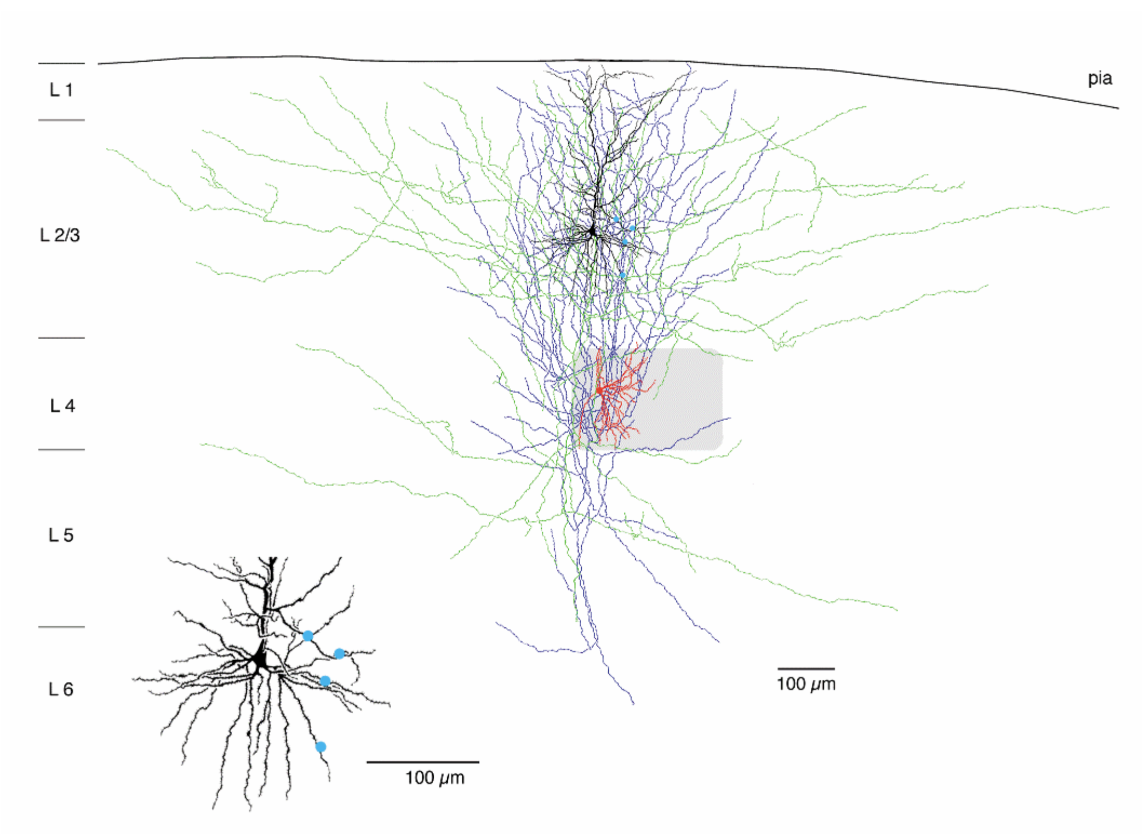
Lecture 1

[

The human brain is immensely complex. It contains about 100 billion (10^8) nerve cells ("neurons") interconnected by a quadrillion (10^{15}) synapses. These connections are not random, but very precisely laid down, by rules we will consider in this course. It is possible that each synapse is laid down with pinpoint precision! (How many different ways can 100 billion neurons be interconnected using a quadrillion different synapses?).

The following picture shows 4 synapses (blue dots) made between 2 neurons in the neocortex of a rat. The input cell body is located in layer 4 (red dot; the red branches are the dendrites of this input cell; the blue branches are from the axon of this cell). The output cell body (black dot) is located in layer 2/3; its dendritic branches are shown in black, its axonal branches are in blue (the small picture or "inset" at bottom left shows details of the placement of the synapses (blue dots) between the axon of the input cell (grey) and the dendrites of the output cell (black). The exact *position* of these synapses probably doesn't matter too much, what matters is the connection they form between the layer 4 cell and the layer 2/3 cell. Note that each of these cells can receive or form thousands of synapses, although only the 4 synapses that are formed specifically between these 2 neurons are shown. Don't worry if you are not familiar yet with terms such as "neuron", "synapse", "neocortex", "dendrite", "axon" – you will be at the end of the course!



Not only is the brain intriguing because of its complexity (for its size, it is probably the most complex object in the universe), but neuroscientists believe that it generates all our thoughts, emotions, memories, behavior, dreams and understanding. (An alternative view is that we possess a mystical “soul” which magically causes our brains to respond like a puppet on a string; such a view is unscientific but until neuroscience makes further progress it will continue to appeal. Indeed, given the great popularity of creationism despite overwhelming scientific evidence for evolution, the magic “soul” view will probably never wither).

In this course we will grapple with modern ideas about how the brain comes to be wired up so as to produce these amazing results. We will not be able to reach a complete picture, partly because there won’t be time, but mainly because neuroscience itself is very far from complete (one of the reasons why it is such an intriguing subject). However, at least it now seems possible to glimpse the sort of mechanisms that are at work, and to have confidence that we are on the right track – even for explaining some of the most “magical” aspects of brain function.

The course is NOT going to be a general introduction to neurobiology, with systematic coverage of all the topics found in a typical neuroscience textbook. (A good general introduction at Stony Brook would be Principles of Neuroscience”, BIO 334; obviously that course would provide good background to this course, but even if you haven’t taken BIO 334, or a similar course such as BIO 208 or BIO 328, I will explain all the key ideas from scratch). Instead, we will focus on the essential topics that are needed to understand how the brain comes to perform miracles like memory and understanding.

The title of the course “From synapse to circuit: selforganisation of the brain” tries to capture the aim of the course: we will be interested in neurons and synapses not so much as sophisticated biological devices assembled from myriad interacting molecules but as the key constituents of circuits that process information. In particular we will explore how mechanisms at the level of individual synapses lead to the formation of circuits that can actually do things (such as identify an image as the face of a friend). Obviously, to see how such useful circuits might *form*, we will also need to consider how such circuits *work*. An obvious analogy would be an automobile: we need to know how to operate it; if we are a mechanic we also need to know how it works; and if we are to build one we need to know how to assemble it. These are 3 different descriptions of automobiles. In the case of the automobile rather little knowledge at one level of description is needed at another level: the director of an assembly plant does not have to know how to drive! In the case of the brain, a knowledge of neuroscience does not help one think, dream or feel, but most of us are at least as curious about the phenomena that underlie these mental events as we are about our gearboxes or carburetors. It remains an open question to what extent understanding the nuts and bolts of neurons will be essential to progress in elucidating cognition and behavior. Many scientists feel that such messy details are irrelevant, or even distracting. For example, “Artificial Intelligence” experts at first believed that computer science would solve these problems, but gradually they have come to realize that even the simplest problems are incredibly difficult, and surpass the power of our best supercomputers. “AI” is now seeking inspiration from the brain. In this course we will focus more on how useful circuits are built, rather than how they work, but the 2 questions are inextricably linked. The reason why they are linked is suggested by the second half of the course's title : "selforganisation". If there is a really good way to recognize faces, or execute complex movements, or any of the other things that brains do, using sloppy biological devices like neurons rather than precise physical devices like transistors, one could in principle imagine that Darwinian evolution could wire up neurons into appropriate circuits, in much the same way that evolution has engineered lungs, hearts and blood vessels to distribute oxygen. Neuroscience would then be just a branch of physiology, and ultimately of molecular biology. However, we will see that there are limits to Darwinian evolution, and that brains have to make use of additional organizing principles. Of course, ultimately genes do specify the proteins that embody these organizing principles – in that sense brains are as much prisoners of our genes as genes are of the laws of physics and chemistry – but most of the heavy lifting is done by "learning" not evolution, by "nurture", not "nature". We will be exploring the link between "nature" and "nurture" not in the sense of looking at the relative contribution of each factor to specific psychological abilities or dispositions like intelligence or violence, but in the deeper sense of the relationship between the machinery of "nature" and "nurture", the limitations of each, and their commonalities.

What does the term "selforganisation" mean? The brain is very elaborately organized – each synapse may be placed with the precision of a note in a colossal symphony, or a tessera in a vast mosaic. But the genome only contains rough outline instructions : put square green pieces in the top lefthand corner. The mosaic has to assemble itself within general guidelines set by the genes. We say that the brain "self-organises" because it wires itself up using a combination of internal rules (which are embodied by proteins which are in turn specified genetically) and experience – experience which is to some extent unique to each individual. A good analogy for this "selforganising" process is the economy in a capitalist country like the US, as opposed to the "planned" economy of communist countries like the former Soviet Union or or Cuba. At one extreme, laissez-faire, individuals are free to try to maximize their own economic benefit, and the "invisible hand" of the market leads to efficient production and employment. At the other extreme, all economic decisions are taken centrally, often by bureaucrats following some arbitrary plan. Of course, all real economies have both elements, and the brain uses central planning (by the genome, and ultimately by Darwinian evolution) and laissez-faire (synapses following their own internal logic).

Another way to think about selforganisation is to return to the mosaic analogy. Imagine finding a collection of individual tessarae arranged in a certain pattern – a mosaic. One could have 3 different theories as to how the mosaic formed. First, someone might just have dumped the pieces out at random (we'll imagine the pieces are made so they all fall flat, rightside up, with no overlap and no gaps). Second, someone might have carefully assembled the pieces to form the pattern. Third, there might be hidden forces that cause pieces of certain colors and shapes to adhere in specific ways (red squares attach to green circles etc). The first case would be disorganization, the second, organization, and the third, selforganisation. Any particular mosaic could use all three mechanisms.

But saying the brain self-organises does not really explain the brain. We need to consider how rules at the level of synapses can lead to circuit selforganisation, in such a way that the circuits perform usefully. (If the rules did not lead to useful circuits, brains would not help genomes survive, and such rules would not arise and persist). It is not obvious what form such rules should take, nor how biological devices such as synapses can implement such rules. In this course we will explore these issues.

It is important to realize (and this point will only clearly emerge as we move deeper into these issues) that even though the brain selforganises by a *synaptic* rather than a *genetic* process, Darwinian evolution of the genome is itself a self-organising process. Indeed the central idea of modern biology is that organisms are the result of Darwinian evolution acting on self-replicating polynucleotides. The genome of organisms selforganises, as a result of selection (the "experience" of the species) acting on genes. Genomes were NOT created in a few days of intense theocracy, but by a slow logic of selfreplication and interaction with an evolving environment. (Some physicists think that selforganisation occurs at even deeper levels, beyond synapse and genes, and has led to the present structure of the universe).

Particularly in the case of humans, self-organisation of the brain by a combination of gene-embodied rules and experience in turn makes possible an entirely new level of selforganisation, involving language, culture, science and technology. This new level of complexity is beyond the scope of this course, but clearly it still involves the brain.

A Simple Example of Self-Organisation : Ferromagnetism.

A magnet is a piece of (typically) iron in which the iron atoms, each of which is itself a tiny magnet, line up in the same direction, so that their magnetic fields add up and can strongly influence other magnets. If you heat up such a magnet, its magnetism decreases because the thermal agitation of the atoms disrupts the alignment. At a specific temperature, known as the critical temperature (Tc), the magnetism disappears completely.

[**Insert graph**]

(Magnets that are not made of iron have different critical temperatures). The iron below the critical temperature is organized, because the atomic magnets are ordered, while above Tc the atomic magnets are disordered. The order can arise because the iron was exposed to a strong external magnetic field. But it can also arise spontaneously – if a hot piece of iron is cooled below Tc in the absence of a magnetic field, it spontaneously magnetises – in other words, it self-organises. What is happening?

To understand this, we study a simple "model" of the physical system. A model is representation that leaves out the inessential elements (iron versus other ferromagnetic materials such as cobalt; the shape of the magnet etc) but keeps the key elements. In this case the key elements are (1) the individual atoms, which have a quantum mechanical property called "spin" that makes the atom magnetic. Because of quantization, these atomic magnets or spins can only point either “up” or “down” and cannot point in intermediate directions. (2) the interaction between “spins”. A spin that points up generates a tiny magnetic field that tends to force nearby items also to point up, and likewise a downward spin tends to twist other nearby spins into the downward direction

[In these Notes passages in square brackets represent details that are added for completeness, but which I do not expect you to know. Here, it should be noted that the force that causes nearby spins to lie in the *same* direction is purely quantum mechanical in origin, operates over very short (atomic) distances and reflects the operation of the Pauli exclusion principle. Actual magnets, made up of billions of spins, behave quite differently: a magnet whose north pole is pointing up tends to twist nearby magnets so their north poles point *down*.]

(3) the disordering effect of temperature. Each spin is being randomly hit by nearby atoms which tend to flip its orientation. This disordering is more pronounced at higher temperatures.

In a specific version of this model called the Ising model the spin-spin interaction is only between next door neighbors.

Let us first consider the 1 dimensional Ising model. (Of course a row of single iron atoms does not physically exist, but nevertheless the model is instructive). Suppose all the spins are initially up: ..UUUUUUUUUU….
All the spins exert, via their magnetic fields, a twisting force on their neighbors, but this twist tends to keep the neighbors pointing in the same direction. In other words, all the spins are happy, because they all agree with each other.
Now suppose one of the spins (number 3) defects – it flips downwards, despite the upward force from its 2 neighbors, because of thermal bombardment:
……..UUDUUUU….
Consider the situation of atomic spin number 2 , just to the left of spin 3, the one that just flipped. It is now flanked by an up spin and a down spin, whose effects cancel out. So there is nothing keeping this spin up, and it too can flip, destabilizing spin 1. In this way a single flip can cascade through the entire system – and it doesn’t matter how probable the first flip was. The result is that all the spins keep flipping back and forth randomly, and there is no net magnetization – at any temperature above absolute zero. Though spins like to agree with their neighbors, there are not enough neighbors to restrain renegade spins. So in the 1D Ising model, net magnetization (of the whole set of iron atoms) is not possible above absolute zero. Since we know that iron *can* be magnetized, the 1D Ising model is *too* simple – it does not capture the physical reality of the situation we are interested in.

In the 2 dimensional Ising model (again, not physically realistic, but a step in the right direction)), each spin is surrounded by several neighbors (for example, one above, one below, one to the left, one to the right). Even if a spin flips, it will not destabilize its neighbors (which are each restrained by 3 unflipped spins. Thus individual defections do not cascade through the ranks, and the initial organized state can be retained. But if 2 neighboring spins should flip, this can, in principle, cascade through the whole system. So we need to consider how likely double flips are at any particular temperature. Clearly double flips will be more likely at higher temperature, but we need a quantitative relation. Unfortunately, whether double flips occur will depend on what is happening to the neighbors of the neighbors, and so on, and this calculation rapidly becomes exceedingly complicated – in fact it defeated the best mathematicians and physicists for decades (and, in the case of the 3D Ising model, continues to defeat them).

But we can make progress by changing the model slightly. Instead of trying to figure out how all the various possible configurations of flipped and unflipped spins will influence each other, let us assume that the force acting on each spin reflects the average state of all the other spins. (In effect we are assuming that all spins affect each other equally). This is often called the "mean field approximation" because we ignore the details of the state of the spins, and consider only their mean (or average) state.

First, we must consider how a downward pointing magnetic field H will affect the probability p_u that a single isolated spin points up compared to the probability p_d that it points down. At absolute zero, in the absence of thermal motion, the spin will always align with the field (however weak), so $p_u = 0$. At very high temperatures, the thermal buffeting will be so strong that however strong the field, a spin is as likely to point up as down: $p_u = 0.5$. But at intermediate temperatures, the effect of the field (which depends both on H and on J, the strength of the single-spin magnet) can be calculated using a basic relation discovered by Boltzmann. He showed that at equilibrium the probability that a particle sits in a high energy state compared to a low energy state depends exponentially on the difference in energy (ΔE) between the 2 states:

$p_{hi}/p_{lo} = \exp(- \Delta E/kT)$ where k is Boltzmann's constant.

(note : exp (x) is another way of writing e^x , i.e the number e raised to the xth power. The number e is 2.718….; it is very important in science because the rate of change of exp (x) is equal to x. This *exponential* function is the only function whose derivative (rate of change) is equal to itself. A function is simply a rule for calculating variables from other variables. Other examples of functions are x^2, sin x,1/x, tanh x. In the Boltzmann formula, the dependent variables are _E and T, the absolute temperature. k is the Boltzmann constant. kT is roughly the thermal energy possessed by a molecule at the temperature T; the formula says that that it is the relative importance of the potential energy of a molecule and its thermal energy which determines the actual state of a molecule).

Note that if the difference between the low energy and high energy states exactly equals the thermal energy (ΔE = kT), the probability that the particle occupies the high energy state is only 1/e or about 37%, and at sufficiently high temperatures the particle distributes equally between the 2 states (make sure you understand why $e^{-0} = 1$). At absolute zero, all the particles must be in the low energy state, and $p_{hi} = 0$.

Applying this to our problem, we find that the magnetisation *m* of a lump of iron composed of **independent** spins each obeying Boltzmann's formula would be given by

$$m = \tanh (HJ/kT)$$

We define magnetization such that when all the spins point up m = 1, and when they all point down m = -1. If equal numbers of spins point up and down, then m = 0. Of course in a real piece of iron the spins are NOT independent – they interact. We will include the interaction in a later, final step.

Remember, H is the external magnetic field which tends to *align* the spins; J is the strength of the spins; T is the absolute temperature, which tends to *disorder* the spins.The tanh function is closely related to the exponential function:
tanh x = (exp x – exp –x)/(exp x + exp-x)

This is shown plotted as a set of graphs. At any given temperature the magnetization points down (i.e. is negative) in a strong downward pointing field, and points up in a strong upward field. The magnetization saturates in strong fields, because all the spins have completely aligned (represented by m = +/- 1). The transition from saturated down to saturated up is sigmoid, with no net magnetisation (half the spins up, half down) in the absence of a field (remember, we are assuming in this case the spins are INDEPENDENT i.e. non-interacting). The steepness of the sigmoid depends on the ratio of J to T. The graphs show curves for low (blue) high (red) and intermediate (green) temperatures. As the temperature goes down the curves get steeper and steeper.

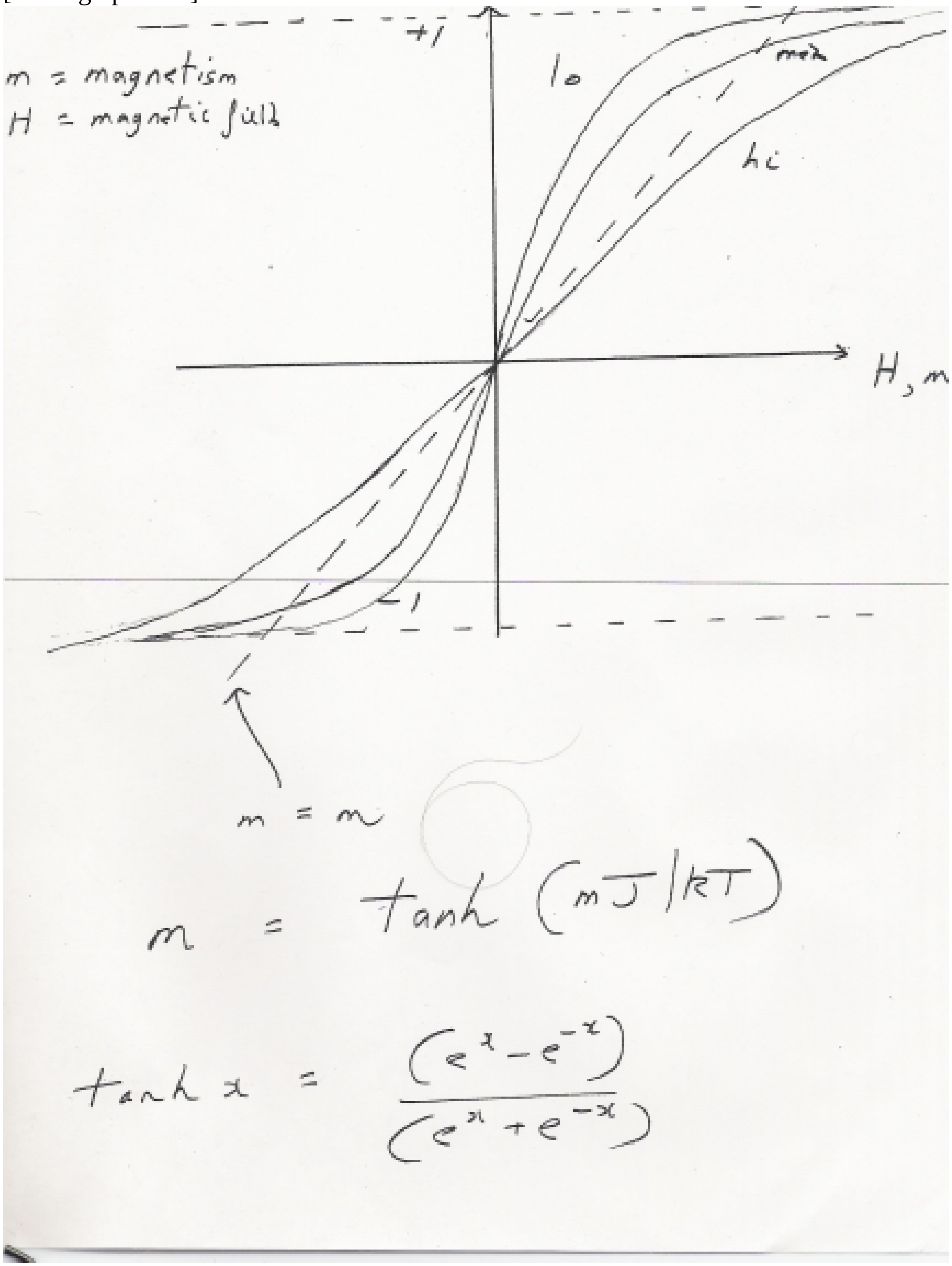
The final step is to introduce the assumption that the field acting on each spin is simply the average field *internally* generated by all the spins, rather than some *externally* generated field. In other words, we are asking whether the magnetic field generated by the piece of iron itself, in the absence of an externally imposed field, could be strong enough to align enough of the spins to generate an internal field strong enough to align enough of the spins……..and so on forever. Or, in essence, we are asking whether the iron can keep itself magnetized, and in particular under what quantitative circumstances this possibility is not excluded.

We are thus assuming that the magnetic field experienced by each spin within the iron reflects the combined magnetic field of all the other spins. This means we can replace H by m, resulting in the equation

$$m = \tanh (mJ/kT).$$

This equation is a celebrated equation in physics, the Weiss mean field equation. The solutions of the equation must correspond to the magnetization that appears *spontaneously* in the absence of an external field (since this was our assumption). They can be found by realizing that, since the equation m = m is obviously true, the solutions are given by the intersections of the curves generated by these 2 equations. The line corresponding to m=m is a straight line passing through the origin, with a 45 degree slope (why?).

[Insert graph here]



Graph Legend The graph shows plots of relative magnetization (m, plotted as the ordinate or y-direction) against magnetic field (abscissa or x-direction) for 3 different values of the temperature (high, medium and low) for a ferromagnetic (such as a lump of iron). When all the spins in the iron point up, the relative magnetization is 1, when all the spins point down the relative magnetization is –1. This happens at sufficiently strong magnetic fields (large positive or negative abscissal values). With zero magnetic field, in response to an external field, and IGNORING the effect that neighboring spins have on each other, half the spins will point up (on average) and half down. At all temperatures the relationship is sigmoidal, but is steeper at lower temperatures (since the ordering effect of a magnetic field is more pronounced at lower temperatures). If however, there is no external magnetic field, there could still be a ("spontaneous") internal magnetic field, if a majority of the spins happen to point up or down. This internal magnetic field would be proportional to the relative magnetization m, so we can think of the external magnetic field H being replaced by the possible internal magnetic field proportional to m (which is why the abscissa in this case is also labeled m). Under these conditions the relationship between magnetization and magnetization would still be sigmoidal (solid curves); since also clearly m must equal m (dotted straight line), the actual spontaneous magnetization must lie both on the sigmoid curve and on the dotted straight line i.e. it must lie on the intersections of these 2 lines.

One of these intersections, where m = 0, occurs for all values of J/T. It is found at the origin, where both m and tanh (mJ/kT) are zero. It corresponds to half the spins up, half down. However, this state is not physically stable, because if one of these exactly balanced spins flips, it will cause a mass flipping of the now minority spins. This situation is like a pencil balanced on its point – the tiniest disturbance will cause collapse.

The other intersections occur only if the slope of the midpoint of the sigmoid is greater than 45 degrees, which will occur only below some critical temperature Tc. There are then 2 symmetrically disposed points corresponding to equal up and down net magnetizations. If the temperature is decreased, these points separate, corresponding to greater degrees of spontaneous equilibrium magnetization (spontaneous magnetization increases at lower temperatures). The two intersections represent physically stable situations, since if a small fluctuation away from these points occurs (a few spins flip spontaneously), it will tend to return to the point. If the temperature is increased too far, there are no stable intersections, and no net magnetization, as we saw in the first graph.

Below Tc the strength of the interaction between the spins is always enough to partially overcome the disordering effect of thermal agitation, but above Tc it is too weak to do so. Because the iron undergoes a transition from an ordered state to a disordered state at Tc we say it has undergone a phase transition (somewhat like the melting of ice). If we cool a sample of iron in the absence of an external field, the magnetization suddenly develops at Tc, but whether it follows the up branch or down branch is quite random. The completely symmetrical state above Tc (equal numbers of up and down spins) is destroyed below Tc, as the system adopts one of the 2 (equally likely) asymmetrical states. This process is called symmetry breaking and is often encountered in selforganising systems. They have several alternative ordered states, which one is chosen depends either on chance, or on some some external biasing force. In the present case, if the disordered iron is cooled in the presence of a weak external field (for example, the earth's own magnetic field), it will follow the branch that agrees with the external field. In fact the developing magnetisation will be extremely sensitive to the field, since only a very weak external force biases the choice. This illustrates another feature common to self-organising systems – they use internal *amplification* of external forces.

[This mean-field version of the 2D Ising model does predict qualitatively the right result – that above a critical temperature spontaneous magnetization disappears – but the shape of the predicted curve of *m* versus *T* is slightly wrong. The reason is the meanfield assumption itself is wrong: the state of a spin depends not on the average effect of all the other spins but on the state of its neighbors. A far more sophisticated analysis is required, and has been achieved, though it cannot be extended to 3 dimensions.]

This simple model shows a crucial feature of many self-organising systems: they are disordered under some conditions, and ordered under others. Notice the rather devious style of argument we used: in order to *prove* that iron spontaneously magnetises below a critical temperature, we *assumed* that it *does* spontaneously magnetise! We then used the resulting “internal” magnetic field (i.e. the field within the iron resulting from the assumed spontaneous magnetization) to calculate what fraction of the spins would point up (or down) – i.e. the magnetization. We are thus essentially asking whether the assumption of spontaneous magnetization is self-consistent – compatible with our knowledge that the fraction of spins that are up or down is a sigmoidal (i.e. tanh) function of internal magnetic field strength. We saw that if T is below Tc, then the spontaneous magnetization assumption *is* self-consistent (but only if the spontaneous magnetisation takes one of two equal but opposite values), whereas above Tc it is not self consistent, and so cannot occur. This is related to a common proof technique in mathematics: one attempts to show that the opposite of what one wants to prove is selfcontradictory. This technique was used for example by Euclid 2000 years ago to prove that there are an infinite number of prime numbers (he showed that assuming there is a finite number is self-contradictory).

A qualitatively very similar situation exists in the nearest neighbor 2D Ising model, except that the critical temperature is slightly different. However, this model is quite difficult to analyse, and will not be considered here. The 3D Ising model is too complex to analyse completely, but it still has the same qualitative behavior as our “mean-field” model. And real iron also behaved qualitatively like this.

It is important to realize that iron, and other self-organising systems, are qualitatively (and not merely quantitatively) different on either side of the phase transition, in much the way that ice is qualitatively different from water. (Indeed if you have never seen ice melting you might not realize that ice and water are actually different states of the same substance).

These models seem to be unrelated to anything going on in the brain (the electrical activity of the brain does generate weak magnetic fields, which can be detected by superconducting devices called “SQUIDS”; likewise strong magnetic pulses can interfere with the electrical activity of brain tissue; however these magnetic effects play no role in the normal operation of the brain). But they do contain the key “brainlike” *element of many interacting units*. We can think of the spins as neurons which are either firing ("up") or not firing ("down"). The firing of a neuron influences its "neighbors" – the cells to which it is connected by synapses. However, in the ferromagnet all spins have the same effects on neighbors, while in the brain each neuron has a different effect on its targets (because the strengths and numbers of the synapses very). This gives the brain much more complex behavior.

The development of order in the ferromagnet is a **collective** property which is due to the interactions between the spins. It is not seen in a collection of non-interacting spins. It is also a *cooperative* phenomenon since it emerges as a result of cooperation between individuals. Finally it is also an *emergent* phenomenon: the phase transition is only manifest when there are large numbers of interacting spins, and nothing like it occurs at the level of individual spins. But although spontaneous ferromagnetism, and the associated phase transition, is only manifest as a collective, cooperative and emergent property of the whole system, it is nevertheless a direct consequence of the properties of the individual atomic magnets. Our account in a sense "reduced" the phenomenon to elementary interactions, there is nothing "mysterious" about the behavior (although in some ways it is quite subtle). Our account was reductionist but still, in a way, the whole is definitely more than the sum of the parts – the magnet exhibits properties which are qualitatively different from those of its components. Likewise, the neuroscientist seeks to "reduce" the brain to interactions of neurons, synapses, molecules etc, but that does not imply that thought, mind, behavior, consciousness etc are "simply" the consequence of neural firings. We anticipate that "mind" emerges from "neurons" in the same way that spontaneous bulk magnetization emerges from "spins", though of course mind is a much richer phenomenon than magnetism. However, until we can account for *how* it emerges, the analogy does not tell us much.

In ferromagnetism the order is extremely simple, a choice between up or down. In more complex self-organising systems the number of possible ordered states is much richer. In the next lecture we will examine another, much more spectacular, self-organising process, Darwinian evolution.

Useful Web Links

http://www.physics.uiuc.edu/research/ElectronicStructure/389/slides/389-lect23/sld014.htm

http://www.tf.uni-kiel.de/matwis/amat/elmat_en/kap_6/backbone/r6_3_1.html

SUMMARY

A ferromagnet is made up of lots of atoms that behave as little magnets, also called spins. These spins can either point up or down. If most of them point the same way (either up or down) the iron is magnetized. The force tending to make the spins point in the same direction is the magnetic field. It can be either externally generated (by a large magnet for example) or internally generated (by the combined

effect of all the aligned spins). The force tending to mix up, or randomize, the spin directions, is thermal agitation (collisions with other atoms).
If only the external magnetic field is in play (for example the atoms are so far apart they do not feel each others' magnetic field) then the stronger the field the more aligned the spins (and the stronger the magnetization of the iron), and the higher the temperature the more random the spins (and the weaker the magnetization). But if only the internally generated magnetic field is in play, then the spontaneous magnetization resulting from spin-spin interactions (the tendency of neighboring spins to adopt the same alignment) disappears completely above a critical temperature.