Introduction: statistical mechanical formulation of graph partionning, k-SAT and coding

In these notes, after a brief review of the basic principles of statistical mechanics, we define some important models namely: the Ising, Edwards-Anderson and Sherrington-Kirkpatrick models. Finally the statistical mechanical formulation of graph partionning, random k SAT (satisfiability) , random XOR SAT, and LDPC (low density parity check) codes is introduced.

1. Brief review of statistical physics formalism

1.a. Statistical physics is the use of probabilistic approaches for the study of large systems or more precisely systems with a large number N of degrees of freedoms. These systems may be:

- a gas or a fluid in which case the degrees of freedom are the positions and velocities of the constituent atoms (r_i, v_i), $i = 1, ..., N$; (here typically $N = O10^{23}$).
- a magnetic material where the degrees of freedom are "spins" $(s_1, ..., s_N)$, $s_i = \pm 1$;
- a string of information bits $(x_1, ..., x_N)$, $x_i = 0, 1$, that we want to estimate, based on output observations of a channel.
- a random graph or network, ect...

1.b. The basic hypothesis of equilibrium statistical mechanics states that: an isolated system will, after a long enough time, reach a stationary (i.e. time independent) macroscopic state (the "equilibrium state") which can be described by *an appropriate measure* over the space of degrees of freedom (the "phase space"). In other words in order to describe this equilibrium state one can replace Newton's dynamical laws (a complicated set of 10^{23} differential equations) by an appropriate time independent measure over phase space. Loosely speaking, this is the content of the *ergodic hypothesis*.

1.c. How to choose the correct measure ? Dynamics is governed by an energy function or hamiltonian (this will turn out to be a cost function in optimization problems).

$$H(x_1, ..., x_N)$$

which is a function of the N degrees of freedom $(x_1, ..., x_N) \in \Gamma_N$. The discussion below is restricted to the case of discrete variables (spins or bits...). For an isolated system the energy is conserved so whatever the dynamics is, it takes place on the energy surface $\Sigma_E \subset \Gamma_N$ defined by the constraint $E = H(x_1, ..., x_N)$. Since there is no cost in energy for the system to go from one state to another on the energy surface, the most natural measure is the uniform measure on Σ_E,

$$\frac{1}{W_N(E)} \delta(E - H(x_1, ..., x_N)) = \mu_{microcanonical}(x_1, ..., x_N)$$

The normalization factor

$$W_N(E) = \sum_{(x_1,...,x_N)\in\Gamma_N} \delta(E - H(x_1, ..., x_N))$$

is called the "microcanonical partition function". This simply counts the number of allowed states of energy E and, since there is usually an exponential number of them as $N \to \infty$, it is natural to measure it on a logarithmic scale,

$$S_{Boltzmann}(E, N) = \ln W_N(E)$$

This is the famous Boltzmann entropy.

Boltzmann's postulate: The purely combinatorial quantity $S_{Boltzman}(E, N)$ is equal to the entropy of thermodynamics $S_{thermodynamics}(E, N)$ discovered by Carnot, Clausius and others. In particular it follows that the temperature (measured by thermometers) can be computed as $1/T = \frac{\partial S}{\partial E}$ (in appropriate units).

The Gibbs measure. Usually systems of interest are not isolated but in contact with some thermal bath (typically the thermal bath is the environment). Since the total universe (system $\cup$ environment) is isolated it can be described by the microcanonical measure and there is a well defined temperature $T^{-1} = \frac{\partial S}{\partial E}$. It is in principle possible to compute the induced measure (the marginal) on the degrees of freedom of the system of interest. If the environment is supposed to be much bigger than the system of interest, one finds for the induced measure

$$\mu_\beta(x_1, ..., x_N) = \frac{1}{Z_N(\beta)} \exp(-\beta H(x_1, ..., x_N))$$

where $\beta = T^{-1}$ is the inverse temperature and the normalisation factor is called the "Partition Function",

$$Z_N(\beta) = \sum_{(x_1,...,x_N)\in\Gamma_N} \exp(-\beta H(x_1, ..., x_N))$$

Note that in optimisation problems where the hamiltonian is some cost function that one wants to minimise there is no "environment" or "temperature" but the above is still a natural measure to study for large β ("low temperature") because it concentrates on parts of Γ_N which are close to the minimum cost. This measure is called the Gibbs measure.

It follows from Boltzmann's postulate that

$$F(\beta, N) = -\frac{1}{\beta} \ln Z_N(\beta)$$

is equal to the thermodynamic free energy.

1.d. Equivalence of ensembles. The free energy and the entropy are related by a Legendre transform and as a consequence the choice of the microcanonical or Gibbs measure

is in principle a matter of convenience. This "equivalence of ensembles" suffers however from exceptions depending on the physical or mathematical situation. But here is the heuristics:

$$Z_N(\beta) = \sum_{(x_1,...,x_N)\in\Gamma_N} \exp(-\beta H(x_1,...,x_N)) = \sum_E W_N(E) e^{-\beta E} = \sum_E e^{N(\frac{S(E,N)}{N} - \beta\frac{E}{N})}$$

The Laplace (or saddle point method or large deviation principle) then implies that

$$\lim_{N\to\infty} \frac{1}{N} F(\beta, N) = -\frac{1}{\beta} \max_e(s(e) - \beta e) = \min_e(e - \beta^{-1} s(e))$$

where $s(e) = \lim_{N\to\infty, \frac{E}{N}=e} S(E,N)$ is the entropy per degree of freedom, and $e = \frac{E}{N}$ the energy per degree of freedom.

In well behaved systems (i.e. where interactions are local) one can show that the entropy is a concave function of e. When this is true the above Legendre transform can be inverted so that

$$s(e) = \min_\beta(e\beta - \beta f(\beta))$$

where $f(\beta) = \lim_{N\to\infty} \frac{1}{N} F(\beta, N)$.

The limits $N \to \infty$ are called thermodynamic limits and in statistical physics constitute an excellent idealisation of reality where $N = 10^{23}$. Finite size corrections are also a relevant and important subject.

2. A few basic models of spin systems.

2.a. Collection of free spins

The degrees of freedom are $s_i = \pm 1$. A configuration or a state is $(s_1, ..., s_N)$. The phase space or configuration space is $\Gamma_N = \{-1, +1\}^N$. Suppose the spins do not interact between themselves, but are submitted to an external magnetic field h (a real number). Then their Hamiltonian or energy function reads

$$H(s_1, ..., s_N) = -h \sum_{i=1}^{N} s_i$$

The Gibbs measure takes a factorized form,

$$\mu_\beta(s_1, ..., s_N) = \frac{1}{Z_N(\beta)} \prod_{i=1}^{N} e^{\beta h s_i}$$

where

$$Z_N(\beta) = (2 \cosh \beta h)^N$$

and the free energy in the thermodynamic limit is

$$f(\beta) = -\beta^{-1} \ln(2 \cosh \beta h)$$

The marginal distribution function of one spin (or "one point function") is

$$\mu^{(1)}(s) = \frac{e^{\beta h s}}{2 \cosh \beta h}$$

¿From this marginal one can compute an important quantity: the magnetisation defined as the expected or average value of a spin

$$m = \langle s \rangle_\beta = \tanh \beta h$$

Note that the log-likelyhood parameter of this distribution is

$$\ln \frac{\mu^{(1)}(s = +1)}{\mu^{(1)}(s = -1)} = 2\beta h$$

So basically one can think of the magnetic field as a log-likelihood.

One can also compute the Boltzmann entropy directly from its definition or by taking the Legendre transform of $\beta f(\beta)$. Both calculations lead to

$$s(e) = \ln 2 - \frac{1}{2}(1 + \frac{e}{h}) \ln(1 + \frac{e}{h}) - \frac{1}{2}(1 - \frac{e}{h}) \ln(1 - \frac{e}{h})$$

the binary entropy function.

2.b. Interacting spins

The spins s_i are attached to the N vertices $i \in V$ of a graph. To each edge $(ij) \in E$ of the graph we associate a real number J_{ij} measuring the strength of the interaction. The interaction energy between two spins connected by an edge is

$$-J_{ij}s_is_j$$

The hamiltonian is (if we also include a set of magnetic fields h_i)

$$H(s_1, ..., s_N) = -\sum_{(ij)\in E} J_{ij}s_is_j - \sum_{i\in V} h_is_i$$

A basic problem of statistical mechanics is to study the Gibbs measures associated to this class of Hamiltonians for different graphs (V, E) and values of the parameters β, J_{ij}, h_i. Stated in this way the problem is much too general, and we have to restrict ourselves to more specific models.

The Ising model. Here the graph has a spatial structure and represent the underlying crystalline structure of the solid. The spins are the magnetic moments of the atoms located at the crystalline sites. The simplest case is that of a cubic graph $\Lambda \subset \mathbf{Z}^d$ (d the dimension of space). The edges are simply the nearest neighbourd bonds defined by $|i - j| = 1$, for $i, j \in Z^d$. The interaction and magnetic field are uniform $J_{ij} = J$, $h_i = h$. So the hamiltonian becomes

$$H(s_1, ..., s_N) = -J\sum_{|i-j|=1} s_is_j - h\sum_{i=1}^{N} s_i$$

For $J > 0$ and $h = 0$ there are two ground states (lowest energy configurations) $s_i = +1$ for all i and $s_i = -1$ for all i. This is called the ferromagnetic model because the alignment of spins has something to do with ferromagnets. For $J < 0$ and $h = 0$ on the other hand the ground states consist of the two possible chessboard configurations $s_i = \pm(-1)^{|i|}$. The model is called anti-ferromagnetic because this is observed in so called anti-ferromagnets.

Computing the marginals, the magnetisation and more generaly studying the Gibbs measure of the Ising model is a non trivial and much studied problem. Much is known but some questions are still open despite more than seventy years of effort.

The Edwards-Anderson model.

For various reasons some materials have inhomogeneous coupling constants J_{ij}. Moreover these may vary from sample to sample. One then models this situation by assuming that for a given sample the set of couplings J_{ij}, $(ij) \in E$, is a realisation of i.i.d random variables. Popular models include a Gaussian distribution

$$P(J_{ij}) = \frac{1}{\sqrt{2\pi V^2}} \exp -\frac{J_{ij}^2}{2V^2}$$

or

$$P(J_{ij}) = \frac{1}{2}\delta(J_{ij} - J) + \frac{1}{2}\delta(J_{ij} + J)$$

The Gibbs measure, the free energy and other quantities are now random quantities. For a given sample the realisation is fixed ("quenched" in the language of physicist) so that one of the relevant objects is the free energy for the given realisation

$$F(\beta, N) = -\frac{1}{\beta} \ln Z_N(\beta)$$

Obviously this is very hard to compute, but fortunately it turns out that this quantity is often self averaging, i.e as $N \to \infty$ its probability distribution concentrates on its average. More precisely, with probability one

$$\lim_{N\to\infty} \frac{1}{N} F(\beta, N) = \lim_{N\to\infty} \mathbf{E}\big[\frac{1}{N} F(\beta, N)\big]$$

Therefore one has to devise methods to compute the r.h.s, a still very difficult task because the expectation over the quenched variables is outside of the logarithm. At least an upper bound is provided by Jensen's inequality

$$\mathbf{E}\big[\frac{1}{N} \ln Z_N(\beta)\big] \leq \frac{1}{N} \ln \mathbf{E}\big[Z_N(\beta)\big]$$

The right hand side is called an annealed average because it corresponds to a partition function where the J_{ij} behave as degrees of freedom instead of frozen (quenched) variables. The replica method, as we will see, is a method to compute the left hand side (the quenched average).

The infinite range Ising model

In the above models the couplings connect nearest neighbords. Although such short range couplings are a realistic modelling, the models are often intractable. An interesting simplification is to abandon the spatial structure and assume that the graph (V, E) is complete. So every pair of points is connected. The hamiltonian of the long range Ising model is

$$H(s_1, ..., s_N) = -\frac{J}{N} \sum_{i<j} s_i s_j - h \sum_{i=1}^{N} s_i$$

Here the coupling constant is scaled by N^{-1} because we want the energy to be an extensive quantity $O(N)$ in order to have a sensible thermodynamic limit. This model is easily exactly solvable and one learns a lot from it. We will go through the solution later. *A discussion of it can be found in chap 1 of Nishimori's book.*

The Sherrington-Kirkpatrick model

We now move on to a ground that is related to the models arising in optimisation, coding, ect... The Sherrington-Kirkpatrick model is simply obtained from the Edwards-Anderson model by removing the spatial structure, i.e: again one defines the model on a

complete graph. In the standard SK model the distribution of couplings are i.i.d gaussian with zero mean

$$P(J_{ij}) = \frac{1}{\sqrt{2\pi V^2}} \exp - \frac{J_{ij}^2}{2V^2}$$

and the Hamiltonian is

$$H(s_1, ..., s_N) = -\sum_{i<j} \frac{J_{ij}}{\sqrt{N}} s_i s_j - h \sum_{i=1}^{N} s_i$$

Note that here the correct scaling involves a $\sqrt{N}$.

We will go through (a least part of) the "replica method", in order to solve this model, which was developped by Edwards, Anderson and Parisi. Another method giving the same results is the "cavity method" and was developped later on by Mezard and Parisi (following work of Thouless, Anderson and Palmer; the TAP approach). A major open problem for 25 years was to decide if the final expression (called the Parisi solution) for the free energy is exact or not. Note that the theory also provides much insight into the structure of the Gibbs measure (see the books of Nishimori and Mezard, Parisi, Virasoro).

These methods seem to apply to problems of coding and optimization. In fact:

- The computation of cost functions can be done by methods developed for computing the free energy.

- density evolution equations in LDPC codes are kinds of complicated replica equations.

- belief propagation algorithms are intimately related to the cavity method with "replica symmetry".

- survey propagation is also related to the cavity method with "replica symmetry breaking".

Recently progress has been made by Guerra and Talagrand and it is now known that the Parisi formula for the free energy is exact. (who developped the "interpolation methods" *Quadratic replica coupling in the Sherrington-Kirkpatrick mean field spin glass model , F. Guerra, F. L. Toninelli, J. Math. Phys; Broken replica symmetry bounds in the mean field spin glass model, F. Guerra, Comm. Math. Phys. 233, 1-12 (2003)*). However the structure of the Gibbs measure is still not fully understood at the rigorous level.

The diluted Sherrington-Kirkpatrick model (also called Bethe spin glass by some authors)

In the SK model the number of neigbords of a site is $O(N)$. This makes things simpler because the law of large number is at work. Yet, it is more realistic to have a finite number of neighbords. A way to do this, while abandonning the spatial structure, is to define the model on an Erdos-Renyi random graph: each bond (ij) of the complete graph is kept with probability $p = k/N$ and removed with probability $1 - p = 1 - k/N$. The average number of neighbords of a given site is $pN = k$. So if $k = O(1)$ this number is finite, while

if $k = O(N)$ we are basically back to the situation of the SK model. The model is

$$H(s_1, ..., s_N) = -\sum_{i<j} \frac{J_{ij}}{\sqrt{k}} s_i s_j - h \sum_{i=1}^{N} s_i$$

with a diluted bond distribution

$$P(J_{ij}) = \frac{k}{2N}\delta(J_{ij} - J) + \frac{k}{2N}\delta(J_{ij} + J) + (1 - \frac{k}{N})\delta(J_{ij})$$

It turns out that the replica and cavity methods become much more involved technically, but can be applied (*The Bethe spin glass revisited , M. Mezard, G. Parisi, Eur Phys J B 20, 217 (2001); The cavity method at zero temperature, M. Mezard, G. Parisi, arXiv cond-mat/0207121*).

In the LDPC codes and in the random k-SAT problem the models are of the type discussed here with the additional complication that the couplings couple more than two spins !

3. Statistical mechanical formulation of partioning, SAT and coding

3.a. Graph Partitioning

This is a well known NP complete problem. An instance of the problem is a graph with a set V of N vertices and some set of edges E. A configuration is a partition of the set of vertices $V = U \cup D$ in two disjoint and equal sets $|U| = |D|$. The cost function is $C =$ *the total number of edges connecting* U *to* D.

Statistical mechanical methods apply to the randomised problem (we do not look at a worst case analysis) where the instance (the graph) is drawn in an ensemble according to some probability distribution. The replica method for the SK model turns out to be well suited when the ensemble of graphs is the Erdos-Renyi random graph model.

The configuration space can be described by spins $s_i = +1$ for $i \in U$ and $s_i = -1$ for $i \in D$ with the constraint

$$\sum_{i=1}^{N} s_i = 0$$

An instance is specified by a matrix $J_{ij} = 0$ if $(ij) \notin E$ and $J_{ij} = 1$ if $(ij) \in E$. The cost of a configuration or of a partition is

$$C_N = \sum_{i<j} \frac{J_{ij}}{2}(1 - s_i s_j)$$

This is nothing else but the hamiltonian of the diluted SK model (up to some additive term). The probability distribution for the iid couplings is

$$P(J_{ij}) = p\delta(J_{ij} - 1) + (1 - p)\delta(J_{ij})$$

For a given instance the minimum cost can be found as

$$\min C_N = \lim_{\beta\to\infty} \frac{1}{\beta} \ln Z_N(\beta)$$

This quantity is expected to be selfaveraging and its average $\mathbf{E}\big[\ln Z_N(\beta)\big]$ has been computed by the replica method (*Application of statistical mechanics to NP complete problems in combinatorial optimisation, Y. Fu, P. W. Anderson, J. Phys. A: Math Gen 19 (1986) 1605-1620*) when $p = O(1)$,

$$\min C = \frac{N^2}{4} p + N^{3/2} \frac{E_0}{2} \sqrt{p(p-1)}$$

where $E_0 = -0.7633$ is the ground state energy ($\beta = \infty$) of the SK model computed from the Parisi solution. It seems that this result is in agreement with numerical simulations.

Random k-SAT problem

The satisfiability problem involves N boolean variables $x_i = 0$ or 1. The phase space is restricted by M constraints (clauses) of the form ($\bar{x}_{i_1}$ or x_{i_2} or ... or x_{i_k}) that we want to be all satisfied (SAT). An instance is a logical formula of the type

$$F = (x_1 or \bar{x}_2 or x_8) and (\bar{x}_3 or x_5 or \bar{x}_4) and (x_5 or x_8 or x_6)$$

In this example $k = 3$, $M = 3$ and $N = 8$. The problem is SAT if $F = 1$ has a solution and UNSAT if $F = 1$ has no solution. An instance can be specified by a Tanner graph (or factor graph as follows. Put N variable nodes on a vertical line and M check nodes on another vertical line. Connect variable node i to check node c if the variable x_i belongs to the clause c.

The cost function (or hamiltonian) can be defined as *the number of unsatisfied clauses.* For $k = 2$ it is known that this is a polynomial problem while for $k \geq 3$ it is NP complete.

The methods of statistical mechanics are useful if one deals with the randomized version of satisfiability:

- the variables are negated or not negated with probability $\frac{1}{2}$.

- the instances are chosen among an ensemble.

One possible ensemble is provided by the random Tanner graphs: fix the number of edges coming out of a check node to k and connect them uniformly over the variable nodes $1, ..., N$. This model has been approached rigorously by the interpolation methods of Guerra (*Replica bounds for optimisation problems and diluted spin systems, S. Franz, M. Leone, J. Stat. Phys. 111, 535-564 (2003)*).

The formalism can be set up as follows. First we go to the spin language by setting $(-1)^{x_i} = s_i$. A given clause c is unsatisfied by a choice of configuration $(s_1, ..., s_N)$ iff

$$\prod_{l \in \partial c} \frac{1}{2}(1 + J_l^c s_l) = 1$$

where $J_l^c = -1$ $(+1)$ if the variable x_l is negated (not negated) in clause c. The $J_l^c = +1$ are i.i.d with probability $\frac{1}{2}$. The cost function is therefore

$$H(s_1, ..., s_N) = \sum_{c=1}^{M} \prod_{l \in \partial c} \frac{1}{2}(1 + J_l^c s_l)$$

This is the hamiltonian of a complicated diluted spin system where the couplings couple k spins (instead of 2 in the graph partitionning) The replica or cavity or interpolation methods can give us some information on the ground state energy through a computation of

$$\lim_{\beta \to \infty} \mathbf{E} \frac{1}{\beta} \big[\ln Z_N(\beta) \big]$$

where

$$Z_N(\beta) = \sum_{s_1 ... s_N} \exp(-\beta H(s_1, ..., s_N))$$

and the expectation is over the "frozen disorder" i.e: all random Tanner graphs and over random J's. Moreover cavity equations provide insight into the developpement of new algorithms to solve particular instances. This has led to the so called survey propagation algorithm (*Random K-satisfiability problem: from an analytic solution to an efficient algorithm, M. Mezard, R. Zecchina , Phys Rev E 66, 056126-1 (2002)*).

3.c. Random XOR SAT problem

This problem is much simpler than the previous one from the computational point of view. Its formulation is quite similar: the logical clauses are replaced by constraints of the form

$$x_1 \oplus x_2 \oplus x_8 = \epsilon_1$$

$$x_3 \oplus x_5 \oplus x_4 = \epsilon_2$$

$$x_5 \oplus x_8 \oplus x_6 = \epsilon_3$$

where the i.i.d $\epsilon_i = 0, 1$ with probability $\frac{1}{2}$. The set of constraints is drawn from an ensemble of Tanner graphs as before. Going to the spin language $s_i = (-1)^{x_i}$ and $J_c = (-1)^{\epsilon_c}$, the cost function (the number of unsat equations) is

$$H(s_1, ..., s_N) = \sum_{c=1}^{M} (1 - J_c \prod_{l \in \partial c} s_l)$$

For $\beta = +\infty$ the associated Gibbs measure can be written as

$$\mu_{\beta=\infty}(s_1, ..., s_N) = \frac{1}{Z_N(\infty)} \prod_{c=1}^{M} \frac{1}{2}(1 + J_c \prod_{l \in \partial c} s_l)$$

In statitical physics this is called the diluted k spin glass model. The coupling is between k spins, the frozen disorder is the graph and the J's. A recent interesting paper is (*Alternative solutions to diluted p-spin models and XORSAT problems, M. Mezard, F. Ricci-Tersenghi, R. Zecchina , J. Stat. Phys 111, 505-533 (2003)*).

4.d. Coding

The translation of the channel coding problem into a random spin glass problem was originally developped by N. Sourlas. Here we do not follow his original approach but the spirit is the same. Assume communication through a noisy symmetric and memoryless channel with transition probability $p(y|x)$ where x is a binary input bit and y the output belonging to some alphabet. We will suppose that the a priori distribution of the source is uniform and that the input is a code word of some linear code C. So the input has distribution

$$p(x_1, .., x_n) = \frac{1}{|C|} 1_C(x_1, ..., x_N)$$

Given an observed output string $y_1, ..., y_N$, we want to estimate the original input binary string of bits $x_1, ..., x_N$. The ideal decoder uses a maximum likelihood estimate for the input bit:

$$\hat{x}_i = \arg\max \sum_{x_1, ..., x_{i-1}, x_{i+1}, ..., x_N} p(x_1, ..., x_N | y_1, ..., y_N)$$

We now show that for linear low density parity check codes this is intimately related to the magnetisation of a random spin system.

¿From the Bayes formula

$$p(x_1, ..., x_N | y_1, ..., y_N) = \frac{p(y_1, ..., y_N | x_1, ..., x_N) p(x_1, ..., x_N)}{\sum_{x_1, ..., x_N} p(y_1, ..., y_N | x_1, ..., x_N) p(x_1, ..., x_N)}$$

For a memoryless channel we rewrite this as

$$p(x_1, ..., x_N | y_1, ..., y_N) = \frac{1}{Z} \prod_{1=1}^{N} p(y_i | x_i) 1_C(x_1, ..., x_N)$$

where the normalisation is

$$Z = \sum_{x_1,...,x_N} \prod_{1=1}^{N} p(y_i|x_i) 1_C(x_1, ..., x_N)$$

In the case of a LDPC code the constraints over $x_1, ..., x_N$ may be defined from a Tanner graph as in the XOR SAT problem except that now all the ϵ's on the r.h.s are equal to 0. So going to the spin language $s_i = (-1)^{x_i}$ we have

$$1_C(s_1, ..., s_N) = \prod_{1}^{M} \frac{1}{2}(1 + \prod_{i \in \partial c} s_i)$$

Here M is the number of linear constraints, so the design rate of the code is $R = (N - M)/N = 1 - N/M$. If the degree of the check nodes is k and the degree of the variable nodes is l, we get $R = 1 - l/k$; the number of edges of the graph is $O(N)$ if k and l are $O(1)$.

Let us now write the transition probability of the channel in a form that is more suggestive for statistical physicist. First note (with $s = (-1)^x$)

$$p(y|x) = \frac{1}{2}(p(y|0) + p(y|1)) + \sigma \frac{1}{2}(p(y|0) - p(y|1))$$

Setting

$$h(y) = \frac{1}{2} \ln \frac{p(y|0)}{p(y|1)}$$

it is easy to check that

$$p(y|x) = \frac{1}{2} p(y|0) \frac{1 + e^{-2h(y)}}{\cosh h(y)} \exp(sh(y))$$

Gathering these formulas we obtain

$$p(s_1, ..., s_N|y_1, ..., y_N) = \frac{1}{Z_N} \prod_{1}^{M} \frac{1}{2}(1 + \prod_{i \in \partial c} s_i) \exp(\sum_{i}^{N} h(y_i) s_i)$$

This is the Gibbs measure of a diluted random spin system. It is diluted because the graph is sparse. Randomness (frozen or quenched disorder) appears in two ways:

- the Tanner graph which defines the code is drawn from an ensemble (to be chosen in a judicious way).
- the observations y_i at the output are random; so the magnetic fields $h_i = h(y_i)$ are i.i.d random variables.

What is the distribution of magnetic fields ? We must keep in mind that the output $y_1, ..., y_N$ is the result of a particular input signal $x_1^{input}, ..., x_N^{input}$.

$$p(h_i)dh_i = p(y_i|x_i^{input})dy_i$$

It should be clear that, if we assume the channel is symmetric (which we have not yet used), it is not a loss of generality to take $x_1^{input}, ..., x_N input = 0, ..., 0$. (So correct decoding means that the estimated bit must be equal to zero). For a binary input additive white noise gaussian channel one then finds

$$p(h)dh = \frac{dh}{\sqrt{2\pi\sigma}} \exp(-\frac{(h-\sigma)^2}{2\sigma})$$

where σ is the intensity of the noise.

In order to compute the estimate of the i th bit we have to look at the marginal $\mu^{(1)}(s_i)$ and compute

$$\arg\max \mu^{(1)}(s_i) = sgn\langle s_i \rangle$$

where $\langle s_i \rangle = \mu^{(1)}(+1) - \mu^{(1)}(-1)$ is the magnetisation of spin i. (Recall $\langle - \rangle$ is the average with respect to the Gibbs measure. The fraction of errors is

$$P_{bit} = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{2}(1 - sgn\langle s_i \rangle)$$

If we assume it is selfaveraging then one may compute the quenched average.

$$\mathbf{E}(sgn\langle s_i \rangle)$$

Remark here that the presence of the sgn function may be a nuisance if one is to apply methods of statistical mechanics.

It is important to note that there is no parameter β, or more precisely $\beta = 1$. So ML bit decoding is done with the Gibbs measure at $\beta = 1$. However it is fruitful to "expand our view" by introducing β in front of $h(y)$, and study the Gibbs measure in the whole plane (β, σ). For example block decoding corresponds to finding to block that maximises $p(x_1, ..., x_N | y_1, ..., y_N)$. This means minimising the "Hamiltonian" or looking at the measure for $\beta \to \infty$. One finds for the block error probability,

$$P_{block} = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{2}(1 - \langle s_i \rangle(\beta = \infty))$$

Finally it is also possible to relate the conditional Shannon entropy $H(X_1, ..., X_N | Y_1, ..., Y_N)$ with the magnetisation $\langle s_i \rangle_{\beta=1}$ and the correlation function $\langle s_i s_j \rangle_{\beta=1} - \langle s_i \rangle_{\beta=1} \langle s_j \rangle_{\beta=1}$. For a gaussian channel

$$\frac{d}{d\sigma} H(X_1, ..., X_N | Y_1, ..., Y_N) = \frac{1}{2} \sum_{i=1}^{N} (1 - \mathbf{E}\langle s_i \rangle_{\beta=1})$$

and

$$\frac{d^2}{d\sigma^2} H(X_1, ..., X_N | Y_1, ..., Y_N) = \frac{1}{2} \sum_{i,j=1}^{N} \mathbf{E}\big[(\langle s_i s_j \rangle_{\beta=1} - \langle s_i \rangle_{\beta=1} \langle s_j \rangle_{\beta=1})^2\big]$$

These two formulas are non trivial because they rely on a special symmetry (called a gauge symmetry) of the spin glass valid on the special line ($\beta = 1, \sigma$) (called the Nishimori line in the spin glass context).

An important problem is the relationship between iterative and maximum likelihood decoding (*The dynamic phase transition for decoding algorithms, S. Franz, M. Leone, A. Montanari, F. Ricci-Tersenghi, Phys. Rev. E 66, 046120 (2002)*).

Another theme is the derivation of bounds on the magnetisation. Correlation inequalities like the Griffith inequality seem to apply (*in Phase Transitions and Critical Phenomena vol 1 ed Domb and Green (1972) (London Academic) Griffiths inequalities for the gaussian spin glass , S. Morita, H. Nishimori, P. Contucci, arXiv cond-mat/0403625*).