QCB 505/PHY 555: Topics in biophysics and quantitative biology

Fall 2020: Statistical mechanics for real biological networks

---

A first look at models — Sep 8, 2020

---

1 The Hopfield model

The idea that the brain is made out of discrete cells, connected by synapses, dates from the start of the twentieth century (Cajal 1906). The electrical signals from individual nerve cells (neurons) were first recorded in the 1920s, starting with the cells in sense organs that provide the input to the brain (Adrian 1928). The crucial observation is that neurons communicate by generating discrete, identical pulses of voltage across their membranes; these pulses are called action potentials or, more colloquially, spikes, and by the 1950s the dynamics underlying the generation and propagation of spikes was clear (Hodgkin and Huxley 1952). For guides to all of this, see Aidley (1998), Dayan and Abbott (2001), and Kandel et al (2012), as well as Rieke et al (1997).

Adrian 1928 *The Basis of Sensation.* ED Adrian (ChristoperÕs, London, 1928).

Aidley 1998 *The Physiology of Excitable Cells, Fourth Edition.* DJ Aidley (Cambridge University Press, Cambridge, 1998).

Cajal 1906 The structure and connexions of neurons. S Ramon y Cajal. Nobel lecture 1906, in *Nobel Lectures, Physiology or Medicine 1901–1921* (Elsevier, Amsterdam, 1967). Also available at http://www.nobelprize.org.

Dayan and Abbott 2001 *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems* P Dayan and LF Abbott (MIT Press, Cambridge, 2001).

Hodgkin and Huxley 1952 A quantitative description of membrane current and its application to conduction and excitation in nerve. AL Hodgkin and AF Huxley, *J Physiol (Lond)* **117,** 500–544 (1952).

Kandel et al 2001 *Principles of Neural Science, Fifth Edition.* ER Kandel, T Jessell, J Schwartz, S Siegelbaum, and AJ Hudspeth (McGraw–Hill, New York, 2012).

Rieke et al 1997 *Spikes: Exploring the Neural Code.* F Rieke, D Warland, RR de Ruyter van Steveninck, and W Bialek (MIT Press, Cambridge MA, 1997).

Even before the mechanisms were clear, people began to think about how the quasi–digital character of spiking could be harnessed to do computations, and this line of thought was crystallized in the remarkable paper by McCulloch and Pitts (1943). Their work made clear that computation in the brain must be the global output of activity in many neurons, and indeed we know that our perceptions and actions typically are the result of activity in many hundreds, perhaps even hundreds of thousands of neurons. It is impossible to see how these perceptions and actions could be coherent if all the neurons acted independently: there must be something coordinated or collective going on. The idea that collective neural activity in the brain might be described with statistical mechanics goes back at least to Norbert Wiener [see Wiener (1958), especially Lecture

8], who was very much influenced by observations on the electroencephalogram (EEG). The EEG is a macroscopic measure of activity, traditionally done simply by placing electrodes on the scalp, and the existence of the EEG is prima facie evidence that the electrical activity of many, many neurons must be correlated.[1] In the 1960s and 70s there was a more focused effort to create statistical physics models for networks of neurons, notably by Cooper (1973) and by Little, Shaw, and colleagues (Little 1974; Shaw and Vasudevan 1974; Little and Shaw 1975, 1978; Shaw and Roney 1979).[2]

Cooper 1973 A possible organization of animal memory and learning. LN Cooper, in *Collective Properties of Physical Systems: Proceedings of Nobel Symposium 24*, B Lundqvist and S Lundqvist, eds, pp. 252–264 (Academic Press, New York, 1973).

Little 1974 Existence of persistent states in the brain. WA Little, *Math Biosci* **19,** 101–120 (1974).

Little and Shaw 1975 A statistical theory of short and long term memory. WA Little and GL Shaw, *Behav Biol* **14,** 115 (1975).

Little and Shaw 1978 Analytic study of the memory storage capacity of a neural network. WA Little and GL Shaw, *Math Biosci* **39,** 281–290 (1978).

McCulloch and Pitts 1943 A logical calculus of ideas immanent in nervous activity. WS McCulloch and W Pitts, *Bull Math Biophys* **5,** 115–133 (1943).

Shaw and Roney 1979 Analytic solution of a neural network theory based on an Ising spin system analogy. GL Shaw and KJ Roney, *Phys Lett A* **74,** 146–150 (1979).

Shaw and Vasudevan 1974 Persistent states of neural networks and the random nature of synaptic transmission. GL Shaw and R Vasudevan, *Math Biosci* 21, 207 (1974).

Wiener 1958 *Nonlinear Problems in Random Theory.* N Wiener (MIT Press, Cambridge MA, 1958).

The key step forward was the work of Hopfield (1982, 1984). As has been widely noted, he was able to step forward because he took a giant step backward. Specifically, he studied highly schematic models for the dynamics of a neural network, not so different from those studied by McCulloch and Pitts and many others, but he made the additional simplifying assumption that if neuron i is connected to neuron j, then the opposite also is true, and the strength of the connection is the same. Thus, the network is symmetric, and in this case the dynamics have a Lyapunov function: the dynamics of neural computation corresponds to sliding downhill on an energy surface. In the first examples, the goal of computation was to recover a stored pattern from partial information (associative memory), but Hopfield and Tank (1985) soon showed that one could construct networks that solve classical optimization problems, and that many biologically relevant problems could be cast in this form (Hopfield and Tank 1986). Textbook accounts of these and related ideas appeared quickly (Amit 1989, Hertz et al 1991), and there is much to be explored beyond what I have summarized here.

[1] There is also the remarkable story of a demonstration by Lord Adrian, in which he sat quietly with his eyes closed with electrodes attached to his head. The signals, sent to an oscilloscope, showed the characteristic "alpha rhythm" that occurs in resting states, roughly an oscillation at $\sim 10\,\mathrm{Hz}$. When asked to add two numbers in his head, the rhythm disappeared, replaced by less easily described patterns of activity. Just in case you had any doubts that your mental life is related to the electrical activity of your brain.

[2] If you are interested in the history of physics, you should get your hands on the volume in which Cooper's paper appears. A Nobel Symposium on collective phenomena, it was held just after the discovery of superfluid ^{3}He and the crystallization of renormalization group ideas. In addition to papers on these topics, there is Cooper on neural networks, Hopfield on cooperativity in hemoglobin, and more, all summarized at the end by Anderson.

Amit 1989 *Modeling Brain Function: The World of Attractor Neural Networks.* DJ Amit (Cambridge University Press, Cambridge, 1989).

Hertz et al 1991 *Introduction to the Theory of Neural Computation.* J Hertz, A Krogh, and RG Palmer (Addison–Wesley, Redwood City, 1991).

Hopfield 1982 Neural networks and physical systems with emergent collective computational abilities. JJ Hopfield, *Proc Natl Acad Sci USA* **79,** 2554–2558 (1982).

Hopfield 1984 Neurons with graded response have collective computational properties like those of two-state neurons. JJ Hopfield, *Proc Natl Acad Sci USA* **81,** 3088–3092 (1984).

Hopfield and Tank 1985 "Neural" computation of decisions in optimization problems. JJ Hopfield and DW Tank, *Biol Cybern* **52,** 141–152 (1985).

Hopfield and Tank 1986 Computing with neural circuits: A model. JJ Hopfield and DW Tank, *Science* **233,** 625–633 (1986).

Hopfield's work ignited an explosion. It's not always easy to see why these things happen, but in this case I think we can point not only to the ideas in Hopfield's paper, but also to many things that were happening at the same time in nearby fields. The Lyapunov function in the Hopfield model of associative memory is a special sort of spin glass—a collection of Ising spins with competing ferromagnetic and anti–ferromagnetic interactions, with frustration generating multiple local minima of the energy, corresponding to the multiple stored memories. Shortly before Hopfield's work, Parisi had solved the mean–field spin glass, and his discovery/invention of replica symmetry breaking itself touched off a huge stream of work on the statistical mechanics of disordered systems (Mézard et al 1987). These tools were immediately applied, for example, to the calculate the memory capacity of the Hopfield model (Amit et al 1985, 1987). At the same time, the idea of simulated annealing (Kirkpatrick et al 1983) led people to take much more seriously the mapping between "computational" problems of optimization and the "physical" problems of finding minimum energy states of many–body systems. From the biological point of view, the Hopfield model stores memories in the pattern of synaptic connections, and there is a straightforward algorithm for updating these parameters to program in a set of desired patterns; this "learning rule" matched surprisingly well with classical ideas from Hebb (1949) and with emerging data on the plasticity of real synapses. From an engineering point of view, models for neural networks connected immediately to the possibility of using modern chip design methods to build analog, rather than digital circuits (Mead 1989). Thus, the Hopfield model formed a nexus among statistical physics, computer science, neurobiology, and engineering.

Amit et al 1985 Spin–glass models of neural networks. DJ Amit, H Gutfreuend, and H Sompolinsky, *Phys Rev A* **32,** 1007–1018 (1985).

Amit et al 1987 Statistical mechanics of neural networks near saturation. DJ Amit, H Gutfreuend, and H Sompolinsky, *Ann Phys* **173,** 30–67 (1987).

Hebb 1949 *The Organization of Behavior: A Neuropsychological Theory.* DO Hebb (John Wiley and Sons, New York, 1949).

Kirkpatrick et al 1983 Optimization by simulated annealing. S Kirkpatrick, CD Gelatt Jr, and MP Vecchi, *Science* **220,** 671–680 (1983).

Mead 1989 *Analog VLSI and Neural Systems.* CA Mead (Addison–Wesley, Redwood City CA, 1989).

Mézard et al 1987 *Spin Glass Theory and Beyond.* M Mézard, G Parisi, and MA Virasoro (World Scientific, Singapore, 1987).

2 Monte Carlo

These notes (and in the following section) are taken from my lectures in PHY301. When I taught, I assigned problems, and for consistency we did everything in MATLAB. If starting again today I would probably do this in Python, but I hope that isn't an obstacle to understanding.

If we are going to understand why simple descriptions of macroscopic systems are possible, then we need a way of exploring more complex models. You'll notice that, so far, we have described models of ideal gases, where there are many particles but they don't interact. Suppose I told you that there were interactions among all the particles in a system, perhaps even strong interactions. How would you calculate the partition function?

Except is special cases, exact solutions of interacting models are impossible. There are many approximation schemes, and we will see that the renormalization group helps us, substantially, to understand what is possible even if it remains difficult to connect particular microscopic models to one of these possibilities. But, there is also an alternative, and this is to use the computer.

The fundamental object in equilibrium statistical mechanics is the Boltzmann distribution. If we could draw states out of this distribution, many times, then we could do a "numerical experiment" and calculate anything we want to know by averaging over these many samples. If we had a general way of doing this it would be even more widely applicable, since there are many situations in which it would be useful to be able to draw samples out of a complicated distribution. Today, people are making models for the probability distribution of things ranging from DNA sequences to sentences to internet traffic to stock prices, and being able to simulate these models is central to all of this work. The basic tool is called "Monte Carlo," a reference to the casino.

In the simplest version of these ideas, we want to "walk" at random through the possible states of a system, but we want to be sure that, in the long run, we visit each state with a probability equal to that predicted by a particular model, in our case the Boltzmann distribution. As we prepare to explore how this is done, notice that the Boltzmann distribution doesn't tell us the absolute probability of being in a state unless we can calculate the partition function, and this is *very* hard in the general case. So we will need to find a way of using just the energy of each state, not the free energy of the system as a whole.

Let's be concrete and label states by s, each with energy E_s. Our walk through the possible states will be described by transition probabilities $\mathcal{T}(s \to s')$, which give the probability of jumping from state s to state s'. Clearly this walk has a chance of simulating an arbitrary distribution only if, eventually, it can visit every state with nonzero probability. This is not so hard. But how can we encode the energies of the states in the transition probabilities?

Imagine that we could simulate many systems in parallel, so many, in fact, that just by counting we could determine the probabilities of each and every state P_s. Now we let every system take one jump according to the transition probability $\mathcal{T}(s \to s')$. How much does the probability P_s for one particular state change? It goes down, because there are many states s' that this one state could jump into. This loss in probability is proportional to the initial probability, so we have

$$\Delta P_s(\text{loss}) = \sum_{s'} \mathcal{T}(s \to s') P_s. \tag{1}$$

On the other hand, systems that were in state s' can jump into state s with probability $\mathcal{T}(s' \to s)$, and this leads to a gain

$$\Delta P_s(\text{gain}) = \sum_{s'} \mathcal{T}(s' \to s) P_{s'}. \tag{2}$$

If the system is really in equilibrium, with states drawn from the Boltzmann distribution, we will have

$$P_s^{\rm eq} = \frac{1}{Z} e^{-\beta E_s}, \tag{3}$$

and all the jumping around will leave this probability distribution invariant, which means that loss and gain balance each other:

$$\begin{aligned}
\Delta P_s(\text{loss}) &= \Delta P_s(\text{gain}) && (4)\\
\sum_{s'} \mathcal{T}(s \to s') \frac{1}{Z} e^{-\beta E_s} &= \sum_{s'} \mathcal{T}(s' \to s) \frac{1}{Z} e^{-\beta E_{s'}} && (5)\\
\sum_{s'} \mathcal{T}(s \to s') &= \sum_{s'} \mathcal{T}(s' \to s) e^{-\beta (E_{s'} - E_s)} && (6)\\
0 &= \sum_{s'} \left[\mathcal{T}(s \to s') - \mathcal{T}(s' \to s) e^{-\beta (E_{s'} - E_s)} \right]. && (7)
\end{aligned}$$

This balancing condition is global: it states that *all* the ways of leaving state s should balance *all* the ways of entering this state. As long as this is true, the Boltzmann distribution will be a stationary distribution of the "walk" defined by $\mathcal{T}(s \to s')$. But this doesn't mean that if we start with an arbitrary state, the system will find its way to this stationary distribution.

It's useful (and very physical, as we can discuss) to impose a much stronger condition, namely that the balancing condition in Eq (7) works for every pair of states separately. That is, we insist that when an ensemble of systems is in the Boltzmann distribution, the transitions from $s \to s'$ are balanced, exactly, by the transitions from $s' \to s$, for every pair of states s and s'. This is called *detailed balance*, and corresponds to the statement that every individual term in Eq (7) be zero on its own, not just the sum:

$$\mathcal{T}(s \to s') = \mathcal{T}(s' \to s) e^{-\beta (E_{s'} - E_s)}. \tag{8}$$

If this condition is met, then certainly the Boltzmann distribution is a stationary distribution. It takes more work to prove that this is the unique stationary distribution (not sure yet how much detail to give here).

It might be useful to pause for a moment and think about what happens when detailed balance is violated. Suppose we have three states, and the transitions are in a cycle, $1 \to 2 \to 3 \to 1$. The distribution over the three states can still come to a steady state, for example if each transition has the same probability per unit time, then if we start with a probability that is uniform over all three states then this won't change in time. But this does not obey detailed balance, because the transition $1 \to 2$ is *not* balanced by the reverse transition $2 \to 1$. Thus, if we take a snapshot of the system we are equally likely to find it in any of the three states, but if we find it in state 1 we can be sure that the next state will be 2 and not 3. This means that if we take a movie, it will look different when we run it backwards ($1 \to 3 \to 2 \to 1$). As we watch a system coming to equilibrium, we can tell which way time is flowing (entropy is increasing). But once we are in equilibrium, detailed balance is the statement that we can't tell the direction of time by watching the fluctuations from state to state.

The condition in Eq (8) is powerful, but still leaves lots of room for choices of the transition probabilities $\mathcal{T}(s \to s')$. Let's try something simple. Assume that we have some natural way of defining neighbors in the space of states. For example, if our system consists of many particles, we can think about neighboring states as being those that involve only a small displacement of

one particle; if our system is a set of spins that can point up or down, then neighboring states are those that differ by only one spin flip; and so on. We'll assume that all transitions with non–zero probability occur between neighboring states. In all the cases just mentioned, every state has the same number of neighbors, which is important (though one can deal with more complicated cases as well).

So we start at one state s, and we consider the possibility of jumping to a neighboring state s'. Should we accept this move? If the move lowers the energy, it is plausible that this will get us closer to the Boltzmann distribution, so let's accept the move if $E_{s'} < E_s$. But what if the move is "uphill" in energy? Then we should accept the move with a probability that matches the Boltzmann distribution, $P_{\text{accept}} = \exp[-\beta(E_{s'} - E_s)]$. If the move is rejected, then we just stay in the same state and try again. So the overall probability of a transition $\mathcal{T}(s \to s')$ is given by

$$\mathcal{T}(s \to s') = \frac{1}{\mathcal{N}_s} \min\left[1,\, e^{-\beta(E_{s'} - E_s)}\right], \tag{9}$$

where $\mathcal{N}_s$ is the number of neighbors of state s, and

$$\mathcal{T}(s \to s) = 1 - \sum_{s' \neq s} \mathcal{T}(s \to s'). \tag{10}$$

Suppose that $E_s > E_{s'}$. Then

$$\mathcal{T}(s \to s') = \frac{1}{\mathcal{N}_s}, \tag{11}$$

whereas

$$\mathcal{T}(s' \to s) = \frac{1}{\mathcal{N}_{s'}} e^{-\beta(E_s - E_{s'})}, \tag{12}$$

so that

$$\mathcal{T}(s \to s') = \mathcal{T}(s' \to s) \frac{\mathcal{N}_s}{\mathcal{N}_{s'}} e^{-\beta(E_s - E_{s'})} = \mathcal{T}(s' \to s) e^{-\beta(E_s - E_{s'})}, \tag{13}$$

where in the last step we make use of $\mathcal{N}_s = \mathcal{N}_{s'}$, which means that every state has the same number of neighbors, as noted above. Thus, this walk through the states satisfies the detailed balance condition in Eq (8).

This general strategy for simulation is called "Monte Carlo," with reference to the famous casino. Recall that the understanding of probability started with thinking about gambling. The particular algorithm described here is called "Metropolis," not after cities in general but after Nicholas Metropolis.[3]

Let's see how this works in an incredibly simple example. Suppose the states are labelled by an integer s and we assign each state a "position" $x_s = s/L$ and an energy $E = \kappa x_s^2/2$. We'll take a walk with `Nsteps` steps, and neighboring states are just neighboring integers. Here is MATLAB code for a Monte Carlo simulation in which we keep track of both the state and the energy at every step:

[3] N Metropolis, AW Rosenbluth, MN Rosenbluth, AH Teller, and E Teller, Equation of state calculations by fast computing machines. *J Chem Phys* **21,** 1087–1092 (1953). Historical note: it still is rare to find a physics paper with an author list including two couples.

```
L = 5;
kappa = 1;
beta = 0.5;
Nsteps = 1e6;
EE = zeros(Nsteps,1);
S = zeros(Nsteps,1);
for t=2:Nsteps;
    sold = S(t-1);
    Eold = (kappa/2)*((sold/L)^2);
    snew = sold;
    step = sign(randn(1));
    snew = sold + step;
    Enew = (kappa/2)*((snew/L)^2);
    if exp(-beta*(Enew - Eold)) > rand(1);
        S(t) = snew;
        EE(t) = Enew;
    else
        S(t) = sold;
        EE(t) = Eold;
    end
end
```

Some results are shown in Fig 1.

Admittedly this example isn't very interesting, but we can check, for instance, that the mean energy is one, as it should be, and we can check the fluctuations in energy, which are related to the heat capacity ... .

This is all slightly more interesting in a model of classical spins in a magnetic field. We imagine that each spin can point either up or down; this is a vestige of quantization, but we won't consider coherent superpositions. Spin up is $\sigma = +1$, and spin down is $\sigma = -1$. Let there be a magnetic field h, so that the energy is

$$E(\{\sigma_i\}) = -h \sum_{i=1}^{N} \sigma_i \tag{14}$$

for a collection of N spins. In this simple case we'll assume that each spin feels the same magnetic field, and that the spins don't interact with one another (but that's coming soon). We want to do a Monte Carlo simulation of this model, following the same strategy as before. The individual steps of the walk now are taken by flipping individual spins, and we'll choose which spin we flip at random; the alternative would be to sweep systematically through the system, flipping spins in order. We'll keep a record of every state along the way, which usually is not done, but it's easy here since the system is not too large. There are other inefficiencies as well in this code, but hopefully being a little less efficient makes it easier to see what we are doing:

```
N = 100;
Nsteps = 1e4;
s = zeros(Nsteps,N);
s(1,:)  = sign(randn(1,N));
h = 1;
beta = 1;
```

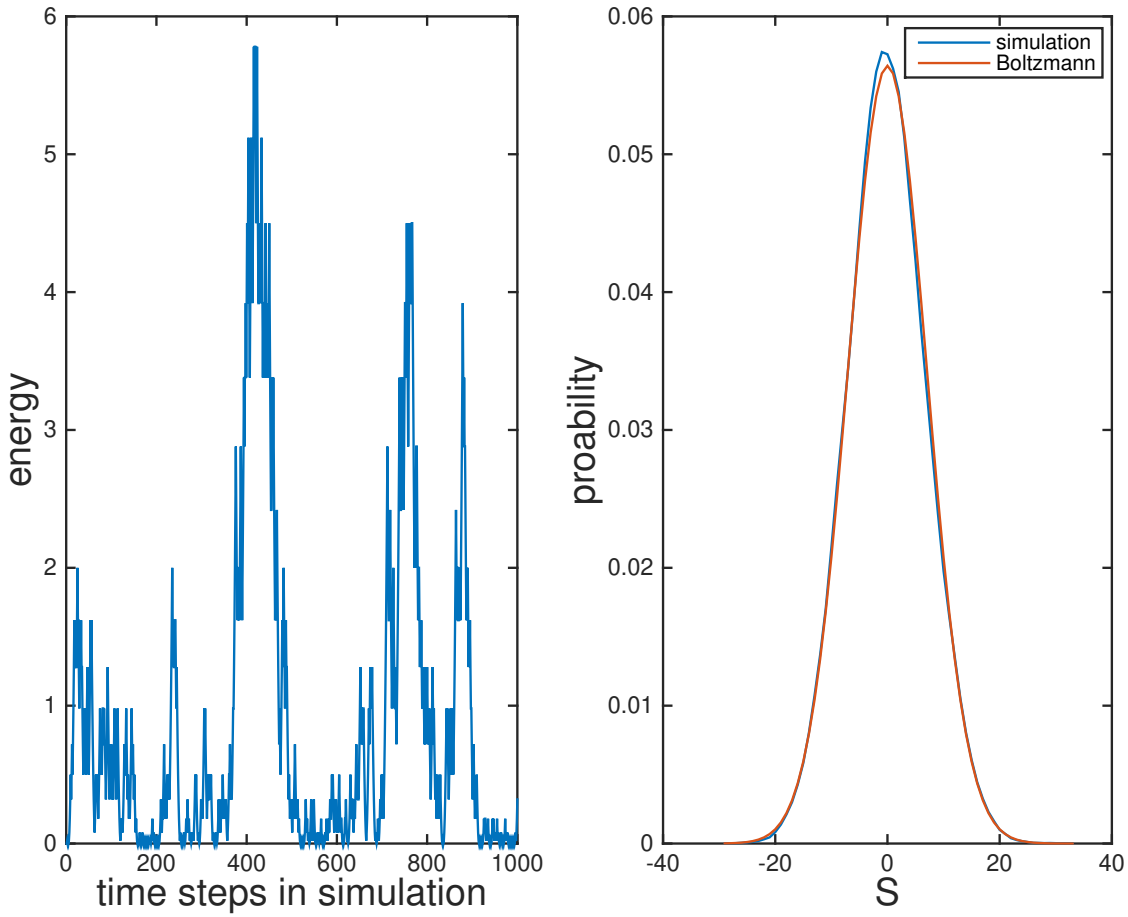

Figure 1: Results of the simplest Monte Carlo simulation, as described in the text. At left, the energy vs time step in the algorithm, showing how the system hovers near the lowest energy but makes occasional excursions to much higher energies. At right, the probability distribution of the variable `S`, compared with the analytic prediction from the Boltzmann distribution.

```
EE = zeros(Nsteps,1);
for t=2:Nsteps
    sold = s(t-1,:);
    Eold = -h*sum(sold);
    snew = sold;
    flip = randperm(N);
    snew(flip(1)) = -sold(flip(1));
    Enew = -h*sum(snew);
    if exp(-beta*(Enew - Eold)) > rand(1);
        s(t,:)  = snew;
        EE(t) = Enew;
    else
        s(t,:)  = sold;
        EE(t) = Eold;
    end
end
```

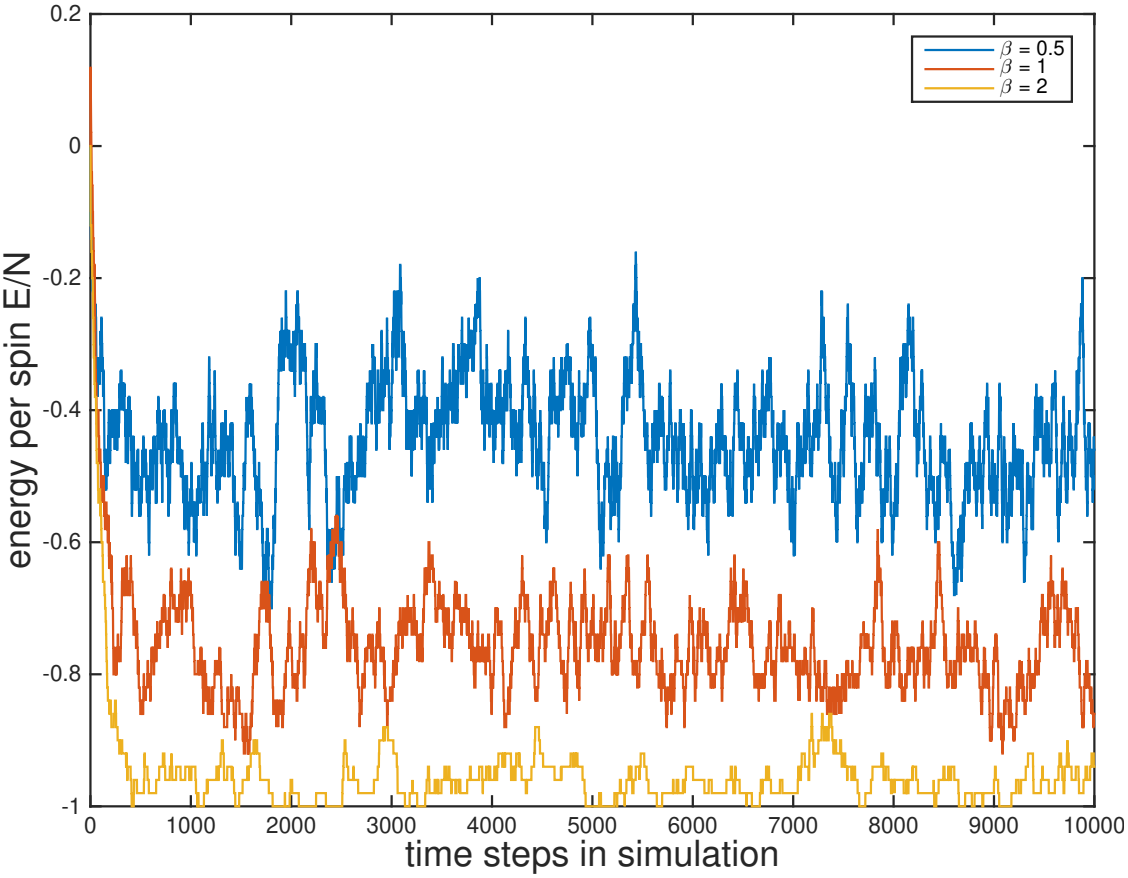

Figure 2: Monte Carlo simulation of $N = 100$ independent spins, with the Hamiltonian from Eq (14) and $h = 1$. Results at three temperatures, as indicated by the inset.

Figure 2 shows results of these simulations at three different values of the temperature, $T = 1/\beta = 0.5,\ 1,\ 2$. Notice that since we start in some random configuration, it takes some "time" before the energy settles down to fluctuating around its mean value. As we lower the temperature, this mean value decreases. At zero temperature the system should get trapped in its lowest energy state, in which all spins are aligned with the magnetic field and hence the energy per spin is $-h = -1$; we see signs of this already at $T = 0.5$. We also see that the variance of the energy goes down as the temperature goes down, and this should mean that the heat capacity also decreases at low temperatures.

The model we just discussed involved "spins" that respond to the magnetic field but do not interact with one another. This is in line with the discussion of ideal gases, both in the classical case and in the quantum (Bose and Fermi) cases. But real particles *interact*. We'd like to capture this, and use the Monte Carlo method as a way of exploring what happens in the presence of such interactions.

Let's imagine a collection of spins $\sigma_{\rm i} = \pm 1$ as before, but they can interact with their neighbors. In the simplest case we would put these spins in a chain, or one–dimensional lattice. If spins interact only with their neighbors, the energy has the form

$$E\left(\{\sigma_{\rm i}\}\right) = -J\sum_{\rm i}\sigma_{\rm i}\sigma_{\rm i+1} - h\sum_{\rm i}\sigma_{\rm i}, \tag{15}$$

where with $J > 0$ the spins prefer to be aligned to their neighbors. We've assumed that all the interactions have the same strength, so there is only one parameter J. The one–dimensional Ising model, as this is called, can be solved exactly, but alas it doesn't do anything very interesting. The first interesting case is in two dimensions, so let's attack this using Monte Carlo.

By analogy with the 1D case, the energy involves only terms where spins interact with their neighbors. But now the positions of spins are labelled by two indices (along the x and y axes), and "neighbor" includes neighbors along both axes. Thus

$$E\left(\{\sigma_{\mathrm{i}}\}\right) = -J \sum_{\mathrm{i,j}} \left(\sigma_{\mathrm{i,j}}\sigma_{\mathrm{i+1,j}} + \sigma_{\mathrm{i,j}}\sigma_{\mathrm{i,j+1}}\right) - h \sum_{\mathrm{i}} \sigma_{\mathrm{i}}, \tag{16}$$

where again we assume that all the interactions are the same. It would be useful to label all the spins with just one index, which we could call nn, and then write

$$E\left(\{\sigma_{\mathrm{i}}\}\right) = -\sum_{\mathrm{nn,mm}} JJ(\mathrm{nn},\mathrm{mm})\sigma_{\mathrm{nn}}\sigma_{\mathrm{mm}}, \tag{17}$$

where for now we'll leave out the magnetic field. To set up this indexing system in a 10×10 two–dimensional lattice, we can do this:

```
L = 10;
N = L^2;
for i=1:L;
    for j=1:L;
        nn(i,j) = (i-1)*L + (j-1) + 1;
    end
end;
```

and then we fill in the matrix JJ assuming periodic boundary conditions:

```
JJ = zeros(N);
for i=1:L;
    for j=1:L;
        if j~=L;
            JJ(nn(i,j),nn(i,j+1)) = 1;
        end
        JJ(nn(i,L),nn(i,1)) = 1;
        if i~=L;
            JJ(nn(i,j),nn(i+1,j)) = 1;
        end
        JJ(nn(L,j),nn(1,j)) = 1;
    end
end
JJ = JJ + JJ';
```

Implicitly this sets $J = 1$, but since all that ever matters is the combination J/T this is not a problem. So now we have the ingredients to compute the energy of any state, which is all we need in order to do Monte Carlo.

The Monte Carlo simulation itself can be done as

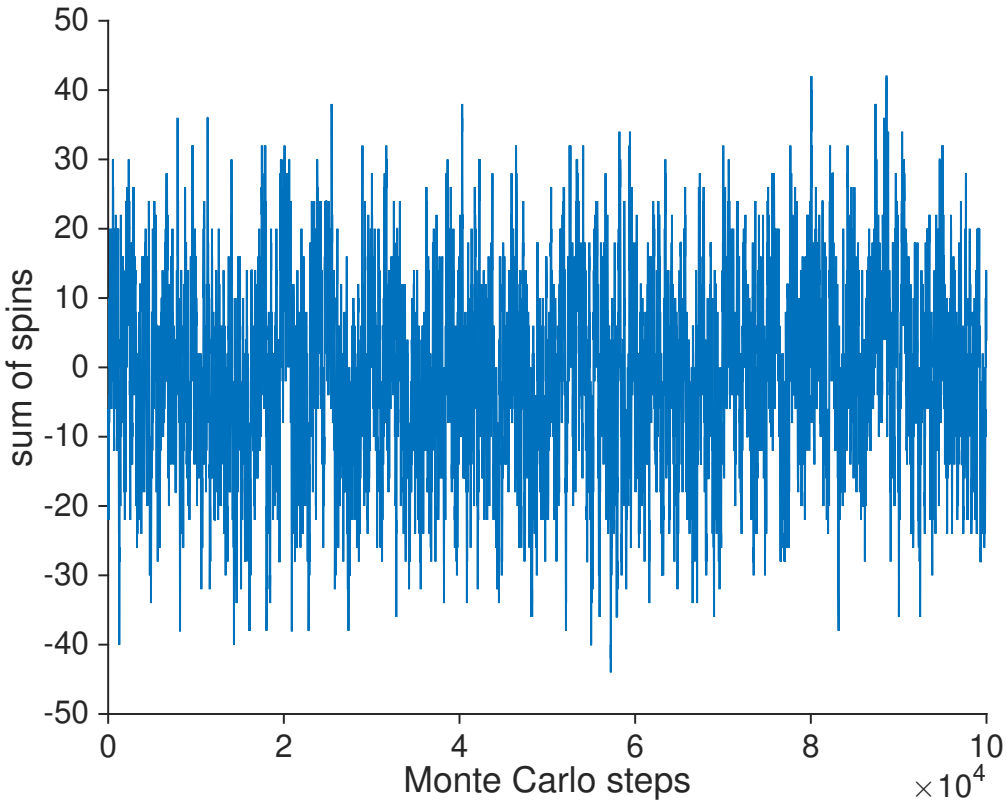

Figure 3: Results of a Monte Carlo simulation of a 10×10 2D Ising model, at $\beta = 0.1$.

```
Nsteps = 1e5;
S = zeros(N,Nsteps);
beta = 0.1;
EE = zeros(Nsteps,1);
S(:,1) = sign(randn(N,1));
for t=2:Nsteps
    sold = S(:,t-1);
    Eold = -(1/2)*sold'*JJ*sold;
    snew = sold;
    flip = randperm(N);
    snew(flip(1)) = -sold(flip(1));
    Enew = -(1/2)*snew'*JJ*snew;
    if exp(-beta*(Enew - Eold)) > rand(1);
        S(:,t) = snew;
        EE(t) = Enew;
    else
        S(:,t) = sold;
        EE(t) = Eold;
    end
end
```

where in this case we will actually keep a record of all the states we encounter along the way. In many cases we won't want to do this, because it takes too much space and because successive

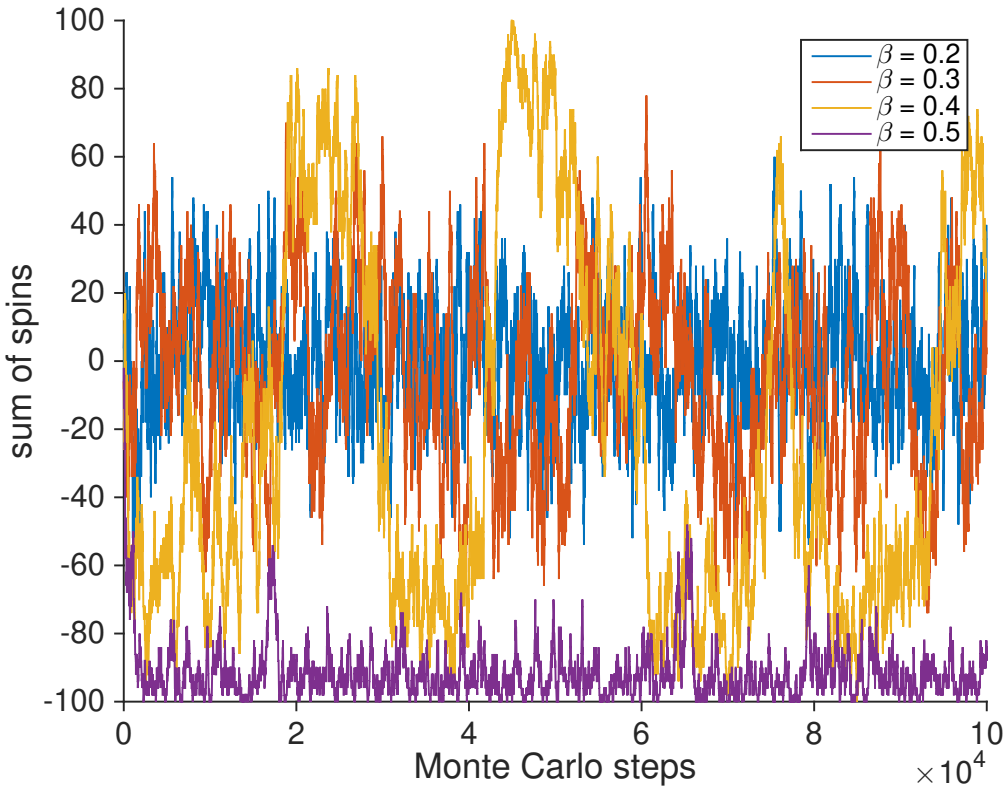

Figure 4: Results of a Monte Carlo simulation of a 10×10 2D Ising model, at different values of β.

states are correlated with one another, so we don't really learn much new from keeping all of them. Nevertheless, this is the simplest approach, and works well on small problems.

In the code written above, the value of $\beta = 0.1$, which is small. If we run the simulation, and keep track of the sum of all the spins, we get the results shown in Fig 3. This is not yet very interesting: we see the sum of all spins fluctuating around zero, and the typical size of these fluctuations is not so different from $\sqrt{N}$, which we would get if all the spins just independently decided to point up or down. Things get more interesting as we lower the temperature, in Fig 4.

We see that, as β increases, the fluctuations in the total magnetization (sum of spins) become larger, eventually so large that they start to hit the limits at $\pm N$ where all the spins are lined up. If we lower the temperature still further, the system quickly relaxes to be very close to one of these saturated states (in this case $\sum_i \sigma_i \to -N$), and stays there. If we do many such runs, we find that the system chooses the "all up" or "all down" state at random. This is starting to look interesting.

Figure 4 is hinting that there is some critical value of the temperature, below which the system spontaneously develops a nonzero average value for the magnetization, while above this critical temperature all we see is fluctuations around zero. We could show this more convincingly with more detailed Monte Carlo simulations, not just running for more steps at more different values of the temperature, but also looking at larger systems, where the "critical" behavior we are discussing becomes sharper.

3 Ising models and mean–field theory more generally

It is possible to solve, exactly, the two–dimensional Ising model that we have been discussing, but this is a very difficult calculation. To get a sense for what is happening, we can try a much simpler problem. Suppose that, instead of interacting with nearest neighbors on a square lattice, each spin interacts with the average of all the other spins in the system. Then

$$E = -\frac{\tilde{J}}{2}\sum_{\mathrm{i}=1}^{N}\sigma_{\mathrm{i}}\left(\frac{1}{N}\sum_{\mathrm{j}=1}^{N}\sigma_{\mathrm{j}}\right) - h\sum_{\mathrm{i}=1}^{N}\sigma_{\mathrm{i}}, \tag{18}$$

and we write $\tilde{J}$ because this isn't quite the same parameter as J above—the models are slightly different, but presumably have related qualitative behaviors. This is called a "mean–field" model.

Now let's try to compute the partition function,

$$Z = \sum_{\{\sigma_{\mathrm{i}}\}}\exp\left[\beta\frac{\tilde{J}}{2}\sum_{\mathrm{i}=1}^{N}\sigma_{\mathrm{i}}\left(\frac{1}{N}\sum_{\mathrm{j}=1}^{N}\sigma_{\mathrm{j}}\right) + \beta h\sum_{\mathrm{i}=1}^{N}\sigma_{\mathrm{i}}\right]. \tag{19}$$

The key trick is to remember that

$$\int\frac{d\phi}{\sqrt{2\pi A}}e^{-\phi^2/(2A)+\phi X} = e^{+AX^2/2}. \tag{20}$$

Using this backwards, if we see the square of something in the exponential, then we can "uncomplete the square" at the cost of introducing a new variable ϕ. But

$$\sum_{\mathrm{i}=1}^{N}\sigma_{\mathrm{i}}\left(\frac{1}{N}\sum_{\mathrm{j}=1}^{N}\sigma_{\mathrm{j}}\right) = \frac{1}{N}\left(\sum_{\mathrm{i}=1}^{N}\sigma_{\mathrm{i}}\right)^2. \tag{21}$$

Thus we have

$$\begin{aligned}
Z &= \sum_{\{\sigma_{\mathrm{i}}\}}\exp\left[\beta\frac{\tilde{J}}{2N}\left(\sum_{\mathrm{i}=1}^{N}\sigma_{\mathrm{i}}\right)^2 + \beta h\sum_{\mathrm{i}=1}^{N}\sigma_{\mathrm{i}}\right] && (22)\\
&= \int\frac{d\phi}{\sqrt{2\pi\beta\tilde{J}/N}}\sum_{\{\sigma_{\mathrm{i}}\}}\exp\left[-\frac{N}{2\beta\tilde{J}}\phi^2 + \phi\sum_{\mathrm{i}=1}^{N}\sigma_{\mathrm{i}} + \beta h\sum_{\mathrm{i}=1}^{N}\sigma_{\mathrm{i}}\right] && (23)\\
&= \int\frac{d\phi}{\sqrt{2\pi\beta\tilde{J}/N}}\sum_{\{\sigma_{\mathrm{i}}\}}\exp\left[-\frac{N}{2\beta\tilde{J}}\phi^2 + (\phi+\beta h)\sum_{\mathrm{i}=1}^{N}\sigma_{\mathrm{i}}\right]. && (24)
\end{aligned}$$

Now it looks as if each spin sees a magnetic field $h + \phi/\beta$, but the spins don't interact with each other directly. That means we can do the sum over states:

$$\begin{aligned}
\sum_{\{\sigma_{\mathrm{i}}\}}\exp\left[-\frac{N}{2\beta\tilde{J}}\phi^2 + (\phi+\beta h)\sum_{\mathrm{i}=1}^{N}\sigma_{\mathrm{i}}\right] &= e^{-N\phi^2/2\beta\tilde{J}}\sum_{\{\sigma_{\mathrm{i}}\}}e^{(\phi+\beta h)\sum_{\mathrm{i}=1}^{N}\sigma_{\mathrm{i}}} && (25)\\
&= e^{-N\phi^2/2\beta\tilde{J}}\left(\sum_{\sigma_{\mathrm{i}}=\pm1}e^{(\phi+\beta h)\sigma_{\mathrm{i}}}\right)^N && (26)\\
&= e^{-N\phi^2/2\beta\tilde{J}}\left(2\cosh(\phi+\beta h)\right)^N. && (27)
\end{aligned}$$

Thus we have

$$Z = 2^N \sqrt{\frac{N}{2\pi\beta\tilde{J}}} \int d\phi \, e^{-Nf(\phi,h)}, \tag{28}$$

where

$$f(\phi, h) = \frac{1}{2\beta\tilde{J}}\phi^2 - \ln\cosh(\phi + \beta h). \tag{29}$$

Now recall that if we are doing an integral of the form

$$I = \int dx \, e^{-Ng(x)}, \tag{30}$$

then as N becomes large the integral is dominated by values of x near the minimum of the function $g(x)$. More precisely, if x_* is the location of the minimum, then

$$I = \int dx \, e^{-Ng(x_*) - (N/2)g''(x_*)(x-x_*)^2 + \cdots} \tag{31}$$

$$= e^{-Ng(x_*)} \sqrt{\frac{2\pi}{Ng''(x_*)}}. \tag{32}$$

As usual, we find the minimum by solving the equation

$$\left.\frac{dg(x)}{dx}\right|_{x=x_*} = 0. \tag{33}$$

Let's use this to evaluate Z from Eqs (28, 29).

To find the dominant value of ϕ, we have to solve the equation

$$0 = \frac{\partial f(\phi, h)}{\partial \phi} = \frac{\phi}{\beta\tilde{J}} - \tanh(\phi + \beta h). \tag{34}$$

In particular, if $h = 0$ we have

$$\phi = \beta\tilde{J}\tanh(\phi). \tag{35}$$

We see that if the temperature is high—$\beta\tilde{J}$ is small—then this equation has only one solution, $\phi = 0$. But at low temperatures—$\beta\tilde{J}$ is large—there are three solutions, $\phi = \pm\phi_0$ and $\phi = 0$. In this regime, $\phi = 0$ is a maximum of $f(\phi, h)$, and so we are interested only in the other two solutions. These nonzero solutions emerge only for $\beta\tilde{J} \leq 1$, so this is a "critical point" between two very different behaviors.

To see what's going on it's useful to plot $f(\phi, h = 0)$, as shown in Fig 5. We see that for high temperatures ($\beta\tilde{J} < 1$) the free energy $f(\phi)$ has just a single minimum at $\phi = 0$. But at $\beta\tilde{J} = 1$ this minimum flattens, and when we lower the temperature further we start to see two degenerate minima at $\pm\phi_*(\beta\tilde{J})$. If we look back at Eq (24), we see that ϕ contributes to the effective magnetic field acting on each spin. So if the external field is zero, and $\phi = 0$, we can be sure that the spins are equally likely to point up or down, and the average magnetization will be zero. But when we cool the system below the critical point at $\beta\tilde{J} = 1$, there is a nonzero ϕ which dominates the integral, and hence there will be a nonzero magnetization. Crucially, this is what we saw in the

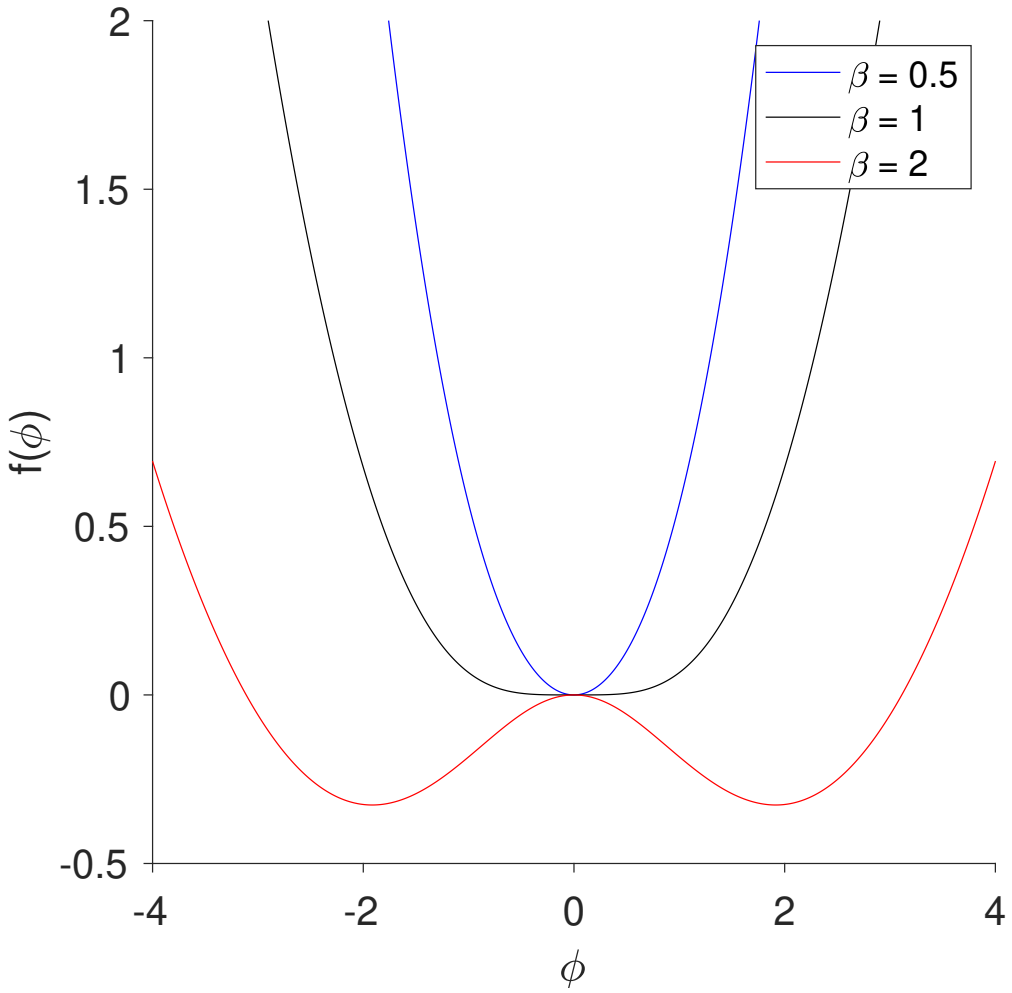

Figure 5: Free energy as a function of ϕ, from Eq (29) with $h = 0$. We choose units of energy where $\tilde{J} = 1$.

simulations of the 2D Ising model: cool below a critical temperature, and we the system develops a nonzero mean magnetization, which could be positive or negative.

We can sketch the magnetization vs temperature, as shown in Fig 6, remembering that below the critical temperature we have solutions with both positive and negative magnetization. There are many details to explore here, such as the precise behavior in the neighborhood of the critical point.

But let us try a different direction. We have interpreted the Ising model as being a model for magnets, imagining that the variables σ_i are, in effect, microscopic magnetic moments of single atoms or perhaps even single electrons. It's interesting that we don't have much microscopic detail in this model. It would be difficult to tell exactly what magnetic material we are describing, since in a real sense the only parameter is J, and since we only see the combination βJ this really just serves to set the scale of temperature. So there is hint here of universality. In particular, if we could plot magnetization vs temperature for many magnets, and scale the axes correctly, this model predicts that we should see the same behavior in all the magnets, or at least in some large class of magnets to which this model applies. This is true, experimentally.

Suppose that instead of magnets, we try to describe a liquid. We can imagine dividing space

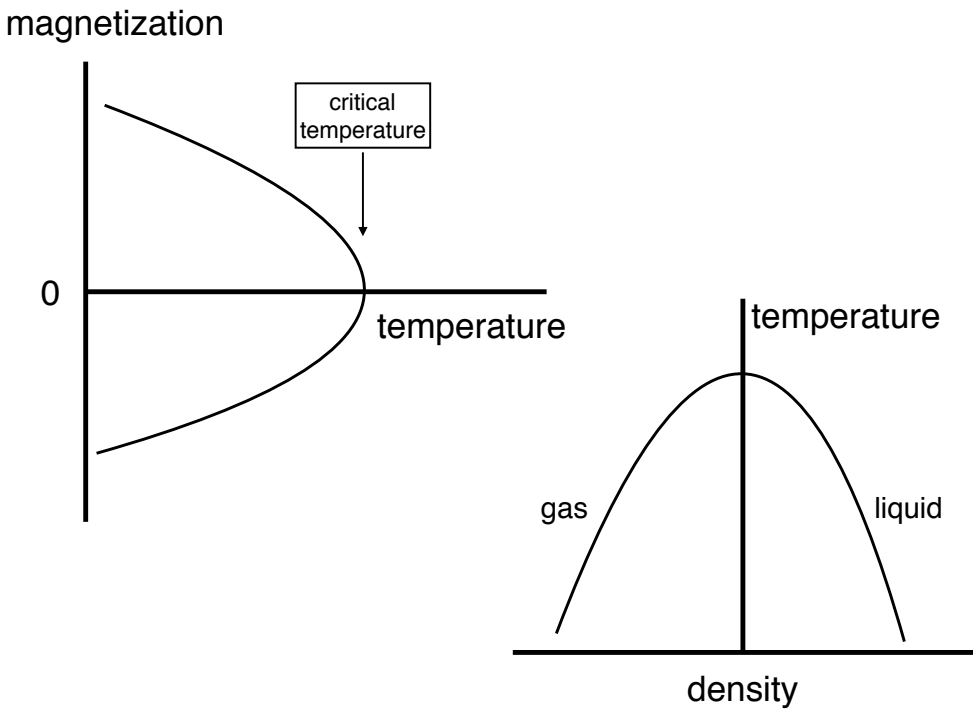

Figure 6: Upper left, magnetization vs temperature in the mean–field ferromagnet. Lower right, the same graph turned on its side, and interpreted for the lattice gas as temperature vs density.

into small boxes, and in each box we have the choice to put a molecule or to leave it empty. Presumably there is some energy difference between these two possibilities. In addition, if we put a molecule in the box, it will interact with the neighboring molecules, so that putting two molecules next to one another has an energy different from putting them far apart. But notice that we are describing a system of binary variables—box with one molecule vs box empty—and interactions among pairs of neighbors. This is again the Ising model, though now called a "lattice gas."

In a lattice gas, the the average value of a spin tells us the local density of the gas or liquid. If the average is zero, then we have some typical density (the boxes are equally likely to be full or empty), and when the average moves away from zero this density rises or falls. Instead of magnetization vs temperature, we could plot density vs temperature, or the other way around, as shown in Fig 6. What does it mean that we can have two values of density at a given temperature?

Think about what happens as you heat a pot of water on the stove. Before it boils, there is a layer of vapor above the liquid in the pot. So, at this temperature and pressure, you can have water in both a liquid state and a vapor or gaseous state. These differ in density, so at one temperature you have two densities, just as in the plot. Notice that in order to make this work, the pressure has to be in the right range.

Now as we change the temperature, we will find that the density difference between gas and liquid changes also. If the pressure is right, there is a critical temperature at which the difference vanishes, just as for the magnetization of a magnet in zero magnetic field. But if our discussion makes sense, then again there are no microscopic details to identify the material. So if we choose the right units, we should get the same curve of temperature vs density for all liquids. This is true,

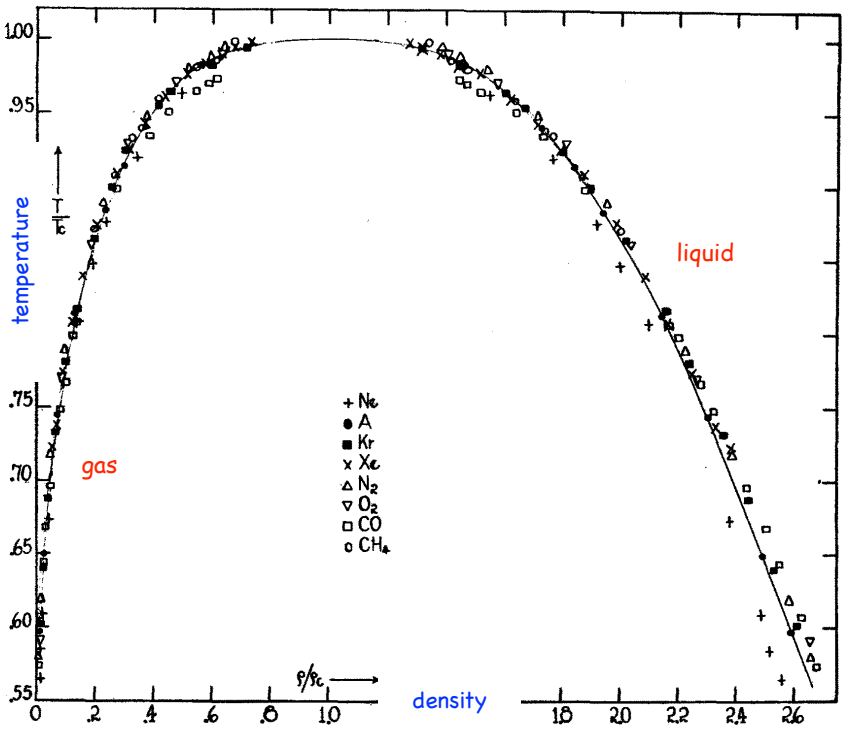

Figure 7: Temperature vs density for a variety of liquid/gas materials, from EA Guggenheim, *J Chem Phys* **13,** 253–261 (1945).

as shown in Fig 7, and is one aspect of the "law of corresponding states." You see that this has been known for quite some time, and really is quite a beautiful fact about the world.

Suggested exercises

Both of these problems are a bit open–ended, and phrased a little less crisply than for a regular course. I hope this is OK, but please let me know if you need clarification.

Numerics. Especially if you have never done a Monte Carlo simulation, now is the time. Try working on the 2D Ising model from the notes above. Do everything more carefully, and explore. Try to plot, for example, the mean magnetization $m = (1/N)\sum_i \sigma_i$ as a function of β. At large β you'll need to think carefully about how to separate the $m \sim \pm 1$. Maybe a solution is to plot

$$m^2 = \left\langle \left(\frac{1}{N} \sum_i \sigma_i \right)^2 \right\rangle, \tag{36}$$

where the average is over a long run. How do you convince yourself that your run is long enough so that you are really sampling the probability distribution? Can you push to larger N?

Analytics. Do a careful job with the mean–field calculation.

a. Start by understanding what would happen if we hadn't inserted the strategic factor of $1/N$ in Eq (18). Can you explain this intuitively?

b. Step back to the general case where states of the system are labelled by s, and each has energy E_{s}. The Boltzmann distribution says that the probability of the system being in state s is

$$P_{\mathrm{s}} = \frac{1}{Z} e^{-\beta E_{\mathrm{s}}} \tag{37}$$

$$Z \equiv \sum_{\mathrm{s}} e^{-\beta E_{\mathrm{s}}}. \tag{38}$$

Show that, in general, you can compute the mean energy from the partition function,

$$\langle E \rangle \equiv \sum_{\mathrm{s}} P_s E_s = -\frac{\partial \ln Z}{\partial \beta}. \tag{39}$$

c. Use the result in [b] to find the mean energy in the mean–field model (stick with $h = 0$ for simplicity). What happens at the critical point $\beta \tilde{J} = 1$?

d. We saw that at $N \to \infty$ the calculation is dominated by a particular value of ϕ. How does this connect with something you observe, such as the mean magnetization?

e. Show that the mean magnetization of the system is also the mean of each spin. Then calculate the correlations between the spins,

$$C_{\mathrm{ij}} = \langle \sigma_{\mathrm{i}} \sigma_{\mathrm{j}} \rangle - \langle \sigma_{\mathrm{i}} \rangle \langle \sigma_{\mathrm{j}} \rangle, \tag{40}$$

to leading order in $1/N$. In particular, show that something special happens to these correlations at the critical temperature, and connect with the hints from Fig 4. This is much harder than the rest of the problem.