Lesson 9 : The Ising model

Notes from Prof. Susskind video lectures publicly available on YouTube

Introduction

We began to study last time the Ising model. It is actually a class of models sharing certain features. They began as the following problem given by the physicist Wilhelm Lenz to his doctoral student Ernst Ising in the early 1920's at the University of Hamburg : In a model of magnet made of a linear sequence of little magnetic moments where only neighbors are energetically coupled is there a phase transition in the magnetism of the system ?

Ising solved the one-dimensional problem correctly, and showed that it didn't have a phase transition. On the basis of that he believed and wrote in his thesis that the Ising model in any dimension does not have a phase transition. But that was incorrect.

We are going to solve the Ising model in one dimension. We are not going to solve the higher dimension Ising models. They are hard maths and not within the scope of this course. But we will use an approximation method that is physically very intuitive. And we will see that in higher dimensions, when their number is sufficiently high, there is a phase transition. As we already said, there is a phase transition, when we lower the temperature, as soon as the number of dimensions is two, but that we won't show.

Generally speaking, a phase transition means a sudden change in the properties of the system as you vary a parameter, for example the temperature. But let's go slow, taking steps. So let us first go back to the simple example – even simpler than an Ising model – that we studied last time.

We studied a problem in which there was a collection of

little magnets independent of each other, i.e. with no coupling, plunged in a magnetic field H. Each magnet could be up or down. And there was an elementary energy function E for each little magnet. The energy function differed between up and down. Its expression was

$$E = \mu H \sigma \tag{1}$$

where μ was the magnetic moment of the little magnet, H was the magnitude of the external magnetic field, and σ was $+1$ or -1 depending on whether the little magnet pointed up or down.

We shall change a little bit the notation for this energy function. First of all, since μ and H appear only through their product, we denote it j. It is a standard notation for what we will be doing later. And by convention we will put a minus sign in front. Thus the elementary energy function becomes

$$E = -j\sigma \tag{2}$$

The minus sign doesn't make any difference to the physics, because it really is just a redefinition of what we mean as up and down. If the original energy favored down, then, if we change the sign, it favors up. But, other than that interchange of up and down, there really is no difference.

Physical systems always tend to favor lower energy. In equation (2), when σ is negative, E is positive. And when σ is positive, E is negative, that is lower. So the system favors σ positive, or the little magnet pointing up.

We studied a whole group of such little magnets, figure 1. And we did some combinatorics. How many states are there, given that there are n little magnets up, and m down ? It is the customary number $\binom{N}{n}$, etc.

Figure 1 : System made of a collection of elementary magnetic moments. The elements are uncoupled among each other. The whole system is plunged in a magnetic field.

We did it that way to see carefully all the details. But we could have done something much simpler.

We have already learned that we can always think of a system as being a small system plus a heat bath.

Figure 2 : System viewed as a small system plus a heat bath. The highlighted magnetic moment is the small system. The rest is the heat bath.

So we focus on anyone of the elementary magnetic moments, for instance the one highlighted in figure 2. We assume that it is in thermal equilibrium with its environment – its environment being the rest of the magnets. Those ma-

gnets of the environment are the heat bath. And they bring the highlighted little magnet to equilibrium.

Now that the system of interest has become only one spin which we focus on, we can write the partition function for it. It is very simple. From equation (2) we get

$$Z = \sum_i \ e^{+\beta j \sigma_i} \tag{3}$$

The plus sign in the exponent is because there is a minus sign in equation (2), and, as we are familiar with, the general form of the partition function is $Z = \sum e^{-\beta E_i}$.

In equation (3), the sum is over all the possible configurations for the little magnet. There are only two : σ equals $+1$ or -1. So equation (3) simplifies into

$$Z = e^{\beta j} + e^{-\beta j} \tag{4}$$

That is the partition function for just one spin.

When we calculated it, thinking of the whole system as one system, do you remember the answer we got ? We got equation (11) of chapter 8, which was

$$Z = \left(\ e^{\beta j} + e^{-\beta j} \ \right)^N \tag{5}$$

It is also the right-hand side of equation (4) to the N-th power. This makes sense because, whenever we have *independent systems*, the partition function of the big system they form is simply the product of the partition functions

of each system. And indeed, in the simplest model we studied in the last chapter, eauch spin is independent of the other spins. There is no coupling between spins. So we got a factor $e^{\beta j} + e^{-\beta j}$ for each spin. And the partition function for the whole collection of spins is that factor to the N-th power.

In other words, we can really short-circuit all that stuff about combinatorics and focus on one spin at a time.

So let's concentrate on the partition function of one spin, that is equation (3). It can be rewritten

$$Z = 2\cosh(\beta j) \tag{6}$$

When we did it last time, we got $2\cosh(\beta j)$ raised to the power the number of spins.

When we take the logarithm of that – which is the interesting thing – it simply becomes the sum.

For example the energy of the collection of spins, which is just minus the derivative of the logarithm of their partition function, is nothing but the sum of the energies of the individual spins.

Here we are concentrating on one spin at a time, one small magnet at a time, and we calculate exactly the same energy for each of them. So let's calculate the energy :

$$E = -\frac{1}{Z}\frac{\partial Z}{\partial \beta} \tag{7}$$

which is the same as minus the derivative of the logarithm.

The derivative of Z is twice the derivative of the cosh. That is 2 sinh. But then we have to multiply by the derivative of the argument βj with respect to β, which gives another j. So we get

$$\begin{aligned} E &= -\frac{1}{Z}\ 2\ \sinh(\beta j)\ j \\ &= -j\ \tanh(\beta j) \end{aligned} \tag{8}$$

That is the average energy of one of these little spins in figure 2.

We can also ask : What is the probability that the spin is up versus down ? Or what is the average ? σ is the variable that can take the value $+1$ or -1. But what is the average σ ?

Well, from equation (2), the energy in each configuration of the spin is $-j\sigma$. Equation (8) is the average energy. Then it is pretty clear that the average σ is the right-hand side of equation (8) without the $-j$. When we write E, unless specified otherwise [1], we talk about the average energy. For the average σ, for maximum clarity, we can go back to our heavier notation $< \sigma >$. So let's summarize this

$$\begin{aligned} E \quad &= -j\ \tanh(\beta j) \\ < \sigma > &= \tanh(\beta j) \end{aligned} \tag{9}$$

We are going to need this. But let's just draw a picture of it. Figure 3 is a picture of the function hyperbolic tangent.

1. as in equation (2).

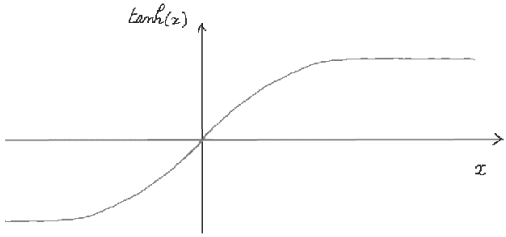

Figure 3 : Function hyperbolic tangent.

This is hyperbolic tangent of x. In the expression of the average of σ, x is βj. j is just a number. We could have taken it to be equal to 1. So $< \sigma >$ is hyperbolic tangent of something proportional to the inverse temperature.

The slope of $\tanh(x)$ at $x = 0$ is 1. Then it has the two asymptotic values -1 and $+1$, when x respectively goes to $-\infty$ and $+\infty$. But, if we took j to be positive, which is entirely conventional and up to us, we are only concerned with the positive branch of its graph.

Now let's look at what that says physically. When β *is very large*, that means *very low temperature*, and since j is positive, then βj is very large and positive. We are way out on the right in figure 3.

What it is telling us is that at very low temperature the spin will point up with probability very close to 1. Indeed, if the two possible values of σ are -1 and $+1$ and its average is close to $+1$, this implies that the probability that it take the value $+1$ must be almost 100%. And we built the model, with equation (2), in such a way that this corres-

ponds to the low energy state of the spin.

Let's turn to very high temperatures. Then βj is very small. We are near the origin. So $< \sigma >= \tanh(\beta j)$ is near zero too, on the positive side. The magnetic spin is being constantly kicked around with equal probabilility between its values $+1$ and -1, and its average value is near 0. At very high temperature it overcomes any bias it could have.

That was a very simple system.

One-dimensional Ising model

Now we want to come to the next least complicated big magnet. It is the one-dimensional Ising model, which we began to talk about at the end of last chapter.

Again, at each point along a one-dimensional array, there is a little magnet with a σ, figure 4. It is either up or down, corresponding to $\sigma = +1$ or -1.

Figure 4 : One-dimensional Ising model.

It looks like the preceding model, and indeed figure 4 is the same figure as figure 1, but the new thing is that the energy

now depends on the relationship between neighboring spins.

And there is no more external magnetic field. We could do it with an external magnetic field too. And we will discuss that at some point. But for the moment no external magnetic field.

Equation (8) was the average energy of one spin. But now the energy of the system comes from the coupling of neighboring spins. So we can no longer simply look at the energy of each individual spins and add them up. Hence we look directly at the energy of the whole shebang. It is given by

$$E \;=\; \sum_i \; -j \; \sigma_i \; \sigma_{i+1} \tag{10}$$

As before, j is only a positive number. And the minus sign means that the model favors neighboring spins pointing in the same direction [2]. In that case, we say that they are *parallel*, or *aligned*.

Thus the energy of one pair is lower when the two spins are parallel, either both up or both down. And it is higher when they are unaligned, i.e. when one is pointing up and

2. Notice that at macroscopic scale this is not what ordinary ferromagnets do. But in ordinary ferromagnets, at microscopic scale, the spins do line up when the temperature and the energy go down. It is a matter of very intricate detail whether the little elementary magnets in the material prefer to be aligned or antialigned. Most of the times they prefer to be antialigned, but in iron they prefer to be aligned. We are not going to go into these details, which are not relevant to this course. Remember that the "big magnets" made of "elementary magnets" we study here are only mathematical models, remotely related to real magnets. We use them as simple examples displaying or not phase transitions.

the other down.

If all the spins are parallel, all the products $\sigma_i\ \sigma_{i+1}$ are equal to $+1$. Then because of the $-j$ in front, the total energy of the system is minimum.

So what is the ground state of the system like ? It is when all the spins point in the same direction. But which direction ? Answer : It doesn't matter. There are two ground states, with the same lowest energy.

Now what would happen if we changed the sign in front of j, that is if we consider another model where the total energy was $\sum\ j\ \sigma_i\ \sigma_{i+1}$?

Is it really a different system ? No. It is really the same system. All you have to do to see it is the same system is to redefine every other spin by changing its sign. On every other spin what you called up you now call down. You redefine the variable. And then what you find out is the ground state wants to have them all antiparallel. And again there are two ground states.

So it is mathematically identical. In other words, the minus sign in front of j is really only a convention, justified because it is simpler to think of grounds states with all the spins aligned, up or down, rather than to think of the two states where they perfectly alternate.

What we want to calculate is the usual thing : our friend the partition function. What is it now ? It is

$$Z\ =\ \sum\ e^{\beta j \sum \sigma_i \sigma_{i+1}} \tag{11}$$

The small sign Σ in the exponent is the sum over i for *one configuration*. And the big sign Σ, is the sum over *all possible configurations*.

Equation (11) looks terrible and very hard. But it is the partition function we must work with.

Let us first present a question that we can ask. We would like to know the answer to the following. Suppose we know that the spin in one of the positions along the chain is up, figure 5.

Figure 5 : We know that the spin at position i is up.

Then, knowing that, we can ask : What is the probability that n links down the chain it is also up ? It is a conditional probability, given that the spin in figure 5 is up, that at some other specified place it is also up ? It can be expressed in terms of correlation [3].

We could also ask it in a different way, but it is the same question really : What is the average of the product of spins

3. The correlation between the r.v. X_i and the r.v. X_j is by definition

$$\text{corr}(X_i,\ X_j) = \frac{< X_i X_j > - < X_i >< X_j >}{\text{sd}(X_i)\ \text{sd}(X_j)}$$

where $< ... >$ stands for mean, and sd(...) for standard deviation.

at two different locations ? [4]

Now you might think – and in this case you'd be right – that a spin being up at some position i would have very little effect on a spin at some position j far down the chain.

You might expect then that the average of the product would be zero. Because whatever σ_i is, σ_j would have equal probability to be up or down, independently of the value of i. So their correlation would be zero.

But it might be wrong too. There might be an effect of just having one spin up at i that would propagate all the way through the system and tell you that there is a net bias throughout the whole sample.

You could diagnose that by looking at the average value of the spin at i times the value of the spin at j, where j is n positions down from position i, that is $j = i + n$. In other words, you would be looking at

$$< \sigma_i \ \sigma_{i+n} > \tag{12}$$

If there existed this kind of memory where, if the spin at location i was up, it biased the system all the way to infinity,

4. Note that, if for all i σ_i is equally probably $+1$ or -1, then all the means are zero, and all the standard deviations are one. In that case the above formula reduces to

$$\text{corr}(\sigma_i, \ \sigma_j) \ = < \sigma_i \sigma_j >$$

paying attention that here σ_i and σ_j are not the usual notation for standard deviations, but are random variables.

then expression (12) would not go to 0 with large distances.

If, on the other hand, that bias does go to 0, then σ_i being up would not bias σ_{i+n}. And on the average the product with a distant spin would be 0.

So this is an interesting diagnostic test of the effectiveness of one spin being up on its neighbors, and how far that propagates down the chain.

Let's see if we can guess what the answer is in our one-dimensional model. Instead of a chain of little magnets, think of it as a game of telephone, the game where you whisper to your neighbor.

So we have a long chain of people and a message is going to start at one end somewhere. It could be anywhere, for instance at location i. And the message is going to be a very simple message. The message is either "0", or "1".

We are not going to make fancy messages like the kids who play this game – the kind of message like : "my dog had a heart attack and therefore I'm off dog food for the rest of my life". Just "0" or "1".

And people hear pretty well, as well as talk pretty well, among this group. But the fidelity is not absolutely perfect. It is pretty good, yet there is a small probability that the information be changed.

The question is : How far down the chain does the signal propagate before it gets lost ? In other words, how far do we have to go before it just becomes equally likely that the

person agrees with the starting point or disagrees.

The answer is no matter how good the fidelity, as long as it is not absolutely perfect, you lose memory sufficiently far down. If there is a probability for an error of 1% in each time, then the probability for no error is 99%. But, when we go two steps down the chain, it is 99% of 99%. And that is 98.01%.

Each time it goes down by a factor of 0.99. And if you go far enough down the chain that product will get arbitrarily small. The probability a hundred units down the chain that the signal is remembered with good fidelity is only 36.6%. And down 10 000 units it is 2.3×10^{-42} %, that is essentially zero. That is basically no memory.

Exercise 1 : Given that the signal started "1", show that 10 000 steps down the chain the probability that it is "1" is 50%, and the probability that it is "0" is 50%.

It is the same thing in our model. There is no signal to be transmitted, but there is an energy bias. If the spin σ_i is up, the system prefering lower energy, the spin σ_{i+1} will have a bias to be up too. At zero temperature, it is like the perfectly accurate signal, infinite fidelity. That is like the situation where there is no loss of information whatsoever. If one spin is up, then everybody just lines up. It is perfect positive correlation.

But if there is the slightest systematic bias, or infidelity, the correlation far down will go to zero. That is what will

happen in the Ising model.

So let's see if we can calculate E in equation (10). There is a marvellous trick. Instead of focussing on the spins, it is to focus on the links between the spins.

We are going to imagine for the moment that the chain is finite. Then we will let it get very big. The first spin at the left extremity can be up or down. Let's start by assuming it is up. And later on we will add back the configuration where it is down. Thus we are going to write the partition function as the sum of two terms. In the first term, the first spin is up. In the second term, the first spin is down. But let's concentrate on the term where the first spin σ_1 is up to begin with, figure 6.

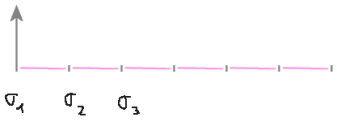

Figure 6 : Collection of spins, and their links. We know that the first spin is up.

Now when considering the second spin, we can either work with the value of the spin itself or we can work with the product $\sigma_1\sigma_2$. Since σ_1 is known to be up, $\sigma_1\sigma_2$ will tell us everything we want to know about σ_2. For any i, let's define

$$\mu_i = \sigma_i \ \sigma_{i+1} \tag{13}$$

Thus μ_1 is a variable that has to do with the relationship

between the first spin and the second spin. But if we know it, then you know what the second spin is.

What about the next one ? $\mu_2 = \sigma_2\ \sigma_3$. Instead of thinking of the spins, think of the links. The links have two possibilities : either parallel or antiparallel. Knowing the links would tell us nothing about any individual spin, unless we knew the value of the first spin. But if we know σ_1 is up and we know μ_1, then we know σ_2. Indeed, if $\sigma_1 = +1$, then, from $\mu_1 = \sigma_1\sigma_2$, we deduce that

$$\sigma_2 = \mu_1 \tag{14}$$

Now supposing we know μ_1 and also μ_2, then we know what σ_3 is. What is it ? To calculate it, we can start from

$$\begin{aligned} \mu_2 &= \sigma_2\sigma_3 \\ &= \mu_1\sigma_3 \end{aligned} \tag{15}$$

Multiply both sides by μ_1

$$\mu_1\mu_2 = \mu_1^2\sigma_3 \tag{16}$$

But whatever μ_1 is, its square is 1. So we get

$$\sigma_3 = \mu_1\mu_2 \tag{17}$$

The general formula is left for reader to establish. It is the following exercise.

Exercise 2 : Given that $\sigma_1 = +1$, show that for all n

$$\sigma_n = \mu_1 \mu_2 \ ... \ \mu_{n-1} \tag{18}$$

In other words, knowing the first spin, if we know the μ's for all of the links in between, we know all the spins. And there is no redundancy, there is no double counting. As long as we know that the first spin is up, it is equally good to know the μ's, which live so to speak on the bonds between the spins, as it is to know the spins themselves.

That is very useful. Why ? Because the energy is just made up out of these bond variables. From equation (10), we can write

$$E \ = \ -j \ \sum_i \ \sigma_i \ \sigma_{i+1} \ = \ -j \ \sum_i \ \mu_i \tag{19}$$

We are not even multiplying anything anymore. With $-j$ in front, the energy E of a particular configuration is just the sum of the values of the individual bonds.

That is a powerful new expression, because the individual bonds are all independent, and that will come in handy when we calculate the partition function of the system.

What are the possible values of the bonds ? It is $+1$ or -1. It is $+1$ if the two spins they link are aligned, and -1 if the spins are antialigned.

Going back to the partition function Z, from equation (11) we can write

$$Z \;=\; 2 \sum e^{\beta j \sum_i \mu_i} \tag{20}$$

The factor 2 in front is because our above reasoning was only in the case σ_1 was up. But, as said, we must add the case when σ_1 is down. The values of the two terms are the same, whence the factor 2. But, as we know, a multiplicative factor is no big deal in a partition function.

As before, in equation (20), the sum in the exponent is over the bonds, and the big sum is over all the possible configurations. Notice there is one fewer bond than spins, although, if the number of elementary magnets is very large, it doesn't matter much.

Equation (20) resembles equation (3) we obtained in the very simple model studied at the outset, before the Ising model. The differences are that the σ_i which appeared in equation (3) is replaced by $\sum_i \mu_i$ in equation (20). This $\sum_i \mu_i$ is for one configuration. Then there is the big sum of $\exp(\beta j \sum_i \mu_i)$ over all the possible configurations, the calculation of which appears daunting. And there is the factor 2.

True, equation (20) looks much messier than equation (3) or even equation (3) to the power N, that is equation (5), but the important point we are going to use is that the μ's are independent of each other.

Remember that we established equation (5) following two methods : first, in the previous chapter, we looked at all the configurations of the N spins ; secondly we looked only at one spin, and wrote that the partition function for the

whole collection was that of one spin raised to the N-th power. We could do that because the spins in that simple model were all independent.

Now in the Ising model, where the *coupled spins* were astutely replaced by the *independent bonds*, we use the second method. That allows us to avoid working with the messy equation (20), and reach directly the simple answer. The partition function for the collection of bonds, is that of one bond raised to the $(N-1)$-st power, because there are $N-1$ bonds. And we have a factor 2 in front. So at one fell swoop we get

$$Z \;=\; 2 \;[\; 2\cosh(\beta j)\;]^{N-1} \tag{21}$$

We have reduced our problem to exactly the same problem as before. The only difference is the physical meaning of the degrees of freedom in the collection of spins or bonds that is a little bit different.

Now what is the average ? Let's not ask about the average of one spin, but about the average of one bond, that is the average of the product of two neighboring spins.

We can either write it as $< \mu_i >$, for any given μ_i, or as $< \sigma_i \; \sigma_{i+1} >$, for any given pair of adjacent spins. In any case, it is

$$< \mu_i > = < \sigma_i \; \sigma_{i+1} > = \; \tanh(\beta j) \tag{22}$$

It is not zero. And if we take j to be positive, it is positive. The fact that the average is non 0 and positive tells us that there is a net – in the sense of average – tendency for the

$(i+1)$-st spin to line up in the same direction as the i-th spin.

That is the same thing as saying that there is a bias for μ_i to be positive. If the first spin is found up, the next one has a better than even chance of also being found up. And what is the better than even chance ? It is $\tanh(\beta j)/2$ above $1/2$.

Exercise 3 : Using the fact that

$$< \sigma_1 \ \sigma_2 > = \ \tanh(\beta j)$$

prove that

$$\text{Prob } \{ \ \sigma_2 = +1 \mid \sigma_1 = +1 \ \} = \frac{1}{2} + \frac{\tanh(\beta j)}{2}$$

The left hand side is the usual notation for the conditional probability that $\sigma_2 = +1$, given that $\sigma_1 = +1$.

This tells us that there is a correlation between neighbors [5].

5. Remember, from elementary probability theory, that two random variables are independent if and only if the unconditional distribution of one of them – also called it *marginal distribution* – is the same as its *conditional distribution* given that we know the value of the other random variable.

The reader may also remember the nagging little fact that uncorrelation and independence are not exactly equivalent concepts. They are equivalent within the family of Gaussian r.v.'s. But they are not quite equivalent in general. It is possible to build two random variables which are dependent but whose correlation coefficient is 0. Counte-

But let's now go far down the line. And let's ask exactly the question which we asked when we started : What is the correlation between the i-th spin and one n units down the chain ?

How are we going to get at that ? The answer is very simple. We want to calculate

$$< \sigma_i \ \sigma_{i+n} > \tag{23}$$

We can write this as follows

$$< \sigma_i \ \sigma_{i+1} \ \sigma_{i+1} \ \sigma_{i+2} \ \sigma_{i+2} \ ... \ \sigma_{i+n-1} \ \sigma_{i+n-1} \ \sigma_{i+n} > \tag{24}$$

Why is that ? Because inside we multiplied by σ_{i+1}^2, σ_{i+2}^2, etc. But these squared random variables are always equal to 1, so we didn't change anything.

The nice thing is that expression (24) can be rewritten [6]

$$< \mu_i \ \mu_{i+1} \ \mu_{i+2} \ ... \ \mu_{i+n-1} > \tag{25}$$

Now again we shall use the fact all of the μ's are independent of each other. The energy is the sum of the energies. The problem completely factorizes. Expression (25) is also

$$< \mu_i > < \mu_{i+1} > < \mu_{i+2} > \ ... \ < \mu_{i+n-1} > \tag{26}$$

---

rexample : take $X = \text{Uniform}(-1, +1)$, and $Y = X^2$. Then you can check that $\text{Corr}(X, Y) = 0$, but they are obvioulsy not independent. However such counterexamples don't concern us here.

6. Notice that we could also have used expression (23) and equation (18) to establish expression (25).

And each of these averages is $\tanh(\beta j)$. So we have finally found that

$$< \sigma_i \ \sigma_{i+n} > = \ \tanh(\beta j)^n \tag{27}$$

n being the number of steps between i and $(i + n)$. This formula is expressing exactly the idea of the game of telephone. The coefficient $\tanh(\beta j)$, with is akin to fidelity, is always positive and less than or equal to 1.

When $\tanh(\beta j) = 1$, β is infinite, that is zero temperature, perfect fidelity. Then, no matter how far down the line you go, $< \sigma_i \ \sigma_{i+n} >$ in equation (27) will be one. This is the statement that at zero temperature if you know that μ_i is lined up, every one of them is lined up. There are only two possibilities, everybody parallel upward or everybody parallel downward. And in either case the product (26) is one.

But if there is any loss of fidelity at all, which means in this case the temperature is not absolutely 0, then $\tanh(\beta j) < 1$. And each time we go another step we multiply by a factor of $\tanh(\beta j)$. It like what we saw when fidelity in the telephone game was not perfect but only say 99% at each step. After two steps it was only about 98%, etc.

So, no matter how close $< \mu >$ is to 1, as long as it is less than 1, when we go sufficiently far down the chain the correlation $< \sigma_i \ \sigma_{i+n} >$, as a function of n, will become arbitrarily small. There would be no or negligible memory. Indeed $\tanh(\beta j)^n$ is an exponential function of n. And when the temperature is above zero, it is a number less than 1 raised to the n-th power. So the correlation falls exponentially fast with n.

Incidentally, the big trick of focussing on the bonds instead of the spins in figure 6, thereby replacing messy coupled random variables – the spins – by nice independent ones – the bonds –, and transforming the problem of finding the partition function, the energy, etc. into a tractable one, is a pattern in physics. It is called a *duality*.

In modern physics this could be thought of as the first duality, that is equivalence of different systems. We found an equivalence between a theory of spins which are connected to nearest neighbors, with another theory of just spins which are uncoupled, because we saw that the bonds in figure 6 behave like the spins of the very first problem where they were independent of each other, just plunged in a magnetic field.

This involved a clever change of variables which basically interchanged sites to bonds. Here is, for instance, a picture of sites and bonds in a 3D lattice, figure 7.

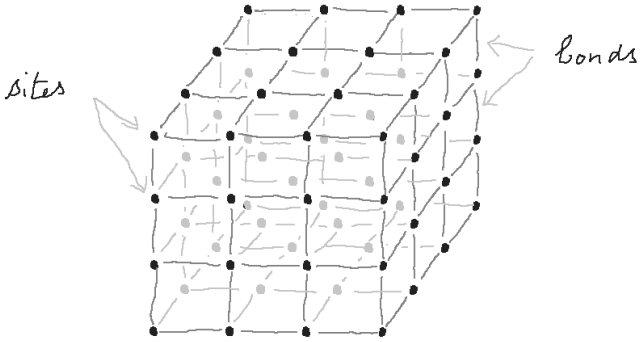

Figure 7 : 3D cubic lattice, with sites and bonds.

The sites and bonds are also sometimes called vertices and edges.

So we met the first example of a duality between different statistical mechanical systems. It is not always an interchange between edges and vertices, but in this case it is.

We found out some things about the *one-dimensional Ising model*. We found out that it is never magnetized, except at zero temperature. Magnetized can translate into the statement that there is this long-range memory : if one spin is up all the others will be biased up till very far away. That is called magnetization.

It is equivalent, in fact, to the statement that if we put a tiny magnetic field on one site of the system to bias it, it will cause everything to line up. We will see that with a little calculation in a moment.

In summary the one-dimensional Ising model is a boring system. It has no phase transition. It does exactly what we might have expected it to do. Correlations fade as we go down the line.

And the reason is simple. It is the same reason as in the game of telephone. There is a probability that we make an error when we go from location i to location $i + 1$. Once we make an error, we start over again with a new message. And we wait until we make another error. Then we start over again, and so forth and so on. And that is the way it goes.

Nevertheless there will be some clumping, a tendency to dis-

play clumps of spins in the same directions. It is just like in the game of telephone where there will be long stretches of agreement. And then statistically, on the average, every so often, there is a switch. But then a long chain of agreement again. Then a switch, etc.

Of course high temperatures are like the fidelity being very bad. The clumps or stretches are short then. Infinite temperature is like the situation where you just can't hear your neighbor. You get no information. He or she whispers in a din, and on top of that there is a dense fog, so that you can't even read his or her lips. You wind up, whatever your neighbor says, making a random guess. That is the statement that there is not even correlation between nearest neighbors. $T = +\infty$, $\beta = 0$, $\tanh(\beta j) = 0$, the probability calculated in exercise 3 is $1/2$, and $\mathrm{Corr}(\sigma_i,\ \sigma_{i+1}) = 0$.

How do we make the system more interesting? What is wrong in a linear chain is the dimension is too low. Once we make a mistake our neighbor doesn't have any support to know what the right answer is.

What about higher dimensions?

Ising model in two or more dimensions

Now, instead of playing telephone, let's play a different game. Players are sitting in a room, for example a lecture hall, with rows after rows of seats covering the entire floor. So there are also columns of seats. Each player has

four neighbors. They can be seen as the vertices of a two-dimensional square lattice.

Kevin over here starts a message. He sends it to the four fellows respectively to his right, his left, in front of him, and behind him. Then they each send the message to all their neighbors. In particular they will send their message to some new people who will receive them from several different directions, for example Seth in figure 8.

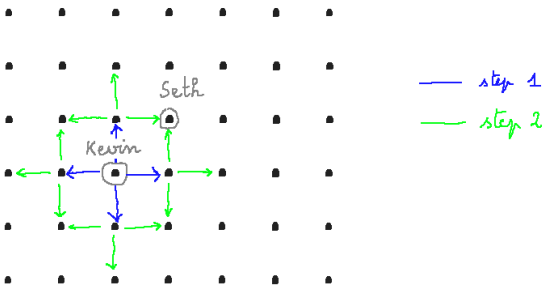

Figure 8 : Message transmission in a 2D square lattice. Players receive messages from several different directions.

As the message spreads throughout the lattice, new people will be getting messages from different sides. They will have to make their judgments about what is the right message, and therefore in turn what to transmit.

If after having sent his message a player becomes somehow new again and can again receive messages and transmit them, then as the game unfolds everybody will get basically four messages from all the people around them.

We are a player, and suppose three of those messages arriving at us say one thing, but one of them says another thing, what do we do ? We will do what computer scientists call error correction. It just means we will take the majority vote.

This system works much better. And in fact, if the fidelity is reasonably good, the initial message will spread throughout and off to infinity.

The same thing happens in the two-dimensional Ising model. The bias of anyone spin will bias the rest of the sample. Said another way, putting a little magnetic field on one spin in two-dimensions will bias the whole sample.

This is not completely obvious. Surprisingly, there is no easy solution. One way to prove it is going through the analysis of the two-dimensional Ising model following what Kramers and Wannier did.

That Ising got it wrong in two-dimensions is forgivable. But he thought and stated that in every dimension, if the fidelity was not 100%, then the message would always eventually get lost, as in one dimension. That is also wrong. And in high dimensions, it is not too hard to disprove.

If we think about it for a minute, we realize that it is very dimension dependent. Why is that ? In one dimension each person has only two neighbors [7]. That is not very many.

7. We studied a model where the information propagated in only one direction. We could have studied another one where information propagated in both directions. But there would be no way to use the rule of majority vote when the messages received differed.

The player at site i doesn't have much of a support system. He gets the message really only from one direction.

How many neighbors does a two-dimensional system have ? Four if it is a square lattice like in figure 8. So that means we will be getting messages from four people. We have a pretty good chance, using the majority rule, of registering the right answer even if the fidelity is not so good.

If we use some weighing procedure such as taking the majority, we do pretty well. We do much much better than if we were just getting the message from one person.

How about in three dimensions ? We have six neighbors, see figure 7. How about 10 000 dimensions. Whatever it is, 20 000 neighbors if the lattice is square.

I used to have this friend Art Harris. Art Harris was the first black Mr America. Anything you asked him, he would say : A hundred men can't be wrong. I don't know what he had in mind, but in high dimensions it is true a hundred neighbors won't be wrong very often. Fluctuations among large numbers of variables tend to be very small in comparison with the net average.

By that, we mean precisely the following. Consider n Bernoulli trials [8], that is n random variables X_i, independent of each other and each one taking value $+1$ with probability p, and -1 with probability $(1-p)$. Then the average of each of them is $2p-1$, and so is the average of $\sum_i X_i/n$.

8. named after Jacob Bernoulli (1654 -1705), Swiss mathematician, member of the the famous Bernoulli family of mathematicians and mathematical physicists.

But *the fluctuation of the average of the X_i's is much smaller than that of each X_i*. By fluctuation we mean the standard deviation. The variance of X_i is $4p(1-p)$. So the standard deviation of X_i is $2\sqrt{p(1-p)}$. Then, because the X_i are independent with the same distribution, the variance of the sum of the X_i's is the sum of the variances, that is

$$\text{Var}\left(\sum_i X_i\right) = 4np(1-p) \tag{28}$$

And if we divide the sum by n, the variance will be divided by n^2.

$$\text{Var}\left(\frac{1}{n}\sum_i X_i\right) = \frac{4p(1-p)}{n} \tag{29}$$

Finally, taking the square root, we reach

$$\text{Std dev}\left(\frac{1}{n}\sum_i X_i\right) = \frac{1}{\sqrt{n}}\ \text{Std dev}(X_i) \tag{30}$$

If $n = 100$, we see that the fluctuation of the average of the X_i is ten times smaller than the fluctuation of each X_i. This is in fact also a direct consequence of the law of large numbers, which is proved along similar lines, using Chebyshev inequality [9].

9. If we have a positive r.v. Y with mean μ, we prove first that, for any $t > 0$, $\Pr\{\ Y \geq t\ \} \leq \mu/t$. Then, if X is a random variable, which can be positive or negative, with mean μ and variance σ^2, Chebyshev inequality says that, for any $t > 0$,

$$\Pr\{\ |X - \mu| \geq t\ \} \leq \frac{\sigma^2}{t^2}$$

It is the familiar and intuitive result that the fluctuation of the average of n measures is smaller than that of each measure, and goes to zero like $1/\sqrt{n}$.

If p is higher than $1/2$, then $EX_i = 2p - 1$ is positive, and standard deviation of $X_i = 2\sqrt{p(1-p)}$. The expectation of the average of the X_i is also $2p - 1$. But, as we saw, the standard deviation of $\sum_i X_i/n = 2\sqrt{p(1-p)/n}$. So it is highly likely that the random average will be positive and therefore the message we will register in a transmission, which we select to be either $+1$ or -1, will be $+1$.

So, the higher the dimensionality, the better our shot at being able to propagate the information of one spin throughout the entire lattice, because then n will be high in the above calculations.

That leads us to ask the question : What would the Ising model be like in very high dimensions ? We consider some very large number of dimensions, ten thousand or whatever it is. Then what can we say about the propagation of the information ? We will observe a phenomenon.

You might even ask : Above what value is the number n of dimensions considered to be large ? It turns out 2 is high enough. For the study of the propagation of the information, using the majority rule, which is reinforced by the number of dimensions, it turns out that 2 behaves like 10 000.

As soon as the number of dimensions is more than one, there is a *phase transition*. Only the one-dimensional Ising model does not have a transition. When $n \geq 2$, if the pro-

bability p, i.e. the fidelity, is high enough, the information is transmitted throughout the lattice. That is not obvious, and could not have been predicted easily. It also turns out that 3 is already very large in this game. That can be calculated, or checked numerically. But anyway $n = 2$ is enough.

As p goes down from 1 to 1/2, at some point there is a phase transition from order to disorder. We won't prove it for $n = 3$, let alone for $n = 2$, because it is hard. But for sufficiently high dimensions there is an awfully good argument to say that there is a transition.

At high temperatures everything is random. As we lower the temperature, everybody wants to line up. And that tendency to lineup propagates throughout the system to infinity.

There are many possible tricks to do this thing. We are going to use something called *mean field approximation*. Instead of n, the number of dimensions will be denoted d. Each site is surrounded by how many neighbors ? In 3D the answer is 6, figure 9.

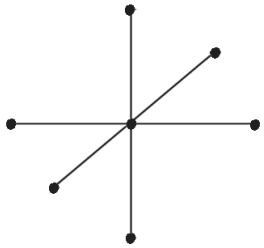

Figure 9 : Neighbors of a site when $d = 3$.

Each site is surrounded by $2d$ neighbors. Imagine that d is very large. And because d is very large, imagine that all of the neighbors define a *field* whose fluctuation is much smaller than its average value.

As we spelled out in the case of the Bernoulli trials, if we have a large number of variables, and they are biased, so that their biasses [10] add up to something of order the number of variables, then the fluctuation of the sum will typically be much smaller than its bias, because the fluctuation will grow like $\sqrt{n}$ compared to the bias which will grow like n. So that is what we are going to do.

Let's focus on the spin in the center of figure 9. For all the other spins surrounding it, we are going to make an approximation about them.

Let's write the energy of the spin in the center. It is

$$E = -j\sigma \sum_{neighbors} \sigma_{neighbor} \tag{31}$$

where σ is the value of the spin at the center. This formula is the generalization of $-j\sigma_i\sigma_{i+1}$ which appeared in equation (10). But be careful, the sum in equation (31) still concerns only one spin, whereas the sum in equation (10) came from the fact that we were looking at the whole sample of spins.

Now let's suppose that there is a bit of bias and that the average of the spins is not 0. Of course we are going to check

10. The bias of a random variable here means the departure of its expectation from 0. We use the word bias to emphasize that, unlike when it has mean 0, the variable carries information.

that. We are going to write a formula, that would allow the spin to be 0 too. But let's suppose throughout the lattice that the spin has an average denoted $\overline{\sigma}$.

$\overline{\sigma}$ is a notation, read *sigma bar*, meaning the average spin in the whole sample. It is lighter than $< \sigma >$, but is the same thing.

We don't know what the value of $\overline{\sigma}$ is. It might be 0, it might be positive, it might be negative. We don't know. But, in equation (31), we are going to replace the sum over neighbors by simply $2d$ times this average. That is a pretty good approximation if the number of neighbors is large. And the larger the number of neighbors, the better the approximation. When we have large numbers, as a typical rule [11] we are allowed to average and the fluctuations are small.

So equation (31 becomes

$$E = -2dj\sigma\overline{\sigma} \tag{32}$$

That is the energy of one particular spin sitting in the bath, or in the field, of all the others. This is called mean field approximation. Mean is not in the sense of nasty, but in the sense of average. And the field now is the field experienced by one spin in the field of all the others.

We can do the partition function of that one spin at the center of figure 9. In fact we don't even need to do the

11. Of course this depends on what we do with the average. And, in a rigorous proof, it must be checked that it can be done. Here we are following a partly heuristic argument to show that in high dimension, there is a phase transition.

calculation. We know how to do the partition function. It is exactly the same calculation we did starting from equation (2), except wherever we wrote j, we instead write $2dj\overline{\sigma}$.

So we can immediately write down the average value of the spin at the center of figure 9. Let's denote it $\overline{\overline{\sigma}}$, read *sigma double bar*. This average spin is going to be

$$\overline{\overline{\sigma}} = \tanh(2d\beta j\overline{\sigma}) \tag{33}$$

The spin at the center of figure 9 is moving in the background of all the others. And the others all constitute a constant field that we just call $\overline{\sigma}$.

This mean field approximation is also sometimes called *self-consistent field approximation*. Why self-consistent – hopefully self-consistent ? Because the spin at the center of figure 9 is no different than all the others. It is just one of the many spins in the system. And if each of the neighbors in figure 9 has an average equal to $\overline{\sigma}$, the physical intuition would require that the one at the center also should have the same average. That has got to do with translation invariance. Everyone is really the same as everyone else in a big sample.

So the self-consistent field theory implies that $\overline{\overline{\sigma}}$ is the same as $\overline{\sigma}$. That gives us an equation for $\overline{\sigma}$:

$$\overline{\sigma} = \tanh(2d\beta j\overline{\sigma}) \tag{33}$$

$2d\beta j$ is simply a constant factor, inside tanh, in front of $\overline{\sigma}$. And remember that β is the inverse of the temperature.

Equation (31) is an implicit equation for $\overline{\sigma}$. We want to solve it. How do we get an intuition about what it says ?

Question : Regarding the factor $2d$, is it only the number of nearest neighbors, or could it include next nearest neighbors ?

Answer : We dont' look at next nearest neighbors. $2d$ is the number of nearest neighbors in d dimensions. In one dimension it is two. It two dimensions it is four, see figure 8. In three dimensions it is six, see figures 7 or 9. In four dimensions it would be eight. Etc.

Next nearest neighbors would go in the same direction as going in higher dimension. That would just change the factor inside tanh in equation (31).

We picked a particular model and study it as a function of dimension. We could study it as a function of the nature of second nearest neighbor couplings, third nearest neighbor couplings. We can do all these things.

The model with a high number of dimensions we are studying is just one way of making a spin have a lot of nearest neighbors. And then this enables us to astutely use mean field approximation.

The physics of high dimensions is not something we realize in the laboratory [12]. But we do this procedure all time : We

12. Notice though that the experiments which we can do in the laboratory – as opposed to the gedanken experiments Einstein was fond of – are always studied, analyzed, interpreted with *models* of a mathematical nature. Sometimes these models involve high or even infinite

want to prove something about a system. It is a little too hard. We can't get our hands on it. But we can go, mixing mathematical and physical reasoning, to some limit where it becomes much easier. And we prove it in the limit.

We haven't proved it for the physical case of interest but at least we can prove that somewhere between, in our case, one dimension and a hundred thousand dimensions, a certain new kind of behavior happens that wasn't present in one dimension – namely the phase transition that we are going to show.

Once we see how that works, we can then ask where does the change happen ? And in our case it happens between dimension one and dimension two.

So when d is large, we get a nice simple example where we can view what happens physically without having to solve a difficult mathematical problem.

So let's try to solve equation (33).

The first thing to do is to change variables. To change variables often seems to obscure a problem, because we go from a variable we have an intuition for to one which seems

dimensions. Nature provides only raw facts, not elaborate models or interpretations. And even those raw facts are not devoid of puzzles. Is it really possible to perceive without any implicit models ? Can there be raw facts or phenomena we cannot perceive ? It is tempting to answer no. But then we are part of the experiments in some incomfortable way. These considerations may eventually lead to the next scientific paradigm which would look differently at Man looking at Nature, and would solve the riddles of the hypothetical multiverses and such things. See the appendix on the anthropic principle for more.

artificial and we can't interpret as easily. But there is a good reason. Equation (33) involves hyperbolic tangent of a complicated thing. We don't want a complicated thing in tanh. That is a nuisance.

So we are going to make $2d\beta j\overline{\sigma}$ our new variable :

$$y = 2d\beta j\overline{\sigma} \tag{34}$$

Equation (33) becomes

$$\frac{y}{2d\beta j} = \tanh y \tag{35}$$

Now to solve this equation, we are just going to graph both sides and see where they intersect. That is the crudest way to solve an equation : to draw a graph of both sides and see where they intersect.

On the left-hand side, instead of β in the denominator, let's put the temperature T in the numerator because it is a more intuitive quantity to follow. Equation (35) becomes

$$\frac{Ty}{2dj} = \tanh y \tag{36}$$

So we shall draw on the same graph $\tanh y$ and the simple linear function $Ty/2dj$ for various values of the coefficient $T/2dj$.

We begin with the first case when the temperature is high. Then the straight line $Ty/2dj$ intersects $\tanh y$ only at the origin, figure 10.

Of course, since d is a large number, T must be large too. How high the temperature must be depends on d. But eventually for a high enough temperature the slope of the straight line is more than 1, i.e. more than the slope of the hyperbolic tangent of y at 0.

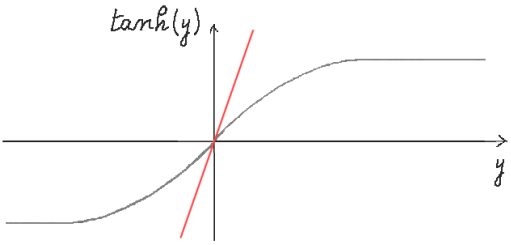

Figure 10 : $\tanh y$ and $Ty/2dj$ when T is high.

So we see that for high temperature, the only possible answer to the self-consistent field theory is that $y = 0$. But what is y ? y is proportional to the average of σ.

Thus we learned the first result that at high enough temperature, the average of σ must be zero, as expected. That is not too surprising.

Now let's start lowering the temperature. As T goes down, the slope of the straight line decreases. We eventually reach a temperature where the slope is equal to 1. We see that when

$$T = 2dj \tag{37}$$

the straight line becomes tangent to the hyperbolic tangent.

Something happens at that temperature. When we go beyond that, there is some new solution to equation (36), or equivalently to equation (33), figure 11.

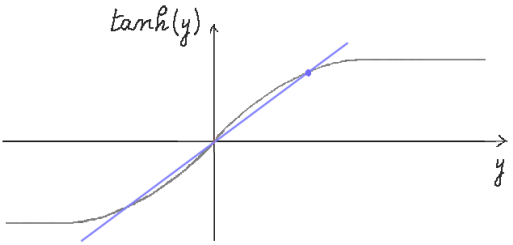

Figure 11 : $\tanh y$ and $Ty/2dj$ when $T < 2dj$.

Just to be clear let's assume j is positive, so we are only interested in the right part of the graph.

So now, beside the solution $y = 0$, there is another solution with y strictly positive, therefore where $\overline{\sigma}$ is not zero.

In other words the system can have an overall average of spin pointing all in the same direction. That is the phenomenon of *magnetization*.

What happens when T crosses $2dj$ is called a *phase transition*. In terms of T, below that point there is an average magnetization, an average field that just permeates the whole system forever.

Now you might say : Hold on, how do I know that the right physical solution wasn't the one at the origin ? Indeed for any value of the constants, the straight line does also intersect the tanh curve at the origin. So how do we know which

one is right ?

We are going to do a little calculation. We shall add to this problem a teeny little magnetic field that biases the system. And we shall discover that the teeniest little magnetic field will tell us that we should be on the right branch in figure 11, and not at the origin.

But we can see it another way. Let's go to zero temperature. Zero temperature is way out to the right in figure 11. And we know what to expect at zero temperature. We expect alignement. Everybody must be locked into a parallel configuration.

Then, as we raise the temperature, we slide along the right branch of tanh leftward toward the origin. Eventually when the increasing T reaches $2dj$, we hit the origin, that is the transition point. Beyond that temperature, the straight line is steeper than 1. Thence, there is only one solution, $y = 0$, or equivalently $\overline{\sigma} = 0$.

That is the nature of a phase transition, which we'll talk about a little more.

We have been working with d high so that we could apply the mean field approximation. We haven't shown that there is a phase transition even for low d as long as it is greater than or equal to 2.

We have already seen that there is no phase transition in one dimension, because as soon as $< \mu >$ is less that one – see equation (22) and infra –, the correlation between distant spins falls off to zero, so they cannot obey like an infi-

nite army of soldiers. But let's see, with a different explanation, why a phase transition can't happen in one dimension.

At absolute zero temperature everybody wants to align, figure 12.

Figure 12 : One dimensional collection of small magnets at zero temperature.

Now let's imagine writing the partition function as a sum over all the configurations. How can we enumerate the configurations starting with the configuration in figure 12 ? We can enumerate it by the number of spins that are flipped relative to this initial configuration.

So writing the partition function, we would start with e to the minus beta times the energy of the initial configuration. We are now familiar with this.

$$Z = \sum_i e^{-\beta E_i} \tag{38}$$

Starting from zero, figure 12 is the dominant configuration. Then we raise the temperature a little tiny bit. Some new configuration start to become important too. For example we flip one spin in figure 12.

To understand what will happen and not happen, let's

also do the same thing in two dimensions. We start with a two dimensional square lattice at zero temperature. All the spins are parallel, for instance all up, marked with a plus sign, figure 13.

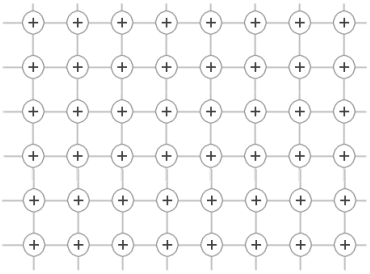

Figure 13 : Two dimensional square lattice of small magnets at zero temperature.

This is one of the two lowest energy states. Remember that in the Ising models the energy is stored in the *bonds* linking the sites. And two sites pointing in the same direction correspond to a bond with low energy, see equation (10) and the comments thereafter. So here we start with everybody pointing in the same direction.

Now let's ask what is the next state, the next higher energy ? What happens if we flip one spin ? How much energy do we put ? How many bonds have we broken ? Clearly if, in figure 13, we flip only one spin, we break four bonds [13].

13. Imagine of course the spin among all the spins, not along a border of the figure. And breaking a bond means to make it have two antialigned spins. In this case the energy of the bond becomes high.

And how many places are there where we can do that ? Basically on each site. The number of places where you can do that is the total number of sites. So that means there is a number of configurations that we can add in with four extra units of energy.

Now supposing we want to flip two spins, the next configuration. If the spins are next to each other, as in figure 14, then we break six bonds.

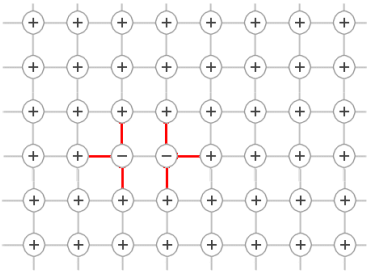

Figure 14 : Two dimensional square lattice, two adjacent spins flipped down, and the six broken bonds.

Notice, in that case, that the bond between the two flipped spins is parallel again, i.e. not broken. Each broken bond costs a certain amount of energy. So we have increased the energy by six broken bonds.

What would have happened if we would have put the second minus sign away from the first one ? Then we get eight broken bonds.

So each time we increase the number of flipped spins, it

costs us something. For each new configuration, a further Boltzmann factor, e to the minus beta times the energy, is added in the partition function. Each time we flip another spin, the energy goes up.

Now let's compare that with the one-dimensional case. From the configuration shown in figure 12, how much energy does it cost to flip one spin ? How many bonds get broken ? Two.

How about to flip two spins ? If they are adjacent, it is still just two broken bonds.

How about three spins ? If they are adjacent, it is still two. You can flip any number of them and it still only costs two units of energy. That means there is a lot of configurations all with the same energy.

In the two-dimensional case, the number of configurations with four extra units of energy would be just proportional to the number of sites.

In the one-dimensional case, the number of configurations with two extra units of energy is proportional to the square of the number of sites, because you can pick any two bonds and flip all the spins in between them. So it is a lot of configurations where a lot of spins are flipped.

That is why the one-dimensional case is unstable – in the sense that it doesn't require more engergy – with respect to flipping lots and lots of spins. And as soon as the temperature is turned on, it costs very little energy to flip whole big loads of them.

Whereas in the two-dimensional case to flip one more spin always costs some energy.

That is the basic mathematical effect that is going on.

Q. : Can't we see the effect of the difference in the number of dimensions from equation (36) which in one case is $Ty/2j = \tanh y$ because $d = 1$, and in the other case is $Ty/4j = \tanh y$ because $d = 2$?

A. : No we can't. Equation (36) only makes sense when d is large. We cannot use it for $d = 1$ nor $d = 2$.

The whole physics of the mean field approximation, the whole argument only make sense in high dimension.

Effect of a small external magnetic field

Now let's turn on a tiny magnetic field and see what happens. Let's still concentrate on one spin, the one at the center of figure 9 for instance. We saw that its energy function was

$$E = -2dj\overline{\sigma}\sigma \tag{39}$$

But now, in addition, the whole system has an external magnetic field. That means that each spin has an extra energy not related to its neighbors.

Let's denote the magnetic field B [14]. The new energy function for one spin is

$$E = -2dj\overline{\sigma}\sigma - B\sigma \tag{40}$$

where $B\sigma$ is favoring up, that is, B is positive. However we are going to imagine eventually that B is very small.

$B\sigma$ is the extra energy of the spin under consideration, not related to its neighbors as said. And each of the neighbors in figure 9 also has such an extra term.

That is the whole difference : instead of the hyperbolic tangent term, in equation (36), being $\tanh(2dj\overline{\sigma})$, it is going to be $\tanh(2dj\overline{\sigma} + B\beta)$. Equation (36) now becomes

$$\frac{Ty}{2dj} = \tanh(y + B\beta) \tag{41}$$

The right-hand side has a new graph, figure 15.

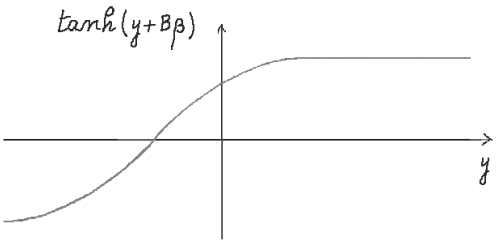

Figure 15 : Graph of $\tanh(y + B\beta)$.

14. Earlier it was denoted H, but it is the same.

We have assumed $B\beta > 0$, so that shifts the curve to the left, compared to that of figures 10 or 11.

Let's draw again the straight lines which we drew in figure 10 and figure 11, figure 16.

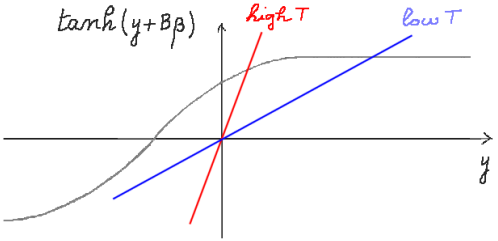

Figure 16 : Graphic solution to $\frac{Ty}{2dj} = \tanh(y + B\beta)$.

Equation (41) is a little more complicated than equation (36) because β is actually $1/T$. We can rewrite it

$$\frac{Ty}{2dj} = \tanh\left(y + \frac{B}{T}\right) \tag{42}$$

But anyway, at any given positive temperature, and therefore slope of the straight line, β, and shift B/T, there is no solution at the origin anymore. Only the intersection point on the right branch, for positive j, is possible [15].

Remember that, when we rotate it clockwise, *the straight line goes from vertical to horizontal as the temperature goes*

15. To be accurate, in figure 16, the tanh curve should be slightly different, with a different shift, for two different temperatures. But this doesn't change the conclusion.

from infinity to zero. The clockwise knob, so to speak, turns the temperature down.

In the high-dimensional case without a magnetic field, we saw that there is a phase transition : starting from infinite temperature, i.e. the straight line vertical, and slowly rotating it, i.e. lowering the temperature, we remained for a while in a zone of no magnetization, until T reached $2dj$. At that point there was a phase transition. Magnetization appeared, and of course remained till T was zero.

Now, when the whole system is plunged into a magnetic field, however tiny, we no longer have this phenomenon of phase transition appearing at some temperature. At any positive temperature, there is magnetization, that is there is a bias in $\overline{\sigma}$.

When T is infinite, the slope of the straight line is vertical, and at the same time the tanh curve in figure 16 is not shifted at all, then $y = 0$ is still the solution. There is no magnetization, as expected. But as soon as T is less than infinite, the only solution is magnetization.

Now we can understand better, in the case of no outside magnetic field what happens when the temperature is below the phase transition point, figure 11. The three solutions $y = 0$, $y > 0$, and the symmetric $y < 0$, are possible. But *the solution $y = 0$, i.e. no magnetization, is unstable.* Add the tiniest little magnetic field, no matter how small it is, and it will bias one way or the other the whole collection of spins. It is called *spontaneous symmetry breaking.*

The argument is that the solution where each spin is equally

likely to be up or down is unstable. The solution is that there is fifty fifty probability of everybody being up or everybody being down. If you turn on the tiniest magnetic field all the spins will turn one way, and the difference of energy to go to the other way can be enormous. Why ? Because if you have a tiny stray magnetic field up and you have a zillion spins, the total energy of up will be much lower than down, just because you have this great number of spins [16]. The symmetric argument works if you have the tiniest stray magnetic field pointing down.

That is what happens with a real magnet too. And that may help understand what is going on. You could imagine that you have a real ferromagnet which has equal probability of pointing everywhere. And then you bring it into the field of the Earth. Pretty quickly that magnet knows that the Earth's magnetic field is there. And it orients itself in that direction.

So the configuration where it is equally likely to be in every direction is unstable. A small stray field will orient it. And the stronger or the bigger the magnet, the more it is unstable.

In fact the Earth magnetic field itself flips once in a while. There must be an instability there too. But that is for geologists to explain why.

16. Remember that we are in several dimensions. The mean field approximation requires d to be high. But, as we said, the phenomenon of phase transition already manifests itself as soon as $d \geq 2$.