

Quenched Disorder: Understanding Glasses using a Variational Principle and the Replica Method

Jonathan S. Yedidia

Lyman Physics Laboratory

Harvard University

Cambridge, MA 02138

Lectures delivered at the Santa Fe Summer School
on Complex Systems, June 15-19, 1992

Outline

- I. Introduction and Elements of Statistical Mechanics
- II. Variational Approach Applied to Proteins and Magnetic Spin Systems
- III. Averaging over Disorder and the Replica Method
- IV. Variational Replica Approach and Replica Symmetry Breaking
- V. Replicas and the Impure Superconductor in a Magnetic Field

Lecture I: Introduction and Elements of Statistical Mechanics

A. Introduction: the goal of these lectures is to learn about the new techniques that have been developed to compute thermodynamic properties (such as the free energy, specific heat, and correlation functions) of physical systems with quenched disorder.

B. Physical Systems: a physical system is characterized by the states it can be in, and the observable quantities which are a function of the state. The Hamiltonian (energy function) is a very important observable quantity. The partition function, if known exactly, gives complete information about the thermodynamics of a physical system.

C. Two Exactly Soluble Models: we compute the free energy exactly for two kinds of physical systems: non-interacting Ising spin systems, and particle systems with quadratic interactions.

D. The Variational Approach: we can define a "trial free energy" which is a quantity that will be greater than the true free energy for any any "trial Hamiltonian," and which will approach the true free energy as the "trial Hamiltonian" approaches the true Hamiltonian.

E. Trial Hamiltonians: our exactly-soluble systems provide a source of trial Hamil-

tonians which permit the analytical computation of trial free energies.

F. Toy Example of the Variational Approach: we use the variational approach for a simple one-dimensional model, and interpret the results.

Lecture II: Variational Approach Applied to Proteins and Magnetic Spin Systems

A. The Protein-folding Problem: we want to predict the shape of a protein, knowing only the sequence of amino acids. We assume that the energy function is known and is a sum of two-body terms.

B. Quadratic Trial Hamiltonian: we assume a trial Hamiltonian in which every monomer is linked to every other monomer by a spring, and try to optimize the spring constants to minimize the trial free energy.

C. Heteropolymer Self-consistent Equations: we derive a set of self-consistent equations which, when solved, predict the position and correlated fluctuations of all the monomers, given the temperature and two-body potentials.

D. The Ising Spin Glass Hamiltonian: this is a model of Ising spins on the lattice in which the interactions between neighboring spins can be either ferromagnetic or antiferromagnetic. The interactions are chosen from a probability distribution, and are quenched.

E. Mean-field Theory: when applied to ferromagnetic spin systems, the variational approach generates a self-consistent equation (mean-field theory) which becomes exact when the number of dimensions of space approaches infinity.

F. Corrections to Mean-Field Theory: an alternative derivation of mean-field theory (an expansion in powers of the inverse temperature at fixed magnetization) provides a way to systematically approach the true free energy. The first term correcting ordinary mean-field theory is particularly important for spin glasses, and gives the Thouless-Anderson-Palmer free energy for the Sherrington-Kirkpatrick model.

Lecture III: Averaging over Disorder and the Replica Method

A. Averaging over Disorder: we are interested in computing the average free energy for an ensemble of systems (or samples) where each sample has a Hamiltonian which is chosen from a probability distribution.

B. A Very Simple Toy Model: we compute directly disorder averages for a very simple ensemble of systems; one in which each sample consists of a single quadratic well, with its center at a random position. We distinguish between correlation functions measuring thermal and disorder fluctuations.

C. The Replica Method: we explain the mathematical identities underlying the replica method, and work out some interesting results for n by n replica-symmetric matrices.

D. Check of the Replica Method: we re-derive disorder averages for our very simple toy model using the replica method. The replica method involves averaging over disorder first, and leaving any additional computations for later. A quadratic effective replica Hamiltonian with a replica-symmetric Green's function has a straightforward interpretation in terms of the original ensemble of samples.

E. Random Potentials: we learn how to mathematically describe rough random potentials in terms of Gaussian probability distributions. The first two moments of the probability distribution provide complete information about it.

F. Averaging over Random Potentials with the Replica Method: we derive an effective replica Hamiltonian resulting from an average over a random potential.

Lecture IV: The Variational Replica Approach and Replica Symmetry Breaking (RSB)

A. Variational Approach to a Toy Model: we derive the replica-symmetric trial free energy for a toy model of a particle in a rough random potential. The result for the average fluctuations is pathological; the explanation is that a replica-symmetric trial Hamiltonian describes a particle in a single well, and the toy model describes a particle in a potential with many metastable minima.

B. One-step RSB: the idea of one-step RSB is that the off-diagonal elements representing correlations between different replicas need not be identical. The replicas can be grouped into families, and intra-family matrix elements will have different values than inter-family matrix elements.

C. Mathematics of one-step RSB: we show how to manipulate one-step RSB matrices, starting with their multiplication.

D. Physical interpretation of One-Step RSB: one-step RSB has a straightforward interpretation in terms of an ensemble of samples which have intra-sample disorder including metastable minima.

E. Full RSB: the infinite-step generalization of replica symmetry breaking actually has a very convenient mathematical form in terms of a function of a variable which ranges from 0 to 1. It can be interpreted in terms of each sample being constructed as a infinite hierarchy of wells within wells.

F. The Full RSB Solution of the Toy Model: we show how the Full RSB solution of the Toy model cures its pathologies, and describes a "freezing" phenomenon as the temperature is lowered.

Lecture V: The Variational Replica Approach to Impure Superconductors in a Magnetic Field

A. Type-II Superconductors and the Abrikosov Crystal: a type-II superconductor in a magnetic field will exhibit an intermediate phase in which the magnetic flux

penetrates the sample as a triangular lattice of flux lines.

B. Perfect Elastic Crystals: the Abrikosov Crystal of flux lines can be described in terms of a perfectly quadratic classical Hamiltonian. We discuss the microscopic and continuum versions of this model, and introduce the most general Hamiltonian consistent with the triangular symmetry.

C. Random Pinning Potentials: we discuss oxygen vacancies in cuprate superconductors as an example of a quenched defect which could give rise to a rough random potential.

D. Trial Free Energy and Self-consistent Equations: we average over disorder and introduce a quadratic trial Hamiltonian, with full RSB. The solution for the Green's function is a function of both momentum and the replica variable.

E. Physical Correlation Functions: we compute various physical correlation functions using the variational replica method. The agreement with Bitter pattern decoration experiments is good.

I: Introduction and Elements of Statistical Mechanics

In these lectures, I will be discussing the statistical mechanics of various disordered physical systems which have been called "glasses" because of some similarities of their properties to those of more familiar glasses. I will only cover a small portion of the subject, governed partly by my own personal idiosyncratic tastes and mainly by a desire to give a pedagogical introduction to some remarkable theoretical ideas which may at first sight appear overly intimidating. These lectures will concentrate on some general methods of calculation that have proven to be especially useful for systems with quenched disorder. The advantage of focusing on general methods is that when one understands them, one can use them on many different problems in the future. These lectures will be unashamedly technical—we will be striving for a mathematical understanding of the physical problems we consider. On the other hand, because these lectures are being delivered at a summer school where many of the students are not physicists, they will also be unashamedly pedagogical and will assume no mathematical background beyond calculus and matrix algebra. I will discuss technical subjects, but I will try my best to introduce all the technical matter in as gentle and comprehensible a way as possible, assuming no previous exposure to the subject of these lectures at all. As you shall see, it will still be possible to address problems at the frontiers of current research.

The general goal in these lectures will be to learn how to compute thermodynamic properties of disordered physical systems. Physicists have long understood how to compute properties like the specific heat of a classical crystal or the magnetic susceptibility of a ferromagnet, or long-distance correlation functions in either system, as long as the crystal or ferromagnet is perfectly regular. They have taken advantage of the symmetries in these systems to invent such important theoretical concepts as phonons and spin waves. In a glass, however, the randomness is intrinsic; each atom in the system is in a different complicated environment, and it seems at first

like an impossible goal to compute, say, the specific heat or correlation functions, as precisely as we are used to for ordered systems. Fortunately, since 1975, there has been substantial progress in learning how to make analytical computations for glasses; these lectures are devoted to teaching you about some of the exciting new ideas and concepts that have been invented.

We will make combined use of two major tools for these computations: a variational principle, and the replica method. In the first two lectures we will introduce the variational principle without using any replicas, and then in the remaining lectures explain, in order of increasing complexity, the ideas behind the replica method. Variational principles are very well established in physics, but the power of the replica method is still not as generally appreciated. For a non-technical introduction to the history of the replica method and its manifold applications, the reader can consult P.W. Anderson's series of articles on the spin glass in *Physics Today* [1]. For a much more technical treatment, together with a collection of reprints of the more important replica articles up to 1987, see reference [15]. In this set of lectures, I will not cover the many interesting applications of the replica method to a variety of problems like neural network theory or optimization problems. Instead, I will concentrate on trying to explain the method itself as simply as possible.

In this first lecture we will review some fundamentals of statistical mechanics and solve exactly two completely trivial models. The reason that we care about these models is that as we go on, we will be studying much more complicated models which we have no hope of solving exactly. The trivial models will prove useful as building blocks for powerful methods which give us approximate results about the more complicated models. We will also introduce one of these approximation methods in this first lecture. Even experienced physicists should find it worthwhile to review these simple models, because I will be presenting them in a way that will ultimately make esoteric theories like the TAP equations for spin glasses and Gaussian replica

field theory much more transparent. In future lectures, we will be applying the approximation techniques to such complex physical systems as proteins, spin glasses, and impure superconductors in a magnetic field.

Enough generalities; let's begin studying the statistical mechanics of some physical systems. A physical system will be characterized by the different states it can be in, and various "observable" quantities which can be measured and which are a function of the state of the system. One very important observable quantity is the energy, or "Hamiltonian." Our first trivial example of a physical system is a single *Ising spin* in a magnetic field. An Ising spin, denoted by the variable S_1 , can be in two states: "up," when $S_1 = 1$, or "down," when $S_1 = -1$. (The subscript 1 in " S_1 " is just there to indicate that it is our first spin. If we had two spins, we would label them S_1 and S_2 .) For a system consisting of a single Ising spin in a magnetic field, the Hamiltonian is

$$H = -h_1 S_1 \tag{1}$$

where h_1 is a magnetic field which tends to align the spin to point "up" if the field is positive, and "down" if the field is negative. The fundamental principle of statistical mechanics is that the probability that a system is in each possible state is proportional to the Boltzmann weight $e^{-H(\text{state})/T}$ of that state, where T is the temperature. The central object of study in statistical mechanics, from which we can compute all thermodynamic quantities of interest, is the partition function Z , which is the sum of the Boltzmann weights of all the states of the system:

$$Z \equiv \sum_{\text{states}} e^{-H(\text{state})/T}. \tag{2}$$

For the model of a single Ising spin with Hamiltonian given by equation (1) we easily find that

$$Z = e^{-h_1/T} + e^{h_1/T}. \tag{3}$$

As I mentioned, when one knows the partition function exactly, as we do for this trivial model, one can then calculate all the thermodynamics exactly. Thus, the *free energy* F is defined by

$$F = -T \ln Z, \quad (4)$$

the *entropy* S is given by

$$S = -\frac{\partial F}{\partial T}, \quad (5)$$

the *internal energy* U is given by

$$U = \frac{T^2}{Z} \frac{\partial Z}{\partial T} = F + TS \quad (6)$$

and the *specific heat* C is given by

$$C = \frac{\partial U}{\partial T} = -T \frac{\partial^2 F}{\partial T^2}. \quad (7)$$

Notice that if we add an over-all constant to the Hamiltonian, it is also added to the free energy and the internal energy, but does not affect the entropy or the specific heat.

The probability that the system is in any of its states is just equal to the Boltzman weight of that state divided by the sum of the Boltzman weights of all the states:

$$p_{state} = \frac{1}{Z} e^{-H(state)/T} \quad (8)$$

We can define the *thermal expectation value* of any state-dependent quantity as the average (weighted by the probability of each state) of that quantity. For example, you can check that the internal energy is actually just the thermal expectation value of the Hamiltonian itself, while the entropy is the thermal expectation value of the negative of the logarithm of the probability of the state, and is thus a measure of how "spread out" the system is between its possible states:

$$U = \sum_{states} p_{state} H(state) = \langle H \rangle \quad (9)$$

$$S = \sum_{states} -p_{state} \ln(p_{state}) = \langle -\ln(p_{state}) \rangle \quad (10)$$

where the angular brackets are a convenient short-hand notation for the thermal expectation value.

There are other thermal expectation values that we might be interested in. For example, the *magnetization* m_1 of the spin is just the average value of the spin, which for this model we can easily relate to the magnetic field and temperature:

$$m_1 \equiv \langle S_1 \rangle = \tanh(h_1/T). \quad (11)$$

Finally, the *susceptibility* χ_1 is defined as the response of the magnetization to a change in the magnetic field:

$$\chi_1 = \frac{\partial m_1}{\partial h_1}. \quad (12)$$

So far, it has been natural to think of the free energy and all the other thermodynamic quantities as functions of the magnetic field h_1 . But equation (11) gives a simple relation between the field and the magnetization, so if we prefer, it is a simple matter to replace the magnetic field by the magnetization and define a magnetization-dependent free energy. We will see later that this is often a convenient thing to do.

Now let us introduce another trivial model. In this model, we consider a particle which moves in one dimension, and which can be located at any position r_1 from negative to positive infinity. (The subscript "1" is again a label indicating that this is the first particle.) The Hamiltonian will be a function of the position r_1 . The partition function is

$$Z = \int_{-\infty}^{\infty} dr_1 e^{-H(r_1)/T}. \quad (13)$$

To make this model exactly soluble, we restrict ourselves to a Hamiltonian function $H(r_1)$ which is quadratic in r_1 :

$$H(r_1) = \frac{1}{2G}(r_1 - a_1)^2. \quad (14)$$

Using the well-known formula for Gaussian integrals (in the appendix, I give a table of Gaussian integrals which will be useful in these lectures), we find that

$$Z = \sqrt{2\pi TG} \quad (15)$$

Using formulae (4)-(7), we can again compute the free energy, entropy, internal energy, and specific heat. We can also compute (again using the appendix) a couple of thermal expectation values which are particularly relevant for this model. Namely, the *average position* is given by

$$\langle r_1 \rangle = \frac{1}{Z} \int_{-\infty}^{\infty} dr_1 r_1 e^{-(r_1 - a_1)^2 / 2GT} = a_1 \quad (16)$$

while the *average fluctuation* in the position is given by

$$\langle (r_1 - a_1)^2 \rangle = \frac{1}{Z} \int_{-\infty}^{\infty} dr_1 (r_1 - a_1)^2 e^{-(r_1 - a_1)^2 / 2GT} = TG. \quad (17)$$

Physically, one can imagine that the particle described by this model is attached to a spring which is nailed to the position a_1 . The amount that it bounces around its average position a_1 is determined by the combination of the temperature T and the softness of the spring G .

In statistical mechanics, we are usually most interested in systems which have a very large number of degrees of freedom, rather than just one as in the examples given. So how can we generalize these models so that they are still exactly soluble but concern a large number of spins or particles? For the Ising spin system, there is not much that we can do beyond considering the system of many *non-interacting* spins, each under the influence of its own private magnetic field, for which the Hamiltonian would be:

$$H = - \sum_{i=1}^N h_i S_i \quad (18)$$

where N is the total number of spins. The free energy for this model is just the sum of the free energies for all the individual spins.

We can generalize the other model in a somewhat more interesting way. A Gaussian integral over many *interacting* degrees of freedom will still be soluble as long as all the interactions are quadratic. Thus, we can consider the generalized Hamiltonian for N particles:

$$H = \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N (r_i - a_i) (G^{-1})_{ij} (r_j - a_j) \quad (19)$$

where $(G^{-1})_{ij}$ is a symmetric matrix of the spring constants connecting all the particles together. We have written this matrix as an inverse matrix to agree with the common convention. This generalized model represents N particles, each attached with a spring to a nail at its own position, but now also connected by springs to every other particle. We have left the values of all the spring constants as general parameters. The wonderful thing about Gaussian integrals is that we can still compute our partition function; we find, using a formula from the appendix, that

$$Z = \sqrt{(2\pi T)^N \det G} \quad (20)$$

where $\det G$ is the determinant of the G matrix. Another extremely useful result is that the expectation value of the correlated fluctuations of two particles around their average positions is simply related to the G matrix:

$$\langle (r_i - a_i)(r_j - a_j) \rangle = TG_{ij} \quad (21)$$

The G matrix is often referred to as the "Green's function."

Now you may be thinking that these exactly soluble models are great, but most interacting systems we know of do not have these Hamiltonians, so what use are they? We shall see that these models can actually be used as inputs for a couple of different approximation schemes.

The first of these approximation schemes is based on a general mathematical inequality. Assume that we have some arbitrary physical system which can be in, say

K different states. The probability of each state is some number p_α ($\alpha = 1, 2, \dots, K$) where

$$\sum_{\alpha=1}^K p_\alpha = 1. \quad (22)$$

Let us also imagine that there is some observable quantity X (like the energy) which depends on which state the system is in. We label the different possible values of the quantity by X_α . By our previous notation, the thermal average of X is

$$\langle X \rangle \equiv \sum_{\alpha=1}^K p_\alpha X_\alpha. \quad (23)$$

The mathematical inequality that I will assert (without proof) is that

$$\langle e^{-X} \rangle \geq e^{-\langle X \rangle} \quad (24)$$

or written out more explicitly

$$\sum_{\alpha=1}^K p_\alpha e^{-X_\alpha} \geq e^{-\sum_{\alpha=1}^K p_\alpha X_\alpha}. \quad (25)$$

You can check for yourself that this inequality follows from the convexity of the exponential function.

Let us now return to our model of a single particle whose energy depends upon its position. We want to compute the partition function

$$Z = \int_{-\infty}^{\infty} dr_1 e^{-H(r_1)/T} \quad (26)$$

Before, we had an especially convenient quadratic function for $H(r_1)$, but imagine that we now have some more complicated function which makes the integral difficult or impossible to compute analytically. We can use our inequality to obtain an approximate solution as follows. The partition function is obviously equal to

$$Z = \int_{-\infty}^{\infty} dr_1 e^{-H(r_1)/T} \frac{\int_{-\infty}^{\infty} dr_1 e^{-H_0(r_1)/T}}{\int_{-\infty}^{\infty} dr_1 e^{-H_0(r_1)/T}} \quad (27)$$

where $H_0(r_1)$ is *any* function at all. We can rewrite this as

$$Z = \frac{\int_{-\infty}^{\infty} dr_1 e^{-[H(r_1)-H_0(r_1)]/T} e^{-H_0(r_1)/T}}{\int_{-\infty}^{\infty} dr_1 e^{-H_0(r_1)/T}} \int_{-\infty}^{\infty} dr_1 e^{-H_0(r_1)/T} \quad (28)$$

or

$$Z = \left\langle e^{-(H-H_0)/T} \right\rangle_0 \int_{-\infty}^{\infty} dr_1 e^{-H_0(r_1)/T} \quad (29)$$

where the notation $\langle X \rangle_0$ means the thermal average of X using the function $H_0(r_1)$ as a so-called *trial Hamiltonian*. We can now use our inequality to assert that

$$Z \geq e^{-\langle (H-H_0)/T \rangle_0} \int_{-\infty}^{\infty} dr_1 e^{-H_0(r_1)/T} \quad (30)$$

for *any* function $H_0(r_1)$. In terms of the free energy $F \equiv -T \ln Z$, we can equivalently assert that

$$F \leq -T \ln \int_{-\infty}^{\infty} dr_1 e^{-H_0(r_1)} + \langle H - H_0 \rangle_0 \equiv \tilde{F} \quad (31)$$

where we define the quantity on the right-hand side of the inequality as the *trial free energy* \tilde{F} corresponding to the trial Hamiltonian H_0 . This is our fundamental variational principle. (R.P. Feynman was one of the first physicists to make use of this principle in his famous treatment of the polaron problem [9]. Another interesting application of the variational principle is to the excluded volume problem in polymer physics [6] [3].) It says that the true free energy will always be less than the trial free energy no matter what trial Hamiltonian we choose. Thus, if from some class of trial functions $H_0(r_1)$, we find one that gives a minimal trial free energy, we know that that is our best estimate of the free energy. Note that if H_0 is equal to H , the trial free energy is automatically equal to the true free energy.

To be able to use this variational principle in practice, of course, we must restrict ourselves to a class of trial functions H_0 for which we can analytically compute \tilde{F} . The best trial Hamiltonian will be the one in this class which is closest to the real Hamiltonian. Finding a class of analytically tractable trial Hamiltonians is precisely

where our previously analysed exactly soluble models become useful. Let us look at a relatively simple example. Consider the Hamiltonian function (we drop the subscript "1" on r_1 for brevity)

$$H(r) = Cr^2 + r^4. \quad (32)$$

If $C \geq 0$, then the function has the form of a "single-well" potential, while if $C < 0$, $H(r)$ has the form of a "double-well" potential with a barrier of height $C^2/4$ separating the two valleys. In either case, the free energy $F \equiv -T \ln \int_{-\infty}^{\infty} dr e^{-H(r)}$ is some perfectly well-defined function of T and C , but this function is rather difficult to compute analytically. We will compute the function approximately using our variational principle with a class of trial Hamiltonian functions of the now-familiar form

$$H_0(r) = \frac{1}{2G}(r - a)^2 \quad (33)$$

where G and a are now arbitrary variational parameters that we will vary in order to minimize the trial free energy.

The trial free energy is

$$\tilde{F} = -T \ln \int_{-\infty}^{\infty} dr e^{-H_0(r)/T} + \langle H - H_0 \rangle_0 \quad (34)$$

$$\tilde{F} = -T \ln \int_{-\infty}^{\infty} dr e^{-(r-a)^2/2GT} + C \langle r^2 \rangle_0 + \langle r^4 \rangle_0 - \frac{1}{2G} \langle (r-a)^2 \rangle_0 \quad (35)$$

$$\begin{aligned} \tilde{F} = & -T \ln \int_{-\infty}^{\infty} dr e^{-(r-a)^2/2GT} + C \int_{-\infty}^{\infty} dr r^2 e^{-(r-a)^2/2GT} + \\ & \int_{-\infty}^{\infty} dr r^4 e^{-(r-a)^2/2GT} - \frac{1}{2G} \int_{-\infty}^{\infty} dr (r-a)^2 e^{-(r-a)^2/2GT} \end{aligned} \quad (36)$$

Now comes the key step, where we take advantage of the fact that our trial Hamiltonian has such a convenient form. Because it is quadratic, all these integrals are

simple to compute. Using the integrals in the appendix, we find that

$$\tilde{F} = -\frac{T}{2} \ln(2\pi TG) + C(a^2 + TG) + (a^4 + 6a^2TG + 3T^2G^2) - \frac{T}{2}. \quad (37)$$

It is interesting to note that adding a constant to the trial Hamiltonian would not change the trial free energy. The constant would have been added in the

$$-T \ln \int_{-\infty}^{\infty} dr e^{-H_0(r)/T}$$

piece of the trial free energy but subtracted in the $\langle -H_0 \rangle_0$ piece.

Minimizing the trial free energy with respect to a , we find that $a = 0$ or

$$a^2 = -\left(\frac{C}{2} + 3TG\right) \quad (38)$$

while minimizing with respect to G gives

$$-\frac{1}{2G} + C + 6a^2 + 6TG = 0 \quad (39)$$

Let us examine the solution in more detail. When $C \geq 0$ (the single-well case), we find that the only solution is $a = 0$ and

$$G = \frac{1}{12T} \left[\sqrt{C^2 + 12T} - C \right]. \quad (40)$$

Returning to the trial free energy, which is our estimate for the true free energy, we finally find that

$$\begin{aligned} \tilde{F} = & -T \left(\frac{1}{4} + \frac{1}{2} \ln \left(\frac{\pi}{6} \right) \right) - \frac{T}{2} \ln \left(\sqrt{C^2 + 12T} - C \right) \\ & + \frac{C}{24} \left(\sqrt{C^2 + 12T} - C \right) \end{aligned} \quad (41)$$

On the other hand, in the double-well case when $C < 0$, there can be two possible solutions, depending on the temperature. For temperatures higher than some critical temperature T_c , we have a functionally identical solution to that for $C \geq 0$, with

$a = 0$, G given by equation (40), and \tilde{F} given by equation (41). For temperatures below T_c , there is another solution with a lower trial free energy. The variational parameters corresponding to this other solution are

$$a^2 = -\frac{1}{8} \left(5C + \sqrt{C^2 + 12T} \right) \quad (42)$$

and

$$G = \frac{1}{24T} \left[\sqrt{C^2 + 12T} + C \right]. \quad (43)$$

and the trial free energy is

$$\begin{aligned} \tilde{F} = & -T \left(\frac{5}{8} + \frac{1}{2} \ln \left(\frac{\pi}{12} \right) \right) - \frac{T}{2} \ln \left(\sqrt{C^2 + 12T} + C \right) \\ & - \frac{5C}{48} \left(\sqrt{C^2 + 12T} + C \right) - \frac{C^2}{4} \end{aligned} \quad (44)$$

The critical temperature itself is simply determined by setting the high-temperature trial free energy of equation (41) equal to the low-temperature trial free energy of equation (44).

What is the physical meaning of this transition between two solutions? First, it should be made clear that the real free energy does *not* have the "cusp" that our trial free energy exhibits, so that is an artifact of our approximation. Nevertheless, it is not a stupid artifact, because there really are two temperature regimes for the true free energy—it is just that the cross-over between them is a gradual one. The Hamiltonian represents a ball in a double-well potential, and that ball is jiggled around by random hits from some back-ground "stuff." When the temperature is high, the jiggling is strong, while when the temperature is low, the ball will just sit at the bottom of the potential, as the jiggling will be weak. Clearly, there will be a high-temperature regime, when the particle bounces back and forth between the two valleys easily, because there is enough thermal energy to get over the barrier. There will also be low-temperature regime, when the particle tends to spend a very long time in one

hill before it bounces over the barrier. These two regimes are represented in our solutions. In our low-temperature solution, the ball has some average position at a , with a fluctuation G which represents how much the ball jiggles around that. In the high-temperature solution, $a = 0$ which means the ball is bouncing back and forth between the two wells, and G is much larger than in the low temperature solution, which again corresponds to the larger fluctuations.

In general, one should be aware that in any variational method, the results one gets for the quantity one is minimizing over (in our case, the free energy) can be very accurate, but the results for other quantities which one deduces from the minimization (like in our case, the size of the fluctuations) will not be as accurate. To understand this, let us imagine that we are minimizing the quantity F over some multidimensional space which we represent by the vector \vec{x} . The minimum of $F(\vec{x})$ will always be quadratic in \vec{x} . That means that if we miss the optimal \vec{x} by some small amount $\delta\vec{x}$ because we have restricted ourselves to a certain portion of \vec{x} -space where we can compute $F(\vec{x})$ analytically, our estimate for the value of F will only be wrong by $(\delta\vec{x})^2$. This does not mean that one should ignore the variational solution except for the upper bound it gives for the free energy (unless you are a rigorous mathematical physicist, in which case it means precisely that), as it is still true that the best \vec{x} found will be closest in the sub-space chosen to the optimal \vec{x} . It just means that one should beware that a poorly chosen subspace (or class of trial functions) can produce misleading results.

II: Variational Approach Applied to Proteins and Magnetic Spin Systems

In this lecture, we will begin by applying the variational method we learned about in the last lecture to the important problem of protein folding. The work that I will describe was done in collaboration with Jean-Philippe Bouchaud and Marc Mézard [4] at the Ecole Normale Supérieure in Paris. I will be discussing a highly simplified model of "proteins" in this lecture; but the ideas presented here could be generalized to a more realistic model.

A protein is a polymer which can be specified by the sequence of amino acids which make up its monomeric units. This sequence is stored biologically in the DNA segment which is ultimately translated into the protein. The sequence of amino acids determines the three-dimensional shape of the protein, and that shape in turn determines how well the protein performs its biological function. The "protein folding problem" is the problem of predicting the three-dimensional shape of a protein given only the sequence of amino acids. It is attracting considerable interest because present technology makes it much easier to sequence proteins than to determine their shape.

The shape of a protein can be specified at varying levels of precision. One could for example, specify bond-angles between neighboring amino acids, or one could specify the position of every atom in the protein. Let us, for the purposes of simplicity, consider a generic model of an N -monomer linear heteropolymer in which the position of the i th monomer in the chain is given by the D -dimensional vector \vec{r}_i . (In ordinary space, $D = 3$ of course, but there is no particular difficulty caused by keeping the dimension of space arbitrary.) We will make the huge assumption that the Hamiltonian for our heteropolymer is known and can be reduced to a sum of two-body monomer-monomer potentials; i.e.

$$H = \sum_{1 \leq i < j \leq N} V_{ij}(|\vec{r}_i - \vec{r}_j|) \quad (45)$$

The effect of the solvent is taken into account in this Hamiltonian only insofar as the two-monomer potentials are affected by it.

Now suppose that at the temperature we are interested in (physiological temperatures for a protein) the heteropolymer has a shape that is well-defined. That is, up to global rotations and translations, each monomer has an average position and some typical fluctuation around that position. This is certainly the case for globular proteins at physiological temperatures, although it is interesting that the physiological temperature is usually not much less than the temperature at which proteins undergo a transition to a fluctuating "coil state" with no definite shape, which suggests that thermal fluctuations are rather significant. If the heteropolymer does have a definite shape, it should not be too bad an approximation to consider a quadratic trial Hamiltonian which assumes that each monomer has some average position with Gaussian fluctuations around that position. Thus we re-introduce the trial Hamiltonian that we first mentioned in the last lecture:

$$H_0 = \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N (G^{-1})_{ij} (\vec{r}_i - \vec{a}_i) \cdot (\vec{r}_j - \vec{a}_j) \quad (46)$$

where \vec{a}_i and G_{ij} are two sets of variational parameters which have straightforward interpretations: from our discussion of this Hamiltonian in the first lecture, we know that \vec{a}_i is the average position of the i th monomer, while G_{ij} is proportional to the correlated fluctuation of monomer i and monomer j . As usual, we will eventually choose these variational parameters to minimize the trial free energy. Because we have so many variational parameters to vary, and because our trial Hamiltonian describes a system which is physically close to the true state of our system, the results we derive from our variational approach should be rather reliable. In fact, one can introduce an even more general and realistic trial Hamiltonian with a tensorial structure for G :

$$H_0 = \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \sum_{\alpha=1}^D \sum_{\beta=1}^D (G^{-1})_{ij}^{\alpha\beta} (r_i^\alpha - a_i^\alpha) (r_j^\beta - a_j^\beta) \quad (47)$$

where α and β are spatial indices. Such a trial Hamiltonian allows the protein to have anisotropic fluctuations, which is obviously desirable. For the sake of simplicity, we shall keep here to the form of equation (46) which is isotropic in space.

We now compute the trial free energy. The following derivation may seem complicated, but keep in mind that it is actually a straightforward generalization of the example from the last lecture. We have

$$\tilde{F} \equiv F_0 + \langle H - H_0 \rangle_0 \quad (48)$$

where

$$F_0 = -T \ln Z_0 \equiv -T \ln \int d\vec{r}_1 d\vec{r}_2 \dots d\vec{r}_N e^{-H_0/T} \quad (49)$$

and $\langle X \rangle_0$ denotes the expectation value of the observable X with respect to the Boltzman measure $\frac{1}{Z_0} \exp(-H_0/T)$:

$$\langle X \rangle_0 \equiv \frac{1}{Z_0} \int d\vec{r}_1 d\vec{r}_2 \dots d\vec{r}_N e^{-H_0/T} X \quad (50)$$

Some of these integrals can actually be done very easily. Using the appendix, we find that

$$F_0 = -T \ln \left([(2\pi T)^N \det G]^{D/2} \right). \quad (51)$$

Using the matrix identity $\ln \det G = \text{Tr} \ln G$, we find

$$F_0 = -\frac{NDT}{2} \ln(2\pi T) - \frac{DT}{2} \text{Tr} \ln G \quad (52)$$

For $\langle -H_0 \rangle_0$, we have

$$\langle -H_0 \rangle_0 = -\frac{1}{2} \sum_{ij} (G^{-1})_{ij} \langle (\vec{r}_i - \vec{a}_i) \cdot (\vec{r}_j - \vec{a}_j) \rangle_0 \quad (53)$$

$$\langle -H_0 \rangle_0 = -\frac{DT}{2} \sum_{ij} (G^{-1})_{ij} G_{ij} = -\frac{NDT}{2} \quad (54)$$

Finally, the most interesting and difficult integral to compute is

$$\langle H \rangle_0 = \sum_{1 \leq i < j \leq N} \langle V_{ij}(|r_i - r_j|) \rangle_0. \quad (55)$$

To compute it, it helps to first learn a general fact about Gaussian integrals. We clearly need to be able to compute quantities like $\langle f(\vec{r}_i - \vec{r}_j) \rangle_0$, where f is some arbitrary function. $\vec{z} \equiv \vec{r}_i - \vec{r}_j$ is a *Gaussian variable*, which technically means that you can determine all of its higher moments from the first two. Its first two moments are

$$\langle \vec{z} \rangle_0 = \vec{a}_i - \vec{a}_j \quad (56)$$

and

$$\begin{aligned} \langle (\vec{z} - \langle \vec{z} \rangle_0)^2 \rangle_0 &= \langle [(\vec{r}_i - \vec{a}_i) - (\vec{r}_j - \vec{a}_j)]^2 \rangle_0 \\ &= DT(G_{ii} + G_{jj} - 2G_{ij}) \end{aligned} \quad (57)$$

The important general fact that we need to know is that when one computes the average of some function of a Gaussian variable over all the original N variables in the problem, one can replace the measure over all the original variables with a new measure over only the Gaussian variable, as long as the new measure gives the same first two moments. Perhaps it is clearer if I just write it in equations; if \vec{z} is a D -dimensional Gaussian variable with first two moments \vec{M}_1 and DM_2 then

$$\langle f(\vec{z}) \rangle_0 = \int \frac{d\vec{z}}{[2\pi M_2]^{D/2}} \exp \left[-\frac{(\vec{z} - \vec{M}_1)^2}{2M_2} \right] f(\vec{z}) \quad (58)$$

Similarly, although we will not need it here, an average of a function of two Gaussian variables can be reduced to an integral over just those two variables if one uses a measure which gives the same first two moments. By the way, we know \vec{z} is a Gaussian variable because linear combinations of Gaussian variables are themselves Gaussian and the original variables of a quadratic Hamiltonian are always Gaussian variables.

Given the general fact described above, the computation of $\langle H \rangle_0$ follows immediately. We find

$$\langle H \rangle_0 = \sum_{1 \leq i < j \leq N} \int \frac{d\vec{z}}{[2\pi T(G_{ii} + G_{jj} - 2G_{ij})]^{D/2}} \exp \left[-\frac{(\vec{z} - (\vec{a}_i - \vec{a}_j))^2}{2T(G_{ii} + G_{jj} - 2G_{ij})} \right] V_{ij}(|\vec{z}|) \quad (59)$$

To summarize our computations, we have found that

$$\begin{aligned} \tilde{F} = & -\frac{NDT}{2}(\ln(2\pi T) + 1) - \frac{DT}{2} \text{Tr} \ln G \\ & + \sum_{1 \leq i < j \leq N} \int \frac{d\vec{z}}{[2\pi T(G_{ii} + G_{jj} - 2G_{ij})]^{D/2}} \exp \left[-\frac{(\vec{z} - (\vec{a}_i - \vec{a}_j))^2}{2T(G_{ii} + G_{jj} - 2G_{ij})} \right] V_{ij}(|\vec{z}|) \end{aligned} \quad (60)$$

Minimizing the trial free energy with respect to all the \vec{a}_i , we find that for all $i = 1, 2, \dots, N$

$$\begin{aligned} 0 = & \sum_{j=1, N(j \neq i)} \int \frac{d\vec{z}}{(2\pi)^{D/2} [T(G_{ii} + G_{jj} - 2G_{ij})]^{(D/2+1)}} [\vec{z} - (\vec{a}_i - \vec{a}_j)] \\ & \exp \left[-\frac{(\vec{z} - (\vec{a}_i - \vec{a}_j))^2}{2T(G_{ii} + G_{jj} - 2G_{ij})} \right] V_{ij}(|\vec{z}|) \end{aligned} \quad (61)$$

while minimizing the trial free energy over all the G_{ij} tell us that for all $i \neq j$,

$$\begin{aligned} (G^{-1})_{ij} = & 2 \int \frac{d\vec{z}}{(2\pi)^{D/2} [T(G_{ii} + G_{jj} - 2G_{ij})]^{(D/2+1)}} V_{ij}(|\vec{z}|) \\ & \exp \left[-\frac{(\vec{z} - (\vec{a}_i - \vec{a}_j))^2}{2T(G_{ii} + G_{jj} - 2G_{ij})} \right] \left[1 - \frac{(\vec{z} - (\vec{a}_i - \vec{a}_j))^2}{DT(G_{ii} + G_{jj} - 2G_{ij})} \right] \end{aligned} \quad (62)$$

and for $i = j$,

$$(G^{-1})_{ii} = - \sum_{j \neq i} (G^{-1})_{ij}. \quad (63)$$

These equations are obviously complicated, and the only way to solve them for some arbitrary set of two-body potentials V_{ij} would be numerically on the computer. (They simplify considerably and can be dealt with analytically in the case when the potentials V_{ij} are identical, corresponding to an ordinary homopolymer [6] [3].) At a

given temperature, there may well be more than one solution to these equations, which would correspond to the different possible metastable states of the heteropolymer. The advantage of this approach is that it naturally accounts for thermal fluctuations while giving a great deal of useful information: given the input of the temperature and the two-body potentials, one gets as output the positions and correlated fluctuations of all the monomers. It would certainly be interesting if someone went ahead and used these equations for a relatively small system, and then compared the results to a more conventional, but time-consuming, Monte Carlo simulation.

The protein is our first example of a system with "quenched disorder." The disorder in this case simply comes from the two-body potentials, which depend in some complicated way on the precise amino-acids that the chain is made out of. The disorder is "quenched" in the sense that the two-body potentials are fixed once and for all for any given protein. Of course, different proteins will be made out of different amino acids, and will therefore have different two-body potentials and ultimately different shapes, but for a given protein, the potentials are quenched. The concept of "quenched disorder" is easy to understand, but it is important to continue to learn how to deal with it on a technical level, as it is an intrinsic aspect of many physical systems. For our next example, we will first apply our variational approach, and then introduce a new and potentially even more powerful technique.

The next system that we will consider for which quenched disorder is important is the *Ising spin glass*. Imagine that we have some lattice (for concreteness we will restrict ourselves to D -dimensional hypercubic lattices like the linear, square, or cubic lattice) of N points, and on each point of the lattice we put an Ising spin which can point up or down. In the first lecture, we only considered Ising spin systems in which each spin was independent of every other spin, but let us now consider what happens if each spin influences its nearest neighbors. In a *ferromagnet* each spin will tend to make its nearest neighbors point in the same direction that it is pointing. A

commonly used Hamiltonian for the ferromagnet is

$$H = -J \sum_{(ij)} S_i S_j \quad (64)$$

where $J > 0$ and the (ij) notation means that the sum is over nearest neighbors in the lattice. This Hamiltonian clearly favors configurations in which all the spins point in the same direction. In an *antiferromagnet*, each spin will tend to make its nearest neighbor point in the direction opposite to its own. The same Hamiltonian as the one used for the ferromagnet will also describe an antiferromagnet if $J < 0$. A *spin glass* is a system in which the interaction between any pair of nearest neighbors is fixed, and is randomly chosen to be either ferromagnetic or antiferromagnetic. The canonical Hamiltonian for a spin glass is

$$H = - \sum_{(ij)} J_{ij} S_i S_j \quad (65)$$

In practice, if one wants to make a computer simulation of a spin glass for example, one chooses each J_{ij} to be equal to $+1$ or -1 with equal probability, or chooses the J_{ij} 's from some other probability distribution. One very popular probability distribution is the Gaussian one, for which

$$p(J_{ij}) = \frac{e^{-(J_{ij}-J_0)^2/(2\tilde{J}^2)}}{\sqrt{(2\pi\tilde{J}^2)}} \quad (66)$$

where J_0 is the average value of a bond and \tilde{J} is the standard deviation in the bond strengths. Once the J_{ij} 's are chosen from whichever probability distribution being used, they are quenched, and cannot be changed. They play a role analogous to that of the V_{ij} 's in the protein Hamiltonian, while the S_i 's play a role analogous to the r_i 's as the degrees of freedom in the problem. The ferromagnetic and antiferromagnetic Hamiltonians are obviously just special cases of the more general spin glass Hamiltonian, although in practice, a system is only called a "spin glass" if the J_{ij} 's are chosen

at random and from a probability distribution which contains both positive and negative J 's. (In 1975, Sherrington and Kirkpatrick introduced their famous model of a spin glass with the above Hamiltonian on the very special lattice in which each spin is a nearest neighbor of every other spin [19]. A spin glass on such a lattice turns out to have thermodynamic properties that are identical to those of a spin glass on a $D = \infty$ -dimensional hypercubic lattice.)

Faced with this Hamiltonian, I would hope that you would first consider trying the variational approach on it. Let's see how that would work. We are interested in computing

$$F = -T \ln \text{Tr} \exp(-H/T) \quad (67)$$

where Tr is a short-hand notation for a sum over all possible states of the system:

$$\text{Tr} \equiv \sum_{S_1=\pm 1} \sum_{S_2=\pm 1} \dots \sum_{S_N=\pm 1} \quad (68)$$

We know that we can make exact computations with a trial Hamiltonian consisting of non-interacting spins, so we take

$$H_0 = - \sum_{i=1}^N h_i S_i \quad (69)$$

where the h_i variables are now variational parameters that we will try to optimize.

The trial free energy is as usual

$$\tilde{F} \equiv F_0 + \langle H - H_0 \rangle_0 \quad (70)$$

where

$$F_0 = -T \ln Z_0 \equiv -T \ln \text{Tr} \exp(-H_0/T) \quad (71)$$

and the expectation value $\langle X \rangle_0$ is taken with respect to the trial Hamiltonian H_0 :

$$\langle X \rangle_0 \equiv \frac{1}{Z_0} \text{Tr} X \exp(-H_0/T) \quad (72)$$

We did the trivial computation of F_0 last lecture:

$$F_0 = -T \sum_i \ln \cosh(h_i/T) \quad (73)$$

The other pieces of \tilde{F} are almost as simple to compute:

$$\langle -H_0 \rangle_0 = \sum_i h_i \langle S_i \rangle_0 = \sum_i h_i \tanh(h_i/T) \quad (74)$$

$$\begin{aligned} \langle H \rangle_0 &= - \sum_{(ij)} J_{ij} \langle S_i S_j \rangle_0 \\ &= - \sum_{(ij)} J_{ij} \langle S_i \rangle_0 \langle S_j \rangle_0 = - \sum_{(ij)} J_{ij} \tanh(h_i/T) \tanh(h_j/T) \end{aligned} \quad (75)$$

We have used the fact that the spins are independent in the trial Hamiltonian to factorize the correlation function $\langle S_i S_j \rangle_0$.

Putting it all together, we find

$$\tilde{F} = -T \sum_i \ln \cosh(h_i/T) - \sum_{(ij)} J_{ij} \tanh(h_i/T) \tanh(h_j/T) + \sum_i h_i \tanh(h_i/T) \quad (76)$$

Minimizing the trial free energy with respect to h_i , we find

$$\begin{aligned} -\tanh(h_i/T) - \frac{1}{T} \sum_{j(i)} J_{ij} \tanh(h_j/T) (1 - \tanh^2(h_i/T)) \\ + \tanh(h_i/T) + \frac{1}{T} h_i (1 - \tanh^2(h_i/T)) = 0 \end{aligned} \quad (77)$$

or

$$h_i = \sum_{j(i)} J_{ij} \tanh(h_j/T). \quad (78)$$

where the notation $j(i)$ means all spins j neighboring spin i .

Notice that these self-consistent equations are considerably simpler than the corresponding equations we derived for the protein problem. We can rewrite them in a

different and slightly more conventional way by changing variables to the local magnetizations $m_i \equiv \langle S_i \rangle_0 = \tanh(h_i/T)$. In terms of the magnetizations, we have the trial free energy

$$\tilde{F} = -T \sum_i \left[\frac{1+m_i}{2} \ln \left(\frac{1+m_i}{2} \right) + \frac{1-m_i}{2} \ln \left(\frac{1-m_i}{2} \right) \right] + \sum_{(ij)} J_{ij} m_i m_j \quad (79)$$

(The first term on the right is the entropy, while the second term is the internal energy of the system in this approximation.) In these variables, the self-consistent equations are

$$m_i = \tanh \left(\sum_{j(i)} \frac{J_{ij} m_j}{T} \right) \quad (80)$$

Let us now consider the special case of the ferromagnet where all the J_{ij} are equal. We will choose the particular scaling $J_{ij} = 1/(2D)$, where D is the dimension of our hypercubic lattice, in order that the ground state energy density (when all spins point in the same direction) will be $E/N = -1$ irrespective of the dimension. Since all the J_{ij} 's are equal, we expect the magnetization at each site to be equal, since there is nothing to distinguish one site from another. Setting $m_i = m$, we find the famous *mean field* equation for the magnetization of an Ising ferromagnet:

$$m = \tanh(m/T) \quad (81)$$

According to this "variational," or "mean-field" approximation, the ferromagnet will have a transition at $T = 1$. For $T > 1$, $m = 0$, but for $T < 1$, the magnetization is non-zero, with a magnitude approaching 1 at $T = 0$. We shall eventually see that the mean-field approximation becomes exact for the ferromagnet when D approaches infinity.

To demonstrate this fact, we will need to develop a new and different technique for calculating the free energy. This technique is conceptually very simple—it is based on the idea of expanding the magnetization-dependent free energy in powers of the

inverse temperature. We shall see that the form for the trial free energy given in equation (79) actually corresponds to just the first two terms in such an expansion. Thus, by computing the higher-order terms in this expansion, we can systematically approach the true free energy. My collaborator on the work that I am about to describe was Antoine Georges at the Ecole Normale Supérieure in Paris.

The starting point of this technique is the *magnetization-dependent free energy*. As I have defined it so far, the free energy is just the logarithm of the partition function, and the magnetization will have some equilibrium value at any temperature. The magnetization has not been a free variable—we have been given the temperature and we have computed the magnetization. One can turn the magnetization into a free variable and define a magnetization-dependent free energy by adding to the physical system a set of external auxiliary fields which are used to insure that the magnetizations are at their desired values. Of course, when the magnetizations are at their *equilibrium* values, no auxiliary fields will be necessary. Let's see how this works in equations, using our spin glass Hamiltonian. The magnetization-dependent free energy is

$$-\beta F(\beta, m_i) = \ln \text{Tr} \exp \left(\beta \sum_{(ij)} J_{ij} S_i S_j + \sum_i \lambda_i(\beta) (S_i - m_i) \right) \quad (82)$$

where $\beta \equiv 1/T$ is the inverse temperature. The $\lambda(\beta)$ are our auxiliary fields (or Lagrange multipliers). Note that they depend explicitly on the inverse temperature, which is just a reflection of the important (and obvious) fact that the fields necessary to fix a certain set of magnetizations will change as the temperature changes. As usual, the magnetizations m_i are defined as $\langle S_i \rangle$, where the expectation value is taken with respect to an effective Hamiltonian which is the sum of the original Hamiltonian and the auxiliary fields: if X is some observable, then

$$\langle X \rangle \equiv \frac{\text{Tr} X \exp(\beta \sum_{(ij)} J_{ij} S_i S_j + \sum_i \lambda_i(\beta) (S_i - m_i))}{\text{Tr} \exp(\beta \sum_{(ij)} J_{ij} S_i S_j + \sum_i \lambda_i(\beta) (S_i - m_i))} \quad (83)$$

Eventually, we are going to minimize the free energy with respect to the magnetizations (set $\partial F/\partial m_i = 0$). You can work out that this condition, when combined with the constraint that $m_i = \langle S_i \rangle$, ensures that the auxiliary fields $\lambda_i(\beta) = 0$, precisely as they should at equilibrium.

We are going to expand $-\beta F(\beta, m_i)$ around $\beta = 0$ using a Taylor expansion. You can already see that this trick will be useful because at $\beta = 0$, the spins will be entirely controlled by their corresponding auxiliary fields, and we will thus have again reduced our problem to one of independent spins. Since m_i is fixed equal to $\langle S_i \rangle$ for any inverse temperature β , it is in particular equal to $\langle S_i \rangle$ when $\beta = 0$, which gives us the important relation

$$m_i = \langle S_i \rangle_{\beta=0} = \frac{\text{Tr} S_i \exp(\lambda_i(0) S_i)}{\text{Tr} \exp(\lambda_i(0) S_i)} = \tanh(\lambda_i(0)). \quad (84)$$

We now expand the $-\beta F(\beta, m_i)$ around $\beta = 0$ using a Taylor expansion:

$$-\beta F(\beta) = -(\beta F)_{\beta=0} - \frac{\partial(\beta F)}{\partial \beta} \Big|_{\beta=0} \beta - \frac{\partial^2(\beta F)}{\partial \beta^2} \Big|_{\beta=0} \frac{\beta^2}{2} - \dots \quad (85)$$

where we have temporarily suppressed the dependence of F on m_i . From the definition of $-\beta F(\beta, m_i)$ given in equation (82), we find that

$$-\beta F(\beta, m_i)_{\beta=0} = \sum_i \ln[\cosh(\lambda_i(0))] - \lambda_i(0) m_i. \quad (86)$$

At this point, we can choose to work with either the variables m_i or the variables $\lambda_i(0)$, which are directly related to the m_i through equation (84). We will choose to eliminate the $\lambda_i(0)$ (note that the formal manipulations are very similar to some of those we did previously when using the variational approach, but the meanings of our variables are somewhat different), and thereby recover

$$-\beta F(\beta, m_i)_{\beta=0} = - \sum_i \left[\frac{1+m_i}{2} \ln \left(\frac{1+m_i}{2} \right) + \frac{1-m_i}{2} \ln \left(\frac{1-m_i}{2} \right) \right] \quad (87)$$

which is the entropy of non-interacting Ising spins constrained to have magnetizations m_i . (Compare with the formula

$$S = \sum_{states} -p_{state} \ln(p_{state}) \quad (88)$$

from last lecture.) Considering next the first derivative in equation (85), we find that

$$-\beta \frac{\partial(\beta F)}{\partial \beta} \Big|_{\beta=0} = \beta \left\langle \sum_{(ij)} J_{ij} S_i S_j \right\rangle_{\beta=0} + \beta \langle S_i - m_i \rangle_{\beta=0} \frac{\partial \lambda_i}{\partial \beta} \Big|_{\beta=0}. \quad (89)$$

At $\beta = 0$, the spin-spin correlation functions factorize so we find that

$$-\beta \frac{\partial(\beta F)}{\partial \beta} \Big|_{\beta=0} = \beta \sum_{(ij)} J_{ij} m_i m_j. \quad (90)$$

This is of course the "variational" internal energy, so we see that as claimed, the first two terms in our expansion give the variational trial free energy.

Naturally, we can continue our expansion, and to arbitrarily high order. If you are interested in some formal details and tricks which make the computation easier, you can refer to reference [10]. To order β^4 , one finds that

$$\begin{aligned} -\beta F(\beta, m_i) = & - \sum_i \left[\frac{1+m_i}{2} \ln \left(\frac{1+m_i}{2} \right) + \frac{1-m_i}{2} \ln \left(\frac{1-m_i}{2} \right) \right] + \beta \sum_{(ij)} J_{ij} m_i m_j \\ & + \frac{\beta^2}{2} \sum_{(ij)} J_{ij}^2 (1-m_i^2)(1-m_j^2) + \frac{2\beta^3}{3} \sum_{(ij)} J_{ij}^3 m_i (1-m_i^2) m_j (1-m_j^2) \\ & \quad + \beta^3 \sum_{(ijk)} J_{ij} J_{jk} J_{ki} (1-m_i^2)(1-m_j^2)(1-m_k^2) \\ & - \frac{\beta^4}{12} \sum_{(ij)} J_{ij}^4 (1-m_i^2)(1-m_j^2)(1+3m_i^2+3m_j^2-15m_i^2 m_j^2) \\ & \quad + 2\beta^4 \sum_{(ijk)} J_{ij}^2 J_{jk} J_{ki} m_i (1-m_i^2) m_j (1-m_j^2)(1-m_k^2) \\ & + \beta^4 \sum_{(ijkl)} J_{ij} J_{jk} J_{kl} J_{li} (1-m_i^2)(1-m_j^2)(1-m_k^2)(1-m_l^2) + \dots \end{aligned} \quad (91)$$

where the notation (ij) , (ijk) , or $(ijkl)$ means that one should sum over all distinct pairs, triplets, or quadruplets of spins.

For the ferromagnet on a hypercubic lattice, all these terms can be re-organized according to their power in $1/D$. It is easy to show that only the zeroth and first order term contribute in the limit $D \rightarrow \infty$, and to generate $1/D$ expansions for all the thermodynamic quantities, including the magnetization [10]. In 1977, Thouless, Anderson and Palmer (TAP) pointed out that the "mean-field" theory for the Sherrington-Kirkpatrick spin-glass model should also include the second-order term (in β) [20]. Unfortunately, solving the N equations obtained from minimizing the TAP free energy is still no easy task for the spin glass problem, although they were nevertheless able to make a number of interesting deductions based on their equations. The interested reader is referred directly to their paper for the details. In the next lecture, we will finally begin studying an even more powerful technique for dealing with quenched disorder, the famous replica method.

III: Averaging over Disorder and the Replica Method

In this lecture, we will introduce a new subject, *averaging over the disorder* of a physical system, and a technique to do it, the *replica method*. The idea of averaging over disorder may have already occurred to you when you saw the results of the variational method for the spin glass Hamiltonian. For N spins, there were N self-consistent equations to solve! This compared very unfavorably to the single self-consistent equation that we needed to solve in the ferromagnetic mean-field theory. Of course, if one is interested in some very specific system; for example the shape of a specific protein, then it makes sense that one will have to solve a lot of equations to get the answers—if you want a lot of detailed information, you need to do a lot of work. But if instead one is satisfied to know the *average* value of thermodynamic quantities for some typical system with quenched disorder chosen from some probability distribution, then it makes sense that one can simplify the problem. Sam Edwards and Philip Anderson were the first to attempt to compute such average quantities using the replica method for the spin glass in 1975 [7], and Giorgio Parisi first gave the correct solution of the replica mean field theory for spin glasses in 1981 [17]. We will introduce the replica method on models which are somewhat simpler than the spin glass, so that the ideas will be clearer. In fact, we will begin with a model which is sufficiently simple that the average over disorder can easily be done without replicas, so that we can check that the replica method does indeed give the correct answer.

Suppose that we have a single particle governed by the Hamiltonian

$$H = \frac{r^2}{2} - f r \quad (92)$$

where f is a quenched random force field. Or rather, we actually have an ensemble of such physical systems, each with a different value of f . A single exemplar system from the ensemble is called a *sample*. Suppose that the samples are assigned values

of f chosen from the probability distribution

$$p(f) = \frac{e^{-f^2/2f_0^2}}{\sqrt{2\pi f_0^2}} \quad (93)$$

For this ensemble of systems, we can compute thermodynamic quantities like the free energy F_f for any particular sample with a certain value of f , but what is more, we can compute the free energy \overline{F} averaged over the entire ensemble, with each value of f weighted by its probability:

$$\overline{F} \equiv \int_{-\infty}^{\infty} df p(f) F_f \quad (94)$$

Let's see how this works. We have

$$H = \frac{r^2}{2} - f r = \frac{1}{2}(r - f)^2 - \frac{f^2}{2} \quad (95)$$

so we see that if the force field is f , the particle will actually be in a quadratic well centered at $r = f$ with minimum $H = -f^2/2$. We have

$$Z_f = \int_{-\infty}^{\infty} dr e^{-\frac{1}{2T}[(r-f)^2 - f^2]} \quad (96)$$

$$Z_f = e^{f^2/2T} \int_{-\infty}^{\infty} dr e^{-\frac{1}{2T}(r-f)^2} = e^{f^2/2T} \sqrt{2\pi T}. \quad (97)$$

The free energy F_f for a sample with force field f is

$$F_f = -T \ln Z_f = -\frac{f^2}{2} - \frac{T}{2} \ln(2\pi T). \quad (98)$$

This is as we would expect: the free energy is shifted from the $f = 0$ free energy by the same amount as the Hamiltonian was shifted. Now we can compute the desired average free energy:

$$\overline{F} = \int_{-\infty}^{\infty} \frac{df}{\sqrt{2\pi f_0^2}} e^{-f^2/2f_0^2} \left(-\frac{f^2}{2} - \frac{T}{2} \ln(2\pi T) \right) \quad (99)$$

$$\overline{F} = -\frac{f_0^2}{2} - \frac{T}{2} \ln(2\pi T) \quad (100)$$

We can compute other disorder-averaged quantities. For example, the thermal average of the position for a given force field f is

$$\langle r \rangle_f = \frac{\int_{-\infty}^{\infty} dr \, r \, e^{-\frac{1}{2T}(r^2 - 2fr)}}{\int_{-\infty}^{\infty} dr \, e^{-\frac{1}{2T}(r^2 - 2fr)}} = f \quad (101)$$

Again this makes sense given that our quadratic well is centered at $r = f$. If we now do the disorder average, we find that the disorder average of the thermal average of the position is zero:

$$\overline{\langle r \rangle} \equiv \int_{-\infty}^{\infty} df \, p(f) \langle r \rangle_f = 0. \quad (102)$$

This result is the consequence of the fact that our probability distribution for f is even so that the samples with negative and positive force fields cancel each other out. Nevertheless, the disorder-average of the *square* of the average position will be non-zero:

$$\overline{\langle r \rangle^2} = \overline{f^2} = \int_{-\infty}^{\infty} \frac{df}{\sqrt{2\pi f_0^2}} e^{-f^2/2f_0^2} f^2 = f_0^2 \quad (103)$$

You can easily work out other averages; for example the disorder average of the thermal average of the squared position is

$$\overline{\langle r^2 \rangle} = T + f_0^2 \quad (104)$$

Note that this correlation function, which measures how much the particles fluctuate around the origin, actually has two contributions. The disorder contribution, $\overline{\langle r \rangle^2} = f_0^2$, does not depend on the temperature, and is proportional to the strength of the disorder. On the other hand, the thermal part of the fluctuation, given by

$$\overline{\langle (r - \langle r \rangle)^2 \rangle} = \overline{\langle r^2 \rangle} - \overline{\langle r \rangle^2} = T \quad (105)$$

does not depend on the disorder and is proportional to the temperature. This thermal part has the form of a so-called *connected* correlation function; note how it automatically subtracts away the shift in the average position caused by the random force field.

We will now re-derive these results using the replica method. For this problem, using replicas is certainly over-kill, but of course the point of the method is that it will work for many other problems where a direct computation is impossible. Our direct computation relied on the fact that we could compute the free energy for any particular sample. Of course, we have seen previously that for other problems, computing the free energy of a particular sample meant solving a large number of self-consistent equations. The idea of the replica method is to reverse the order of the computations—we want to average over disorder *first*, leaving any additional computations for later. The additional computations may still be possible, or even actually simpler, if we do them in this reverse order. In our example, we have

$$\overline{F} = -T \int_{-\infty}^{\infty} \frac{df}{\sqrt{2\pi f_0^2}} e^{-f^2/2f_0^2} \ln \int_{-\infty}^{\infty} dr e^{-\frac{1}{T}(\frac{r^2}{2} - fr)} \quad (106)$$

Instead of doing the r integration first, we want to start with the disorder average represented by the f integration. This would be possible if the logarithm were not there—the f integral would then be a soluble Gaussian. The replica method is based on the mathematical identity

$$\ln Z = \lim_{n \rightarrow 0} \frac{Z^n - 1}{n} \quad (107)$$

which enables us to pull the logarithm out of the way. Let us rewrite our average free energy as

$$\overline{F} = -T \int_{-\infty}^{\infty} df p(f) \ln Z_f \quad (108)$$

We use the mathematical identity

$$x = \lim_{n \rightarrow 0} \frac{1}{n} \ln(1 + nx) \quad (109)$$

to rewrite this as

$$\bar{F} = -T \lim_{n \rightarrow 0} \frac{1}{n} \ln \left(1 + n \int_{-\infty}^{\infty} df p(f) \ln Z_f \right). \quad (110)$$

Using the fact that $\int_{-\infty}^{\infty} df p(f) = 1$, we have

$$\bar{F} = -T \lim_{n \rightarrow 0} \frac{1}{n} \ln \int_{-\infty}^{\infty} df p(f) (1 + n \ln Z_f) \quad (111)$$

Finally, using the identity (107), we have

$$\bar{F} = -T \lim_{n \rightarrow 0} \frac{1}{n} \ln \int_{-\infty}^{\infty} df p(f) Z_f^n \quad (112)$$

which is a form that we can work with; we now will be able to perform the f integration. One way to think of the term Z_f^n is to imagine a new physical system consisting of n identical replicas of the old system; then Z_f^n is just the partition function of the new system. If we label the replicas by the index a , where a can run from 1 to n , then

$$Z_f^n = \int_{-\infty}^{\infty} \prod_{a=1}^n dr_a \exp \left(-\frac{1}{T} \sum_{a=1}^n \left(\frac{r_a^2}{2} - f r_a \right) \right) \quad (113)$$

Finally, we have transformed our original equation (106) for \bar{F} into the mathematically equivalent expression

$$\bar{F} = -T \lim_{n \rightarrow 0} \frac{1}{n} \ln \int_{-\infty}^{\infty} \frac{df}{\sqrt{2\pi f_0^2}} e^{-f^2/2f_0^2} \int_{-\infty}^{\infty} \prod_{a=1}^n dr_a \exp \left(-\frac{1}{T} \sum_{a=1}^n \left(\frac{r_a^2}{2} - f r_a \right) \right) \quad (114)$$

The advantage of this horrible-looking expression is that we can now complete the square, perform the Gaussian f integration, and arrive at the formula

$$\bar{F} = -T \lim_{n \rightarrow 0} \frac{1}{n} \ln \int_{-\infty}^{\infty} \prod_{a=1}^n dr_a \exp \left(-\frac{1}{T} \left(\sum_{a=1}^n \frac{r_a^2}{2} - \frac{f_0^2}{2T} \sum_{a=1}^n \sum_{b=1}^n r_a r_b \right) \right) \quad (115)$$

By averaging over disorder, we have converted our original problem into the mathematically equivalent problem of a system of n particles with no disorder, and interacting according to the effective Hamiltonian

$$H_{eff} = \sum_{a=1}^n \frac{r_a^2}{2} - \frac{f_0^2}{2T} \sum_{a=1}^n \sum_{b=1}^n r_a r_b \quad (116)$$

Now we have to do the "additional computations" that we have postponed—that is, we have to compute the free energy for this new system. Fortunately, in this case the integration over the r_a variables is Gaussian so it can be performed exactly. (For other problems this next step can only be done approximately, for example by a variational approximation; we shall see how this works in the next lecture.) We write the effective Hamiltonian in the form

$$H_{eff} = \frac{1}{2} \sum_{a=1}^n \sum_{b=1}^n (G^{-1})_{ab} r_a r_b \quad (117)$$

where $(G^{-1})_{aa} = 1 - f_0^2/T$ and $(G^{-1})_{a \neq b} = -f_0^2/T$. Doing the Gaussian integral, we find

$$\bar{F} = -T \lim_{n \rightarrow 0} \frac{1}{n} \ln(\sqrt{(2\pi T)^n \det G}) = -\frac{T}{2} \ln(2\pi T) - \frac{T}{2} \lim_{n \rightarrow 0} \frac{1}{n} \text{Tr} \ln G \quad (118)$$

Now we have to take to logarithm of the determinant of our G matrix, or equivalently the trace of the logarithm of that matrix. At this point, it is worthwhile to make a digression to study some needed matrix algebra. We are interested in n by n matrices of the form

$$\begin{pmatrix} \tilde{a} & a & a & a \\ a & \tilde{a} & a & a \\ a & a & \tilde{a} & a \\ a & a & a & \tilde{a} \end{pmatrix}$$

We call these matrices *replica-symmetric matrices*, and we will adhere to the convention that the diagonal elements of the replica-symmetric matrix A_{ab} will be denoted by \tilde{a} , while the off-diagonal elements will be denoted by a . We want to derive rules for multiplying such matrices which are correct for arbitrary n , so that we can take the $n \rightarrow 0$ limit. We will then be able to formally deal with 0×0 matrices!

If we multiply two replica-symmetric matrices A_{ab} and B_{ab} , the result is a new replica symmetric matrix C_{ab} with

$$\tilde{c} = \tilde{a}\tilde{b} + (n - 1)ab \quad (119)$$

and

$$c = \tilde{a}b + a\tilde{b} + (n - 2)ab. \quad (120)$$

Notice that we derive these formulas by thinking of n as an integer, but that we can then extend their validity to all real n . By requiring that $\tilde{c} = 1$ and $c = 0$, we get the conditions which must be satisfied if A_{ab} is the inverse matrix of B_{ab} . In the $n \rightarrow 0$ limit, these conditions are

$$\tilde{a} - a = \frac{1}{\tilde{b} - b}, \quad a = \frac{-b}{(\tilde{b} - b)^2}, \quad \tilde{a} = \frac{\tilde{b} - 2b}{(\tilde{b} - b)^2} \quad (121)$$

Applying these formulae to our G_{ab} matrix, of which we only knew the inverse until now, we find that $\tilde{g} = 1 + f_0^2/T$ and $g = f_0^2/T$.

The trace of the logarithm of a replica symmetric matrix can be worked out using the Taylor expansion

$$\ln(1 + X) = X - \frac{X^2}{2} + \frac{X^3}{3} - \frac{X^4}{4} + \dots \quad (122)$$

in its matrix form. We leave it as an exercise for the reader to work out that for the replica symmetric matrix A

$$\lim_{n \rightarrow 0} \frac{1}{n} \text{Tr} \ln A = \ln(\tilde{a} - a) + \frac{a}{\tilde{a} - a} \quad (123)$$

Using this result in equation (118), we finally obtain

$$\overline{F} = -\frac{T}{2} \ln(2\pi T) - \frac{f_0^2}{2} \quad (124)$$

in agreement with our previous direct computation.

Other disorder averages can also be worked out using the replica method. Again, one does the average over disorder *first*, leaving any additional computations until the end. In this way, one can show that

$$\overline{\langle r^2 \rangle} = \left\langle \sum_{a=1}^n r_a^2 \right\rangle_{eff} \quad (125)$$

where the "effective" expectation value is taken with respect to the replica Hamiltonian: we define an effective partition function

$$Z_{eff} \equiv \int_{-\infty}^{\infty} \prod_{a=1}^n dr_a \exp\left(-\frac{H_{eff}}{T}\right) \quad (126)$$

and the effective expectation value of some observable X is then

$$\langle X \rangle_{eff} \equiv \lim_{n \rightarrow 0} \frac{1}{n} \frac{1}{Z_{eff}} \int_{-\infty}^{\infty} \prod_{a=1}^n dr_a X \exp\left(-\frac{H_{eff}}{T}\right) \quad (127)$$

Completing the computation, we find

$$\overline{\langle r^2 \rangle} = \lim_{n \rightarrow 0} \frac{1}{n} \sum_{a=1}^n TG_{aa} = T\tilde{g} = T + f_0^2 \quad (128)$$

This result is very important—it tells us that the diagonal element of our Green's function matrix in replica space is proportional to the correlation function measuring the combined fluctuations caused by thermal and disorder effects. Similarly, we can show that

$$\begin{aligned} \overline{\langle r^2 \rangle - \langle r \rangle^2} &= \left\langle \sum_{a=1}^n \sum_{b=1}^n r_a r_b \right\rangle_{eff} = \lim_{n \rightarrow 0} \frac{1}{n} (n TG_{aa} + n(n-1) TG_{a \neq b}) \\ &= T(\tilde{g} - g) = T. \end{aligned} \quad (129)$$

This tells us that the purely thermal fluctuations given by the "connected" correlation function are proportional the difference of the diagonal and off-diagonal elements of the Green's function matrix.

We can summarize what we have learned by saying that an effective replica Hamiltonian of the form

$$H_{eff} = \frac{1}{2} \sum_{a=1}^n \sum_{b=1}^n (G^{-1})_{ab} r_a r_b \quad (130)$$

with a replica-symmetric G matrix represents an exact description of an ensemble of physical systems, each one of which is a particle in a quadratic well with identical thermal fluctuations equal to $T(G_{aa} - G_{a \neq b})$, but with the center of the well assigned

a different position in each sample, with fluctuations in the position of the well equal to $TG_{a \neq b}$. The total combined fluctuations in the position averaged over temperature and from sample to sample will be equal to TG_{aa} .

We will now move on to a much more challenging example of averaging over quenched disorder, for which a direct computation is impossible. Our first example was not really so "disordered," as each sample was still just a single particle in a perfectly quadratic well. The disorder was just sample-to-sample disorder, not disorder within a given sample. Much more interesting would be an ensemble of physical systems such that even within a given sample, there exists intrinsic disorder. This is really what most physicists have in mind when they think of glasses. To be specific, let us imagine an ensemble of systems, each one again consisting of a single particle in a potential. The Hamiltonian for the particle will again consist of a quadratic piece plus a random term:

$$H = \frac{r^2}{2} + V(r), \quad (131)$$

but now the random term $V(r)$ represents a whole random potential landscape, looking something like the profile of a one-dimensional mountain landscape, or perhaps the graphical history of a (purely random) stock market. A particle in such a potential will certainly feel intrinsic disorder within the sample, and if each sample has a *different* random potential, we will still have sample-to-sample disorder. Before we can try to solve for the average thermodynamics of such an ensemble of systems (which we will eventually do), we first have to understand how one can even describe such random potentials in a mathematically precise way, which is the problem we turn to now.

One convenient (and realistic) way to make an ensemble of random potentials is to imagine a potential landscape which is actually very slowly fluctuating according to some Hamiltonian at the effective temperature T_{eff} , so that the potential chosen

for a particular sample is just a snapshot of the fluctuating potential at some time. We assume that the fluctuations of the particle on the potential are very much faster than the fluctuations of the potential, so that on time scales relevant to the particle the potential still appears quenched. Technically, this means that we do the thermal average over the Hamiltonian of the particle first, and only afterwards do the disorder average, which is interpreted as a thermal average over the slow Hamiltonian of the potential. One reasonable "slow Hamiltonian" is

$$H_{slow} = \frac{T_{eff}}{2} \int_{-\infty}^{\infty} dr \left(\frac{\partial V}{\partial r} \right)^2 \quad (132)$$

Such a Hamiltonian favors potential landscapes for which near-by points are correlated. A more general quadratic slow Hamiltonian is

$$H_{slow} = \frac{T_{eff}}{2} \int_{-\infty}^{\infty} dr \int_{-\infty}^{\infty} dr' (K^{-1})_{rr'} V(r)V(r') \quad (133)$$

This slow Hamiltonian will give us a Gaussian probability distribution

$$p(V(r)) = \frac{1}{Z_0} \exp \left(-\frac{1}{2} \int_{-\infty}^{\infty} dr \int_{-\infty}^{\infty} dr' (K^{-1})_{rr'} V(r)V(r') \right) \quad (134)$$

where Z_0 is a normalization constant. (To make sense of these expressions, it may help you to think of discretizing space so that the number of points where the particle can sit is finite. Then, for example, $K_{rr'}$ will become a finite matrix. At the end, one can go back to the continuum limit.) One can take this probability distribution as a starting definition of a random potential—the description in terms of a slow Hamiltonian was merely to help give some qualitative understanding. This Gaussian probability distribution has for its first two moments

$$\overline{V(r)} = 0 \quad , \quad \overline{V(r)V(r')} = K_{rr'} \quad (135)$$

Clearly, the specifications of a Gaussian probability distribution by its slow Hamiltonian or by its first two moments are equivalent, as one can switch from one to the other by simply inverting the K matrix.

Normally, $K_{rr'}$ is just a function of $|r - r'|$:

$$K_{rr'} = K(|r - r'|). \quad (136)$$

In that case,

$$\overline{(V(r) - V(r'))^2} = 2 (K(0) - K(|r - r'|)) \quad (137)$$

It is reasonable that $K(|r|)$ should monotonically decrease from its value at $|r| = 0$, as that implies from equation (137) that the closer two points are, the more closely their potentials are correlated. One very reasonable and technically convenient form that for $K(|r|)$ is a Gaussian decay:

$$K(|r - r'|) = W \exp\left(-\frac{(r - r')^2}{2\Delta^2}\right). \quad (138)$$

Such a form describes a random potential with typical magnitude W and correlation length Δ . In the "mountain landscape" analogy, W corresponds to the typical height of the mountain peaks, while Δ is the typical distance between mountain peaks. Another very popular form is a linear decay:

$$K(|r - r'|) = W - f|r - r'|. \quad (139)$$

In contrast to the Gaussian decay, the linear form has no characteristic spatial or energy scale. Instead the typical squared difference in potential will grow linearly with distance at all scales; such a form corresponds to a Brownian random walk for the potential.

We will now begin the computation of the average free energy for an ensemble of systems with a Gaussian random potential by averaging over the disorder with the replica method. In this case, we have no hope of succeeding with the "direct approach" of computing the free energy for an arbitrary sample, and then averaging over the disorder. We must try to average over the disorder first, and leave any

additional computations for later. I will first tell you the result of the average over disorder, and then give a derivation of this result.

The result is: if we have an ensemble of systems with Hamiltonian

$$H = \frac{r^2}{2} + V(r) \quad (140)$$

where $V(r)$ is a Gaussian random potential with first two moments $\overline{V(r)} = 0$ and $\overline{V(r)V(r')} = K_{rr'}$, then

$$\overline{F} = -T \lim_{n \rightarrow 0} \frac{1}{n} \ln \int_{-\infty}^{\infty} \prod_{a=1}^n dr_a \exp\left(-\frac{H_{eff}}{T}\right) \quad (141)$$

where

$$H_{eff} = \frac{1}{2} \sum_{a=1}^n r_a^2 - \frac{1}{2T} \sum_{a=1}^n \sum_{b=1}^n K_{r_a r_b} \quad (142)$$

The derivation: we have (all the initial steps are the same as our previous example)

$$\overline{F} = -T \lim_{n \rightarrow 0} \frac{1}{n} \ln \overline{Z_V^n} \quad (143)$$

where Z_V represents the partition function of the system with a given potential $V(r)$.

We take the n th power of the partition function by replicating the system n times:

$$Z_V^n = \int_{-\infty}^{\infty} \prod_{a=1}^n dr_a \exp\left(-\frac{1}{T} \sum_{a=1}^n \left(\frac{r_a^2}{2} + V(r_a)\right)\right) \quad (144)$$

Since we have n identical replicas of the system with potentials $V(r_a)$, the correlation function of the disorder will not depend on which two replicas we choose, so that

$$\overline{V(r_a)} = 0 \quad , \quad \overline{V(r_a)V(r'_b)} = K_{r_a r'_b} \quad (145)$$

and we find

$$\begin{aligned} \overline{F} &= -T \lim_{n \rightarrow 0} \frac{1}{n} \ln \frac{1}{Z_0^n} \int DV(r) \exp\left(-\frac{1}{2} \sum_{a=1}^n \sum_{b=1}^n \int_{-\infty}^{\infty} dr_a dr'_b (K^{-1})_{r_a r'_b} V(r_a)V(r'_b)\right) \\ &\quad \int_{-\infty}^{\infty} \prod_{a=1}^n dr_a \exp\left(-\frac{1}{T} \sum_{a=1}^n \left(\frac{r_a^2}{2} + V(r_a)\right)\right) \end{aligned} \quad (146)$$

where the notation $DV(r)$ represents a *functional integral* over the function $V(r)$. (Again, if you have trouble with this, just imagine that space is discretized into a finite number of points. The functional integral then becomes a multiple integral over the values of the potential at the discretized points.) The integral over $V(r)$ is Gaussian, and can be done (using the functional integral extension of the formulae in the appendix) yielding the desired result of equations (141) and (142). In the next lecture we will start from this point and use the variational method to try to complete, at least approximately, the calculation of the average free energy of an ensemble of systems with Gaussian random potentials. We shall discover that very interesting problems will arise in such a computation.

IV: Variational Replica Approach and Replica Symmetry Breaking

In this lecture we shall attempt an approximate computation of the average free energy of the ensemble of physical systems consisting of a single particle in a Gaussian random potential. To be specific, we choose the Hamiltonian to be

$$H = \frac{r^2}{2} + V(r) \quad (147)$$

where $\overline{V(r)} = 0$ and

$$\overline{V(r)V(r')} = W - f|r - r'| \quad (148)$$

As we learned in the last lecture, the average over disorder can be done by the replica method, yielding the expression

$$\overline{F} = -T \lim_{n \rightarrow 0} \frac{1}{n} \ln \int_{-\infty}^{\infty} \prod_{a=1}^n dr_a \exp\left(-\frac{H_{eff}}{T}\right) \quad (149)$$

where in this case, the effective replica Hamiltonian is

$$H_{eff} = \frac{1}{2} \sum_{a=1}^n r_a^2 - \frac{1}{2T} \sum_{a=1}^n \sum_{b=1}^n W - f|r_a - r_b|. \quad (150)$$

Notice that the average over disorder induces an effective *attractive* interaction between particles from different replicas.

Unfortunately, the form of the interaction makes an exact integration over the r_a variables impossible, so we must resort to approximate methods. Fortunately, our variational method can still be applied (although because of the $n \rightarrow 0$ limit, the trial average free energy becomes a lower bound on the true average free energy rather than an upper bound). Recalling the formalism that we learned in the first two lectures, we define a trial average free energy \tilde{F} by

$$\tilde{F} \equiv \overline{F_0} + \langle H_{eff} - H_0 \rangle_0 \quad (151)$$

where H_0 is a trial replica Hamiltonian and the thermal average $\langle \rangle_0$ is taken with respect to it:

$$\overline{F_0} \equiv -T \lim_{n \rightarrow 0} \frac{1}{n} \ln Z_0 \quad (152)$$

$$Z_0 \equiv \int_{-\infty}^{\infty} \prod_{a=1}^n dr_a \exp\left(\frac{-H_0}{T}\right) \quad (153)$$

$$\langle X \rangle_0 \equiv \lim_{n \rightarrow 0} \frac{1}{n} \frac{1}{Z_0} \int_{-\infty}^{\infty} \prod_{a=1}^n dr_a X \exp\left(\frac{-H_0}{T}\right) \quad (154)$$

As the class of trial replica Hamiltonians, we use the quadratic form that we solved exactly in the last lecture:

$$H_0 = \frac{1}{2} \sum_{a=1}^n \sum_{b=1}^n (G^{-1})_{ab} r_a r_b \quad (155)$$

where G is a replica symmetric matrix with diagonal elements \tilde{g} and off-diagonal elements g . These matrix elements are now the variational parameters which we will vary to optimize the trial free energy.

(The idea of using quadratic trial replica Hamiltonians (with a more general replica-symmetry-broken form which we will learn about later) in a variational approach was first suggested in a paper by Shakhovich and Gutin [18] on the replica approach to the heteropolymer problem. Mézard and Parisi [13] significantly developed the ideas of this approach, and in a recent preprint [14], applied it directly to the problem we are studying today. I have chosen to discuss their work because it is a particularly illuminating example of the replica method from the pedagogical point of view.)

Fortunately, we have already done most of the work necessary to compute the various pieces of the trial average free energy. For example, from our computation of the free energy in the last lecture we have (see equation (118))

$$\overline{F}_0 = -\frac{T}{2} \ln(2\pi T) - \frac{T}{2} \lim_{n \rightarrow 0} \frac{1}{n} \text{Tr} \ln G \quad (156)$$

and using the result we derived last lecture for the trace of the logarithm of a replica-symmetric matrix (equation (123)), we find

$$\overline{F}_0 = -\frac{T}{2} \ln(2\pi T) - \frac{T}{2} \left(\ln(\tilde{g} - g) + \frac{g}{\tilde{g} - g} \right) \quad (157)$$

For $\langle -H_0 \rangle_0$, we have

$$\langle -H_0 \rangle_0 = -\frac{1}{2} \left\langle \sum_{a=1}^n \sum_{b=1}^n (G^{-1})_{ab} r_a r_b \right\rangle_0 \quad (158)$$

$$\langle -H_0 \rangle_0 = -\frac{T}{2} \lim_{n \rightarrow 0} \frac{1}{n} \sum_{a=1}^n \sum_{b=1}^n (G^{-1})_{ab} G_{ab} = -\frac{T}{2} \quad (159)$$

$\langle H_{eff} \rangle_0$ is itself composed of a few pieces:

$$\left\langle \frac{1}{2} \sum_{a=1}^n r_a^2 \right\rangle_0 = \frac{T}{2} \lim_{n \rightarrow 0} \frac{1}{n} \sum_{a=1}^n G_{aa} = \frac{T\tilde{g}}{2} \quad (160)$$

The $a = b$ piece of the interaction term is

$$\left\langle -\frac{W}{2T} \sum_{a=1}^n 1 \right\rangle_0 = -\frac{W}{2T} \lim_{n \rightarrow 0} \frac{1}{n} n = -\frac{W}{2T} \quad (161)$$

For the $a \neq b$ piece of the interaction term, we note that $z \equiv r_a - r_b$ is a Gaussian variable, with

$$\langle z \rangle_0 = 0 \quad , \quad \langle z^2 \rangle_0 = B_{ab} \equiv T(G_{aa} + G_{bb} - 2G_{ab}) = 2T(\tilde{g} - g) \quad (162)$$

Using what we learned about Gaussian variables in the second lecture (see the discussion around equation (58)), that means

$$\begin{aligned} & \frac{f}{2T} \left\langle \sum_{a \neq b} |r_a - r_b| \right\rangle_0 \\ &= \frac{f}{2T} \lim_{n \rightarrow 0} \frac{1}{n} \sum_{a \neq b} \int_{-\infty}^{\infty} \frac{dz}{\sqrt{2\pi B_{ab}}} \exp\left(-\frac{z^2}{2B_{ab}}\right) |z| \end{aligned} \quad (163)$$

$$= -\frac{f}{2T} \int_{-\infty}^{\infty} \frac{dz}{\sqrt{2\pi[2T(\tilde{g} - g)]}} \exp\left(-\frac{z^2}{2[2T(\tilde{g} - g)]}\right) |z| \quad (164)$$

$$= -f \sqrt{\frac{(\tilde{g} - g)}{\pi T}} \quad (165)$$

Note that the sign of this term switched when taking the $n \rightarrow 0$ limit because there are $n(n-1)$ off-diagonal matrix elements. Collecting all the terms, we find the trial free energy

$$\tilde{F} = -\frac{T}{2}(\ln(2\pi T) + 1) - \frac{W}{2T} - \frac{T}{2} \left(\ln(\tilde{g} - g) + \frac{g}{\tilde{g} - g} \right) + \frac{T\tilde{g}}{2} - f\sqrt{\frac{(\tilde{g} - g)}{\pi T}} \quad (166)$$

Setting the derivatives with respect to \tilde{g} and g equal to zero, we find the two equations

$$\frac{T}{2} \left(-\frac{1}{\tilde{g} - g} + \frac{g}{(\tilde{g} - g)^2} + 1 \right) - \frac{f}{2\sqrt{\pi T(\tilde{g} - g)}} = 0 \quad (167)$$

$$\frac{T}{2} \left(\frac{1}{\tilde{g} - g} - \frac{g}{(\tilde{g} - g)^2} - \frac{1}{\tilde{g} - g} \right) + \frac{f}{2\sqrt{\pi T(\tilde{g} - g)}} = 0 \quad (168)$$

Adding these two equations, we find $\tilde{g} - g = 1$, and plugging that back into the second equation, we find $g = f/\sqrt{\pi T^3}$. We recall from the last lecture (equation (129)) that the physical correlation function corresponding to the thermal fluctuations is

$$\overline{\langle r^2 \rangle - \langle r \rangle^2} = T(\tilde{g} - g) = T \quad (169)$$

which tells us that the thermal fluctuations are proportional to temperature, as we expect. The correlation function measuring the average of both thermal and disorder-induced fluctuations is

$$\overline{\langle r^2 \rangle} = T\tilde{g} = T + \frac{f}{\sqrt{\pi T}} \quad (170)$$

which is a disaster! Our result suggests that the typical displacement caused by disorder diverges as the temperature approaches zero, which makes no sense. We know in fact that at zero temperature, the particle will sit at the bottom of the lowest well in each sample, and the lowest well should always be some finite distance from the origin, given the quadratic term we have included in the Hamiltonian. Therefore we should have gotten a finite answer for the fluctuations at zero temperature.

What went wrong? In fact, it should have been obvious that we were headed for trouble given what we learned in the last lecture about the physical meaning of the

replica-symmetric trial Hamiltonian. As we learned, a replica-symmetric Hamiltonian gives an exact description of an ensemble of systems, each consisting of a particle in a *single* quadratic well, with the position of the well distributed from sample to sample according to a Gaussian distribution. This is very far from our ensemble of systems, for which every sample has many metastable minima. We should not be surprised that we get nonsense from an approach which approximates rough random potentials by a single quadratic well. Garbage in, garbage out.

What is not so obvious is how to make a better approximation. We need an approximation which includes the possibility of disorder within a sample, as well as sample-to-sample disorder. In fact, such an approximation is possible, even retaining the quadratic form of our trial replica Hamiltonian, using an amazing idea due originally to Giorgio Parisi, which he proposed in the context of a replica approach to the spin glass problem [17]. The idea, called *replica symmetry breaking*, technically amounts to widening the class of n by n replica matrices G_{ab} considered in the trial replica Hamiltonian to include matrices for which the off-diagonal elements are not necessarily equal. We will first explain how to do this technically, and then describe the physical meaning of an ensemble of systems represented by a trial quadratic replica Hamiltonian with replica-symmetry-broken matrices. As we will see, replica symmetry breaking is the key ingredient to describing ensembles of systems in which each sample has many metastable states.

We will begin our description of replica-symmetry-broken (RSB) matrices with the simplest such possibility, called *one-step replica symmetry breaking*. In a one-step RSB matrix, the n replicas are grouped into n/m families of m members each. There are three kinds of matrix elements: the diagonal elements, off-diagonal elements for which the row and column replicas belong to the same family, and off-diagonal elements for which the row and column replicas belong to different families. For

example, in the following one-step RSB matrix A_{ab} , $n = 4$, while $m = 2$:

$$\begin{pmatrix} \tilde{a} & a_1 & a_0 & a_0 \\ a_1 & \tilde{a} & a_0 & a_0 \\ a_0 & a_0 & \tilde{a} & a_1 \\ a_0 & a_0 & a_1 & \tilde{a} \end{pmatrix}$$

We have adopted the convention that for a one-step RSB matrix A_{ab} , the n diagonal elements are denoted \tilde{a} , the $n(m - 1)$ "intra-family" matrix elements are denoted a_1 , and the $n(n - m)$ "inter-family" matrix elements are denoted by a_0 .

If we multiply two one-step RSB matrices which share the same values of n and m , the result is also a one-step RSB matrix with the same n and m values. In fact, it is easy to work out that if

$$C_{ab} = \sum_{c=1}^n A_{ac}B_{cb}, \quad (171)$$

then

$$\tilde{c} = \tilde{a}\tilde{b} + (n - m)a_0b_0 + (m - 1)a_1b_1 \quad (172)$$

$$c_0 = \tilde{a}b_0 + (m - 1)a_1b_0 + \tilde{b}a_0 + (m - 1)b_1a_0 + (n - 2m)a_0b_0 \quad (173)$$

$$c_1 = \tilde{a}b_1 + a_1\tilde{b} + (m - 2)a_1b_1 + (n - m)a_0b_0 \quad (174)$$

We derived these equations thinking of n and m as positive integers with $n \geq m$, but we can consider the equations to be valid for arbitrary real n and m . We can take the $n \rightarrow 0$ limit of these equations and then use them as we did last lecture to find the set of conditions that must be satisfied if A_{ab} is the inverse of B_{ab} . We can also use them to compute the trace of the logarithm of a one-step RSB matrix A_{ab} ; the result is

$$\lim_{n \rightarrow 0} \frac{1}{n} \text{Tr} \ln A = \ln(\tilde{a} - \langle a \rangle) + \frac{a_0}{\tilde{a} - \langle a \rangle} - \frac{m - 1}{m} \ln \left(\frac{\tilde{a} - a_1}{\tilde{a} - \langle a \rangle} \right) \quad (175)$$

where

$$\langle a \rangle \equiv ma_0 + (1 - m)a_1. \quad (176)$$

We recover our previous replica-symmetric results in either of the limits $m \rightarrow n$ or $m \rightarrow 1$, corresponding to all the off-diagonal elements being "intra-family" or "inter-family" respectively.

We have developed enough "replica technology" to be able to use a trial Hamiltonian of the same quadratic form as in equation (155) but with a one-step RSB G_{ab} matrix. The one-step RSB trial Hamiltonian has four variational parameters: \tilde{a} , a_0 , a_1 , and m , and must give a trial free energy at least as good as the replica-symmetric Hamiltonian, because it is more general. When we compute thermal replica expectation values and encounter sums over off-diagonal elements, we must remember to break the sum into intra-family and inter-family parts. This has implications for the computation of disorder averages. For example, the connected correlation function measuring thermal fluctuations is within a one-step RSB trial Hamiltonian (compare equation(129))

$$\begin{aligned} \overline{\langle r^2 \rangle - \langle r \rangle^2} &= \left\langle \sum_{a=1}^n \sum_{b=1}^n r_a r_b \right\rangle_0 \\ &= \lim_{n \rightarrow 0} \frac{1}{n} (nT\tilde{g} + n(n-m)Tg_0 + n(m-1)Tg_1) \\ &= T(\tilde{g} - \langle g \rangle). \end{aligned} \tag{177}$$

In the computation of the trial free energy, the only terms that change from their replica-symmetric forms are the term giving the trace of the logarithm of the G_{ab} matrix and the term from the $a \neq b$ piece of the interaction expectation value. The average square of the Gaussian variable $z \equiv r_a - r_b$ depends on whether replica a and b are in the same family or not; $\langle z^2 \rangle_0 = 2T(\tilde{g} - g_1)$ if they are in the same family, and $\langle z^2 \rangle_0 = 2T(\tilde{g} - g_0)$ if they are in different families. We thus find (compare with the replica symmetric result of equation (165)) that

$$\frac{f}{2T} \left\langle \sum_{a \neq b} |r_a - r_b| \right\rangle_0 = -\frac{fm}{\pi} \sqrt{\frac{(\tilde{g} - g_0)}{T}} - f(1-m) \sqrt{\frac{(\tilde{g} - g_1)}{\pi T}} \tag{178}$$

If we collect all the pieces together, the one-step RSB trial free energy is

$$\begin{aligned} \tilde{F} = & -\frac{T}{2} \left(\ln(2\pi T) + 1 - \tilde{g} + \ln(\tilde{g} - \langle g \rangle) + \frac{g}{\tilde{g} - \langle g \rangle} + \frac{m-1}{m} \ln \left(\frac{\tilde{g} - g_1}{\tilde{g} - \langle g \rangle} \right) \right) \\ & - \frac{W}{2T} - fm \sqrt{\frac{(\tilde{g} - g_0)}{\pi T}} - f(1-m) \sqrt{\frac{(\tilde{g} - g_1)}{\pi T}} \end{aligned} \quad (179)$$

Now of course, we can optimize the free energy with respect to the variational parameters \tilde{g} , g_0 , g_1 , and m . It is important to realize that in the $n \rightarrow 0$ limit, the optimal values of the parameters will give a *maximum* of the free energy, rather than a minimum [15].

The physical meaning of these variational parameters has been worked out by Mézard and Parisi in their paper [13]. It can be understood in much the same way that we understood the physical meaning of the variational parameters \tilde{g} and g in the replica-symmetric case: by comparing results of a "direct" computation of some disorder average with the results of the replica computation. Specifically, Mézard and Parisi showed that if we have an ensemble of physical systems for which the average free energy is given by the expression

$$\bar{F} = -T \lim_{n \rightarrow 0} \frac{1}{n} \ln \int_{-\infty}^{\infty} \prod_{a=1}^n dr_a \exp \left(-\frac{H_{replica}}{T} \right) \quad (180)$$

with

$$H_{replica} = \frac{1}{2} \sum_{a=1}^n \sum_{b=1}^n (G^{-1})_{ab} r_a r_b \quad (181)$$

with a one-step RSB G_{ab} matrix, then the ensemble is equivalent to one constructed in the following way:

1. For each sample we determine a "central point" r_0 by choosing it from the probability distribution

$$p(r_0) = \frac{1}{\sqrt{2\pi g_0}} \exp \left(-\frac{r_0^2}{2g_0} \right) \quad (182)$$

2. Around the central point of each sample we generate an infinite number of quadratic wells. The positions r_α chosen for the center of each quadratic well are

uncorrelated with each other; they are chosen from the probability distribution

$$p(r_\alpha) = \frac{1}{\sqrt{2\pi(g_1 - g_0)}} \exp\left(-\frac{(r_\alpha - r_0)^2}{2(g_1 - g_0)}\right) \quad (183)$$

The potential V_α at the minimum of each well is also a random variable uncorrelated with the position r_α or with any of the other wells. The V_α are chosen from a probability distribution such that the average fraction of wells with minima below the level V is e^{mV} . (That is, all the wells have minima below $V = 0$; while a fraction e^{-m} have minima below $V = -1$, and so on. m is normally between 0 and 1.) In the total Boltzman sum for the sample, each well will have a weight

$$W_\alpha = \frac{e^{-V_\alpha/T}}{\sum_\alpha e^{-V_\alpha/T}} \quad (184)$$

3. The thermal fluctuations within each well are equal to $T(\tilde{g} - g_1)$; that is, the effective well Hamiltonian is

$$H_{well} = \frac{1}{2} \frac{(r - r_\alpha)^2}{(\tilde{g} - g_1)} \quad (185)$$

As I mentioned, this "interpretation" was arrived at by Mézard and Parisi [13] by demonstrating that the disorder averages that one computes using the interpretation agree with those obtained from the replica method.

An ensemble such as that described above clearly exhibits the kind of intra-sample disorder that we want. The many quadratic wells mimic the many metastable minima in our original problem, and our various variational parameters give a clear quantitative measure of the disorder.

Nevertheless, one can do better than the one-step RSB scheme. In the one-step scheme, the replicas are organized into a very simple hierarchy of families. Parisi [17] proposed an even more general scheme called *k-step replica symmetry breaking* in which the n replicas are organized into n/m_1 "1-families" of m_1 elements each. The m_1 replicas in a "1-family" are then further organized into m_1/m_2 "2-families" of m_2

elements each, and so on until we reach the level of "k-families," which are ordinary families consisting of m_k elements, and "k+1-families" which are simple elements. We use the convention that all replicas belong to the same "0-family" so that $m_0 = n$ and $m_{k+1} = 1$. Off-diagonal elements of the k -step RSB matrix A_{ab} for which replicas a and b belong to the same " l -family" (the maximal l being chosen) are labeled a_l .

As an example, we give a 2-step RSB A_{ab} matrix, with $m_0 \equiv n = 8$, $m_1 = 4$, and $m_2 = 2$:

$$\begin{pmatrix} \tilde{a} & a_2 & a_1 & a_1 & a_0 & a_0 & a_0 & a_0 \\ a_2 & \tilde{a} & a_1 & a_1 & a_0 & a_0 & a_0 & a_0 \\ a_1 & a_1 & \tilde{a} & a_2 & a_0 & a_0 & a_0 & a_0 \\ a_1 & a_1 & a_2 & \tilde{a} & a_0 & a_0 & a_0 & a_0 \\ a_0 & a_0 & a_0 & a_0 & \tilde{a} & a_2 & a_1 & a_1 \\ a_0 & a_0 & a_0 & a_0 & a_2 & \tilde{a} & a_1 & a_1 \\ a_0 & a_0 & a_0 & a_0 & a_1 & a_1 & \tilde{a} & a_2 \\ a_0 & a_0 & a_0 & a_0 & a_1 & a_1 & a_2 & \tilde{a} \end{pmatrix}$$

Two k -step RSB matrices with the same values of m_i will, when multiplied together give another k -step RSB matrix with the same values of m_i . As we did above for one-step RSB matrices, we can write down the multiplication rules for k -step RSB matrices, and then derive formulas for the inverse and the trace of the logarithm of a k -step RSB matrix. We can thus compute the trial free energy, and optimize it with respect to all the variational parameters in the problem (\tilde{g} , the $k + 1$ off-diagonal elements g_i , and the k parameters m_i .) Clearly, a trial Hamiltonian based on a k -step RSB matrix will always be more general than one based on a $k - 1$ -step RSB matrix, and should thus provide a trial free energy that is at least as good. The physical interpretation of a such a Hamiltonian is a rather straightforward generalization of the interpretation of a one-step RSB Hamiltonian. For example, in the case of a two-step RSB Hamiltonian, in each sample, one first determines a "central point" from a

Gaussian probability distribution. Around that central point one distributes an infinite number of "wells" as we did in the one-step case. Now, however, the position of the center of each well is just a "central point" around which one distributes an infinite number of quadratic "sub-wells," again with spatial and energy displacements determined by probability distributions using the g_i and m_i parameters. Finally, the particle fluctuates thermally in the sub-wells. In three-step RSB, the sub-wells are further broken down into sub-sub-wells, and so on. For more details, see [13].

The best we could do within the k -step RSB scheme would be to take the $k \rightarrow \infty$ limit. Incredibly enough, this limit can be taken, and is referred to as *full replica symmetry breaking*. Actually, in order to make sense of this limit, we need the following fact, which has been empirically observed in all known examples of k -step replica symmetry breaking with finite k . The parameters m_i should, in the $n \rightarrow 0$ limit, obey the inequalities

$$0 \leq m_1 \leq m_2 \leq \dots \leq m_{k-1} \leq m_k \leq 1. \quad (186)$$

We can "justify" these inequalities in the following way. In each row of a RSB matrix, there are $n - 1$ off-diagonal elements, which in the $n \rightarrow 0$ limit, equals -1 . Say, for example, that we had the two-step RSB matrix A_{ab} . Then $n - m_1$ of the off-diagonal elements on each row would equal a_0 , $m_1 - m_2$ of the off-diagonal elements would equal a_1 , and $m_2 - 1$ of those elements would equal a_2 . If we want a certain fraction of the elements to have each of the possible values, and if all the fractions should be between 0 and 1, then we require for example that $n - m_1$ should be between 0 and -1 (so that the number of elements equal to a_0 be a fraction of -1), which means (for $n \rightarrow 0$) that $0 \leq m_1 \leq 1$. The other inequalities similarly follow. If the above inequalities are obeyed, then we can represent all the parameters m_i and a_i in terms of a single function $a(x)$, where x ranges from 0 to 1. We construct the function piece by piece, with $a(x) = a_i$ for $m_i \leq x < m_{i+1}$. For any finite k , the function will $a(x)$

will thus consist of a series of steps.

The variable x in the function $a(x)$ is a measure of the relatedness of the two replicas denoted by the row and column indices of a matrix element of A_{ab} . When $x \rightarrow 1$, the relatedness is high, while when $x \rightarrow 0$, the relatedness is low. x is the generalization of the one-step variable m , which you recall was physically interpreted in terms of the difference in potentials between the different wells. Physically, therefore, the small x regime refers to correlations between wells that have a big potential difference.

The rules for multiplying two full RSB 0×0 matrices A_{ab} and B_{ab} to get a new matrix C_{ab} can be written in terms of the parameter representing the diagonal element and the function representing all the off-diagonal elements [17]. The following formulae are taken from [13], Appendix II:

$$\tilde{c} = \tilde{a}\tilde{b} - \langle ab \rangle \quad (187)$$

$$c(u) = (\tilde{b} - \langle b \rangle)a(u) + (\tilde{a} - \langle a \rangle)b(u) - \int_0^u dv (a(u) - a(v))(b(u) - b(v)) \quad (188)$$

where

$$\langle a \rangle \equiv \int_0^1 du a(u) \quad (189)$$

In order that B_{ab} be the inverse of A_{ab} we require that $\tilde{c} = 1$ and $c(u) = 0$. The resulting conditions are written out in detail in reference [13]. They have also worked out the formula for the trace of the logarithm of a full RSB matrix:

$$\lim_{n \rightarrow 0} \frac{1}{n} \text{Tr} \ln A = \ln(\tilde{a} - \langle a \rangle) + \frac{a(0)}{\tilde{a} - \langle a \rangle} - \int_0^1 \frac{du}{u^2} \ln \left(\frac{\tilde{a} - \langle a \rangle - [a](u)}{\tilde{a} - \langle a \rangle} \right) \quad (190)$$

where

$$[a](u) \equiv - \int_0^u dv a(v) + ua(u) \quad (191)$$

You can check that our previous formulae in the replica symmetric or one-step RSB cases are just special cases of these more general formulae.

We can now employ a quadratic trial Hamiltonian with a full RSB G_{ab} matrix. The variational parameters will be \tilde{g} and the full function $g(x)$. Note that whenever we encounter a sum over off-diagonal elements, the sum can be replaced by an integral over x (with a minus sign to account for the fact that there are -1 off-diagonal elements per row.) For example, the connected correlation function measuring thermal fluctuations is

$$\begin{aligned} \overline{\langle r^2 \rangle} - \langle r \rangle^2 &= \left\langle \sum_{a=1}^n \sum_{b=1}^n r_a r_b \right\rangle_0 = \left\langle \sum_{a=1}^n r_a^2 + \sum_{a \neq b} r_a r_b \right\rangle_0 \\ &= T \left(\tilde{g} - \int_0^1 dx g(x) \right) = T(\tilde{g} - \langle g \rangle) \end{aligned} \quad (192)$$

With this information, it is easy to re-compute the trial free energy assuming a full RSB G matrix; the final result is

$$\begin{aligned} \tilde{F} &= -\frac{T}{2} \left(\ln(2\pi T) + 1 - \tilde{g} + \ln(\tilde{g} - \langle g \rangle) + \frac{g(0)}{\tilde{g} - \langle g \rangle} - \int_0^1 \frac{du}{u^2} \ln \left(\frac{\tilde{g} - \langle g \rangle - [g](u)}{\tilde{g} - \langle g \rangle} \right) \right) \\ &\quad - \frac{W}{2T} - f \int_0^1 du \sqrt{\frac{\tilde{g} - g(u)}{\pi T}} \end{aligned} \quad (193)$$

We can now optimize this free energy with respect to \tilde{g} and $g(u)$. Actually, Mézard and Parisi used the equivalent and technically simpler procedure of differentiating with respect to the G_{ab} matrix first, before taking the $n \rightarrow 0$ limit, and then breaking replica symmetry on the saddle point equations. Because the inverse of G naturally arises in this procedure, they chose to write the trial Hamiltonian in terms of the "self-energy" matrix σ_{ab} :

$$H_0 = \frac{1}{2} \sum_{a=1}^n r_a^2 - \frac{1}{2} \sum_{a=1}^n \sum_{b=1}^n \sigma_{ab} r_a r_b \quad (194)$$

so that

$$G_{ab} = \left((1 - \sigma)^{-1} \right)_{ab} \quad (195)$$

Mézard and Parisi [14] found the following solution: For $t > 1$, where t is the "reduced temperature"

$$t \equiv T \left(\frac{2\sqrt{\pi}}{f} \right)^{2/3} \quad (196)$$

the replica symmetric solution given previously is valid, with $g(u) = \sigma(u) = 2/t^{3/2}$.

For $t \leq 1$, the result is

$$\sigma(u) = \begin{cases} 2/t, & 0 < u < 3t/4, \\ 32u^2/9t^3, & 3t/4 < u < 3/4, \\ 2/t^3, & 3/4 < u < 1, \end{cases} \quad (197)$$

and

$$\tilde{\sigma} - \langle \sigma \rangle = 1 \quad (198)$$

Their final result for the average of combined thermal and disorder-induced fluctuations is

$$\overline{\langle r^2 \rangle} = \begin{cases} T \left(1 + 2t^{-3/2} \right), & t > 1, \\ \frac{3T}{t} = 3 \left(\frac{f}{2\sqrt{\pi}} \right)^{2/3}, & t < 1, \end{cases} \quad (199)$$

which makes much more sense than our previous replica symmetric result. According to this solution, the system "freezes" at the critical temperature $T_c = \left(\frac{f}{2\sqrt{\pi}} \right)^{2/3}$ into its low-temperature configuration, and at zero temperature, the fluctuations are finite. Of course, we learned in the first lecture that we should not trust a variational approach to the extent of believing that there is actually a sharp transition as described here. Indeed Mézard and Parisi [14] performed numerical simulations of this system and showed that in the true system, the cross-over from the low to high temperature regimes is actually smoothed out. Nevertheless, their comparison does show that the predictions of the full RSB variational approach are not too far from reality.

V: Replicas and the Impure Superconductor in a Magnetic Field

In this last lecture, we will apply the ideas we have learned about to a real physical system which has attracted considerable experimental interest, especially in the last five years—the impure superconductor in a magnetic field. The work that I will describe was done in collaboration with Jean-Philippe Bouchaud and Marc Mézard at the Ecole Normale Supérieure in Paris [5].

As you may know, a superconductor has the property that sufficiently small magnetic fields cannot penetrate inside of it. Nevertheless, for all superconductors, there is a critical magnetic field (which will depend on the temperature) above which the superconductor cannot maintain itself. In type-I superconductors, there is a single transition from low magnetic fields when the sample will superconduct, to high magnetic fields when it will not. In type-II superconductors, the type which we will discuss today (the new high- T_c superconductors are all type-II), there exists an intermediate regime in which the magnetic field partially penetrates into the sample. In this intermediate regime, the magnetic field inside the superconductor is organized into a triangular lattice of flux lines (called an "Abrikosov lattice") parallel to the direction of the external field. As the magnetic field is increased from zero in a type-II superconductors, two transitions occur. At the first transition, called H_{c1} , the magnetic field first begins to penetrate and the Abrikosov lattice is formed. As the magnetic field is further increased, the density of flux lines will continually increase, until at a second transition called H_{c2} the superconductivity is finally destroyed.

The Abrikosov lattice can be experimentally identified in a *Bitter decoration experiment*. The experimenter lays nickel filings on the superconductor, and they are attracted to the flux lines. A photograph is taken of the filings, and they indicate the structure of the lattice. A striking feature of the new high-temperature superconductors has been the extent of the disorder in the Bitter patterns of some samples; disorder which perhaps is caused by intrinsic quenched impurities in the sample. In

this lecture, we will be using our replica methods to compute the properties a model that assumes such impurities, and we shall see that one can make detailed predictions that can be compared to the Bitter decoration experiments. We will begin, however by constructing a model of an Abrikosov lattice in a perfectly pure superconductor.

An Abrikosov lattice is much like any other classical crystal, except that it is constructed from flux lines rather than atoms. Nevertheless it will have thermal fluctuations which can be understood using classical statistical mechanics. The simplest model that describes a classical crystal is the perfectly quadratic elastic solid. One can think of an elastic solid in terms of either a microscopic atomic picture or a more macroscopic continuum picture. In the two-dimensional version of the microscopic atomic picture, the atoms are arranged in a triangular array, with a lattice spacing a , and each atom is linked to its nearest neighbors by a spring. The springs are taken to be perfectly quadratic harmonic oscillators, with possibly different spring constants for longitudinal and transverse fluctuations. Each atom has some unperturbed equilibrium position \vec{x} . You can call \vec{x} the "label" of the atom, because no matter how much it jiggles around, its unperturbed equilibrium position \vec{x} will never change. At any given time, each atom will actually be at some position $\vec{r}(\vec{x})$. The displacement of each atom from its equilibrium position is $\vec{u}(\vec{x}) \equiv \vec{r}(\vec{x}) - \vec{x}$. The energy of the system is just the sum of all the two-body terms corresponding to stretching each spring.

This microscopic picture is useful to keep in mind, especially when discussing experiments that can measure the microscopic structure of the lattice. But it is often more useful and general to think in terms of a continuum description—such a description will be appropriate when we work at length scales very long compared to the atomic spacing. In this description, we imagine following some point \vec{x} in the solid as we stretch it—again it gets displaced by an amount $\vec{u}(\vec{x})$ to the new position $\vec{r}(\vec{x})$. Quite generally, we know that when we make a perturbation of a system around

its minimum, the energy will grow quadratically with the perturbation—in this case $\vec{u}(\vec{x})$. In fact, we expect that the continuum generalization of the atomic Hamiltonian will look like

$$H_{elastic} = \int d^2\vec{x} \left[(C_{11} - C_{66}) \left(\sum_{\alpha} \frac{\partial u_{\alpha}}{\partial x_{\alpha}} \right)^2 + C_{66} \sum_{\alpha\beta} \left(\frac{\partial u_{\alpha}}{\partial x_{\beta}} \right)^2 \right] \quad (200)$$

The Greek indices α and β refer to the x and y directions, and the derivatives are the generalizations of nearest neighbor energy cost terms. This is actually the most general quadratic form consistent with the symmetries of a triangular lattice. C_{11} is called the "bulk modulus" and tell you how hard it is to squeeze the solid; C_{66} is called the "shear modulus" and tell you how difficult it is to shear the solid. For a solid like rubber which is easy to shear, $C_{66} \ll C_{11}$. This will also be true for superconductors in the regime that we are interested in.

For a three-dimensional solid of triangular lines, the continuum elastic Hamiltonian is

$$H_{elastic} = \int d^2\vec{x} dz \left[(C_{11} - C_{66}) \left(\sum_{\alpha} \frac{\partial u_{\alpha}}{\partial x_{\alpha}} \right)^2 + C_{66} \sum_{\alpha\beta} \left(\frac{\partial u_{\alpha}}{\partial x_{\beta}} \right)^2 + C_{44} \sum_{\alpha} \left(\frac{\partial u_{\alpha}}{\partial z} \right)^2 \right] \quad (201)$$

where C_{44} is the "tilt modulus" which measures how hard it is to tilt the lines as they travel in the z direction. This is the standard Hamiltonian used to describe the Abrikosov lattice of vortex lines in type-II superconductors. It should generally give a good description when the displacements \vec{u} are not too large—that is, when the temperature is low enough.

Because these Hamiltonians are quadratic, we can solve for all the thermodynamics exactly. In particular, we can calculate any correlation function that we want. One interesting example is the "translational correlation function" or "Debye-Waller factor"

$$g_{\vec{K}}(\vec{x}) \equiv \left\langle e^{i\vec{K}(\vec{u}(\vec{x}) - \vec{u}(\vec{0}))} \right\rangle \quad (202)$$

where \vec{K} is some reciprocal lattice vector. We can calculate this correlation function exactly; in two dimensions we find the interesting result [16]

$$g_{\vec{K}}(\vec{x}) \sim x^{-\eta_{\vec{K}}(T)} \quad (203)$$

where $\eta_{\vec{K}}(T)$ is a temperature-dependent exponent:

$$\eta_{\vec{K}}(T) = \frac{T(2C_{66} + C_{11})|\vec{K}^2|}{4\pi C_{66}(C_{66} + C_{11})} \quad (204)$$

In three dimensions, this correlation function does not decay to zero for large \vec{x} , but only goes to a constant. This correlation function measure density-density correlations; it can be related to the probability that if we have an atom at some position, there is an atom exactly x lattice spacings away. The fact that it goes to a constant at long-distances in three-dimensions means that the crystal has "long-range order." We will be able to compare the above results with our computations in the disordered case. Even more interestingly, we will be able to compare with experiments where these correlation functions are measured in real systems.

So far, our Hamiltonians have described the Abrikosov lattice in a perfectly pure superconductor. We will now add a term which models the effect of defects through a random pinning potential. I will motivate the form of this potential using the example of a thin film cuprate superconductor in a magnetic field that points in a direction perpendicular to the film. In this example, the triangular lattice of flux points (flux points in two dimensions; flux lines in three dimensions) will sit on square lattice of copper and oxygen atoms. These atoms never move, but the flux points do. The spacing of the lines will be on the order of 1000 lattice spacings of the square copper-oxide lattice. We assume that there are some "point" defects in the copper-oxide lattice, like for example oxygen vacancies. We assume that whatever the defects are, their typical correlation length Δ_{xy} will be much less than the lattice spacing of the flux lines a . Note that while the flux points move around, the defects in the copper-oxide lattice are "quenched": they are frozen in and never move.

The defects in the copper-oxide lattice will attract flux points. To properly describe the lattice of flux points including their attraction to the defects, we should add to our elastic Hamiltonian a pinning potential

$$H = H_{elastic} + \sum_{\vec{x}} V(\vec{r}(\vec{x})). \quad (205)$$

We idealize the pinning potential a little by assuming that it is given by a Gaussian probability distribution similar to those we have been using in the last two lectures; in the two-dimensional case, for example, the first two moments are chosen to be

$$\overline{V(\vec{r})} = 0 \quad , \quad \overline{V(\vec{r})V(\vec{r}') } = U_{pin}^2 \exp\left(-\frac{(\vec{r} - \vec{r}')^2}{2\Delta_{xy}^2}\right) \quad (206)$$

Such a form implies that the typical magnitude of the potential is the pinning energy U_{pin} and the correlation length is Δ_{xy} .

The reasonableness of our assumptions is in fact bolstered by a quick examination of the results of some recent Bitter decoration experiments on three-dimensional cuprate superconductors [11]. In the 69 Gauss experiment of the Bell Labs group, one can see that a perfect triangular topological structure is maintained within the camera's field of view. The physical idea behind our calculation is that the visible local distortions of the triangular lattice are caused by the attraction flux lines feel for microscopic defects like oxygen vacancies. This hypothesis can be checked by comparing the quantitative predictions that we derive for correlation functions with those obtained by the experimentalists for their samples. We shall see that the agreement is quite good.

We are interested in computing the average free energy of an ensemble of systems described by the above Hamiltonian, and we begin by averaging over the disorder using the replica method. As usual, the average over disorder converts our problem into a mathematically equivalent one of n identical crystals, for which the atoms no longer feel the random potential, but instead feel an inter-replica attraction. In

particular, we have

$$\bar{F} = -T \lim_{n \rightarrow 0} \frac{1}{n} \exp \left(-\frac{H_{eff}}{T} \right) \quad (207)$$

with (in two dimensions)

$$H_{eff} = \frac{1}{2} \sum_{a=1}^n \int d^2 \vec{x} \left[(C_1 - C_6) \left(\sum_{\alpha} \frac{\partial u_{\alpha}^a}{\partial x_{\alpha}} \right)^2 + C_6 \sum_{\alpha\beta} \left(\frac{\partial u_{\beta}^a}{\partial x_{\alpha}} \right)^2 \right] - \frac{W}{2T} \sum_{a=1}^n \sum_{b=1}^n \sum_{\vec{x}\vec{x}'} \delta(\vec{r}_a(\vec{x}) - \vec{r}_b(\vec{x}')) \quad (208)$$

where we have taken the limit $\Delta_{xy} \ll a$ to convert our Gaussian into a delta-function and where now all distances are written in units of the lattice spacing a . We define

$$W \equiv \frac{2\pi U_{pin}^2 \Delta_{xy}^2}{a^2} \quad (209)$$

and

$$C_1 \equiv C_{11} a^2 \quad , \quad C_6 \equiv C_{66} a^2 \quad (210)$$

The first part of the effective Hamiltonian is just the elastic crystal replicated n times. In the other part, one has an effective attraction between all the atoms in the system. Of course, they will not all sit on top of one another because the elastic term keeps them apart.

We cannot compute the average free energy exactly because the effective Hamiltonian is not quadratic. As usual, we will use the variational approach with a trial quadratic Hamiltonian which is as general as possible. Thus, we choose

$$H_0 = \frac{1}{2} \sum_{ab} \sum_{\alpha\beta} \sum_{\vec{x}\vec{x}'} \left(G^{-1} \right)_{\alpha\beta}^{ab} (\vec{x} - \vec{x}') u_a^{\alpha}(\vec{x}) u_b^{\beta}(\vec{x}') \quad (211)$$

Physically we are coupling the fluctuation of every atom (or flux point) in the system with every other atom, taking into account with the α and β indices the difference between longitudinal and transverse fluctuations. Because the system is translationally

invariant, we can diagonalize the spatial part of the G matrix by going into Fourier space, so that

$$H_0 = \frac{1}{2} \sum_{ab} \sum_{\alpha\beta} \int \frac{d^2\vec{q}}{(2\pi)^2} u_a^\alpha(\vec{q}) (G^{-1})_{\alpha\beta}^{ab}(\vec{q}) u_b^\beta(-\vec{q}) \quad (212)$$

We can diagonalize the spatial indices of the Green's function by breaking it down into longitudinal and transverse fluctuations

$$G_{\alpha\beta}(\vec{q}) \equiv \left(\delta_{\alpha\beta} - \frac{q_\alpha q_\beta}{q^2} \right) G_T(q) + \left(\frac{q_\alpha q_\beta}{q^2} \right) G_L(q) \quad (213)$$

Finally, we can compute the trial free energy \tilde{F} and try to optimize it with respect to the Green's functions $G_{L,T}^{ab}(q)$. I will not write out the trial free energy or the saddle-point equations here; they are long and not too enlightening and you can find them in reference [5]. Of course, when we optimize with respect to the replica indices, we use a full RSB matrix. That means that we ultimately optimize with respect to the Green's functions $G_{L,T}(q, v)$, where v is a real number ranging from 0 to 1. In fact, if one is lazy and tries to use a replica-symmetric ansatz, one runs into precisely the same type of trouble as we saw in the last lecture with a one-particle problem— all the fluctuations seem to diverge at zero temperature. Again a full-RSB approach is necessary to account for the fact that the many possible metastable configurations of the atoms.

The final form of the Green's functions are (for $C_{66} \ll C_{11}$, which is reasonable for the superconductors)

$$G_{L,T} = \frac{1}{q^{2+2\nu}} g_{L,T} \left(\frac{v}{q^\omega} \right) \quad (214)$$

where $\omega = 2\nu$ in two dimensions ($\omega = 2\nu + 1$ in three dimensions) and $g_{L,T}(x)$ is a complicated function given explicitly in [5]. ν is an exponent whose value and physical significance we will discuss shortly. Of course, the Green's function in and of itself is not too interesting. We are more interested in using the Green's function to com-

pute various disorder averages which have a more obvious physical and experimental significance.

Imagine, for example, that we take two atoms some distance x apart along the x axis and look at their typical squared longitudinal and transverse fluctuations. These are given by the disorder-average correlation functions $\tilde{B}_L(x)$ and $\tilde{B}_T(x)$:

$$\tilde{B}_L(x) = \overline{\langle (u_x(\vec{x}) - u_x(\vec{0}))^2 \rangle} \quad (215)$$

$$\tilde{B}_T(x) = \overline{\langle (u_y(\vec{x}) - u_y(\vec{0}))^2 \rangle} \quad (216)$$

We find that these correlation functions increase with distance with a power-law form (with possible logarithmic corrections)

$$\tilde{B} \sim x^{2\nu} \quad (217)$$

In particular, a full variational calculation gives the result (in two dimensions)

$$\tilde{B}_L(x) = \frac{3}{5}\tilde{B}_T(x) = \frac{3\Gamma(2/3)^2}{8\pi^{2/3}2^{1/3}} \left(\frac{x}{\xi}\right)^{2/3} \simeq 0.25 \left(\frac{x}{\xi}\right)^{2/3} \quad (218)$$

for $1 \ll x \ll \xi = C_6/\sqrt{W}$ and

$$\tilde{B}_T(x) = \frac{1}{2}\tilde{B}_T(x) = \frac{\sqrt{\pi}}{6} \frac{2^{1/2}}{3^{1/4}} \left(\frac{x}{\xi\sqrt{\ln x}}\right) \simeq 0.32 \left(\frac{x}{\xi\sqrt{\ln x}}\right) \quad (219)$$

for $x \gg \xi$. In three dimensions, we find

$$\tilde{B}_L(x) = \frac{3}{4}\tilde{B}_T(x) = \frac{3\Gamma(2/3)}{7\pi^{2/3}} \left(\frac{x}{\xi}\right)^{1/3} \simeq 0.27 \left(\frac{x}{\xi}\right)^{1/3} \quad (220)$$

for $1 \ll x \ll \xi = C_4^{1/2}C_6^{3/2}/W$ and

$$\tilde{B}_L(x) = \frac{2}{3}\tilde{B}_T(x) = \frac{2^{1/2}4}{3^{1/4}5} \left(\frac{x}{\xi \ln x}\right)^{1/2} \simeq 0.43 \left(\frac{x}{\xi \ln x}\right)^{1/2} \quad (221)$$

for $x \gg \xi$. Notice that the ratio of transverse to longitudinal fluctuations is always equal to $2\nu + 1$. Of course we do not expect these variational results to be exact,

but there are various arguments, including other more qualitative approaches [12] [8], which indicate that these results are reasonable.

The correlation length ξ is the distance at which fluctuations become equal to a full lattice spacing. Until that length scale, all the atoms see essentially independent potentials, but beyond that length scale, two different atoms can take advantage of the same attractive pinning potential. ξ can actually be quite large, so that for many experiments, the relevant result is intermediate distance regime $1 \ll x \ll \xi$.

Finally, we get to the promised density correlation function $g_{\vec{K}}(\vec{x})$. We find

$$g_{\vec{K}}(\vec{x}) = \exp\left(-\frac{K^2}{2}(\tilde{B}_L(x) \cos^2 \theta + \tilde{B}_T(x) \sin^2 \theta)\right) \quad (222)$$

where θ is the angle between \vec{K} and \vec{x} . Compared to the behavior in a pure system—power law in two dimensions or decay to a constant in three dimensions, the decay of $g_{\vec{K}}(\vec{x})$ is quicker; it is a stretched exponential with

$$g_{\vec{K}}(\vec{x}) \sim e^{-x^{2\nu}} \quad (223)$$

radial behavior. This prediction appears to agree quite well with experimental results [11], with values of ν that are also consistent with our predictions.

For a much more detailed account of these results, the reader is again referred to reference [5]. Some extensions of these ideas to more complicated physical situations have also been worked out [2]. The main message of this lecture is that the replica technology that we have learned about in previous lectures really can be a working tool of physicists, who can thereby make theoretical predictions which can be compared with experiment.

Appendix

Some useful Gaussian integrals:

$$\begin{aligned} \int_{-\infty}^{\infty} \frac{\prod_{i=1}^n dx_i}{\sqrt{(2\pi)^n \det A}} \exp \left(-\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n x_i (A^{-1})_{ij} x_j + \sum_{i=1}^n x_i J_i \right) \\ = \exp \left(\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n J_i A_{ij} J_j \right) \end{aligned} \quad (224)$$

$$\int_{-\infty}^{\infty} \frac{\prod_{i=1}^n dx_i}{\sqrt{(2\pi)^n \det A}} x_i x_j \exp \left(-\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n x_i (A^{-1})_{ij} x_j \right) = A_{ij} \quad (225)$$

$$\int_{-\infty}^{\infty} \frac{dx}{\sqrt{2\pi A}} e^{-x^2/2A} = 1 \quad (226)$$

$$\int_{-\infty}^{\infty} \frac{dx}{\sqrt{2\pi A}} x e^{-x^2/2A} = 0 \quad (227)$$

$$\int_{-\infty}^{\infty} \frac{dx}{\sqrt{2\pi A}} |x| e^{-x^2/2A} = \sqrt{\frac{2A}{\pi}} \quad (228)$$

$$\int_{-\infty}^{\infty} \frac{dx}{\sqrt{2\pi A}} x^2 e^{-x^2/2A} = A \quad (229)$$

$$\int_{-\infty}^{\infty} \frac{dx}{\sqrt{2\pi A}} x^4 e^{-x^2/2A} = 3A^2 \quad (230)$$

Acknowledgements

I thank my long-time friends and collaborators, Jean-Philippe Bouchaud, Antoine Georges, and Marc Mézard, with whom I have enjoyed exploring the issues described in these lectures. I thank Dan Stein for inviting me to give these lectures and the members of the Santa Fe Institute for their hospitality during my visit. My work has been supported by the Harvard Society of Fellows, where I am a Junior Fellow, and by the National Science Foundation, through Grant No. DMR91-15491.

References

- [1] P.W. Anderson. "Reference Frame" articles appearing in *Physics Today* from January, 1988 to March, 1990.
- [2] J.-P. Bouchaud, A. Georges. "Competition between Lattice Pinning and Impurity Pinning: Variational Theory and Physical Realizations." *Phys. Rev. Lett.* 68 (1992): 3908-3911.
- [3] J.-P. Bouchaud, M. Mézard, G. Parisi, and J. S. Yedidia. "Polymers with Long-Ranged Self-Repulsion: A Variational Approach." *J. Phys. A* 24 (1991): L1025-L1030.
- [4] J.-P. Bouchaud, M. Mézard, and J. S. Yedidia. "Some Mean Field Like Equations Describing the Folding of Heteropolymers at Finite Temperature." LPTENS preprint 91/27.
- [5] J.-P. Bouchaud, M. Mézard, and J. S. Yedidia. "Variational Theory for Disordered Vortex Lattices." *Phys. Rev. Lett.* 67 (1991): 3840-3843; "A Variational Theory for the Pinning of Vortex Lattices by Impurities." LPTENS preprint 92/3, to be published in *Phys. Rev. B* (1992).
- [6] J. des Cloizeaux. "The Statistics of Long Chains with Non-Markovian Repulsive Interactions and the Minimal Gaussian Approximation." *J. de Physique* 31 (1970): 715-736.
- [7] S. Edwards and P. W. Anderson. "Theory of Spin Glasses." *J. Phys. F* 5 (1975): 965-974.
- [8] M. Feigel'man, V. B. Geshkenbein, A. Larkin, and V. Vinokur. "A Theory of Collective Flux Creep." *Phys. Rev. Lett.* 63 (1989): 2303-2306.

- [9] R. P. Feynman. "Slow Electrons in a Polar Crystal." *Phys. Rev.* 97 (1955): 660-665.
- [10] A. Georges and J. S. Yedidia. "How to Expand Around Mean-Field Theory Using High-Temperature Expansions." *J. Phys. A* 24 (1991): 2173-2192.
- [11] D. G. Grier, C. A. Murray, C. A. Bolle, P. L. Gammel, D. J. Bishop, D. B. Mitzi, and A. Kapitulnik. "Translational and Bond-Orientational Order in the Vortex Lattice of the High- T_c Superconductor $Bi_{2.1}Sr_{1.9}Ca_{0.9}Cu_2O_{8+\delta}$." *Phys. Rev. Lett.* 66 (1991): 2270-2273.
- [12] T. Halpin-Healy. "Diverse Manifolds in Random Media." *Phys. Rev. Lett.* 62 (1989): 442-445; "Disorder-Induced Roughening of Diverse Manifolds." *Phys. Rev. A* 42 (1990): 711-722.
- [13] M. Mézard and G. Parisi. "Replica Field Theory for Random Manifolds." *J. Physique I* 1 (1991): 809-836.
- [14] M. Mézard and G. Parisi. "Manifolds in Random Media: Two Extreme Cases." LPTENS preprint 92/12 (1992).
- [15] M. Mézard, G. Parisi, and M.A. Virasoro. *Spin Glass Theory and Beyond*. Singapore: World Scientific, 1987.
- [16] D. R. Nelson. In *Phase Transitions and Critical Phenomena*, Vol 7, p. 5, edited by C. Domb and M. S. Green. London: Academic Press, 1983.
- [17] G. Parisi. "Toward a Mean Field Theory for Spin Glasses." *Phys. Lett* 73A (1979): 203-205; "A Sequence of Approximated Solutions to the S-K Model for Spin Glasses." *J. Phys. A* 13 (1980) L115-L121; "Order Parameter for Spin Glasses." *J. Phys. A* 13 (1980): 1101-1112; "Magnetic Properties of Spin Glasses in a New Mean Field Theory" *J. Phys A* 13 (1980): 1887-1895.

- [18] E. I. Shakhnovich and A. M. Gutin. "Frozen States of a Disordered Globular Heteropolymer." *J. Phys. A* 22 (1989): 1647-1659.
- [19] D. Sherrington and S. Kirkpatrick. "Solvable Model of a Spin Glass." *Phys. Rev. Lett.* 35 (1975): 1792-1796.
- [20] D. J. Thouless, P. W. Anderson, and R. G. Palmer. "Solution of 'Solvable Model of a Spin Glass.'" *Phil. Mag.* 35 (1977): 593-601.