}}(t) = f_{2m}/2$ .)

There are lots of other uses for generating functions: analyzing partial sums, finding moments, taking convolutions...

(c) Argue that  $F(1)$  is the probability that a particle will eventually return to the origin, and that  $F'(1)$  is  $\langle m \rangle$ , half the expected time to return to the origin. What is the probability that our one-dimensional walk will return to the origin? What is the mean time to return to the origin?

<sup>24</sup>This problem was created with the assistance of Archishman Raju.

<sup>25</sup>The continuum limit of this problem is discussed in Exercise 2.18, which uses the method of images to solve for the Green's function for a diffusion equation with an absorbing boundary condition.

## Chapter 3: Temperature and equilibrium

### Exercises

#### 3.2 Large and very large numbers. *i*

The numbers that arise in statistical mechanics can defeat your calculator. A googol is  $10^{100}$  (one with a hundred zeros after it). A googolplex is  $10^{\text{googol}}$ .<sup>26</sup>

Consider a monatomic ideal gas with one mole of particles ( $N = \text{Avogadro's number}, 6.02 \times 10^{23}$ ), room temperature  $T = 300\text{ K}$ , and volume  $V = 22.4\text{ liters}$  (at atmospheric pressure).

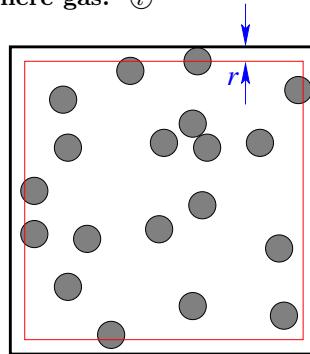
(a) Which of the properties ( $S$ ,  $T$ ,  $E$ , and  $\Omega(E)$ ) of our gas sample are larger than a googol? A googolplex? Does it matter what units you use, within reason?

If you double the size of a large equilibrium system (say, by taking two copies and weakly coupling them), some properties will be roughly unchanged; these are called *intensive*. Some, like the number  $N$  of particles, will roughly double; they are called *extensive*. Some will grow much faster than the size of the system.

(b) To which category (intensive, extensive, faster) does each property from part (a) belong?

For a large system of  $N$  particles, one can usually ignore terms which add a constant independent of  $N$  to extensive quantities. (Adding  $2\pi$  to  $10^{23}$  does not change it enough to matter.) For properties which grow even faster, overall *multiplicative* factors often are physically unimportant.

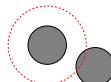
#### 3.5 Hard sphere gas. *i*



**Fig. 3.8 Hard sphere gas.**

We can improve on the realism of the ideal gas by giving the atoms a small radius. If we make the potential energy infinite inside this radius (*hard spheres*), the potential energy is simple (zero unless the spheres overlap, which is forbidden). Let us do this in two dimensions; three dimensions is no more complicated, but slightly harder to visualize.

A two-dimensional  $L \times L$  box with hard walls contains a gas of  $N$  distinguishable<sup>27</sup> hard disks of radius  $r \ll L$  (Fig. 3.8). The disks are dilute; the summed area  $N\pi r^2 \ll L^2$ . Let  $A$  be the effective area allowed for the disks in the box (Fig. 3.8):  $A = (L - 2r)^2$ .



**Fig. 3.9 Excluded area around a hard disk.**

<sup>26</sup>The firm Google is named for the number googol, with a slight spelling change. They named their corporate headquarters the googleplex.

<sup>27</sup>Using indistinguishable particles adds a factor of  $1/N!$  to all the estimates for  $\Omega$ , the volume in configuration space, and  $-k_B \log(N!)$   $\approx -k_B(N \log N - N)$  to all the entropy estimates. It does not change the pressure, and does make the calculation slightly less illuminating.

<sup>28</sup>Again, ignore small corrections when the excluded region around one disk overlaps the excluded regions around other disks, or near the walls of the box.

(a) The area allowed for the second disk is  $A - \pi(2r)^2$  (Fig. 3.9), ignoring the small correction when the excluded region around the first disk overlaps the excluded region near the walls of the box. What is the allowed  $2N$ -dimensional volume in configuration space<sup>28</sup>  $\Omega_{\text{HD}}^{\mathbb{Q}}$  of allowed zero-energy configurations of hard disks, in this dilute limit? Leave your answer as a product of  $N$  terms.

Our formula in part (a) expresses  $\Omega_{\text{HD}}^{\mathbb{Q}}$  strangely, with each disk in the product only feeling the excluded area from the former disks. For large numbers of disks and small densities, we can rewrite  $\Omega_{\text{HD}}^{\mathbb{Q}}$  more symmetrically, with each disk feeling the same excluded area  $A_{\text{excl}}$ .

(b) Write  $\log(\Omega_{\text{HD}}^{\mathbb{Q}})$  as a sum over the number of disks. Use  $\log(1 - \epsilon) \approx -\epsilon$  and<sup>29</sup>  $\sum_{m=0}^{N-1} m = N(N-1)/2 \approx N^2/2$  to approximate  $\log(\Omega_{\text{HD}}^{\mathbb{Q}}) \approx N \log A - N\epsilon$ , solving for  $\epsilon$  and evaluating any sums over disks. Then use  $-\epsilon \approx \log(1-\epsilon)$  to write  $\log(\Omega_{\text{HD}}^{\mathbb{Q}})$  in terms of  $A - A_{\text{excl}}$ . Interpret your formula for the excluded area  $A_{\text{excl}}$ , in terms of the range of excluded areas you found in part (a) as you added disks.

(c) Find the pressure for the hard-disk gas in the large  $N$  approximation of part (b). Does it reduce to the ideal gas law for  $b = 0$ ? (Hint: Constant energy is the same as constant temperature for hard particles, since the potential energy is zero.)

### 3.10 Triple product relation. (Thermodynamics, Mathematics) ⑦

In traditional thermodynamics, there are many useful formulas like

$$dE = T dS - P dV, \quad (3.36)$$

(see Section 6.4 and the inside front cover of this text). For example, if  $V$  is held constant (and hence  $dV = 0$  then  $dE = T dS$  from eqn 3.36), giving  $(\partial S / \partial E)|_V = 1/T$  (the definition of temperature, eqn 3.29).

(a) Use eqn 3.36 to rederive the traditional formula for the pressure  $P$ .

Let us consider a general formula of this type,

$$A dx + B dy + C df = 0. \quad (3.37)$$

<sup>29</sup>If this formula is not familiar, you can check the exact formula for the first few  $N$ , or convert the sum to an integral  $\int_0^{N-1} mdm \approx \int_0^N mdm = N^2/2$ .

<sup>30</sup>They are really *differential forms*, which are mathematically subtle (see note 23 on p. 142).

<sup>31</sup>One can derive the formula by solving  $S = S(N, V, E)$  for  $E$ . It is the same surface in four dimensions as  $S(N, V, E)$  (Fig. 3.4) with a different direction pointing ‘up’.

(b) What is  $(\partial f / \partial x)|_y$ ?  $(\partial f / \partial y)|_x$ ?  $(\partial x / \partial y)|_f$ ? Use these to derive the triple product relation eqn 3.33,  $(\partial x / \partial y)|_f (\partial y / \partial f)|_x (\partial f / \partial x)|_y = -1$ .

The author has always been uncomfortable with manipulating dXs.<sup>30</sup> How can we derive these relations geometrically, with traditional partial derivatives? Our equation of state  $S(E, V, N)$  at fixed  $N$  is a surface embedded in three dimensions. Figure 3.4 shows a triangle on this surface, which we can use to derive the general triple-product relation between partial derivatives.

(b) Derive the triple product relation  $(\partial f / \partial x)|_y (\partial x / \partial y)|_f (\partial y / \partial f)|_x = -1$ . (Hint: Consider the triangular path in Fig. 3.4, viewing  $f = S$ ,  $x = E$ , and  $y = V$ , so the surface is  $S(E, V)$ . The first side starts at the lower right at  $(E_0, V_0, S_0)$  and moves along the hypotenuse at constant  $S$  to  $V_0 + \Delta V$ . The resulting vertex at the upper left will thus be at  $(E_0 + (\partial E / \partial V)|_S \Delta V, V_0 + \Delta V, S_0)$ . The second side runs at constant  $E$  back to  $V = V_0$ , and the third side runs at constant  $V$  back to  $(E_0, V_0, S_0)$ . The curve must close to make  $S$  a single-valued function; the resulting equation should imply the triple-product relation.)

### 3.11 Maxwell relations. (Thermodynamics) ⑧

Consider the microcanonical formula for the equilibrium energy  $E(S, V, N)$  of some general system.<sup>31</sup> One knows that the second derivatives of  $E$  are symmetric; at fixed  $N$ , we get the same answer whichever order we take partial derivatives with respect to  $S$  and  $V$ .

Use this to show the Maxwell relation

$$\left. \frac{\partial T}{\partial V} \right|_{S,N} = - \left. \frac{\partial P}{\partial S} \right|_{V,N}. \quad (3.38)$$

(This should take two lines of calculus or less.) Generate two other similar formulae by taking

other second partial derivatives of  $E$ . There are many of these relations [22].

**3.13 Weirdness in high dimensions.** p

We saw in momentum space that most of the surface area of a high-dimensional sphere is along the equator. Consider the volume of a high-dimensional sphere.

*Is most of the volume near the center, or the surface? How might this relate to statistical mechanics treatments of momentum space, which in some texts approximate the volume of an energy shell with the volume of the entire sphere?*

**3.14 Entropy maximum and temperature.** p

*Explain in words why, for two weakly coupled systems, that eqn 3.23*

$$\rho(E_1) = \Omega_1(E_1)\Omega_2(E - E_1)/\Omega(E) \quad (3.39)$$

*is intuitive for a system where all states of energy  $E$  have equal probability density. Using  $S = k_B \log(\Omega)$ , show in one step that maximizing the probability of  $E_1$  makes the two temperatures  $1/T = \partial S/\partial E$  the same, and hence that maximizing  $\rho(E_1)$  maximizes the total entropy.*

**3.15 Taste and smell and chemical potential.** p

In note 42 on page 60, do we breathe hard because of the chemical potential of carbon dioxide, or because of the concentration of carbon dioxide in our blood? They both go up when you hold your breath, and usually you would say your body is responding to the concentration. The same could be said of smell and taste. Usually the chemical potential goes up monotonically with concentration – how can we guess what is being measured? Smell and taste measure the binding of molecules to certain receptors.

*Suppose a flavor molecule binds tightly to alcohol but weakly to water. At a given concentration, would you expect to taste it less in an alcoholic drink? If chemical potential  $\mu$  is the energy cost of adding a particle to the fluid at constant entropy, would you imagine the change in binding to the receptor will be controlled by  $\mu$ ? Do you taste chemical potential, or concentration?*

Our sensations of taste and particularly smell are more complicated and nonlinear than our sensation of heat and pressure, sound and light. Partly this is due to the gigantic number of molecules being sensed; partly it might be due to the fact that many molecules can compete for the same receptor, and several receptors can sense the same molecule.

**3.16 Undistinguished particles.** p

*Why should we divide the phase-space volume by  $N!$ , whenever we do not keep track of the differences between the  $N$  particles? Look up ‘entropy of mixing’. In Fig. 5.4, can you see how to extract work from the mixing using methods that cannot tell white from black particles? If we had a door in the partition wall that let through only white particles, what would happen? Could we then extract work from the system?*

### 3.17 Random energy model.<sup>32</sup> ④

The nightmare of every optimization algorithm is a random landscape; if every new configuration has an energy uncorrelated with the previous ones, no search method is better than systematically examining every configuration. Finding ground states of disordered systems like spin glasses and random-field models, or equilibrating them at non-zero temperatures, is challenging because the energy landscape has many features that are quite random. The random energy model (REM) is a caricature of these disordered systems, where the correlations are completely ignored. While optimization of a single REM becomes hopeless, we shall see that the study of the ensemble of REM problems is quite fruitful and interesting.

The REM has  $M = 2^N$  states for a system with  $N$  ‘particles’ (like an Ising spin glass with  $N$  spins), each state with a randomly chosen energy. It describes systems in limit when the interactions are so strong and complicated that flipping the state of a single particle completely randomizes the energy. The states of the individual particles then need not be distinguished; we label the states of the entire system by  $j \in \{1, \dots, 2^N\}$ . The energies of these states  $E_j$  are assumed independent, uncorrelated variables with a Gaussian probability distribution

$$P(E) = \frac{1}{\sqrt{\pi N}} e^{-E^2/N} \quad (3.40)$$

of standard deviation  $\sqrt{N/2}$ .

*Microcanonical ensemble.* Consider the states in a small range  $E < E_j < E + \delta E$ . Let the number of such states in this range be  $\Omega(E)\delta E$ .

(a) Calculate the average

$$\langle \Omega(N\epsilon) \rangle_{\text{REM}} \quad (3.41)$$

over the ensemble of REM systems, in terms of the energy per particle  $\epsilon$ . For energies per particle near zero, show that this average density of states grows exponentially as the system size  $N$  grows. In contrast, show that  $\langle \Omega(N\epsilon) \rangle_{\text{REM}}$  decreases exponentially for  $\epsilon = E/N < -\epsilon_*$  and for  $\epsilon > \epsilon_*$ ,

where the limiting energy per particle

$$\epsilon_* = \sqrt{\log 2}. \quad (3.42)$$

(Hint: As  $N$  grows, the probability density  $P(N\epsilon)$  decreases exponentially, while the total number of states  $2^N$  grows exponentially. Which one wins?)

What does an exponentially growing number of states mean? Let the entropy per particle be  $s(\epsilon) = S(N\epsilon)/N$ . Then (setting  $k_B = 1$  for notational convenience)  $\Omega(E) = \exp(S(E)) = \exp(Ns(\epsilon))$  grows exponentially whenever the entropy per particle is positive.

What does an exponentially decaying number of states for  $\epsilon < -\epsilon_*$  mean? It means that, for any particular REM, the likelihood of having *any* states with energy per particle near  $\epsilon$  vanishes rapidly as the number of particles  $N$  grows large. How do we calculate the entropy per particle  $s(\epsilon)$  of a typical REM? Can we just use the annealed<sup>33</sup> average

$$s_{\text{annealed}}(\epsilon) = \lim_{N \rightarrow \infty} (1/N) \log \langle \Omega(E) \rangle_{\text{REM}} \quad (3.43)$$

computed by averaging over the entire ensemble of REMs?

(b) Show that  $s_{\text{annealed}}(\epsilon) = \log 2 - \epsilon^2$ .

If the energy per particle is above  $-\epsilon_*$  (and below  $\epsilon_*$ ), the expected number of states  $\Omega(E)\delta E$  grows exponentially with system size, so the fractional fluctuations become unimportant as  $N \rightarrow \infty$ . The typical entropy will become the annealed entropy. On the other hand, if the energy per particle is below  $-\epsilon_*$ , the number of states in the energy range  $(E, E + \delta E)$  rapidly goes to zero, so the typical entropy  $s(\epsilon)$  goes to minus infinity. (The annealed entropy is not minus infinity because it gets a contribution from exponentially rare REMs that happen to have an energy level far into the tail of the probability distribution.)

Hence

$$\begin{aligned} s(\epsilon) &= s_{\text{annealed}}(\epsilon) = \log 2 - \epsilon^2 & |\epsilon| < \epsilon_* \\ s(\epsilon) &= -\infty & |\epsilon| > \epsilon_*. \end{aligned} \quad (3.44)$$

Notice why these arguments are subtle. Each REM model in principle has a different entropy. For large systems as  $N \rightarrow \infty$ , the entropies of different REMs look more and more similar to one

<sup>32</sup>This exercise draws heavily from [32, chapter 5].

<sup>33</sup>Annealing a disordered system (like an alloy or a disordered metal with frozen-in defects) is done by heating it to allow the defects and disordered regions to reach equilibrium. By averaging  $\Omega(E)$  not only over levels within one REM but also over all REMs, we are computing the result of equilibrating over the disorder—an annealed average.

another<sup>34</sup> (the entropy is self-averaging) whether  $|\epsilon| < \epsilon_*$  or  $|\epsilon| > \epsilon_*$ . However,  $s(\epsilon)$  is not self-averaging for  $|\epsilon| > \epsilon_*$ , so the typical entropy is not given by the ‘annealed’ logarithm  $\langle \Omega(E) \rangle_{\text{REM}}$ . The REM has a *glass transition* at the temperature  $T_c$  corresponding to  $\epsilon_*$ . Above  $T_c$  the entropy is extensive and the REM acts much like an equilibrium system. Below  $T_c$  one can show [32, eqn

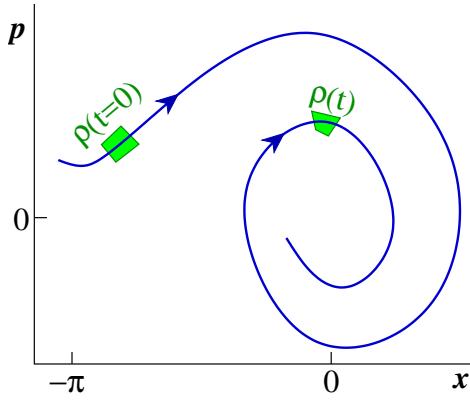
5.25] that the REM thermal population condenses onto a finite number of states (i.e., a number that does not grow as the size of the system increases), which goes to zero linearly as  $T \rightarrow 0$ .

The mathematical structure of the REM also arises in other, quite different contexts, such as combinatorial optimization and random error correcting codes [32, chapter 6].

<sup>34</sup>Mathematically, the entropies per particle of REM models with  $N$  particles approach that given by eqn 3.44 with probability one [32, eqn 5.10].

## Exercises

**4.2 Liouville vs. the damped pendulum.** (Mathematics, Dynamical systems) ⓘ



**Fig. 4.10 Total derivatives.** The total derivative gives the local density as measured by a particle moving with the flow:  $d\rho/dt = d/dt(\rho(x(t), p(t), t))$ . Applying the chain rule gives the definition of the total derivative,  $d\rho/dt = \partial\rho/\partial t + \partial\rho/\partial x\dot{x} + \partial\rho/\partial p\dot{p}$ .

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

$$\begin{aligned}\dot{x} &= p/M, \\ \dot{p} &= -\gamma p - K \sin(x).\end{aligned}\quad (4.45)$$

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

(a) In the steps leading from eqn 4.5 to eqn 4.7, why does Liouville's theorem not apply to the

damped pendulum? More specifically, what are  $\partial\dot{p}/\partial p$  and  $\partial\dot{q}/\partial q$ ?

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

**4.5 No attractors in Hamiltonian systems.** (Dynamical systems, Astrophysics) ⓘ

Planetary motion is one of the prime examples of a Hamiltonian dynamical system. In the three-body problem with one body of small mass, Lagrange showed that there were five configurations (L1, L2, L3, L4, and L5) that rotated as a unit – forming fixed points in a rotating frame of reference. Three of these (L1-L3) are unstable, and the other two are stable. In dissipative dynamical systems, ‘stable fixed point’ implies that neighboring initial conditions will converge to the fixed point at long times.

*Is that true here? What does it mean to be stable, or unstable, in a Hamiltonian system?*

**4.6 Perverse initial conditions.** ⓘ

If we start a gas of classical spherical particles in a square box all in a vertical line, all moving exactly in the vertical direction, they will bounce back and forth vertically forever.

*Does that mean a gas of particles in a square box is not ergodic? Why or why not?*

**4.7 Bayesian priors.** <sup>35</sup> (Statistics) ⓘ

In this exercise, we shall explore an analogy between statistical mechanics and Bayesian statistics. As in Exercise 1.13, we consider the problem of fitting a Gaussian probability distribution to a collection of measurements. (See also Exercise 1.15 for an information-geometry analysis of this same problem.)

Consider the population the heights of women in the United States. Several Web sites quote a mean height of  $\mu_0 = 162\text{cm}$  for US women, but neglect to mention the variance. We will assume

<sup>35</sup>This exercise was developed in collaboration with Colin Clement.

$\mu = \mu_0$  is known, and we would like to estimate the probability distribution of the unknown standard deviation  $\sigma$ , given a single uncorrelated sample of  $N$  women. We know

$$\begin{aligned} P(\{x_n\}|\sigma) &= \prod_{n=1}^N \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x_n - \mu_0)^2/(2\sigma^2)} \\ &= \frac{e^{-\sum_{n=1}^N (x_n - \mu_0)^2/2\sigma^2}}{\sqrt{2\pi\sigma^2}^N} \\ &= (2\pi\sigma^2)^{-N/2} \exp(-S_N/2\sigma^2), \end{aligned} \quad (4.46)$$

where the value  $S_N = \sum_{n=1}^N (x_n - \mu_0)^2$  provides sufficient statistics for estimating  $\sigma$ .

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 ( $\theta = (\sigma)$ ), 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  $(\sigma, \sigma + \Delta\sigma)$ ?

In Bayesian statistics, we estimate the probability of a given set of parameters  $\theta$  given data  $\mathbf{d}$  by using Bayes' theorem (see Exercise 6.14):

$$P(\theta|\mathbf{d}) = P(\mathbf{d}|\theta)P(\theta)/P(\mathbf{d}). \quad (4.47)$$

with  $P(\{x_n\}|\sigma)$  from eqn 4.46. Here the probability of the data,  $P(\mathbf{d})$ , is independent of the parameters and basically acts to normalize  $P(\theta|\mathbf{d})$  to one. The probability density  $P(\theta)$  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  $P(\theta|\mathbf{d})$  is analogous to the Boltzmann factor  $\exp(-\mathcal{H}/k_B T)/Z$  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 [44]) using the same techniques we discuss in Chapter 8.

But what is the prior  $P(\theta)$ ? 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  $\alpha$  of particular interest.

- $\alpha = 0$ , the 'uniform prior' for  $\sigma$  where every interval  $(\sigma, \sigma + \Delta\sigma)$  is equally likely.
- A value for  $\alpha$ , where every interval  $(v, v + \Delta v)$  in the variance  $\sigma^2$  is equally likely (uniform prior for  $\sigma^2$ ).
- 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  $\mu_0$ . Suppose for simplicity that in each case their sample gave the population average<sup>36</sup>  $S_N = N\sigma_{\text{pop}}^2$ .

(a) Plot  $P_0(\sigma|S_N)$  versus  $\sigma/\sigma_{\text{pop}}$  for these three samples  $N = 4, 40$ , and  $400$ , assuming uniform prior  $\alpha = 0$  for  $\sigma$  and using  $S_N = N\sigma_{\text{pop}}^2$ . The normalization ( $P(\mathbf{d})$  in eqn 4.47) can be computed either numerically, or analytically in terms of  $\Gamma(z) = \int_0^\infty x^{z-1} \exp(-x) dx$ . How does the maximum likelihood  $\sigma_{\text{ML}}$ , where  $P_0(\sigma_{\text{ML}})$  is maximum, vary with  $N$ ? Is it biased, compared to the naive estimate  $\sigma_{\text{pop}}$ ? Finally, explain why the curve appears so asymmetric for small  $N$ . Is the average  $\sigma$  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.

<sup>36</sup>The standard deviation of women's heights in the US turns out to be about  $\sigma_{\text{pop}} = 6.9\text{cm}$ .

Note that the bias  $\langle X \rangle_{\text{samp}}$  we found in Exercise 1.13, eqn 1.8 is quite different than the bias  $\langle X \rangle_{\text{BayesAv}}$  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 ‘true distributions’ that could have produced the data.

As we mentioned earlier, uniform priors seem unbiased. But a prior uniform in the standard deviation  $\sigma$  is not uniform in the variance  $\sigma^2$ !

(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  $P(\sigma)d\sigma = P(v)dv$ .) What is  $\alpha$  for a uniform prior on the variance? Calculate the mode and the average for our three samples  $S_N = N\sigma_{\text{pop}}^2$ , ( $N = 4, 40$ , and  $400$ ,  $\mu = \mu_0$ ) for a uniform prior on the variance. Compare with the naive estimate.

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  $\tau$  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.

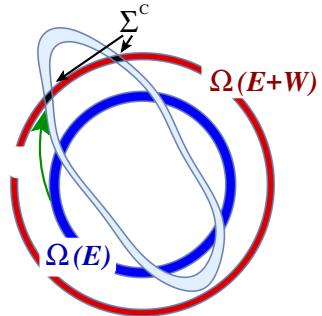
(c) Consider a uniform prior in the log of the width  $\log(\sigma)$ . Show that  $P(\sigma) \propto 1/\sigma$ , so  $\alpha = -1$ . Check that the prior  $P_\alpha(\sigma)$  integrated over a range  $(\sigma_0, 10\sigma_0)$  is indeed constant, independent of  $\sigma_0$ . Calculate the mode and the average for our three samples  $S_N = N\sigma_{\text{pop}}^2$  ( $N = 4, 40$ , and  $400$ ,  $\mu = \mu_0$ ) for Jeffrey’s prior.

#### 4.8 Crooks. ③

Just at the end of the 20th century, striking new fluctuation theorems were developed for broad classes of systems *driven out of equilibrium*. We can derive a version of Crooks’ fluctuation theorem [15] using Liouville’s theorem in the microcanonical ensemble. See also Exercise 4.9.

Consider a system in microcanonical equilibrium in the energy shell  $(E, E + \delta E)$ . (Think of a piston of gas, insulated from the outside world). Let the system be subject to an external forcing, giving it some time-dependent potential energy taking the system from an initial state to a final state. In particular, if the system starts

at a point in phase-space  $(\mathbb{P}, \mathbb{Q})$  when the external force starts, let us denote the final point  $U(\mathbb{P}, \mathbb{Q})$  as in Fig. 4.11. (Consider compressing the gas, perhaps very rapidly, sending it out of equilibrium.) The external force will do different amounts of work  $W$  on the system depending on the initial condition. (Consider a system with just a few gas atoms. As the gas is compressed, how strongly the atoms collide with the moving piston will depend on their initial state.)



**Fig. 4.11 Crooks fluctuation theorem: evolution in time.** The inner energy shell  $\Omega(E)$  gets mapped by a time-dependent Hamiltonian evolution  $U$  into a region spanning a variety of energies (wiggly region); different initial conditions exchange different amounts of work. The region  $\Sigma^C$  denotes the final states that landed in the energy shell  $\Omega(E + W)$  (absorbing work  $W$ ).

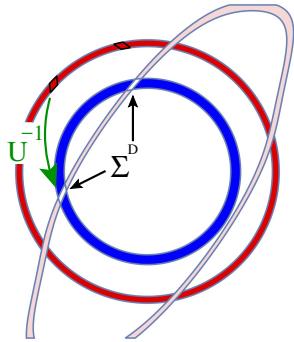
(a) Check that the derivation of Liouville’s theorem, eqns (4.5-4.7), also applies for time-dependent potential energies; that they also conserve phase space volume.

(b) The system starts in microcanonical equilibrium in the energy shell  $(E, E + \delta E)$ , and the probability density evolves under  $U$  into the wiggly region in Fig. 4.11. Inside this wiggly region, what is the final phase-space probability density? Conversely, if the system starts in microcanonical equilibrium in the energy shell  $(E + W, E + \delta E + W)$ , what is the final phase space density under the evolution  $U^{-1}$  (Fig. 4.12)? Express your answer in terms of the function  $\Omega(E)$ .

The microscopic equations of motion are invariant under time reversal. The volume  $\Sigma^C$  in Fig. 4.11 includes the trajectories that started near energy  $E$  and absorbed work  $W$  under the time evolution  $U$ . Under the reverse evolution  $U^{-1}$  these

trajectories all map to the volume  $\Sigma^D$  that emitted work  $W$  under the reverse evolution  $U^{-1}$  to end near  $E$ . Liouville's theorem thus tells us that the phase-space volume of  $\Sigma^C$  must therefore be equal to the phase-space volume of  $\Sigma^D$ .

An experiment starts in a microcanonical state in the energy shell  $(E, E + \delta E)$ , and measures the probability  $p^C(W)$  that the evolution  $U$  leaves it in the energy shell  $(E + W, E + \delta E + W)$  (roughly doing work  $W$ ). Another experiment starts in  $(E + W, E + \delta E + W)$ , and measures the probability  $p^D(-W)$  that the evolution  $U^{-1}$  leaves it in the energy shell  $(E, E + \delta E)$ .



**Fig. 4.12 Crooks fluctuation theorem: evolution backward in time.** Starting from the outer energy shell  $\Omega(E + W)$  transformed under the inverse evolution  $U^{-1}$  (wiggly region), the region  $\Sigma^C$  gets mapped back into  $\Sigma^D$  (and vice-versa under  $U$ ).

(c) What is  $p^C(W)$ , in terms of  $\Omega(E)$ ,  $\Omega(E + W)$ , and  $\Sigma^C$ ? What is  $p^D(-W)$ , in terms of the energy-shell volumes and  $\Sigma^D$ ? (See Figs. 4.11 and 4.12. Hint: the microcanonical ensemble fills its energy shell with uniform density in phase space.)

Now let us use Liouville's theorem to derive a way of using this *non-equilibrium* experiment to measure the *equilibrium* entropy difference  $S(E + W) - S(E)$ .

(d) Use your result from part (c) to write  $\Omega(E + W)/\Omega(E)$  in terms of the measured  $p^C(W)$  and  $p^D(-W)$ . Use this to derive the entropy difference  $S(E + W) - S(E)$ . (Hint:  $S(E) = k_B \log(\Omega(E))$ .) Who would have thought that anything could be proven about an arbitrary time-dependent system, driven out of equilibrium? Who would have thought that anything physical like  $p^C(W)$  would be ever given in terms of the very-large-number

$\Omega(E)$  (instead of its log)?

Your result in part (d) is a version of Crooks fluctuation theorem [15], which states that

$$p^D(-W)/p^C(W) = \exp(-\Delta S/k_B). \quad (4.48)$$

#### 4.9 Jarzynski. ③

You cannot create a perpetual motion machine – a machine that extracts energy from thermal vibrations to do useful work. Tiny machines, however, fluctuate – sometimes doing work on their environment. This exercise discusses a remarkable result, the Jarzynski equality [24], that quantifies the effects of fluctuations in small non-equilibrium systems. (See also Exercise 4.8.)

Suppose we compress a piston from length  $L$  to length  $L'$ , doing work  $W_c \geq 0$  on the gas in the piston, heating the gas. As we allow the gas to expand back to length  $L$ , the net work done by the outside world  $W_e \leq 0$  is negative, cooling the gas. Perpetual motion is impossible, so no net work can be done extracting kinetic energy from the gas:

$$\overline{W_c + W_e} \geq 0. \quad (4.49)$$

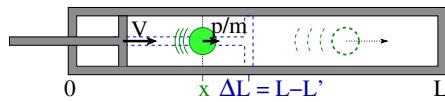
Note the average in eqn 4.49. Especially for small systems compressed quickly, there will be fluctuations in the work done. Sometimes it will happen that the gas will indeed do net work on the outside world. One version of the Jarzynski equality states that a thermally isolated system, starting in equilibrium at temperature  $T$ , evolved under a time-dependent Hamiltonian back to its initial state, will always satisfy

$$\overline{\exp(-(W_c + W_e)/k_B T)} = 1. \quad (4.50)$$

(a) Can this be true if the work  $W_c + W_e > 0$  for all initial conditions? Must every system, evolved to a non-equilibrium state under a time-dependent Hamiltonian, sometimes do net work on the outside world?

Note that the Jarzynski equality eqn 4.50 implies the fact that the average work must be positive, eqn 4.49. Because  $e^{-x}$  is convex,  $\overline{e^{-x}} \geq e^{-\overline{x}}$ . (Roughly, because  $e^{-x}$  is curved upward everywhere, its average is bigger than its value at the average  $x$ .) Applying this to eqn 4.50 gives us  $1 = \overline{e^{-(W_c + W_e)/k_B T}} \geq e^{-\overline{W_c + W_e}/k_B T}$ . Taking logs of both sides tells us  $0 \geq -\overline{W_c + W_e}/k_B T$  so  $0 \leq \overline{W_c + W_e}$ ; the net work done is not negative. Other forms of the Jarzynski equality give the change in free energy for non-equilibrium systems that start and end in different states.

We shall test a version of this equality in the smallest, simplest system we can imagine – one particle of an ideal gas inside a thermally insulated piston, Fig. 4.13, initially confined to  $0 < x < L$ , compressed into a region  $L - L' < x < L$  of length  $L'$  at velocity  $V$ , and then immediately allowed to expand back to  $L$  at velocity  $-V$ . For simplicity, we shall assume that the particles which are hit by the piston during compression do not have time to bounce off the far wall and return.<sup>37</sup>



**Fig. 4.13 One particle in a piston.** The piston is pushed forward at velocity  $V$  (not assumed small) a distance  $\Delta L$ , changing the piston length from  $L$  to  $L' = L - \Delta L$ . The particle is initially at position  $x$  and momentum  $p$ . If it is hit by the piston, it will recoil to a new velocity, absorbing work  $W_c(p)$  from the piston.

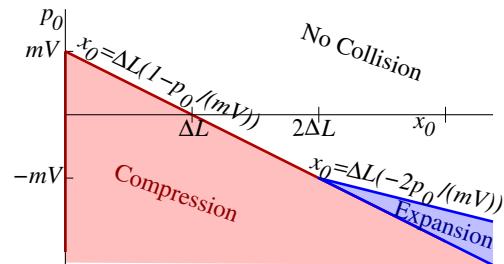
First, we carefully work through some elementary mechanics. How much energy is transferred during a collision with the piston? Which initial conditions will collide during contraction? During expansion?

(b) *If a particle with momentum  $p$  elastically collides with the compressing piston moving with velocity  $V$ , as shown in Fig. 4.13, what is the final kinetic energy after the collision? (Hint: It is easy in the moving reference frame of the piston.) What work  $W_c(p)$  is done by the piston on the particle? If a particle with momentum  $p$  elastically collides with the expanding piston moving with velocity  $-V$ , what is the work done  $W_e(p)$  by the piston on the particle? Can the piston do negative work during compression? (Hint: What is the direction of the impulsive force during a collision?) Can  $W_e(p)$  be positive?*

Many initial condition  $(x_0, p_0)$  for the particle at  $t = 0$  will not collide with the piston during compression, many will not collide with it during the expansion. Fig. 4.14 shows the initial conditions in phase space from which a particle will collide with the piston during the compression and ex-

pansion.

(c) *Which boundary describes the initial conditions for a particle that collides with the piston at the very start of the compression cycle? At the end of the compression cycle? Which boundary describes the initial conditions which collide at the beginning of the expansion cycle? At the end of the expansion cycle? Will any particle collide with the piston more than once, given our assumption that the particle does not have time to bounce off of the wall at  $x = L$  and return? Why or why not?*



**Fig. 4.14 Collision phase diagram.** Initial conditions  $(x_0, p_0)$  in phase space for the particle, showing regions where it collides with the piston during the compression cycle and during the expansion cycle. The equations determining the boundaries are shown.

Our macroscopic intuition is that the fluctuations that do net negative work should be rare. But these fluctuations, weighted exponentially as in the Jarzynski equality eqn 4.50, must balance those that do positive work.

(d) *Which of the initial conditions of part (c) will do negative net work on the particle? At low temperatures and high velocities  $V$ , where  $k_B T \ll \frac{1}{2} m V^2$ , are these initial conditions rare?* The theoretical argument for the Jarzynski equality is closely related to that for the Crooks thermodynamic relation of Exercise 4.8. Instead of deriving it, we shall test the theorem numerically for our piston. Choose  $m = 2$ ,  $k_B T = 8$ ,  $V = 2$ ,  $L = 20$ , and  $L' = 19$  (so  $\Delta L = 1$ ).

We could calculate the expectation values  $\langle \dots \rangle$  by doing an integral over phase space (Fig. 4.14) weighted by the Boltzmann distribution. Instead, let us do an ensemble of experiments, starting particles with a Boltzmann distribution of initial con-

<sup>37</sup>Note that, to stay in equilibrium, the particle must bounce off the piston many times during the compression – we are working in the other extreme limit. Also note that we could just assume an infinitely long piston, except for the computational part of the exercise.

ditions in phase space. (This is an example of a Monte-Carlo simulation; see Chapter 8).

- (e) Write a routine that generates  $N$  initial conditions  $(x_0, p_0)$  with a Boltzmann probability distribution for our ideal gas particle in the piston, and calculates  $W_c(x_0, p_0) + W_e(x_0, p_0)$  for each. Calculate the average work done by the piston  $\bar{W}_c + \bar{W}_e$ . How big must  $N$  be before you get a reliable average? Does your average obey eqn 4.49?
- (f) Use your routine with  $N = 1000$  to calculate Jarzynski's exponentially weighted average  $\exp(-\bar{W}/k_B T)$  in eqn 4.50, and compare to  $\exp(-\bar{W}/k_B T)$ . Do this 200 times, and plot a histogram of each. Does the Jarzynski equality appear to hold?

For big systems, the factor  $\exp(-W/k_B T)$  in the Jarzynski equality balances extremely rare fluctuations with low work against the vast majority of instances where the work has small fluctuations about the average.

#### 4.10 2D turbulence and Jupiter's great red spot.<sup>38</sup> (Astrophysics, Computation, Dynamical systems) ③

Fully-developed turbulence is one of the outstanding challenges in science. The agitated motion of water spouting through a fire hose has a complex, fluctuating pattern of swirls and eddies spanning many length scales. Many have studied the velocity-velocity correlation functions in turbulence, and there are close analogies to the scaling behavior seen at continuous phase transitions, which remain topics of active research.

Turbulence in two-dimensional fluids is much better understood, with the latest fashion using conformal field theories [40]. Our inspiration [33] is based on an ancient point vortex model introduced by Onsager and others.

The model describes the motion of point vortices, describing local rotational eddies in the fluid. It provides an analogy to Liouville's theorem, an example of a system with an infinite number of conserved quantities, and a system which can have a negative temperature. And it provides a plausible explanation of Jupiter's Giant Red Spot, and cyclones, hurricanes, and typhoons on Earth.<sup>39</sup>

<sup>38</sup>This exercise was developed in collaboration with Jaron Kent-Dobias. Hints for the computations can be found at the book Web site [49].

<sup>39</sup>Cyclones, hurricanes, and typhoons are different names used for the same weather patterns.

<sup>40</sup>More generally, if the vorticity is defined in a non-simply connected region  $S$  with  $n$  holes, and if  $\mathbf{u}$  is parallel to the boundaries, then knowing  $\omega(\mathbf{r})$  and the circulations  $\int_{\partial S_n} \mathbf{u} \cdot d\ell$  around the edges  $\partial S_n$  of the holes specifies the velocity  $\mathbf{u}$  inside.

The flow of most liquids is almost incompressible; if  $\mathbf{u}(\mathbf{x})$  is the velocity field, then to an excellent approximation  $\nabla \cdot \mathbf{u} = 0$ . Helmholtz's theorem tells us that a smooth field  $\mathbf{u}$  is determined by its divergence and curl, if it dies away fast enough at infinity and the region has no holes.<sup>40</sup> So knowing how the vorticity  $\boldsymbol{\omega} = \nabla \times \mathbf{u}$  evolves is enough to determine the evolution of the fluid. In two dimensions, the vorticity is a scalar  $\omega(x) = \partial u_y / \partial x - \partial u_x / \partial y$ .



**Fig. 4.15 Jupiter's Red Spot.** Jupiter's atmosphere is stormy, with winds that can exceed 400 miles per hour in a complex, turbulent pattern. Jupiter's 'red spot' is a giant storm  $3\frac{1}{2}$  times the size of the earth, that has lasted for at least 186 years. (Image from Voyager I, courtesy of NASA/JPL [35].)

- (a) Show that

$$\mathbf{u}_0(\mathbf{x}) = \frac{\Gamma}{2\pi} \frac{\hat{\theta}}{r} \quad (4.51)$$

in two dimensions has both zero curl and zero divergence, except at zero. (This is the velocity field around one of Onsager's vortices.) Use Stokes' theorem to show that the curl of this velocity field is  $\omega(\mathbf{x}) = \Gamma \delta(\mathbf{x}) = \Gamma \delta(x) \delta(y)$ . Argue using this solution that the velocity field in an annulus is not uniquely defined by its divergence and curl alone (as described by footnote 40). (Hint: Consider

the annulus bounded by two concentric circles surrounding our vortex.)

We provide a simulation of the dynamics of interacting vortices. Download the hints file.

(b) *Run the simulation with one vortex with  $\Gamma_1 = 1$  and a ‘tracer vortex’ with  $\Gamma = 0$ . Does the tracer vortex rotate around the test vortex?*

At low velocities, high viscosity, and small distances, fluids behave smoothly as they move; turbulence happens at high Reynolds numbers, where velocities are big, distances small, and viscosities are low. If we reduce the viscosity to zero, the kinetic energy of the fluid is conserved.

We can write  $\mathbf{u}$  as a nonlocal function of  $\omega$

$$u_x(\mathbf{r}) = -\frac{1}{2\pi} \int d\mathbf{r}' \frac{y - y'}{(\mathbf{r} - \mathbf{r}')^2} \omega(\mathbf{r}') \quad (4.52)$$

$$u_y(\mathbf{r}) = \frac{1}{2\pi} \int d\mathbf{r}' \frac{x - x'}{(\mathbf{r} - \mathbf{r}')^2} \omega(\mathbf{r}') \quad (4.53)$$

(the Biot-Savart law).

(c) *Show that the Biot-Savart law agrees with your answer from part (a) for the case of a  $\delta$ -function vortex.*

We can write the kinetic energy  $H$  in terms of the vorticity, which turns out to be a non-local convolution

$$\begin{aligned} H &= \int \frac{1}{2} \mathbf{u}(\mathbf{r})^2 d\mathbf{r} \\ &= -\frac{1}{4\pi} \int d\mathbf{r} \int d\mathbf{r}' \omega(\mathbf{r}) \omega(\mathbf{r}') \log |\mathbf{r} - \mathbf{r}'| \end{aligned} \quad (4.54)$$

where we set the density of the fluid equal to one and ignore an overall constant.<sup>41</sup> We can also write the equations of motion for the vorticity  $\omega$  in terms of itself and the velocity  $\mathbf{u}$

$$\frac{\partial \omega}{\partial t} = -\mathbf{u} \cdot \nabla \omega, \quad (4.55)$$

Note that this continuum vorticity has total derivative zero  $d\omega/dt = \partial\omega/\partial t + \mathbf{u} \cdot \nabla \omega = 0$ ; the vorticity measured along a packet of fluid remains constant in time. This means that the amount of fluid  $\rho(\omega, t)d\omega$  in a small range  $(\omega, \omega + d\omega)$  of vorticity stays constant in time, *for each possible  $\omega$*  – an infinite number of conservation laws.

To set up our simulation, we need to find a numerical representation for the vorticity field. Onsager suggested a discretization into vortices (part (a)),

$\omega(\mathbf{r}) = \sum_{i=1}^N \Gamma_i \delta(\mathbf{r} - \mathbf{r}_i)$ . For point vortices, the energy becomes

$$\mathcal{H} = -\frac{1}{2\pi} \sum_i \sum_{j \neq i} \Gamma_i \Gamma_j \log ((\mathbf{r}_i - \mathbf{r}_j)^2). \quad (4.56)$$

Note that there is no ‘kinetic energy’ for the vortices. The energy in the velocity field is purely kinetic energy; the ‘potential’ energy of the vortices is the kinetic energy of the fluid.

The vortices move according to the local velocity field,

$$d\mathbf{r}_i/dt = \mathbf{u}(r_i). \quad (4.57)$$

This just means that the fluid drags the vortices with it, without changing the strength of the vortices. (This agrees with the continuum law that the total derivative of the vorticity is zero, eqn 4.55). Hence

$$dx_i/dt = \frac{1}{2\pi} \sum_{j \neq i} \Gamma_j (y_j - y_i) / (\mathbf{r}_i - \mathbf{r}_j)^2 \quad (4.58)$$

$$dy_i/dt = -\frac{1}{2\pi} \sum_{j \neq i} \Gamma_j (x_i - x_j) / (\mathbf{r}_i - \mathbf{r}_j)^2.$$

If there is no kinetic energy for the vortices, what are the ‘conjugate variables’ analogous to  $x$  and  $p$  for regular particles?

(d) *Check from eqn 4.58 and eqn 4.56 that  $dx_i/dt = (1/4\Gamma_i) \partial \mathcal{H} / \partial y_i$  and  $dy_i/dt = -(1/4\Gamma_i) \partial \mathcal{H} / \partial x_i$ . Thus  $2\sqrt{\Gamma_i} x_i$  and  $2\sqrt{\Gamma_i} y_i$  are analogous to  $x$  and  $p$  in a regular phase space.*

Launch the simulation.

(e) *Start with  $n = 20$  vortices with a random distribution of vortex strengths  $\Gamma_i \in [-1, 1]$  and random positions  $\mathbf{r}_i$  within the unit circle.<sup>42</sup> Print the original configuration. Run for a time  $t = 10$ , and print the final configuration. Do you see the spontaneous formation of a giant whirlpool? Are the final positions roughly also randomly arranged? Measure and report the energy  $H$  of your vortex configuration.* (Warning: Sometimes the differential equation solver crashes. Just restart again with a new set of random initial conditions.)

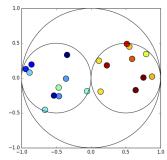
Note: we are not keeping the vortices inside the unit circle during our dynamical evolution. Hurricanes and Jupiter’s Red Spot can be thought of as large concentrations of vorticity – all the plus or minus vortices concentrated into one area, making a giant whirlpool. What do we need to do

<sup>41</sup>Miller, Weichman, and Cross tell us that  $\log(|\mathbf{r} - \mathbf{r}'|/R_0)$  corresponds to free boundary conditions, where  $R_0$  is a constant with dimensions of length.

<sup>42</sup>You can generate random points in the unit circle by choosing  $\theta \in (-\pi, \pi]$  and picking  $r = \sqrt{p}$  with  $p$  uniformly distributed  $\in [0, 1]$ .

to arrange this? Let's consider the effect of the total energy.

(f) For a given set of vortices of strength  $\Gamma_i$ , would clustering the positive vortices and the negative vortices each into their own clump (counter-rotating hurricanes) be a high-energy or a low energy configuration, as the size of the clumps goes to zero? (Hint: You will be able to check this with the simulation later.) Why?



**Fig. 4.16 Circles.** It is unlikely that the vortices would collect into clumps by accident.

You chose points at random for part (e), but did not see the vortices separated into clumps. How did we know this was likely to happen?

(g) Estimate the entropy difference between a state where twenty vortices are confined within the unit circle, and a state where ten positive vortices and ten negative vortices each are confined inside the particular circles of radius  $R = \frac{1}{2}$  shown in Fig. 4.16.<sup>43</sup> Leave your answer as a multiple of  $k_B$ . (Note: The vortices have different values of  $\Gamma$ , so are distinguishable.) How many tries would you need to see this clumping happen?

Onsager's problem is one of the best examples of negative temperature.<sup>44</sup>

(h) Using your results from part (f) and part (g), in the energy range where the hurricanes will form, will the change in entropy be positive or negative as the energy increases? Is the temperature negative?

In most simulations of Onsager's vortices, one selects for states that form a big hurricane by start-

ing with several small ones, and watching them combine. (The small hurricanes must be tightly packed; as they combine they gain entropy because the vortices can spread out more.) Instead, we shall set the energy of our configuration by doing a Metropolis Monte-Carlo simulation at a fixed temperature.

(i) Thermalize the Monte-Carlo simulation for your  $n = 20$  vortices at low temperature  $\beta = 1/(k_B T) = 2$ , report the final energy, and print out your configuration. Does the thermalized vortex state look similar to the initial conditions you generated in part (e)? Thermalize again at temperature  $\beta = -2$ , report the energy, and print out your final configuration. Do the vortices separate out into clumps of positive and negative vorticity?

The time it takes to run a simulation is roughly determined by the minimum distance between vortices. We can use this to keep our simulation from crashing so much.

(j) Re-run the Monte Carlo simulation with  $n = 20$  vortices until you find a configuration with a clear separation of positive and negative vortices, but where the minimum distance between vortices is not too small (say, bigger than 0.01). (This should not take lots of tries, so long as you thermalize with an energy in the right region.) Is the energy you need to thermalize to positive, or negative? Print this initial configuration of vortices. Animate the simulation for  $t = 10$  with this state as the initial condition. How many hurricanes do you find? Print out the final configuration of vortices. (Hint: In Mathematica, one can right-click in the animation window to print the current configuration (Save Graphic As...). In Python, just plot  $xOft[-1]$ ,  $yOft[-1]$ ; an array evaluated at [-1] gives the last entry. You can copy the axis limits and colors and sizes from the animation, to make a nice plot.)

<sup>43</sup>It is another interesting question how many small circles could be formed, but that is beyond the scope of this question.

<sup>44</sup>Because the phase space is finite, the volume of the energy shell at high energies goes to zero, instead of continuing to increase.

## Chapter 5: Entropy

### Exercises

**5.8 The Arnol'd cat map.** (Mathematics, Dynamical systems) ③

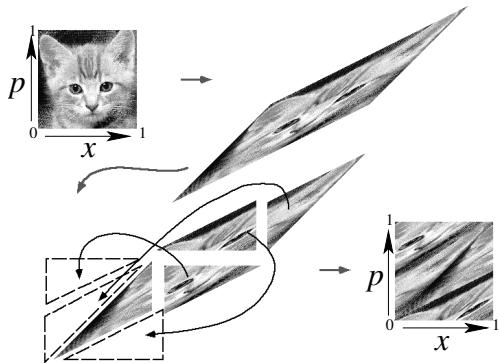
Why do we suppose equilibrium systems uniformly cover the energy surface? Chaotic motion has *sensitive dependence on initial conditions*; regions on the energy surface get stretched into thin ribbons that twist and fold in complicated patterns, losing information about the initial conditions and leaving many systems with a uniform distribution of probability over all accessible states. Since energy is the only thing we know is conserved, we average over the energy surface.

Arnol'd developed a simple illustration of this stretching and folding using a function taking a two-dimensional square into itself (called the “cat map”, Figure 5.17). Liouville’s theorem will tell us that Hamiltonian dynamics pre-

serves the  $6N$ -dimensional phase-space volume; Arnol'd map preserves area in this  $1 \times 1$  square, taking  $[0, 1] \times [0, 1]$  in the plane onto itself.<sup>45</sup>

$$\begin{aligned}\Gamma(x, p) &= \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x \\ p \end{pmatrix} \bmod 1 \\ &= \begin{pmatrix} 2x + p \\ x + p \end{pmatrix} \bmod 1 \\ &= \text{Mod}_1 \left( M \begin{pmatrix} x \\ p \end{pmatrix} \right)\end{aligned}\quad (5.59)$$

where  $M = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$  stretches and squeezes our square ‘energy shell’ and  $\text{Mod}_1$  cuts and pastes it back into the square. We imagine a trajectory  $(x_{n+1}, p_{n+1}) = \Gamma(x_n, p_n)$  evolving with a discrete ‘time’  $n$ . This map vividly illustrates how Hamiltonian dynamics leads to the microcanonical ensemble.



**Fig. 5.17 Arnol'd cat map.** Arnol'd, in a take-off on Schrödinger's cat, paints a cat on a 2D phase space, which gets warped and twisted as time evolves. From <http://www-chaos.umd.edu/misc/catmap.html>.

The cutting and pasting process confuses our analysis as we repeatedly apply the map  $(x_n, p_n) = \Gamma^n(x_0, p_0) = \text{Mod}_1(M(\text{Mod}_1(M(\dots(x_0, p_0))))).$  Fortunately, we can view the time evolution as many applications of the stretching process, with a final cut-and-paste:

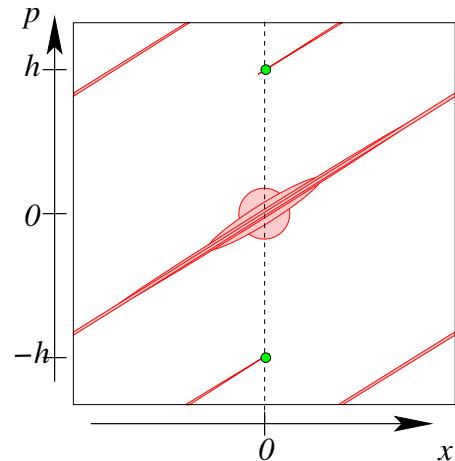
$$(x_n, p_n) = \text{Mod}_1(M^n(x_0, p_0)). \quad (5.60)$$

(a) Show that the last equation 5.60 is true. (Hint: Work by induction. You may use the fact that

$y \bmod 1 = y + n$  for some integer  $n$  and that  $(z + n) \bmod 1 = z \bmod 1$  for any integer  $n.$ ) Thus the stretching by  $M$  can be done in the entire  $(x, p)$  plane, and then the resulting thin strip can be folded back into the unit square.

Now we analyze how the map stretches and squeezes the rectangle.

(b) Verify that  $(\gamma, 1)$  and  $(-\frac{1}{\gamma}, 1)$  are eigenvectors of  $M$ , where  $\gamma = (\sqrt{5} + 1)/2$  is the Golden Ratio. What are the eigenvalues? Consider a square box, centered anywhere in the plane, tilted so that its sides are parallel to the two eigenvectors. How does its height and width change under application of  $M$ ? Show that its area is conserved. (Hint: If you are doing this by hand, show that  $1/\gamma = (\sqrt{5} - 1)/2$ , which can simplify your algebra.)



**Fig. 5.18 Evolution of an initially concentrated phase-space density.** Suppose we know our particle has initial position and momentum confined to a small region  $|(x, p)| < r.$  This small region is stretched along an irrational angle under the Arnol'd cat map. (This figure shows the origin  $x = 0, p = 0$  as the center of the figure; in the map of eqn 5.59 shown in Fig. 5.17 the origin is at the corner.)

The next question explores the map iteration when the initial point is at or near the origin  $(x_0, p_0) = (0, 0).$  As illustrated in Figure 5.18, a small, disk-shaped region of uncertainty, centered on the origin, evolves with time  $n$  into a thin strip.

<sup>45</sup>Such maps arise as Poincaré sections of Hamiltonian flows (Fig. 4.9), or as the periodic map given by a Floquet theory. There are many analogies between the behavior of dynamical systems with continuous motion and discrete maps.

When the iteration would make thin strip hit the square's boundary, it gets cut in two; further iterations stretch and chop our original circle into a bunch of parallel thin lines. (In the version of our map given by eqn 5.60,  $M^n$  stretches the initial circle into a growing long thin line, which gets folded in by  $\text{Mod}_1$  to a series of parallel line segments.)

(c) Calculate the momentum  $h$  at which the direction of the thin strip (as measured by the expanding eigenvector in part (b)) first crosses the line  $x = 0$ . Note that  $h$  is irrational. The multiples of an irrational number, taken mod 1, can be shown to be dense in the interval  $[0, 1]$ . Using this, argue that the resulting thin strip, in the limit of many iterations, is dense in the unit square (i.e., the maximum gap between line segments goes to zero as time goes to infinity).

These lines eventually cover the square (our ‘energy shell’) densely and uniformly – intuitively illustrating how the microcanonical ensemble typically describes the long-term behavior. (The small circle represents an uncertainty in the initial position. As for chaotic motion, Arnol'd's cat map amplifies the uncertainty in the initial position, which leads to the microcanonical ensemble probability distribution spread uniformly over the square.)

But there are many initial conditions whose trajectories do not cover the energy surface. The most obvious example is the origin,  $x_0 = p_0 = 0$ , which maps into itself. In that case, the time average of an operator<sup>46</sup>  $O$ ,  $\lim_{T \rightarrow \infty} 1/T \sum_0^T O(x_n, p_n) = O(0, 0)$ , clearly will usually be different from its microcanonical average  $\int_0^1 dx \int_0^1 dp O(x, p)$ . The next part of the question asks about a less trivial example of a special initial condition leading to non-ergodic behavior.

(d) If  $x_0$  and  $p_0$  can be written as fractions  $a/q$  and  $b/q$  for positive integers  $q$ ,  $a \leq q$ , and  $b \leq q$ , show that  $x_n$  and  $p_n$  can be written as fractions with denominator  $q$  as well. Show thus that this trajectory must eventually<sup>47</sup> settle into a periodic orbit, and give an upper bound to the period. Must the time average of an operator  $O$  for such a trajectory equal the microcanonical average?

You probably have heard that there are infinitely many more irrational numbers than rational ones. So the probability of landing on one of the periodic orbits is zero (points with rational  $x$  and  $p$  coordinates are of measure zero). But perhaps there are even more initial conditions which do not equilibrate?

(e) Find an initial condition with a non-periodic orbit which goes to the origin as  $n \rightarrow \infty$ . What will the time average of  $O$  be for this initial condition?

Arnol'd's argument that the cat map (eqn 5.59) is ergodic shows that, even taking into account points like those you found in parts (d) and (e), the probability of landing on a non-ergodic component is zero. For ‘almost all’ initial conditions, or any initial condition with a small uncertainty, the Arnol'd cat map trajectory will be dense and uniformly cover its square phase space at long times.

### 5.12 Rubber band. (Condensed matter) ⓘ

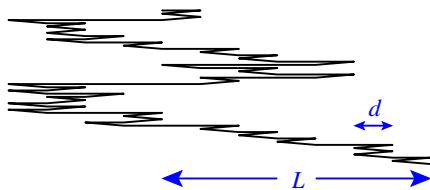
Figure 5.19 shows a one-dimensional model for rubber. Rubber is formed from long polymeric molecules, which undergo random walks in the undeformed material. When we stretch the rubber, the molecules respond by rearranging their random walk to elongate in the direction of the external stretch. In our model, the molecule is represented by a set of  $N$  links of length  $d$ , which with equal energy point either parallel or antiparallel to the previous link. Let the total change in position to the right, from the beginning of the polymer to the end, be  $L$ .

As the molecule extent  $L$  increases, the entropy of our rubber molecule decreases.

(a) Find an exact formula for the entropy of this system in terms of  $d$ ,  $N$ , and  $L$ . (Hint: How many ways can one divide  $N$  links into  $M$  right-pointing links and  $N - M$  left-pointing links, so that the total length is  $L$ ?)

<sup>46</sup>The operator  $O$  is some function of  $x$  and  $p$ . In a physics problem, it could be the energy, the momentum, or the square of the position, or here something like  $O(x, p) = \sin^2(2\pi x)(\cos(4\pi p) + 1)$  (to respect the periodic boundary conditions).

<sup>47</sup>Since the cat map of eqn 5.59 is invertible, the orbit will be periodic without an initial transient, but that's harder to argue.



**Fig. 5.19 Rubber band.** Simple model of a rubber band with  $N = 100$  segments. The beginning of the polymer is at the top; the end is at the bottom; the vertical displacements are added for visualization.

The external world, in equilibrium at temperature  $T$ , exerts a force pulling the end of the molecule to the right. The molecule must exert an equal and opposite entropic force  $F$ .

(b) *Find an expression for the force  $F$  exerted by the molecule on the bath in terms of the bath entropy.* (Hint: The bath temperature  $1/T = \partial S_{\text{bath}}/\partial E$ , and force times distance is energy.) *Using the fact that the length  $L$  must maximize the entropy of the Universe, write a general expression for  $F$  in terms of the internal entropy  $S$  of the molecule.*

(c) *Take our model of the molecule from part (a), the general law of part (b), and Stirling's formula  $\log(n!) \approx n \log n - n$ , write the force law  $F(L)$  for our molecule for large lengths  $N$ . What is the spring constant  $K$  in Hooke's law  $F = -KL$  for our molecule, for small  $L$ ?*

Our model has no internal energy; this force is entirely entropic. Note how magical this is – we never considered the mechanism of how the segments would generate a force. Statistical mechanics tells us that the force generated by our segmented chain is *independent* of the mechanism. The joint angles in the chain may jiggle from thermal motion, or the constituent polymer monomers may execute thermal motion – so long as the configuration space is segment orientations and the effective potential energy is zero the force will be given by our calculation. For the same reason, the pressure due to compressing an ideal gas is independent of the mechanism. The kinetic energy of particle collisions for real dilute gases gives the same pressure as the complex water-solvent interactions give for osmotic pressure (Section 5.2.2).

(d) *If we increase the temperature of our rubber*

*band while it is under tension, will it expand or contract? Why?*

In a more realistic model of a rubber band, the entropy consists primarily of our configurational random-walk entropy plus a vibrational entropy of the molecules. If we stretch the rubber band without allowing heat to flow in or out of the rubber, the total entropy should stay approximately constant. (Rubber is designed to bounce well; little irreversible entropy is generated in a cycle of stretching and compression, so long as the deformation is not too abrupt.)

(e) *True or false?*

(T) (F) *When we stretch the rubber band, it will cool; the configurational entropy of the random walk will decrease, causing the entropy in the vibrations to decrease, causing the temperature to decrease.*

(T) (F) *When we stretch the rubber band, it will cool; the configurational entropy of the random walk will decrease, causing the entropy in the vibrations to increase, causing the temperature to decrease.*

(T) (F) *When we let the rubber band relax, it will cool; the configurational entropy of the random walk will increase, causing the entropy in the vibrations to decrease, causing the temperature to decrease.*

(T) (F) *When we let the rubber band relax, there must be no temperature change, since the entropy is constant.*

This more realistic model is much like the ideal gas, which also had no configurational energy.

(T) (F) *Like the ideal gas, the temperature changes because of the net work done on the system.*

(T) (F) *Unlike the ideal gas, the work done on the rubber band is positive when the rubber band expands.*

You should check your conclusions experimentally; find a rubber band (thick and stretchy is best), touch it to your lips (which are very sensitive to temperature), and stretch and relax it.

### 5.13 How many shuffles? (Mathematics) ⑦

For this exercise, you will need a deck of cards, either a new box or sorted into a known order (conventionally with an ace at the top and a king at the bottom).<sup>48</sup>

<sup>48</sup>Experience shows that those who do not do the experiment in part (b) find it challenging to solve part (c) by pure thought.

How many shuffles does it take to randomize a deck of 52 cards?

The answer is a bit controversial; it depends on how one measures the information left in the cards. Some suggest that seven shuffles are needed; others say that six are enough.<sup>49</sup> We will follow reference [55], and measure the growing randomness using the information entropy.

We imagine the deck starts out in a known order (say, A♣, 2♣, ..., K♣).

(a) *What is the information entropy of the deck before it is shuffled? After it is completely randomized?*

(b) *Take a sorted deck of cards. Pay attention to the order; (in particular, note the top and bottom cards). Riffle it exactly once, separating it into two roughly equal portions and interleaving the cards in the two portions.. Examine the card sequence, paying particular attention to the top few and bottom few cards. Can you tell which cards came from the top portion and which came from the bottom?*

The mathematical definition of a *riffle shuffle* is easiest to express if we look at it backward.<sup>50</sup> Consider the deck after a riffle; each card in the deck either came from the top portion or the bottom portion of the original deck. A riffle shuffle makes each of the  $2^{52}$  patterns *tbbtbtb...* (denoting which card came from which portion) equally likely.

It is clear that the pattern *tbbtbtb...* determines the final card order: the number of *t*'s tells us how many cards were in the top portion, and then the cards are deposited into the final pile according to the pattern in order bottom to top. Let us first pretend the reverse is also true: that every pattern corresponds one-to-one with a unique final card ordering.

(c) *Ignoring the possibility that two different riffles could yield the same final sequence of cards, what is the information entropy after one riffle?*

You can convince yourself that the only way two riffles can yield the same sequence is if all the cards in the bottom portion are dropped first, followed by all the cards in the top portion.

(d) *How many riffles drop the entire bottom portion and then the entire top portion, leaving the card ordering unchanged? What fraction of the  $2^{52}$  riffles does this correspond to?* (Hint:  $2^{10} = 1024 \approx 10^3$ . Indeed, this approximation underlies measures of computing resources: a gigabyte is not  $10^9$  bytes, but  $(1024)^3 = 2^{30}$  bytes.) *Hence, what is the actual information entropy after one riffle shuffle?*

We can put a lower bound on the number of riffles needed to destroy all information by assuming the entropy increase stays constant for future shuffles.

(e) *Continuing to ignore the possibility that two different sets of  $m$  riffles could yield the same final sequence of cards, how many riffles would it take for our upper bound for the entropy to pass that of a completely randomized deck?*

### 5.24 The Dyson sphere. (Astrophysics) ⓘ

Life on Earth can be viewed as a heat engine, taking energy from a hot bath (the Sun at temperature  $T_S = 6000^\circ\text{K}$ ) and depositing it into a cold bath (interstellar space, at a microwave background temperature  $T_{MB} = 2.725\text{K}$ , Exercise 7.15). The outward solar energy flux at the Earth's orbit is  $\Phi_S = 1370\text{W/m}^2$ , and the Earth's radius is approximately 6400 km,  $r_E = 6.4 \times 10^6\text{m}$ .

(a) *If life on Earth were perfectly efficient (a Carnot cycle with a hot bath at  $T_S$  and a cold bath at  $T_{MB}$ ), how much useful work (in watts) could be extracted from this energy flow? Compare that to the estimated world marketed energy consumption of  $4.5 \times 10^{20}\text{J/year}$ . (Useful constant: There are about  $\pi \times 10^7\text{s}$  in a year.)*

Your answer to part (a) suggests that we have some ways to go before we run out of solar energy. But let's think big.

(b) *If we built a sphere enclosing the Sun at a radius equal to Earth's orbit (about 150 million kilometers,  $R_{ES} \approx 1.5 \times 10^{11}\text{m}$ ), by what factor would the useful work available to our civilization increase?*

This huge construction project is called a *Dyson sphere*, after the physicist who suggested [17] that

<sup>49</sup>More substantively, as the number of cards  $N \rightarrow \infty$ , some measures of information show an abrupt transition near  $\frac{3}{2} \log_2 N$ , while by other measures the information vanishes smoothly and most of it is gone by  $\log_2 N$  shuffles.

<sup>50</sup>We could mimic an expert riffler splitting the deck into two roughly equal portions by choosing  $n$  cards on top with a binomial probability  $2^{-52} \binom{52}{n}$ . Then the expert drops cards with probability proportional to the number of cards remaining in its portion. This makes each of the  $2^{52}$  choices in the backward definition equally likely.

we look for advanced civilizations by watching for large sources of infrared radiation.

Earth, however, does not radiate at the temperature of interstellar space. It radiates roughly as a black body at near  $T_E = 300^\circ \text{K} = 23^\circ \text{C}$  (see, however, Exercise 7.22).

(c) *How much less effective are we at extracting work from the solar flux, if our heat must be radiated effectively to a  $300^\circ \text{K}$  cold bath instead of one at  $T_{MB}$ , assuming in both cases we run Carnot engines?*

There is an alternative point of view, though, which tracks entropy rather than energy. Living beings maintain and multiply their low-entropy states by dumping the entropy generated into the energy stream leading from the Sun to interstellar space. New memory storage also intrinsically involves entropy generation (Exercise 5.2); as we move into the information age, we may eventually care more about dumping entropy than about generating work. In analogy to the ‘work effectiveness’ of part (c) (ratio of actual work to the Carnot upper bound on the work, given the hot and cold baths), we can estimate an entropy-dumping effectiveness (the ratio of the actual entropy added to the energy stream, compared to the entropy that could be conceivably added given the same hot and cold baths).

(d) *How much entropy impinges on the Earth from the Sun, per second per square meter cross-sectional area? How much leaves the Earth, per second per cross-sectional square meter, when the solar energy flux is radiated away at temperature  $T_E = 300^\circ \text{K}$ ? By what factor  $f$  is the entropy dumped to outer space less than the entropy we could dump into a heat bath at  $T_{MB}$ ? From an entropy-dumping standpoint, which is more important, the hot-bath temperature  $T_S$  or the cold-bath temperature ( $T_E$  or  $T_{MB}$ , respectively)?*

For generating useful work, the Sun is the key and the night sky is hardly significant. For dumping the entropy generated by civilization, though, the night sky is the giver of life and the realm of opportunity. These two perspectives are not really at odds. For some purposes, a given amount of work energy is much more useful at low temperatures. Dyson later speculated about how life could make efficient use of this by running at much colder tempeartures (Exercise 5.1). A hyper-advanced information-based civilization would hence want not to radiate in the infrared, but in the microwave range.

To do this, it needs to increase the area of the Dyson sphere; a bigger sphere can re-radiate the Solar energy flow as black-body radiation at a lower temperature. Interstellar space is a good insulator, and one can only shove so much heat energy through it to get to the Universal cold bath. A body at temperature  $T$  radiates the largest possible energy if it is completely black. We will see in Exercise 7.7 that a black body radiates an energy  $\sigma T^4$  per square meter per second, where  $\sigma = 5.67 \times 10^{-8} \text{ J}/(\text{s m}^2 \text{ K}^4)$  is the Stefan–Boltzmann constant.

(e) *How large a radius  $R_D$  must the Dyson sphere have to achieve 50% entropy-dumping effectiveness? How does this radius compare to the distance to Pluto ( $R_{PS} \approx 6 \times 10^{12} \text{ m}$ )? If we measure entropy in bits (using  $k_S = (1/\log 2)$  instead of  $k_B = 1.3807 \times 10^{-23} \text{ J/K}$ ), how many bits per second of entropy can our hyper-advanced civilization dispose of?* (You may ignore the relatively small entropy impinging from the Sun onto the Dyson sphere, and ignore both the energy and the entropy from outer space.)

The sun wouldn’t be bright enough to read by at that distance, but if we had a well-insulated sphere we could keep it warm inside—only the outside need be cold. Alternatively, we could just build the sphere for our computers, and live closer in to the Sun; our re-radiated energy would be almost as useful as the original solar energy.

### 5.18 Undistinguished particles and the Gibbs factor. ②

If we have  $N$  particles in a box of volume  $V$  and do not distinguish between the particles, then the effective configurational entropy is  $S_Q = k_B \log(V^N/N!)$ , where  $N!$  is sometimes called the *Gibbs factor*. We saw that this factor keeps the entropy of mixing small for undistinguished particles.

(a) *Use Stirling’s formula to show that this configurational entropy is related to  $N$  times the entropy of a particle in a region roughly given by the distance to neighboring particles.*

(b) *In eqn 3.58 for the ideal gas entropy,  $S = Nk_B(5/2 - \log(\rho\lambda^3))$ , what part corresponds to this configurational entropy?* (Hint: Expand the logarithm.)

### 5.19 Entropy of socks. ②

*If a billion children neaten their bedrooms, how much does the entropy of their toys and clothing decrease? Assume that they each have one hundred distinguishable items weighing 20gm, initially*

*randomly placed in a bedroom with  $V=5m^3$ , and that they end up located in their place in a drawer with center-of-mass location accuracy of 1cm. For this exercise, assume that the internal entropies of the objects are unchanged in the process (atomic vibrations, folding entropy for socks, etc.) Compare this entropy change to that of contracting the air in a one liter balloon by 1% in volume.*

**5.20 Loaded dice and sticky spheres.** 

(a) *A loaded three-sided die has probability 1/2 of rolling a one, and probability 1/4 of rolling a two or a three. What is the entropy (in units of  $k_B$ ) per roll?*

(b) *Let  $\theta$ ,  $\phi$  be the latitude and longitude on the Earth, which we assume to be a sphere. If comets hit the earth with uniform probability density  $\rho_c(\theta, \phi) = 1/(4\pi)$  and asteroids hit the earth with probability density  $\rho_a(\theta, \phi) = \cos(\theta)/\pi^2$  that is maximum at the equator and zero at the poles, which distribution has higher entropy? What is the entropy difference between the distribution of asteroids and comets?* (Hints: The integral of  $f(\theta)$  over a sphere is  $\int_{-\pi/2}^{\pi/2} 2\pi \cos(\theta) f(\theta) d\theta$ . You may also want to know that  $\int_{-\pi/2}^{\pi/2} \cos^2(\theta) d\theta = \pi/2$  and  $\int_{-\pi/2}^{\pi/2} \cos^2(\theta) \log(\cos(\theta)) d\theta = \pi(1 - \log(4))/4$ .)

**5.21 Gravity and entropy.** (Astrophysics) 

The motion of gravitationally bound collections of particles would seem a natural application of statistical mechanics. The chaotic motion of many stars could maximize a ‘stellar entropy’ of the positions and momenta of the stars treated as particles. But there are challenges.

The gravitational interaction is long range. One of our key assumptions in deriving the entropy was that it should be extensive: the entropy of weakly interacting subsystems adds up to the entropy of the system.

(a) *Can we divide up a galaxy of stars into weakly interacting subsystems? If not, must the entropy of the positions and momenta of the stars in the galaxy be a sum of the stellar entropy of pieces of the galaxy? Do you conclude that Shannon entropy is applicable to stellar dynamics?*

(b) *The Coulomb force is also long range. Why can we describe distributions of atoms using Shannon entropy, when they are composed of electrons and nuclei with long-range forces?*

The  $1/r^2$  potential shared by Newton and Coulomb also has a strong singularity at short distances. The Boltzmann distribution for two

stars diverges at  $r = 0$ , so in ‘equilibrium’ all stars would fall into one another, another problem with applying statistical mechanics to stars. For electrons and nuclei, this divergence is cut off by quantum mechanics.

**5.23 Aging, entropy, and DNA.** (Biology) 

Is human aging inevitable? Does the fact that entropy must increase mean that our cells must run down? In particular, as we get older the DNA in our cells gradually builds up damage – thought to be a significant contribution to the aging process (and a key cause of cancer). Can we measure DNA damage using entropy, to quantify the challenge of keeping ourselves young?

There are roughly thirty trillion ( $3 \times 10^{13}$ ) cells in the human body, and about three billion ( $3 \times 10^9$ ) nucleotides in the DNA of each cell. Each nucleotide comes in four types (C, T, A, and G). The damaged DNA of each cell will be different. The repair job for fixing all of our DNA cannot be worse than changing totally random sequences back into exact copies of the correct sequence.

(a) *How many bits of information is it possible to store in the nucleotide sequence of the DNA in an entire human body? How much entropy, in joules/Kelvin, is associated with a completely randomized sequence in every cell?* (Boltzmann’s constant is  $1.38 \times 10^{-23}$  J/K.)

Life exists by consuming low-entropy sources and turning them into higher-entropy biproducts. A small cookie has 100 Calories. (Be warned: a Calorie is 1000 calories. Food calories are measured in kilocalories, and then the kilo is conventionally dropped in favor of a capital C.) Body temperature is about 310K.

(b) *Calculate the minimum free energy needed to repair a human’s DNA if it starts in a completely scrambled state. How many cookies would one need to consume?* (There are 4.184 joules per calorie, and 1000 calories per Calorie.)

Entropy does not discriminate between important and trivial information. Knowing that air molecules are confined in a balloon is much less useful than knowing that your kid’s toys are put away neatly (or knowing the contents of the Library of Congress), and there are a lot of air molecules...

**5.22 Entropy of the galaxy.**<sup>51</sup> (Astrophysics) ①

What dominates the entropy of our Galaxy? The known sources of entropy include stars, interstellar gas, the microwave background photons inside the Galaxy, and the supermassive black hole at the galactic center.

- (a) **Stars.** *Presume there are 100 billion stars ( $10^{11}$ ) in our galaxy, with mass  $m = 2 \times 10^{30} \text{ kg}$ , volume  $V = 10^{45} \text{ m}^3$ , and temperature  $T = 10^7 \text{ K}$ . Assume that they are made of hydrogen, and form an ideal gas (a poor approximation). What is the entropy of one star? Compare to Bekenstein's estimate of  $10^{35} \text{ J/K}$  [3]. What is your estimate for the total entropy  $S_{\text{stars}}$  of the stars in the Galaxy?*
- (b) **Gas.** *If the volume of the galaxy is  $V = 10^{61} \text{ m}^3$ , with interstellar gas of hydrogen atoms of total mass 10% that in the stars, and temperature  $10K$ , what is its entropy  $S_{\text{gas}}$ ?*
- (c) **Microwave photons.** *What is the total entropy for the cosmic microwave background radiation at  $T_{\text{CMB}} = 2.75 \text{ K}$  in the volume  $V$  of the galaxy?*
- (d) **Galactic black hole.** *What is the entropy for the supermassive black hole at the center of the galaxy, if its mass is  $M_{\text{BH}} = 10^{37} \text{ kg}$ ?*
- (e) **Where to live?** *Where would a technological civilization locate, if finding a place to deposit entropy was the bottleneck for their activities?*

**5.25 Nucleosynthesis and the arrow of time.**<sup>52</sup>

(Astrophysics) ③

In this exercise, we shall explain how the first few minutes after the Big Bang set up a low-entropy state that both fuels our life on Earth and provides an arrow of time. To do so, we shall explore the statistical mechanics of reaction rates in the early universe, use the equilibrium number

density and entropy of the non-interacting photon gas, and model the Universe as a heat engine.

The microscopic laws of physics are invariant under time reversal.<sup>53</sup> It is statistical mechanics that distinguishes future from past – the future is the direction in which entropy increases.

This must arise because the early universe started in a low-entropy state. Experimentally, however, measurements of the cosmic microwave background radiation tell us that the matter in the universe<sup>54</sup> started out as nearly uniform, hot, dense gas (Exercises 7.15 and 10.1) – an *equilibrium state* and hence of *maximum* entropy. How can a state of maximal entropy become our initial low-entropy state?

Some cosmologists suggest that gravitational effects that lead matter to cluster into galaxies, stars, and eventually black holes explain the problem; the initial uniform distribution of matter and energy maximizes the entropy of the particles, but is a low entropy state for space-time and gravitational degrees of freedom. It is certainly true on a cosmological scale that the formation of black holes can dominate the creation of entropy (Exercise 5.22). However, gravity has not been a significant source of free energy<sup>55</sup> (sink for entropy) apart from the vicinity of black holes; a thorough critique is given by Wallace [56]. The dominant source of free energy in the Solar System is not gravitation, but the hydrogen which fuels the Sun. Hydrogen predominates now because the Universe *fell out of equilibrium* as it expanded.

When thermally isolated systems are changed slowly enough to remain in equilibrium, we say the evolution is *adiabatic*.<sup>56</sup> Closed systems evolving adiabatically do not change in entropy.

- (a) *If a closed, thermally isolated system changes*

<sup>51</sup>This exercise was developed in collaboration with Katherine Quinn.

<sup>52</sup>This exercise was developed in collaboration with Katherine Quinn.

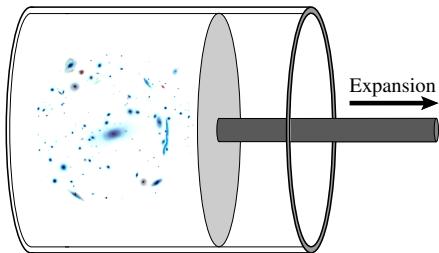
<sup>53</sup>Technically, the weak interaction is not time-reversal symmetric. But the laws of physics *are* invariant under CPT, where charge-conjugation reverses matter and antimatter and parity reflects spatial coordinates. There are no indications that entropy of antimatter decreases with time – the arrow of time for both appears the same.

<sup>54</sup>We do not have direct evidence for equilibration of neutrinos with the other massive particles, as they must have fallen out of equilibrium before the decoupling of light from matter which led to the microwave background. But estimates suggest that they too started in equilibrium.

<sup>55</sup>Helmholtz and Kelvin in the last half of the 1800's considered whether the Sun could shine from the energy gained by gravitational collapse. They concluded that the Sun could last for a few million years. Even then, geology and biology made it clear that the Earth had been around for billions of years; gravity is not enough.

<sup>56</sup>'Adiabatic' is sometimes used to describe systems that do not exchange heat with their surroundings, and sometimes to describe systems that change slowly enough to stay in equilibrium or in their quantum ground states. Here we mean both.

*too rapidly to stay in equilibrium, will the entropy change be positive or negative? Does a fall from equilibrium by itself explain the low-entropy initial state that determines the arrow of time?*

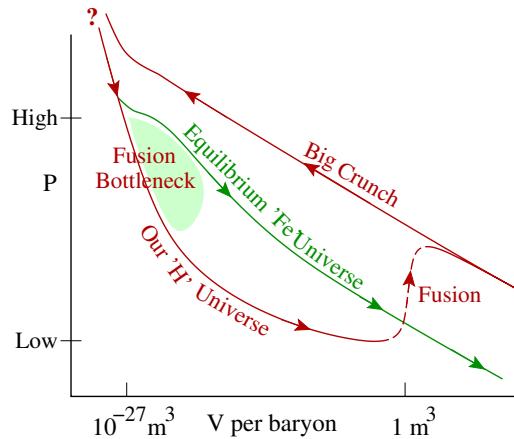


**Fig. 5.20 The Universe as a piston.** Imagine the expansion of the Universe as the decompression of a gas of photons and nucleons. The Universe does not exchange heat with the ‘space-time piston’ as it expands, but it can exchange work with the piston. (The image is galaxy cluster MACS 1206 [34]. We shall ignore the formation of galaxies and stars, however, and compare the formation of an H-Universe of hydrogen gas to a Fe-Universe of iron atoms.)

The very early Universe was so hot that any heavier nuclei quickly evaporated into protons and neutrons. Between a few seconds and a couple of minutes after the Big Bang, protons and neutrons began to fuse into light nuclei – mostly helium. The rates for conversion to heavier elements were too slow, however, by the time they became entropically favorable. Almost all of the heavier elements on Earth were formed later, inside stars (see Exercise 6.22).

How would our Universe compare with one with faster nucleosynthesis reactions, which remained in equilibrium? To simplify the problem, we shall ignore gravity (except insofar as it drives the expansion of the Universe), chemistry (no molecule or iron lump formation), and the distinction between protons and neutrons (just keep track of the baryons). We shall compare two universes (see Fig. 5.21). In the Fe-Universe nucleosynthesis is very fast; this universe begins as a gas of protons at high temperatures but stays in equilibrium, forming iron atoms early on in the expan-

sion. In the H-Universe we imagine nucleosynthesis is so slow that no fusion takes place; it evolves as an adiabatically<sup>57</sup> expanding gas of hydrogen atoms.<sup>58</sup>



**Fig. 5.21 PV diagram for the universe.** In our Universe, the baryons largely did not fuse into heavier elements, because the reaction rates were too slow. Only in the eons since have stars been fusing hydrogen into heavier elements, glowing with non-equilibrium flows of low-entropy sunlight. Here we compare a simplified ‘H’ version of our Universe with an equilibrium ‘Fe’-universe where nucleosynthesis finished at early times. This schematic shows the pressure-volume paths as the ‘piston’ of Fig. 5.20 is moved. The ‘H-Universe’ has nucleosynthesis completely suppressed; the proton gas adiabatically cools into hydrogen atoms, until a time near the present when hydrogen is fused (dashed line). In the ‘Fe-Universe’, nucleosynthesis stayed in equilibrium as the universe expanded, leaving only iron atoms. Finally, in the ‘Big Crunch’ the H-Universe re-compresses, re-forming a hot proton gas.

We want to compare the entropy of the H-Universe with that of the Fe-Universe.

(b) *How much does the entropy of the Fe-Universe change as it adiabatically expands from an initial gas of hot protons, through nucleosynthesis, to a gas of iron atoms at the current baryon density? How much does the entropy of the H-Universe change as it expands adiabatically into a gas of*

<sup>57</sup>The nuclear degrees of freedom, which do not evolve, can be viewed as decoupled from the photons and the atomic motions.

<sup>58</sup>Without stars, the intelligent beings in the H-Universe will need to scoop up hydrogen with Bussard ramjets, fuse it into heavier elements, and use the resulting heavy elements (and abundant energy) to build their progeny and their technological support system. But they are better off than the inhabitants of the Fe-Universe.

*hydrogen atoms at the current density?* (No calculations should be needed, but a cogent argument with a precise answer is needed for both entropy changes.)

Some facts and parameters. Wikipedia tells us that nucleosynthesis happened “A few minutes into the expansion, when the temperature was about a billion ... Kelvin and the density was about that of air”. The current microwave background radiation temperature is  $T_{\text{CMB}} = 2.725\text{K}$ ; let us for convenience set the temperature when nucleosynthesis would happen in equilibrium to a billion times this value,  $T_{\text{reaction}} = 2.725 \times 10^9\text{K}$ . (We estimate the actual temperature in Exercise 6.22.) The thermal photons and the baryons are roughly conserved during the expansion, and the baryon-to-(microwave background photon) ratio is estimated to be  $\eta = [\text{Baryons}]/[\text{Photons}] \sim 5 \times 10^{-10}$  (Exercise 6.22). Hence the temperature, energy density, and pressure are almost completely set by the photon density during this era (see part (c) below). The red shift of the photons increases their wavelength proportional to the expansion, and thus reduces their energy (and hence the temperature) inversely with the expansion rate. Thus the number density of photons and baryons goes as  $T^3$ . The current density of baryons is approximately one per cubic meter,<sup>59</sup> and thus we assume that the equilibrium synthesis happens at a density of  $10^{27}$  per cubic meter. Fusing 56 protons into  $^{56}\text{Fe}$  releases an energy of about  $\Delta E = 5 \times 10^8 \text{ eV}$  or about  $\Delta E/56 = 1.4 \times 10^{-12} \text{ joules/baryon}$ .

Truths about the photon gas (see Exercise (7.15)). The number density for the photon gas is  $N/V = (2\zeta(3)/\pi^2)(k_B T)^3/(c^3 \hbar^3)$ , the internal energy density is  $E/V = (1/15)\pi^2(k_B T)^4/(c^3 \hbar^3) = 4T^4\sigma/c$ , and the pressure is  $P = E/3V$ .<sup>60</sup> The baryons do not contribute a significant pressure or kinetic energy density.

(c) Calculate the temperature rise of the Fe-Universe when the hydrogen fuses into iron in equilibrium at high temperature, pressure, and density, releasing thermal photons. Calculate the difference in temperature for the Fe-Universe filled with iron and the H-Universe filled with hydrogen gas, when both reach the current baryon density. (Hint: How does the temperature change with volume?)

(d) Calculate the temperature rise in the H-Universe if we fuse all the hydrogen into iron at current densities and pressures (dashed line in Figure 5.21), using our heat engines to turn the energy into thermal photons. Calculate the pressure rise. Calculate how much entropy would be released per baryon, in units of  $k_B$ .

The key is not that our Universe started in a low entropy state. It is that it has an untapped source of energy. Why, by postponing our fusion, do we end up with so much more energy? The energy for the matter and photons in an expanding universe is *not conserved*.

(e) Calculate the difference in work done per baryon by the photon gas during the expansion, between the Fe-Universe and the H-Universe. Compare it to the energy released by nucleosynthesis. (Hint: Calculate the temperature  $T(V)$  and the difference  $\Delta T(V)$  as a function of volume. Use that to determine  $\Delta P(V)$ .)

(f) Consider a hypothetical ‘big crunch’, where the universe contracts slowly after expanding (and nothing irreversible happens at low densities). In the post-burn H-Universe, will the pressure and volume return to its original value at volumes per baryon much smaller than  $10^{-27}$ ? (That is, will the P-V loop close in Fig. 5.21?) If not, why not? What will the ‘Fe-Universe Big Crunch’ return path be, assuming the return path is also in equilibrium?

<sup>59</sup>Very roughly.

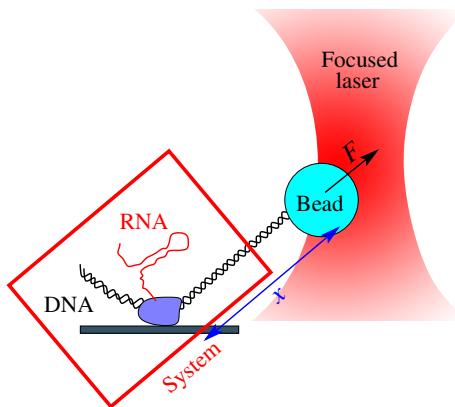
<sup>60</sup>Here  $\sigma = \pi^2 k_B^4 / (60 \hbar^3 c^2) = 5.67 \times 10^{-8} \text{ J/m}^2 \text{sK}^4$  is the Stefan-Boltzmann constant.

## Chapter 6: Free energies

### Exercises

**6.4 Molecular motors and free energies.**<sup>61</sup> (Biology) ⓘ

Figure 6.22 shows the set-up of an experiment on the molecular motor RNA polymerase that transcribes DNA into RNA.<sup>62</sup> Choosing a good ensemble for this system is a bit involved. It is under two constant forces ( $F$  and pressure), and involves complicated chemistry and biology. Nonetheless, you know some things based on fundamental principles. Let us consider the optical trap and the distant fluid as being part of the external environment, and define the ‘system’ as the local region of DNA, the RNA, motor, and the fluid and local molecules in a region immediately enclosing the region, as shown in Fig. 6.22.



**Fig. 6.22** RNA polymerase molecular motor attached to a glass slide is pulling along a DNA molecule (transcribing it into RNA). The opposite end of the DNA molecule is attached to a bead which is being pulled by an optical trap with a constant external force  $F$ . Let the distance from the motor to the bead be  $x$ ;

thus the motor is trying to move to decrease  $x$  and the force is trying to increase  $x$ .

(a) *Without knowing anything further about the chemistry or biology in the system, which of the following must be true on average, in all cases?*

(T) (F) *The total entropy of the Universe (the system, bead, trap, laser beam, ...) must increase or stay unchanged with time.*

(T) (F) *The entropy  $S_s$  of the system cannot decrease with time.*

(T) (F) *The total energy  $E_T$  of the Universe must decrease with time.*

(T) (F) *The energy  $E_s$  of the system cannot increase with time.*

(T) (F)  *$G_s - Fx = E_s - TS_s + PV_s - Fx$  cannot increase with time, where  $G_s$  is the Gibbs free energy of the system.*

Related formula:  $G = E - TS + PV$ .

(Hint: Precisely two of the answers are correct.)

The sequence of monomers on the RNA can encode information for building proteins, and can also cause the RNA to fold into shapes that are important to its function. One of the most important such structures is the hairpin (Fig. 6.23). Experimentalists study the strength of these hairpins by pulling on them (also with laser tweezers). Under a sufficiently large force, the hairpin will unzip. Near the threshold for unzipping, the RNA is found to jump between the zipped and unzipped states, giving *telegraph noise*<sup>63</sup> (Fig. 6.24). Just as the current in a telegraph signal is either on or off, these systems are bistable and make transitions from one state to the other; they are a two-state system.

<sup>61</sup>This exercise was developed with the help of Michelle Wang.

<sup>62</sup>RNA, ribonucleic acid, is a long polymer like DNA, with many functions in living cells. It has four monomer units (A, U, C, and G: Adenine, Uracil, Cytosine, and Guanine); DNA has T (Thymine) instead of Uracil. Transcription just copies the DNA sequence letter for letter into RNA, except for this substitution.

<sup>63</sup>Like a telegraph key going on and off at different intervals to send dots and dashes, a system showing telegraph noise jumps between two states at random intervals.

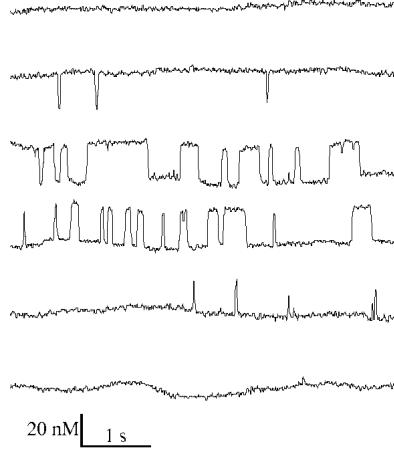


**Fig. 6.23 Hairpins in RNA.** (Reprinted with permission from Liphardt et al. [27], ©2001 AAAS.) A length of RNA attaches to an inverted, complementary strand immediately following, forming a hairpin fold.

The two RNA configurations presumably have different energies ( $E_z, E_u$ ), entropies ( $S_z, S_u$ ), and volumes ( $V_z, V_u$ ) for the local region around the zipped and unzipped states, respectively. The environment is at temperature  $T$  and pressure  $P$ . Let  $L = L_u - L_z$  be the extra length of RNA in the unzipped state. Let  $\rho_z$  be the fraction of the time our molecule is zipped at a given external force  $F$ , and  $\rho_u = 1 - \rho_z$  be the unzipped fraction of time.

- (b) Of the following statements, which are true, assuming that the pulled RNA is in equilibrium?
- (T) (F)  $\rho_z/\rho_u = \exp((S_z^{\text{tot}} - S_u^{\text{tot}})/k_B)$ , where  $S_z^{\text{tot}}$  and  $S_u^{\text{tot}}$  are the total entropy of the Universe when the RNA is in the zipped and unzipped states, respectively.
  - (T) (F)  $\rho_z/\rho_u = \exp(-(E_z - E_u)/k_B T)$ .
  - (T) (F)  $\rho_z/\rho_u = \exp(-(G_z - G_u)/k_B T)$ , where  $G_z = E_z - TS_z + PV_z$  and  $G_u = E_u - TS_u + PV_u$  are the Gibbs energies in the two states.
  - (T) (F)  $\rho_z/\rho_u = \exp(-(G_z - G_u + FL)/k_B T)$ , where  $L$  is the extra length of the unzipped RNA and  $F$  is the applied force.

<sup>64</sup>Joseph-Louis Lagrange (1736–1813). See [31, section 12, p. 331].



**Fig. 6.24 Telegraph noise in RNA unzipping.** (Reprinted with permission from Liphardt et al. [27], ©2001 AAAS.) As the force increases, the fraction of time spent in the zipped state decreases.

## 6.6 Lagrange.<sup>64</sup> (Thermodynamics) ⑦

In this exercise, we explore a statistical mechanics from an information-theory perspective. We shall derive the canonical and grand canonical perspective by maximizing the entropy subject to constraints on the energy and the number of particles. In the process, we introduce the mathematics of *Lagrange multipliers*.

Lagrange multipliers allow one to find the extremum of a function  $f(\mathbf{x})$  given a constraint  $g(\mathbf{x}) = g_0$ . One sets the derivative of

$$f(\mathbf{x}) + \lambda(g(\mathbf{x}) - g_0) \quad (6.61)$$

with respect to  $\lambda$  and  $\mathbf{x}$  to zero. The derivatives with respect to components of  $\mathbf{x}$  then include terms involving  $\lambda$  which push  $\mathbf{x}$ . Setting the derivative with respect to  $\lambda$  to zero determines the value  $\lambda$  needed to enforce the constraint.

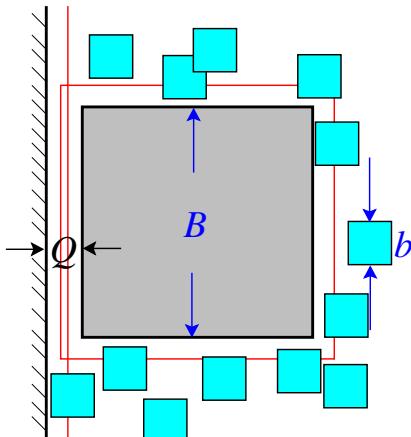
We will use Lagrange multipliers to find the maximum of the non-equilibrium entropy

$$S = -k_B \sum p_i \log p_i$$

constraining the normalization, energy, and number.

- (a) Microcanonical. Using a Lagrange multiplier to enforce the normalization  $\sum_i p_i = 1$ , show that the probability distribution that extremizes the entropy is a constant (the microcanonical distribution).
- (b) Canonical. Add another Lagrange multiplier to fix the mean energy  $\langle E \rangle = \sum_i E_i p_i$ . Show that the canonical distribution maximizes the entropy given the constraints of normalization and fixed energy.
- (c) Grand canonical. Summing over different numbers of particles  $N$  and adding the constraint that the average number is  $\langle N \rangle = \sum_i N_i p_i$ , show that you get the grand canonical distribution by maximizing the entropy.
- (d) Links to ensembles. Write the grand partition function  $\Xi$ , the temperature  $T$ , and the chemical potential  $\mu$  in terms of your three Lagrange multipliers in part (c).

### 6.13 Pollen and hard squares. ①



**Fig. 6.25** Square pollen grain in fluid of oriented square molecules, next to a wall. The thin lines represent the exclusion region around the pollen grain and away from the wall.

Objects embedded in a gas will have an effective attractive force at short distances, when the gas molecules can no longer fit between the objects. This is called the *depletion force*, and is a common tool in physics to get large particles to clump together. One can view this force as a pressure imbalance (no collisions from one side) or as an entropic attraction.

Let us model the entropic attraction between a pollen grain and a wall using a two-dimensional ideal gas of classical indistinguishable particles as the fluid. For convenience, we imagine that the pollen grain and the fluid are formed from square particles lined up with the axes of the box, of lengths  $B$  and  $b$ , respectively (Fig. 6.25). We assume *no interaction* between the ideal gas molecules (unlike in Exercise 3.5), but the potential energy is infinite if the gas molecules overlap with the pollen grain or with the wall. The container as a whole has one pollen grain,  $N$  gas molecules, and total area  $L \times L$ .

Assume the pollen grain is close to only one wall. Let the distance from the surface of the wall to the closest face of the pollen grain be  $Q$ . (A similar square-particle problem with interacting small molecules is studied in [21].)

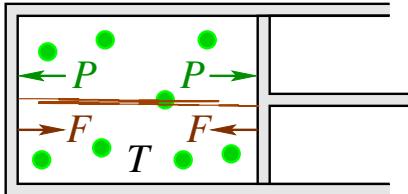
(a) Plot the area  $A(Q)$  available for the gas molecules, in units of  $(\text{length})^2$ , in three steps. Plot  $A(Q \gg 0)$  when the pollen grain is far from the wall. At what distance  $Q_0$  does this change? Plot  $A(0)$ , the area available when the pollen grain is touching the wall. Finally, how does the area vary with  $Q$  (linearly, quadratically, step function ...) between zero and  $Q_0$ ? Plot this, and give formulae for  $A(Q)$  as a function of  $Q$  for the two relevant regions,  $Q < Q_0$  and  $Q > Q_0$ .

(b) What is the configuration-space volume  $\Omega(Q)$  for the gas, in units of  $(\text{length})^{2N}$ ? What is the configurational entropy of the ideal gas,  $S(Q)$ ? (Write your answers here in terms of  $A(Q)$ .) Your answers to part (b) can be viewed as giving a free energy for the pollen grain after integrating over the gas degrees of freedom (also known as a partial trace, or coarse-grained free energy).

(c) What is the resulting coarse-grained free energy of the pollen grain,  $\mathcal{F}(Q) = E - TS(Q)$ , in the two regions  $Q > b$  and  $Q < b$ ? Use  $\mathcal{F}(Q)$  to calculate the force on the pollen grain for  $Q < b$ . Is the force positive (away from the wall) or negative? Why?

(d) Directly calculate the force due to the ideal gas pressure on the far side of the pollen grain, in terms of  $A(Q)$ . Compare it to the force from the partial trace in part (c). Why is there no balancing force from the other side? Effectively how ‘long’ is the far side of the pollen grain?

### 6.15 Entropic forces: Gas vs. rubber band. ③



**Fig. 6.26 Piston with rubber band.** The gas in the piston exerts an outward force. A rubber band stretches across the piston, exerting an inward tensile force. The piston adjusts to its thermal equilibrium distance, with no external force applied.

In Fig. 6.26, we see a piston filled with a gas which exerts a pressure  $P$ . The volume outside the piston may be assumed to be empty (vacuum,  $P_{\text{ext}} = 0$ ), and no external forces are applied. A rubber band attached to the piston and the far wall resists the outward motion with an inward force  $F$ . The piston moves to its equilibrium position.

Assume the gas and the rubber band are modeled as both having zero potential energy  $E_{\text{int}} = 0$  for all accessible states. (So, for example, perhaps the gas was modeled as hard spheres, similar to Exercise 3.5, and the rubber band as zero-energy random walk chain as in Exercise 5.12).

*What will happen to the position of the piston if the temperature is increased by a factor of two? Why? Give an elegant and concise reason for your answer.* (Your answer should involve no model-specific calculations. It should probably refer to  $P_{\text{ext}}$  and  $E_{\text{int}}$ . Hint: Consider the name of this exercise.)

### 6.16 Rubber band free energy. (Condensed matter) ①

This exercise illustrates the convenience of choosing the right ensemble to decouple systems into independent subunits. Consider again the random-walk model of a rubber band in Exercise 5.12 –  $N$  segments of length  $d$ , connected by hinges that had zero energy both when straight and when

bent by  $180^\circ$ . We there calculated its entropy  $S(L)$  as a function of the folded length  $L$ , used Stirling's formula to simplify the combinatorics, and found its spring constant  $K$  at  $L = 0$ .

Here we shall do the same calculation, without the combinatorics. Instead of calculating the entropy  $S(L)$  at fixed  $L$ ,<sup>65</sup> we shall work at fixed temperature and fixed energy, calculating an appropriate free energy  $\chi(T, F)$ . View our model rubber band as a collection of segments  $s_n = \pm 1$  of length  $d$ , so  $L = d \sum_{n=1}^N s_n$ . Let  $F$  be the force exerted by the band on the world (negative for positive  $L$ ).

(a) *Write a Hamiltonian for our rubber band under an external force.<sup>66</sup> Show that it can be written as a sum of uncoupled Hamiltonians  $\mathcal{H}_n$ , one for each link of the rubber band model. Solve for the partition function  $X_n$  of one link, and then use this to solve for the partition function  $X(T, F)$ .* (Hint: Just as for the canonical ensemble, the partition function for a sum of uncoupled Hamiltonians is the product of the partition functions of the individual Hamiltonians.)

(b) *Calculate the associated thermodynamic potential  $\chi(T, F)$ . Derive the abstract formula for  $\langle L \rangle$  as a derivative of  $\chi$ , in analogy with the calculation in eqn 6.11. Find the spring constant  $K = \partial F / \partial L$  in terms of  $\chi$ . Evaluate it for our rubber band free energy at  $F = 0$ .*

The calculations of this exercise, and of the original rubber band Exercise 5.12, are closely related to Exercise 6.3 on negative temperatures.

### 6.17 Rubber band: Illustrating the formalism. (Condensed matter) ②

Consider the rubber band of Exercise 5.12. View it as a system that can exchange ‘length’  $L_M = d(2M - N)$  with the external world, with the force  $F$  being the restoring force of the rubber band (negative for positive  $L$ ).

(a) *Use the number of configurations  $\Omega(L_M)$  at fixed length to write a formal expression for  $X(T, F)$  as a sum over  $M$ .* (See the analogous third line in eqn 6.37.)

(b) *Write the corresponding thermodynamic potential  $\chi = -k_B T \log(X(T, F))$  in terms of  $E$ ,  $T$ ,  $S$ ,  $F$ , and  $L$ .* (For example, eqn 6.17 says

<sup>65</sup>Usually we would write  $S(E, L)$ , but our model rubber band is purely entropic –  $E = 0$  for all states.

<sup>66</sup>Our model rubber band has no internal energy. This Hamiltonian includes the energy exchanged with the outside world when the length of the rubber band changes, just as the Gibbs free energy  $G(T, P) = E - TS + PV$  includes the energy needed to borrow volume from the external world.

that a system connected to an external bath is described by  $\langle E \rangle - TS$ ; what analogous expression applies here?)

### 6.18 Langevin dynamics. (Computation, Dynamical systems) $\oplus$

Even though energy is conserved, macroscopic objects like pendulums and rubber balls tend to minimize their potential and kinetic energies unless they are externally forced. Section 6.5 explains that these energies get transferred into heat – they get lost into the  $6N$  internal degrees of freedom. (See also Exercise 10.7.) Equation 6.49 suggests that these microscopic degrees of freedom can produce both friction  $\gamma$  and noise  $\xi(t)$ .

(a) First consider the equation  $\ddot{h} = \xi(t)$  for the velocity of a particle of mass  $m$ . Assume  $\xi(t)$  gives a ‘kick’ to the particle at regular intervals separated by  $\Delta t$ :  $\xi(t) = \sum_{j=-\infty}^{\infty} \xi_j \delta(t-j\Delta t)$ , with  $\langle \xi_i^2 \rangle = \sigma^2$  and  $\langle \xi_i \xi_j \rangle = 0$  for  $i \neq j$ . Argue that this leads to a random walk in momentum space. Will this alone lead to a thermal equilibrium state?

(b) Now include the effects of damping, which reduces the momentum by  $\exp(-\gamma\Delta t)$  during each time step, so  $p_j = \exp(-\gamma\Delta t)p_{j-1} + m\xi_j$ . By summing the geometrical series, find  $\langle p_j^2 \rangle$  in terms of the mass  $m$ , the noise  $\sigma$ , the damping  $\gamma$ , and the time step  $\Delta t$ .

(c) What is the relation between the noise and the damping needed to make  $\langle p^2 / 2m \rangle = \frac{1}{2}k_B T$ , as needed for thermal equilibrium? How does the relation simplify in the small time-step limit  $\Delta t \rightarrow 0$ ?

This is *Langevin dynamics*, which both is a way of modeling the effects of the environment on macroscopic systems and a numerical method for simulating systems at fixed temperature (*i.e.*, in the canonical ensemble).

### 6.19 Newton’s theory of sound. $\oplus$

Note that our derivation of the diffusion equation 6.67 for perfume in still air ignored the air molecule degrees of freedom. Without saying so, we have integrated out the air degrees of freedom – allowing the perfume molecules to exchange momentum and energy locally with a ‘bath’ of air molecules. Let us consider the effective equation of motion for an ideal gas with a locally conserved momentum, but at constant temperature

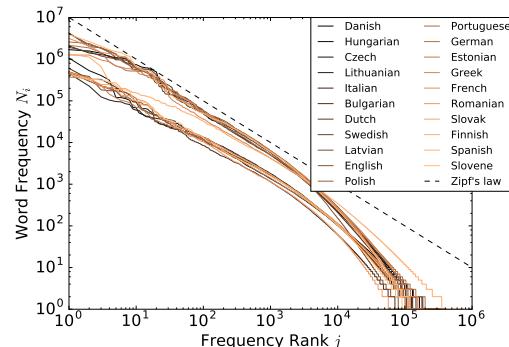
(exchanging heat but not momentum with its environment.)

If the mass of the ideal gas atoms is  $m$ , the momentum density is  $\Pi(x) = mJ(x)$ , where  $J(x) = \rho(x)v(x)$  is the particle current and  $v(x)$  is the mean particle velocity. Newton’s law  $F = \dot{p} = ma$  turned into densities tells us that  $\dot{\Pi} = \mathcal{F}(x)$ , where the force density  $\mathcal{F}(x) = -\rho(x)\partial\mu/\partial x$ , the density of particles times the force per particle.

Using  $\partial\rho/\partial t = -\partial J/\partial x$ , find  $\partial^2\rho/\partial t^2$  in terms of  $\rho$  and  $\partial\mu/\partial x$ . Use the ideal gas law eqn 6.65 for  $\mu(x)$  to find  $\partial^2\rho/\partial t^2$ . What law emerges? What is the speed of sound?

(Newton originally proposed this calculation in the *Principia*; it was later pointed out by Laplace that sound vibrations are too fast to exchange heat with their surroundings.)

### 6.20 Word frequency: Zipf’s law. (Linguistics) $\circledcirc$



**Fig. 6.27 Zipf’s law for word frequencies.** Frequency of the  $j$ th most used word, versus rank  $j$  (courtesy Colin Clement). Data is from the Proceedings of the European Parliament, 1996–2011 [25].

The words we use to communicate with one another show a fascinating pattern. Naturally, we use some words more frequently than others. In English<sup>67</sup> ‘the’ is the most frequently occurring word (about 7% of all words), followed by ‘of’ (3.5%), followed by ‘and’ (about 3%). Following down the list, the word ranked  $j$  in usage has frequency roughly proportional to  $1/j$  – Zipf’s law

<sup>67</sup>Sampled in the Brown corpus, according to Wikipedia. Notice, surprisingly, that the rankings of even the first few words differ depending on the corpus; the Oxford English corpus says the highest ranked words are ‘the’, ‘be’, ‘to’, ...

(Fig. 6.27). This law holds approximately for all human languages. There are many explanations offered for this, but no clear consensus.

Here we explore a version of one of many explanations [36, 39] for Zipf's law – that it maximizes the communication needed for a given effort. This will build on our understanding of Shannon entropy and communications, and will provide an interesting analogue to our derivation in Exercise 6.6 of different statistical ensembles as maximizing entropy subject to constraints on probability, energy, and number.

First, how much information is communicated by a common word? Surely the word 'primeval' tells us more than the word 'something' – the former immediately tells us our message is about primitive times or traits, even though the latter is longer. How can we measure the information in a word? Shannon tells us that a communication channel used to transmit message # $i$  with probability  $p_i$  has an information capacity of  $-k_s \sum p_i \log p_i$  bits/message.

(a) Argue that the information from message # $i$  is given by  $k_s \log 1/p_i$ . This quantity is called the *surprise*: the number of bits of information learned from the message. Using each word of our language as a message, argue that

$$I = k_s \sum_j N_j \log(1/p_j) \quad (6.62)$$

is the total information transmitted in a series of  $N$  words with word # $j$  being transmitted with probability  $p_j$ .

Second, how much effort is needed to communicate a given word? Here the presumption is that an unusual word (like 'primeval',  $j = 17627$ ) will take more effort to transmit and receive than a simple word (like 'something',  $j = 287$ ). Shannon would say that a good encoding scheme will use a short string to compress the common word 'something' and a longer string to compress 'primeval'. Even though the two words are of the same length in letters, we imagine that our brains engage in a kind of compression algorithm too. If we restrict

our compression to whole words, and ignore correlations between words, then Shannon would say that the optimal encoding would take the smallest words into the shortest strings, so 'the' would be given the string '0', 'something' would be given '100011111' (287 base 2), and 'primeval' would be '100010011011011'.

We are not interested in maximizing the information per word, but maximizing the information per unit of effort.

Let  $C_j$  be the effort needed to speak and hear the word of rank  $j$  (the number of characters needed to transmit the word). Thus the effort for a series of  $N$  words with frequency  $N_j$  of word  $j$  is

$$E = \sum_j N_j C_j. \quad (6.63)$$

Notice that in this exercise we measure effort in *characters*, while we measure information in transmitted *letters*. Usually our characters and letters both have two states ('0' and '1'), but later we shall add another space character ' ' to our transmission channel.

While most of the exercises will leave the costs undetermined, we shall occasionally test our calculation on two special cases. One is from Exercise 5.15, where the letters (instead of words) are encoded in strings of length  $C_1^{\text{Abc}} = 1$ ,  $C_2^{\text{Abc}} = 2$ , and  $C_3^{\text{Abc}} = 2$ , and the frequencies in a long message are  $N_j = N/2$ ,  $N/4$ , and  $N/4$ , respectively. The other is a rough approximation to an optimal encoding of words, as above for 'something' and 'primeval'. The number of bits needed to transmit a word of rank  $j$  will be about  $\log_2 j$ .<sup>68</sup> Our brains do not encode in binary, but we shall pretend that the effort for hearing word # $j$  in the ranking is  $C_j^{\text{word}} = \log_2 j$ .

We shall study the optimal probability distribution for the words in our language by maximizing the information  $I$  transmitted with respect to the  $N_j$ , given a total effort  $E$ . We assume our message is long enough that  $N_j$  is well approximated by  $N p_j$ .<sup>69</sup> We have two constraints, eqn 6.63 and the total normalization  $N = \sum_j N_j$ .

<sup>68</sup>This depends on how words are encoded. For example, if there is a separate stop character separating words, then we can use '0' for  $j = 1$ , '00' for  $j = 3$ , '000' for  $j = 7$ , leading to  $C_j = \lceil \log_2(j+2) \rceil + s - 1$ , where  $s$  is the effort of the stop character, and  $\lceil x \rceil = \text{ceil}(x)$  is the lowest integer larger than or equal to  $x$ . In this encoding, 'something' becomes '00100000', and 'primeval' becomes '00010011011100'. But this demands a special way of transmitting the stop character.

<sup>69</sup>This is analogous to, say, pressure  $P$  in thermodynamics, which is well approximated in a given large system by  $\langle P \rangle$ .

Remember that we can derive the Boltzmann distribution by optimizing the entropy  $-k_B \sum p_i \log p_i$  while constraining the mean effort  $\sum E_i p_i$  and the normalization  $\sum p_i = 1$  using Lagrange multipliers (Exercise 6.6). We shall use the same strategy here to examine the optimal word frequency distribution. We thus want to find the extremum of

$$\begin{aligned} \mathcal{L} = k_S \sum_j N_j \log(N/N_j) \\ - \beta(\sum_j N_j C_j - E) \\ - \lambda(\sum_j N_j / N - 1) \end{aligned} \quad (6.64)$$

with respect to  $N_j$ ,  $N$ ,  $\beta$ , and  $\lambda$ . Remember  $k_S = 1/\log(2)$ .

(b) Give the equations found by varying  $\mathcal{L}$  with respect to  $N$ . Solve for  $\lambda$ .

(c) Give the equations found by varying  $\mathcal{L}$  with respect to  $N_j$ . Using your result from part (b), solve for  $N_j$  in terms of  $\beta$ ,  $C_j$ , and  $N$ .

(d) Using the three-letter language effort  $C_j^{\text{Abc}}$  above, use your result from part (c) to solve for  $\beta$ . (Hint: Make  $N = N_1 + N_2 + N_3$ .) Do the resulting optimal letter frequencies agree with those given above? Using the effort  $C_j^{\text{word}} = \log_2 j$ , give the power law governing the word distribution, in terms of  $\beta$ ,  $N$ , and  $\lambda$ . If the word distribution is presumed to continue to  $j \rightarrow \infty$ , at what value of  $\beta$  does it become impossible to keep the distribution normalized? How does this extreme value compare with Zipf's law?

You might ask what value for  $\beta$  we find for our logarithmic compression  $C^{\text{word}}$ , by using the normalization (as you did for  $C^{\text{Abc}}$  in part (d)). Note that the unmodified use of  $C_j^{\text{word}} = \log_2(j)$  does not make sense, since the most common word with  $j = 1$  costs no effort ( $\log_2(1) = 0$ ). In our brain-compressed series of words, one must imagine some marker corresponding to the extra space we use between each word. Suppose we model the extra marker as a third option for a 'letter', so each symbol being transmitted has three 'character states' (0, 1, ' ') rather than two. If we add one extra space marker separating every word, so  $C_j = C^{\text{word}} + 1 = 1 + \log_2(j)$ , simulations with 100 total words give  $\beta \sim 1.4$  – cutting off the probability of rare words pretty drastically. (Oddly, if one

adds two extra space characters between words, one gets pretty close to Zipf's law and optimal compression.)

Now we turn to the information  $I$ , the effort  $E$ , and the analytical calculation of  $\beta$ . We argued in Exercises 5.15 that a communication channel reliably transmitting bits can transmit no more than one unit of Shannon information entropy per bit. In this exercise, we are optimizing the information transfer over the frequencies  $N_j$  for a given language with words of cost  $C_j$ . A poor choice for the  $C_j$  will not saturate the Shannon bound.

(e) What does this general bound for any language imply about the ratio  $I/E$ , for effort measured in two-state characters? If we include an extra marker as an allowed 'character state' in the channel, what would the maximum Shannon entropy be per character? (Now  $I$  is measured in two-state letters, and  $E$  is measured in three-state characters.) What bound do we thus derive for  $I/E$  if spaces are allowed?

Instead of determining  $\beta$  from the normalization, we can take derivatives of  $\mathcal{L}$  to find it analytically. It turns out to be related to the information  $I$  and the effort  $E$  (eqns 6.62 and 6.63).

(f) Give the equation found by varying  $\mathcal{L}$  with respect to  $\beta$ . Using your results from part (c), solve for the effort  $E$  in terms of  $\beta$ , the number of words  $N$  and the  $C_i$ . Is the number of words linear in the effort budget?

(g) Solve for the information  $I$ . Solve for the information per unit effort  $I/E$ , in terms of  $\beta$ . How close does our three-word language  $C^{\text{Abc}}$  (which needs no spaces between words) approach Shannon's fundamental bound? How close does my computation for  $\beta$  using  $C^{\text{word}} + 1$  approach the bound for three allowed characters?

It is not clear from this analysis why natural languages form an extremal limit of optimal information transmission. After all,  $C^{\text{Abc}}$ , or even a two-word language of 0's and 1's, can be used to optimize the Shannon information transmission per symbol. Perhaps we need a large vocabulary to support an advanced civilization.<sup>70</sup> In that case, given the logarithmic nature of compression algorithms (and a perhaps unlikely assumption that our brain uses word compression to process language), Zipf's law arguably arises as the distribution which optimizes communication per unit

<sup>70</sup>Perhaps a large vocabulary enhances reproductive success? Communicating with patterns of two types of grunts might not compete in the social arena, no matter how efficient.

effort while also maximizing the vocabulary size.

**6.21 Epidemics and zombies.<sup>71</sup>** (Biology, Epidemiology) ③

This exercise is based on Alemi and Bierbaum's class project, published in *You can run, you can hide: The epidemiology and statistical mechanics of zombies* [1]. See also the Zombietown site [9] and simulator [10].

Epidemics are studied by disease control specialists using statistical methods, modeling the propagation of the disease as susceptible people are infected, infect others, and recover. The SIR model is the simplest commonly studied model, with three coupled differential equations:  $\dot{S} = -\beta SI$  reflects the rate  $\beta$  at which each infected person  $I$  infects each susceptible person  $S$ , and  $\dot{R} = \kappa I$  reflects the rate  $\kappa$  that each infected person joins the recovered population  $R$ .

(a) *What is the equation for  $\dot{I}$  implied by these first two equations, assuming no infected people die or shift groups other than by new infections or recoveries?*

We shall use a less common, but even simpler SZR model [1], designed to predict the evolution of a zombie outbreak.

$$\begin{aligned}\dot{S} &= -\beta SZ \\ \dot{Z} &= (\beta - \kappa)SZ \\ \dot{R} &= \kappa SZ.\end{aligned}\quad (6.65)$$

Here the zombie population  $Z$  never recovers, but if it is destroyed by a member of the surviving population  $S$ , it joins the removed population  $R$ . The bite parameter  $\beta$  describes the rate at which a zombie bites a human it encounters, and the kill parameter  $\kappa$  gives the rate that a human may destroy a zombie it finds.

The SZR model is even simpler than the SIR model, in that one can write an explicit solution for the evolution as a function of time. We do so in two steps.

(b) *Argue that the only stationary states have all zombies or all humans.* Both  $\dot{S}$  and  $\dot{Z}$  are linear in  $SZ$ , so there must be a linear combination of the two that has no time dependence. *Show that  $P = Z + (1 - \kappa/\beta)S$  satisfies  $\dot{P} = 0$ .* Argue from these two facts that for  $P < 0$  the zombies must lose.

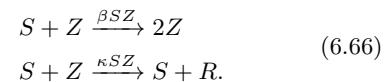
(c) *Show that  $\chi = S/Z$  satisfies  $\dot{\chi} = \gamma\chi$ , and so  $\chi(t) = \chi_0 \exp(\gamma t)$ .* Show that  $\gamma = -\beta P$ .<sup>72</sup> Check

that this answer concurs with your criterion for human survival in part (b).

So the fraction of the doomed species exponentially decays, and the population of the surviving species is determined by  $P$ . If desired, one could use the added equation  $S(t) + Z(t) + R(t) = N$  with your answers to parts (b) and (c) to solve analytically for the explicit time evolution of  $S$ ,  $Z$ , and  $R$ . We shall solve these equations numerically instead (below).

Suppose now that we start with a single zombie  $Z_0 = 1$ , and the number of humans  $S_0$  is large. It would seem from our equation for our invariant  $P$  from part (b) that if the bite parameter  $\beta$  is greater than the kill parameter  $\kappa$  that the humans are doomed. But surely there is some chance that we will be lucky, and kill the zombie before it bites any of us? This will happen with probability  $\kappa/(\beta + \kappa)$ . If we fail the first time, we can hope to destroy two zombies before either bites again...

Here is where the statistical fluctuations become important. The differential eqns 6.65 are a continuum approximation to the discrete transitions between integer numbers of the three species. These three equations are similar to reaction rate equations in chemistry (as in eqn 6.50) with molecules replaced by people:



Just as for disease outbreaks, if the number of molecules is small then chemical reactions exhibit important statistical fluctuations. These fluctuations are important, for example, for the biology inside cells, where the numbers of a given species of RNA or protein can be small, and the number of DNA sites engaging in creating RNA is usually either zero or one (see Exercises 8.10 and 8.11). We can simulate each individual bite and kill event for a population of  $S$  humans and  $Z$  zombies. (Indeed, this can be done rather efficiently for the entire population of the USA; see [1, 9, 10].) The time to the next event is an exponential random variable given by the total event rate. Which event happens next is then weighted by the individual rates for the different events.

<sup>71</sup>Hints for the computations can be found at the book Web site [49].

<sup>72</sup>Note that some of the solutions in reference [1] have typographical errors.

It is well known that decay rates add. Let us nonetheless derive this. Let the probability density of event type  $\#n$  be  $\rho_n(t)$ , with survival probability  $S_n(t) = 1 - \int_0^t \rho_n(\tau) d\tau$ . Let there be two types of events.

(d) Write the probability density of the first event,  $\rho_{\text{tot}}(t)$ , in terms of  $\rho_1$ ,  $\rho_2$ ,  $S_1$ , and  $S_2$ . (The final answer should involve no integrals or derivatives.) Specialize to systems with constant rates, which have an exponentially decaying survival time,  $S_n(t) = \exp(-\gamma_n t)$ . Show that the total event rate  $\gamma_{\text{tot}}$  is the sum  $\gamma_1 + \gamma_2$ . Show that the probability of the next event being of type  $\#1$  is  $\gamma_1/\gamma_{\text{tot}}$ .

To simulate one step of our discrete SZR model, we

- (i) find the total rate of events  $\gamma_{\text{tot}}$ ,
- (ii) increment  $t$  by  $\Delta t$ , a random number pulled from an exponential distribution with decay rate  $\gamma_{\text{tot}}$ ,
- (iii) choose to bite or to kill by choosing a random number uniform in  $(0, \gamma_{\text{tot}})$ , and checking if it is less than  $\gamma_1$ ,
- (iv) change  $S$ ,  $Z$ , and  $R$  appropriately for the event, and
- (v) perform any observations needed.

This is a simple example of the *Gillespie* algorithm, discussed in more detail in Exercises 8.10 and 8.11.

(e) Write a routine to use the *Gillespie* algorithm to solve the discrete SZR model for eqns 6.66, keeping track of  $t$ ,  $S$ ,  $Z$ , and  $R$  for each event, ending when  $S = 0$  or  $Z = 0$ , and adding an extra point at  $t_{\text{max}}$  if the system terminates early. Write a routine numerically solve the continuum equations 6.65. Use  $\beta = 0.001$ ,  $\kappa = 0.0008$ , and  $t_{\text{max}} = 5$ . (Should the zombies win?) Plot the two for initial conditions  $Z_0 = 100$ ,  $S_0 = 9900$ , and  $R_0 = 0$  (a simultaneous outbreak of a hundred zombies). Is the continuum limit faithfully describing the behavior for large numbers of zombies and humans?

Now let us examine the likelihood of zombies being stamped out despite their advantage in biting, if we start with only one zombie.

(f) Plot the continuum solution with the zombie simulation for twenty initial conditions using  $Z_0 = 1$ ,  $S_0 = 9999$ , and  $R_0 = 0$ . Are the simulations suffering from more fluctuations than they

did for larger number of zombies? Do you see evidence for zombie extinction early in the outbreak? What fraction of the initial outbreaks appear to have killed off all the humans? Zoom in and plot at early times ( $Z \leq 5$ ,  $t < 0.5$ ) and note a few trajectories where the first zombie is killed before biting, and trajectories where the zombie population goes extinct after reaching a peak population of two or three.

We can estimate the probability  $P_{\text{ext}}^\infty$  that a single zombie with bite rate larger than kill rate will go extinct before taking over a large initial human population  $S_0 \rightarrow \infty$ . As described in reference [1], the probability  $P_{\text{ext}}$  that the zombies go extinct, in the limit of many humans, is equal to the probability that the first one is destroyed, plus the probability that it bites first times the probability that both zombie lines go extinct:

$$P_{\text{ext}}^\infty = \frac{\kappa}{\beta + \kappa} + \frac{\beta}{\beta + \kappa} (P_{\text{ext}}^\infty)^2. \quad (6.67)$$

This can be solved to show  $P_{\text{ext}}^\infty = \kappa/\beta$ . Exactly this same argument holds for regular disease outbreaks. Similar arguments can be used to determine the likelihood that an advantageous gene mutation will take over a large population.

(g) Was the fraction of extinctions you observed in part (f) roughly given by the calculation above? Write a simpler routine that simulates the *Gillespie* algorithm over  $n$  epidemics, reporting the fraction in which zombies go extinct (observing only whether  $Z = 0$  happens before  $S = 0$ , ignoring the time and the trajectory). For  $S_0 = 9999$  and 1000 epidemics, how good is the prediction? Draw a plot of  $P_{\text{ext}}$  versus  $\log S_0$ , for  $S_0 = 1, 2, 4, \dots, 512$ , using 1000 epidemics each. How large must the human population be before our estimate  $P_{\text{ext}}^\infty$  is within about 10% of the correct answer?

## 6.22 Nucleosynthesis as a chemical reaction.<sup>73</sup> (Astrophysics) ③

The very early Universe was so hot that any heavier nuclei quickly evaporated into protons and neutrons. Between a few seconds and a couple of minutes after the Big Bang, protons and neutrons began to fuse into light nuclei – mostly helium. The energy barriers for conversion to heavier elements were too slow, however, by the time they became entropically favorable. Almost all of the heavier elements on Earth were formed later, inside stars.

<sup>73</sup>This exercise was developed in collaboration with Katherine Quinn.

In this exercise, we explore a hypothetical universe with faster nucleosynthesis reactions, or slower expansion, so that the reactions remained in equilibrium. Clearly the nucleons for the equilibrated Universe will all fuse into their most stable form.<sup>74</sup> The binding energy  $\Delta E$  released by fusing nucleons into  $^{56}\text{Fe}$  is about  $5 \times 10^8$  eV (about half a proton mass). In this exercise, we shall ignore the difference between protons and neutrons,<sup>75</sup> ignore nuclear excitations (assuming no internal entropy for the nuclei, so  $\Delta E$  is the free energy difference), and ignore the electrons (so protons are the same as hydrogen atoms, etc.) The creation of  $^{56}\text{Fe}$  from nucleons involves a complex cascade of reactions. We argued in Section (6.6) that however complicated the reactions, they must in net be described by the overall reaction, here



releasing an energy of about  $\Delta E = 5 \times 10^8$  eV or about  $\Delta E/56 = 1.4 \times 10^{-12}$  joules/baryon. We used the fact that most chemical species are dilute (and hence can be described by an ideal gas) to derive the corresponding law of mass-action, here

$$[\text{Fe}]/[\text{p}]^{56} = K_{\text{eq}}(T). \quad (6.69)$$

(Note that this equality, true in equilibrium no matter how messy the necessary reactions pathways, can be rationalized using the oversimplified picture that 56 protons must simultaneously collect in a small region to form an iron nucleus.) We then used the Helmholtz free energy for an ideal gas to calculate the reaction constant explicitly  $K_{\text{eq}}(T)$  for a similar reaction, in the ideal gas limit.

(a) Give symbolic formulas for  $K_{\text{eq}}(T)$  in eqn 6.69, assuming the nucleons form an approx-

imate ideal gas. Reduce it to expressions in terms of  $\Delta E$ ,  $k_B T$ ,  $h$ ,  $m_p$ , and the atomic number  $A = 56$ . (For convenience, assume  $m_{\text{Fe}} = A m_p$ , ignoring the half-proton-mass energy release. Check the sign of  $\Delta E$ : should it be positive or negative for this reaction?) Evaluate  $K_{\text{eq}}(T)$  in MKS units, as a function of  $T$ . (Note: Do not be alarmed by the large numbers.)

Wikipedia tells us that nucleosynthesis happened “A few minutes into the expansion, when the temperature was about a billion ... Kelvin and the density was about that of air”. To figure out when half of the protons would have fused ( $[\text{Fe}] = [\text{p}]/A$ ), we need to know the temperature and the net baryon density  $[\text{p}] + A[\text{Fe}]$ . We can use eqn 6.69 to give one equation relating the temperature to the baryon density.

(b) Give the symbolic formula for the net baryon density<sup>76</sup> at which half of the protons would have fused, in terms of  $K_{\text{eq}}$  and  $A$ .

What can we use for the other equation relating temperature to baryon density? One of the fundamental constants in the Universe as it evolves is the baryon to photon number ratio. The total number of baryons has been conserved in accelerator experiments of much higher energy than those during nucleosynthesis. The cosmic microwave background (CMB) photons have mostly been traveling in straight lines<sup>77</sup> since the *decoupling time*, a few hundred thousand years after the Big Bang when the electrons and nucleons combined into atoms and became transparent (decoupled from photons). The ratio of baryons to CMB photons  $\eta \sim 5 \times 10^{-10}$  has been constant since that point: two billion photons per nucleon.<sup>78</sup> Between the nucleosynthesis and decoupling eras, photons were created and scattered

<sup>74</sup>It is usually said that  $^{56}\text{Fe}$  is the most stable nucleus, but actually  $^{62}\text{Ni}$  has a higher nuclear binding energy per nucleon. Iron-56 is favored by nucleosynthesis pathways and conditions inside stars, and we shall go with tradition and use it for our calculation.

<sup>75</sup>The entropy cost for our reaction in reality depends upon the ratio of neutrons to protons. Calculations imply that this ratio falls out of equilibrium a bit earlier, and is about 1/7 during this period. By ignoring the difference between neutrons and protons, we avoid considering reactions of the form  $M\text{p} + (56 - M)\text{n} + (M - 26)\text{e} \rightarrow ^{56}\text{Fe}$ .

<sup>76</sup>Here we ask you to assume, unphysically, that all baryons are either free protons or in iron nuclei – ignoring the baryons forming helium and other elements. These other elements will be rare at early (hot) times, and rare again at late (cold) times, but will spread out the energy release from nucleosynthesis over a larger range than our one-reaction estimate would suggest.

<sup>77</sup>More precisely, they have traveled on geodesics in space-time.

<sup>78</sup>It is amusing to note that this ratio is estimated using the models of nucleosynthesis that we are mimicking in this exercise.

by matter, but there were so many more photons than baryons that the number of photons (and hence the baryon to photon ratio  $\eta$ ) was still approximately constant.

We know the density of blackbody photons at a temperature  $T$ ; we integrate eqn (7.66) for the number of photons per unit frequency to get

$$\rho_{\text{photons}}(T) = (2\zeta(3)/\pi^2)(k_B T/\hbar c)^3; \quad (6.70)$$

the current density  $\rho_{\text{photons}}(T_{\text{CMB}})$  of microwave background photons at temperature  $T_{\text{CMB}} = 2.725\text{K}$  is thus a bit over 400 per cubic centimeter, and hence the current density of a fraction of a baryon per cubic meter.

(c) *Numerically matching your formula for the net baryon density at the halfway point in part (b) to  $\eta\rho_{\text{photons}}(T_{\text{reaction}})$ , derive the temperature  $T_{\text{reaction}}$  at which our reaction would have occurred.* (You can do this graphically, if needed.) *Is it roughly a billion degrees Kelvin? Is the nucleon density roughly equal to that of air?* (Hint: Air has a density of about  $1\text{ kg/m}^3$ .)

### 6.23 Entropy cost for controlling engines.<sup>79</sup> (Mathematics, Thermodynamics, Information Geometry) ④

The Carnot cycle (the heat engine that can be run backward as a refrigerator without net inputs) is the origin of our concept of entropy. Expanding a piston at fixed temperature; expanding at fixed energy; contracting at fixed temperature; contacting at fixed energy — is there any entropy cost to guiding the machinery through this complicated path in parameter space?

Machta [29] argues<sup>80</sup> that this cost is given by a *geodesic distance* in parameter space. If Machta is correct, his result pokes holes in two of the key ideas we have learned in this class. (1) The Carnot cycle has a (small) entropy cost. (2) There may be obstacles to running an engine by burning information (Exercise 5.2).

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<sup>81</sup>

$$\begin{aligned} \langle \Delta S_{\text{control}} \rangle &= 2k_B \int_{P_i}^{P_f} \sqrt{g_{PP} dP} \\ &= 2 \int_{P_i}^{P_f} \sqrt{g_{PP}} |dP|. \end{aligned} \quad (6.71)$$

More generally, the entropy cost for leading a system around a path in the space of external conditions is the arc length in the space of probability distributions. The metric tensor in this space, as discussed in Exercise 1.15, is

$$g_{\mu\nu} = - \left\langle \frac{\partial^2 \log(\rho)}{\partial \theta_\mu \partial \theta_\nu} \right\rangle, \quad (6.72)$$

the *Fisher information metric*, giving the natural distance between two nearby probability distributions.

For example, in a Gibbs ensemble<sup>82</sup> at constant pressure  $P = \theta_1$  and inverse temperature  $\beta = 1/k_B T = \theta_2$ , the squared distance between two nearby external conditions is

$$d^2(\rho(\mathbf{X}|\boldsymbol{\theta}), \rho(\mathbf{X}|\boldsymbol{\theta} + d\boldsymbol{\theta})) = g_{\mu\nu} d\theta_\mu d\theta_\nu. \quad (6.73)$$

In our case,  $d^2(\rho(\mathbf{X}|P, T), \rho(\mathbf{X}|P + dP, T)) = g_{PP} dP^2$ , leading directly to eqn 6.71.

<sup>79</sup>This exercise was developed in collaboration with Ben Machta, Archishman Raju, Colin Clement, and Katherine Quinn

<sup>80</sup>Machta [29] studies this by adding a control system to the piston (or a general statistical mechanics ensemble), providing a driving force around the loop. So, to change the pressure on a piston, Machta imagines a continuously variable transmission connecting the piston volume to a mass under gravitational force. To compress the piston, one would introduce an entropy gradient in the transmission favoring gear ratios that allow the mass to fall farther for a given piston volume change.

At fixed gear ratio, the piston fluctuates around an equilibrium volume causing the mass height to fluctuate about some mean. If the gear ratio is also allowed to fluctuate in the absence of an entropy gradient, the mass will gradually drift downward, wasting its free energy. So, if the entropy gradient is too small, free energy is wasted in by the fluctuations; if this entropy gradient is too large, entropy is wasted directly. By balancing the two, he finds a minimum entropy cost for compressing the piston.

<sup>81</sup>We shall follow Machta and set  $k_B = 1$  in this exercise, writing it explicitly only when convenient.

<sup>82</sup>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  $\delta$  function. So we study pistons varying  $P$  and  $\beta = 1/k_B T$ , rather than at fixed volume or energy.

We shall start with computing the metric tensor  $g_{\mu\nu}^{(P,\beta)}$  in the coordinates  $(P, \beta)$  for the ideal gas. We shall analyze the cost for thermodynamic control for Szilard's burning information in Exercise 5.2. Then we will find an elegant thermodynamics approach for  $g_{\alpha\beta}$  for a general Hamiltonian using different coordinates. We then return to the ideal gas piston, to calculate the entropy cost for controlling the Carnot cycle (Section 5.1), both computationally and graphically.

The degrees of freedom for our piston will be  $\mathbf{X} = \{\mathbb{P}, \mathbb{Q}, V\}$ , 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 our piston is the probability density

$$\rho = (1/\Gamma) \exp(-\beta\mathcal{H}(\mathbb{P}, \mathbb{Q}) - \beta PV). \quad (6.74)$$

Here  $\Gamma$  is the partition function for the Gibbs ensemble, normalizing the distribution to one.

(a) 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.) What is the partition function  $\Gamma(P, \beta)$  for the Gibbs ensemble? What is the joint probability density  $\rho_{\text{Gibbs}}(\mathbb{P}, \mathbb{Q}, V | \beta, P)$  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 6.74)? Check that  $\int \rho_{\text{Gibbs}} d\mathbb{Q} dV d\mathbb{P} = 1$ . (Hint: The integral over  $\mathbb{Q}$  depends on  $V$ : do it first.)

(b) For  $\boldsymbol{\theta} = (P, \beta)$ , find the  $2 \times 2$  tensor  $g_{\mu\nu}^{(P,\beta)}$  from eqn 6.72. (Hints: Remember to take the expectation value:  $g_{\mu\nu} = -\int \rho(\mathbf{X}) \partial_\mu \partial_\nu \log(\rho) d\mathbf{X}$ ; again, you will want to do the integral over  $\mathbb{Q}$  first. Your answer should have  $g_{PP} = (1+N)/P^2$ , and  $g_{P\beta} = (1+N)/(\beta P)$ .)

(c) 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 estimate appears to challenge Szilard's argument that information entropy and thermodynamic entropy can be exchanged.

Doing the calculation in the coordinates  $(P, \beta)$  is straightforward, but is not particularly intuitive

and somewhat involved. We can simplify the calculation by changing variables. Let  $p = \beta P = P/k_B T$ ; we change coordinates from  $\boldsymbol{\theta} = (P, \beta)$  to  $\Phi = (p, \beta)$ .

(d) Compute  $g_{\mu\nu}^{(p,\beta)} = -\langle \partial^2 \log(\rho) / \partial \phi_\mu \partial \phi_\nu \rangle$  directly in the new coordinates. Is it less complicated? Using  $dp = \beta dP + P d\beta$ , check that  $g^{(P,\beta)}$  gives the same distance squared as  $g^{(P,\beta)}$  for the same shift in parameters  $(dP, d\beta)$ .

Why are these coordinates special? The log of the Gibbs probability distribution for an arbitrary interacting collection of particles with Hamiltonian  $\mathcal{H}$  (eqn 6.74) is

$$\begin{aligned} \log(\rho) &= -\beta\mathcal{H}(\mathbb{P}, \mathbb{Q}) - \beta PV - \log \Gamma \\ &= -\beta\mathcal{H}(\mathbb{P}, \mathbb{Q}) - pV - \log \Gamma. \end{aligned} \quad (6.75)$$

This is the logarithm of the partition function  $\Gamma$  plus terms *linear* in  $p$  and  $\beta$ .<sup>83</sup> So the second derivatives with respect to  $p$  and  $\beta$  only involve  $\log(\Gamma)$ . We know that the Gibbs free energy  $G(p, \beta) = -k_B T \log(\Gamma) = (1/\beta) \log(1/\Gamma(p, \beta))$ , so  $\log(\Gamma) = \beta G(p, \beta)$ . The first derivatives of the Gibbs free energy  $dG = -SdT + VdP + \mu dN$  are related to things like volume and entropy and chemical potential; our metric is given by the second derivatives (compressibility, specific heat, ...)

(e) For a collection of particles interacting with Hamiltonian  $\mathcal{H}$ , relate the four terms  $g_{\mu\nu}^{(P,\beta)}$  in terms of physical quantities given by the second derivatives of  $G$ . Write your answer in terms of  $N$ ,  $p$ ,  $\beta$ , 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,  $c_P = (T/N)(\partial S / \partial T)|_P$ . (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 more usual coordinates.

(f) Show that

$$g^{(P,\beta)} = N \begin{pmatrix} \beta\kappa/\rho & \alpha/\beta\rho \\ \alpha/\beta\rho & c_P/\beta^2 \end{pmatrix}$$

and

$$g^{(P,T)} = N \begin{pmatrix} \kappa/\rho T & -\alpha/\rho T \\ -\alpha/\rho T & c_P/T^2 \end{pmatrix}.$$

<sup>83</sup>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.

(g) Calculate  $g^{(p,\beta)}$  for the ideal gas using your answer from part (e). Compare with your results calculating  $g^{(p,\beta)}$  directly from the probability distribution in part (d). 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  $\lambda$  depends on temperature.)

The standard formulas for an ideal gas do not include the piston wall as a degree of freedom, so part (g) has one fewer positional degree of freedom than part (d). That is, the macroscopic calculation neglects the entropic contribution of the fluctuations in volume (the position of the piston inside the cylinder).

Now let us apply Machta's entropic control cost estimate to the Carnot cycle, Fig. 5.3. It will be convenient to parameterize the isothermal portions of the cycle by the heat flow  $Q$ , and the adiabatic portions of the cycle with the temperature  $T$ . We label the four steps of the Carnot cycle as in Fig. 5.3.

(h) Find the formulas of the path  $(P_{ab}(Q), \beta_{ab}(Q))$  for the isothermal expansion  $a \rightarrow b$  at the hot temperature  $T_1$  as  $Q$  varies from 0 to  $Q_1$ . Find the path  $(P_{bc}(T), \beta_{bc}(T))$  for the adiabatic expansion  $b \rightarrow c$  as  $T$  cools from  $T_1$  to  $T_2$ . Find the two remaining paths. Plot the  $P$ - $V$  diagram for your Carnot cycle, setting  $T_1 = 1$ ,  $T_2 = 0.8$ ,  $N = 1$ , and  $P_a = 1$ , and  $P_b = 1/2$  (so the initial isothermal expansion is by a factor of two). (Feel free to look up the Carnot cycle paths.)

(i) Calculate the geodesic path length around the four steps, using your answer from part (b) or part (d) so that you include the contribution from the fluctuations in the volume. Express your an-

swer in terms of the number of particles  $N$ , the temperature ratio  $T_1/T_2$ , and the pressure ratio  $P_a/P_b$ . (Be careful to ensure the contribution from each leg is positive: for example, the adiabatic expansion has  $T_2 < T_1$ .) What is  $\langle \Delta S_{\text{control}} \rangle$  for the Carnot cycle?

Machta finds an entropy cost to control the engine of<sup>84</sup>  $4\sqrt{N} \log(P_a/P_b) + 2\sqrt{15N} \log(T_2/T_1)$ . The work done by the Carnot cycle is proportional to  $N$ , so this cost becomes negligible for macroscopic engines.

(j) Is your entropic cost for control also proportional to  $\sqrt{N}$  for large  $N$ ? To  $O(\sqrt{N})$ , do you get Machta's result?

The metric tensor  $g^{(p,\beta)}$  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?

We can answer this by changing coordinates again, as we did in part (d).

(k) Show that one can turn the metric tensor into the identity  $g_{\mu\nu}^{(x,y)} = \delta_{\mu\nu}$  by a coordinate transformation  $(p, \beta) \rightarrow (x = A \log(p), y = B \log(\beta))$ . 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.

(l) What would the path be in the  $(x, y)$  plane for your calculation in part (c) of the control cost for expanding a piston by a factor of two? Is the arc length in  $(x, y)$  equal to your estimate in part (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  $N = 1$ .

<sup>84</sup>This differs slightly from his the published formula.

## Chapter 7: Quantum statistical mechanics

### Exercises

#### 7.4 Does entropy increase in quantum systems? (Mathematics, Quantum) ④

We saw in Exercise 5.7 that in classical Hamiltonian systems the non-equilibrium entropy  $S_{\text{nonequil}} = -k_B \int \rho \log \rho$  is constant in a classical mechanical Hamiltonian system. Here you will show that in the microscopic evolution of an isolated quantum system, the entropy is also time independent, even for general, time-dependent density matrices  $\rho(t)$ .

Using the evolution law (eqn 7.19)  $\partial\rho/\partial t = [\mathcal{H}, \rho]/(i\hbar)$ , prove that  $S = \text{Tr}(\rho \log \rho)$  is time independent, where  $\rho$  is any density matrix. (Two approaches: (1) Go to an orthonormal basis  $\psi_i$  which diagonalizes  $\rho$ . Show that  $\psi_i(t)$  is also orthonormal, and take the trace in that basis. (2) Let  $U(t) = \exp(-i\mathcal{H}t/\hbar)$  be the unitary operator that time evolves the wave function  $\psi(t)$ . Show that  $\rho(t) = U(t)\rho(0)U^\dagger(t)$ . Write  $S(t)$  as a formal power series in  $\rho(t)$ . Show, term-by-term in the series, that  $S(t) = U(t)S(0)U^\dagger(t)$ . Then use the cyclic invariance of the trace.)

#### 7.9 Bosons are gregarious: superfluids and lasers. (Quantum, Optics, Atomic physics) ③

*Adding a particle to a Bose condensate.* Suppose we have a non-interacting system of bosonic atoms in a box with single-particle eigenstates  $\psi_n$ . Suppose the system begins in a Bose-condensed state with all  $N$  bosons in a state  $\psi_0$ , so

$$\Psi_N^{[0]}(\mathbf{r}_1, \dots, \mathbf{r}_N) = \psi_0(\mathbf{r}_1) \cdots \psi_0(\mathbf{r}_N). \quad (7.76)$$

Suppose a new particle is gently injected into the system, into an equal superposition of the  $M$  lowest single-particle states.<sup>85</sup> That is, if it were in-

jected into an empty box, it would start in state

$$\phi(\mathbf{r}_{N+1}) = \frac{1}{\sqrt{M}} (\psi_0(\mathbf{r}_{N+1}) + \psi_1(\mathbf{r}_{N+1}) + \dots + \psi_{M-1}(\mathbf{r}_{N+1})). \quad (7.77)$$

The state  $\Phi(\mathbf{r}_1, \dots, \mathbf{r}_{N+1})$  after the particle is inserted into the non-interacting Bose condensate is given by symmetrizing the product function  $\Psi_N^{[0]}(\mathbf{r}_1, \dots, \mathbf{r}_N)\phi(\mathbf{r}_{N+1})$  (eqn 7.30).

(a) *Symmetrize the initial state of the system after the particle is injected.* (For simplicity, you need not calculate the overall normalization constant  $C$ .) Calculate the ratio of the probability that all particles in the symmetrized state are in the state  $\psi_0$  to the probability that one of the particles is in another particular state  $\psi_m$  for  $0 < m < M$ . (Hint: First do it for  $N = 1$ . The enhancement should be by a factor  $N + 1$ .)

So, if a macroscopic number of bosons are in one single-particle eigenstate, a new particle will be much more likely to add itself to this state than to any of the microscopically populated states.

Notice that nothing in your analysis depended on  $\psi_0$  being the lowest energy state. If we started with a macroscopic number of particles in a single-particle state with wavevector  $\mathbf{k}$  (that is, a superfluid with a supercurrent in direction  $\mathbf{k}$ ), new added particles, or particles scattered by inhomogeneities, will preferentially enter into that state. This is an alternative approach to understanding the persistence of supercurrents, complementary to the topological approach (Exercise 9.7).

*Adding a photon to a laser beam.* This ‘chummy’ behavior between bosons is also the principle behind lasers.<sup>86</sup> A laser has  $N$  photons in a particular mode. An atom in an excited state emits a

<sup>85</sup>For free particles in a cubical box of volume  $V$ , injecting a particle at the origin  $\phi(\mathbf{r}) = \delta(\mathbf{r})$  would be a superposition of *all* plane-wave states of equal weight,  $\delta(\mathbf{r}) = (1/V) \sum_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{r}}$ . (In second-quantized notation,  $a^\dagger(\mathbf{x} = 0) = (1/V) \sum_{\mathbf{k}} a_{\mathbf{k}}^\dagger$ .) We ‘gently’ add a particle at the origin by restricting this sum to low-energy states. This is how quantum tunneling into condensed states (say, in Josephson junctions or scanning tunneling microscopes) is usually modeled.

<sup>86</sup>Laser is an acronym for ‘light amplification by the stimulated emission of radiation’.

photon. The photon it emits will prefer to join the laser beam than to go off into one of its other available modes by a factor  $N + 1$ . Here the  $N$  represents *stimulated emission*, where the existing electromagnetic field pulls out the energy from the excited atom, and the  $+1$  represents *spontaneous emission* which occurs even in the absence of existing photons.

Imagine a single atom in a state with excitation energy energy  $E$  and decay rate  $\Gamma$ , in a cubical box of volume  $V$  with periodic boundary conditions for the photons. By the energy-time uncertainty principle,  $\langle \Delta E \Delta t \rangle \geq \hbar/2$ , the energy of the atom will be uncertain by an amount  $\Delta E \propto \hbar\Gamma$ . Assume for simplicity that, in a cubical box without pre-existing photons, the atom would decay at an equal rate into any mode in the range  $E - \hbar\Gamma/2 < \hbar\omega < E + \hbar\Gamma/2$ .

(b) *Assuming a large box and a small decay rate  $\Gamma$ , find a formula for the number of modes  $M$  per unit volume  $V$  competing for the photon emitted from our atom. Evaluate your formula for a laser with wavelength  $\lambda = 619\text{ nm}$  and the line-width  $\Gamma = 10^4\text{ rad/s}$ . (Hint: Use the density of states, eqn 7.65.)*

Assume the laser is already in operation, so there are  $N$  photons in the volume  $V$  of the lasing material, all in one plane-wave state (a *single-mode* laser).

(c) *Using your result from part (a), give a formula for the number of photons per unit volume  $N/V$  there must be in the lasing mode for the atom to have 50% likelihood of emitting into that mode.*

The main task in setting up a laser is providing a population of excited atoms. Amplification can occur if there is a *population inversion*, where the number of excited atoms is larger than the number of atoms in the lower energy state (definitely a non-equilibrium condition). This is made possible by *pumping* atoms into the excited state by using one or two other single-particle eigenstates.

### 7.17 Eigenstate thermalization. (Quantum) Ⓜ

Footnote 10 on page 175 discusses many-body eigenstates as ‘weird delicate superpositions of states with photons being absorbed by the atom and the atom emitting photons’. Many-body eigenstates with a *finite energy density* can be far more comprehensible – they often correspond to *equilibrium* quantum systems.

*Look up ‘eigenstate thermalization’. Find a discussion that discusses a single pure state of a large system  $A+B$ , tracing out the ‘bath’  $B$  and leaving*

*a density matrix for a small subsystem  $A$ . Is the discussion near the footnote still correct? What should be added or changed?*

### 7.18 Is sound a quasiparticle? (Condensed matter) Ⓜ

Sound waves in the harmonic approximation are non-interacting – a general solution is given by a linear combination of the individual frequency modes. Landau’s Fermi liquid theory (footnote 23, page 180) describes how the non-interacting electron approximation can be effective even though electrons are strongly coupled to one another. The quasiparticles are electrons with a screening cloud; they develop long lifetimes near the Fermi energy; they are described as poles of Greens functions.

(a) *Do phonons have lifetimes? Do their lifetimes get long as the frequency goes to zero? (Look up ‘ultrasonic attenuation’ and Goldstone’s theorem.)*

(b) *Are they described as poles of a Green’s function? (See Section 9.3 and Exercise 10.9.)*

Are there analogies for phonons to the screening cloud around a quasiparticle? A phonon screening cloud would be some kind of collective, non-linear movement of atoms that behaved at long distances and low frequencies like an effective, harmonic interaction. In particular, the effective scattering between these quasiphonons should be dramatically reduced from that one would calculate assuming that the phonons are harmonic vibrations.

At low temperatures in perfect crystals (other than solid helium), anharmonic interactions are small except at high sound amplitudes. But as we raise the temperature, anharmonic effects lead to thermal expansion and changes in elastic constants. We routinely model sound in crystals in terms of the density and elastic constants at the current temperature, not in terms of the ‘bare’ phonon normal modes of the zero-temperature crystal.

### 7.19 Drawing wavefunctions. (Quantum) ⓘ

(a) *Draw a typical first excited state  $\Psi(x, y)$  for two particles at positions  $x$  and  $y$  in a one-dimensional potential well, given that they are (i) distinguishable, (ii) bosons, and (iii) fermions. For the fermion wavefunction, assume the two spins are aligned. Do the plots in the 2D region  $-L/2 < x, y, < L/2$ , emphasizing the two regions where the wavefunction is positive and negative, and the nodal curve where  $\Psi(x, y) = 0$ .*

(b) Which nodal lines in your plots for part (a) are fixed? Which can vary depending on the manybody wavefunction? In particular, for the bose excited state, must the nodal curve be straight? Which nodal lines in the fermion excited state must remain in position independent of the Hamiltonian?

**7.20 Universe of light baryons.** (Quantum) ②

Presume a parallel universe where neutrons and protons were as light as electrons (2000 times lighter than they are now).

Estimate the superfluid transition temperature  $T_c$  for water in the parallel universe. Will it be superfluid at room temperature? (Presume 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 stays about the same size in the parallel world.)

**7.21 Many-fermion wavefunction nodes.** (Quantum) ①

Consider an  $N$ -electron atom or molecule in three dimensions, treating the nuclei as fixed (and hence the Coulomb attraction to the electrons as a fixed external potential with electron ground-state wavefunction  $\Psi(\mathbf{R})$ ,  $\mathbf{R} = \{\mathbf{r}_1, \dots, \mathbf{r}_N\} \in \mathbb{R}^{3N}$ ). We shall explore the structure of the zeros of this  $3N$ -dimensional wave function.

In high-dimensional spaces, one often measures the *co-dimension* of hypersurfaces by counting down from the dimension of the embedding space. Thus the unit hypersphere is co-dimension one, whatever the dimension of space is.

Assume that there are no magnetic fields, and that all spins are aligned, so that  $\Psi(\mathbf{r})$  can be assumed to be both real and totally antisymmetric under interchange of particles.

(a) What is the co-dimension of the sets where  $\Psi(\mathbf{R}) = 0$ ?

(b) What zeros are ‘pinned in place’ by the Fermi statistics of the electrons? What is the co-dimension of this lattice of pinned nodes?

I view the nodal surfaces as being like a soap bubble, pinned on the fixed nodes and then smoothly interpolating to minimize the energy of the wavefunction. Diffusion quantum Monte Carlo, in fact, makes use of this. One starts with a carefully constructed variational wavefunction  $\Psi_0(\mathbf{r})$ ,

optimizes the wavefunction in a ‘fixed-node’ approximation (finding the energy in a path integral restricted to one local connected region where  $\Psi_0(\mathbf{r}) > 0$ ), and then at the end relaxes the fixed-node approximation.

(c) Consider a connected region in configuration space surrounded by nodal surfaces where  $\Psi(\mathbf{r}) > 0$ . Do you expect that  $\Psi$  in this region will define, through antisymmetry,  $\Psi$  everywhere in space? (Hint: For bosons and distinguishable particles, the ground state wavefunction has no nodes.)

**7.22 The greenhouse effect and cooling coffee.** (Astrophysics, Ecology) ②

Vacuum is an excellent insulator. This is why the surface of the Sun can remain hot ( $T_S = 6000^\circ \text{K}$ ) even though it faces directly onto outer space at the microwave background radiation temperature  $T_{MB} = 2.725 \text{ K}$ , (Exercise 7.15). The main way<sup>87</sup> in which heat energy can pass through vacuum is by thermal electromagnetic radiation (photons). We will see in Exercise 7.7 that a black body radiates an energy  $\sigma T^4$  per square meter per second, where  $\sigma = 5.67 \times 10^{-8} \text{ J}/(\text{s m}^2 \text{ K}^4)$ .

A vacuum flask or Thermos bottle<sup>TM</sup> keeps coffee warm by containing the coffee in a *Dewar*—a double-walled glass bottle with vacuum between the two walls.

(a) Coffee at an initial temperature  $T_H(0) = 100^\circ \text{C}$  of volume  $V = 150 \text{ mL}$  is stored in a vacuum flask with surface area  $A = 0.1 \text{ m}^2$  in a room of temperature  $T_C = 20^\circ \text{C}$ . Write down symbolically the differential equation determining how the difference between the coffee temperature and the room temperature  $\Delta(t) = T_H(t) - T_C$  decreases with time, assuming the vacuum surfaces of the dewar are black and remain at the current temperatures of the coffee and room. Solve this equation symbolically in the approximation that  $\Delta$  is small compared to  $T_C$  (by approximating  $T_H^4 = (T_C + \Delta)^4 \approx T_C^4 + 4\Delta T_C^3$ ). What is the exponential decay time (the time it takes for the coffee to cool by a factor of e), both symbolically and numerically in seconds? (Useful conversion:  $0^\circ \text{C} = 273.15^\circ \text{K}$ .)

Real Dewars are not painted black! They are coated with shiny metals in order to minimize this radiative heat loss. (White or shiny materials not only absorb less radiation, but they also emit less radiation, see exercise 7.7.)

<sup>87</sup>The sun and stars can also radiate energy by emitting neutrinos. This is particularly important during a supernova.

The outward solar energy flux at the Earth's orbit is  $\Phi_S = 1370 \text{ W/m}^2$ , and the Earth's radius is approximately 6400 km,  $r_E = 6.4 \times 10^6 \text{ m}$ . The Earth reflects about 30% of the radiation from the Sun directly back into space (its *albedo*  $\alpha \approx 0.3$ ). The remainder of the energy is eventually turned into heat, and radiated into space again. Like the Sun and the Universe, the Earth is fairly well described as a black-body radiation source in the infrared. We will see in Exercise 7.7 that a black body radiates an energy  $\sigma T^4$  per square meter per second, where  $\sigma = 5.67 \times 10^{-8} \text{ J/(s m}^2 \text{ K}^4)$ .

(b) *What temperature  $T_A$  does the Earth radiate at, in order to balance the energy flow from the Sun after direct reflection is accounted for? Is that hotter or colder than you would estimate from the temperatures you've experienced on the Earth's surface?* (Warning: The energy flow in is proportional to the Earth's cross-sectional area, while the energy flow out is proportional to its surface area.)

The reason the Earth is warmer than would be expected from a simple radiative energy balance is the *greenhouse effect*.<sup>88</sup> The Earth's atmosphere is opaque in most of the infrared region in which the Earth's surface radiates heat. (This frequency range coincides with the vibration frequencies of molecules in the Earth's upper atmo-

sphere. Light is absorbed to create vibrations, collisions can exchange vibrational and translational (heat) energy, and the vibrations can later again emit light.) Thus it is the Earth's atmosphere which radiates at the temperature  $T_A$  you calculated in part (b); the upper atmosphere has a temperature intermediate between that of the Earth's surface and interstellar space.

The vibrations of oxygen and nitrogen, the main components of the atmosphere, are too symmetric to absorb energy (the transitions have no dipole moment), so the main greenhouse gases are water, carbon dioxide, methane, nitrous oxide, and chlorofluorocarbons (CFCs). The last four have significantly increased due to human activities; CO<sub>2</sub> by  $\sim 30\%$  (due to burning of fossil fuels and clearing of vegetation), CH<sub>4</sub> by  $\sim 150\%$  (due to cattle, sheep, rice farming, escape of natural gas, and decomposing garbage), N<sub>2</sub>O by  $\sim 15\%$  (from burning vegetation, industrial emission, and nitrogen fertilizers), and CFCs from an initial value near zero (from former aerosol sprays, now banned to spare the ozone layer). Were it not for the Greenhouse effect, we'd all freeze (like Mars)—but we could overdo it, and become like Venus (whose deep and CO<sub>2</sub>-rich atmosphere leads to a surface temperature hot enough to melt lead).

<sup>88</sup>The glass in greenhouses also is transparent in the visible and opaque in the infrared. This, it turns out, isn't why it gets warm inside; the main insulating effect of the glass is to forbid the warm air from escaping. The greenhouse effect is in that sense poorly named.

## Chapter 8: Calculation and computation

### Exercises

#### 8.5 Detailed balance. *i*

In an equilibrium system, for any two states  $\alpha$  and  $\beta$  with equilibrium probabilities  $\rho_\alpha^*$  and  $\rho_\beta^*$ , detailed balance states (eqn 8.14) that

$$P_{\beta \leftarrow \alpha} \rho_\alpha^* = P_{\alpha \leftarrow \beta} \rho_\beta^*, \quad (8.78)$$

that is, the equilibrium flux of probability from  $\alpha$  to  $\beta$  is the same as the flux backward from  $\beta$  to  $\alpha$ . It is both possible and elegant to reformulate the condition for detailed balance so that it does not involve the equilibrium probabilities. Consider three states of the system,  $\alpha$ ,  $\beta$ , and  $\gamma$ .

(a) Assume that each of the three types of transitions among the three states satisfies detailed balance. Eliminate the equilibrium probability densities to derive<sup>89</sup>

$$P_{\alpha \leftarrow \beta} P_{\beta \leftarrow \gamma} P_{\gamma \leftarrow \alpha} = P_{\alpha \leftarrow \gamma} P_{\gamma \leftarrow \beta} P_{\beta \leftarrow \alpha}. \quad (8.79)$$

Viewing the three states  $\alpha$ ,  $\beta$ , and  $\gamma$  as forming a circle, you have derived a relationship between the rates going clockwise and the rates going counter-clockwise around the circle.

Can we show the converse to part (a), that if every triple of states in a Markov chain satisfies the condition 8.79 then it satisfies detailed balance? Given the transition matrix  $P$ , can we construct a probability density  $\rho^*$  which makes the probability fluxes between all pairs of states equal?

In most problems, most of the rates are zero: the matrix  $P$  is sparse, connecting only nearby states. This makes the cyclic condition for detailed balance stricter than than Equation 8.79. (Detailed balance demands that the forward and reversed products for cycles of all lengths must be equal, the *Kolmogorov criterion*.) There are  $N^3$  equations in Equation 8.79 ( $\alpha$ ,  $\beta$ , and  $\gamma$  each running over all  $N$  states), and we need to solve for  $N$  unknowns  $\rho^*$ . But in a sparse matrix most of these ‘triangle’ conditions tell us only that  $0 = 0$ . However, if we assume that the transition matrix has enough positive entries, the three-state

cyclic eqn 8.79 is enough to construct the stationary state  $\rho^*$  from  $P$ , satisfying detailed balance (eqn 8.78).

(c) Suppose  $P$  is the transition matrix for some Markov chain satisfying the condition 8.79 for every triple of states  $\alpha$ ,  $\beta$ , and  $\gamma$ . Assume that there is a state  $\alpha_0$  with non-zero transition rates from all other states  $\delta$ . Construct a probability density  $\rho^*$  that demonstrates that  $P$  satisfies detailed balance (eqn 8.78). (Hint: Assume you know  $\rho_{\alpha_0}^*$ ; use some of the eqns 8.78 to write a formula for each of the other  $N - 1$  elements  $\rho_\delta^*$  that ensures detailed balance for the pair. Then solve for  $\rho_{\alpha_0}^*$  to make the probability distribution normalized. Show that this candidate stationary state  $\rho^*$  satisfies detailed balance for any two states  $\beta$  and  $\delta$ , using the cyclic condition eqn 8.79.)

#### 8.17 Solving the Ising model with parallel updates. *(Computation) p*

Our description of the heat-bath and Metropolis algorithm equilibrates one spin at a time. This is inefficient even in Fortran and C++ (memory nonlocality); in interpreted languages like Python or Mathematica conditional loops like this are really slow.

Could we update all the spins at once, thermalizing them according to their current neighborhood? If not, can you figure out a way of bringing many of the spins to equilibrium with their neighborhoods in one set of array operations (without ‘looping’ over spins)? No implementation is needed – just a description of the method. (Hint: Consider the Ising hard disks in Exercise 8.16, at low temperatures where overlaps are rare. In a long region of down spins, if a parallel update flips a spin up with probability  $p$ , how often will it flip two adjacent spins?)

<sup>89</sup>Note that, as in eqn 8.78, we do *not* sum over repeated indices here; this equation must hold for all triples of states  $\alpha$ ,  $\beta$ , and  $\gamma$ . Hence these form  $N^3$  equations for the  $N^2$  variables  $P_{\alpha \leftarrow \beta}$ .

**8.18 Low temperature expansion of the Ising model.** 

Consider the low-temperature cluster expansion of eqn 8.19. Which zero-temperature ground state (spin-up or spin-down) is it perturbing about? What cluster gives the first term? Explain the power  $x^6$  and the coefficient  $-2$ . What cluster(s) contribute to the second term, proportional to  $x^{10}$ ?

**8.19 Cluster expansions: the 2D Ising model.**  ③

In this exercise, we shall derive the first two terms of the low-temperature expansion for the free energy and the magnetization, corresponding to eqn 8.19 but for the two-dimensional Ising model on a square lattice (rather than the 3D cubic lattice). The lattice is finite, having periodic boundary conditions and dimensions  $L$  columns by  $L$  rows, so that there are  $N = L \times L$  sites.

(a) Let  $E_g$  be the energy per spin of the ground state at zero temperature, and  $\delta$  be the extra energy needed to flip one spin away from the ground state. What is the partition function  $Z$  at temperatures low enough that only these two types of states need to be considered? (Hint: How many states share the energy  $NE_g + \delta$ ?) For later use, what is the ground state energy  $E_g$  per spin and the spin-flip energy  $\delta$ , as a function of the nearest-neighbor bond strength  $J$  and the external field  $H$ ?

At low temperatures, it will be a rare event that a particular spin will flip. But for  $N \gg \exp(-\delta/k_B T)$ , a typical state will have many spins flipped – a *dilute gas* of flipped spins.<sup>90</sup> Can we find the partition function in that limit? Flipping two spins costs  $2\delta$  in energy, but happens more often.

(b) Argue that the number of configurations with two isolated spins flipped is  $N(N-5)/2 = N^2/2 + O(N)$ . (The two spins flipped cannot be the same spin, and if two neighboring spins flip the energy is not  $2\delta$ . Both corrections can be incorporated into higher-order terms in the expansion, analogous to the term  $14x^{12}$  in eqn 8.19.) Argue that the number of configurations with  $m$  isolated spins is  $N^m/m! + O(N^{m-1})$ . Knowing the energy to flip  $m$  isolated spins is  $m\delta$ , write the low-temperature partition function as

a sum over  $m$ , in the dilute gas approximation where all spins flipping are isolated. Show that  $Z = \exp(-NE_g/k_B T + N \exp(-\delta/k_B T))$ . What is the free energy  $f$  per spin?

This fact, that summing over an infinite number of clusters for  $Z$  is equal to a single cluster calculation for  $\log Z$ , is an example of the *linked cluster theorem*, used also to simplify Feynman diagram calculations.

(c) Using your answer for  $E_g$  and  $\delta$ , find the magnetization per spin  $m(T)$  at zero field by evaluating the appropriate derivative of the free energy per spin  $f(T, H)$ . Compare to the first two terms of eqn 8.19.

**8.20 Fruit flies, Markov matrices, and free energies.**  ① (Biology) ③

Fruit flies exhibit several stereotyped behaviors. A Markov model for the transitions between these states has been used to describe the transitions between behaviors [5, 6]. Let us simplify the fruit fly behavior into three states: idle ( $\alpha = 1$ ), grooming ( $\alpha = 2$ ), and locomotion ( $\alpha = 3$ ). Assume that measurements are done at regular time intervals (say one minute), and let  $P_{\beta\alpha}$  be the probability that a fly in state  $\alpha$  in time interval  $n$  goes into state  $\beta$  in the next interval  $n+1$ , so the probability distribution  $\rho_\beta(n+1) = \sum_\alpha P_{\beta\alpha}\rho_\alpha(n)$ .

<sup>90</sup>This dilute gas approximation is quite common: for example, it is used in nucleation theory (Section 11.3) to turn the free energy cost of forming a raindrop into a rate per unit volume of raindrop formation, and in quantum mechanics it turns the instanton action into a tunneling rate per unit time.

<sup>91</sup>This problem was created with the assistance of Gordon Berman.

***Idle/Slow***

**Fig. 8.28 Fruit fly behavior map**, from “Hierarchy and predictability in *Drosophila* behavior” [4]. Fruit fly stereotyped behaviors were analyzed using an unsupervised machine-learning platform and projected into a two dimensional space to preserve local similarity. The resulting behavior clusters were labeled by hand. Black lines represent transitions between states, with right-handed curvatures indicating direction of transmission, and line thickness proportional to the transition rates. Anterior, abdomen, and wing motions are basically all grooming. We lump together slow motions (consisting of small leg twitches and proboscis extension) with the idle state. The transition matrix here does not faithfully represent the actual behavior. Courtesy of G.J. Berman.

Assume that the transition matrix is

$$P = \begin{pmatrix} 29/32 & 1/8 & 1/32 \\ 1/12 & 55/64 & 1/32 \\ 1/96 & 1/64 & 15/16 \end{pmatrix}. \quad (8.80)$$

- (a) Show that probability is conserved under the transition matrix  $P$ , both directly and by checking the appropriate left eigenvector and eigenvalue.
- (b) Is the dynamics given by the transition matrix  $P$  Markov? Ergodic? Must the dynamics of a fruit fly be Markov? Why or why not? Is real fruit fly dynamics ergodic? If so, why? If not, give a state of the fly (not included in our model) that can never return to the locomotion state. (Hint: Flies are not immortal.)
- (c) Find the stationary state giving the probability distribution  $\rho_\alpha^* = \rho_\alpha(\infty)$  of fruit fly behavior at long times. Is it a left or a right eigenvector? What is its eigenvalue?
- (d) Does the transition matrix  $P$  satisfy detailed balance? First check this directly by demanding that no net probability flows around cycles

$$P_{\alpha \leftarrow \beta} P_{\beta \leftarrow \gamma} P_{\gamma \leftarrow \alpha} = P_{\alpha \leftarrow \gamma} P_{\gamma \leftarrow \beta} P_{\beta \leftarrow \alpha}. \quad (8.81)$$

(eqn 8.79 in Exercise 8.5). Then check it using your stationary solution in part (c) by demanding that the stationary state has equal probability flows forward and backward

$$P_{\beta \leftarrow \alpha} \rho_\alpha^* = P_{\alpha \leftarrow \beta} \rho_\beta^*, \quad (8.82)$$

- (eqn 8.14). (Hint: the second method should serve as a check for your answer to part (c).)
- (e) Must fruit flies obey detailed balance in their transition rates? If so, why? If not, discuss how the fruit fly might plausibly have a net probability flow around a cycle from idle to grooming to locomotion.

In the morning, when the lights are turned on, 60% of the fruit flies are idle ( $\alpha = 1$ ), 10% are grooming ( $\alpha = 2$ ), and 30% are undergoing locomotion ( $\alpha = 3$ ).

(f) What is the net entropy per fly of this initial probability distribution? What is the entropy per fly of the stationary state  $\rho^*$ ? Does fly entropy increase or decrease from morning to mid-day? Plot the entropy per fly for the first hour, by iterating our Markov transition matrix directly on the computer. Is the entropy change monotonic? Fruit flies are not isolated systems: they can exchange energy with their environment. We would like to check if the fruit fly evolves to minimize its free energy  $F = E - TS$ .

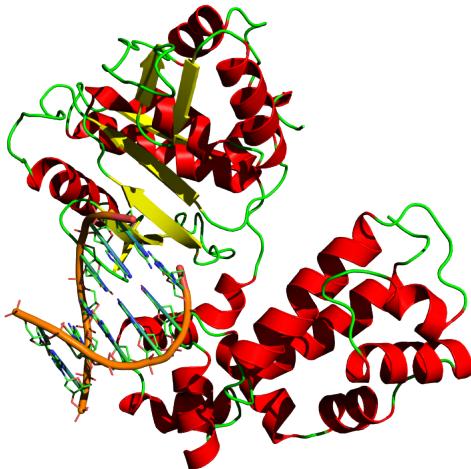
(g) What quantities are exchanged between the fruit fly and its environment? Is it feasible to describe the real fruit fly with a free energy, integrating out the rest of the world as a heat bath? Must fruit flies minimize their free energy as time progresses? Why or why not?

How can we estimate the energy for each fly state?

(h) Assume that the stationary state of the flies is a Boltzmann distribution  $\rho_\alpha^* \propto \exp(-E_\alpha/k_B T)$ . Calculate  $E_\alpha/k_B T$  for the three fly states. (Note: Only the differences are determined; choose the zero of energy to make the partition function equal to one.) What is the average fly energy first thing in the morning? In the stationary state? Does the fly energy increase or decrease from morning to mid-day? Plot the energy for the first hour; is it monotonic?

(i) Let the free energy be  $F = E - TS$  for the flies in our simple model. Plot  $F/k_B T$  for the first hour. Does it monotonically decrease?

**8.21 Kinetic proofreading in cells.<sup>92</sup> (Biology) ③**



**Fig. 8.29 DNA polymerase**, guiding replication of DNA. First, helicases break the hydrogen bonds between base pairs and unwind the DNA strand. Then DNA polymerase catalyzes the reaction which adds complementary base pairs to the nucleotide sequence. The process of matching these base pairs is done with extremely high accuracy. The mechanism with which this is achieved is called proofreading. Kinetic proofreading [23] is one explanation for the high accuracy of replication. (Image from [https://en.wikipedia.org/wiki/DNA\\_polymerase](https://en.wikipedia.org/wiki/DNA_polymerase) reproduced here under the CC BY-SA 3.0 license).

Many biological processes in our body require that we distinguish between two different pathways with very high accuracy. During DNA replication, for example, DNA polymerase (Fig. 8.29), an enzyme that synthesizes the copy of the existing DNA sequence, can sometimes make mistakes. Our body needs an error-correcting mechanism to make sure that incorrect base pairs are not added to the copied DNA sequence and replication happens with high fidelity.

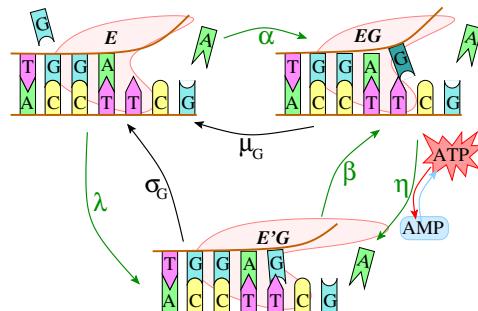
Typically, there is an energy difference between a correct base pair and an incorrect base pair being added to the copy of the DNA strand. However, these energy differences are usually not enough to explain the extremely high accuracy with which DNA replication takes place. Thus, the copying

<sup>92</sup>This exercise was developed by Archishman Raju in collaboration with Colin Clement and the author.

<sup>93</sup> We shall use a discrete time step  $\Delta t$  to model these reaction networks as Markov chains. Taking the time step to zero leads to a continuous time Markov process —

is followed by some activity which adds a further correction. Here, we will explore one particular mechanism for this proposed by Hopfield called *kinetic proofreading* [7, 8, 23].

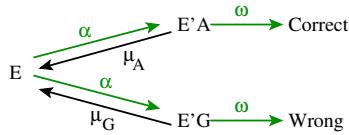
DNA replication is shown in Fig. 8.30. You imagine the lower end of the DNA (the parent strand) extends up and to the right to a Y-junction, at which another protein (a helicase) holds it apart from its original spousal complementary strand. The upper strand is being assembled one nucleotide at a time, with *A* matched to *T* and *G* matched to *C* for the four possible DNA nucleotides Adenine, Thymine, Guanine, and Cytosine. This matching is done with high fidelity, with error ratios (incorrect/correct) in yeast and humans quoted between  $f = 10^{-7}$  to less than  $10^{-9}$ . The challenge for the enzyme is to attach *A* to a parent *T* with that accuracy, and then also to attach *G* to *C* with the same accuracy (Fig. 8.30).



**Fig. 8.30 DNA proofreading** during replication, here showing a potential copying error, a *G* rather than *A* attached to a *T*. The proofreading involves two states of the enzyme/DNA/nucleotide complex,  $E'G$  and  $EG$ , each of which can abandon the error and return to the unbound enzyme/DNA complex  $E$ .

Suppose the enzyme-DNA complex  $E$  is bound to a parent strand with the next parent nucleotide  $T$  desiring a partner  $A$ , and wanting to avoid pairing with  $G$ . Clearly the complex must associate the potential partner  $X \in \{A, G\}$  with  $T$  to find out which matches better. Let this associated bound state be denoted  $E'X$ , and let the free energies of the three be  $\mathcal{E}_E$ ,  $\mathcal{E}_{E'A}$ , and  $\mathcal{E}_{E'G}$ . We may assume that the rate of attachment  $E \rightarrow E'A$  and

$E \rightarrow E'G$  are both the same: in a time-step<sup>93</sup>  $\Delta t$ , an unbound complex will bind either  $A$  or  $G$  with probability  $\alpha$ . Because  $E'A$  is more tightly bound than  $E'G$ , the fraction of unbinding  $\mu_A$  is smaller than  $\mu_G$  (Fig. 8.31).



**Fig. 8.31 Naive DNA replication** reaction diagram. Dark arrows are reactions whose rates are assumed to depend on the nucleotide. Notation from [28].

Throughout this exercise, we shall take limits to simplify our analysis. In Fig. 8.31 as drawn, we shall assume the final reaction of rate  $\omega$  completing the nucleotide insertion is small, so we can assume the populations of  $E$ ,  $E'A$ , and  $E'G$  come to steady state before the complex moves on to other tasks. This makes the error ratio equal to the steady state  $f = \rho_{E'G}^*/\rho_{E'A}^*$ , where  $\rho^*$  is the steady state concentrations of the various components. Similarly,  $A$  and  $G$  compete in Fig. 8.31 for the enzyme complex  $E$  — a complex bound to  $G$  is not available to bind to  $A$ , and will reduce the overall concentration of  $E'A$ . If we assume  $\alpha \ll \mu_x$  for both reactions, we may assume  $E$  is rarely bound to either, and assume  $\rho_E^*$  stays close to one. This allows us to analyze the upper and lower branches of Fig. 8.31 separately.<sup>94</sup>

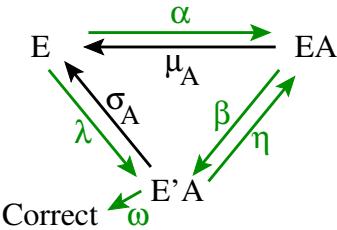
(a) Calculate  $\rho_{E'A}^*$  and  $\rho_{E'G}^*$  in terms of  $\mathcal{E}_E$ ,  $\mathcal{E}_{E'A}$ , and  $\mathcal{E}_{E'G}$ , assuming the system comes to equilibrium, given the approximations above. What must be true for most of  $E$  to remain unbound? Write a formula  $f_{\rho^*}$  for the failure rate in terms of the free energies. Calculate separate formulae for the steady state  $\rho_{E'A}^*$  and  $\rho_{E'G}^*$  in terms of  $\alpha$ ,  $\mu_A$ , and  $\mu_G$ , again using the approximations above. Write the failure rate  $f_{\text{rate}}$  also in terms of the reaction rates. Use the two-state detailed balance condition, eqn 8.14, to show that your reaction-rate failure ratio is correct even if  $\rho_E$  is not approximately one. (The answers should not be complicated.)

As a rough estimate, the difference in the binding

energy between a correct and incorrect nucleotide in DNA is of the order of 5 kcal/mol. At body temperature,  $k_B T \approx 0.616$  kcal/mol.

(b) What is  $f_{\text{naive}}$  for our naive model at body temperature, assuming  $\mathcal{E}_{E'G} - \mathcal{E}_{E'A} = 5$  kcal/mol? How much would the enzyme need to magnify this free energy difference to make the naive error ratio equal to the experimental value  $10^{-9}$ ?

Now let us consider the more elaborate reaction diagram, shown in Fig. 8.32. The basic idea is to add an intermediate state  $EA$  between the free enzyme  $E$  and the final bound state  $E'A$  before nucleotide incorporation. Can this give us a lower error rate, for a given free energy difference  $\mathcal{E}_{E'G} - \mathcal{E}_{E'A}$ ?



**Fig. 8.32 Equilibrium DNA replication** reaction diagram. Dark arrows are reactions whose rates are assumed to depend on the binding nucleotide — those that change the association of  $A$  or  $G$  with the parent nucleotide  $T$ . Adding the reaction rates for the alternative pathway (incorrect addition of  $G$ ) would double the diagram, as in Fig. 8.31.

(c) Consider the reaction network shown in Fig. 8.32, including an intermediate state  $EA$ . (The reaction network for incorporation of the wrong nucleotide  $G$  is similar, with only the labeled rates  $\mu$  and  $\sigma$  changed.) Argue, if the system is in equilibrium, that the failure ratio remains  $f_{\rho^*}$  in terms of the free energies, as in part (a). Argue, if the reaction rates all satisfy the two-state detailed balance condition (eqn 8.14) with  $\rho_x^* = \exp(-\beta \mathcal{E}_x)$ , then the error ratio is still  $f_{\text{rate}}$  from part (a). Use the two-state detailed balance condition between the (slow) reaction rate  $\lambda$  and the two forward reactions  $\sigma_A$  and  $\sigma_G$  to write the error ratio in terms of  $\sigma_A$  and  $\sigma_G$ . Finally, use the cyclic detailed balance condition (eqn 8.79 in

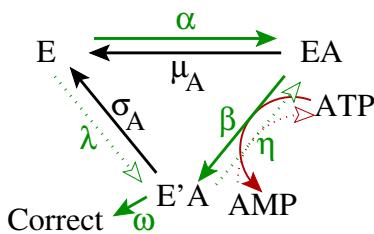
more realistic, no more complicated, but not covered in this text.

<sup>94</sup>The failure ratio, eqn 8.85, is unchanged by this decoupling approximation, since the competition reduces all steady-state probabilities by the same factor.

Exercise 8.5)

$$P_{a \leftarrow c} P_{c \leftarrow b} P_{b \leftarrow a} = P_{a \leftarrow b} P_{b \leftarrow c} P_{c \leftarrow a} \quad (8.83)$$

to show that  $\sigma_A/\sigma_G = \mu_A/\mu_G$ . (Hint: In equilibrium, the probability of being in a state is independent of the dynamics — given only by the energy of the state. This is a profound truth in statistical mechanics.)

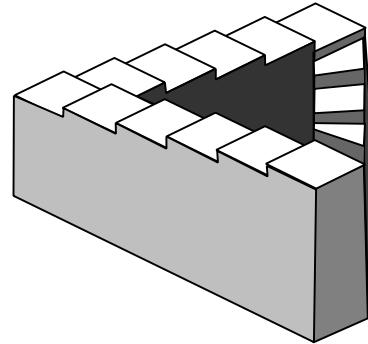


**Fig. 8.33 Kinetic proofreading** reaction diagram. Here the reactions connecting  $EA$  to  $E'A$  are driven by the fuel source of the cell, converting ATP (adenosine triphosphate) to AMP (adenosine monophosphate) by removing (hydrolysing) two phosphate groups in the form of  $PP_i$ . Dark arrows are reactions whose rates are assumed to depend on the nucleotide.

The key is to make use of a fuel source that drives the reaction out of equilibrium (Fig. 8.33). For example, the reaction  $EA \rightarrow E'A$  now becomes  $EA + ATP \rightarrow E'A + AMP + PP_i$ . Converting ATP to AMP releases a large amount of free energy, both because ATP as a single molecule has a larger free energy, and because the concentration of ATP in a cell is kept much higher than AMP and ADP by the action of mitochondria (which use glucose to convert AMP and ADP to ATP.) In particular, the free energy released in the reaction  $EA + ATP \rightarrow E'A + AMP$  is

$$\delta\Xi = \delta G_0 + kT \log \frac{[ATP][EA]}{[AMP][PP_i][E'A]}. \quad (8.84)$$

Detailed balance between the new reactions with rates  $\beta$  and  $\eta$  now tell us that  $\beta/\eta = \exp(-(\mathcal{E}_{E'A} - \delta\Xi)/k_B T)$ .



**Fig. 8.34 Impossible stair.** The addition of the free energy of ATP hydrolysis to one of the reactions is like providing an elevator connecting the ground floor to one two floors up. The effective staircase clockwise goes downhill forever; the system no longer satisfies detailed balance. (Thanks to Escher and the Penrose stair.)

The typical energy shifts  $\delta\Xi$  associated with the removal of phosphate groups is large ( $20-30 k_B T$  even for a single phosphate). This inspires another approximation: we may assume  $(\mathcal{E}_{E'A} - \mathcal{E}_E) \gg k_B T$  (so the back reaction rate  $\lambda$  is small), and yet we may assume  $(\mathcal{E}_{EA} - \mathcal{E}_{E'A} - \delta\Xi) \gg k_B T$  (so the back reaction rate  $\eta$  is small).

(d) Does the cyclic condition for detailed balance, eqn 8.83, relating the product of clockwise and counter-clockwise reaction rates, hold for our kinetic proofreading reaction diagram? Why not? (See Fig. 8.34.) What changes in the environment when one goes clockwise around the reaction, back to the ‘same’ initial state?

We shall make another approximation — that the conversion rate  $\beta$  from  $EA$  to  $E'A$  is slow enough that it does not significantly change the concentration of  $EA$ . Just as making  $\alpha$  slow compared to  $\mu$  allowed the correct and incorrect nucleotide reactions to decouple, this approximation allows the reactions  $E \leftrightarrow EA$  to decouple from the reactions  $EA \rightarrow E'A \rightarrow E$ .

(e) Use your answers from part (a) to solve for the steady-state behavior of the kinetic proofreading reactions  $E \leftrightarrow EA$  in Fig. 8.33, in these approximations. Solve for the steady state behavior of  $\rho_{E'A}^*/\rho_{EA}^*$  and use your answers from part (c) to simplify it. What is the error ratio of the kinetic proofreading reaction diagram?

Kinetic proofreading can square the error fraction, at the expense of using many ATPs (one for each error-correcting cycle around the triangle before

the nucleotide is accepted at rate  $\omega$ ).

Finally, let us use the machinery of Markov chains to solve the problem in greater generality. We shall still ignore the competition between the two nucleotides for  $E$  and assume that the fixation rate  $\omega$  is small.

(f) *Explicitly write the  $3 \times 3$  Markov transition matrix for Figs 8.32 and 8.33, evolving the systems by a small time step  $\Delta t$ . (See footnote 93 on page 67.)*

By explicitly solving for the steady-state probabilities for this matrix, the error ratio can be calculated and is given by Hopfield's formula [23,

Equation 5]

$$f = \frac{(\alpha\beta + \beta\lambda + \lambda\mu_G)(\eta\mu_A + \beta\sigma_A + \mu_A\sigma_A)}{(\alpha\beta + \beta\lambda + \lambda\mu_A)(\eta\mu_G + \beta\sigma_G + \mu_G\sigma_G)} \quad (8.85)$$

(g) *Look back to part (c) and use the cyclic condition of detailed balance (eqn 8.83) to show that the error rate  $f$  is equal to the reaction-rate form of  $f_{\text{naive}}$  if detailed balance is satisfied. (Hint: Write an expression for  $\sigma_G$  and  $\sigma_A$  in terms of the other coefficients and substitute them in  $f$ . The calculation is a bit complicated.)*

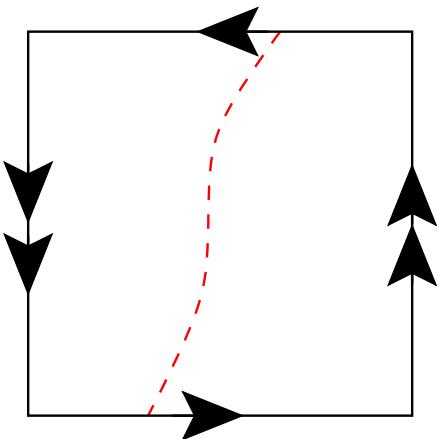
(h) *What is the error ratio when we set  $\lambda = \eta = 0$ ? Does it approach your answer from part (e) when  $\beta$  is small?*

## Chapter 9: Order parameters, broken symmetry, and topology

### Exercises

#### 9.9 Nematic order parameter half space.<sup>95</sup> *p*

Let us develop some intuition for the projective plane, the order parameter space of a nematic.



**Fig. 9.35 Projective Plane.** Just like the torus, the order parameter space  $\mathbb{RP}^2$  for a nematic can be described as a square with opposite sides identified, but now with a twist. (Think of the square as a top-view of the hemisphere, when the latter is pinched outward along the four diagonals.) The dashed line, for example, connects two points on the edge of the square that correspond to the same nematic orientation. What geometrical figure (disk, cylinder, Möbius strip, sphere with a hole, two disks) is formed when you cut along the dashed line?

(a) If one cuts the order parameter space for a two-dimensional crystal (Fig. 9.7) around the edge, it becomes a cylinder. What topological object is formed by cutting the order parameter space for a nematic along the curved path shown in

#### Fig. 9.5b?

This provides us an excuse to exercise our skills with scissors and tape. Unfortunately, the projective plane cannot be pasted together in ordinary three dimensions: we will need to improvise.

(b) Get an printed, elongated version of Fig. 9.35, suitable for cutting.<sup>96</sup> Tape the short ends together, aligning the arrows. What object do you form? Cut along the dotted line. Can you now tape the long ends together? (Remember, in higher dimensions one can pass one paper strip through another (moving it into the fourth dimension). You can mimic this by cutting a strip and passing it (without twisting) past another before taping it back together.) What object do you form in the end?

#### 9.10 Ising order parameter. *p*

What would the topological order parameter space be for the Ising model? (Hint: what is the space of possible ground states?)

#### 9.11 Pentagonal order parameter.<sup>97</sup> (Mathematics) *i*

Glasses are rigid, like crystals, but have disordered atomic positions, like liquids. We do not have a definitive theory of glasses, but one feature appears to be frustration – good glass-forming systems can form local low energy structures that are difficult to arrange into crystalline patterns. The classic example is the packing of spheres. Four spheres can neatly pack into a low-energy regular tetrahedron, but tetrahedra cannot fill space (Fig. 12.19).

The regular tetrahedron has six edges, each with an dihedral angle of  $\arccos(1/3)=70.5^\circ$ .<sup>98</sup> Undistorted tetrahedra cannot fit together around an edge without leaving a gap.

<sup>95</sup>A printable version of Fig. 9.35 can be found at the book Web site [49].

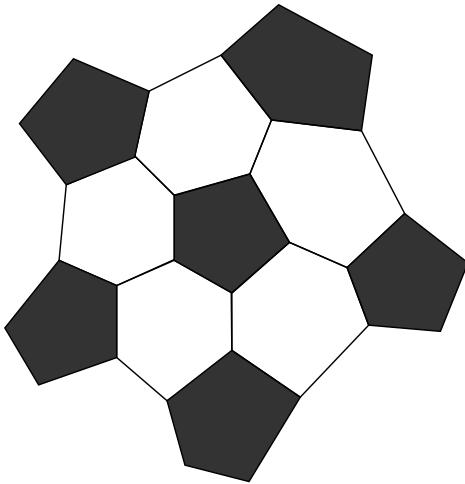
<sup>96</sup>If you are not provided with one, you can download it from the computer exercises site [49].

<sup>97</sup>A printable version of Figs. 9.36 and 9.37 can be found at the book Web site [49].

<sup>98</sup>A dihedral angle is the angle between two intersecting planes – the ‘opening angle’ of the edge.

- (a) How many tetrahedra meet at a vertex in the frustrated icosahedron of Fig. 12.19a? What is the angle of the gap formed by this many tetrahedra meeting at an edge?

We can create a frustrated order parameter for this kind of glassy system, by considering an ‘ideal glass’ formed by curving space to close the gap. Consider the analogous problem – a two-dimensional material whose lowest-energy local configurations form the faces of a soccer ball, Fig. 9.36.



**Fig. 9.36 Frustrated soccer ball.** A material that prefers to form alternating pentagons and hexagons cannot fill two-dimensional flat space, but does live happily on the soccer-ball sphere  $S^2$  in three dimensions. Note the distortion of the pentagons and hexagons becomes more and more severe as one works outward from the origin.

Figure 9.36 shows a low energy region. To describe the order at a point  $\mathbf{x}$  on the plane, we must both identify the orientations of the polygons and the translational positions within the polygons. How can we combine both into a single order parameter field?

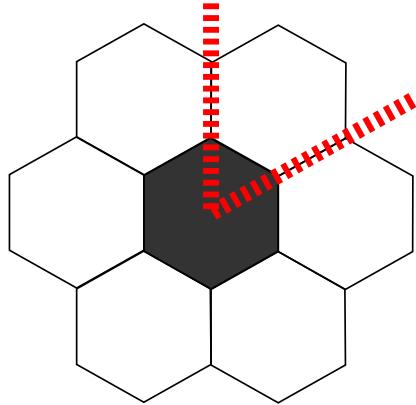
- (b) Get a soccer ball. Generate a copy of Fig. 9.36 blown up so that the polygons roughly match the size of the faces of the soccer ball. Pick a point  $\mathbf{x}$  on the sheet, and align the ball atop the sheet so as to match the local material structure to the ideal soccer-ball template. What is the order pa-

rameter space you need to describe the local order?

(Hint: This is a challenging question. You need a three-dimensional rotation matrix to define the orientation of the soccer ball tangent to  $\mathbf{x}$ . Is there more than one rotation matrix that can give an equivalent local fit to the structure?)

Constructing this order parameter, and exploring the consequences, consumed a large portion of the author’s post-doctoral years. In particular, the natural low energy structure in Fig. 9.36 is given by rolling the sphere without slipping (a kind of ‘parallel transport’).

One cannot fill 2D space by flattening the sphere without stretching (a longstanding complaint of map-makers). One must introduce a disclinations, by adding a wedge of material chosen to sew naturally into the gaps created by flattening the ideal template (Fig. 12.19(a) and Fig. 9.37). For our soccer-ball material, one disclination is needed at the center of each pentagon.



**Fig. 9.37 Cutting and sewing: disclinations.** The natural defect needed to flatten the soccer ball is a disclination, introducing a wedge of material at the center of a pentagon. This produces a net change in angle as one follows the order parameter around a loop containing the pentagon’s center.

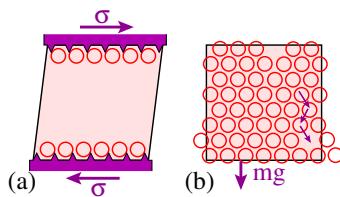
- (c) Get a copy of Fig. 9.37 blown up so that the polygons match the faces on the soccer ball.<sup>99</sup> Place a pentagon of the soccer ball atop the central black polygon, tilted so that one edge on the ball lies atop the corresponding edge in the material. Roll the soccer ball around the central polygon so

<sup>99</sup>If you are not provided with one, you can download it from the computer exercises site [49].

as to trace out its perimeter. As you traverse a path  $\mathbf{x}(\theta)$  an angle  $\theta = 2\pi$  around the center of a pentagon in the plane, what is the net angle  $\phi$  that the soccer ball rotates? If we define the strength of the defect line as  $s = \phi/2\pi$  in analogy with the nematic defect line of Fig. 9.19, what is  $s$ ?

- 9.12 Rigidity of crystals.** (Order parameters)  $\textcircled{p}$   
 Rocks, bridges, and our bones are rigid; unlike fluids, they resist shear. This is because bridges and most rocks are polycrystalline. The free energy of a crystal rises if the atoms near one surface are displaced relative to those on a distant surface (Fig. 9.38a), because the atoms in a crystal have a broken translational symmetry.<sup>100</sup> In an unstressed, equilibrium crystal the atoms on average lie on a regular grid of crystalline lattice positions. Shifting the top layer with respect to the bottom layer forces the grid to distort; this elastic deformation increases the energy. Rigidity of the order parameter is a hallmark of broken symmetry states of matter.

Mountains are built of rocks, and they hold up under gravity. Is this fundamentally because they are rigid? Or is a rock placed on your desk flowing into a puddle like a liquid? It turns out that gravity can transport mass *through* a crystalline lattice without shearing planes of atoms past one another. In most crystals, the dominant mechanism is through *vacancy diffusion* (Fig. 9.38b).



**Fig. 9.38 Crystal rigidity [46].** (a) Crystals are rigid to shears  $\sigma$  that couple to the lattice (*i.e.*, the broken translational symmetry). (b) Under a force like gravity that couples to the mass density, an equilibrium crystal *can* flow like a liquid (at a rate linear in the gravitational force), via vacancy diffusion. (This is true at temperatures above the *roughening transition* [14], where the equilibrium surfaces all have steps that can absorb and emit vacancies.)

As vacancies move upward, there is a net current of atoms downward. The rate will be some prefactor times the probability  $\exp(-\beta f_{\text{trans}})$  per site that the atoms in the crystal have a vacancy poised at the transition state saddle-point<sup>101</sup> between two lattice sites (see Exercise 6.11). It is traditional to think of this probability as the product of the density of vacancies times the probability that a given vacancy will move to its saddle-point. Let the free energy cost of a vacancy be  $F_V = E_V - TS_V$ . Let the free energy barrier for the motion of a vacancy be  $F_B = E_B - TS_B$ . Then we see that  $f_{\text{trans}} = F_V + F_B$ .

The experimentalists [52] estimate that  $S_{\text{trans}} = S_V + S_B = 10.9k_B$ . A commonly considered contribution to the entropy of a defect, or the entropy of a transition state, is the change in the vibrational frequencies of nearby atoms. For simplicity, let us model the entropy change for each of these two as due to a change of frequency of one local normal mode. So, *e.g.*, for  $S_V$ , the atoms might vibrate radially outward from a vacancy more slowly ( $\omega_V$  than they did when an atom occupied the site ( $\omega_{\text{noV}}$ ), with all the other modes unchanged.

(a) Assume the temperature is high compared to  $\hbar\omega$  for all the oscillators, so that the entropies are all equal to the classical entropies of the harmonic oscillators. Calculate  $S_V$  in terms of  $\omega_V$  and  $\omega_{\text{noV}}$ . How much would the frequency need to change to account for half of the experimentally measured entropy? Does this frequency change seem plausible? Discuss.

The free energy of a vacancy, or the barrier to its motion, can depend on temperature for reasons other than the entropy of local phonons. For example, the thermal expansion of our silicon crystal will lower the energy needed for the companion atom to squeeze between neighbors across the barrier. It is likely that silicon vacancy diffusion has a large effective prefactor partly due to this mechanism.

How long will it take for, say, an equilibrium silicon crystal to flow into a puddle? The self-diffusion constant for silicon is experimentally measured to be  $D = D_0 \exp(-E_{\text{trans}}/k_B T)$ , with  $D_0 \approx 0.04 \text{ m}^2/\text{s}$  and the activation barrier at the transition state  $E_{\text{trans}} \approx 4.7 \text{ eV}$  [52].

<sup>100</sup>The breakdown of rigidity in crystals is also discussed in Exercise 11.5 (dislocation-mediated plasticity) and Exercise 11.14 (fracture of crystals).

<sup>101</sup>As an atom falls one lattice site into a vacancy below it, the saddle-point is crossed at the midpoint of the transition.

(b) Using the Einstein relation relating the mobility to the diffusion constant, give a formula for the upward current  $J$  of vacancies due to gravity, as a function of the diffusion constant  $D$ , the temperature  $k_B T$ , the mass  $m_{\text{Si}}$  of a silicon atom, the number density  $\rho_{\text{Si}}$  of silicon atoms per unit volume, and the force  $g$  due to gravity. Give the formula involving  $J$  giving the velocity  $v$  at which the top surface of silicon will shrink downward due to gravity. Evaluate  $v$  for silicon just below its melting point,  $1414^\circ\text{C} = 1687^\circ\text{K}$ . How big is it in Ångstroms/year? (Useful:  $1\text{eV} = 1.6 \cdot 10^{-19}\text{J}$ ;  $k_B = 1.38 \cdot 10^{-23}\text{J/K}$ ;  $m_{\text{Si}} = 28.085\text{au} = 4.66 \cdot 10^{-26}\text{kg}$ ;  $g = 9.8\text{m}^2/\text{s}$ , and the density of silicon atoms is  $\rho_{\text{Si}} = 5 \cdot 10^{28}\text{atoms/m}^3$ . The current you derived in (b) should be linear in the applied force – the same linear response one would find in a liquid. Apart from flowing slowly, are there any other differences in the behavior between the flow of our silicon crystal and than of an extremely viscous liquid?<sup>102</sup>

(c) Will the crystalline axes be preserved under the flow? Would an single-crystalline ice cube, placed with its hexagonal axis vertical on a surface kept below freezing, flow into a circular, squashed droplet (like a liquid) after reaching its equilibrium shape? Into a snowflake, with complex dendritic branches? Roughly what shape would you expect? (Hint: The shape of a water drop on a surface is set by the balance of gravity and surface tension. Should the interfacial tension between a crystal and gas be isotropic?)

Vacancy diffusion is usually the dominant mechanism for linear response bulk flow in perfect crystals, but there are usually much faster mechanisms for moving atoms in solids. Ice in your freezer sublimates (evaporates directly from the solid to gas), explaining why your ice cubes become rounded with time, especially with self-defrosting freezers that occasionally raise their temperature above freezing. Diffusion on surfaces is much faster than bulk diffusion, but scales differently with system size  $L$  (since the current downward is proportional to the perimeter  $L$  and not the area  $L^2$ ). Non-equilibrium defects in the crystal (grain boundaries and dislocations) also often have much higher diffusion rates, and also can move to change the crystal shape. (Note that these defects do not exist in an unstrained equilibrium crystal, and their nucleation rate is not

linear in the applied force, Exercise 11.5.)

### 9.13 Chiral wave equation. ②

The evolution of a physical system is described by a field  $\Xi$ , obeying a partial differential equation

$$\partial\Xi/\partial t = A \partial\Xi/\partial x. \quad (9.86)$$

(a) Symmetries. Give the letters corresponding to ALL the symmetries that this physical system appears to have:

- (1) Spatial inversion ( $x \rightarrow -x$ ).
- (2) Time-reversal symmetry ( $t \rightarrow -t$ ).
- (3) Order parameter inversion ( $\Xi \rightarrow -\Xi$ ).
- (4) Homogeneity in space ( $x \rightarrow x + \Delta$ ).
- (5) Time translational invariance ( $t \rightarrow t + \Delta$ ).
- (6) Order parameter shift invariance ( $\Xi \rightarrow \Xi + \Delta$ ).

(b) Traveling Waves. Show that our equation  $\partial\Xi/\partial t = A \partial\Xi/\partial x$  has a traveling wave solution. If  $A > 0$ , which directions can the waves move?

### 9.15 Superfluid second sound. ③

There are two different order parameters for superfluids. The ‘soft-spin’ Landau order parameter is a complex number as a function of position. (Remember the Bose condensate  $\Psi(\mathbf{r}) = \psi_1(x_1)\psi_2(x_2)\dots$  of non-interacting bosons; the quantum single-particle state  $\psi(x)$  gives the Landau order parameter.) The ‘topological’ order parameter is the phase  $\phi(x)$  of this complex function  $\psi(x) = \psi_0 \exp(i\phi(x))$ . Superfluids break gauge symmetry (the arbitrary phase  $\phi$  of the quantum wavefunction).

Under most circumstances, for every continuous broken symmetry there is an elementary excitation consisting of a long wavelength, sinusoidal variation of the broken symmetry direction of the corresponding order parameter (Goldstone’s theorem).

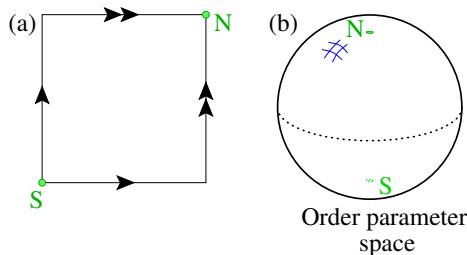
Argue that the elementary excitation for the superfluid order is a slowly-varying small amplitude oscillation in the phase, as in  $\phi(x) = \phi_0 + a \cos(kx)$ . Using the single-particle formula  $\mathbf{J} = (\hbar/2mi)(\psi^* \nabla \psi - \psi \nabla \psi^*)$  (treating the superfluid as a Bose condensate as in Exercise 9.7a), argue that the oscillation  $\phi(x)$  will change the density of superfluid bosons periodically in space.

There is already another density wave in fluids – the longitudinal phonon that gives ordinary sound

<sup>102</sup>This part can be done independently of (a) and (b).

in liquids and gases. One often uses a *two-fluid model* with a superfluid density and a density of ‘normal fluid’ (quasiparticles); sound waves are an in-phase oscillation of the two fluids (with net density changes), and ‘second sound’ has the normal and superfluid components oscillating out of phase (with little net density change). Since the superfluid carries no heat, the ‘normal fluid’ component of second sound forms an oscillation of heat entropy...

- 9.16 Can’t lasso a basketball.**<sup>103</sup> (Mathematics) ⑦ The magnetization  $\mathbf{M}$  of a Heisenberg ferromagnet is of fixed magnitude  $|\mathbf{M}| = M_0$ , but can point with equal energy in any direction. Thus the order parameter space for this magnet is a sphere (Fig. 9.39(b)). Mathematicians sometimes construct the sphere topologically by pasting together a square as in Fig. 9.39(a), with the corresponding sides matched according to the style of arrows (single or double), and with the sides pasted so that the arrows point in the same directions.



**Fig. 9.39 Sphere.** The sphere (b) can be topologically constructed by gluing together the edges of a square (a) as shown.

(a) To test that you understand this cutting and pasting, sketch directly onto Fig. 9.39b the seams where the edges are pasted together, turning the square (a) into the sphere (b). Include the arrows along the seams, with the correct styles and orientation. Assume that the seams are predominantly on the visible (front) side of the sphere, and that the north pole  $N$  and south pole  $S$  are as shown. (Did you draw the arrows?)

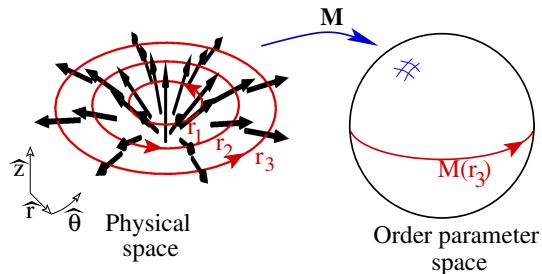
There are no topological line defects for Heisenberg ferromagnets. This is because  $\Pi_1(\mathbb{S}^2) = 0$ ;

any closed path on the sphere can be contracted to a point, so a loop in real space never surrounds something that cannot be smoothly filled in. We can give a tangible example of this – sometimes called ‘escape into the third dimension’.

A two-dimensional Heisenberg ferromagnet has an order parameter field

$$\mathbf{M} = M_0(\hat{z}\lambda(r) + \hat{r}\sqrt{1 - \lambda^2}), \quad (9.87)$$

expressed in cylindrical coordinates. (Fig. 9.40). Here  $\lambda(r)$  decreases rapidly from one at  $r = 0$  to zero at larger  $r$ , so that  $\mathbf{M}$  points radially outward at long distances from the origin. One may thus measure a winding number  $s = 1$  for this magnetization pattern at large  $r$ , but it nonetheless has a smooth magnetic field everywhere inside.



**Fig. 9.40 Escape in the third dimension.** The magnetic field for a planar two-dimensional material points upward (along  $\hat{z}$ ) at the origin, and points almost radially outward (along  $\hat{r}$ ) far away (radius  $r_3$ ). As one circles  $\vec{r}$  around the origin counter-clockwise at distance  $r_3$ , the magnetization  $\mathbf{M}(\vec{r})$  circles the order parameter space around the equator, as shown.

(b) Just as the image  $\mathbf{M}(\vec{r})$  of the magnetization along  $r = r_3$  is shown in order parameter space, sketch on the right the approximate curves described by the magnetization at distances  $r_2$  and  $r_1$  from the origin, as denoted in real space on the left. Label these  $M(r_1)$  and  $M(r_2)$ , in analogy with  $M(r_3)$  shown.

(c) Does the path in order parameter space contract to a point, as the loop in real space shrinks to  $r = 0$ ? What is the homotopy group element of the sphere,  $g \in \Pi_1(\mathbb{S}^2)$ , corresponding to  $M(r_3)$ ?

<sup>103</sup>Topologically, one must admit, you also cannot lasso a cow. Practically, you cannot lasso a doughnut.

## Chapter 10: Correlations, response, and dissipation

### Exercises

**10.9 Quasiparticle poles and Goldstone's theorem.** (Condensed matter) ③

We saw in Exercise 9.14 that sound waves obey Goldstone's theorem. As the fundamental excitations associated with a continuous (translational) symmetry, they have low frequency excitations at long wavelengths, whose damping disappears as the frequency goes to zero. Here we want to explore the complex susceptibility of damped sound, and look for analogues of quasiparticle in the behavior of sound.

We add forcing  $f(x, t)/\rho$  to eqn 9.29 for the damping of sound waves:

$$\frac{\partial^2 u}{\partial t^2} = c^2 \frac{\partial^2 u}{\partial x^2} + d^2 \frac{\partial}{\partial t} \frac{\partial^2 u}{\partial x^2} + \frac{f(x, t)}{\rho}. \quad (10.88)$$

Here again  $c$  is the speed of sound  $c$ ,  $\rho$  the density, and  $d^2$  is the Kelvin damping term consistent with Galilean invariance.

If we force the undamped system with a forcing  $f(x, t) = \cos(kx) \cos(ckt)$  that matches the wavevector and frequency  $\omega = ck$  of one of the phonons, the amplitude of that mode of vibration will grow without limit; the susceptibility diverges. Let us try it with damping.

(a) *What is the susceptibility  $\tilde{\chi}(k, \omega)$  for our damped system (eqn 10.88)?* (Hint: Change to Fourier variables  $k$  and  $\omega$ , and the exercise should reduce to simple algebra.) *Check that your answer without dissipation ( $d = 0$ ) has poles<sup>104</sup> at  $\omega_k = \pm ck$  for each wavevector  $k$ ; the susceptibility diverges when excited on resonance. Show that  $\Omega_k = \omega_k - i\Gamma_k$  when dissipation is included.* (If you have done Exercise 9.14, you may want to check your formula for  $\Gamma_k$  with the dispersion relation you derived there.)

(b) *Check that the poles in the susceptibility are all in the lower half-plane, as required by causality (Section 10.9).*

Neutron scattering can be used to measure the Fourier transform of the correlation function,

$\tilde{C}(k, \omega)$  in a material. Suppose in our material  $c = 1000 \text{ m/s}$ ,  $d = 10^{-3} \text{ m/s}^{1/2}$ ,  $\rho = 1 \text{ g/cm}^3$ , and  $T = 300 \text{ K}$ ; for your convenience,  $k_B = 1.3807 \times 10^{-23} \text{ J/K}$ .

(c) *Use the fluctuation-dissipation theorem*

$$\chi''(\mathbf{k}, \omega) = \frac{\beta\omega}{2} \tilde{C}(\mathbf{k}, \omega) \quad (10.89)$$

(see eqn 10.65) to calculate  $\tilde{C}(k, \omega)$ , the correlation function for our material. Plot  $\tilde{C}(k, \omega)$  at  $\omega = 10^8$  (about 16 MHz), for  $k$  from zero to  $2 \times 10^5$ . Does it appear difficult to estimate the dispersion relation  $\omega_k$  from the correlation function?

In strongly-interacting systems, the elementary excitations are *quasiparticles*: dressed electrons, photons, or vibrational modes which are not many-body eigenstates because they eventually decay (note 23 on p. 180). These quasiparticles are *defined* in terms of poles in the propagator or quantum Green's function, closely related to our classical correlation function  $\tilde{C}(k, \omega)$ .

**10.10 Human correlations.** ④

Consider a person-to-person correlation function for the horizontal center-of-mass positions  $\mathbf{h}_\alpha = (x_\alpha, y_\alpha)$  of people labeled by  $\alpha$ , in various public places. In particular, let  $C_{\text{crowd}}(\mathbf{r}) = \langle \rho(\mathbf{h})\rho(\mathbf{h}+\mathbf{r}) \rangle$ , where  $\rho(\mathbf{h}) = \sum_\alpha \delta(\mathbf{h} - \mathbf{h}_\alpha)$ , and the ensemble average  $\langle \dots \rangle$  is over a large number of similar events.

(a) *Sketch a rough, two-dimensional contour plot for  $C_{\text{crowd}}(\mathbf{r})$ , for a subway car of length  $L = 10\text{m}$  with two benches along the edges  $x = 0$  and  $x = 3\text{m}$ . Assume the people are all seated, with random distances  $\delta y = 0.5 \pm 0.2\text{m}$  between neighbors.*

(b) *Sketch a rough contour plot for  $C_{\text{crowd}}(\mathbf{r}, t)$  in the same subway car, where  $t = 20$  minutes is long enough that most people will have gotten off at their stops.*

<sup>104</sup>A function  $\psi(z)$  has a *pole* at  $z_0$  if it diverges at  $z_0$  like a constant times  $z - z_0$ . The poles are easy to find if  $\psi$  is a ratio of two functions which themselves do not go to infinity; the poles are the zeros of the denominator.

(c) Sketch a rough contour plot for the Fourier transform  $\tilde{C}_{\text{crowd}}(\mathbf{k})$  for the correlation function in part (a).

### 10.11 Onsager regression hypothesis. i

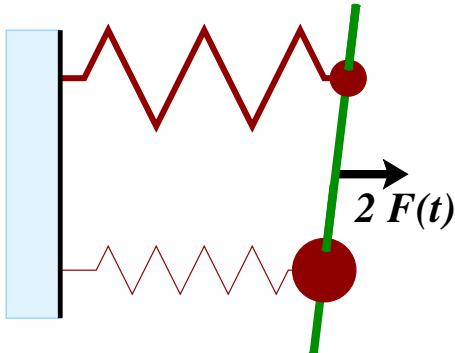
A harmonic oscillator with mass  $M$ , position  $Q$  and momentum  $P$  has frequency  $\omega$ . The Hamiltonian for the oscillator is thus  $\mathcal{H} = P^2/2M + \frac{1}{2}M\omega^2Q^2$ . The correlation function  $C_Q(\tau) = \langle Q(t+\tau)Q(t) \rangle$ .

(a) What is  $C_Q(0)$ ? (Hint: Use equipartition.) The harmonic oscillator is coupled to a heat bath, which provides a dissipative force  $-\dot{Q}/\eta$ , so under an external force  $F(t)$  we have the macroscopic equation of motion  $\ddot{Q} = -\omega^2Q - \dot{Q}/\eta + F/M$  (ignoring thermal fluctuations). We assume the harmonic oscillator is overdamped, so

$$\frac{dQ}{dt} = -\lambda Q + (\lambda/M\omega^2)F, \quad (10.90)$$

where  $\eta = \lambda/M\omega^2$ . We first consider the system without an external force.

(b) What is  $C_Q(\tau)$ , in terms of  $\lambda$  and your answer from (a)? (Hint: Use the Onsager regression hypothesis to derive an equation for  $dC_Q/d\tau$ .)



**Fig. 10.41 Two springs.** Two harmonic oscillators, with an external force  $2F(t)$  coupled to the average displacement  $(Q(t)+q(t))/2 = X(t)/2$  (equivalent to  $F(t)$  coupled to the sum  $Q+q=X$ ).

Now consider a system of two uncoupled harmonic oscillators, one as described above and another with mass  $m$ , position  $q$ , momentum  $p$ , and frequency  $\Omega$ , coupled to a heat bath with overdamped decay rate  $\Lambda$ . (Capital letters are meant to indicate size: this is a small mass with a high frequency and a quick decay time.) Here we only observe<sup>105</sup>  $X = Q + q$ . Over the following few exercises, we shall compute properties of this combination  $X$ .

(c) What is  $C_X(0)$ ?  $C_X(\tau)$ ? Roughly sketch a graph of the latter, assuming  $\lambda \approx \Lambda/10$  and  $M\omega^2 \approx m\Omega^2/2$ . (Hint: They are completely uncoupled, and so also uncorrelated random variables.)

A consequence of the Onsager regression hypothesis is that the correlation function decays ‘according to the same laws as [a deviation] that has been produced artificially’. But clearly there are different ways in which  $X(t)$  could evolve from an artificially imposed initial condition, depending on the way we start  $Q$  and  $q$ .

(d) How would  $X(t)$  decay from an initial state  $Q(0) = X(0)$ ,  $q(0) = 0$ ? From a state  $q(0) = X(0)$ ,  $Q(0) = 0$ ? Does either decay the way  $C_X(t)$  decays from part (c)? How would  $X(t)$  decay from an initial state prepared with a sudden impulsive force  $F(t) = J_{\text{imp}}\delta(t)$ ? From an initial state prepared with a constant force  $F(t) = F_0\Theta(-t)$ ? Does either decay in the same fashion as  $C_X(t)$ ?

Onsager’s macroscopically perturbed system must be perturbed slowly compared to the internal dynamical relaxation times, for its future evolution to be determined by its current state, and for his hypothesis to be well posed. Stated another way, Onsager hypothesizes that we have included all of the slow variables into the macroscopic observables whose dynamics and correlations are under consideration.

### 10.12 Liquid free energy. p

The free energy  $F$  of a liquid (unlike an ideal gas), resists rapid changes in its density fluctuations  $\rho(x)$ :

$$F(\rho) = \int \frac{1}{2}K(\nabla\rho)^2 + \frac{1}{2}\alpha(\rho - \rho_0)^2 dx.$$

(See Exercise (9.4), where a similar term led to a surface tension for domain walls in magnets.)

<sup>105</sup>That is,  $Q$  and  $q$  represent microscopic degrees of freedom in a material, and  $X$  represents a macroscopic property whose correlation function Onsager is trying to estimate.

Thus in a periodic box of length  $L = 1$ ,

$$F(\tilde{\rho}) = \sum_m \frac{1}{2}(Kk_m^2 + \alpha)|\tilde{\rho}_m|^2 L,$$

with  $k_m = 2\pi m/L$ .

(a) *Show that equal-time correlation function is  $\tilde{C}(k_m) = |\tilde{\rho}(k_m)|^2 = k_B T / (Kk_m^2 + \alpha)L$ . (Hint:  $F$  is a sum of harmonic oscillators; use equipartition.) How should we express it in the continuum limit, where  $L \rightarrow \infty$ ? (Hint:  $\sum_m \equiv L/(2\pi) \int dk$ . Alternatively, as a function of continuous  $k$ ,  $\tilde{C}(k) = k_B T / (Kk^2 + \alpha)$ ).*

(b) *Transform to real space. Does the surface tension introduce correlations (as opposed to those in the ideal gas)? Do the correlations fall off to zero at long distances? (Remember the  $1/2\pi$  in the continuum inverse Fourier transform of eqn A.7. Hint:  $\int_{-\infty}^{\infty} \exp(iqy)/(1+q^2) dq = \pi \exp(-|y|)$ ).<sup>106</sup>*

The density of a liquid moving through a porous medium will macroscopically satisfy the diffusion equation.

(c) *According to Onsager's regression hypothesis, what will the correlation  $\hat{C}(k, t)$  be? (See note 12 on page 271 and note 13 on page 272.)*

It is non-trivial to Fourier transform this back into real space, but the short-distance fluctuations (as they do in the ideal gas) will get smoothed out even more with time.

### 10.13 Harmonic susceptibility and dissipation. (i)

This is a continuation of exercise 10.11, where we considered damped harmonic oscillators. Consider one harmonic oscillator  $(P, Q)$  with  $\mathcal{H} = P^2/2M + \frac{1}{2}M\omega_0^2Q^2 - F(t)Q(t)$ , overdamped with decay rate  $\lambda$ .

The static susceptibility  $\chi_0$  is defined as the ratio of response  $Q$  to force  $F$ .

(a) *Under a constant external force  $F_0$ , what is the equilibrium position  $Q_0$ ? What is the static susceptibility  $\chi_0$ ? (Hint: Ignoring fluctuations, this is just freshman mechanics. What is the spring constant for our harmonic oscillator?) Generally speaking, a susceptibility or compliance will be the inverse of the elastic modulus or stiffness – sometimes a matrix inverse.*

Consider our harmonic oscillator held with a constant force  $F$  for all  $t < 0$ , and then released. For  $t > 0$ , the dynamics is clearly an exponential decay (releasing the overdamped harmonic oscillator), and can also be written as an integral of the time-dependent susceptibility  $\chi(\tau)$  for the oscillator:

$$\begin{aligned} Q(t) &= Q_0 \exp(-\lambda t) \\ &= \int_{-\infty}^0 dt' \chi(t-t') F \end{aligned} \quad (10.91)$$

(Note that the integral ends at  $t' = 0$ , where the force is released.)

(b) *Solve eqn 10.91 for  $\chi(\tau)$ . (Hint: You can take derivatives of both sides and integrate the right-hand-side by parts. Or you can plug in  $\chi(\tau)$  as an exponential decay  $A \exp(-Bt)$  and solve.)*

We can now Fourier transform  $\chi$  to find  $\tilde{\chi}(\omega) = \int dt \exp(i\omega t) \chi(t) = \chi'(\omega) + i\chi''(\omega)$ .

(c) *What is  $\tilde{\chi}(\omega)$ ?  $\chi'(\omega)$ ?  $\chi''(\omega)$ ? Is the power dissipated for an arbitrary frequency dependent force  $f_\omega$  (given by eqn 10.37) always positive?*

### 10.14 Liquid susceptibility and dissipation. (p)

(Continuation of Liquid free energy exercise.) Looking ahead to eqn 10.60, calculate the liquid susceptibility in  $k$ -space  $\hat{\chi}(k, t)$  using your Onsager's regression analysis from the Liquid free energy exercise, and doing Fourier transform to find the real and imaginary parts  $\hat{\chi}(k, \omega) = \hat{\chi}'(k, \omega) + i\hat{\chi}''(k, \omega)$ . Show in particular that

$$\hat{\chi}''(k, \omega) = \frac{Dk^2\omega}{(Kk^2 + \alpha)(D^2k^4 + \omega^2)}.$$

(Remember that  $\chi = 0$  for  $t < 0$ .) Is the imaginary part positive at all  $k$  and  $\omega$ ? Is the power dissipated (eqn 10.37) positive?

### 10.15 Harmonic fluctuation dissipation. (i)

This is a continuation of Exercise 10.11, where you derived the correlation function  $C_Q(\tau) = A \exp(-\lambda\tau)$  for the overdamped harmonic oscillator with rate  $\lambda$ , and the correlation function  $C_X(\tau) = A \exp(-\lambda\tau) + B \exp(-\Lambda\tau)$  for the sum  $X = Q + q$  of two uncoupled oscillators with overdamped decay rates  $\lambda$  and  $\Lambda$ . In Exercise 10.13 we calculated the susceptibility

<sup>106</sup>The function  $1/(1+q^2)$  is often called a Lorentzian or a Cauchy distribution. Its Fourier transform can be done using Cauchy's residue theorem, using the fact that  $1/(1+q^2) = 1/((q+i)(q-i))$  has a pole at  $q = \pm i$ . So  $\int_{-\infty}^{\infty} \exp(iqy)/(1+q^2) dq = \pi \exp(-|y|)$ . The absolute value happens because one must close the contour differently for  $y > 0$  and  $y < 0$ .

directly. Now we shall calculate it using the fluctuation-dissipation theorem.

(a) Using the classical fluctuation-dissipation theorem, eqn 10.60, derive the time-dependent susceptibilities  $\chi_Q(\tau)$  and  $\chi_X(\tau)$ . You may leave your answer in terms of  $A$  and  $B$ .

(b) Sketch a graph of  $\chi_X(\tau)$  for  $-1/\lambda < \tau < 1/\lambda$ , assuming that  $\lambda \approx \Lambda/10$  and  $A \approx 2B$ .

### 10.16 Liquid static susceptibility and fluctuation-dissipation theorem. $\oplus$

(Continuation of Liquid free energy exercise.)

(a) Use your equal-time correlation function for the liquid free energy to derive the static density profile  $\rho(x)$  of our liquid if a static force  $f\delta(x)$  is applied at  $x = 0$ .

(b) Check the AC version of the classical fluctuation-dissipation theorem using the correlation function  $\tilde{C}(k, t)$  and  $\chi''(k, \omega)$ .

### 10.17 Harmonic Kramers-Krönig. $\circledast$

In general, susceptibilities are causal:  $\chi(\tau) = 0$  for  $\tau < 0$  because the response cannot precede the force. We saw in general that this implied that  $\tilde{\chi}(\omega)$  had to be analytic in the upper half plane,  $\text{Im}(\omega) > 0$ .

Let us check this for the overdamped harmonic oscillator, where  $\chi(\tau) = C \exp(-\lambda\tau)$ .

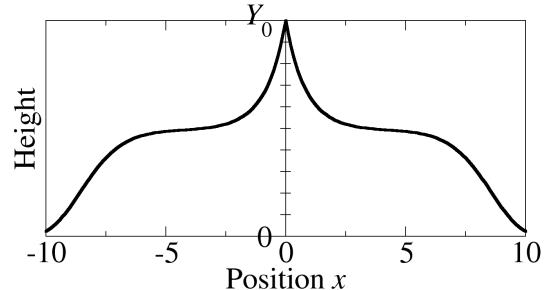
What is  $\tilde{\chi}(\omega)$ ? Where are its poles? Are they in the upper half plane? Are there any singularities in the lower half plane? (Beware: The conventions on Fourier transforms are pretty strange (Appendix 13)). Whether the poles are in the upper or lower half plane depends on whether going from  $t$  to  $\omega$  multiplies by  $\exp(i\omega t)$  or by  $\exp(-i\omega t)$ . Equation A.6 tells us the former choice is correct, but eqn A.9 tells us the opposite convention holds for spatial Fourier transforms.)

### 10.18 Liquid Kramers Krönig. $\oplus$

(Continuation of Liquid free energy exercise.)

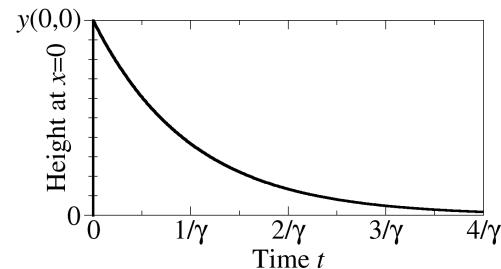
Check your formula for the liquid susceptibility  $\tilde{\chi}(k, \omega)$ . Is it analytic in the upper half plane? Explicitly check the Kramers-Krönig formulas for the real and imaginary parts  $\chi'$  and  $\chi''$ . (Hint: Integrating through a pole, using the principle value, gives the average of an integral passing below the pole and above the pole.)

### 10.19 Susceptibilities and correlations. $\circledast$



**Fig. 10.42 Pulling.** An order parameter  $y$  is pulled upward slightly by a time-independent point force  $F$  at the origin; the resulting height  $Y(x)$  is shown. Without the force, the order parameter exhibits small fluctuations about height zero.

(a) The height fluctuations are measured in the same system of Fig. 10.42, at a temperature  $T$ , in the absence of an external force. Calculate the thermal ensemble average  $\langle y(0, t)y(\Delta x, t) \rangle$  of these height fluctuations without the force, in terms of  $Y(x)$ ,  $F$ , and  $k_B T$ . Plot it over  $x \in (-10, 10)$ , labeling the limits of your vertical axis in terms of these same variables. (Hints: The fluctuations should be larger for higher temperature, and the response should be larger for higher forces. Section (10.7) may be useful.)



**Fig. 10.43 Hitting.** The height  $y(x, t)$  at the origin  $x = 0$  after the same system is hit upward with an impulsive force  $g\delta(x)\delta(t)$ . The decay is exponential,  $y(0, t) = y(0, 0) \exp(-\gamma t)$  for  $t > 0$ .

(b) The same system is hit (Fig. 10.43) and the time decay at the origin is measured. The height fluctuations are measured in the system at  $x = 0$  as a function of time in the absence of an external hit. Calculate the thermal ensemble average  $\langle y(0, t)y(0, t + \Delta t) \rangle$  without the hit, in terms of  $y(0, 0)$ ,  $g$ ,  $k_B T$ ,  $\Delta t$ , and  $\gamma$ . Plot it

for  $t \in (-4/\gamma, 4/\gamma)$ , labeling the limits of your vertical axis in terms of these same variables.

- (c) Give a formula for  $Y_0$  in Fig. 10.42 (the response to a time-independent pull), in terms of  $y(0,0)$  (the response to a  $\delta$ -function hit),  $g$ ,  $F$ ,  $\gamma$ , and  $k_B T$ .

### 10.20 Subway bench correlations. ③

We shall study the correlation function between the centers of passengers sitting on a bench alongside the windows in a subway car. We model the system as a one-dimensional equilibrium hard-disk gas of  $N$  disks with length  $a$ , contained in a car  $0 < y < L$ .<sup>107</sup>

If  $Na = L$ , there is no free space on the bench, and the passengers have centers of mass at packed positions

$$y_{\text{packed}} = [a/2, 3a/2, \dots, (2N-1)a/2]. \quad (10.92)$$

Call the free space  $L_{\text{free}} = L - Na$ . We can relate the ensemble of hard disks in one dimension in a box of length  $L$  and an ideal gas in a box of length  $L_{\text{free}}$ .

- (a) Show that the configuration  $y_1 < y_2 < \dots < y_N$  of passengers sitting on the bench of length  $L$  corresponds precisely to legal configurations  $z_1 < \dots < z_N$  of ideal gas particles in a box of length  $L_{\text{free}}$ , with  $z_n$  measuring the free space to the left<sup>108</sup> of passenger  $n$ , by giving an invertible linear mapping from  $\mathbf{z} \rightarrow \mathbf{y}$ . Is the configurational entropy the same for the ideal gas and the subway car? Why? (Feel free to look up the entropy of the Tonks gas to check your answer. But do explain your reasoning. Would the entropy be the same if the mapping were nonlinear, but invertible?)

The correlation functions for the ideal gas and the subway car are definitely not the same. But the mapping to the ideal gas provides a very efficient way of sampling the equilibrium ensemble of subway passengers:

- (b) For  $L = 10$ ,  $a = 0.4$ , and  $N = 20$ , generate a sample from the equilibrium ideal gas ensemble (that is,  $N$  uniformly distributed random numbers  $z_n$  in the range  $[0, L_{\text{free}}]$ ). Generate a sample  $y_n$  from the subway car ensemble. (Hint: Use  $z_n$  and  $y_{\text{packed}}$ .) Plot a function that is one

if a person is occupying the space, and zero if the space is empty, by generating a curve with four points  $(0, y_n - a/2), (1, y_n - a/2), (1, y_n + a/2), (0, y_n + a/2)$  for each passenger.

We can now collect the distances between each pair of passengers for this snapshot.

- (c) Take your sample  $y_n$ , and make a list of all the (signed) separations  $\Delta y = y_n - y_m$  between pairs of passengers. Plot a histogram of this distribution, using the standard normalization (the number of counts in each bin) and a bin-width of one. Explain why it has an overall triangular shape. Zoom in to  $-2 < \Delta y < 2$ , using a bin width of 0.1. Explain briefly the two striking features you see near  $\Delta y = 0$  – the spike, and the gap.

To explore the correlation function more quantitatively, we now want to take an average of an ensemble of many passenger packings.

- (d) Generate  $n = 100$  ensembles of passengers as in part (b), and collect all the pair separations. Make a histogram of the pair separations, in the range  $-2 < \Delta y < 2$  with bin width of 0.02. Adjust the vertical range of the plot to show the entire first bump (but clipping the central spike). Explain how the bumps in the histogram are related to the distances to neighbors seated in passengers to the right.

At short distances, you can show that your histogram is related<sup>109</sup> to the density-density correlation function  $C(r) = \langle \rho(y)\rho(y+r) \rangle$  evaluated at  $\Delta y = r$ , as follows.

- (e) In an infinite system with  $\rho$  particles per unit length, given there is a particle at  $y$ , write in terms of  $C(r)$  the probability per unit length  $P(y+r|y)$  that there is another particle at  $y+r$ . If we take  $n$  samples from the ensemble with  $N$  passengers each, and compute all the pair separations, what is the number per unit length of separations we expect at  $r$ , in terms of  $C(r)$ ? Show that the expected number of separations in a narrow bin of width  $B$  centered at  $r$  is  $(nNB/\rho)C(r)$ . (Hint: The probability density  $P(A|B)$  of  $A$  (person at  $r$ ) given  $B$  (person at 0) is the probability density  $P(A \& B)$  of both happening divided by the probability density  $P(B)$  of  $B$  happening. Multiply by the number of peo-

<sup>107</sup> In one dimension, a disk of diameter  $a$  is really a rod of length  $a$ , so this is often called a hard-rod gas. It is also known as the Tonks gas.

<sup>108</sup>That is, in the direction of decreasing  $y$ .

<sup>109</sup>  $C(r)$  and our histogram are also closely related to the pair distribution function  $g(r)$  we introduced in Exercise 10.2.

ple and the bin size to get the expected number in the bin.)

Next, we will decompose  $C(r)$  between all pairs into a sum

$$C(r) = \sum_m C_m(r) \quad (10.93)$$

of terms collecting distances to the  $m^{\text{th}}$  neighbor to the right (with negative  $m$  corresponding to the left). We calculate the shapes of these bumps for an infinite system at density  $\rho$ .

Let us start with  $C_1(r)$ . The separations between nearest neighbors  $m = 1$  in the subway is just  $a$  larger than the separations between nearest neighbors in the corresponding one-dimensional ideal gas,  $\Delta y_{\text{nn}} = a + \Delta z_{\text{nn}}$ . The corresponding ideal gas, remember, has density  $\rho_{\text{ideal}} = N/L_{\text{free}}$ .

(f) Argue that the  $m = 1$  nearest-neighbor contribution to the correlation function is an exponential decay

$$C_1(r) = \rho \rho_{\text{ideal}} \exp[-\rho_{\text{ideal}}(r - a)] \Theta(r - a), \quad (10.94)$$

where  $\Theta$  is the Heaviside function (one for positive argument, zero otherwise). (Hint: How is it related to  $C_1^{\text{ideal}}(r)$  for the ideal gas?) Using the relation between  $C(r)$  and your histogram derived in part (d), plot your prediction against the first bump at positive  $\Delta y$  in your histogram.

(g) Calculate  $\tilde{C}_1(k)$  from eqn 10.94, using the conventions of eqn A.9 in the Fourier appendix. (Hint: It is the integral of an exponential over a half-space.)

The gaps between passengers in the infinite subway are just the same as the gaps between particle in the ideal gas – and hence are independent. The second-neighbor distance is thus the sum of two first-neighbor distances independently drawn from the same distribution, so the probability distribution  $P_2(y + r|y)$  is the convolution of two first-neighbor probability distributions. Fourier transforms become useful here (see Exercise 12.11 and eqn A.23);

(h) Using this and your answer to part (e), write the Fourier spectrum of the second neighbor peak  $\tilde{C}_2(k)$ , first in terms of the Fourier spectrum of the first neighbor peak  $\tilde{C}_1(k)$ , and then using your answer from part (g). What is  $\tilde{C}_n(k)$ ? (Hint: Start with calculating  $\tilde{P}_2(k)$ .)

By either doing inverse Fourier transforms back to  $r$  or convolutions, one may derive the correlation function for the subway problem ignoring the finite size of the car.<sup>110</sup> We present the answer here:

$$\begin{aligned} C_m(r) &= \rho \rho_{\text{ideal}}^m \frac{(r - ma)^{m-1}}{(m-1)!} \\ &\quad e^{-\rho_{\text{ideal}}(r - ma)} \Theta(r - ma) \quad m > 0, \\ C_0(r) &= \rho \delta(0) \\ C_m(r) &= C_{(-m)}(-r) \quad m < 0. \end{aligned} \quad (10.95)$$

(i) Plot<sup>111</sup> the theoretical prediction of eqn 10.95 against your measured histogram, for  $-2 < r < 2$ . Explain any discrepancy you find. (Hint: Remember the triangle plot in part (c).)

<sup>110</sup>These are just the  $m^{\text{th}}$  neighbor correlations in the ideal gas, shifted to the right by  $ma$ , the sum of the widths of  $m$  passengers.

<sup>111</sup>In Mathematica, you may need to increase PlotPoints and MaxRecursion to avoid missing features in the curve.

## Chapter 11: Abrupt phase transitions

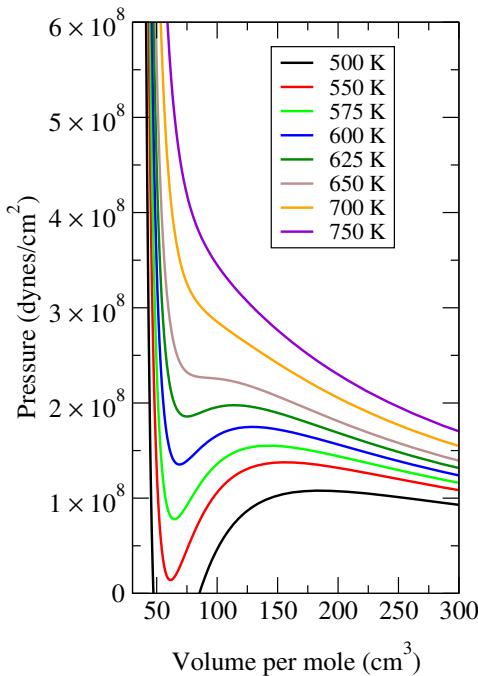
### Exercises

**11.1 Maxwell and van der Waals.**<sup>112</sup> (Chemistry) 

The van der Waals (vdW) equation

$$(P + N^2 a/V^2)(V - Nb) = Nk_B T \quad (11.96)$$

is often applied as an approximate equation of state for real liquids and gases. The term  $V - Nb$  arises from short-range repulsion between molecules (Exercise 3.5); the term  $N^2 a/V^2$  incorporates the leading effects<sup>113</sup> of long-range attraction between molecules.



**Fig. 11.44**  $P$ - $V$  plot: van der Waals. Van der Waals (vdW) approximation (eqn 11.96)

to  $\text{H}_2\text{O}$ , with  $a = 0.55 \text{ J m}^3/\text{mol}^2$  ( $a = 1.52 \times 10^{-35} \text{ erg cm}^3/\text{molecule}$ ), and  $b = 3.04 \times 10^{-5} \text{ m}^3/\text{mol}$  ( $b = 5.05 \times 10^{-23} \text{ cm}^3/\text{molecule}$ ), fit to the critical temperature and pressure for water.

Figure 11.44 shows the pressure versus volume curves for one mole of  $\text{H}_2\text{O}$ , within the vdW model. A real piston of water held at constant temperature would in equilibrium pass through three regimes as it expanded – first a decompressing liquid, then coexisting liquid and vapor at constant pressure (as the liquid evaporates or boils to fill the piston), and then a decompressing gas. The Maxwell construction tells us what the vapor pressure and the two densities for the coexisting liquid and gas is at each temperature. *Get a copy of Fig. 11.44. By hand, roughly implement the Maxwell construction for each curve, and sketch the region in the  $P$ - $V$  plane where liquid and gas can coexist.*

**11.6 Coarsening in the Ising model.**<sup>114</sup> (Computation) 

Coarsening is the process by which phases separate from one another; the surface tension drives tiny fingers and droplets to shrink, leading to a characteristic length scale that grows with time. Start up the Ising model. Run with a fairly large system, demagnetize the system to a random initial state ( $T = \infty$ ), and set  $T = 1.5$  (below  $T_c$ ) and run one sweep at a time. (As the pattern coarsens, you may wish to reduce the graphics refresh rate.) The pattern looks statistically the same at different times, except for a typical *coarsening length* that is growing. How can we define and measure the typical length scale  $L(t)$  of this pattern?

(a) Argue that at zero temperature the total energy above the ground-state energy is propor-

<sup>112</sup>A printable version of Fig. 11.44 can be found at the book Web site [49].

<sup>113</sup>These corrections are to leading orders in the density; they are small for dilute gases.

<sup>114</sup>A link to the software can be found at the book Web site [49].

<sup>115</sup>At finite temperatures, there is a contribution from thermally flipped spins, which should not really count as perimeter for coarsening.

tional to the perimeter separating up-spin and down-spin regions.<sup>115</sup> Argue that the inverse of the perimeter per unit area is a reasonable definition for the length scale of the pattern.

(b) With a random initial state, set temperature and external field to zero. Measure the mean energy  $E(t)$  per spin as a function of time. Plot your estimate for  $L(t)$  in part (a) versus  $t$  on a log-log plot. (You might try plotting a few such curves, to estimate errors.) What power law does it grow with? What power law did we expect?

### 11.10 What is it unstable to? ④

In Fig. 11.3(a), suppose a system held at constant pressure  $P < P_v$  and constant temperature starts in the metastable liquid state. Draw this initial state on a copy of the figure, and draw an arrow to the final state.

### 11.11 Droplet Nucleation in two dimensions. ④

Abrupt phase transitions can also arise in two-dimensional systems. Lipid bilayer membranes, such as the ones surrounding your cells, often undergo a miscibility phase transition. At high temperatures, two types of lipids may dissolve in one another, but at low temperatures they may be immiscible. If the transition is abrupt, it will happen via the same kind of critical droplet nucleation that Section 11.3 analyzed in three dimensions.

Calculate the critical radius  $R_c$  and the free energy barrier  $B$  for the critical droplet in a two-dimensional abrupt phase transition. Let the transition be at temperature  $T_m$ , with latent heat per unit area  $\ell$ , and line tension  $\lambda$  per unit length between the two phases.

It is something of a surprise that living cell membranes appear to be quite near this kind of miscibility transition (five percent above the critical temperature). What is even more interesting is that the transition upon cooling is not abrupt, but almost continuous – in the Ising universality class we shall study in Chapter 12.

### 11.12 Linear stability of a growing interface. (Ecology) ④

A plant disease has infested one side of a corn field in Nebraska, infecting all plants with  $y < u(x)$ , where  $u(x)$  is small (a nearly flat front). The disease propagates at different speeds depending on whether the surrounding plants are already infected. The curvature of the front tells one whether the nearby plants along the  $x$  direction are sick or not.

(a) In regions where  $u''(x) > 0$ , are the surrounding plants more or less healthy than regions where  $u''(x) < 0$ ?

We model the initial time evolution of the front with a simple, linear growth law:

$$\partial u / \partial t = v_0 + w \partial^2 u / \partial x^2. \quad (11.97)$$

Let us assume that the disease spreads more rapidly in regions surrounded by uninfected plants (there are more untapped opportunities for infection).

(b) According to your answer in part (a), is  $w$  positive or negative under this assumption?

To solve this evolution law, it is convenient to work in Fourier space, writing

$$\tilde{u}(k) = (1/L) \int_0^L u(x) \exp(-ikx) dx. \quad (11.98)$$

where the field extends from 0 to  $L$ . You may assume that all derivatives of  $u(x)$  are zero at the edges of the field (that is, ignore the boundary terms).

(c) Find the differential equation for  $\partial \tilde{u}(k_m)/\partial t$ , giving the evolution of the wiggles in the crop growth boundary as it grows, where  $k_m = 2\pi m/L$ . (Hint: You may want to consider  $k_m = 0$  separately. You will need to integrate by parts twice.)

(d) Solve the differential equation for  $\tilde{u}_k(t)$  in terms of the initial  $\tilde{u}_k$  at  $t = 0$ . For what values of  $w$  does an initial irregularity in the front smoothen out with time? For what values does it grow? (Hint: With the assumption that the disease spreads more rapidly when surrounded by healthy plants, the tips of a sinusoidal  $u(x)$  should advance faster than the valleys, leading to the growth of small initial waves.)

### 11.13 Gibbs free energy barrier. (Chemistry) ③

Figure 11.14 shows the chemical potential for a theory (due to van der Waals) for the liquid-gas transition, as a function of density  $\rho = N/V$ , at a pressure and temperature where gas and liquid coexist.

(a) Which well is the gas? Which is the liquid? How can we tell that the two coexist?

This figure is analogous to Fig. 11.2, except that instead of the Helmholtz potential  $A$  we plot the Gibbs free energy per particle  $\mu = G/N$ . In Exercise 11.3, we shall use the chemical potential barrier between the two wells to estimate the surface tension of water.

But what does this plot mean? The caption says the temperature and pressure are fixed. That explains why we use the Gibbs free energy  $G(T, P, N)$ . We know  $G(T, P, N) = N\mu(T, P)$ . (Remember,  $G$  is extensive, and  $T$  and  $P$  are intensive, so it makes sense that it must be linear in the only other extensive variable  $N$ .) But the caption tells us that the graph is for fixed  $T = 373\text{K}$  and  $P = 1.5 \times 10^7 \text{dynes/cm}^2$ . How can  $\mu$  also depend on  $\rho = N/V$ ? *Most of the points on Figure 11.14 represent states that are not in equilibrium with the external world.*

(b) *In equilibrium, show that  $(d/dV(A + PV))|_{T,N} = 0$ , and hence  $d\mu/d\rho = 0$ . Which points on this graph are in local equilibrium? Which local equilibrium is unstable to uniform changes in density?*

We often want to study the free energy of systems that are not in a global equilibrium. In Section 6.7, for example, we studied the free energy of a spatially varying ideal gas by constraining the local density  $\rho(x)$  and minimizing the free energy with respect to other degrees of freedom. Here we want to study the free energy inside a domain wall between the liquid and the gas, interpolating at densities between the two equilibrium states. Usually one would use the Helmholtz free energy  $A(T, V, N)$  to study systems with fixed  $\rho = N/V$ . But here we also have an external pressure, which biases the free energy toward the liquid.

Knowing the equilibrium Helmholtz free energy  $A(T, V, N)$  for a local volume  $\rho = N/V$ , can we calculate the non-equilibrium, constrained Gibbs free energy density  $G(T, P, N, V)$  in terms of the four variables? We can think of this as two subsystems (as in Fig. 6.4), (the gas/liquid mixture and the environment) each in internal equilibrium, but where we have not yet allowed the subsystems to exchange volume.

(c) *Argue that  $G(T, P, N, V) = A(T, V, N) + PV$ , where we do not vary  $V$  to minimize the sum, is the free energy for the local subsystem incorporating the cost of stealing volume from the environment at pressure  $P$ .*

The Helmholtz free energy for the van der Waals

gas is

$$A(T, V, N) = -aN^2/V + Nk_B T \left( \log \left( \frac{N\lambda^3}{V - bN} \right) - 1 \right). \quad (11.99)$$

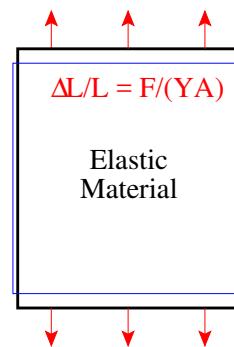
(d) *Verify eqn 11.22 from Exercise 11.3 using eqn 11.99 and the discussion in part (c).*

**11.14 Elastic theory has zero radius of convergence.**<sup>116</sup> (Engineering, Condensed matter) ③

In this exercise, we shall use methods from quantum field theory to tie together two topics which American science and engineering students study in their first year of college: Hooke's law and the convergence of infinite series.

Consider a large steel cube, stretched by a moderate strain  $\epsilon = \Delta L/L$  (Figure 11.45). You may assume  $\epsilon \ll 0.1\%$ , where we can ignore plastic deformation.

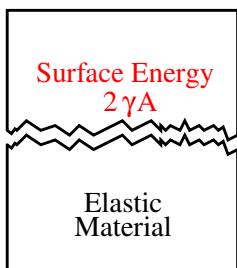
(a) *At non-zero temperature, what is the equilibrium ground state for the cube as  $L \rightarrow \infty$  for fixed  $\epsilon$ ? (Hints: Remember, or show, that the free energy per unit (undeformed) volume of the cube is  $\frac{1}{2}Y\epsilon^2$ . Notice figure 11.46 as an alternative candidate for the ground state.) For steel, with  $Y = 2 \times 10^{11} \text{ N/m}^2$ ,  $\gamma \approx 2.5 \text{ J/m}^2$ ,<sup>117</sup> and density  $\rho = 8000 \text{ kg/m}^3$ , how much can we stretch a beam of length  $L = 10 \text{ m}$  before the equilibrium length is broken in two? How does this compare with the amount the beam stretches under a load equal to its own weight?*



<sup>116</sup>This exercise draws heavily on Alex Buchel's work [11, 12].

<sup>117</sup>This is the energy for a clean, flat [100] surface, a bit more than 1eV/surface atom [50]. The surface left by a real fracture in (ductile) steel will be rugged and severely distorted, with a much higher energy per unit area. This is why steel is much harder to break than glass, which breaks in a brittle fashion with much less energy left in the fracture surfaces.

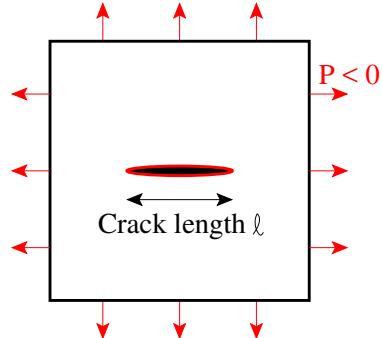
**Fig. 11.45** Stretched block of elastic material, length  $L$  and width  $W$ , elongated vertically by a force  $F$  per unit area  $A$ , with free side boundaries. The block will stretch a distance  $\Delta L/L = F/YA$  vertically and shrink by  $\Delta W/W = \sigma \Delta L/L$  in both horizontal directions, where  $Y$  is Young's modulus and  $\sigma$  is Poisson's ratio, linear elastic constants characteristic of the material. For an isotropic material, the other elastic constants can be written in terms of  $Y$  and  $\sigma$ ; for example, the (linear) bulk modulus  $\kappa_{\text{lin}} = Y/3(1 - 2\sigma)$ .



**Fig. 11.46** Fractured block of elastic material, as in figure 11.45 but broken in two. The free energy here is  $2\gamma A$ , where  $\gamma$  is the free energy per unit area  $A$  of (undeformed) fracture surface.

Why don't bridges fall down? The beams in the bridge are in a *metastable state*. What is the barrier separating the stretched and fractured beam states? Consider a crack in the beam, of length  $\ell$ . Your intuition may tell you that tiny cracks will be harmless, but a long crack will tend to grow at small external stress.

For convenient calculations, we will now switch problems from a stretched steel beam to a taut two-dimensional membrane under an isotropic tension, a negative pressure  $P < 0$ . That is, we are calculating the rate at which a balloon will spontaneously pop due to thermal fluctuations.



**Fig. 11.47** Critical crack of length  $\ell$ , in a two-dimensional material under isotropic tension (negative hydrostatic pressure  $P < 0$ ).

The crack costs a surface free energy  $2\alpha\ell$ , where  $\alpha$  is the free energy per unit length of membrane perimeter. A detailed elastic theory calculation shows that a straight crack of length  $\ell$  will release a (Gibbs free) energy  $\pi P^2(1 - \sigma^2)\ell^2/4Y$ .

(b) *What is the critical length  $\ell_c$  of the crack, at which it will spontaneously grow rather than heal? What is the barrier  $B(P)$  to crack nucleation? Write the net free energy change in terms of  $\ell$ ,  $\ell_c$ , and  $\alpha$ . Graph the net free energy change  $\Delta G$  due to the the crack, versus its length  $\ell$ .*

The point at which the crack is energetically favored to grow is called the *Griffiths threshold*, of considerable importance in the study of brittle fracture.

The predicted fracture nucleation rate  $R(P)$  per unit volume from homogeneous thermal nucleation of cracks is thus

$$R(P) = (\text{prefactors}) \exp(-B(P)/k_B T). \quad (11.100)$$

One should note that thermal nucleation of fracture in an otherwise undamaged, undisordered material will rarely be the dominant failure mode. The surface tension is of order an eV per bond ( $> 10^3 \text{ eV}/\text{\AA}$ ), so thermal cracks of area larger than tens of bond lengths will have insurmountable barriers even at the melting point. Corrosion, flaws, and fatigue will ordinarily lead to structural failures long before thermal nucleation will arise.

*Advanced topic: Elastic theory has zero radius of convergence.*

Many perturbative expansions in physics have zero radius of convergence. The most precisely

calculated quantity in physics is the gyromagnetic ratio of the electron [45]

$$\begin{aligned}(g-2)_{\text{theory}} = & \alpha/(2\pi) - 0.328478965\dots(\alpha/\pi)^2 \\ & + 1.181241456\dots(\alpha/\pi)^3 \\ & - 1.4092(384)(\alpha/\pi)^4 \\ & + 4.396(42) \times 10^{-12} \quad (11.101)\end{aligned}$$

a power series in the fine structure constant  $\alpha = e^2/\hbar c = 1/137.035999\dots$ . (The last term is an  $\alpha$ -independent correction due to other kinds of interactions.) Freeman Dyson gave a wonderful argument that this power-series expansion, and quantum electrodynamics as a whole, has zero radius of convergence. He noticed that the theory is sick (unstable) for any negative  $\alpha$  (corresponding to a pure imaginary electron charge  $e$ ). The series must have zero radius of convergence since any circle in the complex plane about  $\alpha = 0$  includes part of the sick region.

How does Dyson's argument connect to fracture nucleation? Fracture at  $P < 0$  is the kind of instability that Dyson was worried about for quantum electrodynamics for  $\alpha < 0$ . It has implications for the convergence of nonlinear elastic theory.

Hooke's law tells us that a spring stretches a distance proportional to the force applied:  $x - x_0 = F/K$ , defining the spring constant  $1/K = dx/dF$ . Under larger forces, the Hooke's law will have corrections with higher powers of  $F$ . We could define a 'nonlinear spring constant'  $K(F)$  by

$$\frac{1}{K(F)} = \frac{x(F) - x(0)}{F} = k_0 + k_1 F + \dots \quad (11.102)$$

Instead of a spring constant, we'll calculate a nonlinear version of the bulk modulus  $\kappa_{\text{nl}}(P)$  giving the pressure needed for a given fractional change in volume,  $\Delta P = -\kappa \Delta V/V$ . The linear isothermal bulk modulus<sup>118</sup> is given by  $1/\kappa_{\text{lin}} = -(1/V)(\partial V/\partial P)|_T$ ; we can define a nonlinear

generalization by

$$\begin{aligned}\frac{1}{\kappa_{\text{nl}}(P)} &= -\frac{1}{V(0)} \frac{V(P) - V(0)}{P} \\ &= c_0 + c_1 P + c_2 P^2 + \dots + c_N P^N + \dots \quad (11.103)\end{aligned}$$

This series can be viewed as higher and higher-order terms in a nonlinear elastic theory.

(c) *Given your argument in part (a) about the stability of materials under tension, would Dyson argue that the series in eqn 11.103 has a zero or a non-zero radius of convergence?*

In Exercise 1.5 we saw the same argument holds for Stirling's formula for  $N!$ , when extended to a series in  $1/N$ ; any circle in the complex  $1/N$  plane contains points  $1/(-N)$  from large negative integers, where we can show that  $(-N)! = \infty$ . These series are *asymptotic expansions*. Convergent expansions  $\sum c_n x^n$  converge for fixed  $x$  as  $n \rightarrow \infty$ ; asymptotic expansions need only converge to order  $O(x^{n+1})$  as  $x \rightarrow 0$  for fixed  $n$ . Hooke's law, Stirling's formula, and quantum electrodynamics are examples of how important, powerful, and useful asymptotic expansions can be.

Buchel [11, 12], using a clever trick from field theory [59, Chapter 40], was able to calculate the large-order terms in elastic theory, essentially by doing a Kramers–Krönig transformation on your formula for the decay rate (eqn 11.100) in part (b). His logic works as follows.

- The Gibbs free energy density  $\mathcal{G}$  of the metastable state is complex for negative  $P$ . The real and imaginary parts of the free energy for complex  $P$  form an analytic function (at least in our calculation) except along the negative  $P$  axis, where there is a branch cut.
- Our isothermal bulk modulus for  $P > 0$  can be computed in terms of  $\mathcal{G} = G/V(0)$ . Since  $dG = -S dT + V dP + \mu dN$ ,  $V(P) = (\partial G / \partial P)|_T$

<sup>118</sup>Warning: For many purposes (e.g. sound waves) one must use the *adiabatic* elastic constant  $1/\kappa = -(1/V)(\partial V/\partial P)|_S$ . For most solids and liquids these are nearly the same.

<sup>119</sup>Notice that this is not the (more standard) pressure-dependent linear bulk modulus,  $\kappa_{\text{lin}}(P)$  which is given by  $1/\kappa_{\text{lin}}(P) = -(1/V)(\partial V/\partial P)|_T = -(1/V)(\partial^2 \mathcal{G} / \partial P^2)|_T$ . This would also have a Taylor series in  $P$  with zero radius of convergence at  $P = 0$ , but it has a different interpretation;  $\kappa_{\text{nl}}(P)$  is the nonlinear response at  $P = 0$ , while  $\kappa_{\text{lin}}(P)$  is the pressure-dependent linear response.

and hence<sup>119</sup>

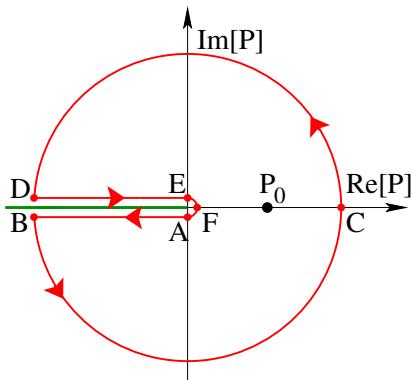
$$\begin{aligned}\frac{1}{\kappa_{nl}(P)} &= -\frac{1}{V(0)} \frac{(\partial G/\partial P)|_T - V(0)}{P} \\ &= -\frac{1}{P} \left( \frac{\partial G}{\partial P} \Big|_T - 1 \right).\end{aligned}\quad (11.104)$$

- (d) Write the coefficients  $c_n$  of eqn 11.103 in terms of the coefficients  $g_m$  in the nonlinear expansion

$$\mathcal{G}(P) = \sum g_m P^m.\quad (11.105)$$

- The decay rate  $R(P)$  per unit volume is proportional to the imaginary part of the free energy  $\text{Im}[\mathcal{G}(P)]$ , just as the decay rate  $\Gamma$  for a quantum state is related to the imaginary part  $i\hbar\Gamma$  of the energy of the resonance. More specifically, for  $P < 0$  the imaginary part of the free energy jumps as one crosses the real axis:

$$\text{Im}[\mathcal{G}(P \pm i\epsilon)] = \pm(\text{prefactors})R(P).\quad (11.106)$$



**Fig. 11.48 Contour integral in complex pressure.** The free energy density  $\mathcal{G}$  of the elastic membrane is analytic in the complex  $P$  plane except along the negative  $P$  axis. This allows one to evaluate  $\mathcal{G}$  at positive pressure  $P_0$  (where the membrane is stable and  $\mathcal{G}$  is real) with a contour integral as shown.

- Buchel then used Cauchy's formula to evaluate the real part of  $\mathcal{G}$  in terms of the imaginary part, and hence the decay rate  $R$  per unit volume:

$$\begin{aligned}\mathcal{G}(P_0) &= \frac{1}{2\pi i} \oint_{ABCDEF} \frac{\mathcal{G}(P)}{P - P_0} dP \\ &= \frac{1}{2\pi i} \int_B^0 \frac{\mathcal{G}(P + i\epsilon) - \mathcal{G}(P - i\epsilon)}{P - P_0} dP \\ &\quad + \int_{EFA} + \int_{BCD} \\ &= \frac{1}{\pi} \int_B^0 \frac{\text{Im}[\mathcal{G}(P + i\epsilon)]}{P - P_0} dP \\ &\quad + (\text{unimportant})\end{aligned}\quad (11.107)$$

where the integral over the small semicircle vanishes as its radius  $\epsilon \rightarrow 0$  and the integral over the large circle is convergent and hence unimportant to high-order terms in perturbation theory.

The decay rate (eqn 11.100) for  $P < 0$  should be of the form

$$R(P) \propto (\text{prefactors}) \exp(-D/P^2),\quad (11.108)$$

where  $D$  is some constant characteristic of the material. (You may use this to check your answer to part (b).)

- Using eqns. 11.106, 11.107, and 11.108, and assuming the prefactors combine into a constant  $A$ , write the free energy for  $P_0 > 0$  as an integral involving the decay rate over  $-\infty < P < 0$ . Expanding  $1/(P - P_0)$  in a Taylor series in powers of  $P_0$ , and assuming one may exchange sums and integration, find and evaluate the integral for  $g_m$  in terms of  $D$  and  $m$ . Calculate from  $g_m$  the coefficients  $c_n$ , and then use the ratio test to calculate the radius of convergence of the expansion for  $1/\kappa_{nl}(P)$ , eqn 11.103. (Hints: Use a table of integrals, a computer algebra package, or change variable  $P = -\sqrt{D/t}$  to make your integral into the  $\Gamma$  function,

$$\Gamma(z) = (z - 1)! = \int_0^\infty t^{z-1} \exp(-t) dt.\quad (11.109)$$

If you wish, you may use the ratio test on every second term, so the radius of convergence is the value  $\lim_{n \rightarrow \infty} \sqrt{|c_n/c_{n+2}|}$ .

(Why is this approximate calculation trustworthy? Your formula for the decay rate is valid only up to prefactors that may depend on the pressure; this dependence (some power of  $P$ ) won't change the asymptotic ratio of terms  $c_n$ . Your formula for the decay rate is an approximation, but one which becomes better and better for smaller values of  $P$ ; the integral for the high-order terms  $g_m$  (and hence  $c_n$ ) is concentrated at small  $P$ , so your approximation is asymptotically correct for the high order terms.)

Thus the decay rate of the metastable state can be used to calculate the high-order terms in perturbation theory in the stable phase! This is a general phenomena in theories of metastable states, both in statistical mechanics and in quantum physics.

## Chapter 12: Continuous phase transitions

## Exercises

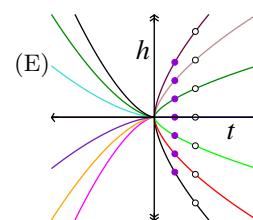
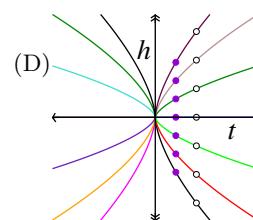
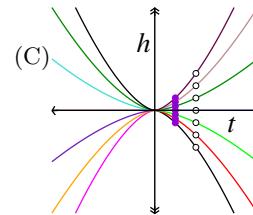
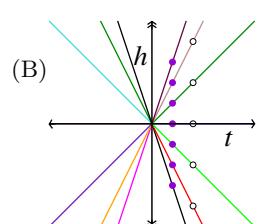
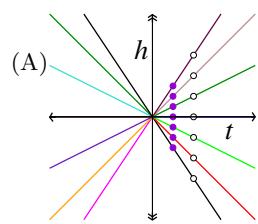
12.7 Renormalization-group trajectories. 

This exercise provides an early introduction to how we will derive power laws and universal scaling functions in Section 12.2 from universality and coarse-graining.

An Ising model near its critical temperature  $T_c$  is described by two variables: the distance to the critical temperature  $t = (T - T_c)/T_c$ , and the external field  $h = H/J$ . Under coarse-graining, changing lengths to  $x' = (1 - \epsilon)x$ , the system is observed to be similar to itself at a shifted temperature  $t' = (1 + a\epsilon)t$  and a shifted external field  $h' = (1 + b\epsilon)h$ , with  $\epsilon$  infinitesimal and  $a > b > 0$  (so there are two relevant eigendirections, with the temperature more strongly relevant than the external field). Assume  $a > b > 0$ .

The curves shown below connect points that are similar up to some rescaling factor.

(a) Which diagram below has curves consistent with this flow, for  $a > b > 0$ ? Is the flow under coarse graining inward or outward from the origin? (No math should be required. Hint: After coarse-graining, how does  $h/t$  change?)



The solid dots are at temperature  $t_0$ ; the open circles are at temperature  $t = 2t_0$ .

(b) In terms of  $\epsilon$  and  $a$ , by what factor must  $x$  be rescaled by to relate the systems at  $t_0$  and  $2t_0$ ? (Algebraic tricks: Use  $(1 + \delta) \approx \exp(\delta)$  everywhere. If you rescale multiple times until  $\exp(na\epsilon) = 2$ , you can solve for  $(1 - \epsilon)^n \approx \exp(-n\epsilon)$  without solving for  $n$ .) If one of the solid dots in the appropriate figure from part (a) is at  $(t_0, h_0)$ , what is the field  $\hat{h}$  for the corresponding open circle, in terms of  $a$ ,  $b$ ,  $\epsilon$ , and the original coordinates? (You may use the relation between  $\hat{h}$  and  $h_0$  to check your answer for part (a).)

The magnetization  $M(t, h)$  is observed to rescale under this same coarse-graining operation to

$M' = (1 + ce) M$ , so  $M((1 + a\epsilon)t, (1 + b\epsilon)h) = (1 + ce)M(t, h)$ .

(c) Suppose  $M(t, h)$  is known at  $(t_0, h_0)$ , one of the solid dots. Give a formula for  $M(2t_0, \hat{h})$  at the corresponding open circle, in terms of  $M(t_0, h_0)$ , the original coordinates,  $a$ ,  $b$ ,  $c$ , and  $\epsilon$ . (Hint: Again, rescale  $n$  times.) Substitute your formula for  $\hat{h}$  into the formula, and solve for  $M(t_0, h_0)$ .

You have now basically derived the key result of the renormalization group; the magnetization curve at  $t_0$  can be found from the magnetization curve at  $2t_0$ . In Section 12.2, we shall coarse-grain not to  $t = 2t_0$ , but to  $t = 1$ . We shall see that the magnetization everywhere can be predicted from the magnetization where the invariant curve crosses  $t = 1$ .

(d) There was nothing about the factor of two in our shift in temperature that was special. Substitute  $1/t_0$  for 2 in your formula from part (c). Show that  $M(t, h) = t^\beta \mathcal{M}(h/t^{\beta\delta})$  (the standard scaling form for the magnetization in the Ising model). What are  $\beta$  and  $\delta$  in terms of  $a$ ,  $b$ , and  $c$ ? How is  $\mathcal{M}$  related to  $M(t, h)$  where the curve crosses  $t = 1$ ?

Note that we have succeeded in writing  $M(t, h)$  in the two-dimensional plane in terms of its value  $M(1, h)$  along the line  $t = 1$ . A property depending on  $n$  variables near a critical point has a singular part that can be written as a power law in one variable times a scaling function  $\mathcal{M}$  of  $n - 1$  variables. What is more, the power law and the scaling function is universal – shared between all systems that can flow to the same renormalization-group fixed point.

### 12.5 Mean-field theory. (Condensed matter) ①

In Chapter 11 and Exercise 9.5, we make reference to mean-field theories, a term which is often loosely used for any theory which absorbs the fluctuations of the order parameter field into a single degree of freedom in an effective free energy. The original mean-field theory actually used the mean value of the field on neighboring sites to approximate their effects.

In the Ising model on a square lattice, this amounts to assuming each spin  $s_j = \pm 1$  has four neighbors which are magnetized with the average magnetization  $m = \langle s_j \rangle$ , leading to a one-spin mean-field Hamiltonian

$$\mathcal{H} = -4Jms_j. \quad (12.110)$$

(a) At temperature  $k_B T$ , what is the value for  $\langle s_j \rangle$  in eqn 12.110, given  $m$ ? At what temperature  $T_c$  is the phase transition, in mean field theory? (Hint: At what temperature is a non-zero  $m = \langle s \rangle$  self-consistent?) Argue as in Exercise 12.4 part (c) that  $m \propto (T_c - T)^\beta$  near  $T_c$ . Is this value for the critical exponent  $\beta$  correct for the Ising model in either two dimensions ( $\beta = 1/8$ ) or three dimensions ( $\beta \approx 0.32641\dots$ )?

(b) Show that the mean-field solution you found in part (a) is the minimum in an effective temperature-dependent free energy

$$V(m) = 2Jm^2 - k_B T \log(\cosh(4Jm/k_B T)). \quad (12.111)$$

(The first term comes from the mean-field approximation of the energy per bond  $J\langle s_i s_j \rangle \approx J\langle s_i \rangle \langle s_j \rangle = m^2$ , times the number of bonds per spin. The second term is the free energy of a spin in the mean field  $4Jm$  due to the four neighbors. See Exercise 12.24 for a derivation.) On a single graph, plot  $V(m)$  for  $J = 1$  and  $1/(k_B T) = 0.1, 0.25$ , and  $0.5$ , for  $-2 < m < 2$ , showing the continuous phase transition. Compare with Fig. 9.22.

(c) What would the mean-field Hamiltonian be for the square-lattice Ising model in an external field  $H$ ? At  $T = T_c$  from part (a), what is the magnetization  $m(H) \propto H^\delta$  near  $T = T_c$ ,  $H = 0$ ? How does the mean-field critical exponent  $\delta$  compare with the known values  $\delta_{2d} = 15$  and  $\delta_{3d} \approx 4.7898$  for the Ising model in two and three dimensions? Is the agreement improving for higher dimensions?

(d) Show that the mean-field magnetization is given by the minima in

$$V(m, H) = 2Jm^2 - k_B T \log(\cosh((4Jm + H)/k_B T)) \quad (12.112)$$

(derived in Exercise 12.23). On a single graph, plot  $V(m, H)$  for  $\beta = 0.5$  and  $H = 0, 0.5, 1.0$ , and  $1.5$ , showing metastability and an abrupt transition. At what value of  $H$  does the metastable state become completely unstable? Compare with Fig. 11.2(a).

(e) Changing variables to  $t = T - T_c$ , take a Taylor series of  $V(m, H)$  about  $m = t = H = 0$ , up to fourth order in  $m$  and linear order in  $t$  and  $H$ .

*Compare with the Landau theory free energy for the Ising model, eqn 9.19. What is  $\mu(T)$ , the coefficient of  $m^2$ ? What is  $f_0$  and  $g$  at the critical point  $t = H = 0$ ?*

**12.14 Crackling noises.<sup>120</sup> *i***

Listen to the sound file for the earthquakes in 1995 (shown graphically in Fig. 12.3). Each click represents an earthquake, with the sound energy proportional to the energy release and with the year compressed into a couple of seconds.

(a) *Roughly how many clicks can you hear? Consider the histogram in Fig. 12.3(b), down to what magnitude do you estimate you can perceive?*

Now listen to paper crumpling, Rice Krispies<sup>©</sup>, fire, and to our model of magnetic noise. All of these sound files share common features – they are composed of brief pulses or avalanches with a broad range of sizes. Just as for earthquakes, the largest avalanches are rare, and the smaller avalanches are more common, with avalanches of size  $S$  happening with probability proportional to  $S^{-\tau}$ .

These sound files differ, however, in their values for the exponent  $\tau$ . For earthquakes,  $\tau$  can vary from one fault type to another; our data set (Fig. 12.3(b)) shows  $\tau$  between 1.5 and 1.8 (see Exercise 12.16). Milk invading bubbles in puffed rice is likely a special case of the well-studied problem of fluid invasion into porous media (also known as *imbibition*), where in three dimensions estimates of  $\tau$  range between around 1.3 and 1.5. Paper crumpling and fire are not well understood, but we have an excellent understanding of our model of magnetic noise,<sup>121</sup> where  $\tau \sim 2$ .

Changing  $\tau$  should change the sound of the crackling noise – changing the predominance of the loudest crackles over the rest.

(b) *Should larger  $\tau$  correspond to fewer very large avalanches, or more? Does that correspond roughly to your perception of the difference between the sounds from earthquakes and magnets? What about Rice Krispies<sup>©</sup>?*

**12.15 Hearing chaos.<sup>122</sup> (Dynamical systems) *i***

This exercise listens to a system as it transitions from regular to chaotic behavior. Chaos scrambles our knowledge of the initial state of a system (Exercise 5.9), allowing entropy to increase and mediating equilibration. Also, the onset of chaos (Exercises 12.22 and 12.9) is studied using the same renormalization-group methods we use to study continuous phase transitions (Chapter 12). The English word chaos makes us think of billiard balls, bumble bees, bumper cars, or turbulence – the realm of statistical mechanics. The mathematical term chaos in dynamical systems often reflects a more subtle irregularity of motion. Think of an unbalanced dryer, or a faucet that drips at irregular intervals.

One of the characteristics of chaotic motion is a continuous spectrum of frequencies. (Or, perhaps more precisely, systems that are non-chaotic but non-stationary usually have a spectrum that consists of sums and differences of a few frequencies.) Our ears are spectacularly sensitive to frequencies. Listen to the audio files from the book Web site [49]. Figure 12.17 shows the states visited at long times in the logistic map  $f_\mu(x)$ . The audio moves the loudspeaker back and forth in time as  $x(t + \Delta) = f_\mu(x(t))$ ; thus starting at  $\mu_1$  one hears a tone with frequency  $\omega = 2\pi/\Delta$  that grows louder as one approaches  $\mu_2$ . The audio continues to raise  $\mu$ , slowly crossing through a sequence of period-doubling bifurcations and then into the chaotic region.

(a) *If the period of a sound wave doubles, how much does the pitch change? (One note? A perfect fourth? A perfect fifth? An octave?)*

Listen to the two audio files. (One starts closer to the onset of chaos.)

(b) *Does the pitch start sounding like a pure tone? Can you hear the first period doubling? Does the new pitch agree with your answer to part (a)?*

Shortly after the first period doubling, you should begin to hear a chaotic signal.

<sup>120</sup>The computer exercises link on the book Web site [49] provides links to audio files of various types of crackling noise for this exercise.

<sup>121</sup>Barkhausen noise in magnets is usually studied not using sound emission, but by measuring jumps in the magnetization as the external field is increased. Also, real magnets are not described by our model, but are in the same universality class as fluid invasion.

<sup>122</sup>The book Web site [49] provides links to the audio files for this exercise; thanks to Erich Mueller for creating them.

(c) *Describe the new sound. Is it musical, or noisy?*

Figure 12.17 shows the states visited at long times by the logistic map, as the parameter  $\mu$  is varied along the horizontal axis. It does not show the dynamics as the system hops between one point and another along the curves, but you can reconstruct that. For example, between  $\mu_1$  and  $\mu_2$  the system hops between the upper and lower branches of the curve, giving a tone with period equal to one iteration of the map. Between  $\mu_2$  and  $\mu_3$  you heard a second tone one octave lower. You probably did not hear the next period doubling, but after  $\mu_\infty$  you began to hear noise.

(d) *Note the ‘windows’ of non-chaotic motion deep inside the chaotic region. In the audio file, do you hear periods of tonal sounds interspersed in the noise?*

### 12.16 The Gutenberg Richter law. (Engineering, Geophysics) ③

Power laws often arise at continuous transitions in non-equilibrium extended systems, particularly when disorder is important. We don’t yet have a complete understanding of earthquakes, but they seem clearly related to the transition between pinned and sliding faults as the tectonic plates slide over one another.

The size  $S$  of an earthquake (the energy radiated, shown in the upper axis of Figure 12.3b) is traditionally measured in terms of a ‘magnitude’  $M \propto \log S$  (lower axis). The Gutenberg-Richter law tells us that the number of earthquakes of magnitude  $M \propto \log S$  goes down as their size  $S$  increases. Figure 12.3b shows that the number of avalanches of magnitude between  $M$  and  $M + 1$  is proportional to  $S^{-B}$  with  $B \approx 2/3$ . However, it is traditional in the physics community to consider the probability density  $P(S)$  of having an avalanche of size  $S$ .

If  $P(S) \sim S^{-\tau}$ , give a formula for  $\tau$  in terms of the Gutenberg-Richter exponent  $B$ . (Hint: The bins in the histogram have different ranges of size  $S$ . Use  $P(M) dM = P(S) dS$ .)

### 12.17 Random walks and critical exponents. ④

Figure 2.2 shows that a random walk appears statistically self-similar. Random walks naturally have self-similarity and power laws: they are said to exhibit *generic scale invariance*.

(a) Argue that the fractal dimension of the random walk is two, independent of dimension (see note 6, p. 27.) (Hint: How does the mass of

the ink  $T$  needed to draw the walk scale with the distance  $d$  between endpoints?)

DNA has a *persistence length* of 50nm; one can model it roughly as a random walk with a step size of the persistence length. The diameter of a DNA molecule is 2nm, and will not overlap with itself during the random walk. Will the thickness change the power law  $d \sim T^\nu$  relating distance  $d$  versus length  $T$ , and hence the fractal dimension  $1/\nu$ ? More generally, are *self-avoiding* random walks a different universality class?

(b) Show that  $d_f = \nu = 1$  for a one-dimensional self-avoiding random walk. (Hint: There are only two allowed walks.)

We can think about self-avoidance in general dimensions by asking whether a typical random walk will intersect itself. In the spirit of the renormalization group, let us divide a long polymer of  $N$  persistence lengths into  $M$  segments, with  $1 \ll M \ll N$ , each of which forms a fuzzy blob of dimension  $d_f$ . Suppose two of the blobs from distant parts of the polymer overlap. Each may have lots of empty space between the polymer segments, so they may or may not intersect even if they overlap. If they typically intersect, intersections would happen on all scales, and the critical exponents would likely change. If overlapping blobs rarely intersect, one would reasonably expect that the polymer would act like an ordinary random walk on long length scales.

(c) What is the likely dimension of the intersection of two  $D$ -dimensional smooth surfaces in the same vicinity of  $\mathbb{R}^d$ ? (Hint: We can define a smooth  $D$  dimensional set in  $d$  dimensions as the solution of  $d-D$  constraint equations. Check your answer for two curves on the plane (which typically intersect in isolated points), and two curves in three dimensions (which typically will miss one another).) Argue by analogy that self-avoiding random walks in dimensions two and three will likely have a different fractal dimension than ordinary random walks (see note 9, p. 29 and Exercise 2.10), but in dimensions above four they could be in the same universality class.

Indeed, it turns out that self-avoiding random walks have  $\nu_{2D} = \frac{3}{4}$  and  $d_f = \frac{4}{3}$  in two dimensions,  $\nu_{3D} \approx 0.588$  and  $d_f = 1/\nu_{3D}$  in three dimensions, and logarithmic corrections in four dimensions; above four dimensions (the ‘upper critical dimension’) they obey the ‘mean-field’ behavior of ordinary random walks at long length scales.

This pattern (varying critical exponents shifting to mean-field above an upper critical dimension) is common to most statistical mechanical models. Also, we should note that the self-avoiding random walk can be viewed as the limit of an  $n$ -component spin model as  $n \rightarrow 0$ ; the Ising model ( $n = 1$ ) also obeys mean-field theory above four dimensions.

**12.18 Hysteresis and Barkhausen Noise.** (Complexity) <sup>①</sup>

Hysteresis is associated with abrupt phase transitions. Supercooling and superheating are examples (as temperature crosses  $T_c$ ). Magnetic recording, the classic place where hysteresis is studied, is also governed by an abrupt phase transition – here the hysteresis in the magnetization, as the external field  $H$  is increased (to magnetize the system) and then decreased again to zero. Magnetic hysteresis is characterized by crackling (Barkhausen) electromagnetic noise. This noise is due to avalanches of spins flipping as the magnetic interfaces jerkily are pushed past defects by the external field (much like earthquake faults jerkily responding to the stresses from the tectonic plates). It is interesting that when dirt is added to this abrupt magnetic transition, it exhibits the power-law scaling characteristic of continuous transitions.

Our model of magnetic hysteresis (unlike the experiments) has avalanches and scaling only at a special critical value of the disorder  $R_c \sim 2.16$  (Figure 12.14). The integrated probability distribution  $D(S, R)$  has a power law  $D(S, R_c) \propto S^{-\bar{\tau}}$  at the critical point (where  $\bar{\tau} = \tau + \sigma\beta\delta$  for our model) but away from the critical point takes the *scaling form*

$$D(S, R) \propto S^{-\bar{\tau}} \mathcal{D}(S^\sigma (R - R_c)). \quad (12.113)$$

Note from eqn 12.113 that at the critical disorder  $R = R_c$  the distribution of avalanche sizes is a power law  $D(S, R_c) = S^{-\bar{\tau}}$ . The scaling form controls how this power law is altered as  $R$  moves away from the critical point. From Figure 12.14 we see that the main effect of moving above  $R_c$  is to cut off the largest avalanches at a typical largest size  $S_{\max}(R)$ , and another important effect is to form a ‘bulge’ of extra avalanches just below the cut-off.

Using the scaling form from eqn 12.113, with what exponent does  $S_{\max}$  diverge as  $r = (R_c -$

$R) \rightarrow 0$ ? (Hint: At what size  $S$  is  $D(S, R)$ , say, one millionth of  $S^{-\bar{\tau}}$ ? Given  $\bar{\tau} \approx 2.03$ , how does the mean  $\langle S \rangle$  and the mean-square  $\langle S^2 \rangle$  avalanche size scale with  $r = (R_c - R)$ ? (Hint: Your integral for the moments should have a lower cutoff  $S_0$ , the smallest possible avalanche, but no upper cutoff, since that is provided by the scaling function  $\mathcal{D}$ . Assume  $\mathcal{D}(0) > 0$ . Change variables to  $Y = S^\sigma r$ . Which moments diverge?)

**12.19 Biggest of bunch: Gumbel.** <sup>123</sup> (Mathematics, Statistics, Computation, Engineering) <sup>③</sup>

Much of statistical mechanics focuses on the average behavior in an ensemble, or the mean square fluctuations about that average. In many cases, however, we are far more interested in the extremes of a distribution.

Engineers planning dike systems are interested in the highest flood level likely in the next hundred years. Let the high water mark in year  $j$  be  $H_j$ . Ignoring long-term weather changes (like global warming) and year-to-year correlations, let us assume that each  $H_j$  is an independent and identically distributed (IID) random variable with probability density  $\rho_1(H_j)$ . The *cumulative distribution function* (cdf) is the probability that a random variable is less than a given threshold. Let the cdf for a single year be  $F_1(H) = P(H' < H) = \int^H \rho_1(H') dH'$ .

(a) Write the probability  $F_N(H)$  that the highest flood level (largest of the high-water marks) in the next  $N = 1000$  years will be less than  $H$ , in terms of the probability  $F_1(H)$  that the high-water mark in a single year is less than  $H$ .

The distribution of the largest or smallest of  $N$  random numbers is described by *extreme value statistics* [51]. Extreme value statistics is a valuable tool in engineering (reliability, disaster preparation), in the insurance business, and recently in bioinformatics (where it is used to determine whether the best alignments of an unknown gene to known genes in other organisms are significantly better than that one would generate randomly).

(b) Suppose that  $\rho_1(H) = \exp(-H/H_0)/H_0$  decays as a simple exponential ( $H > 0$ ). Using the formula

$$(1 - A) \approx \exp(-A) \text{ small } A \quad (12.114)$$

show that the cumulative distribution function

<sup>123</sup>Computational hints can be found at the book Web site [49].

$F_N$  for the highest flood after  $N$  years is

$$F_N(H) \approx \exp \left[ -\exp \left( \frac{\mu - H}{\beta} \right) \right]. \quad (12.115)$$

for large  $H$ . (Why is the probability  $F_N(H)$  small when  $H$  is not large, at large  $N$ ?) What are  $\mu$  and  $\beta$  for this case?

The constants  $\beta$  and  $\mu$  just shift the scale and zero of the ruler used to measure the variable of interest. Thus, using a suitable ruler, the largest of many events is given by a Gumbel distribution

$$\begin{aligned} F(x) &= \exp(-\exp(-x)) \\ \rho(x) &= \partial F / \partial x = \exp(-(x + \exp(-x))). \end{aligned} \quad (12.116)$$

How much does the probability distribution for the largest of  $N$  IID random variables depend on the probability density of the individual random variables? Surprisingly little! It turns out that the largest of  $N$  Gaussian random variables also has the same Gumbel form that we found for exponentials. Indeed, any probability distribution that has unbounded possible values for the variable, but that decays faster than any power law, will have extreme value statistics governed by the Gumbel distribution [32, section 8.3]. In particular, suppose

$$F_1(H) \approx 1 - A \exp(-BH^\delta) \quad (12.117)$$

as  $H \rightarrow \infty$  for some positive constants  $A$ ,  $B$ , and  $\delta$ . It is in the region near  $H^*[N]$ , defined by  $F_1(H^*[N]) = 1 - 1/N$ , that  $F_N$  varies in an interesting range (because of eqn 12.114).

(c) Show that the extreme value statistics  $F_N(H)$  for this distribution is of the Gumbel form (eqn 12.115) with  $\mu = H^*[N]$  and  $\beta = 1/(B\delta H^*[N]^{\delta-1})$ . (Hint: Taylor expand  $F_1(H)$  at  $H^*$  to first order.)

The Gumbel distribution is *universal*. It describes the extreme values for any unbounded distribution whose tails decay faster than a power law.<sup>124</sup> (This is quite analogous to the central limit theorem, which shows that the normal or Gaussian distribution is the universal form for sums of large numbers of IID random variables, so long as the individual random variables have non-infinite variance.)

The Gaussian or standard normal distribution  $\rho_1(H) = (1/\sqrt{2\pi}) \exp(-H^2/2)$ , for example, has a cumulative distribution  $F_1(H) = (1/2)(1 + \text{erf}(H/\sqrt{2}))$  which at large  $H$  has asymptotic form  $F_1(H) \sim 1 - (1/\sqrt{2\pi}H) \exp(-H^2/2)$ . This is of the general form of eqn 12.117 with  $B = 1/2$  and  $\delta = 2$ , except that  $A$  is a slowly varying function of  $H$ . This slow variation does not change the asymptotics.

(d) Generate  $M = 10000$  lists of  $N = 1000$  random numbers distributed with this Gaussian probability distribution. Plot a normalized histogram of the largest entries in each list. Plot also the predicted form  $\rho_N(H) = dF_N/dH$  from part (c). (Hint:  $H^*(N) \approx 3.09023$  for  $N = 1000$ ; check this if it is convenient.)

Other types of distributions can have extreme value statistics in different universality classes (see Exercise 12.20). Distributions with power-law tails (like the distributions of earthquakes and avalanches described in Chapter 12) have extreme value statistics described by *Fréchet distributions*. Distributions that have a strict upper or lower bound<sup>125</sup> have extreme value distributions that are described by Weibull statistics (see Exercise 1.9).

### 12.20 Extreme value statistics: Gumbel, Weibull, and Fréchet.

(Mathematics, Statistics, Engineering) ③

*Extreme value statistics* is the study of the maximum or minimum of a collection of random numbers. It has obvious applications in the insurance business (where one wants to know the biggest storm or flood in the next decades, see Exercise 12.19) and in the failure of large systems (where the weakest component or flaw leads to failure, see Exercise 1.9). Recently extreme value statistics has become of significant importance in bioinformatics. (In guessing the function of a new gene, one often searches entire genomes for good matches (or *alignments*) to the gene, presuming that the two genes are evolutionary descendants of a common ancestor and hence will have similar functions. One must understand extreme value statistics to evaluate whether the best matches are likely to arise simply at random.)

<sup>124</sup>The Gumbel distribution can also describe extreme values for a bounded distribution, if the probability density at the boundary goes to zero faster than a power law [51, section 8.2].

<sup>125</sup>More specifically, bounded distributions that have power-law asymptotics have Weibull statistics; see note 124 and Exercise 1.9, part (d).

The limiting distribution of the biggest or smallest of  $N$  random numbers as  $N \rightarrow \infty$  takes one of three *universal forms*, depending on the probability distribution of the individual random numbers. In this exercise we understand these forms as fixed points in a renormalization group.

Given a probability distribution  $\rho_1(x)$ , we define the *cumulative distribution function* (CDF) as  $F_1(x) = \int_{-\infty}^x \rho(x') dx'$ . Let us define  $\rho_N(x)$  to be the probability density that, out of  $N$  random variables, the largest is equal to  $x$ . Let  $F_N(x)$  to be the corresponding CDF.

(a) Write a formula for  $F_{2N}(x)$  in terms of  $F_N(x)$ . If  $F_N(x) = \exp(-g_N(x))$ , show that  $g_{2N}(x) = 2g_N(x)$ .

Our renormalization group coarse-graining operation will remove half of the variables, throwing away the smaller of every pair, and returning the resulting new probability distribution. In terms of the function  $g(x) = -\log \int_{-\infty}^x \rho(x') dx'$ , it therefore will return a rescaled version of the  $2g(x)$ . This rescaling is necessary because, as the sample size  $N$  increases, the maximum will drift upward—only the form of the probability distribution stays the same, the mean and width can change. Our renormalization-group coarse-graining operation thus maps function space into itself, and is of the form

$$T[g](x) = 2g(ax + b). \quad (12.118)$$

(This renormalization group is the same as that we use for sums of random variables in Exercise 12.11 where  $g(k)$  is the logarithm of the Fourier transform of the probability density.) There are three distinct types of fixed-point distributions for this renormalization group transformation, which (with an appropriate linear rescaling of the variable  $x$ ) describe most extreme value statistics. The Gumbel distribution (Exercise 12.19) is of the form

$$\begin{aligned} F_{\text{gumbel}}(x) &= \exp(-\exp(-x)) \\ \rho_{\text{gumbel}}(x) &= \exp(-x) \exp(-\exp(-x)). \\ g_{\text{gumbel}}(x) &= \exp(-x) \end{aligned}$$

The Weibull distribution (Exercise 1.9) is of the form

$$\begin{aligned} F_{\text{weibull}}(x) &= \begin{cases} \exp(-(-x)^\alpha) & x < 0 \\ 1 & x \geq 0 \end{cases} \\ g_{\text{weibull}}(x) &= \begin{cases} (-x)^\alpha & x < 0 \\ 0 & x \geq 0, \end{cases} \end{aligned} \quad (12.119)$$

and the Fréchet distribution is of the form

$$\begin{aligned} F_{\text{fréchet}}(x) &= \begin{cases} 0 & x \leq 0 \\ \exp(-x^{-\alpha}) & x > 0 \end{cases} \\ g_{\text{fréchet}}(x) &= \begin{cases} \infty & x < 0 \\ x^{-\alpha} & x \geq 0, \end{cases} \end{aligned} \quad (12.120)$$

where  $\alpha > 0$  in each case.

(b) Show that these distributions are fixed points for our renormalization-group transformation eqn 12.118. What are  $a$  and  $b$  for each distribution, in terms of  $\alpha$ ?

In parts (c) and (d) you will show that there are only these three fixed points  $g^*(x)$  for the renormalization transformation,  $T[g^*](x) = 2g^*(ax + b)$ , up to an overall linear rescaling of the variable  $x$ , with some caveats...

(c) First, let us consider the case  $a \neq 1$ . Show that the rescaling  $x \rightarrow ax + b$  has a fixed point  $x = \mu$ . Show that the most general form for the fixed-point function is

$$g^*(\mu \pm z) = z^{\alpha'} p_{\pm}(\gamma \log z) \quad (12.121)$$

for  $z > 0$ , where  $p_{\pm}$  is periodic and  $\alpha'$  and  $\gamma$  are constants such that  $p_{\pm}$  has period equal to one. (Hint: Assume  $p(y) \equiv 1$ , find  $\alpha'$ , and then show  $g^*/z^{\alpha'}$  is periodic.) What are  $\alpha'$  and  $\gamma$ ? Which choice for  $a$ ,  $p_+$ , and  $p_-$  gives the Weibull distribution? The Fréchet distribution?

Normally the periodic function  $p(\gamma \log(x - \mu))$  is assumed or found to be a constant (sometimes called  $1/\beta$ , or  $1/\beta^{\alpha'}$ ). If it is not constant, then the probability density must have an infinite number of oscillations as  $x \rightarrow \mu$ , forming a weird essential singularity.

(d) Now let us consider the case  $a = 1$ . Show again that the fixed-point function is

$$g^*(x) = e^{-x/\beta} p(x/\gamma) \quad (12.122)$$

with  $p$  periodic of period one, and with suitable constants  $\beta$  and  $\gamma$ . What are the constants in terms of  $b$ ? What choice for  $p$  and  $\beta$  yields the Gumbel distribution?

Again, the periodic function  $p$  is often assumed a constant ( $e^\mu$ ), for reasons which are not as obvious as in part (c).

What are the domains of attraction of the three fixed points? If we want to study the maximum of many samples, and the initial probability distribution has  $F(x)$  as its CDF, to which universal form will the extreme value statistics converge? Mathematicians have sorted out these questions.

If  $\rho(x)$  has a power-law tail, so  $1 - F(x) \propto x^{-\alpha}$ , then the extreme value statistics will be of the Frechet type, with the same  $\alpha$ . If the initial probability distribution is bounded above at  $\mu$  and if  $1 - F(\mu - y) \propto y^\alpha$ , then the extreme value statistics will be of the Weibull type. (More commonly, Weibull distributions arise as the smallest value from a distribution of positive random numbers, Exercise 1.9.) If the probability distribution decays faster than any polynomial (say, exponentially) then the extreme value statistics will be of the Gumbel form. (Gumbel extreme-value statistics can also arise for bounded random variables if the probability decays to zero faster than a power law at the bound.)

### 12.21 Diffusion equation and universal scaling functions.<sup>126</sup> ②

The diffusion equation universally describes microscopic hopping systems at long length scales. We will investigate how to write the evolution in a universal scaling form.

The solution to a diffusion problem with a non-zero drift velocity is given by  $\rho(x, t) = 1/\sqrt{4\pi D t} \exp(-((x - vt)^2/(4Dt)))$ . We will coarse grain by throwing away half the time points. We will then rescale the distribution so it looks like the original distribution. We can just write these two operations as  $t' = t/2$ ,  $x' = x/\sqrt{2}$ ,  $\rho' = \sqrt{2}\rho$ .<sup>127</sup> These three together constitute our renormalization group operation. (a) Write an expression for  $\rho'(x', t')$  in terms of  $D$ ,  $v$ ,  $x'$ , and  $t'$  (not in terms of  $D'$  and  $v'$ ). Use it to determine the new renormalized velocity  $v'$  and diffusion constant  $D'$ . Are  $v$  and  $D$  relevant, irrelevant or marginal variables?

Typically, whenever writing properties in a scaling function, there is some freedom in deciding which invariant combinations to use. Here let us use the invariant combination of variables,  $\mathcal{X} = x/\sqrt{t}$  and  $\mathcal{V} = \sqrt{t}v$ . We can then write

$$\rho(x, t) = t^{-\alpha} \mathcal{P}(\mathcal{X}, \mathcal{V}, D), \quad (12.123)$$

a power law times a universal scaling function of invariant combination of variables.

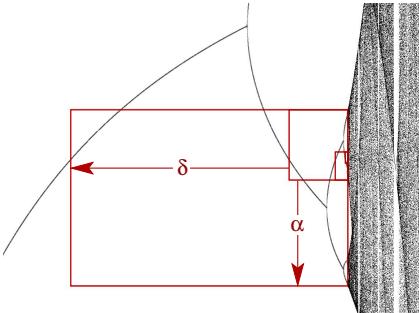
(b) Show that  $\mathcal{X}$  and  $\mathcal{V}$  are invariant under our renormalization group operation. What is  $\alpha$ ? Write an expression for  $\mathcal{P}$ , in terms of  $\mathcal{X}$ ,  $\mathcal{V}$ , and  $D$  (and not  $x$ ,  $v$ , or  $t$ ).

(Note that we need to solve the diffusion equation to find the universal scaling function  $\mathcal{P}$ , but we can learn a lot from just knowing that it is a fixed point of the renormalization group. So, the universal exponent  $\alpha$  and the invariant scaling combinations  $\mathcal{X}$ ,  $\mathcal{V}$ , and  $D$  are determined just by the coarsening and rescaling steps in the renormalization group. In experiments and simulations, one often uses data to extract the universal critical exponents and universal scaling functions, relying on emergent scale invariance to tell us that a scaling form like eqn 12.123 is expected.)

### 12.22 Period doubling and the onset of chaos.

(Dynamical systems) ②

Most of you will be familiar with the *period doubling route to chaos*, and the bifurcation diagram shown below. (See also Section 12.3.3).



**Fig. 12.49 Scaling in the period doubling bifurcation diagram.** Shown are the points  $x$  on the attractor (vertical) as a function of the control parameter  $\mu$  (horizontal), for the logistic map  $f(x) = 4\mu x(1 - x)$ , near the transition to chaos.

The self-similarity here is not in space, but in time. It is discrete instead of continuous; the behavior is similar if one rescales time by a factor of two, but not by a factor  $1 + \epsilon$ . Hence instead of power laws we find a discrete self-similarity as we approach the critical point  $\mu_\infty$ . (a) From the diagram shown, roughly estimate the values of the Feigenbaum numbers  $\delta$  (governing the rescaling of  $\mu - \mu_\infty$ ) and  $\alpha$  (governing the rescaling of  $x - x_p$ , where  $x_p = 1/2$  is the peak of

<sup>126</sup>This exercise was developed in collaboration with Archishman Raju.

<sup>127</sup>Because  $\rho$  is a density, we need to rescale  $\rho' dx' = \rho dx$ .

the logistic map). (Hint: be sure to check the signs.)

Remember that the relaxation time for the Ising model became long near the critical temperature; it diverges as  $t^{-\zeta}$  where  $t$  measures the distance to the critical temperature. Remember that the correlation length diverges as  $t^{-\nu}$ . Can we define  $\zeta$  and  $\nu$  for period doubling?

(b) If each rescaling shown doubles the period  $T$  of the map, and  $T$  grows as  $T \sim (\mu_\infty - \mu)^{-\zeta}$  near the onset of chaos, write  $\zeta$  in terms of  $\alpha$  and  $\delta$ . If  $\xi$  is the smallest typical length scale of the attractor, and we define  $\xi \sim (\mu_\infty - \mu)^{-\nu}$  (as is traditional at thermodynamic phase transitions), what is  $\nu$  in terms of  $\alpha$  and  $\delta$ ? (Hint: be sure to check the signs.)

### 12.23 Ising mean field derivation.<sup>128</sup> (Mathematics) ③

In this exercise, we derive the mean-field free energy  $F^{\text{MF}}$  for the Ising model used in Exercise 12.5. Our mean-field approximation will be a rigorous upper bound on the true free energy  $F$ , as shown by a theorem attributed to Gibbs, Bogoliubov, and Feynman (GBF)

$$F \leq F^{\text{MF}} = F^{[0]} + \langle \mathcal{H} - \mathcal{H}^{[0]} \rangle_{[0]} \quad (12.124)$$

(see Exercise 12.24 for a derivation). Here  $\mathcal{H}^{[0]}$  is a simple trial Hamiltonian with the same states  $\alpha$  (say spin configurations on a square lattice) as the true Hamiltonian  $\mathcal{H}$ .  $F^{[0]}$  is the free energy of that trial Hamiltonian, and  $\langle X \rangle_{[0]}$  is the thermal expectation of the operator  $X$  in the thermal ensemble for the trial Hamiltonian

$$\langle X \rangle_{[0]} = (1/Z^{[0]}) \sum_{\alpha} X_{\alpha} \exp(-\beta \mathcal{H}_{\alpha}^{[0]}). \quad (12.125)$$

If the original Hamiltonian depends on parameters (here  $\mathcal{H}(J, H) = -J \sum_{\langle ij \rangle} s_i s_j - H \sum_i s_i$ ), and the new Hamiltonian has parameters (here  $\mathcal{H}^{[0]}(h) = -h \sum_i s_i$ ), the new parameter(s) can be varied to minimize  $F^{\text{MF}}$ , leading to a temperature-dependent mean-field variational bound  $F^{\text{MF}}(\beta, J, H)$  that is remarkably useful for practical calculations away from critical points.

The key challenge is to find a trial Hamiltonian simple enough that one can calculate both the free energy  $F^{[0]}$  and the expectation values  $\langle \mathcal{H} \rangle_{[0]}$  and  $\langle \mathcal{H}^0 \rangle_{[0]}$ . For the  $N$ -spin nearest-neighbor

Ising model on a square lattice of this problem, we shall use let us use as a trial Hamiltonian a set of  $N$  uncoupled spins, each in an external field  $h$ .

Derive the mean field equations (eqn 12.112 used in Exercise 12.5) for the square lattice Ising ferromagnet  $\mathcal{H} = -J \sum_{\langle ij \rangle} s_i s_j - H \sum_i s_i$  using the GBF inequality (eqn 12.124), where  $\mathcal{H}^{[0]} = -\sum_i h S_i$  is a trial Hamiltonian, with  $h$  chosen to minimize the right-hand side.

### 12.24 Mean-field bound for free energy.<sup>129</sup> (Mean field) ③

Exercise 12.23 uses a rigorous upper bound eqn 12.124,  $F \leq F^{\text{MF}} = F^{[0]} + \langle \mathcal{H} - \mathcal{H}^{[0]} \rangle_{[0]}$  to generate the standard mean-field theory for the Ising model. Again,  $\mathcal{H}^{[0]}$  is a trial Hamiltonian with the same states  $\alpha$  as the true Hamiltonian  $\mathcal{H}$ ,  $F^{[0]}$  is the trial Hamiltonian free energy, and  $\langle X \rangle_{[0]} = (1/Z^{[0]}) \sum_{\alpha} X_{\alpha} e^{-\beta \mathcal{H}_{\alpha}^{[0]}}$  is average of  $X$  in the trial thermal ensemble, since the probability in the trial ensemble of being in state  $\alpha$  is  $\rho_{\alpha}^{[0]} = \exp(-\beta \mathcal{H}_{\alpha}^{[0]})/Z^{[0]}$ . In this exercise, we derive this bound, which is attributed to Gibbs, Bogoliubov, and Feynman. This will be done in two steps.

(a) Show that

$$\begin{aligned} Z &= \sum_{\alpha} e^{-\beta \mathcal{H}_{\alpha}} \\ &= Z^{[0]} \langle e^{-\beta(\mathcal{H} - \mathcal{H}^{[0]})} \rangle_{[0]}. \end{aligned} \quad (12.126)$$

(Hint: Work backward from the answer)

Now  $\exp(x)$  is a convex function: the line connecting two points on the graph lies above the curve:  $\lambda \exp(x_1) + (1 - \lambda) \exp(x_2) \geq \exp(\lambda x_1 + (1 - \lambda)x_2)$ . This is true for sums of many terms (see note 37 on page 108):

$$\begin{aligned} \langle \exp(X) \rangle_{[0]} &= \sum_{\alpha} \rho_{\alpha}^{[0]} \exp(X_{\alpha}) \geq \exp(\rho_{\alpha}^{[0]} X_{\alpha}) \\ &= \exp(\langle x \rangle_{[0]}). \end{aligned}$$

(b) Use this to derive the GBF inequality of eqn 12.124,  $F \leq F^{\text{MF}} = F^{[0]} + \langle \mathcal{H} - \mathcal{H}^{[0]} \rangle_{[0]}$ .

Notice that having a variational bound on the free energy is not as useful as it sounds. The free energy itself is usually not of great interest; the objects of physical importance are almost always derivatives (first derivatives like the pressure, energy and entropy, and second derivatives like the specific heat) or differences (telling

<sup>128</sup>See Cardy [13], equation 2.7 and exercise 2.1.

<sup>129</sup>See Cardy [13], equation 2.7 and exercise 2.1.

which phase is more stable). Knowing an upper bound to the free energy doesn't tell us much about these other quantities. Like most variational methods, mean-field theory is useful not because of its variational bound, but because (if done well) it provides qualitative insight into the behavior.

**12.26 The onset of chaos: Lowest order renormalization-group for period doubling.** (Dynamical systems) ③

In this exercise, we set up a low-order approximate renormalization group to study the period-doubling route to chaos of Fig. 12.17. Our goal is to estimate the scaling factors  $\alpha \sim -2.5029$  and  $\delta \sim 4.6692$  governing the self-similarity in space  $x$  and control parameter  $\mu$  near the onset of chaos  $\mu_\infty$ , as discussed in Exercise 12.22. We shall extend the renormalization group we develop here to higher order in Exercise 12.27. The period-doubling route to chaos is understood by a renormalization group that (as usual) has a coarse-graining step and a rescaling. The coarse-graining step ‘decimates’ the time series  $\{x, g(x), g(g(x)), \dots\}$  by dropping every other point, replacing the function  $g(x)$  by  $g(g(x))$ . The rescaling expands  $x$  about its maximum by a factor  $\alpha$ . If we choose the maximum value of  $g(x)$  to be at zero, the renormalization group sends  $g$  to the function  $T[g]$ , where

$$T[g](x) = \alpha g(g(x/\alpha)). \quad (12.127)$$

(We explore the scaling and universality implied by this renormalization group in Exercise 12.9; our eqn 12.127 is the same as eqn 12.40) in that exercise, except there our functions were symmetric about  $x = \frac{1}{2}$ .)

Our functions  $g(x)$  are ‘one-humped maps’, with a parabolic maximum at zero. To lowest order, let us approximate  $g(x)$  by a parabola centered at the origin

$$g(x) \approx G_0 + G_1 x^2. \quad (12.128)$$

where  $G_1 < 0$ .

We shall approximate the fixed point  $g^*$  of our renormalization group by demanding that  $T[g^*](x) = g^*(x)$  at two points,  $x = 0$  and  $x = 1$ . We have three constants,  $\alpha$ ,  $G_0^*$ , and  $G_1^*$ , so for convenience we set  $G_0^*$  to one.

(a) Use the fixed-point condition at  $x = 0$  to show that  $\alpha = 1/g^*(1) = 1/(1 + G_1^*)$ .

(b) Use the fixed-point condition at  $x = 1$  to give an equation for  $G_1^*$ , substituting in your equation for  $\alpha$  above. (Hint: The equation simplifies to a sixth-order polynomial.)

We expect  $\alpha \approx -2.5$ , so since our approximate  $\alpha = 1/(1 + G_1^*)$ , we expect  $G_1^* \approx 1/\alpha - 1 \approx -1.4$ .

(c) Plot the quantity from part (b) that must be zero: does it have a root near  $G_1 = -1.4$ ? Numerically solve your equation from part (b) for the root closest to  $-1.4$ . What is  $\alpha$  in our approximation? (Your approximation for  $\alpha$  should be within a few percent of the correct value  $\alpha = -2.5029\dots$ ; your value for  $G_1^*$  should be within a few percent of  $-1.4$  and not too far from the true value of the quadratic term at the fixed point,  $-1.5276\dots$ )

In statistical mechanics, we find the universal critical exponents by linearizing the renormalization-group flows about the fixed point and finding directions that grow. Here the exponent  $\delta = 4.669\dots$  describes the fastest growing direction in function space:  $T[g^* + \epsilon\psi](x) - g^*(x) = \delta\psi(x)$ . That is, we add a perturbation  $g(x) = g^*(x) + \epsilon\psi(x)$  and study to linear order in  $\epsilon$  how the perturbation grows under  $T$ . The lowest-order perturbation to our parabola adds an overall constant  $G_0 \rightarrow 1 + \epsilon$ . Our function  $g^*(x)$  is fixed at both  $x = 0$  and  $x = 1$ . Let us check the growth of our perturbation at  $x = 0$ , which should grow by a factor of approximately  $\delta$  when our renormalization-group transformation is applied.

(d) Using  $g(x) = (1+\epsilon)+G_1 x^2$ , write the formula for the term in  $T[g](0)$  linear in  $\epsilon$  as a function of  $G_1$  and  $\alpha$ . Insert your fixed-point values for  $G_1$  and  $\alpha$  from part (c). What is your estimate for  $\delta$ ? (Your approximation for  $\delta$  should be within a few percent of the correct value  $\delta = 4.669\dots$ )

**12.27 The onset of chaos: Full renormalization-group calculation.**<sup>130</sup> (Computation, Dynamical systems) ③

In this exercise, we implement Feigenbaum's numerical scheme [18, pp. 693-694] for finding high-precision values of the universal constants

$$\begin{aligned} \alpha &= -2.50290787509589282228390287322 \\ \delta &= 4.66920160910299067185320382158, \end{aligned} \quad (12.129)$$

that quantify the scaling properties of the period-doubling route to chaos (Fig. 12.17, Ex-

<sup>130</sup>Hints for the computations can be found at the book Web site [49].

ercise 12.22). This extends the lowest-order calculation of Exercise 12.26.

Our renormalization group operation (Exercises 12.9 and 12.26) coarse-grains in time taking  $g \rightarrow g \circ g$ , and then rescales distance  $x$  by a factor of  $\alpha$ . Centering our functions at  $x = 0$ , this leads to  $T[g](x) = \alpha g(g(x/\alpha))$  (eqn 12.127).

We shall solve for the properties at the onset of chaos by analyzing our function-space renormalization-group by expanding our functions in a power series

$$g(x) \approx 1 + \sum_{n=1}^N G_n x^{2n}. \quad (12.130)$$

Notice that we only keep even powers of  $x$ ; the fixed point is known to be symmetric about the maximum, and the unstable mode responsible for the exponent  $\delta$  will also be symmetric.

First, we must approximate the fixed point  $g^*(x)$  and the corresponding value of the universal constant  $\alpha$ . At order  $N$ , we must solve for  $\alpha$  and the  $N$  polynomial coefficients  $G_n^*$ . We can use the  $N+1$  equations fixing the function at equally spaced points in the positive unit interval:

$$T[g^*](x_m) = g^*(x_m), \quad x_m = m/N, \quad m = \{0, \dots, N\}. \quad (12.131)$$

We can use the first of these equations to solve for  $\alpha$ .

(a) Show that the equation for  $m = 0$  sets  $\alpha = 1/g^*(1)$ .

We can use a root-finding routine to solve for  $G_n^*$ . (b) Implement the other  $N$  constraint equations of eqn 12.131 in a form appropriate for your method of finding roots of nonlinear equations, substituting your value for  $\alpha$  from part (a). Check that your routine at  $N = 1$  gives values for  $\alpha \approx -2.5$  and  $G_1^* \approx -1.5$ . (These should reproduce the values from Exercise 12.26(c).)

(c) Use a root-finding routine to calculate  $\alpha$  for  $N = 1, \dots, 9$ . Start the search at  $G_1^* = -1.5$ ,  $G_n^* = 0$  ( $n > 1$ ) to avoid landing at the wrong fixed point. (If it is convenient for you to use high-precision arithmetic, continue to higher  $N$ .)

To how many decimal places can you reproduce the correct value for  $\alpha$  in eqn 12.129?

Now we need to solve for the renormalization group flows  $T[g]$ , linearized about the fixed point  $g(x) = g^*(x) + \epsilon\psi(x)$ . Feigenbaum tells us that  $T[g^* + \epsilon\psi] = T[g^*] + \epsilon\mathcal{L}[\psi]$ , where  $\mathcal{L}$  is the linear

operator taking  $\psi(x)$  to

$$\mathcal{L}[\psi](x) = \alpha\psi(g^*(x/\alpha)) + \alpha g^{*\prime}(g(x/\alpha))\psi(x/\alpha). \quad (12.132)$$

(d) Derive eqn 12.132.

We want to find eigenfunctions that satisfy  $\mathcal{L}[\psi] = \lambda\psi$ . Again, we can expand  $\psi(x)$  in a polynomial

$$\psi(x) = \sum_{n=0}^{N-1} \psi_n x^{2n} \quad (\psi_0 \equiv 1). \quad (12.133)$$

We then approximate the action of  $\mathcal{L}$  on  $\psi$  by its action at  $N$  points  $\tilde{x}_i$ , that need not be the same as the  $N$  points  $x_m$  we used to find  $g^*$ . We shall use  $\tilde{x}_i = (i-1)/(N-1)$ ,  $i = 1, \dots, N$ . (For  $N = 1$ , we use  $\tilde{x}_1 = 0$ .) This leads us to a linear system of  $N$  equations for the coefficients  $\psi_n$ , using eqns 12.133 and 12.134:

$$\begin{aligned} \sum_{n=0}^{N-1} [\alpha g(\tilde{x}_i/\alpha)^{2n} + \alpha g'(g(\tilde{x}_i/\alpha))(\tilde{x}_i/\alpha)^{2n}] \psi_n \\ = \lambda \sum_{n=0}^{N-1} \tilde{x}_i^{2n} \psi_n \end{aligned} \quad (12.134)$$

These equations for the coefficients  $\psi_n$  of the eigenfunctions of  $\mathcal{L}$  is in the form of a *generalized eigenvalue problem*

$$\sum_n L_{in} \psi_n = \lambda \sum_n X_{in} \psi_n. \quad (12.135)$$

The solution to the generalized eigenvalue problem can be found from the eigenvalues of  $X^{-1}L$ , but most eigenvalue routines provide a more efficient and accurate option for directly solving the generalized equation given  $L$  and  $X$ .

(e) Write a routine that calculates the matrices  $L$  and  $X$  implicitly defined by eqns 12.135 and 12.134. For  $N = 1$  you should generate  $1 \times 1$  matrices. For  $N = 1$ , what is your prediction for  $\delta$ ? (These should reproduce the values from Exercise 12.26(d).)

(f) Solve the generalized eigenvalue problem for  $L$  and  $X$  for  $N = 1, \dots, 9$ . To how many decimal places can you reproduce the correct value for  $\delta$  in eqn 12.129?

### 12.28 Singular corrections to scaling. ③

The renormalization group says that the number of relevant directions at the fixed point in system space is the number of parameters we need to tune to see a critical point, and that the critical exponents depend on the eigenvalues of these

relevant directions. Do the irrelevant directions matter?

Let the Ising model in zero field be described by flow equations

$$dt_\ell/d\ell = t_\ell/\nu, \quad du_\ell/d\ell = -yu_\ell \quad (12.136)$$

where  $t_\ell$  describes the renormalization of the reduced temperature  $t = (T_c - T)/T_c$  after a coarse-graining by a factor  $b = \exp(\ell)$ , and  $u$  and  $u_\ell$  represent a slowly-decaying irrelevant perturbation under the renormalization group. In Fig. 12.8, one may view  $t$  as the expanding eigendirection running roughly horizontally, and  $u$  as the contracting, irrelevant coordinate running roughly vertically. Thus our model starts with a value  $u_0$  associated to the distance in system space between our critical point  $R_c$  and the RG fixed point  $S^*$  along the irrelevant coordinate.

(a) What is the invariant combination  $z = ut^\omega$  that stays constant under the renormalization group? What is  $\omega$  in terms of the eigenvalues  $-y$  and  $1/\nu$ ?

Properties near critical points have universal power law singularities, but the corrections to these power laws also have universal properties predicted by the renormalization group. These come in two types – *analytic* corrections to scaling and *singular* corrections to scaling.

Let us consider corrections to the susceptibility. In analogy with other systems we have studied, we would expect that the susceptibility

$$\chi(t, u) = t^{-\gamma} X(z) \quad (12.137)$$

with  $X(z)$  a universal function of the invariant combination you found in part (a). (We shall derive this scaling form in Exercise 12.29.) As a function of  $t$ ,  $\chi(t, u)$  has singularities at small  $t$ . But we expect properties to be analytic as we vary  $u$ , since the irrelevant direction is not being tuned to a special value, so we expect that a Taylor series of  $\chi(t, u)$  in powers of  $u$  should make sense. Since  $z \propto u$ , we thus expect that  $X(z)$  will be an analytic function of  $z$  for small  $z$ .<sup>131</sup>

(b) Show that for small  $t$ , your  $z$  from part (a) goes to zero. Taylor expand  $X(z)$ . What corrections do you predict for the susceptibility from

the first and second-order terms in the series?

These are the singular corrections to scaling due to the irrelevant perturbation  $u$ .

An Ising magnet on a sample holder is loaded into a magnetometer, and the susceptibility is measured<sup>132</sup> at zero external field as a function of reduced temperature  $t = (T - T_c)/T_c$ . It is found to be well approximated by

$$\begin{aligned} \chi(T) = & At^{-1.24} + Bt^{-0.83} + Ct^{-0.42} \\ & + D + Et + \dots \end{aligned} \quad (12.138)$$

You may ignore any errors due to the magnetometer.

(c) The exponent  $\omega \approx 0.407$  for the 3D Ising universality class, and  $\gamma \approx 1.237$ . Which terms are explained as singular corrections to scaling?

(d) Can you provide a physical interpretation for the terms in eqn 12.138 that are not explained by singular corrections to scaling? For example, how do we expect the susceptibility of the sample holder to depend on temperature? These are examples of analytic corrections to scaling.

One must note that it is normally completely infeasible in an experiment or simulation to measure quantities with sufficiently accuracy to identify so many simultaneous corrections to scaling.

### 12.29 Singular corrections to scaling and the renormalization group. ③

In this exercise, we derive the form of the scaling function (eqn 12.137) for the effects of *irrelevant* operators on the properties of systems near critical points (see Exercise 12.28). Remember that irrelevant directions shrink under coarse-graining. Let  $\chi$  be the susceptibility of the Ising model, as a function of the reduced temperature  $t = T_c - T$  and some irrelevant operator  $u$ :

$$\begin{aligned} d\chi_\ell/d\ell &= -(\gamma/\nu)\chi_\ell, \\ dt_\ell/d\ell &= t_\ell/\nu, \\ du_\ell/d\ell &= -yu_\ell \end{aligned} \quad (12.139)$$

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  $\chi$ . Consider a point  $(t_0, u_0, \chi_0)$  in the system space, and the invariant curve defined by  $z = u_0 t_0^\omega$  (dashed lines). Our renormalization

<sup>131</sup>Had we used a scaling variable  $Z = tu^{1/\omega}$ , for example, we would not have expected the corresponding scaling function to be analytic in small  $Z$ .

<sup>132</sup>The accuracy of the quoted exponents is not experimentally realistic.

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  $\chi_{\ell^*}$  where the flow crosses  $t_{\ell^*} = 1$ .

(a) *Solve eqns 12.139 for  $u_\ell$  and  $t_\ell$ . Setting  $t_{\ell^*} = 1$ , what is  $u_{\ell^*}$  in terms of your invariant combination  $z$ ?*

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

(b) *Solve eqns 12.139 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 12.137?*

Remember the critical manifold is co-dimension one (or two, if you include temperature and external field), and the unstable manifold is dimension one (or two) – so we get universal predictions for a huge variety of systems, by observing the outgoing trajectories near a narrow tube or surface emitted from the fixed point.

### 12.30 Nonlinear flows, analytic corrections, and hyperscaling.<sup>133</sup> 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  $\ell$ . One can think of one coarse-graining step by  $b = (1 + \epsilon)$  as incrementing  $\ell \rightarrow \ell + \epsilon$ ; equivalently, coarse-graining to  $\ell$  changes length scales by  $\exp(\ell)$ .

Consider the particular flow equations<sup>134</sup>

$$df_\ell/d\ell = Df_\ell - at_\ell^2 \quad dt_\ell/d\ell = t_\ell/\nu, \quad (12.140)$$

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

Notice there are several free energies here. We shall call the free energy per spin of the actual system either  $f$  or  $f_0$ , and the temperature of the actual system either  $t$  or  $t_0$ . The coarse-grained free energy and temperature are  $f_\ell$  and  $t_\ell$ , 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  $\ell^*$  where  $t_{\ell^*} = 1$ , at which point the coarse-grained free energy is  $f_{\ell^*}$ . Let us start with the linear case  $a = 0$ .

(a) *Solve for  $f_\ell$  and  $t_\ell$  for  $a = 0$ . Setting  $t_{\ell^*} = 1$ , solve for  $f_0$  in terms of  $f_{\ell^*}$ ,  $t_0$ ,  $D$ , and  $\nu$ . Solve for the specific heat per spin  $c = T\partial^2 f/\partial T^2$ , where  $t = t_0 = (T - T_c)/T_c$  and  $f = f_0$ . (Hint: Use the chain rule  $\partial f/\partial T = (\partial f/\partial t)(dt/dT)$ .) 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*

$$c = t^{-\alpha}(c_0 + c_1 t + c_2 t^2 + \dots), \quad (12.141)$$

*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,  $df/d\ell = Df + \dots$ , where all other terms are hard-to-compute critical exponents? There is no completely general answer to this question (although there are arguments for specific models). Indeed, 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.

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

<sup>133</sup>This exercise developed in collaboration with Colin Clement

<sup>134</sup>Note that these are total derivatives. So the first equation tells us the total change in  $f$  after coarsening by a factor  $1 + d\ell$ .  $f(t)$  then will coarse grain to  $f_\ell(t_\ell)$  without needing to worry about the chain rule  $df(t)/d\ell = \partial f/\partial t + \partial f/\partial t dt/d\ell$ .

Look up the critical exponent describing how the correlation length  $\xi$  diverges as  $t \rightarrow 0$ .) 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*.<sup>135</sup> Here it is *subdominant* – near the critical temperature where  $t \rightarrow 0$ , 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 12.141.

Does the introduction of the higher-order nonlinear term  $at_\ell^2$  in eqn 12.140 change the behavior in an important way? Rather than exercising your expertise in analytic solutions of nonlinear differential equations, eqn 12.142 provides not  $f_\ell$  and  $t_\ell$  as functions of  $\ell$ , but the relation between the two:

$$\begin{aligned} f_\ell(t_\ell) &= f_0 \left( \frac{t_\ell}{t_0} \right)^{D\nu} \\ &\quad - \frac{avt_\ell^2}{(2-D\nu)} (1 - (t_\ell/t_0)^{-(2-D\nu)}). \end{aligned} \quad (12.142)$$

(c) Show that  $f_\ell(t_\ell)$  in eqn 12.142 satisfies the differential equation given by eqn 12.140, using

$$\frac{df_\ell}{dt_\ell} = \frac{df_\ell}{d\ell} / \frac{dt_\ell}{d\ell} \quad (12.143)$$

Show that it has the correct initial conditions at  $\ell = 0$ . What is  $f_{\ell^*}$  at  $\ell^*$ , 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  $f_0(t_0)$  that we want to know. Since here we have only one relevant variable (in zero field), all the flows lead to the same<sup>136</sup> final point  $f_{\ell^*}(t_{\ell^*} = 1)$

(d) Solve for  $f_0$  in terms of  $f_{\ell^*}$  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

$$c = c_{\text{+analytic}}(t) + t^{-\alpha} c_{\text{*analytic}}(t) \quad (12.144)$$

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  $f_{\ell^*}$ ,  $T_c$ ,  $a$ ,  $\nu$ , and  $D$ .

Here the nonlinear term  $a$  gives us not only a smooth multiplicative term in the specific heat, but also a smooth additive background. This clearly is also expected in a real physical system – other degrees of freedom uncoupled to the magnetization, or even the box holding the magnet, will contribute a specific heat that is analytic near  $T_c$ .

**12.31 Beyond power laws: Nonlinear flows and logarithms in the 2D Ising model.**<sup>137</sup> 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

$$\begin{aligned} c(T) &= k_B \frac{2}{\pi} \left( \frac{2J}{k_B T_c} \right)^2 \left[ -\log(1 - T/T_c) \right. \\ &\quad \left. + \log(k_B T_c/(2J)) - (1 + \pi/4) \right] \\ &= -\frac{8}{\pi k_B T_c^2} \log \left( \frac{t}{\frac{1}{2}k_B T_c \exp(-(1 + \pi/4))} \right) \\ &= -c_0 \log \left( \frac{t}{\tau} \right) \end{aligned} \quad (12.145)$$

where  $t = (T - T_c)/T_c$  and we set  $J = 1$ . (Remember  $\log(t)$  is negative for small  $t$ .) Linearized flows around the renormalization group fixed point predict  $c \sim t^{-\alpha}$ , and when  $\alpha \rightarrow 0$  one often observes logarithmic corrections. But such

<sup>135</sup>These are distinct from *singular* corrections to scaling that arise, for example, from irrelevant terms under the renormalization group, that would produce subdominant corrections to  $c$  with powers  $t^{-\alpha+\Delta}$  where  $\Delta$  is not an integer, and is bigger than zero (hence subdominant).

<sup>136</sup>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).

<sup>137</sup>This exercise developed in collaboration with Colin Clement

corrections are not predicted by the linearized flows. The key nonlinear term is the term  $at_\ell^2$  of eqn 12.140 in Exercise 12.30.

(a) Is the solution for  $f_\ell(t_\ell)$  in eqn 12.142 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 12.140 for  $D = 2$

$$f_\ell(t_\ell) = f_0(t_\ell/t_0)^2 - at_\ell^2 \log(t_\ell/t_0). \quad (12.146)$$

(b) Show that  $f_\ell(t_\ell)$  in eqn 12.146 satisfies the differential equation given by eqn 12.140, with the correct initial conditions. Solve for  $f_0$  in terms of  $f_{\ell^*}$  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:  $\partial f/\partial T = (\partial f/\partial t)(dt/dT)$ .) Does it agree asymptotically with the exact result in eqn 12.145? What are  $c_0$  and  $\tau$ , in terms of  $a$  and  $f_{\ell^*}$ ?

Thus for the 3D Ising model (Exercise 12.30), 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 [?].

### 12.25 Avalanche size distribution. ③

One can develop a mean-field theory for avalanches in non-equilibrium disordered systems by considering a system of  $N$  Ising spins coupled to one another by an infinite-range interaction of strength  $J/N$ , with an external field  $H$  and each spin also having a local random field  $h$ :

$$\mathcal{H} = -J_0/N \sum_{i,j} S_i S_j - \sum_i (H + h_i) S_i. \quad (12.147)$$

We assume that each spin flips over when it is pushed over; i.e., when its change in energy

$$\begin{aligned} H_i^{\text{loc}} &= \frac{\partial \mathcal{H}}{\partial S_i} = J_0/N \sum_j S_j + H + h_i \\ &= J_0 m + H + h_i \end{aligned}$$

changes sign.<sup>138</sup> Here  $m = (1/N) \sum_j S_j$  is the average magnetization of the system. All spins start by pointing down. A new avalanche is launched when the least stable spin (the unflipped spin of largest local field) is flipped by increasing the external field  $H$ . Each spin flip changes the magnetization by  $2/N$ . If the magnetization change from the first spin flip is enough to trigger the next-least-stable spin, the avalanche will continue.

We assume that the probability density for the random field  $\rho(h)$  during our avalanche is a constant

$$\rho(h) = (1+t)/(2J_0). \quad (12.148)$$

The constant  $t$  will measure how close the density is to the critical density  $1/(2J_0)$ .

(a) Show that at  $t = 0$  each spin flip will trigger on average one other spin to flip, for large  $N$ . Can you qualitatively explain the difference between the two phases with  $t < 0$  and  $t > 0$ ?

We can solve exactly for the probability  $D(S,t)$  of having an avalanche of size  $S$ . To have an avalanche of size  $S$  triggered by a spin with random field  $h$ , you must have precisely  $S-1$  spins with random fields in the range  $\{h, h+2J_0S/N\}$  (triggered when the magnetization changes by  $2S/N$ ). The probability of this is given by the Poisson distribution. In addition, the random fields must be arranged so that the first spin triggers the rest. The probability of this turns out to be  $1/S$ .

(b) (Optional) By imagining putting periodic boundary conditions on the interval  $\{h, h+2J_0S/N\}$ , argue that exactly one spin out of the group of  $S$  spins will trigger the rest as a single avalanche. (Hint from Ben Machta: For simplicity, we may assume<sup>139</sup> the avalanche starts at  $H = m = 0$ . Try plotting the local field  $H^{\text{loc}}(h') = J_0 m(h') + h'$  that a spin with random field  $h'$  would feel if the spins between  $h'$  and  $h$  were flipped. How would this function change if we shuffle all the random fields around the periodic boundary conditions?)

(c) Show that the distribution of avalanche sizes is thus

$$D(S,t) = \frac{S^{S-1}}{S!} (t+1)^{S-1} e^{-S(t+1)}. \quad (12.149)$$

<sup>138</sup>We ignore the self-interaction, which is unimportant at large  $N$

<sup>139</sup>Equivalently, measure the random fields with respect to  $h_0$  of the triggering spin, and let  $m$  be the magnetization change since the avalanche started.

With  $t$  small (near to the critical density) and for large avalanche sizes  $S$  we expect this to have a scaling form:

$$D(S, t) = S^{-\tau} \mathcal{D}(S/t^{-x}) \quad (12.150)$$

for some mean-field exponent  $x$ . That is, taking  $t \rightarrow 0$  and  $S \rightarrow \infty$  along a path with  $St^x$  fixed, we can expand  $D(S, t)$  to find the scaling function.

(d) *Show that  $\tau = 3/2$  and  $x = 2$ . What is the scaling function  $\mathcal{D}$ ?* Hint: You'll need to use Stirling's formula  $S! \sim \sqrt{2\pi S} (S/e)^S$  for large  $S$ , and that  $1+t = \exp(\log(1+t)) \approx e^{t-t^2/2+t^3/3\dots}$ .

This is a bit tricky to get right. Let's check it by doing the plots.

(e) *Plot  $S^\tau D(S, t)$  versus  $Y = S/t^{-x}$  for  $t = 0.2, 0.1$ , and  $0.05$  in the range  $Y \in (0, 10)$ . Does it converge to  $\mathcal{D}(Y)$ ?*

See Reference [48] for more information.

### 12.32 Conformal invariance.<sup>140</sup> (Mathematics, Biology, Computation) ③

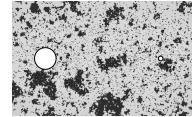
Emergence in physics describes new laws that arise from complex underpinnings, and often exhibits a larger symmetry than the original model. The diffusion equation emerges as a continuum limit from complex random microscopic motion, and diffusion on a square lattice has circular symmetry. Critical phenomena emerges near continuous phase transitions, and the resultant symmetry under dilations is exploited by the renormalization group to predict universal power laws and scaling functions.

The Ising model on a square lattice at the critical point, like the diffusion equation, also has an emergent circular symmetry: the complex patterns of up and down spins look the same on long length scales also when rotated by an angle. Indeed, making use of the symmetries under changes of length scale, position, and angle (plus one spatially nonuniform transformation), systems at their critical points have a *conformal* symmetry group.

In two dimensions, the conformal symmetry group becomes huge. Roughly speaking, any complex analytic function  $f(z) = u(x + iy) +$

$iv(x + iy)$  takes a snapshot of an Ising model  $M(x, y)$  and warps it into a new magnetization pattern at  $(u, v)$  that 'looks the same'. (Here  $u$ ,  $v$ ,  $x$ , and  $y$  are all real.)

You may remember that most ordinary functions (like  $z^2$ ,  $\sqrt{z}$ ,  $\sin(z)$ ,  $\log(z)$ , and  $\exp(iz)$ ) are analytic. All of them yield cool transformations of the Ising model – weird and warped when magnified until you see the pixels, but recognizably Ising-like on long scales. This exercise will generate an example.



**Fig. 12.50 Two proteins in a critical membrane.** The figure shows the pixels of a critical Ising model simulated in a square, conformally warping the square onto the exterior of two circles (representing two proteins in a cell membrane). The warped pixels vary in size – largest in the upper and lower left, smallest near the smaller circle. They also locally rotate and translate the square lattice, but notice the pixels remain looking square – the angles and aspect ratios remain unchanged. The pixels are gray rather than black and white, with only the smallest pixels pure black and white; we must not only warp the pixels conformally, but also rescale the magnetization. Ignoring the pixelation, the different regions look statistically similar. A gray large pixel mimics the average color of similar-sized regions of tiny pixels.

(a) *What analytic function shrinks a region uniformly by a factor  $b$ , holding  $z = 0$  fixed? What analytic function translates the lattice by a vector  $\mathbf{r}_0 = (u_0, v_0)$ ? What analytic function rotates a region by an angle  $\theta$ ?*

(b) *Expanding  $f(z+\delta) = f(z)+\delta f'(z)$ , show that an analytic function  $f$  transforms a local region about  $z$  to linear order in  $\delta$  by first rotating and dilating  $\delta$  about  $z$  and then translating. What complex number gives the net translation?*

Figure 12.50 shows how one can use this conformal symmetry to study the interactions between

<sup>140</sup>This exercise was developed in collaboration with Benjamin Machta. Ising snapshots and hints for the computations can be found at the book Web site [49].

<sup>141</sup>We use this transformation to study the effective interaction between two circular 'proteins' in a two-dimensional cell membrane near its critical point. The energy of attraction between two 'up' proteins is derivable from the energy of the square-lattice Ising model with the two side boundaries set to 'up'. (See B. B. Machta, S. L. Veatch,

circular ‘proteins’ embedded in a two dimensional membrane at an Ising critical point.<sup>141</sup> In the renormalization group, we first coarse grain the system (shrinking by a factor  $b$ ) and then rescale the magnetization (by some power  $b^{y_M}$ ) in order to return to statistically the same critical state:  $\widehat{M} = b^{y_M} M$ . This rescaling turns the larger pixels in Fig. 12.50 more gray; a mostly up-spin ‘white’ region with tiny pixels is mimicked by a large single pixel with the statistically averaged gray color.

We can discover the correct power for  $M$  by examining the rescaling of the correlation function  $C(\mathbf{r}) = \langle M(\mathbf{x})M(\mathbf{x} + \mathbf{r}) \rangle$ . In the Ising model at its critical point the correlation function  $C(r) \sim r^{-(2-d+\eta)}$ . In dimension  $d = 2$ ,  $\eta = 1/4$ . We expect that the correlation function for the conformally transformed magnetization will be the same as the original correlation function.

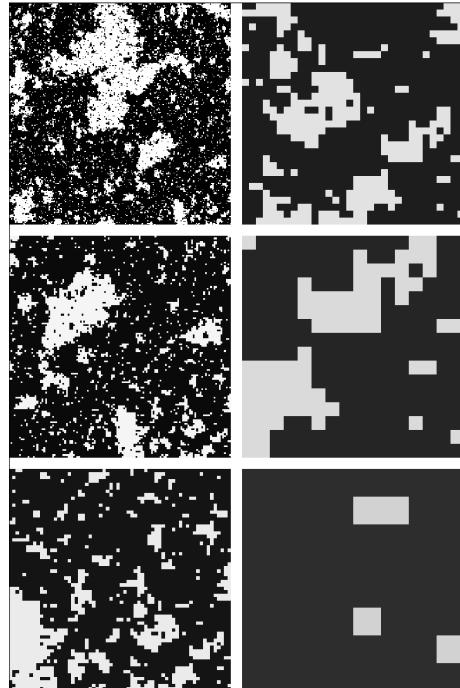
(c) If we coarse-grain by a constant factor  $b$ , what power of  $b$  must we multiply  $M$  by to make  $C(\mathbf{r}) = \langle \widehat{M}(f(z))\widehat{M}(f(z) + \mathbf{r}) \rangle$ ? Explain your reasoning.

When our conformal transformation takes a pixel at  $M(z)$  to a warped pixel of area  $A$  at  $f(z)$ , it rescales the magnetization by

$$\widehat{M}(f(z)) = |A|^{-1/16} M(z). \quad (12.151)$$

The pixel area for a locally uniform compression by  $b$  changes by  $|df/dz|^2 = 1/b^2$ . You may use this to check your answer to part (c).

Figure 12.51 illustrates the self-similarity for the Ising model under rescaling, in analogy with the treatment of scale invariance in Fig. 12.11 for the random-field avalanche model. Here, unlike in that figure, as we zoom in we incorporate the renormalization of the magnetization by changing the grayscale.



**Fig. 12.51** Ising model under conformal rescaling. The ‘powers of two’ rescaling of the avalanches in Fig. 12.11 ignored this rescaling of  $M$ . Here we show the Ising model, again with the lowest right-hand quarter of each square inflated to fill the next – but now properly faded according to eqn 12.151.

Let us now explore the image of the Ising model under various analytic functions. Generate a snapshot of an Ising model, equilibrated at the critical temperature.<sup>142</sup>

Notice that Mathematica’s notebook has trouble with plotting large numbers of polygons. Use Python. Or, if you prefer – save often, try not using a second display (a known problem), try not opening other programs in the background (including second Mathematica files), and avoid typing or otherwise disturbing the Mathematica window while it is running a long task. Debug everything with small systems, directly save your graphs of larger systems to a file, and display them outside Mathematica. In either language,

and J. P. Sethna, ‘Critical Casimir forces in cellular membranes, *PRL* 109, 138101 (2012).)

<sup>142</sup>You can get one from Matt Bierbaum’s Web simulation at <http://mattbierbaum.github.io/ising.js/> of size at least 256x256, or download one from the course Web site (along with hints files) at <http://pages.physics.cornell.edu/~sethna/teaching/562/HW.html>.

512x512 will strain your machine; anything over 128x128 is acceptable, but larger systems will make the physics a bit clearer.

The hints files will allow you to import an Ising image, convert it into a two-dimensional array  $S_{mn} = \pm 1$ , and select an  $L \times L$  subregion (if you wish, especially while you debug your code). We imagine these spins spread over the unit square in the complex plane; our code generates a list of spins and square polygons associated to each, with the spin  $S_{mn}$  sitting at the center<sup>143</sup>  $z_{mn} = ((m + \frac{1}{2}) + i(n + \frac{1}{2})) / L$  of a  $1/L \times 1/L$  square. The code will allow you to provide a function  $f(z)$  of your choice, and will return the deformed quadrilaterals<sup>144</sup> centered at  $f(z_{mn})$  with areas  $A_{mn}$ , and their associated rescaled spins  $A_{mn}^{-1/16} S_{mn}$ .<sup>145</sup> The software will also provide example routines showing how to output and shade in these quadrilaterals.

(d) Generate a conformal transformation with a nonlinear analytic function of your choice, warping the Ising mode in an interesting way. Zoom in to regions occupied by lots of small pixels, where the individual pixels are not obvious.<sup>146</sup> Include both your entire plot and a cropped figure showing the expanded zoom. Discuss your zoomed plot critically – does it appear that the Ising correlations and visual structures have been faithfully retained by your analytic transformation?

(e) Load an Ising model equilibrated above the critical temperature at  $T = 100$  (random noise), and one at  $T = 3$  (short-range correlations). Distort and zoom each using your chosen conformal transformation. Analyze each and include your images. If you ‘blur your eyes’ enough

to ignore the individual pixels, can you tell how much the system has been dilated? Are your conformally transformed images faithfully retaining the correlations and visual structures away from the critical point? For  $T = 3$ , which regions look qualitatively like  $T_c$ ? Which regions look like  $T = 100$ ?

(f) Invent a non-analytic function, and use it to distort your Ising model. (Warning: most functions you write down, like  $\log(\cosh^4(z + 1/z^2))$  will be analytic except at a few singularities. The author tried two methods: inventing real-valued functions  $u(x, y)$  and  $v(x, y)$  and forming  $f = u + iv$ , and picking two different analytic functions  $g(z)$  and  $h(z)$  and using  $f(z) = \text{Re}(g(z)) + i\text{Im}(h(z))$ . Make sure your function is non-analytic almost everywhere (e.g., violates the Cauchy-Riemann equations), not just at a point.)<sup>147</sup> Find an example that makes for an interesting picture; include your images, including a zoom into a region with many pixels that range in size. As above, examine the images critically – does it appear that the Ising correlations and visual structures have been faithfully retained by your non-analytic transformation? Describe the distortions you see. (Are the pixels still approximately square?)

Conformal symmetry in two dimensions was studied in an outgrowth of string theory. The representations of the conformal group allowed them to deduce the exact critical exponents for all the usual two dimensional statistical mechanical models – reproducing Onsager’s result for the 2D Ising model, known solutions for the 2D tri-critical Ising model, 2D percolation, … More recently, field theorists [26] have used conformal

<sup>143</sup>If you work in Mathematica or Fortran, where the indices of arrays run from  $(1 \dots L)$ ,  $z_{mn} = ((m - \frac{1}{2}) + i(n - \frac{1}{2})) / L$ .

<sup>144</sup>The routine will drop quadrilaterals that extend to infinity, and also will remove quadrilaterals with ‘negative area’. The latter are associated with pixels which get inverted by  $f(z)$ ; plotting packages usually do not provide routines that color the exterior of a polygon.

<sup>145</sup>The list of quadrilaterals and spins will be linear, not two-dimensional, since graphics routines plotting polygons usually want flattened lists.

<sup>146</sup>You can zoom either using the graphics software or (sometimes more efficient) by saving a vector-graphics figure (like pdf) and viewing it separately. Start with  $L = 64$  or so to make plots and zooming efficient, but for your final plots use  $L$  as large as feasible.

<sup>147</sup>Warning: The program automatically drops quadrilaterals with ‘negative area’, which usually happen when an internal point goes to infinity (and the polygon should be shaded ‘outside’). This will also happen for the conjugate of an analytic function (e.g.,  $f(x + iy) = y + ix = i\bar{z}$ ); you will get either errors about Transpose[] in Mathematica or an empty plot in Python. If this happens for your choice of  $u(x, y) + iv(x, y)$ , try using  $u - iv$ .

invariance in higher dimensions (with a strategy called ‘conformal bootstrap’) to produce bounds for critical exponents. They now hold the record

on accuracy for the exponents of the 3D Ising model, giving  $\beta = 0.326419(3)$ ,  $\nu = 0.629971(4)$ , and  $\delta = 4.78984(1)$ .

# Advanced Statistical Mechanics

## 3

### Exercises

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

(3.1) **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<sup>1</sup>  $K = J/T$  and a next-neighbor interaction  $K_2 = J_2/T$  lying along the diagonal bonds.

$$\begin{aligned} \mathcal{H} = & -K \sum_{i,j} S_{i,j} S_{i+1,j} + S_{i,j} S_{i,j+1} \\ & - K_2 \sum_{i,j} S_{i,j} S_{i+1,j+1} + S_{i,j} S_{i+1,j-1} \end{aligned} \quad (3.1)$$

If we decimate to the ‘black’ squares of a checkerboard (say,  $i + j$  even), we get a new square-lattice Hamiltonian rotated by  $45^\circ$  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 in-

creased by  $K_2$ . So under one coarse-graining step

$$\begin{aligned} K' - K^* &= b^{1/\nu}(K - K^*) + K_2, \\ K'_2 &= 0. \end{aligned} \quad (3.2)$$

(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  $\lambda_0$  and  $\lambda_1$ ? (Let  $\lambda_1$  be the relevant eigenvalue, greater than one.)*

Our Jacobian matrix is not symmetric (or Hermitian), so it has two sets of eigenvectors – left eigenvectors  $\hat{\ell}_\alpha J = \lambda_\alpha \hat{\ell}_\alpha$ , 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?*

(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  $u_\alpha$ , given by the dot products of the displacement from the

<sup>1</sup>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.

fixed point with the left eigenvectors:

$$u_\alpha = \ell_\alpha \cdot (K - K^*, K_2) \quad (3.3)$$

- (d) Show that the phase boundary has  $u_1 = 0$  (using the convention that  $\lambda_1$  is the relevant direction). How do the coordinates  $u_\alpha$  flow under the renormalization group?

In general, there is a nonlinear transformation between the parameters  $T, H, J_2, \dots$  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?

**(3.2) Is the fixed point unique? Period doubling.** (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  $\phi$  is smooth, monotone increasing, has a smooth inverse  $\phi^{-1}$ .

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  $\tilde{g}(y)$  corresponding to the map  $g(x)$ , in terms of  $g$  and  $\phi$ . 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  $\phi$  and  $\tilde{g}$  is

$$g(x) = \phi^{-1} \tilde{g}(\phi(x)). \quad (3.4)$$

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

Recall from Exercises 12.9 and 12.26 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  $\alpha \approx -2.5$ :

$$T[g](x) = \alpha g(g(\alpha^{-1}x)). \quad (3.5)$$

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 transformation  $\tilde{T}$  for which  $\tilde{g}^*$  is a fixed point. (Hint: Use the formula  $T[g^*] = g^*$ , and the formula for  $g^*$  in terms of  $\tilde{g}^*$  and  $\phi$  from part (a). Change variables to  $y$  and solve for  $\tilde{g}(y)$  on the right-hand side.) Show that  $\tilde{T}$  can be viewed as a decimation plus a nonlinear ‘stretching’ function  $\tilde{\alpha}(y)$ ,  $\tilde{T} = \tilde{\alpha} \circ \tilde{g} \circ \tilde{g} \circ \tilde{\alpha}^{-1}(y)$ . What is  $\tilde{\alpha}(y)$ ?

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  $x \rightarrow \phi(x)$ .

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<sup>2</sup>  $g^{*[2^n]}(x_n) = g^*(g^*(\dots(x_n)\dots)) = x_n$ , with period 2, 4, 8, ... (Beware!  $g^*$  is a fixed point in function space under the transformation  $T$ .  $x^*$

<sup>2</sup>Remember the period doubling bifurcations, 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.

is a fixed point on the real line under the function  $g$ .)

(c) Show that  $dx/dy = \phi^{-1'}(\phi(x)) = 1/(dy/dx) = 1/\phi'(x)$ . If  $x^*$  is a fixed point of  $g^*$ , what is the fixed point  $y^*$  of  $\tilde{g}^*$ ? Find the slope  $d\tilde{g}^*/dy|_{y^*}$  at the new fixed point, and show that it equals the slope  $g'^*(x^*)$  at the old fixed point. Similarly, show that  $dg^{*[2^n]}/dx|_{x_n} = d\tilde{g}^{[2^n]}/dy|_{y_n}$ . (Note that  $dg(g(\dots(x)\dots))/dx = g'(g(\dots))g'(\dots)\dots$ , 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 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  $\tilde{g}^*$ . (It is a fixed point of  $\tilde{T}$  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  $\phi$  that are near to the identity:

$$\phi(x) = x + \epsilon \sum_{p=1}^{\infty} \phi_p x^p. \quad (3.6)$$

(d) Write  $\phi^{-1}(y)$  to linear order in  $\epsilon$ , as a similar sum. Calculate the change  $\tilde{g}^*(x) - g^*(x)$  to linear order in  $\epsilon$ . (Hint: Start with the equation  $\phi(g^*(x)) = \tilde{g}^*(\phi(x))$ .) 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  $g'^*(x) = g'^*(g(x/\alpha))g'^*(x/\alpha)$ . Calculate the change  $T[\tilde{g}^*](x) - g^*(x)$  to linear order in  $\epsilon$ . 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  $\alpha^{1-p}$ . You have also shown that only these eigenvalues and eigenfunctions can be generated by infinitesimal changes of coordinates. Feigenbaum conjectured in his early paper [18, 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  $\lambda_n$  of  $T$ , calculated for perturbations in the even subspace, are approximately

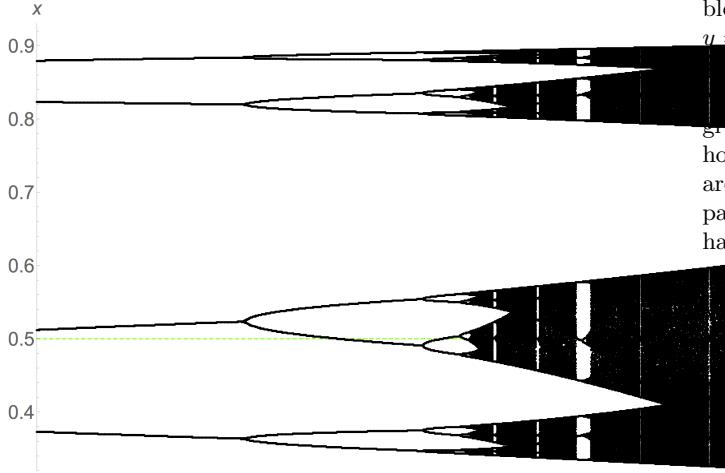
$$\{\lambda_1, \lambda_2, \dots\} = \{4.67, 1.00, 0.160, -0.124, -0.0573, 0.0255, -0.0101, \dots\} \quad (3.7)$$

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

$$\alpha^{1-p} = \{1.00, -0.400, 0.160, -0.0638, 0.0255, -0.0102, 0.00407, \dots\} \quad (3.8)$$

(f) Is  $\lambda_4$  a power of  $\alpha$ ? Was Feigenbaum correct? Give the next few eigenvalues  $\lambda_n$  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 what may be a new nomenclature. We shall call the first redundant critical points and the second (more common) category singular critical points. More generally, there are three sources of corrections to scaling. Smooth coordinate transformations in the control variables (here the parameter  $\mu$  in the logistic map) generate what we call analytic corrections to scaling; these can give corrections with integer power-laws (like the analytic background in the specific heat), or corrections that are combinations of integers and the main critical exponents. 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  $\lambda_4$  in part (f)). Finally, there are redundant corrections to scaling, which correspond to systems that could have been fixed points of a different renormalization group. The redundant corrections to scaling will involve combinations of powers of relevant critical exponents.

How in practice can we tell the special ‘redundant’ critical points from the ‘singular’ ones?



**Fig. 3.1 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  $\alpha$  and horizontally by  $\delta$ ; 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. 3.1); for example, the ratio  $\alpha_n$  of the widths of the  $n^{th}$  and  $(n+1)^{st}$  tier branch will not be precisely  $\alpha$ . Let the  $\psi_n(x)$  be the  $n^{th}$  eigenfunction of  $T$ .<sup>3</sup>

(g) Consider an infinitesimal perturbation of the fixed-point  $g^* + \epsilon\psi_1(x)$  along this direction. How does the leading correction to  $\alpha_n - \alpha$  scale with  $\lambda_n$ ? (Hint: How does  $\alpha_n$  change under one application of the renormalization group operator  $T$ ?)

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

#### Conjectures:

- (a) The allowed renormalization-group fixed point functions are precisely those accessi-

ble by analytic coordinate transformations  $y = \phi(x)$ , which are precisely those generated by moving along the scaling variables associated with the eigenvalues  $\lambda_n$  not given by powers of  $\alpha$  (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  $\alpha$ .

We conjecture that all corrections to scaling for one-humped maps [30] will be given by changing to ‘normal form’ coordinates. 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  $u_\mu(\mu) = \mu - \mu_\infty + b_1(\mu - \mu_\infty)^2 + \dots$ ,  $u_2(\mu) = a_2 + b_2(\mu - \mu_\infty) + c_2 \dots, \dots$ . We are led to conjecture that a further change of coordinates  $x \rightarrow \phi(x)$  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 redundant submanifold in the surface of critical points, but that the critical points can be ‘foliated’ into surfaces where the singular corrections are the same.

- (c) We conjecture that the separation between redundant 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 [20], which answers the analogous question as to whether there is a renormalization group that can make any Ising critical point into its fixed point.
- (d) 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

<sup>3</sup>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  $\psi_2(x) = \Psi_1(x) = \psi_1(x)$ .

cal entropy  $S(E, H)$  or the Helmholtz free energy  $A(T, M)$ . We conjecture that Legendre transformations will convert the analytic corrections to redundant 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.

- (e) 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.)
  - (f) Many of the ideas here were prompted by an analysis of the onset of chaos from quasiperiodic motion [37, 43]. 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 redundant and singular components.
- (3.3) **A fair split? Number partitioning.<sup>4</sup>** (Computer science, Mathematics, Computation, 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 [32]  $S_1 = [10, 13, 23, 6, 20]$ ,  $S_2 = [6, 4, 9, 14, 12, 3, 15, 15]$ ,  $S_3 = [93, 58, 141, 209, 179, 48, 225, 228]$ , and  $S_4 = [2474, 1129, 1388, 3752, 821, 2082, 201, 739]$ . Hint:  $S_1$  has a balanced partition, and  $S_4$  has a minimum 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 can be found. (A random instance with  $N = 2$  and  $M = 10$  has a one chance in  $2^{10} = 1024$  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

<sup>4</sup>This exercise draws heavily from [32, chapter 7]. Hints for the computations can be found at the book Web site [49].

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  $\{1, \dots, 2^M\}$ , 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 `pperf` that can be partitioned evenly (zero cost). Plot `pperf` 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  $\kappa_c$  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)<sup>5</sup> 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 non-zero energy in the thermodynamic limit.

We can understand both the value  $\kappa_c$  of the phase transition and the form of  $p_{\text{perf}}(N, M)$  by studying the distribution of possible ‘signed’ costs  $E_s = \sum_j a_j s_j$ . These energies are distributed over a maximum total range of  $E_{\max} - E_{\min} = 2 \sum_{j=1}^N a_j \leq 2N 2^M$  (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 standard deviation  $\sigma$  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  $a_j \in [1, \dots, 2^M]$ , 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 non-zero 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

<sup>5</sup>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.

approximation.<sup>6</sup> We can use the random-energy approximation to estimate  $p_{\text{perf}}$  that you plotted in part (b).

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

$$p_{\text{perf}}(\kappa, N) \approx 1 - \exp \left[ -\sqrt{\frac{3}{2\pi N}} 2^{-N(\kappa-\kappa_c)} \right]. \quad (3.9)$$

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

$$p_{\text{perf}}^{\text{scaling}}(x) = 1 - \exp \left[ -\sqrt{\frac{3}{2\pi}} 2^{-x} \right] \quad (3.10)$$

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

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

Rigorous results show that this random-energy approximation gives the correct value of  $\kappa_c$ . The entropy of zero-cost states below  $\kappa_c$ , the probability distribution of minimum costs above  $\kappa_c$  (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.<sup>7</sup> It also lends credibility to the conjecture in the computer science community that  $\mathbf{P} \neq \mathbf{NP}$ -complete; any polynomial-time algorithm for NPP would have to ingeniously make use of the seemingly unimportant correlations between energy levels.

#### (3.4) Cardiac dynamics.<sup>8</sup> (Computation, Biology, Complexity) ④

**Reading:** References [38, 58], 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

<sup>6</sup>More precisely, we ignore correlations between the energies of different teams  $\mathbf{s} = \{s_i\}$ , except for swapping the two teams  $\mathbf{s} \rightarrow -\mathbf{s}$ . This leads to the  $N - 1$  in the exponent of the exponent for  $p_{\text{perf}}$  in part (f). Notice that in this approximation, NPP is a form of the random energy model (REM, exercise 3.17), except that we are interested in states of energy near  $E = 0$ , rather than minimum energy states.

<sup>7</sup>The computational cost does peak near  $\kappa = \kappa_c$ . For small  $\kappa \ll \kappa_c$  it’s 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  $\kappa$  grows [32, fig 7.3].

<sup>8</sup>This exercise and the associated software were developed in collaboration with Christopher Myers.

to excite a new pulse, followed by a return to the resting state. The FitzHugh-Nagumo equations provide a simplified model for the excitable heart tissue:<sup>9</sup>

$$\begin{aligned}\frac{\partial V}{\partial t} &= \nabla^2 V + \frac{1}{\epsilon}(V - V^3/3 - W) \\ \frac{\partial W}{\partial t} &= \epsilon(V - \gamma W + \beta),\end{aligned}\quad (3.11)$$

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  $dV/dt$  and  $dW/dt$  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  $\Delta$  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,  $\partial^2 V / \partial x^2 \approx (V(x+dx) - 2V(x) + V(x-dx)) / dx^2$ :

$$\begin{aligned}\nabla_{[5]}^2 V(x_i, y_i) &\approx (V(x_i, y_{i+1}) + V(x_i, y_{i-1}) \\ &\quad + V(x_{i+1}, y_i) + V(x_{i-1}, y_i) \\ &\quad - 4V(x_i, y_i)) / dx^2 \\ &\leftrightarrow \frac{1}{dx^2} \begin{pmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{pmatrix}\end{aligned}\quad (3.12)$$

where  $dx = x_{i+1} - x_i = y_{i+1} - y_i$  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

$$\nabla_{[9]}^2 V(x_i, y_i) \leftrightarrow \frac{1}{dx^2} \begin{pmatrix} 1/6 & 2/3 & 1/6 \\ 2/3 & -10/3 & 2/3 \\ 1/6 & 2/3 & 1/6 \end{pmatrix}. \quad (3.13)$$

We will simulate our partial-differential equation (PDE) on a square  $100 \times 100$  grid with a grid spacing  $dx = 1$ .<sup>10</sup> 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 text web site [49].

(c) *Solve eqn 3.11 for an initial condition equal to the fixed-point  $(V^*, W^*)$  except for a  $10 \times 10$  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  $10 \times 10$  square). Real-

<sup>9</sup>Nerve tissue is also an excitable medium, modeled using different Hodgkin-Huxley equations.

<sup>10</sup>Smaller grids would lead to less grainy waves, but slow down the simulation a lot.

istically, 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  $\Delta$ . How does the system go unstable? Repeat this process, reducing  $\Delta$  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  $dx^2$ , the lattice spacing squared—thus to make  $dx$  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  $dx$  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 [16].

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 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.

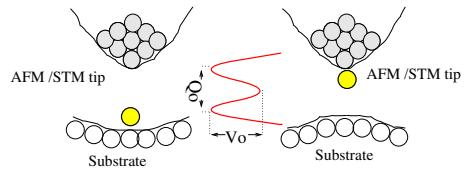
(e) Add the code for interactively stimulating a general rectangle with an increment to  $V$  of size  $\Delta = 3$ . Play with generating rectangles in different places while other pulses are going by: make some spiral waves. Clear the spirals by giving a stimulus that spans the system.

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 web site.

(3.5) **Quantum dissipation from phonons.**  
(Quantum) ②

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  $(V/2\pi)^3 d^3 k$ ).



**Fig. 3.2 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 non-zero 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.<sup>11</sup> 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).

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:

$$\mathcal{H} = \sum_k p_k^2/2m + 1/2 m\omega_k^2 q_k^2 + \mathbf{F} \cdot \mathbf{u}_0. \quad (3.14)$$

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  $u_n = (1/\sqrt{N}) \sum_k q_k \exp(-ikx_n)$  (with  $N$  the number of atoms), so  $q_k = (1/\sqrt{N}) \sum_n u_n \exp(ikx_n)$ .

*Substituting for  $u_0$  in the Hamiltonian and completing the square, find the displacement  $\Delta_k$  of each harmonic oscillator. 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  $\exp(-\Delta_k^2/8a_k^2)$  (with  $a_k = \sqrt{\hbar/2m\omega_k}$  the zero-point motion in that mode). Converting the product to the exponential of a sum, and the sum to an integral  $\sum_k \sim (V/(2\pi)^3 \int dk)$ , do we observe an overlap catastrophe?*

### Mean Field: Introduction

Mean field theory can be derived and motivated in several ways.

- (a) The interaction (field) from the neighbors can be approximated by the average (mean) field of all sites in the system, effectively as described in Cardy section 2.1. This formulation makes it easy to understand why mean-field theory becomes more accurate in high dimensions and for long-range interactions, as a spin interacts with more and more neighbors.
- (b) The free energy can be bounded above by the free energy of a noninteracting mean-field model. This is based on a variational principle I first learned from Feynman (Exercise 12.24, based on Cardy's exercise 2.1).
- (c) The free energy can be approximated by the contribution of one order parameter configuration, ignoring fluctuations. (From a path-integral point of view, this is a 'zero-loop' approximation.) For the Ising model, one needs first to change away from the

Ising spin variables to some continuous order parameter, either by coarse-graining (as in Ginzburg-Landau theory, Sethna 9.5 below) or by introducing Lagrange multipliers (the Hubbard-Stratonovich transformation, not discussed here).

- (d) The Hamiltonian can be approximated by an infinite-range model, where each spin does interact with the average of all other spins. Instead of an approximate calculation for the exact Hamiltonian, this is an exact calculation for an approximate Hamiltonian – and hence is guaranteed to be at least a sensible physical model. See Exercise 12.25 for an application to avalanche statistics.
- (e) The lattice of sites in the Hamiltonian can be approximated as a branching tree (removing the loops) called the *Bethe lattice* (not described here). This yields a different, solvable, mean-field theory, which ordinarily has the same mean-field critical exponents but different non-universal features.

### (3.6) Ising lower critical dimension. (Dimension dependence) ③

What is the lower critical dimension of the Ising model? If the total energy  $\Delta E$  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<sup>12</sup> 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  $L \rightarrow \infty$ ? 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.

(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  $\xi$  (the distance between domain

<sup>12</sup>Energy, not free energy! Think about  $T = 0$ .

walls) grow as  $T \rightarrow T_c = 0$ ? (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  $\nu$  satisfying  $\xi \sim (T - T_c)^{-\nu}$  is  $1$ ,  $\sim 0.63$ , and  $1/2$  in dimensions  $D = 2, 3$ , and  $\geq 4$ , respectively. Is there an exponent  $\nu$  governing this divergence in one dimension? How does  $\nu$  behave as  $D \rightarrow 1$ ?

- (3.7) **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

$$-J \mathbf{S}_i \cdot \mathbf{S}_j. \quad (3.15)$$

Estimate the energy needed to twist the spins at one boundary  $180^\circ$  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  $L \rightarrow \infty$ ? 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  $L \rightarrow \infty$  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 paramagnetic 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  $180^\circ$  costs a free energy that goes to a constant as  $L \rightarrow \infty$ . 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.

- (3.8) **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

$$\mathcal{H} = \sum_{i>j} \frac{J}{|i-j|^\sigma} S_i S_j. \quad (3.16)$$

For what values of  $\sigma$  will a domain wall between up and down spins have finite energy? Suggest a bound on  $\sigma$  analogous to the lower critical dimension (the maximum power  $\sigma$ , 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.

(3.9) **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/sehna/Crystal\\_Shapes/Equilibrium\\_Crystal\\_Shapes.html](http://www.lassp.cornell.edu/sehna/Crystal_Shapes/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 interactions (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  $\theta$  between up spins and down spins. Show that the energy cost per unit length of the interface is

$$\gamma_0(\theta) = 2J(\cos(|\theta|) + \sin(|\theta|)), \quad (3.17)$$

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  $\ell$  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

<sup>13</sup>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.

<sup>14</sup>See Burton, Cabrera, and Frank, Appendix D. This is their equation D4, with  $p \propto \gamma(\theta)$  given by equation D7.

<sup>15</sup>This is the emergent spherical symmetry at the Ising model critical point.

that the entropy per unit length is

$$\begin{aligned} s_0(\theta) = & k_B \left( (\cos(|\theta|) + \sin(|\theta|)) \right. \\ & \log(\cos(|\theta|) + \sin(|\theta|)) \\ & - \cos(|\theta|) \log(\cos(|\theta|)) \quad (3.18) \\ & \left. - \sin(|\theta|) \log(\sin(|\theta|)) \right). \end{aligned}$$

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” Badri-narayan 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’s unstable.<sup>13</sup> The key is to do a simulation below  $T_c$ , but with a net (conserved) negative magnetization slightly closer to zero than 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

$$\begin{aligned} x &= \cos(\theta)\gamma(\theta, T) - \sin(\theta)\frac{d\gamma}{d\theta} \\ y &= \sin(\theta)\gamma(\theta, T) + \cos(\theta)\frac{d\gamma}{d\theta} \end{aligned}$$

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.<sup>14</sup>

(c) Wolff construction. Using the energy of eqn 3.17 and approximating the entropy at low temperatures with the zero-temperature form eqn 3.18, plot the Wulff shape 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<sup>15</sup> 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.<sup>16</sup>

<sup>16</sup>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.



# Numerical Methods

# 4

## Exercises

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.

- (4.1) **Condition number and accuracy.**<sup>1</sup> (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

$$A = \begin{pmatrix} 0.780 & 0.563 \\ 0.913 & 0.659 \end{pmatrix} \quad \text{and} \quad \mathbf{b} = \begin{pmatrix} 0.217 \\ 0.254 \end{pmatrix}. \quad (4.1)$$

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  $\mathbf{r} = \mathbf{b} - A\mathbf{x}$ .) Does the more accurate solution have the smaller residual?

(b) Compute the condition number<sup>2</sup> 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  $2.2 \times 10^{-16}$ ? Is the problem the accuracy of the solution, or rounding errors in calculating  $A$  and  $b$ ? (Hint: try calculating  $A\mathbf{x} - \mathbf{b}$ .)

- (4.2) **Sherman–Morrison formula.** (Numerical) ③

Consider the 5x5 matrices

$$T = \begin{pmatrix} E - t & t & 0 & 0 & 0 \\ t & E & t & 0 & 0 \\ 0 & t & E & t & 0 \\ 0 & 0 & t & E & t \\ 0 & 0 & 0 & t & E - t \end{pmatrix} \quad (4.2)$$

and

$$C = \begin{pmatrix} E & t & 0 & 0 & t \\ t & E & t & 0 & 0 \\ 0 & t & E & t & 0 \\ 0 & 0 & t & E & t \\ t & 0 & 0 & t & E \end{pmatrix}. \quad (4.3)$$

These matrices arise in one-dimensional models of crystals.<sup>3</sup> The matrix  $T$  is *tridiagonal*: its entries are zero except along the central diagonal

<sup>1</sup> Adapted from Saul Teukolsky, 2003.

<sup>2</sup> 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.

<sup>3</sup> 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.

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<sup>4</sup> we can find  $C^{-1}$  from  $T^{-1}$  efficiently using the Sherman-Morrison formula (section 2.7).

*Compute the inverse<sup>5</sup> 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*

$$\Delta = \frac{T^{-1}\mathbf{u} \otimes \mathbf{v}\mathbf{T}^{-1}}{1 + \mathbf{v} \cdot T^{-1} \cdot \mathbf{u}}. \quad (4.4)$$

#### (4.3) 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 Figure 4.1 corresponds with which method?

#### (4.4) Numerical definite integrals. (Numerical) ③

In this exercise we will integrate the function you graphed in the first, warmup exercise:

$$F(x) = \exp(-6 \sin(x)). \quad (4.5)$$

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  $\pi$ . Compare your answer to the analytic integral<sup>6</sup> ( $\approx 0.34542493760937693$ ) 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  $\text{Trap}(h)$  for  $N+1$  points spaced at  $h = \pi/N$ , with  $N = 1, 2, 4, \dots, 2^{10}$ . Plot the estimated integral versus the spacing  $h$ . Does it extrapolate smoothly to the true value as  $h \rightarrow 0$ ? With what power of  $h$  does the error vanish? Replot the data as  $\text{Trap}(h)$  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  $\text{Trap}(h)$  versus  $h^2$  plot. Notice that the extrapolation to  $h \rightarrow 0$  is  $A$ , and show that  $A$  is  $4/3 \text{Trap}(h_0) - 1/3 \text{Trap}(2h_0)$ . 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  $\{h^2, \text{Trap}(h)\}$  for  $h = \pi/2, \dots, \pi/2^M$ , 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?

(e) *Gaussian Quadrature.* Implement Gaussian quadrature with  $N$  points optimally chosen on

<sup>4</sup>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 non-zero 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).

<sup>5</sup>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's not important.

<sup>6</sup> $\pi(\text{Bessel}[0, 6] - \text{StruveL}[0, 6])$ , according to Mathematica.

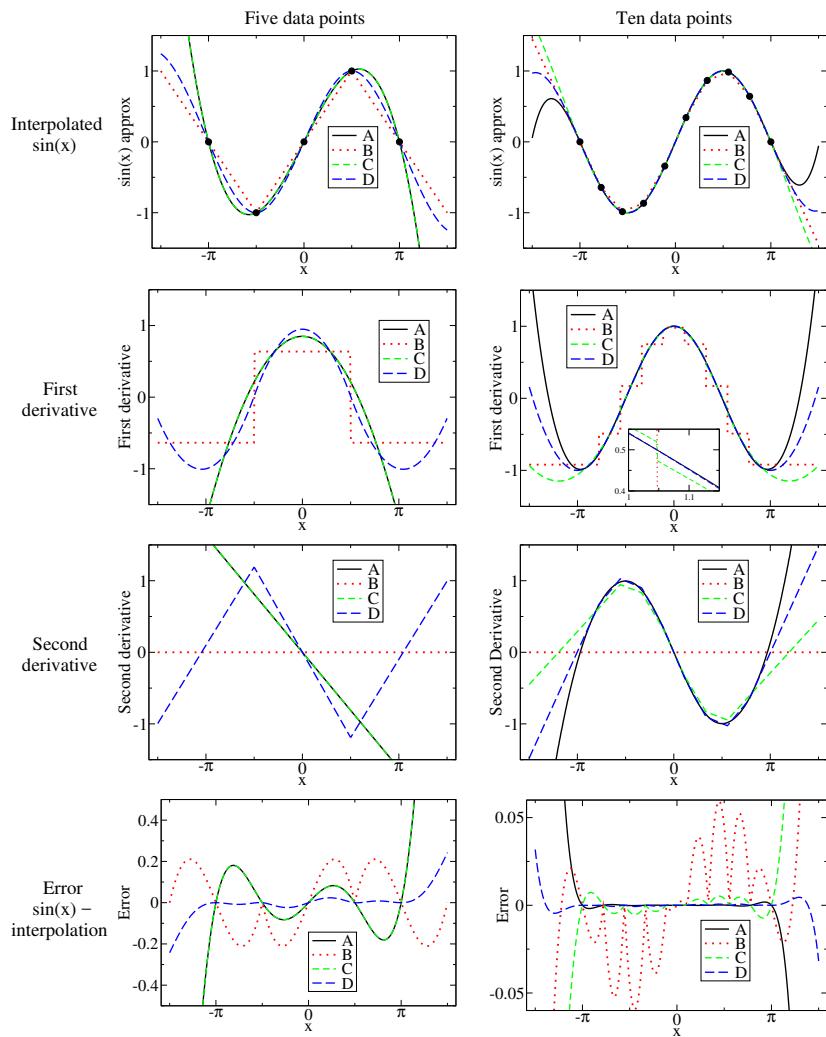


Fig. 4.1 Interpolation methods.

the interval  $(0, \pi)$ , with  $N = 1, 2, \dots, 5$ . (You may find the points and the weights appropriate for integrating functions on the interval  $(-1, 1)$  on the course Web site; 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  $2\pi$ , 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.

(4.5) **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

$$\frac{d^2y}{dx^2} \sim \frac{y(x+h) - 2y(x) + y(x-h)}{h^2}, \quad (4.6)$$

again optimizing  $h$  and making it exactly representable. Estimate your error again, and compare to the observed error.

(4.6) **Summing series.** (Efficiency) ②

Write a routine to calculate the sum

$$s_n = \sum_{j=0}^n (-1)^j \frac{1}{2j+1}. \quad (4.7)$$

As  $n \rightarrow \infty$ ,  $s_\infty = \pi/4$ . About how many terms do you need to sum to get convergence to within  $10^{-7}$  of this limit? Now try using Aitken's  $\Delta^2$  process to accelerate the convergence:

$$s'_n = s_n - \frac{(s_{n+1} - s_n)^2}{s_{n+2} - 2s_{n+1} + s_n}. \quad (4.8)$$

About how many terms do you need with Aitken's method to get convergence to within  $10^{-7}$ ?

(4.7) **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?

(4.8) **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  $V(f)$  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

$$y(x) = \exp(-6 \sin(x)) \quad (4.9)$$

over  $0 \leq x < 2\pi$ . (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  $\langle y^2 \rangle = (1/2\pi) \int_0^{2\pi} y^2(x) dx \approx 18948.9$ .)

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:

$$g(x) = \frac{\epsilon}{2\pi} + \frac{1-\epsilon}{\sqrt{\pi/3}} \exp(-6(x - 3\pi/2)^2/2). \quad (4.10)$$

I found minimum variance around  $\epsilon = 0.005$ .

(c) Importance Sampling (Optional for 480). Generate 1000 random numbers with probability distribution  $g$ .<sup>7</sup> Use these to estimate the integral of  $y(x)$ . How accurate is your answer?

<sup>7</sup>Take  $(1 - \epsilon)M$  Gaussian random numbers and  $\epsilon M$  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.

<sup>8</sup>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.

(4.9) **Washboard potential.** (Solving) ②

Consider a washboard potential<sup>8</sup>

$$V(r) = A_1 \cos(r) + A_2 \cos(2r) - Fr \quad (4.11)$$

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 bifurcation.) 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  $r_c$ .

(4.10) **Sloppy minimization.** (Statistics) ③

“With four parameters I can fit an elephant. With five I can make it waggle its 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 *slopiness* – the parameters in the model are poorly constrained by the data.

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  $\theta_\alpha$ , the radioactivity that you would measure is

$$y_{\theta}(t) = \sum_{\alpha=0}^{N-1} \theta_\alpha \exp(-\theta_\alpha t). \quad (4.12)$$

Now, suppose you do not know the decay rates  $\theta_\alpha$ . 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  $\theta_0 = [1, 2]$  (so the data fall on the curve  $d(t) = \exp(-t) + 2\exp(-2t) = y_{\theta_0}(t)$ ). 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

$$C[\theta] = \int_0^\infty (y_\theta(t) - d(t))^2 dt. \quad (4.13)$$

- (a) Draw a contour plot of  $C$  in the square  $0.5 < \theta_\alpha < 2.5$ , with contours at  $C = \{2^{-12}, 2^{-11}, \dots, 2^0\}$ . You may need to 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  $\theta$ . 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  $\theta = \theta_0$ , show that the cost Hessian is a continuous integral

$$H_{\alpha\beta} = (J^T J)_{\alpha\beta} = J_{t\alpha} J_{t\beta} = \int_0^\infty J(t, \alpha) J(t, \beta) dt \quad (4.14)$$

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  $H(\theta)$  by doing the indefinite integral in eqn 4.14. Find the eigenvalues and the eigenvectors for the cost Hessian  $H$  for your plot in part (b), evaluated at  $\theta_0$ , and check them against your con-

tour 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  $\theta_0 = [1, 2, 3, \dots, 9]$ , construct the Hessian, find its eigenvalues. By what factor does each successive eigenvalue shrink? Are they sloppy (roughly equally spaced in log)?

(4.11) **Sloppy monomials.**<sup>9</sup> (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<sup>10</sup>) fit:

$$\begin{aligned} f(x) &= \sin(2\pi x) \\ f_{\text{Taylor}} &\approx 0.000 + 6.283x + 0.000x^2 - 41.342x^3 \\ &\quad + 0.000x^4 + 81.605x^5 \\ f_{\text{Chebyshev}} &\approx 0.007 + 5.652x + 9.701x^2 - 95.455x^3 \\ &\quad + 133.48x^4 - 53.39x^5 \\ f_{\text{Legendre}} &\approx 0.016 + 5.410x + 11.304x^2 - 99.637x^3 \\ &\quad + 138.15x^4 - 55.26x^5 \end{aligned}$$

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

$$\begin{aligned} \text{Chebyshev} - \text{Legendre} &= (4.15) \\ &- 0.009 + 0.242x - 1.603x^2 \\ &+ 4.182x^3 - 4.67x^4 + 1.87x^5 \end{aligned}$$

a polynomial with coefficients two hundred times larger than the maximum difference!

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  $C^{\text{poly}}$ , the integral of

<sup>9</sup>Thanks to Joshua Waterfall, whose research is described here.

<sup>10</sup>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  $[0, 1]$ .

the squared difference between the polynomial and the function:

$$\begin{aligned} C^{\text{poly}} &= (1/2) \int_0^1 (f(x) - y_{\theta}(x))^2 dx, \\ y_{\theta}(x) &= \sum_{m=0}^{M-1} \theta_m x^m \end{aligned} \quad (4.16)$$

How quickly does this cost increase as we move the  $M$  parameters  $\theta_m$  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 4.15.

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...

(a) Note that the first derivative of the cost  $C^{\text{poly}}$  is zero at the best fit. Show that the Hessian second derivative of the cost is

$$H_{mn}^{\text{poly}} = \frac{\partial^2 C^{\text{poly}}}{\partial \theta_m \partial \theta_n} = \frac{1}{m+n+1}. \quad (4.17)$$

This Hessian is the Hilbert matrix, famous for being ill-conditioned (having a huge range of eigenvalues).<sup>11</sup> Tiny eigenvalues of  $H^{\text{poly}}$  correspond to directions in polynomial space where the fit does not change.

(b) Calculate the eigenvalues of the  $6 \times 6$  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)? Calculate the eigenvalues of larger Hilbert matrices. At what size do your eigenvalues seem contaminated by rounding errors?

Notice from Eqn 4.17 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.

(c) Calculate the eigenvector corresponding to the smallest eigenvalue of  $H^{\text{poly}}$ , checking to

make sure its norm is one (so the coefficients are of order one). Plot the corresponding polynomial in the range  $[0, 1]$ : does it stay small everywhere in the interval? Plot it in a larger range  $[-1, 2]$  to contrast its behavior inside and outside the fit interval.

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  $\theta_n$  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. Indeed, if we change basis from the coefficients  $\theta_n$  of the monomials  $x^n$  to the coefficients  $\ell_n$  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  $\ell_n$  are thus well determined by the data, and indeed the Hessian for the cost  $C^{\text{poly}}$  in terms of this new basis is the identity matrix.

**(4.12) 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

$$V(y) = (1/8) y^2 (-4A^2 + \log^2(y^2)). \quad (4.18)$$

That is,

$$\begin{aligned} \frac{d^2y}{dt^2} &= -dV/dy \\ &= -(1/4)y(-4A^2 + 2\log(y^2) + \log^2(y^2)). \end{aligned} \quad (4.19)$$

<sup>11</sup>For fits to discrete data, at points  $(x_0, \dots, x_{N-1})$ , the Hessian  $H^{\text{poly}} = V^T V$ , where  $V_{i\alpha} = x_i^\alpha$  is the Vandermonde matrix, also famous for being ill-conditioned.

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<sup>12</sup>

$$F(t) = \exp(-6 \sin(t)). \quad (4.20)$$

Note that the potential energy  $V(y)$  is zero at the five points  $y = 0, y = \pm \exp(\pm A)$ .

(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  $4\pi$ , 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 blackbox 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 nu-

merical 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:

$$y(t+h) = 2y(t) - y(t-h) + h^2 f[y(t)] \quad (4.21)$$

where  $f[y] = -dV/dy$ . Given an initial position  $y(0)$  and an initial velocity  $v(0)$ , one can initialize and finalize

$$\begin{aligned} y(h) &= y(0) + h(v(0) + (h/2)f[y(0)]) \\ v(T) &= (y(T) - y(T-h))/h + (h/2)f[y(T)] \end{aligned} \quad (4.22)$$

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:

$$\begin{aligned} v(h/2) &= v(0) + (h/2)f[y(0)] \\ y(h) &= y(0) + h v(h/2) \\ v(t+h/2) &= v(t-h/2) + h f[y(t)] \\ y(t+h) &= y(t) + h v(t+h/2) \end{aligned} \quad (4.23)$$

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 4.21 and 4.22 converge to the true solutions as as  $h \rightarrow 0$  – 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?<sup>13</sup> Implement Leapfrog, and apply it to solving eqn 4.19 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

<sup>12</sup>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  $y = F(t)$ .

<sup>13</sup>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  $N^2h^n \sim h^{N-2}$ .

*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 4.23 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).



# Quantum

# 5

## Exercises

These exercises were designed for a graduate quantum mechanics course at Cornell.

(5.1) **Quantum notation.** (Notation) ②

For each entry in the first column, match the corresponding entries in the second and third column.

Schrödinger

$$\begin{aligned}\psi(x) \\ \psi^*(x) \\ |\psi(x)|^2 \\ \int \phi^*(x)\psi(x)dx \\ |\int \phi^*(x)\psi(x)dx|^2 \\ \int |\psi(x)|^2 dx \\ (\hbar/2mi)(\psi^*\nabla\psi - \psi\nabla\psi^*)\end{aligned}$$

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  $\psi$  is in state  $\phi$
- IV. Amplitude of  $\psi$  at  $x$
- V. Amplitude of  $\psi$  in state  $\phi$
- VI. Current density
- VII. None of these.

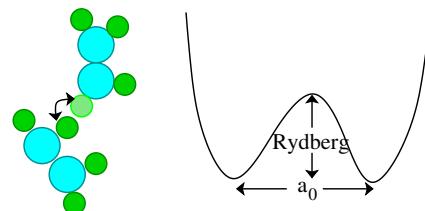
(5.2) **Light proton atomic size.** (Dimensional analysis) ③

In this exercise, we examine a parallel world where the proton and neutron masses are equal to the electron mass, instead of  $\sim 2000$  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:  $[\hbar] = ML^2/T$ ,  $[ke^2] = \text{Energy} * L = ML^3/T^2$ , and  $[m_e] = M$ . Here  $k = 1$  in CGS units, and  $k = 1/(4\pi\epsilon_0)$  in SI units.)

(5.3) **Light proton tunneling.** (Dimensional analysis) ③



**Fig. 5.1 Atom tunneling.** A hydrogen atom tunnels a distance  $a_0$ , breaking a bond of strength  $E_{\text{bind}}$  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  $\sim 2000$  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  $\exp(-\sqrt{2M\bar{V}}Q/\hbar)$ ).

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 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  $M_p \rightarrow m_e$ ? (Simplify your answer as much as possible.)* (Use the real-world<sup>1</sup>  $a_0$  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.)

(5.4) **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  $\sim 2000$  times larger.

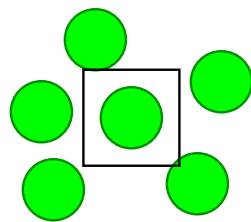


Fig. 5.2 Atoms with imaginary box of size equal to average space per atom

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  $E_{He,ours}$  of Helium in our world and water  $E_{H_2O,light}$  in the parallel world. Helium goes superfluid at  $\sim 2^\circ K$  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.<sup>2</sup> Room temperature is about  $300^\circ K$ . 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.)

(5.5) **Aharonov-Bohm Wire.** (Parallel transport) ③

What happens to the electronic states in a thin metal loop as a magnetic flux  $\Phi_B$  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  $\Phi_B/\Phi_0$ , with  $\Phi_0 = hc/e$  the 'quantum of flux' familiar from the Bohm-Aharonov 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

<sup>1</sup>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.

<sup>2</sup>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.

found (noisier) periods of  $\Phi_0$ . 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  $\Phi_B$ , thin compared to  $R$ , lie along the  $\hat{z}$  axis. Let  $\phi$  be the angle around the circle with respect to the positive  $x$ -axis. (Don't confuse the flux  $\Phi_B$  with the angle  $\phi$ !) 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  $\vec{A}$  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  $\psi$  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  $\vec{A}' = \vec{A} + \nabla\Lambda$  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  $\psi'$  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  $\Phi_B$ . Plot the current in the wire as a function of  $\Phi_B$ . Is it periodic with period  $\Phi_0$ , 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  $hc/e$ .

### (5.6) 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. 5.3). This means that the wavefunction for these particles must return to itself, up to an overall phase, when the particles are permuted:

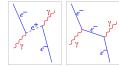
$$\Psi(\mathbf{r}_1, \mathbf{r}_2, \dots) = \exp(i\chi)\Psi(\mathbf{r}_2, \mathbf{r}_1, \dots). \quad (5.1)$$

where  $\dots$  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

<sup>3</sup> 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*

electrons on the left are not only indistinguishable, they are the *same particle!* The antiparticle is also the electron, traveling backward in time.<sup>3</sup>



**Fig. 5.3 Feynman diagram: identical particles.**

In three dimensions,  $\chi$  must be either zero or  $\pi$ , corresponding to bosons and fermions. In two dimensions, however,  $\chi$  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  $\phi$  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  $\phi$  at fixed  $r$ . When  $\phi = 180^\circ \equiv \pi$ , the particles have exchanged positions, leading to a boundary condition on the wavefunction

$$\Psi(\mathbf{R}, r, \phi, \dots) = \exp(i\chi)\Psi(\mathbf{R}, r, \phi + \pi, \dots). \quad (5.2)$$

Permuting them counter-clockwise (backward along the same path) must then<sup>4</sup> give  $\Psi(\mathbf{R}, r, \phi, \dots) = \exp(-i\chi)\Psi(\mathbf{R}, r, \phi - \pi, \dots)$ . This in general makes for a many-valued wavefunction (similar to Riemann sheets for complex analytic functions).

Why can't we get a general  $\chi$  in three dimensions?

(a) *Show, in three dimensions, that  $\exp(i\chi) = \pm 1$ , by arguing that a counter-clockwise rotation and a clockwise rotation must give the same phase.* (Hint: The phase change between  $\phi$  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  $\exp(i\chi)?)$



**Fig. 5.4 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 5.4 illustrates how in two dimensions rotations by  $\pi$  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).

An angular boundary condition multiplying by a phase should seem familiar: it's 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  $\Phi$  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  $\Phi$ . If a composite particle of charge  $e$  and flux  $\Phi$  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*

*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."*

<sup>4</sup>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.

*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*  $\nu$  (electrons per unit flux in units of  $\Phi_0$ ) 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  $\Phi_0/\nu$ . As a perturbation, we can imagine later relaxin the binding and allow the field to spread uniformly.<sup>5</sup>

(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 *non-abelian* statistics, which could become useful for quantum computation.
- The composite particle picture is a central tool both conceptually and in calculations for this field.

<sup>5</sup>This is not nearly as crazy as modeling metals and semiconductors as non-interacting electrons, and adding the electron interactions later. We do that all the time – 'electrons 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 non-interacting 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.

<sup>6</sup>This exercise was developed by Paul Ginsparg, based on an example by Bell '64 with simplifications by Clauser, Horne, Shimony, & Holt ('69).

<sup>7</sup>There's another version for GHZ state, where three people have to get  $a+b+c \bmod 2 = x$  or  $y$  or  $z$ . Again one can achieve only 75% success classically, but they can win every time sharing the right quantum state

(5.7) **Bell.**<sup>6</sup> (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  $(a+b = xy)\bmod 2$ . 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 theres 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  $\leq 75\%$  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.<sup>7</sup>

Suppose Alice and Bob share the entangled state  $\frac{1}{\sqrt{2}}(|\uparrow\rangle_e|\uparrow\rangle_r + |\downarrow\rangle_e|\downarrow\rangle_r)$ , 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  $3/4$ .

(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.).<sup>8</sup>

(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  $R_{\alpha_0}$  or  $R_{\alpha_1}$ , according to whether  $x$  is 0 or 1, and Bob applies  $R_{\beta_0}$  or  $R_{\beta_1}$ , 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.

### (5.8) 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), 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  $90^\circ$  W, latitude  $\sim 40^\circ$  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  $\sim 0^\circ$  W) down

<sup>8</sup>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.

to the equator, turn right, walk back to longitude  $90^\circ W$  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 ( $\sim 40^\circ N$ ) to St. Louis (yes, it's 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  $(D_i v)^j = \partial_i v^j + \Gamma_{ik}^j v^k$  involving the Christoffel symbol  $\Gamma_{ik}^\mu$ . The amount a vector  $v^\nu$  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

$$v'^i - v^i = R_{jk\ell}^i v^j \Delta s^k \Delta t^\ell. \quad (5.3)$$

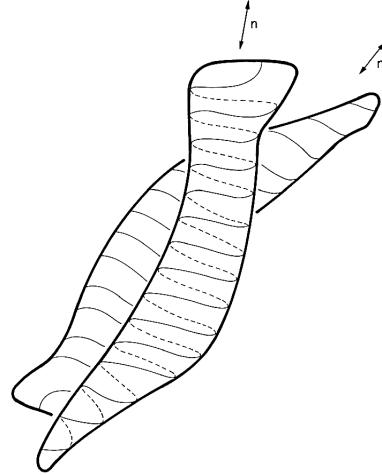
The algebra gets messy on the sphere, though (spherical coordinates are pretty ugly).

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

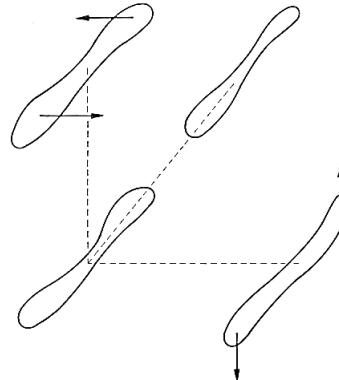
$$(D_i n)^j = \partial_i n^j - q\epsilon_{ijk} n_k, \quad (5.4)$$

where  $\epsilon_{ijk}$  is the totally antisymmetric tensor<sup>9</sup> with three indices (Fig. 5.5 and 5.6). Since the Blue Phases are in flat space, you don't need to worry about the difference between upper and lower indices.

<sup>9</sup>Hence  $\Gamma_{ik}^j = -q\epsilon_{ijk}$ . Warning: this connection has *torsion*: it's not symmetric under interchange of the two bottom indices. This is what allows it to have a non-zero 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.

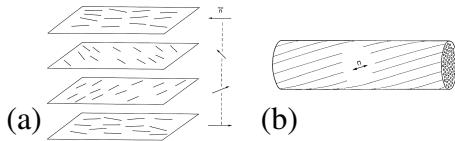


**Fig. 5.5 Blue Phase Molecules.** long thin molecules aligned along an axis  $\mathbf{n}(\mathbf{r})$ , like to sit at a slight angle with respect to their neighbors. To align the threads and grooves on the touching surface between the molecules demands a slight twist. This happens because, like threaded bolts or screws, these molecules are *chiral*.



**Fig. 5.6 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. 5.7(a) (the low-temperature state of this system), or that of Fig. 5.7(b)?



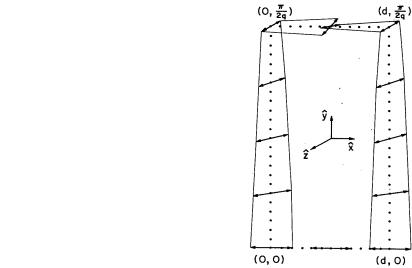
**Fig. 5.7 Local structures** of cholesteric liquid crystals. (a) Helical. (b) Tube.

So, what's the curvature tensor (eqn 5.3) for the blue phase? I figured out at the time that it came out to be

$$R_{ijkl} = q^2(\delta_{i\ell}\delta_{kj} - \delta_{ik}\delta_{\ell j}). \quad (5.5)$$

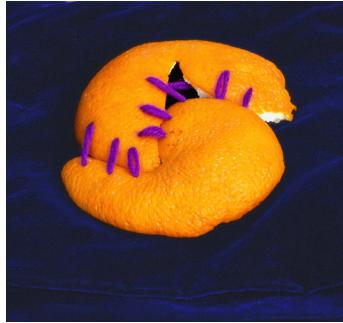
This curvature means that the blue phases are *frustrated*; you can't fill space everywhere with material that's 'happy', with  $D\mathbf{n} = \mathbf{0}$ .

(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 5.3 and 5.5). according to eqn 5.3. Is the shift in  $\mathbf{n}$  along the same direction we observe for the larger loop in Fig 5.8?



**Fig. 5.8 Parallel transport frustration.** We can check your answer to part (c) qualitatively by thinking about a larger loop (Fig. 5.8). 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. 5.7(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. 5.9).



**Fig. 5.9 Orange-peel carpet.** (Copyright Pamela Davis Kivelson)

(5.9) **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  $f(\mathbf{x}) \rightarrow_R f(R^{-1}\mathbf{x})$ . These transformations are linear ( $af(\mathbf{x}) + g(\mathbf{x}) \rightarrow_R af(R^{-1}\mathbf{x}) + g(R^{-1}\mathbf{x})$ ), and obey the group composition rule, and thus form a representation of the rotation group.

(a) Argue that the homogeneous polynomials of degree  $\ell$ ,

$$f(x, y, z) = \sum_{m=0}^{\ell} \sum_{n=0}^{\ell-m} f_{\ell m n} x^m y^n z^{\ell-m-n} \quad (5.6)$$

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  $Y_1^{\pm 1} =$

$\sqrt{3/2\pi}(x \pm iy)/r$ , 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).<sup>10</sup> Electrons in  $d$ -orbitals are much more 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  $5 \times 5$ ) 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  $\phi$  are conjugate to one another.<sup>11</sup>

In class, we found that the character of the  $\ell = 1$  representation by using the Cartesian

<sup>10</sup>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 Greens functions and Feynman diagrams, but it has recently been re-cast as a renormalization-group flow.

<sup>11</sup>If the two rotations have axes  $\hat{n}$  and  $\hat{n}'$ , choose  $S$  to rotate  $\hat{n}'$  into  $\hat{n}$ .

$x, y, z$  basis, where  $R_z(\phi) = \begin{pmatrix} \cos(\phi) & -\sin(\phi) & 0 \\ \sin(\phi) & \cos(\phi) & 0 \\ 0 & 0 & 1 \end{pmatrix}$ .

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  $Y_1^{\pm 1} \rightarrow \exp(\pm i\phi)Y_1^{\pm 1}$ . Here  $R_z(\phi) = \begin{pmatrix} \exp(i\phi) & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \exp(-i\phi) \end{pmatrix}$ , 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 ( $A_1, A_2, E, \dots$ ) 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 in part (d). (Hint: Check how these five orbitals transform under the octahedral symmetries that permute  $x, y$ , and  $z$  among themselves.)

### (5.10) Entangled Spins. (Spins) ③

In class, we studied the entanglement of the singlet spin state  $|S\rangle = (1/\sqrt{2})(|\uparrow\rangle_L|\downarrow\rangle_r - |\downarrow\rangle_L|\uparrow\rangle_r)$  of electrons of a diatomic molecule as the atoms  $L$  and  $R$  are separated;<sup>12</sup> 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:

$$|\chi\rangle = (1/\sqrt{2})(|\uparrow\rangle_L|\downarrow\rangle_r + |\downarrow\rangle_L|\uparrow\rangle_r) \quad (5.7)$$

You should know from angular momentum addition rules that the space of product wavefunctions of two spin  $1/2$  states can be decomposed into a spin 1 and a spin 0 piece:  $1/2 \otimes 1/2 = 1 \oplus 0$ . 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  $|S\rangle$  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  $\Psi(x_L, s_L, x_R, s_R)$  is a product of this spin wavefunction  $\chi(s_L, s_R)$  and a two-particle spatial wavefunction  $\psi(x_L, x_R)$ , what symmetry must  $\psi$  have under interchange of electrons?

We noted in class that two other spin-1 product states,  $|\uparrow\rangle_L|\uparrow\rangle_r$  and  $|\downarrow\rangle_L|\downarrow\rangle_r$  do not form entangled states when  $L$  and  $R$  separate. Is  $|\chi\rangle$  like these spin-1 states, or is it entangled like  $|S\rangle$  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:

$$|\chi\rangle = (\uparrow\rangle_L \downarrow\rangle_L) M \begin{pmatrix} \uparrow\rangle_r \\ \downarrow\rangle_r \end{pmatrix}. \quad (5.8)$$

What is  $M$ ? Give an explicit singular value de-

<sup>12</sup>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.

<sup>13</sup>Remember that the SVD guarantees that  $U$  and  $V$  have orthonormal columns, and  $\Sigma$  is a diagonal matrix whose diagonal elements  $\sigma_i$  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.

composition<sup>13</sup>  $M = U\Sigma V^T$  of the matrix  $M$ . Explain how the SVD gives us the Schmidt decomposition of part (c).

(5.11) **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  $2 \times 2$  matrices with determinant one. Vectors transform under  $SO(3)$ , the ordinary  $3 \times 3$  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  $\{\text{Re}(a), \text{Im}(a), \text{Re}(b), \text{Im}(b)\}$  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  $360^\circ$  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  $\phi$  about an axis  $\hat{\mathbf{n}}$  is given by

$$\exp(-i\mathbf{S}\cdot\hat{\mathbf{n}}\phi/\hbar) = \exp(-i\boldsymbol{\sigma}\cdot\hat{\mathbf{n}}\phi/2) = \exp(-i\mathbf{J}\cdot\hat{\mathbf{n}}\phi/\hbar) \quad (5.9)$$

where the middle formula works for  $SU(2)$  (where  $\mathbf{S} = \hbar\boldsymbol{\sigma}/2$ , because the particles have spin 1/2) and the last formula is appropriate for  $SO(3)$ .<sup>14</sup> 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  $\phi$ ? Since all rotations by  $360^\circ$  are the same, it makes sense to make the length of the vector go to infinity as  $\phi \rightarrow 360^\circ$ . 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  $360^\circ$ , develop a phase change of  $\exp(i\pi) = -1$  (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  $720^\circ$  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  $\infty$ . 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  $\phi$  and a rotation by  $\phi + 2\pi$  should be identified.)*

(5.12) **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*).<sup>15</sup>

$$\mathcal{H}^Z = \sum_{i=1}^Z p_i^2/2m - k_\varepsilon Ze^2/r_i \quad (5.10)$$

Remember that the eigenstates of a single electron bound to a nucleus with charge  $Z$  are the hydrogen levels ( $\psi_n^Z = \psi_{1s}^Z, \psi_{2s}^Z, \psi_{2p}^Z, \dots$ ), except shrunken and shifted upward in binding energy ( $E^Z$  more negative):

$$\begin{aligned} \mathcal{H}^Z \psi_n^Z &= E_n^Z \psi_n \\ \psi_n^Z(\mathbf{r}) &= \psi_n^H(\lambda_r \mathbf{r}) \\ E^Z &= \lambda_E E^H \end{aligned} \quad (5.11)$$

(a) *By what factor  $\lambda_r$  do the wavefunctions shrink? By what factor  $\lambda_E$  do the energies grow?* (Hint: Dimensional arguments are preferred over looking up the formulas.)

<sup>14</sup>The factor of two is there for  $SU(2)$  because the spin is 1/2. For  $SO(3)$ , infinitesimal generators are antisymmetric matrices, so  $\mathbf{S}_{jk}^{(i)} = \mathbf{J}_{jk}^{(i)} = i\hbar\epsilon_{ijk}$  in the  $xyz$  basis; in the usual quantum basis  $m_z = (-1, 0, 1)$  the formula will be different.

<sup>15</sup>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 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*

$$\Psi(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \dots) = \psi_{n_1}^Z(\mathbf{r}_1)\psi_{n_2}^Z(\mathbf{r}_2)\psi_{n_3}^Z(\mathbf{r}_3)\dots \quad (5.12)$$

has energy  $E = \sum_i E_{n_i}^Z$ .

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.<sup>16</sup> 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:

$$\Psi(\mathbf{r}_1, s_1; \mathbf{r}_2, s_2; \mathbf{r}_3, s_3; \dots) \stackrel{?}{=} \psi(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3 \dots) \chi(s_1, s_2, s_3 \dots) \quad (5.13)$$

We'll check this in the first non-trivial 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 ↑<sub>3</sub>:

$$\Psi^0(\mathbf{r}_1, s_1; \mathbf{r}_2, s_2; \mathbf{r}_3, s_3) = \psi_{1s}^{Li}(\mathbf{r}_1)\psi_{1s}^{Li}(\mathbf{r}_2)\psi_{2s}^{Li}(\mathbf{r}_3) \uparrow_1 \downarrow_2 \uparrow_3 . \quad (5.14)$$

But this combination is not antisymmetric under permutations of the electrons.

(d) *Antisymmetrize  $\Psi^0$  with respect to electrons 1 and 2. Show that the resulting state is a singlet with respect to these two electrons. Antisymmetrize  $\Psi^0$  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 5.13?*

**(5.13) 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  $\int dx dy \Psi^* \Psi$ .) 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  $3 \times 3$  matrices  $M_{ij}$ , where  $M =$

<sup>16</sup>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.

$\sum_{i=1}^3 \sum_{j=1}^3 M_{ij} \hat{\mathbf{e}}_i \hat{\mathbf{e}}_j$ . 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 isn't a vector – it is an antisymmetric  $3 \times 3$  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.

(5.14) **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)?

(5.15) **Propagators to path integrals.** (Path Integrals) ③

In class, we calculated the propagator for free particles, which Sakurai also calculates

(eqn 2.6.16):

$$K(x', t; x_0, t_0) = \sqrt{\frac{m}{2\pi i\hbar(t-t_0)}} \exp\left[\frac{im(x'-x_0)^2}{2\hbar(t-t_0)}\right]. \quad (5.15)$$

Sakurai also gives the propagator for the simple harmonic oscillator (eqn 2.6.18):

$$\begin{aligned} K(x', t; x_0, t_0) &= \sqrt{\frac{m\omega}{2\pi i\hbar \sin[\omega(t-t_0)]}} \\ &\times \exp\left[\left\{\frac{im\omega}{2\hbar \sin[\omega(t-t_0)]}\right\}\right. \\ &\left. [(x'^2 + x_0^2) \cos[\omega(t-t_0)] - 2x'x_0]\right]. \end{aligned} \quad (5.16)$$

In deriving the path integral, Feynman approximates the short-time propagator in a potential  $V(x)$  using the ‘trapezoidal’ rule:

$$\begin{aligned} K(x_0 + \Delta x, t_0 + \Delta t; x_0, t_0) & \quad (5.17) \\ &= N_{\Delta t} \exp\left[\frac{i\Delta t}{\hbar} \left\{ \frac{1}{2} m (\Delta x/\Delta t)^2 - V(x_0) \right\} \right], \end{aligned}$$

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  $\Delta t$  for the free particle and the simple harmonic oscillator? 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  $\Delta t$  the terms in the exponential for the free particle propagator (eqn 5.15) and the simple harmonic oscillator (eqn 5.16). 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  $\Delta t$ .)

(5.16) **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 non-interacting, the quantum eigenstates can be written as suitably symmetrized or antisymmetrized single-particle eigenstates. One can use a level diagram, such as in Fig. 5.10, to denote the fillings of the single particle states for each many-electron eigenstate.

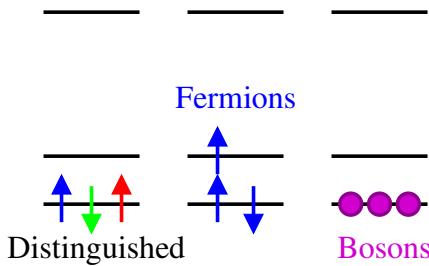


Fig. 5.10 Level diagram, showing one of the ground states for each of the three cases.

(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. 5.10).

(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.

- (5.17) **Rotation matrices.** (Mathematics,  $\times 1$ ) 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  $R$ ?

$$I. R = \begin{pmatrix} u_1 v_1 + v_1 w_1 + w_1 u_1 & \dots \\ & \dots \end{pmatrix}$$

$$II. R = \begin{pmatrix} (\hat{\mathbf{u}}) \\ (\hat{\mathbf{v}}) \\ (\hat{\mathbf{w}}) \end{pmatrix};$$

$$III. R = \begin{pmatrix} (\hat{\mathbf{u}}) & (\hat{\mathbf{v}}) & (\hat{\mathbf{w}}) \end{pmatrix};$$

$$IV. R = \hat{\mathbf{u}} \otimes \hat{\mathbf{v}} + \hat{\mathbf{v}} \otimes \hat{\mathbf{w}} + \hat{\mathbf{w}} \otimes \hat{\mathbf{u}}$$

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  $\langle u_i|x_j\rangle = R_{ij}^T$ . 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) Multiplying 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  $x_\ell$ , derive the matrix transformation formula  $A'_{ij} = R_{ik}^T A_{k\ell} R_{\ell j} = R^T AR$ , where we use the Einstein summation convention over repeated indices..

- (5.18) **Trace.** (Mathematics,  $\times 1$ ) 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 5.17(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  $R^T AR$  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.

(5.19) **Complex exponentials.** (Mathematics) ②

You can prove the double angle formulas using complex exponentials: just take the real and imaginary parts of the identity  $\cos(A + B) + i\sin(A + B) = e^{i(A+B)} = e^{iA}e^{iB} = (\cos A + i\sin A)(\cos B + i\sin B) = \cos A \cos B - \sin A \sin B + i(\cos A \sin B + \sin A \cos B)$ .

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?

$$(A) \cos(3\theta) = \cos^3(\theta) - 3\cos(\theta)\sin^2(\theta), \\ \sin(3\theta) = \sin^3(\theta) - 3\cos^2(\theta)\sin(\theta);$$

$$(B) \sin(3\theta) = \cos^3(\theta) - 3\cos(\theta)\sin^2(\theta), \\ \cos(3\theta) = 3\cos^2(\theta)\sin(\theta) - \sin^3(\theta);$$

$$(C) \cos(3\theta) = \cos^3(\theta) - 3\cos(\theta)\sin^2(\theta), \\ \sin(3\theta) = 3\cos^2(\theta)\sin(\theta) - \sin^3(\theta);$$

$$(D) \sin(3\theta) = 3\cos^2(\theta)\sin(\theta) + \sin^3(\theta); \\ \cos(3\theta) = \cos^3(\theta) + 3\cos(\theta)\sin^2(\theta);$$

$$(E) \cos(3\theta) = \cos^3(\theta) - 3i\cos(\theta)\sin^2(\theta), \\ \sin(3\theta) = 3i\cos^2(\theta)\sin(\theta) - \sin^3(\theta).$$

(5.20) **Dirac  $\delta$ -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:  $\int d\mathbf{x}|\psi(\mathbf{x})|^2 < \infty$ .

But what about unbound states? For example, the propagating plane-wave states  $\psi(x) = |k\rangle \propto \exp(-ikx)$  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  $\phi = \mathbb{1}\phi = \int dk |k\rangle\langle k|\phi\rangle$ . This tells us how we must normalize our continuum wavefunctions: instead of the Kronecker- $\delta$  function  $\langle m|n\rangle = \delta_{mn}$  enforcing orthonormal states, we demand  $|k'\rangle = \mathbb{1}|k'\rangle = \int dk |k\rangle\langle k|k'\rangle = |k'\rangle = \int dk |k\rangle\delta(k - k')$  telling us that  $\langle k|k'\rangle =$

$\delta(k - k')$  is needed to ensure the useful decomposition  $\mathbb{1} = \int dk |k\rangle\langle k|$ .

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  $0 \leq x < L$ , 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  $|n\rangle \propto \exp(-ik_n x)$  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  $\Delta k$  between successive wavevectors? Does it go to zero as  $L \rightarrow \infty$ , 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  $\int_{-\infty}^{\infty} N_k N_{k'} \exp(-i(k' - k)x) dx = \delta(k' - k)$ .

(b) What is the normalization  $\langle x|n\rangle = N_n \exp(ik_n x)$  for the discrete wave-functions in the periodic box, to make  $\langle n|n'\rangle = \int_0^L N_n N_{n'} \exp(-i(k'_n - k_n)x) dx = \delta_{nn}$ ? Write  $\mathbb{1} = \sum_n |n\rangle\langle n|$ , and change the sum to an integral in the usual way ( $\int dk |k\rangle\langle k| \approx \sum_n |n\rangle\langle n|\Delta k$ ). Show that the normalization of the continuum wavefunctions must be  $N_k = 1/\sqrt{2\pi}$ , so  $\psi_k(x) = \langle x|k\rangle = \exp(ikx)/\sqrt{2\pi}$ . (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  $\mathbb{1} = \int dk |k\rangle\langle k|$  (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  $\delta$  function:

$$\delta(k' - k) = \langle k|k'\rangle = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i(k'-k)x} dx. \quad (5.18)$$

Also,  $\delta(x' - x) = 1/(2\pi) \int_{-\infty}^{\infty} dk \exp(ik(x' - x))$ .

II. The same normalization is used for ‘position eigenstates’  $|x\rangle$ , so  $\langle x'|x\rangle = \delta(x' - x)$  and  $\mathbb{1} = \int dx |x\rangle\langle x|$ .

III. The Fourier transform can be viewed as a change of variables from the basis  $|x\rangle$  to the ba-

sis  $|k\rangle$ :

$$\begin{aligned}\tilde{\phi}(k) &= \langle k|\phi \rangle = \langle k|1\!\!1|\phi \rangle \\ &= \int dx \langle k|x\rangle \langle x|\phi \rangle \\ &= 1/\sqrt{2\pi} \int dx \exp(-ikx)\phi(x)\end{aligned}\quad (5.19)$$

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](http://pages.physics.cornell.edu/~sethna/Stat_Mech/EntropyOrderParametersComplexity.pdf)); there the  $2\pi$  is placed entirely on the inverse Fourier transform, which here it is split symmetrically between the two, so the inverse Fourier transform is

$$\phi(x) = \langle x|\phi \rangle = \langle x|1\!\!1|\phi \rangle \quad (5.20)$$

$$= \int dk \langle x|k\rangle \langle k|\phi \rangle \quad (5.21)$$

$$= 1/\sqrt{2\pi} \int dk \exp(ikx)\tilde{\phi}(k).$$

IV. The Dirac  $\delta$ -function can be written in many different ways. It is basically the limit<sup>17</sup> as  $\epsilon \rightarrow 0$  of sharply-peaked, integral-one functions of width  $\epsilon$  and height  $1/\epsilon$  centered at zero. Let's use this to derive the useful relation:

$$\lim_{\epsilon \rightarrow 0+} \frac{1}{x - i\epsilon} = p.v. \frac{1}{x} + i\pi\delta(x). \quad (5.22)$$

Here all these expressions are meant to be inside integrals, and p.v. is the Cauchy principal value of the integral:<sup>18</sup>

$$p.v. \int_{-\infty}^{\infty} = \lim_{\epsilon \rightarrow 0+} \int_{-\infty}^{-\epsilon} + \int_{\epsilon}^{\infty}. \quad (5.23)$$

(c) Note that  $\epsilon/(x^2 + \epsilon^2)$  has half-width  $\epsilon$  at half-maximum, height  $\epsilon$ , and integrates to  $\pi$ , so basically (i.e., in the weak limit)  $\lim_{\epsilon \rightarrow 0} \epsilon/(x^2 + \epsilon^2) = \pi\delta(x)$ . Argue that  $\lim_{\epsilon \rightarrow 0} \int f(x)/(x - i\epsilon) dx = p.v. \int f(x)/x dx + i\pi f(0)$ . (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  $x + i\epsilon$ .)

(5.21) **Eigen Stuff.** (Mathematics,  $\times 1$ ) ②

Consider an operator for a two-state system  $O = \begin{pmatrix} 0 & -4 \\ -4 & 6 \end{pmatrix}$ . Its eigenvectors are  $|e_1\rangle = \frac{1}{\sqrt{5}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$  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 5.17(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 5.17(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?

(5.22) **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<sup>19</sup> in an electronic  $f$ -state (i.e., with  $\ell = 3$ ), adsorbed on a monolayer of graphene (Fig. 5.11). 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- $\ell$  representation for  $SO(3)$  for a rotation by angle  $\theta$  is  $\chi^{(\ell)}(\theta) = \sin[(\ell + \frac{1}{2})\theta]/\sin[\frac{1}{2}\theta]$ . (Check this for the  $\ell = 1$

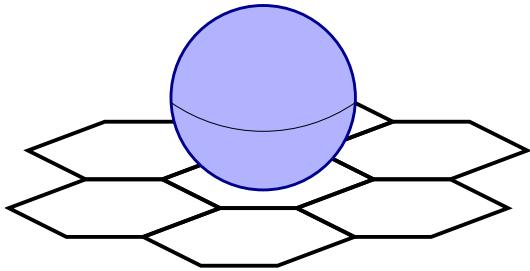
<sup>17</sup>Clearly this is not a limit in the ordinary sense: the difference between functions does not go to zero as  $\epsilon$  goes to zero, but rather (within  $\epsilon$  of the origin) has large positive and negative values that cancel. It is a *weak limit* – when integrated against any smooth functions, the differences go to zero.

<sup>18</sup>If  $f(x)$  is positive at zero,  $\int f(x)/x dx$  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.

<sup>19</sup>The atom is light so that we may ignore the spins of the electrons. A heavy atom would have significant spin-orbit interactions.

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  $\sigma_v$  and  $\sigma'_v$ ) 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. 5.11 Atom adsorbed on graphene.**

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  $\ell$  representation of  $O(3)$  generated by the rotations of the spherical harmonics  $Y_{\ell}^m(\theta, \phi)$ . Remember that  $\theta$  is the angle from the  $\hat{z}$  axis, and  $\phi$  is measured from the  $\hat{x}$  axis.

(c) How does  $Y_{\ell}^m$  transform under the reflection  $\Sigma(\hat{y})$  in the  $x-z$  plane? In the  $(2\ell+1)$ -dimensional space of  $Y_{\ell}^m$  for fixed  $\ell$ , 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  $\ell$ .

Table ?? 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 non-degenerate eigenstates with reps  $A_1$  and  $B_2$ , and three doublet eigenstates, two with reps  $E_2$  and one with rep  $E_1$ .”

- (5.23) **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.<sup>20</sup> 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

$$\mathcal{L} = \frac{1}{2}m\dot{y}^2 - mgy. \quad (5.24)$$

(a) In classical mechanics, if the buckyball starts and ends at  $y = 0$  and travels for a time  $2\Delta t$ , how high  $y_{\text{peak}}$  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  $y(t)$ :

$$\langle y_f, t_f | y_i, t_i \rangle = \iiint_{y_i, t_i}^{y_f, t_f} \mathcal{D}[y(t)] \exp(i/\hbar S[y(t)]) = \iiint_{y_i, t_i}^{y_f, t_f} \mathcal{D}[y(t)] \exp\left(i/\hbar \int \mathcal{L} dt\right) \quad (5.25)$$

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  $y_2$ :

$$S[y(t)] \approx \left[ \frac{1}{2}m \left( \frac{y_3 - y_2}{\Delta t} \right)^2 - \frac{1}{2}mg(y_1 + y_2) - \frac{1}{2}mg(y_2 + y_3) + \frac{1}{2}m \left( \frac{y_2 - y_1}{\Delta t} \right)^2 \right] \Delta t. \quad (5.26)$$

- (b) What intermediate point  $y_2^*$  minimizes the trapezoidal action (eqn 5.26), for general  $y_1$  and  $y_3$ ? For  $y_1 = y_3 = 0$ , how does this compare to

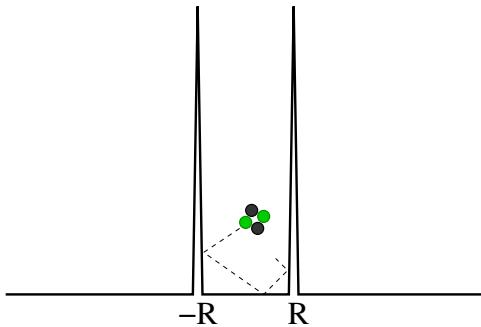
<sup>20</sup>See “Nanomechanical oscillations in a single-C<sub>60</sub> transistor”, by Hongkun Park, Jiwoong Park, Andrew K.L. Lim, Erik H. Anderson, A. Paul Alivisatos, and Paul L. McEuen, *Nature* **407**, 57-60 (2000).

the peak of the trajectory in part (a)? What is the action  $S^* = S[y_2^*]$  for this 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  $\langle y=0, t=2\Delta t | y=0, t=0 \rangle$ ? (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,  $\langle y_f, t_i | y_i, t_i \rangle = \delta(y_f - y_i)$  has units of inverse length. Also,  $\int_{-\infty}^{\infty} dx \exp(-iAx^2) = \sqrt{\pi/iA}$ .)

(5.24) **Resonances:  $\alpha$ -decay.** (Quantum) ③



**Fig. 5.12 One-dimensional nuclear potential.**

In this exercise, we solve a one-dimensional model of radioactive  $\alpha$ -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  $\alpha$  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  $\delta$ -function of strength  $U > 0$ <sup>21</sup> (see Fig. 5.12). Outside the nucleus, the potential is zero:

$$\begin{aligned} V(x) &= 0 & (|x| < R) \\ V(x) &= U\delta(x \pm R) & (x = \mp R) \\ V(x) &= 0 & (R < |x| < \infty) \end{aligned}$$

(The attractive case  $U < 0$  is a model for hydrogen, and is discussed for example in Wikipedia's *Double Delta Potential* article, [http://en.wikipedia.org/wiki/Delta\\_potential#Double\\_Delta\\_Potential](http://en.wikipedia.org/wiki/Delta_potential#Double_Delta_Potential).)

Parts (a)-(d) of this exercise solve analytically for the energy eigenstates, but getting them correct is important for the later parts.<sup>22</sup>

Our Hamiltonian has a symmetry which allows us to choose energy eigenstates that are even ( $\psi_E$ ) or odd ( $\phi_E$ ).

(a) What symmetry of the Hamiltonian is this? Given an energy eigenstate  $\zeta_E(x)$  with mixed symmetry (in particular,  $\zeta_E$  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 states  $\psi_E$  which can be non-zero at  $x = 0$ .

(b) For the eigenstates  $\psi_E$  which are non-zero at  $x = 0$ , what is  $\psi'_E(0) = (\partial\psi_E/\partial x)|_{x=0}$ ?

Next we want to solve for the energy eigenstates. This is best done in three steps. First, we deduce the form of the wavefunction. Note that, away from the  $\delta$ -function, the wavefunction has wavevector  $k(E) = \sqrt{2mE/\hbar}$ ; 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<sup>23</sup>  $\psi_k^{\text{out}} = B \sin(kx + \Delta_k)$ . Note that there is a continuum of  $\psi_k$  eigenstates, so it is proper for us to use the  $\delta$ -function normalization

<sup>21</sup> In one-dimensional quantum mechanics, the first derivative of the wave-function jumps where the potential has a  $\delta$ -function. Find details in a textbook or on the Web.

<sup>22</sup> 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  $\sigma$  as small as is numerically convenient.

<sup>23</sup> For  $x < -R$ , we use the even symmetry of  $\psi_E$  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.

$$\langle \psi_k | \psi_{k'} \rangle = \delta(k - k').$$

(c) 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.)

We then impose the conditions induced by the  $\delta$ -potential at the edge of the nucleus.

(d) Write the condition on  $A_k$  and  $\Delta_k$  given by imposing continuity of  $\psi_k(x)$  at  $x = R$ . Write the conditions on  $A_k$  and  $\Delta_k$  given by the discontinuity of  $\psi'_k(x)$  imposed by the  $\delta$ -function potential (see footnote 21). For convenience, write your answers from here on in terms of the unitless ratio  $\tilde{U} = 2mRU/\hbar^2$ .

We can now solve for the eigenstates of our Hamiltonian that are non-zero at  $x = 0$ .

(e) Use the conditions of part (d), solve for  $A_k^2$ . (Trick: Arrange the two equations of part (d) to be  $\sin(kR + \Delta_k) = \dots$  and  $\cos(kR + \Delta_k) = \dots$ , where  $\dots$  is independent of  $\Delta_k$ . Sum the squares of the right-hand sides: what must the sum be equal to?)

We now consider the decay of an  $\alpha$ -particle injected into this potential at  $x = 0$ . That is, consider an initial wavefunction  $\Psi(x) = \delta(x)$ .<sup>24</sup>

(f) (10 points) What is the probability<sup>25</sup>  $P(k)$  of being in eigenstate  $\psi_k$ ? (Write your answer abstractly in terms of  $\psi_k(x)$ . This you can do without solving parts (a-d).)

(g) (20 points) Plot the probabilities  $P(k)$  versus  $kR$  with  $\tilde{U} = 30$  and for  $0 < kR < 10$ .

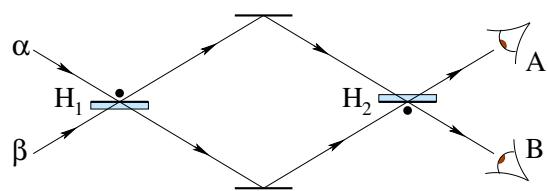
In the limit  $U \rightarrow \infty$ , the nucleus should approximate a particle in a box of size  $2R$ . In that limit, the injection of an  $\alpha$ -particle can only occur at certain discrete energies – the nuclear eigenstates  $\mathcal{E}_m^\infty$  of a free particle in a box of size  $2R$ .

(h) (20 points) Compare the peaks you found in part (g) to the wavevectors for the particle-in-a-box states. Why are you missing half of the

peaks?

(i) (Extra credit, up to 20 points for elegant answers.) Find the density of states  $P(E)$  from  $P(k)$ . Using the FWHM of the peaks, estimate the lifetimes of the first three even resonances 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  $U \rightarrow \infty$ ?

(5.25) Mirror path integrals. (Path Integrals) ③



**Fig. 5.13 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 5.13 shows a photon<sup>26</sup> coming from the left in a superposition  $(\alpha \beta)$ , through a set of mirrors, to two detectors named Alice (A) and Bob (B). We work in the basis  $|1\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$  representing the upper of the two beams at a given position, and  $|0\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$  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.<sup>27</sup>

<sup>24</sup>This is a nuclear version of tunneling from an STM tip;  $P(E) = P(k(E)) (dk/dE)$  measures the local density of states for the  $\alpha$  particle at the center of the nucleus.

<sup>25</sup>The position eigenstate  $\Psi(x) = |x=0\rangle$  is  $\delta$ -function normalized, with  $\langle x|x' \rangle = \delta(x-x')$ . Hence the ‘probability’  $P(k)$  integrates to infinity, and not to one.

<sup>26</sup>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.

<sup>27</sup>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,  $|1\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$  is a vector whose first element is one [zeroth in Python/C], and  $|0\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$  is a vector whose second element is one [first in Python/C].

(a) *What is  $G$ ?* Given the impinging wave  $(\frac{\alpha}{\beta})$  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

$$G = \mathbb{1} H_2 \mathbb{1} H_1 \mathbb{1} = \sum_{x_i=0}^1 \sum_{x_m=0}^1 \sum_{x_f=0}^1 |x_f\rangle \langle x_f| H_2 |x_m\rangle \langle x_m| H_1 |x_i\rangle \langle x_i|. \quad (5.27)$$

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)?*

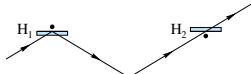
(A)



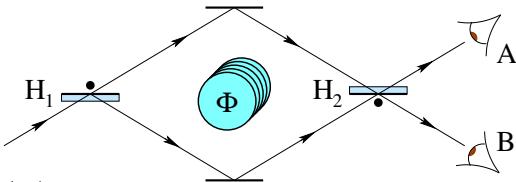
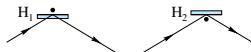
(B)



(C)



(D)

**Fig. 5.14 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  $\Phi$  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  $\Phi$ , what is the probability that an initial electron will be seen by Alice? What values of  $\Phi$ , 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:  $\oint_C \mathbf{A} \cdot d\ell = \Phi$  if the path  $C$  encircles the solenoid counter-clockwise once. The path-integrand amplitude for  $\mathbf{x}(t)$  in a field  $\mathbf{A}$  gains a phase  $\zeta = \int(q/c)\mathbf{A}(x) \cdot d\mathbf{x}/\hbar$ . The charge on an electron is  $q = -e$ .)

(5.26) **Coherent State Evolution.** (Operator algebra) ③

Consider the annihilation operator  $a$  for a simple harmonic oscillator, transformed into the time-dependent Heisenberg-representation operator  $\mathbf{a}_H(t)$ :

$$\mathbf{a}_H(t) = e^{i\mathcal{H}t/\hbar} a e^{-i\mathcal{H}t/\hbar} = U^\dagger(t) a U(t). \quad (5.28)$$

The time evolution for an operator in the Heisenberg representation is given by the commutator with the Hamiltonian, so

$$\frac{d\mathbf{a}_H}{dt} = \frac{i\mathcal{H}}{\hbar} \mathbf{a}_H - \mathbf{a}_H \frac{i\mathcal{H}}{\hbar} = -i/\hbar [\mathbf{a}_H, \mathcal{H}]. \quad (5.29)$$

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  $[\mathbf{a}_H, \mathcal{H}]$ , and write it in terms of  $\mathbf{a}_H$ . What is  $d\mathbf{a}_H/dt$ ?* (Simplify your answers

until they only involve  $\mathbf{a}_H$  and constants, not  $\mathcal{H}$  or  $a$ .)

(b) Show that  $\mathbf{a}_H(t) = \exp(-i\omega t)\mathbf{a}_H(0) = \exp(-i\omega t)a$  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  $\omega$ . 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  $a$ :

$$a|\lambda\rangle = \lambda|\lambda\rangle, \quad (5.30)$$

which is also normalized  $\langle\lambda|\lambda\rangle = 1$ . Here  $\lambda \in \mathbb{C}$  can be any complex number.

(c) In the Schrödinger representation<sup>28</sup> 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. It so happens that, for the standard definition of coherent states, the phase  $C$  is independent of  $\lambda$ , but depends on time.

(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.)

### (5.27) Decoherence.<sup>29</sup> (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  $\omega$  of the oscillation?

Recall from problem 11.1 that any  $2 \times 2$  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  $|\vec{n}| \leq 1$ , forming the solid *Bloch sphere*. (Remember  $\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ ,  $\sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$ , and  $\sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ .)

- (c) Calculate  $\rho(t)$  for  $|\psi(t)\rangle$  from your calculation in part (b), in terms of  $\omega$ . 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<sup>30</sup> to exercise 11.1,  $d\vec{n}/dt \propto -\vec{B} \times \vec{n}$ .
- (d) Show that the eigenvalues of a general  $2 \times 2$  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  $S = -k_B \text{Tr}(\rho \log \rho)$  of a general density matrix in terms of  $\vec{n}$ ? (Hint: use the basis in which  $\rho$  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?

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  $\rho = \begin{pmatrix} \rho_{11} & \rho_{12} \\ \rho_{21} & \rho_{22} \end{pmatrix}$  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.

<sup>28</sup>As opposed to the Heisenberg representation of part (b).

<sup>29</sup>Developed in collaboration with Bhuvanesh Sundar.

<sup>30</sup>That exercise had funny units, but the form of the equation and the sign should agree.

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  $\Gamma$  per unit time. In a small time interval  $\delta t$ , 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  $\delta t$ .

- (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  $\delta t \rightarrow 0$ .*
- (g) *Show that  $n_z$  obeys the second-order differential equation for a damped harmonic oscillator,  $d^2n_z/dt^2 + \eta dn_z/dt + \omega_0^2 n_z = 0$ . What are  $\eta$  and  $\omega_0$  in terms of  $\Gamma$  and  $\omega$ ?*
- (h) *What is the long-time limit for  $\vec{n}$ ? For  $\rho$ ? For the entropy?*

(5.28) **Quantum Algorithms.**<sup>31</sup> (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.<sup>32</sup> 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  $x_0, x_1, \dots, x_{n-1}$ , where  $x_0, x_1, \dots, x_{n-1}$  are all 0 or 1. Let  $\mathbf{x} = \sum_{m=0}^{n-1} 2^m x_m$  denote the integer represented by the bit sequence  $x_0, x_1, \dots, x_{n-1}$ ;  $x_0$  is the least significant bit and  $x_{n-1}$  is the most significant bit. We'll also denote  $|x_0\rangle|x_1\rangle\dots|x_{n-1}\rangle$  as  $|\mathbf{x}\rangle$ . Let  $f$  be a function that maps the  $n$  bits  $x_0, x_1, \dots, x_{n-1}$  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 pro-

<sup>31</sup>Developed in collaboration with Bhuvanesh Sundar, based on an exercise by Paul Ginsparg.

<sup>32</sup>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.

gram 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;<sup>33</sup> 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,  $U_f((\alpha|0\rangle + \beta|1\rangle)|0\rangle|y\rangle) = \alpha|0\rangle|0\rangle|y \oplus f(0, 0)\rangle + \beta|1\rangle|0\rangle|y \oplus f(1, 0)\rangle$ .

(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  $2 \times 2$  unitary matrices.

(d) In the basis<sup>34</sup>  $|0\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$  and  $|1\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ , write the single-Qbit gate  $NOT$  as a  $2 \times 2$  matrix. Show that the Hadamard gate, written  $H = 1/\sqrt{2} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$ , 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<sup>35</sup> 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  $|x_0\rangle, |x_1\rangle, \dots, |x_{n-1}\rangle$  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  $x_0 = 0, x_1 = 0, \dots, x_{n-1} = 0, y = 1$ .

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  $|\Psi_1\rangle = H^{n+1}|\Psi_0\rangle = (H|0\rangle)^n(H|1\rangle)$ .

(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

<sup>33</sup>The reverse operation is  $U^\dagger = U^{-1}$ .

<sup>34</sup>I apologize for the shift in notation: I used the reverse convention in lecture and in exercise 5.29,  $|1\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$  and  $|0\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ .

<sup>35</sup>The special case  $n = 1$  of the Deutsch-Josza algorithm is called the Deutsch's algorithm.

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  $x_0, x_1, \dots, x_{n-1}$  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 ( $|\Psi_2\rangle = \Pi_{0 \leq i < n} |\phi_i\rangle$  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.  $|\Psi_2\rangle = \Pi_{0 \leq i < n} |\phi_i\rangle$ ), and write this product explicitly. Is  $|\Psi_2\rangle$  different from  $|\Psi_1\rangle$ ? What are the measured values of the Qbits  $x_0, \dots, x_{n-1}$  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  $x_0, \dots, x_{n-1}$  are measured to be something other than all zeros. The Deutsch-Josza algorithm states that for any balanced function  $f$ , the probability of measuring  $x_0, \dots, x_{n-1}$  to be all zeros is 0. We'll prove this in the following way.

(i) Show that for an arbitrary balanced function  $f$ ,

$$\begin{aligned} |\Psi_2\rangle &= U_f|\Psi_1\rangle = \frac{1}{2^{\frac{n+1}{2}}} \sum_{0 \leq \mathbf{x} < 2^n} (-1)^{f(\mathbf{x})} |\mathbf{x}\rangle (|0\rangle - |1\rangle) \\ &= \frac{1}{2^{\frac{n+1}{2}}} \left( \sum_{\mathbf{x}: f(\mathbf{x})=0} |\mathbf{x}\rangle - \sum_{\mathbf{x}: f(\mathbf{x})=1} |\mathbf{x}\rangle \right) (|0\rangle - |1\rangle). \end{aligned} \quad (5.31)$$

After the final Hadamard operation, show that the probability of measuring  $x_0 = 0, x_1 = 0, \dots, x_{n-1} = 0, y = 1$  is zero, i.e.  $\langle\Psi_0|(H^{n+1})|\Psi_2\rangle = 0$ . (Hint: Rather than calculating  $\langle\Psi_0|\Psi_3\rangle$ , 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 5.31.)

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.

- (5.29) **Supersymmetric harmonic oscillator.**<sup>36</sup> (Quantum) <sup>3</sup>

One of the main predictions of supersymmetry<sup>37</sup> is that each particle comes with a supersymmet-

<sup>36</sup>Developed in collaboration with John Stout, Fall 2013.

<sup>37</sup>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.

<sup>38</sup>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

ric partner with the same mass but with opposite statistics.<sup>38</sup> 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”<sup>39</sup> example of a supersymmetric Hamiltonian, and try to illustrate each of these features of supersymmetry.<sup>40</sup>

Remember the commutation relations for creation and annihilation operators suitable for bosons

$$[a, a^\dagger] = 1 \quad [a, a] = [a^\dagger, a^\dagger] = 0, \quad (5.32)$$

and fermions

$$\{b, b^\dagger\} = 1 \quad \{b, b\} = \{b^\dagger, b^\dagger\} = 0. \quad (5.33)$$

where  $[A, B] = AB - BA$  is the commutator and  $\{A, B\} = AB + BA$  is the anticommutator. For this simple example, we take our bosons and fermions to be noninteracting, so their creation and annihilation operators commute,

$$[a, b] = [a, b^\dagger] = [a^\dagger, b] = [a^\dagger, b^\dagger] = 0. \quad (5.34)$$

In one dimension, the Hamiltonian of the simple harmonic oscillator of frequency  $\omega$  can be written either in terms of  $x$  and  $p$ :

$$\mathcal{H}_B = p^2/2m + \frac{1}{2}m\omega^2x^2 \quad (5.35)$$

or in terms of the creation and annihilation operators

$$\mathcal{H}_B = \hbar\omega(a^\dagger a + \frac{1}{2}). \quad (5.36)$$

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.

#### Model particles.

<sup>38</sup>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 non-relativistic 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.

<sup>40</sup>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.

<sup>41</sup>I apologize again for the shift back in notation. In this problem, we revert back to the notation used in lecture:  $|1\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$  and  $|0\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ , instead of the quantum computing notation used in Exercise 5.28.

<sup>42</sup>That is a hint for part (b).

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,  $\mathcal{H}_F = \frac{1}{2}\hbar\omega[b^\dagger, b]$ .

(b) What is the ground state energy of  $\mathcal{H}_F$ ? How many fermions are in the ground state? What is the energy of the state with one fermion?

(c) If we write the zero-fermion state as<sup>41</sup>  $|0\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$  and the one-fermion state as  $|1\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ , then write  $b$ ,  $b^\dagger$ , and  $\mathcal{H}_F$  in terms of the three Pauli matrices  $\sigma_x$ ,  $\sigma_y$ , and  $\sigma_z$ . Check that your form for  $b$  and  $b^\dagger$  satisfy the anticommutation relations of eqn 5.33 (Remember  $\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ ,  $\sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$ , and  $\sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ .)

We can write our first supersymmetric Hamiltonian by adding the boson and fermion harmonic oscillators:

$$\mathcal{H}_S = \mathcal{H}_B + \mathcal{H}_F = \frac{1}{2}\hbar\omega \left( \{a^\dagger, a\} + [b^\dagger, b] \right). \quad (5.37)$$

Note that the ground state energy for this Hamiltonian is zero.<sup>42</sup>

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 fermions is a fermion – so half of our eigenstates represent bosons, and half represent fermions.

(d) Which eigenstates represent fermions? Which bosons? Draw the ‘level diagram’ for  $\mathcal{H}_S$ , with the first few boson eigenenergies as a column of horizontal lines on the left, and the first few fermion eigenenergies on the right. Is there a fermion state for each boson state? What state is the exception? If we interpret the energy of a state as the mass of a particle<sup>43</sup>, 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*,

$$Q = b \left( \frac{p}{\sqrt{m}} + i\sqrt{m\omega}x \right) = i\sqrt{2\hbar\omega}ba^\dagger. \quad (5.38)$$

(Remember that  $x = \sqrt{\hbar/2m\omega}(a^\dagger + a)$  and  $p = i\sqrt{m\omega\hbar/2}(a^\dagger - a)$ .)

(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.<sup>44</sup>

Supersymmetry has been shown (by Haag, Lopuszanski, and Sohnius<sup>45</sup>) to be the only way to consistently extend the symmetries of spacetime. Spacetime has a spatial 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

<sup>43</sup>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.

<sup>44</sup> $Q\Psi = 0$  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.

<sup>45</sup>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 symmetries (like the SU(3) of the strong interaction) in a trivial way (so that the total symmetry group is (space-time symmetry)  $\times$  (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.

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

$$Q_W = b \left( \frac{p}{\sqrt{m}} + i\sqrt{m}W'(x) \right) \quad (5.39)$$

where  $W'(x) = dW/dx$ , 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  $\mathcal{H}_W$  as  $2 \times 2$  matrices

$$\mathcal{H}_W = \begin{pmatrix} \mathcal{H}_1 & 0 \\ 0 & \mathcal{H}_2 \end{pmatrix} \quad \text{and} \quad Q_W = \begin{pmatrix} 0 & 0 \\ A & 0 \end{pmatrix}. \quad (5.40)$$

where the elements of the matrices are functions of  $p$  and  $x$ . (Hint: Remember  $p = -i\hbar\partial/\partial x$ . You might check this against the Web, which has different units.)

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.)

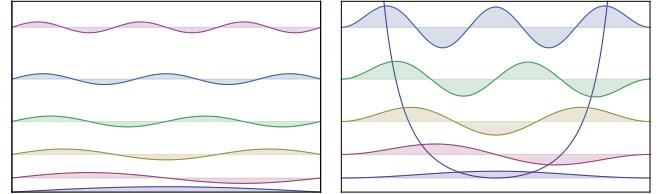
Let us work out a specific example. Consider  $W'(x) = (\pi\hbar/mL)\cot(\pi x/L)$ .

(i) Show that  $\mathcal{H}_1$  is the particle-in-a-box Hamiltonian (Fig. 5.15) shifted by a constant to set its ground state energy to zero. Show that  $\mathcal{H}_2$  is a Hamiltonian with potential<sup>46</sup>

$$V(x) = \frac{\pi^2\hbar^2}{2mL^2} \left( 2\csc^2\left(\frac{\pi x}{L}\right) - 1 \right). \quad (5.41)$$

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. 5.15 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.<sup>47</sup>

### (5.30) 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.

<sup>46</sup>Note that the  $\Psi = 0$  boundary conditions for the two Hamiltonians are the same for both  $\mathcal{H}_1$  and  $\mathcal{H}_2$ .

<sup>47</sup>E. Witten, *Supersymmetry and Morse Theory*, J. Diff Geom. **17**, 661–692 (1982).

<sup>48</sup>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 particu-

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  $\phi$ .<sup>48</sup>

(a) *Show that every different  $g_\phi$  is in its own conjugacy class.* (This is true for any commutative group.)

Thus we may label the conjugacy classes by the angle  $\phi$ .

Consider the action of  $g_\phi$  on a function  $f(\theta)$ :

$$R(g_\phi) : f(\theta) \rightarrow f(\theta - \phi). \quad (5.42)$$

Here  $\theta$  represents a point on a circle, the complex function  $f(\theta)$  is a vector in the Hilbert space of complex functions<sup>49</sup> on the circle, and  $R(g_\phi)$  is a linear mapping of that function space into itself.<sup>50</sup>

(b) *Show that, for any non-negative 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  $2 \times 2$  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  $\phi$ ).

For finite groups, we decomposed representations into irreducible pieces using the ‘little’ orthogonality theorem: for any two irreducible rep-

resentations  $\alpha$  and  $\beta$ , 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 the group,<sup>51</sup> and the number of elements of the group replaced by the ‘volume’ of the group. For two-dimensional rotations, we find

$$\int_0^{2\pi} d\phi \chi^{(\alpha)}(\phi) \chi^{(\beta)}(\phi)^* = 2\pi \delta_{\alpha\beta}. \quad (5.43)$$

(d) *Show that the characters of your irreducible representations from part (c) satisfy the orthogonality relation 5.43. 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

$$P^{(\alpha)} = (f^{(\alpha)} / o(G)) \sum_{g \in G} \chi^{(\alpha)}(g)^* R(g). \quad (5.44)$$

Here<sup>52</sup>  $f^{(\alpha)}$  is the dimension of the representation  $\alpha$ .

For example, any random deformation of the molecule, when averaged over the group, gave a uniform dilation of the triangle. This dilation 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 (5.44) suggests. When Alex multiplied by the characters of the two-dimensional representation  $E$  (using

lar, 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.

<sup>49</sup>Particularly,  $L^2$  functions on the circle.

<sup>50</sup>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.

<sup>51</sup>For  $SO(2)$ , this is just an integral over  $\phi$ . More generally, and in particular for  $SO(3)$ , you have an extra factor in the integral (the *Haar measure*).

<sup>52</sup>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  $\alpha$  equal to the original function. Alex also, I think, missed the complex conjugate (but all his characters were real).

$P^{(E)}$  in eqn 5.44), 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

$$P^{(\alpha)} = (1/2\pi) \int_0^{2\pi} d\phi \chi^{(\alpha)}(\phi)^* R(g_\phi). \quad (5.45)$$

Let  $f(\theta)$  be a particular complex function on the circle. Let  $R(g_\phi)$  be defined on the function as in eqn 5.42).

(e) Show that the projection operator in eqn 5.45, 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<sup>53</sup>  $\tilde{f}_m$  for  $f$ ?

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  $SO(2)$ ?

(5.31) **Solving Schrödinger: WKB, resonances, and lifetimes.** (Computation) ③

We study the problem of quantum tunneling through a barrier. We shall use a potential in the form of a cubic polynomial.<sup>54</sup>

$$V(y) = \frac{1}{2}m\omega^2(y^2 - y^3/Q). \quad (5.46)$$

You should observe that this potential has frequency  $\omega$  for small oscillations in the well, and has a turning point at  $y = Q$ . (The *turning point* is where the potential energy goes to zero again.)

<sup>53</sup>There are many different conventions for Fourier series. Clearly state which one you are using.

<sup>54</sup>The cubic potential is not only convenient for analytic calculations, it also approximates a generic potential when the barrier is low. (More precisely, it approximates a *generic* potential near the *saddle-node* transition when the barrier vanishes.) We saw in exercise (6.1) that tunneling of atoms is very slow unless the barrier is low and narrow compared to the typical scales of one Rydberg and one Bohr radius.

<sup>55</sup>In the symmetric well,  $\psi$  leaks through the barrier. In the decay from the metastable well, probability  $\psi^* \psi$  must escape, leading to two suppression factors. This rough argument can be made precise in simple two-level models.

Remember that the instanton formula for the tunneling decay rate is of the form

$$\Gamma_0 \exp(-S_0/\hbar) = \Gamma_0 \exp\left(-2 \int_0^Q \sqrt{2mV(q)} dq / \hbar\right) \quad (5.47)$$

where  $\Gamma_0$  is a prefactor of order  $\omega$  that we will estimate numerically. Note the factor of two in the exponent compared to the symmetric double well tunnel splitting: the instanton bounce for decay out of a metastable well crosses the barrier twice.<sup>55</sup> Remember that  $a_0 = \sqrt{\hbar/(2m\omega)}$  is the root-mean-square width of the ground state wavefunction in the well in the harmonic approximation.

(a) If we set  $Q = n_0 a_0$ , calculate the barrier height  $V_{\max}$  and the instanton action  $S_0/\hbar$ . Simplify your answer: it should only depend on  $n_0$ . (Hint: If you aren’t using a symbolic manipulation package like Mathematica, you can do the integral for  $S_0/\hbar$  numerically, by changing variables and pulling out all the factors of  $m$ ,  $\hbar$ ,  $\omega$ ,  $n_0$ , etc.)

How tall and wide a barrier can one tunnel through in a reasonable time? It depends on what’s considered reasonable. For our simulation, we shall simulate 1000 periods of the oscillation  $P = 2\pi/\omega$ . In an experiment with a few molecules, one can wait for a few seconds. In a radioactive decay experiment, where  $10^{23}$  potential decays can be monitored, one can measure lifetimes of billions of years. For this calculation, pretend that the true decay rate is given by eqn 5.47 with  $\Gamma_0 = \omega = 10^{12}/\text{sec}$ .

(b) How big can  $n_0$  be to get a lifetime of  $1000P$ ? Of one second? Of a billion years? How big is  $V_{\max}$  and  $Q$ , in units of Rydbergs and Bohr radii? (Hint: There are approximately  $\pi \times 10^7$  seconds in a year. The Bohr radius is  $\hbar/(\alpha m_e c)$  and the Rydberg is  $m_e c^2 \alpha^2/2$ , where the fine structure constant  $\alpha = e^2/\hbar c \approx 1/137.036$ .)

It may be remarkable how little the barrier changes to make such a large difference in the

tunneling.

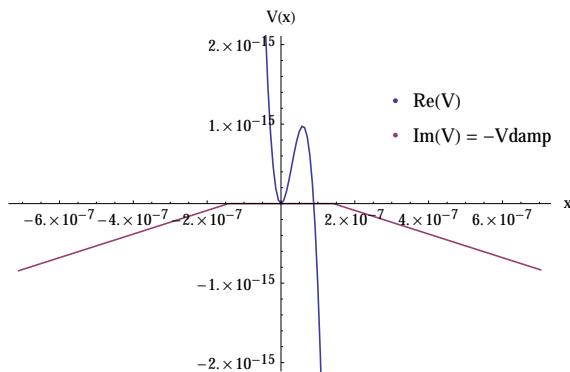
We set  $Q = 5a_0$  (hence  $n_0 = 5$ ) to get a reasonable tunneling rate. We set  $m$  to be the mass of a hydrogen atom and  $\omega = 10^{12}/\text{sec}$ . Our WKB formula is asymptotically exact in the limit when the barrier height is many times the energy splitting  $\hbar\omega$  in the well.

(c) What is the ratio of the barrier height  $V_{\max}$  to  $\hbar\omega$ ? Is  $V_{\max}$  at least larger than the zero point energy  $\frac{1}{2}\hbar\omega$ ?

We'll use a grid of length  $L = 80a_0$  with  $N_p = 200$  points. As the wavefunction leaks out of the well, we need it to disappear before it reflects back into the well. There are several ways of doing this, mathematically and numerically. We shall do it by adding a negative imaginary part to the potential energy  $V_{\text{damp}}$ :

$$U(t) = e^{-i(H-iV_{\text{damp}}[x])t/\hbar} = e^{-V_{\text{damp}}[x]t/\hbar-iHt/\hbar} \quad (5.48)$$

This imaginary part depletes the wavefunction exponentially with a rate  $V_{\text{damp}}[x]/\hbar$  wherever it is non-zero. We shall make  $V_{\text{damp}}$  zero near the well and the barrier, starting linearly at  $|y| > x_D = L/10 = 8a_0 = 1.6Q$  with slope  $-2\hbar\omega/L$ . Plot the real and imaginary parts of  $V(x)$ , and check them against Fig. 5.16.



**Fig. 5.16 Real and imaginary part of  $V(x)$ .**

Letting the time step  $dt = P/50 = 2\pi/(50\omega)$ , make an array  $U_{\text{pot}}(dt/2)$  as usual, but incorporating both the real and imaginary parts of the potential energy. Make the array  $U_{\text{kin}}(dt)$  as usual, and start the wavefunction in the harmonic-oscillator ground state in the well. Using our operator-splitting Fourier method based

on the Baker-Campbell-Hausdorff formula (Exercise 5.6), iterate to  $1000P = 50000dt$ . (This should take a good fraction of a minute.)

(d) On a single plot, show  $|\psi(x,t)|^2$  versus  $x$  for  $t = 0, 200P$ , and  $800P$ . Does it appear to be exponentially decaying?

(e) Write a routine to integrate  $P(t) = \int |\psi(x,t)|^2 dx$  over the system, to see what fraction of the probability has not escaped the well and been eaten by the imaginary potential. Fit your answer, for  $t > 40P$ , to determine the exponential decay rate  $\Gamma$ . Use this to measure the prefactor  $\Gamma_0$  that would make eqn 5.47 correct. What is the ratio of  $\Gamma_0$  to our rough estimate  $\omega$ ?

(5.32) **GHZ and  $\sigma$  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,  $(X_AY_BY_C)(Y_AX_BY_C)(YAY_BX_C)$  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

$$\Omega = (X_AY_BY_C)(Y_AX_BY_C)(YAY_BX_C) \quad (5.49)$$

In this exercise you shall show that  $\Omega$  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  $\pm 1$ , their squares are the identity (e.g.,  $Y_B^2 = \mathbb{1}$ ).

(a) Using only this fact, rewrite  $\Omega$  of eqn (5.49) 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

<sup>56</sup>Thus the spin operator  $\sigma_z = (\hbar/2)Z$ . Also,  $X$ ,  $Y$ , or  $Z$  acting on a spin at another

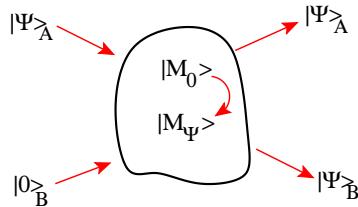
that site<sup>56</sup> gives  $Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ . Similarly, in the  $z$  basis  $X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$  and  $Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$ .

(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  $XXX$ ,  $XY\bar{Y}$ ,  $Y\bar{X}Y$ , 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  $\psi$  will share the property of the GHZ state that  $\Omega|\psi\rangle = -XXX|\psi\rangle$ .

(5.33) **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.<sup>57</sup>



**Fig. 5.17 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  $\psi$ .

Pretend it was possible. In particular, let us start with three subsystems (Fig. 5.17):  $A$  (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  $|M_0\rangle_M = \mathcal{M}_0(\mathbf{Z})$ . 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  $\Psi(\mathbf{x}, \mathbf{y}, \mathbf{Z}) =$

site is the identity.

<sup>57</sup>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

$\psi(\mathbf{x})\beta(\mathbf{y})\mathcal{M}_0(\mathbf{Z})$  into

$$U_c|\Psi\rangle = \psi(\mathbf{x})\psi(\mathbf{y})\mathcal{M}_\psi(\mathbf{Z}) \quad (5.50)$$

where  $\mathcal{M}_\psi(\mathbf{Z})$  is the final state of the machine after it processes the state  $\Psi$ . 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 (5.50) 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.

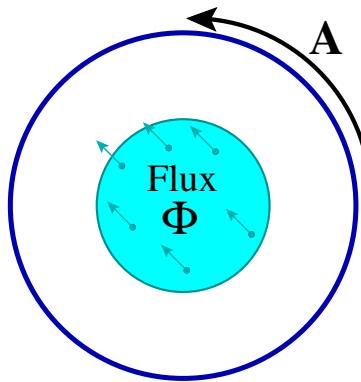
Suppose our initial cloning machine acts on two orthonormal input states  $\psi(\mathbf{x})$  and  $\phi(\mathbf{x})$ , so in addition to eqn (5.50), we have

$$U_c|\Phi\rangle = \phi(\mathbf{x})\phi(\mathbf{y})\mathcal{M}_\phi(\mathbf{Z}) \quad (5.51)$$

(b) Let  $U_c$  act on the superposition  $(1/\sqrt{2})(|\Psi\rangle + |\Phi\rangle)$ ; expand it using linearity and eqns (5.50) and (5.51). 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  $\psi$ ,  $\phi$ , and their superposition. No general-purpose cloning operator is possible over a Hilbert space with more than one dimension.

- (5.34) **Aharonov-Bohm wire: straightforward approach.** (Parallel Transport) ③



**Fig. 5.18 Aharonov-Bohm wire loop.** A circular wire of radius  $R$  encloses a magnetic field of total flux  $\Phi$ .

An electron is confined to a one-dimensional wire loop of radius  $R$  (Fig. 5.18). 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

$$\begin{aligned} E\psi(s) &= -\frac{\hbar^2}{2m} \left( \frac{\partial}{\partial s} - \frac{ieA}{\hbar c} \right)^2 \psi(s) \\ &= -\hbar^2/(2m) \left( \frac{\partial}{\partial s} - \frac{ie\Phi}{2\pi R\hbar c} \right)^2 \psi(s) \end{aligned}$$

where  $e < 0$  is the charge on the electron. The energy eigenstates are of the form  $\psi_n(s) \propto \exp(ik_n s)$ . 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  $\Phi$ .)
- (b) *Show that  $\psi_n$  is an eigenstate of  $(\partial/\partial s) - ieA/\hbar c$ . 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.)

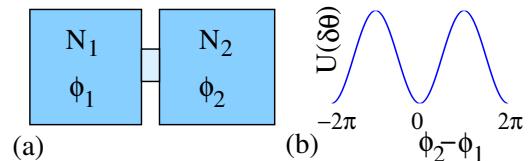
- (5.35) **Number, Phase, and Josephson.** (Quantum commutation) ③

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. (5.19a). Let the number of bosons (charge  $2e$  Cooper pairs) in the two pieces be  $N_1$  and  $N_2$ , and the superconducting phases be  $\phi_1$  and  $\phi_2$ . 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  $1 - \cos(\delta\phi)$ ,

$$U(\delta\phi) = U_0(1 - \cos(\delta\phi)), \quad (5.52)$$

as plotted in Fig. (5.19b).



**Fig. 5.19 Josephson junction schematic.** (a) A Josephson junction is made from a weak link (light blue 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,  $[\phi, N] = i$ , similar to how position and momentum are conjugate  $[x, p] = i\hbar$ . 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  $I/(2e) = dN_2/dt = -dN_1/dt$  which depends on  $\delta\phi$ :

$$I = I_c \sin(\delta\phi). \quad (5.53)$$

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 (5.52) that  $d(\delta N)/dt = I(\delta\phi)/(2e) \propto -dU/d\delta\phi$ , 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  $\delta N$  up to an overall constant  $x$ .

- (a) Use the general Heisenberg relation  $dO/dt = (1/i\hbar)[O, \mathcal{H}]$ , and assume that  $\mathcal{H}$  is some (so far) unknown function of  $\delta N$  (the ‘kinetic energy’) plus  $U(\delta\phi)$  (the ‘potential energy’). Show that  $\delta N = (x/i)\partial/\partial(\delta\phi)$  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  $\delta\phi = (y/i)\partial/\partial\delta N$  by showing that  $[\delta\phi, \delta N^m] = (y/i)m\delta N^{m-1}$  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 5.53), this implies  $d(\delta\phi)/dt = 2eV/\hbar$ . 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  $\delta N$ . We can use this to deduce the constant  $x$ .

- (c) Use  $d(\delta\phi)/dt = (1/i\hbar)[\delta\phi, \mathcal{H}]$  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  $[\delta\phi, \delta N] = i$ , 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 clas-

sical evolution of the current and phase, have important *quantum* fluctuations?

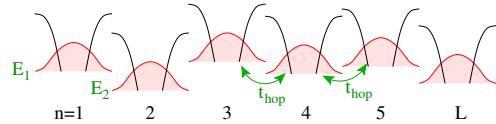
Just as  $[x, p] = i\hbar$  implies  $\sigma_x \sigma_p \geq \hbar/2$ ,  $[\delta\phi, \delta N] = i$  implies  $\sigma_{\delta\phi} \sigma_{\delta N} \geq \frac{1}{2}$ . Under normal circumstances, the number of Cooper pairs  $\delta N$  exchanged between the superconductors is enormous. In many ways, 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. (5.19a) be of thickness  $t = 10\text{nm}$ , 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  $C = 0.02 \pm 0.005\text{pF}$ , and a critical current of  $I_c = 55\mu\text{A}$ . 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.

### (5.36) Localization. (Disorder) ③

In this exercise, we study *localization* – a property of quantum wavefunctions in strongly disordered environments.



**Fig. 5.20 Anderson model of localization.** Our state  $\psi_n$  represents the amplitude that orbital  $n$  is occupied. Our Hamiltonian provides a hopping matrix element  $t_{\text{hop}}$  between neighboring sites on a one-dimensional chain, and a random energy distribution of on-site energies  $E_n$ .

In the study of the electronic properties of solids, it is often useful to work neither in a position basis nor a momentum basis, but in a basis of *orbitals*: orthogonal states situated on atoms numbered  $n$  in the material. These orbitals approximate atomic energy levels. Let us consider a one-dimensional array of atoms, with one orbital per atom. In the Hamiltonian, we have a *hopping* energy  $-t_{\text{hop}}$  for jumping an electron from one atom to a neighboring atom. If these atoms are of different types, or if they are subject to a random potential energy from nearby defects, the energies of the atomic states  $E_n$  would vary one from the other. We shall assume these energies are randomly chosen with standard deviation  $W$  and zero mean. Our single-particle Hamiltonian for non-interacting electrons is thus<sup>58</sup>

$$\mathcal{H} = \begin{pmatrix} E_1 & -t_{\text{hop}} & 0 & 0 & \dots & 0 \\ -t_{\text{hop}} & E_2 & -t_{\text{hop}} & 0 & \dots & 0 \\ 0 & -t_{\text{hop}} & E_3 & -t_{\text{hop}} & \dots & 0 \\ 0 & 0 & -t_{\text{hop}} & E_4 & \dots & 0 \\ \dots & & & & & \\ 0 & 0 & 0 & 0 & -t_{\text{hop}} & E_{L-1} \\ 0 & 0 & 0 & 0 & 0 & -t_{\text{hop}} \end{pmatrix} \quad (5.54)$$

This is known as the *Anderson model of localization*. In this exercise, we ignore the electron spin and electron-electron interactions. Also, we do *not* use periodic boundary conditions: there is no hopping between site  $L$  and site 1.

(a) Write the corresponding many-body Hamiltonian in second quantized form, using the anticommuting Fermion creation and annihilation operators  $c_n^\dagger$  and  $c_n$  (where, for example,  $c_n^\dagger$  creates an electron in the orbital at site  $n$ ). (Hint: You can probably find this in the literature, if you are having trouble, but reference your

sources.) Make sure your Hamiltonian is a Hermitian operator, in particular checking that the anticommuting operators are in the correct order.

The anticommutation relations enforce the antisymmetry for identical fermions. For a non-interacting Hamiltonian, the many-body eigenstates are antisymmetrized products of single-particle eigenstates obeying Pauli exclusion, so we will not need to use second-quantized operators in the rest of this exercise.

Write a routine, in the language of your choice, to generate the  $L \times L$  matrix  $\mathcal{H}$ , with random elements  $E_n$  generated from a Gaussian normal distribution of standard deviation  $W$ . In your routine, please set the seed of your random number generator to 6572 to facilitate grading. Write a routine that calculates the propagator  $K(i, j; t) = \exp(-i\mathcal{H}t)_{ij}$ . (Here we use units so that  $\hbar = 1$ . In most modern programming environments, you should be able to find a built-in function that calculates the matrix exponential.) You can think of  $i$  and  $j$  as corresponding to positions  $x, x'$  in the more traditional propagator (notationally we have changed  $K(x', t', x, t) \rightarrow K(i, j; t' - t)$ ).

How does the smoothly evolving free-particle packet you studied in exercise (3.6) change when we add disorder? Propagating electrons will scatter off the disorder.

0(b) Using  $L = 100$ ,  $t_{\text{hop}} = 1$ , and  $W = 0.1$ , animate the plot of the real and imaginary parts of  $0\psi_n(t) = K(1, n; t)$  for  $t$  running from zero to 10. For Python and other zero-based array environments, of course, use  $(0, \dots)$  rather than  $(1, \dots)$ ,  $-t_{\text{hop}}$  here and in future sections. Plot the real and imaginary parts of  $\psi_n(t = 10)$ . For what initial condition is this the solution to Schrödinger's equation?

What is the difference between a disordered metal and a disordered insulator? In a disordered metal, an electron will diffuse, eventually filling space. In a disordered insulator, an electron starting at one edge of a large system will forever remain near that edge. Is our Hamiltonian a metal or an insulator? Let us examine the current as a function of time.

The rate of change of probability  $|\psi_n|^2$  at a site  $n$  is the difference between the current  $J_{n-1 \rightarrow n}$

<sup>58</sup>For languages like Python, where arrays start at zero rather than one, the diagonal sites would naturally be labeled  $E_0, E_1, \dots$

from the left and the current  $J_{n \rightarrow n+1}$  to the right.

(c) Give the expression for  $d|\psi_n|^2/dt$  using Schrödinger's equation using  $\mathcal{H}$  from eqn (5.54), using the evolution laws for  $\psi_n$  and  $\psi_n^*$ . Derive the formula for the current

$$J_{n \rightarrow n+1} = (t_{\text{hop}}/\hbar)(\psi_n^* \psi_{n+1} - \psi_n \psi_{n+1}^*), \quad (5.55)$$

and note that it is independent of the disorder. (The time evolution of the  $\psi_n$ , naturally, will depend on the disorder.) Show that this is proportional to  $\psi^* D\psi - \psi D\psi^*$ , the term in the continuum current density,<sup>59</sup> for a suitable discrete derivative  $D$ .

Write a routine to calculate  $J$  from the solution  $K(1, n; t)$  of part (b).

(d) Using  $L = 100$ ,  $t_{\text{hop}} = 1$ , and  $W = 0.1$ , animate the plot of  $J_{n \rightarrow n+1}(n, t)$  for  $t$  running from zero to 100, again using  $\psi_n(t) = K(1, n, t)$ . Does it appear to diffuse over the entire system? Change to  $W = 1.0$ . Is the behavior different? For both cases, plot  $J_{n \rightarrow n+1}(n, t = 100)$ .

We see that for large disorder, our one-dimensional chain appears not to transmit current. We can gain insight by looking at the energy eigenstates. *Localized* eigenstates are centered on a few atoms, and die away exponentially away from them. *Extended* eigenstates are spread irregularly over the entire system.

(e) Using  $L = 100$ ,  $t_{\text{hop}} = 1$ , and  $W = 0.1$ , calculate the eigenenergies and eigenstates for  $\mathcal{H}$ . Plot the lowest energy eigenstate. Does it appear localized or extended? Plot the eigenstate closest to  $E = 0$ . Does it appear localized or extended? Repeat both at  $W = 1.0$ .

Our time-dependent wavefunction in parts (b) and (d) started in a superposition of all eigenstates with weight at the origin. Some of the eigenstates you found in part (e) appeared lo-

calized, and some extended.<sup>60</sup> In general, one can show that there is a sharp energy boundary between the localized and extended states. In a metal under a small voltage, it is only the states near the Fermi energy that matter for conductivity; hence if the states at the Fermi energy are extended the system is metallic, if localized it is insulating. Thus we would like to study the propagation of electrons in a way that focuses on particular energy ranges. We could do that by producing wave-packets that have narrow energy ranges, but let us instead consider the Green's function

$$G(i, j; E) = \lim_{\epsilon \rightarrow 0} -i \int_0^\infty \frac{dt}{\hbar} e^{i(E+i\epsilon)t/\hbar} K(i, j, t), \quad (5.56)$$

which takes the time evolution  $K$  from site  $j$  at time zero to site  $i$ , and pulls out the piece with energy  $E$ .

Let  $\mathcal{H}_0$  be the Hamiltonian setting  $t_{\text{hop}} = 0$  in eqn 5.54, and  $G_0$  the corresponding Green's function.

(f) Using eqn (5.56), analytically solve for  $G_0$ . (Hint: It will be a diagonal matrix.) Do you reproduce Anderson's formula  $G = 1/(E - \mathcal{H})$ ? Does the Green's function  $G_0$  have non-zero matrix elements connecting site 1 to site  $L$ ?

Let us use Dyson's formula to calculate perturbative approximations to  $G_0$  in the limit when  $t_{\text{hop}}$  is small.<sup>61</sup> Let  $\mathcal{H} = \mathcal{H}_0 + I$ , where  $I$  consists of the off-diagonal  $-t_{\text{hop}}$  terms and  $\mathcal{H}_0$  is diagonal. Dyson's equation (in the long-winded version) told us that

$$G = G_0 + G_0 I G_0 + G_0 I G_0 I G_0 + G_0 I G_0 I G_0 I G_0 + \dots \quad (5.57)$$

(g) Give an explicit formula for  $G_{ij}$  to second order in  $t_{\text{hop}}$ , keeping the first three terms in Dyson's equation (5.57). It should be zero for most pairs of sites  $(i, j)$  at this order. You only

<sup>59</sup>Remember the continuum current density is  $j = \frac{\hbar}{2mi}(\psi^* \partial\psi/\partial x - \psi \partial\psi^*/\partial x)$ .

<sup>60</sup>As it happens, in one dimension, one can prove that all states are localized for an infinite system. But the behavior seen here, although due to finite size effects, is roughly similar to that one would see in a three-dimensional infinite system.

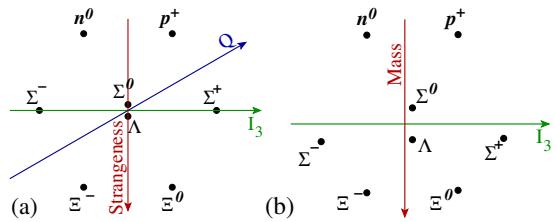
<sup>61</sup>Note that our ‘unperturbed’ Hamiltonian is a set of independent atomic orbitals – an excellent zeroth approximation for a disordered insulator. The more traditional way of simplifying Hamiltonians for metals is to set the disorder  $W = 0$ . That would give us plane-wave states as eigenstates, and is the natural starting point for understanding extended, conducting states. To get from one description to the other demands going beyond perturbation theory (*e.g.*, summing an infinite number of diagrams). This is due to the simple reason that insulators and metals are different phases, and the phase boundary between them has singularities that provide natural limits to perturbation theory. Here the boundary is the *mobility edge* – the single-particle energy dividing the extended and localized states.

need to calculate the matrix elements for an infinite system, or equivalently matrix elements connecting sites far from either boundary. What is the maximum distance  $m$  from site  $n$  that has a non-zero matrix element  $G_{n,n+m}$  to second order? How many terms in perturbation theory would we need to connect the boundary site 1 to the other boundary site  $L$  with a non-zero  $G_{1,L}$ ? (h) Argue from dimensional analysis, and your form for  $G_0$ , that the  $M^{\text{th}}$  order term in Dyson's series for  $I$  will be composed of terms of the form  $t_{\text{hop}}^M / (\prod_{m=1}^{M+1} (E - E_{n_m}))$ , for some set of integers  $n_m \in 1, \dots, L$ . (Hint for part (g): Check that your first two orders in perturbation theory are of this form.)

Note again the low-energy states you found in part (e). Each is centered near some site  $n_0$ , and should decay rapidly as one moves away from  $n_0$ . The states at energy near zero for large disorder should decay rapidly away from their center too, but in an oscillatory and irregular manner. (Hint: Check your answers to see if this is true.) In general, a localized state centered at  $n_0$  will decay very roughly as  $|\psi_n|^2 \sim \exp(-|n - n_0|/\lambda)$ , where  $\lambda$  is the *localization length*.

(i) Make a very rough estimate of the localization length, given  $t_{\text{hop}}/W \ll 1$ . (By rough, I suggest assuming that  $E - E_n \approx W$  in part (h).) What happens to your approximation when  $t_{\text{hop}}/W > 1$ ? In what ranges of  $t_{\text{hop}}/W$  would you expect localization?

(5.37) **Eight-Fold Way.** (Particles, Group reps) ③



**Fig. 5.21 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  $I_3$ .<sup>62</sup> Thus for example the neutron has isospin  $1/2$ , and  $\Sigma^-$  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 5.21a). Figure (5.21b) shows the masses of one family of baryons, indicating the original experimental evidence for the model.

In this exercise,<sup>63</sup> we will explore the geometry shown in Fig. (5.21a), 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  $2^{1/2}$  electron masses different. A little lighter, and the neutron could not decay.) Werner Heisenberg made an analogy with spin-

<sup>62</sup>The nomenclature is confusing. A spin  $1/2$  particle has total spin  $1/2$ , but may have  $z$ -component of spin either  $\pm 1/2$ ; sometimes one says loosely that the spin of a particle is  $-1/2$ . It is always said that the neutron has isospin  $-1/2$ , but we call it  $I_3$  analogous to  $s_z$ . In this exercise, we will say that the proton and neutron have ‘total isospin’  $1/2$ .

<sup>63</sup>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](http://www.ks.uiuc.edu/Services/Class/PHYS480/qm_PDF/chp12.pdf)).

<sup>64</sup>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).

$\frac{1}{2}$  particles – the proton has isospin  $I_3 = +\frac{1}{2}$  and the neutron isospin  $I_3 = -\frac{1}{2}$ .

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  $\frac{1}{2}$  and isospin  $\frac{1}{2}$ , thus would have four degenerate states ( $n^\uparrow, n^\downarrow, p^\uparrow$ , and  $p^\downarrow$ ).<sup>64</sup>

(a) Draw a figure corresponding to Fig. (5.21a) 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 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.<sup>65</sup> 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<sup>66</sup> 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 (5.21a) 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  $J_\pm = J_x \pm iJ_y$  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).<sup>67</sup> 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  $j_z = (-j, \dots, j)$ . We will focus here just on generalizing these ladder operators to SU(3).

The three quarks

$$|u\rangle = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad |d\rangle = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad |s\rangle = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \quad (5.58)$$

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 group and its traditional representation as a set of  $3 \times 3$  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  $I_3 = \begin{pmatrix} \frac{1}{2} & 0 & 0 \\ 0 & -\frac{1}{2} & 0 \\ 0 & 0 & 0 \end{pmatrix}$

<sup>65</sup>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.

<sup>66</sup>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.

<sup>67</sup>As a quick review, this works because  $J_z |j_z \pm 1\rangle = J_z J_\pm |j_z\rangle = (J_\pm J_z + [J_z, J_\pm]) |j_z\rangle = J_\pm \hbar j_z |j_z\rangle \pm \hbar J_\pm |j_z\rangle = \hbar (j_z \pm 1) J_\pm |j_z\rangle = \hbar (j_z \pm 1) |j_z \pm 1\rangle$ .

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. (5.21a),  $I_\pm$  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. (5.21a) up-and-to-the-right at roughly<sup>68</sup>  $60^\circ$ . The subgroup mixing down and strange forms a third pair of ladders  $U_\pm$  moving at roughly  $-60^\circ$ .

(c) Show that  $[I_3, V_+] = \frac{1}{2}V_+$ , using the representation **3** for the generators above.

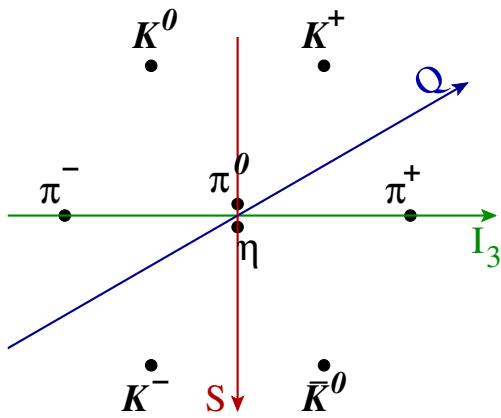
The three-fold symmetry of the raising and lowering operators gives us the three-fold symmetry of the baryon octet of Fig. (5.21a).

(d) If  $V_+|\Sigma^-\rangle = \alpha|n^0\rangle$  for some constant  $\alpha$ , 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  $\Sigma^-$ .) 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. (5.21b) 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<sup>69</sup> under isospin  $SU(2)$ , written in terms of the three quark directions in eqn (5.58)? (Hint: Which quarks are mixed by the isospin symmetry group? This is not subtle.) Show that<sup>70</sup> the  $SU(3)$  representation **3** decomposes into 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. 5.22 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. (5.21a) from the standard convention, which would have the neutron and proton have highest (least negative) strangeness of their family of baryons.)

<sup>68</sup> To make the diagram in Fig. (5.21a) 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  $\Sigma^-$  and  $n^0$  up-and-to-the-right. Because we plot strangeness  $S$  along the vertical axis rather than the more symmetrical  $\sqrt{3}S/2$ , moving along  $V_+$  actually is at  $63.43^\circ$ .

<sup>69</sup>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.

<sup>70</sup> (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 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  $SU(2)$ .)

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. 5.22). 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  $\mathbf{1}$  is a one-dimensional representation corresponding in this case to the  $\eta'$  meson.

We need not refer to character orthogonality relations to find the isospin decomposition of  $\mathbf{8}$ . Given that we know how the two three dimensional  $SU(3)$  reps  $\mathbf{3}$  and  $\bar{\mathbf{3}}$  decompose when broken to  $SU(2)$ , we can deduce the decomposition of  $\mathbf{8}$ . 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  $\mathbf{1}$  must lead to a one-dimensional isospin-zero piece, the isospin subspaces of  $\mathbf{8}$  then can be deduced by elimination.<sup>71</sup>

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:

$$((\mathbf{a} \otimes \mathbf{c}) \oplus (\mathbf{b} \otimes \mathbf{c}))_{\omega,\gamma} = \mathbf{a}_{\omega} \mathbf{c}_{\gamma} \text{ if } \omega \leq d_A \text{ and}$$

$$((\mathbf{a} \otimes \mathbf{c}) \oplus (\mathbf{b} \otimes \mathbf{c}))_{\omega,\gamma} = \mathbf{b}_{\omega-d_A} \mathbf{c}_{\gamma} \text{ if } \omega > d_A.$$

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

<sup>71</sup>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  $a_{\alpha}$  and  $c_{\gamma}$  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  $\mathbf{a}$  with  $\mathbf{b}$  of dimension  $d_B$  is a vector  $\mathbf{a} \oplus \mathbf{b}$  of dimension  $d_A + d_B$  with first  $d_A$  components  $a_{\alpha}$  followed by  $d_B$  components  $b_{\beta}$ . 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  $90^\circ$ , 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  $\Psi(\mathbf{x}, \mathbf{y}) \rightarrow \Psi(Rx, Ry)$  rotates single-particle product wavefunctions  $\chi(\mathbf{x})\phi(\mathbf{y}) \rightarrow \chi(Rx)\phi(Ry)$ ; 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  $3 \times 5 = 15$ .

$(R_A[g]\mathbf{a} \otimes R_c[g]\mathbf{c}) \oplus (R_B[g]\mathbf{b} \otimes R_C[g]\mathbf{c})$ .<sup>72</sup>

$$\begin{aligned} ((R_A[g]\mathbf{a} \oplus R_B[g]\mathbf{b}) \otimes R_c[g]\mathbf{c})_{\omega,\gamma} &= (R_A[g]\mathbf{a})_{\omega}(R_c[g]\mathbf{c})_{\gamma} \\ ((R_A[g]\mathbf{a} \oplus R_B[g]\mathbf{b}) \otimes R_c[g]\mathbf{c})_{\omega,\gamma} &= (R_B[g]\mathbf{b})_{\omega-d_A}(R_c[g]\mathbf{c})_{\gamma} \end{aligned} \quad (5.59)$$

Similarly,

$$\begin{aligned} ((R_A[g]\mathbf{a} \otimes R_c[g]\mathbf{c}) \oplus (R_B[g]\mathbf{b} \otimes R_C[g]\mathbf{c}))_{\omega,\gamma} &= (R_A[g]\mathbf{a})_{\omega}(R_B[g]\mathbf{b})_{\omega-d_A}(R_C[g]\mathbf{c})_{\gamma} \\ ((R_A[g]\mathbf{a} \otimes R_c[g]\mathbf{c}) \oplus (R_B[g]\mathbf{b} \otimes R_C[g]\mathbf{c}))_{\omega,\gamma} &= (R_B[g]\mathbf{b})_{\omega-d_A}(R_C[g]\mathbf{c})_{\gamma} \end{aligned}$$

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 and

reducible representations of the isospin subgroup  $SU(2)$ . (Hint: There should be five invariant

subspaces of this nine-dimensional space.)

How many of the isospin subspaces from Fig. (5.22) should be part of the octet? Give the total isospin and  $I_3$  for each meson, grouping them into isospin multiplets (the mesons forming the different representations).

<sup>72</sup>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.

# References

- [1] Alemi, A. A., Bierbaum, M. K., Myers, C. R., and Sethna, J. P. (2015, Nov). You can run, you can hide: The epidemiology and statistical mechanics of zombies. *Phys. Rev. E*, **92**, 052801. [6.6](#), [6.6](#), [72](#)
- [2] Anderson, P. W. (1972). More is different. *Science*, **177**(4047), 393–396. [1.1](#)
- [3] BekensteinD, J. D. (1973). Black holes and entropy. *Physical Review D*, **7**(8), 2333–46. [5.5](#)
- [4] Berman, G. J., Bialek, W., and Shaevitz, J. W. (2016). Predictability and hierarchy in drosophila behavior. *Proceedings of the National Academy of Sciences*, **113**(42), 11943–11948. [8.28](#)
- [5] Berman, G. J., Choi, D. M., Bialek, W., and Shaevitz, J. W. (2014). Mapping the stereotyped behaviour of freely moving fruit flies. *Journal of The Royal Society Interface*, **11**(99), 20140672. [8.8](#)
- [6] Berman, G. J., Choi, D. M., Bialek, W., and Shaevitz, J. W. (2014). Mapping the structure of drosophilid behavior. *bioRxiv*. [8.8](#)
- [7] Bialek, W. (2012). *Biophysics: searching for principles*. Princeton University Press. [8.8](#)
- [8] Bialek, W. (2012). Biophysics: searching for principles, early draft. <http://www.princeton.edu/~wbialek/PHY562/WB-biophysics110918.pdf>. [8.8](#)
- [9] Bierbaum, M. K., Alemi, A. A., Myers, C. R., and Sethna, J. P. (2015). Zombietown usa. [http://hey.runat.me/pages/programming/zombietown\\_usa.html](http://hey.runat.me/pages/programming/zombietown_usa.html). [6.6](#), [6.6](#)
- [10] Bierbaum, M. K., Alemi, A. A., Myers, C. R., and Sethna, J. P. (2015). Zombietown usa simulator. <http://mattbierbaum.github.io/zombies-usa/>. [6.6](#), [6.6](#)
- [11] Buchel, A. and Sethna, J. P. (1996). Elastic theory has zero radius of convergence. *Physical Review Letters*, **77**, 1520. [116](#), [11.11](#)
- [12] Buchel, A. and Sethna, J. P. (1997). Statistical mechanics of cracks: Thermodynamic limit, fluctuations, breakdown, and asymptotics of elastic theory. *Physical Review E*, **55**, 7669. [116](#), [11.11](#)
- [13] Cardy, J. (1996). *Scaling and renormalization in statistical physics*. Cambridge University Press. [128](#), [129](#)
- [14] Chui, S. T. and Weeks, J. D. (1978, Mar). Dynamics of the roughening transition. *Phys. Rev. Lett.*, **40**, 733–736. [9.38](#)

- [15] Crooks, G. E. (1998). Nonequilibrium measurements of free energy differences for microscopically reversible markovian systems. *Journal of Statistical Physics*, **90**(5), 1481–7. [4.4](#), [4.4](#)
- [16] 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. [3.4](#)
- [17] Dyson, Freeman (1960). Search for artificial stellar sources of infrared radiation. *Science*, **131**, 1667–1668. [5.5](#)
- [18] Feigenbaum, M. (1979, 12). The universal metric properties of nonlinear transformations. *Journal of Statistical Physics*, **21**, 669–706. [12.12](#), [3.2](#)
- [19] Feynman, R. P., Leighton, R. B., and Sands, M. (1963). *The Feynman lectures on physics*. Addison-Wesley, Menlo Park, CA. [2.2](#), [22](#)
- [20] Fisher, M. E. and Randeria, M. (1986, May). Location of renormalization-group fixed points. *Phys. Rev. Lett.*, **56**, 2332–2332. [3.2c](#)
- [21] Frenkel, D. and Louis, A. A. (1992). Phase separation in a binary hard-core mixture. An exact result. *Physical Review Letters*, **68**, 3363. [6.6](#)
- [22] Greiner, W., Neise, L., and Stöcker, H. (1995). *Thermodynamics and statistical mechanics*. Springer, New York. [3.3](#)
- [23] Hopfield, J. J. (1974). Kinetic proofreading: a new mechanism for reducing errors in biosynthetic processes requiring high specificity. *Proceedings of the National Academy of Sciences*, **71**(10), 4135–4139. [8.29](#), [8.8](#), [8.8](#)
- [24] Jarzynski, C. (1997). Nonequilibrium equality for free energy differences. *Physical Review Letters*, **78**(14), 2690. [4.4](#)
- [25] Koehn, P. (2005). Europarl: A parallel corpus for statistical machine translation. In *MT summit*, Volume 5, pp. 79–86. [6.27](#)
- [26] Kos, F., Poland, D., Simmons-Duffin, D., and Vichi, A. (2016). Precision islands in the ising and  $\phi(n)$  models. *Journal of High Energy Physics*, **2016**(8), 36. [12.12](#)
- [27] Liphardt, J., Onoa, B., Smith, S. B., Tinoco, I., and Bustamante, C. (2001). Reversible unfolding of single RNA molecules by mechanical force. *Science*, **292** (5517), 733–7. [6.23](#), [6.24](#)
- [28] Lucas, A. Kinetic proofreading. <http://https://web.stanford.edu/~ajlucas/Kinetic\Proofreading.pdf>. [8.31](#)
- [29] Machta, B. B. (2015). Dissipation bound for thermodynamic control. *Phys. Rev. Lett.*, **115**, 260603. [6.6](#), [80](#)
- [30] Mao, J. and Hu, B. (1987). Corrections to scaling for period doubling. *Journal of Statistical Physics*, **46**, 111. [3.2b](#)
- [31] Mathews, J. and Walker, R. L. (1964). *Mathematical methods of physics*. Addison-Wesley, Redwood City, CA. [64](#)
- [32] Mézard, M. and Montanari, A. (To be published (2006)). *Constraint satisfaction networks in Physics and Computation*. Clarendon Press, Oxford. [32](#), [3.3](#), [34](#), [12.12](#), [3.3](#), [4](#), [7](#)
- [33] Miller, J., Weichman, P. B., and Cross, M. C. (1992). Statistical

- mechanics, Eulers equation, and Jupiters red spot. *Physical Review A*, **45**(4), 2328–59. [4.4](#)
- [34] NASA, ESA, (STScI), M. Postman, and the CLASH Team (2011). Galaxy cluster MACS 1206. <http://https://apod.nasa.gov/apod/ap111017.html>. [5.20](#)
- [35] NASA/JPL (2014). Image from voyager i, nasa's goddard space flight center. <http://http://www.nasa.gov/content/jupiters-great-red-spot-viewed-by-voyager-i>. [4.15](#)
- [36] Newman, M. E. J. (2005). Power laws, pareto distributions and zipf's law. *Contemporary Physics*, **46**(5), 323–351. [6.6](#)
- [37] Orlund, 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**, 303–342. [3.2f](#)
- [38] Otani, N. (2004). Home page. <http://otani.vet.cornell.edu>. [3.4](#)
- [39] Piantadosi, S. T. (2014). Zipfs word frequency law in natural language: A critical review and future directions. *Psychon Bull Rev.*, **21**, 11121130. [6.6](#)
- [40] Polyakov, A. M. (1993). The theory of turbulence in two dimensions. *Nuclear Physics B*, **396**(2-3), 367–85. [4.4](#)
- [41] Purcell, E. M. (1977). Life at low Reynolds number. *American Journal of Physics*, **45**, 3–11. [2.2](#)
- [42] Quinn, K. N., Bernardis, F. De, Niemack, M. D., and Sethna, J. P. (2017). Visualizing theory space: Isometric embedding of probabilistic predictions, from the Ising model to the cosmic microwave background. <http://arxiv.org/abs/1709.02000>. [1.1](#), [1.1](#)
- [43] Rand, D., Orlund, S., Sethna, J. P., and Siggia, E. D. (1982, Jul). Universal transition from quasiperiodicity to chaos in dissipative systems. *Phys. Rev. Lett.*, **49**, 132–135. [3.2f](#)
- [44] Robert, C. P. and Casella, G. (2005). *Monte Carlo Statistical Methods*. Springer Texts in Statistics. Springer New York. [4.4](#)
- [45] Sethna, J. P. (1997). Quantum electrodynamics has zero radius of convergence. <http://www.lassp.cornell.edu/sethna/Cracks/QED.html>. [11.11](#)
- [46] Sethna, J. P., Bierbaum, M. K., Dahmen, K. A., Goodrich, C. P., Greer, J. R., Hayden, L. X., Kent-Dobias, J. P., Lee, E. D., Liarte, D. B., Ni, X., Quinn, K. N., Raju, A., D. Z. Rocklin, Shekhawat, A., and Zapperi, S. (2017). Deformation of crystals: Connections with statistical physics. *Annual Review of Materials Research*, **47**, 217–246. [9.38](#)
- [47] Sethna, J. P., Chachra, R., Machta, B. B., and Transtrum, M. K. (2013). Why is science possible? <http://www.lassp.cornell.edu/sethna/Sloppy/WhyIsSciencePossible.html>. [1.2](#), [1.3](#)
- [48] Sethna, J. P., Dahmen, K. A., Kartha, S., Krumhansl, J. A., Roberts, B. W., and Shore, J. D. (1993, May). Hysteresis and hierarchies: Dynamics of disorder-driven first-order phase transformations. *Phys. Rev. Lett.*, **70**, 3347–3350. [12.12](#)
- [49] Sethna, J. P. and Myers, C. R. (2004). *Entropy, Order Parameters, and Complexity* computer exercises: Hints

- and software. <http://www.physics.cornell.edu/sethna/StatMech/ComputerExercises.html>. 38, 71, 95, 96, 97, 99, 112, 114, 12.12, 120, 122, 123, 130, 140, 4, 3.4
- [50] Skriver, H. (2004). The hls metals database. <http://http://databases.fysik.dtu.dk/hlsPT/>. 117
- [51] Smith, R. L. (2006). *Environmental Statistics*. 12.12, 124
- [52] Südkamp, T. and Bracht, H. (2016, Sep). Self-diffusion in crystalline silicon: A single diffusion activation enthalpy down to 755 °C. *Phys. Rev. B*, **94**, 125208. 9.9
- [53] Transtrum, M. K., Machta, B. B., and Sethna, J. P. (2010). Why are nonlinear fits to data so challenging? *Phys. Rev. Lett.*, **104**, 060201. 1.1
- [54] Transtrum, M. K., Machta, B. B., and Sethna, J. P. (2011). Geometry of nonlinear least squares with applications to sloppy models and optimization. *Phys. Rev. E*, **83**, 036701. 1.1
- [55] Trefethen, L. N. and Trefethen, L. M. (2000). How many shuffles to randomize a deck of cards? *Proceedings of the Royal Society of London A*, **456**, 2561–8. 5.5
- [56] Wallace, D. (2010). Gravity, entropy, and cosmology: In search of clarity. *Brit. J. Phil. Sci.*, **61**, 513–40. 5.5
- [57] Wilczek, F. and Shapere, A. (1989). *Geometric Phases in Physics*. World Scientific. 21
- [58] Winfree, A. T. (1991). Varieties of spiral wave behavior: An experimentalist’s approach to the theory of excitable media. *Chaos*, **1**, 303–34. 3.4
- [59] Zinn-Justin, J. (1996). *Quantum field theory and critical phenomena (3rd edition)*. Oxford University Press. 11.11