The Random Energy Model in a Magnetic Field and Joint Source–Channel Coding

Neri Merhav

Department of Electrical Engineering
Technion - Israel Institute of Technology
Haifa 32000, ISRAEL

Abstract

We demonstrate that there is an intimate relationship between the magnetic properties of Derrida's random energy model (REM) of spin glasses and the problem of joint source–channel coding in Information Theory. In particular, typical patterns of erroneously decoded messages in the coding problem have "magnetization" properties that are analogous to those of the REM in certain phases, where the non–uniformity of the distribution of the source in the coding problem, plays the role of an external magnetic field applied to the REM. We also relate the ensemble performance (random coding exponents) of joint source–channel codes to the free energy of the REM in its different phases.

Keywords: spin glasses, REM, phase transitions, magnetization, information theory, joint source–channel codes.

1 Introduction

In the last few decades it has become apparent that many problems in Information Theory, and coding problems in particular, can be mapped onto (and interpreted as) analogous problems in the area of statistical physics of disordered systems, most notably, spin glass models. Such analogies are useful because physical insights, as well as statistical mechanical tools and analysis techniques (like the replica method), can be harnessed in order to advance the knowledge and the understanding with regard to the information–theoretic problem under discussion (and conversely, information–theoretic approaches to problems in physics may sometimes prove useful to physcists as well). A very small, and by no means exhaustive, sample of works along this line includes references [1]–[25].

In particular, Sourlas [8],[9] was the first to observe that there are strong analogies and parallisms between the behavior of ensembles of error correcting codes and certain spin glass models with quenched parameters, like the p–spin glass model and Derrida's random energy model (REM) [26],[27],[28] at least as far as the mathematical formalism goes. In particular, the REM is an especially attractive model to adopt in this context, as it is, on the one hand, exactly solvable, and on the other hand, rich enough to exhibit phase transitions. As noted in [5, Chap. 6] and [16], ensembles of error correcting codes 'inherit' these phase transitions from the REM when viewed as physical systems whose phase diagram is defined in the plane of the coding rate vs. decoding temperature. In [29] this topic was further investigated and ensemble performance figures of error correcting codes (random coding exponents) were related to the free energies in the various phases of the phase diagram.

While the above–described relation takes place between *pure* channel coding and the REM *without* any external magnetic field, in this work, we demonstrate that there are also intimate relationships between *combined source/channel coding* and the REM *with* such a magnetic field. In particular, it turns out that typical patterns of erroneously decoded messages in the source/channel coding problem have "magnetization" properties that are analogous to those of the REM in certain phases, where the non–uniformity of the distribution of the source in the joint source–channel coding system, plays the role of an external magnetic field applied to the spin glass modeled by the REM. We also relate the ensemble performance (random coding exponents) of joint source–channel codes to the free energy of the REM in its different phases.

The outline of this paper is as follows. In Section 2, we provide some background, both on the information theoretic aspect of this work, which is the problem of joint source channel coding, and the statistical mechanical aspect, which is the REM and its magnetic properties. In Section 3, we present the phase diagram pertaining to finite–temperature decoding of an ensemble of joint source–channel codes and characterize the free energies in the various phases. Finally, in Section 4, we derive random coding exponents pertaining to this emsemble and demonstrate their relationships to the free energies.

2 Background

In this section, we give some very basic background which will be needed in the sequel. In Subsection 2.1, we provide a brief overview of Shannon's fundamental coding theorems, the skeleton of Information Theory: The source coding theorem, the channel coding theorem, and finally the joint source–channel coding theorem. In Subsection 2.2, we review a few models of spin glasses, with special emphasis on the REM.

2.1 Information Theory

2.1.1 Source Coding

Suppose we wish to compress a sequence of N bits, $(u_1, u_2, \ldots, u_N)$, drawn from a stationary memoryless binary source, i.e., each bit is drawn independently, where $\Pr\{u_i = 1\} = 1 - \Pr\{u_i = 0\} = q$. Shannon's *source coding theorem* (see, e.g., [30, Chap. 5]) tells that if we demand that the source sequence would be perfectly reconstructable from the compressed data, then the best achievable compression ratio (i.e., the smallest average ratio between the compressed message length and the original source message length – N), at the limit of large N, is given by the entropy of the source, which in the binary memroyless case considered here, is given by:

$$h(q) = -q \log_2 q - (1 - q) \log_2 (1 - q).$$

Many practical coding algorithms are known to achieve $h(q)$ asymptotically, e.g., Huffman coding, Shannon coding, arithmetic coding, and Lempel–Ziv coding, to name a few [30].

2.1.2 Channel Coding

Shannon's celebrated *channel coding theorem* (see, e.g., [30, Chap. 7]) is about reliable transmission of digital information across a noisy channel: Suppose we wish to transmit a binary messsage of k bits, indexed by m $(0 \leq m \leq 2^k - 1)$, through a noisy binary symmetric channel, which flips the transmitted bit with probability p or conveys it unaltered, with probability $1 - p$. If we wish to convey the message via the channel reliably (i.e., with very small probability of error), then before we transmit the message via the channel, we have to encode it, i.e., map it in a sophisticated manner into a longer binary message of length n $(n \geq k)$ and then transmit the encoded message

$\boldsymbol{x}(m) = (x_1(m), \ldots, x_n(m))$. The ratio $R = k/n$ is called the *coding rate*. It measures how efficiently the channel is used, i.e., how many information bits are conveyed per one channel use. The corresponding channel output sequence, $\boldsymbol{y} = (y_1, \ldots, y_n)$ (with some of the bits flipped by the channel), is received at the decoder.

The optimum decoder, in the sense of minimum probability of error, estimates the message m by the *maximum a–posteriori* (MAP) decoder, i.e., it selects the message $0 \leq m \leq 2^k - 1$ which maximizes posterior probability given $\boldsymbol{y}$, that is, $P(m|\boldsymbol{y})$, or equivalently, it maximizes the product $P(m)P(\boldsymbol{y}|\boldsymbol{x}(m))$, where $P(m)$ the prior probability of message m and $P(\boldsymbol{y}|\boldsymbol{x}(m))$ is the conditional probability of the observed $\boldsymbol{y}$ given that $\boldsymbol{x}(m)$ was transmitted. In the important special case where all messages are *a-priori* equiprobable, that is, $P(m) = 2^{-k}$ for all m, the MAP decoding rule boils down to the maximization of $P(\boldsymbol{y}|\boldsymbol{x}(m))$, which is the *maximum likelihood* (ML) decoding rule.

Channel capacity C is defined as the supremum of all coding rates R for which there still exist encoders and decoders which make the probability of error arbitrarly small provided that n is large enough (keeping $R = k/n$ fixed). Shannon's channel coding theorem provides a formula of the channel capacity, which in the binary case considered here, is given by

$$C = 1 - h(p) = 1 + p \log_2 p + (1 - p) \log_2(1 - p).$$

One of the mainstream efforts in the Information Theory literature has evolved around devising practical coding and decoding schemes, in terms of computational complexity and storage, with rates close to capacity.

2.1.3 Joint Source–Channel Coding

Finally, we consider the problem of *joint source–channel coding* (see, e.g., [30, Sect. 7.13]): Suppose we have a binary memoryless source, as in the first paragraph above, and a binary memoryless channel, as in the second paragraph above. We assume that by the time that the source generates N symbols, the channel can transmit $n = N\theta$ bits ($\theta \geq 0$ is fixed).

A joint source–channel code maps the source sequence $\boldsymbol{u} = (u_1, \ldots, u_N)$ of length N into a channel input sequence $\boldsymbol{x}(\boldsymbol{u})$ of length n. The decoder, that receives the channel output vector $\boldsymbol{y}$, estimates $\boldsymbol{u}$ either by the *symbol MAP* decoder, which minimizes the symbol error probability (or the bit error probability) or the *word MAP* decoder, which as mentioned earlier, minimizes the word error probability. The word MAP decoder works similarly to the above described MAP decoder for a channel code: It estimates the source sequence as a whole by seeking the vector $\boldsymbol{u}$ that maximizes $P(\boldsymbol{u})P(\boldsymbol{y}|\boldsymbol{x}(\boldsymbol{u}))$, where $P(\boldsymbol{u})$ is the probability of the source vector $\boldsymbol{u}$. The symbol MAP decoder, on the other hand, estimates each bit u_i of the source separately by seeking the symbol $u \in \{0,1\}$ that maximizes $\Pr\{u_i = u, \boldsymbol{y}\} = \sum_{\boldsymbol{u}:\ u_i=u} P(\boldsymbol{u})P(\boldsymbol{y}|\boldsymbol{x}(\boldsymbol{u}))$, $i = 1,\ldots,N$.

These two decoders can be thought of as two special cases of a more general class of decoders, referred to as *finite–temperature decoders* [18]. A finite–temperature decoder estimates the i–th symbol u_i by

$$\hat{u}_i = \text{argmax}_{u\in\{0,1\}} \sum_{\boldsymbol{u}:\ u_i=u} [P(\boldsymbol{u})P(\boldsymbol{y}|\boldsymbol{x}(\boldsymbol{u}))]^{\beta},$$

where the parameter β can be thought of as an inverse temperature parameter. The choice $\beta = 1$ corresponds to the symbol MAP decoder, whereas $\beta \to \infty$ gives us the word MAP decoder [5, Chap. 6].

The joint source–channel coding theorem asserts that a necessary and sufficient condition for the existence of codes, that for large enough n and N (with $\theta = n/N$ fixed), $\boldsymbol{u}$ can be decoded with aribrarily small probability of error (both wordwise and symbolwise) is given by

$$h(q) \leq \theta C. \tag{1}$$

One approach to achieve reliable communication, whenever this condition holds, is to apply separate source coding and channel coding: First compress the source to essentially $h(q)$ bits per symbol, resulting in a binary compressed message of length about $Nh(q) = nh(q)/\theta$ bits, as described in the first paragraph above, and then use a reliable channel code of rate $R = h(q)/\theta \leq C$ to convey the compressed message, as described in the second paragraph. The decoder will first decode the message by the corresponding channel decoder and then decompress the resulting message. Another approach is to map $\boldsymbol{u}$ directly to a channel input vector $\boldsymbol{x}(\boldsymbol{u})$. It can be shown [31, Exercise 5.16, p. 534] that by a random selection of a code from the uniform ensemble (i.e., by generating each codeword $\boldsymbol{x}(\boldsymbol{u})$, $\boldsymbol{u} \in \{0,1\}^N$, independently by a sequence of n fair coin tosses), the average probability of error, over this ensemble of codes, tends to zero as the block length goes to infinity, as long as the above necessary and sufficient condition holds.

2.2 The REM

Consider a spin glass with n spins, designated by a binary vector $\sigma = (\sigma_1, \ldots, \sigma_n)$, $\sigma_i \in \{-1, +1\}$, $i = 1, 2, \ldots, n$. The simplest model of this class is that of a *paramagnetic* solid, namely, the one where the only effect is that of the external magnetic field H, whereas the effect of interactions is negligible (cf. [32, Chap. 3]). Assuming that the spin directions are all either parallel or antiparallel to the direction of the external magnetic field, the energy associated with a configuration σ is given (in the appropriate units) by:

$$\mathcal{E}(\sigma) = -H \sum_{i=1}^{n} \sigma_i,$$

which means (according to the Boltzmann distribution) that each spin is independently oriented upward (+1) with probability $e^{\beta H}/[2\cosh(\beta H)]$ or downward (-1) with probability $e^{-\beta H}/[2\cosh(\beta H)]$. This means that the average (net) magnetic moment is

$$m \triangleq (+1) \cdot \frac{e^{\beta H}}{2\cosh(\beta H)} + (-1) \cdot \frac{e^{-\beta H}}{2\cosh(\beta H)} = \tanh(\beta H) \tag{2}$$

and so the average internal energy per particle is $\bar{E} = -H \tanh(\beta H)$ and the free energy per particle is $F = -\frac{1}{\beta} \ln[2\cosh(\beta H)]$.

More involved (and more interesting) situations occur, of course, when the effect of mutual interactions among the spins is appreciable. The simplest model that accounts for interactions is the *Ising model*, given by

$$\mathcal{E}(\sigma) = -J \sum_{i,j} \sigma_i \sigma_j - H \sum_{i=1}^{n} \sigma_i \tag{3}$$

where the second term is the contribution of the external magnetic field as before, and the in the first term, pertaining to the interaction, J describes the intensity of the interaction with the summation being defined over pairs of neighboring spins (depending on the geometry of the problem).

More general models allow interactions not only with immediate neighbors, but also with more distant ones, and then there are different strengths of interaction, depending on the distance between the two spins. In this case, the first term is replaced, by the more general form $-\sum_{i,j} J_{ij}\sigma_i\sigma_j$, where now the sum can be defined over all possible pairs $\{(i,j)\}$. Here, in addition to the ferromagnetic case, where all $J_{ij} > 0$, and the antiferromagnetic case, where all $J_{ij} < 0$, there is also a mixed situation where some J_{ij} are positive and others are negative, which is the case of a *spin glass.*

Here, not all spin pairs can be in their preferred mutual position (parallel/antiparallel), thus the system may be *frustrated.*

To model situations of disorder, it is common to model $\{J_{ij}\}$ as random variables (RV's) with, say, equal probabilities of being positive or negative. For example, in the Edwards–Anderson (EA) model [33], $\{J_{ij}\}$ are taken to be i.i.d. zero–mean Gaussian RV's when i and j are neighbors and set to zero otherwise. In the Sherrington–Kirkpatrick (SK) model [34], all $\{J_{ij}\}$ are i.i.d., zero–mean Gaussian RV's. In the *p–spin–glass model*, the interaction terms consist of all products of combinations of p spins (rather than just pairs) with Gaussian coefficients of the appropriate scaling (cf. e.g., [35]).

In all these models, the system has two levels of randomness: the randomness of the interaction coefficients and the randomness of the spin configuration given the interaction coefficients, according to the Boltzmann distribution. However, the two sets of RV's are normally treated differently. The random coefficients are commonly considered *quenched* RV's, namely, they are considered fixed in the time scale at which the spin configuration may vary. This is analogous to the model of coded communication in a random coding paradigm: A randomly drawn code should normally be thought of as a quenched entity, as opposed to the randomness of the source and/or the channel.

2.2.1 The REM in the Absence of a Magnetic Field

In [26],[27],[28], Derrida took the above described idea of randomizing the (parameters of the) Hamiltonian to an extreme, and suggested a model of spin glass with disorder under which the energy levels $\{\mathcal{E}(\sigma)\}$ are simply i.i.d. RV's, without any structure in the form of (3) or its above–described extensions. It can also be viewed, however, as the asymptotic behavior of the p–spin–glass model when $p \to \infty$ (a limit to be taken after the limit $n \to \infty$, i.e., $p << n$) [35]. In particular, in the absence of a magnetic field, the 2^n RV's $\{\mathcal{E}(\sigma)\}$ are taken to be i.i.d., zero–mean Gaussian RV's, all with variance $nJ^2/2$, where J is a parameter.[1] The beauty of the REM is in that on the one hand, it is very easy to analyze, and on the other hand, it consists of sufficient richness to exhibit phase transitions.

[1]The variance scales linearly with n to match the behavior of the Hamiltonian (3) with a limited number of interacting neighbors and random interaction parameters, which has a number of independent terms that is linear in n.

The basic observation about the REM is that for a typical realization of the configurational energies $\{\mathcal{E}(\sigma)\}$, the density of number of configurations with energy about E (i.e., between E and $E+dE$), $N(E)$, is proportional (up to sub–exponential terms in n) to $2^n \cdot e^{-E^2/(nJ^2)}$, as long as $|E| \le E_0 \stackrel{\triangle}{=} nJ\sqrt{\ln 2}$, whereas energy levels outside this range are typically not populated by spin configurations ($N(E)=0$), as the probability of having at least one configuration with such an energy decays exponentially with n. Thus, the asymptotic (thermodynamical) entropy per spin, which is defined by

$$S(E) = \lim_{n\to\infty} \frac{\ln N(E)}{n}$$

is given by

$$S(E) = \begin{cases} \ln 2 - \left(\frac{E}{nJ}\right)^2 & |E| \le E_0 \\ -\infty & |E| > E_0 \end{cases}$$

The partition function of a typical realization of a REM spin glass is then

$$\begin{aligned} Z(\beta) &\doteq \int_{-E_0}^{E_0} dE \cdot N(E) \cdot e^{-\beta E} \\ &\doteq \int_{-E_0}^{E_0} dE \cdot e^{nS(E)} \cdot e^{-\beta E} \end{aligned} \tag{4}$$

where the notation $\doteq$ designates asymptotic equivalence between two functions of n in the exponential scale.[2] The exponential growth rate of $Z(\beta)$,

$$\phi(\beta) \stackrel{\triangle}{=} \lim_{n\to\infty} \frac{\ln Z(\beta)}{n},$$

behaves according to

$$\begin{aligned} \phi(\beta) &= \max_{|E|\le E_0} \left[S(E) - \beta \cdot \frac{E}{n} \right] \\ &= \max_{|E|\le E_0} \left[\ln 2 - \left(\frac{E}{nJ}\right)^2 - \beta J \cdot \left(\frac{E}{nJ}\right) \right]. \end{aligned} \tag{5}$$

Solving this simple optimization problem, one finds that $\phi(\beta)$ is given by

$$\phi(\beta) = \begin{cases} \ln 2 + \frac{\beta^2 J^2}{4} & \beta \le \frac{2}{J}\sqrt{\ln 2} \\ \beta J \sqrt{\ln 2} & \beta > \frac{2}{J}\sqrt{\ln 2} \end{cases}$$

which means that the asymptotic free energy per spin, a.k.a. the *free energy density*, is given by (cf. [5, Proposition 5.2]):

$$F(\beta) = \begin{cases} -\frac{\ln 2}{\beta} - \frac{\beta J^2}{4} & \beta \le \frac{2}{J}\sqrt{\ln 2} \\ -J\sqrt{\ln 2} & \beta > \frac{2}{J}\sqrt{\ln 2} \end{cases}$$

[2] More precisely, $a_n \doteq b_n$ means that $\lim_{n\to\infty} \frac{1}{n} \log \frac{a_n}{b_n} = 0$.

Thus, the free energy density is subjected to a phase transition at the inverse temperature $\beta_c \triangleq \frac{2}{J}\sqrt{\ln 2}$. At high temperatures ($\beta < \beta_c$), which is referred to as the *paramagnetic phase*, the partition function is dominated by an exponential number of configurations with energy $E = -n\beta J^2/2$ and the entropy grows linearly with n. When the system is cooled to $\beta = \beta_c$ and beyond, which is the *glassy phase*, the system freezes but it is still in disorder – the partition function is dominated by a subexponential number of configurations of minimum energy $E = -E_0$. The entropy, in this case, grows sublinearly with n, namely the entropy per spin vanishes, and the free energy density no longer depends on β. Further details about the REM can be found in [5] and the references mentioned in the Introduction.

2.2.2 The REM in the Presence of a Magnetic Field

The random energy levels of the REM, as described above, represent the interaction energies among the various spins in the absence of an external magnetic field. In the presence of an external uniform magnetic field, H (cf. [26],[27],[28]), the Hamiltonian of the system should be supplemented with the term $-H\sum_i \sigma_i = -nm(\sigma)H$ (cf. eq. (3)), where

$$m(\sigma) = \frac{1}{n}\sum_{i=1}^{n}\sigma_i = \frac{1}{n}\sum_{i=1}^{n}[1\{\sigma_i = 1\} - 1\{\sigma_i = -1\}] = \frac{2n_1(\sigma)}{n} - 1$$

is the magnetization associated with the configuration σ, and $n_1(\sigma)$ is the number of spins up, $\sum_{i=1}^{n} 1\{\sigma_i = 1\}$. As far as the statistical description of the REM goes, this shifts the expectation of the random energy level $\mathcal{E}(\sigma)$ from zero to $-nm(\sigma)H$. Equivalently, we can assign the same zero–mean Gaussian distribution as before to the interaction energy, call it now $\mathcal{E}_I(\sigma)$, and add to each configuration σ the term $-nm(\sigma)H$. The corresponding partition function would then be:

$$\begin{aligned} Z(\beta, H) &= \sum_{\sigma} e^{-\beta[\mathcal{E}_I(\sigma) - nm(\sigma)H]} \\ &= \sum_{m}\left[\sum_{\sigma:\ m(\sigma)=m} e^{-\beta\mathcal{E}_I(\sigma)}\right] e^{\beta nmH} \\ &\triangleq \sum_{m}\zeta(\beta, m)e^{\beta nmH} \end{aligned} \tag{6}$$

where $\zeta(\beta, m)$, referred to as the *partial partition function*, contains only the contributions of configurations whose magnetization $m(\sigma)$ is equal to m. The behavior of $\zeta(\beta, m)$ is exactly like that of the REM without a magnetic field, except that instead of 2^n configurations, it has only $|\{\sigma: m(\sigma) = m\}| \doteq 2^{nh((1+m)/2)}$ configurations, where $h(\cdot)$ is the binary entropy function. By carrying out a similar analysis as in the previous subsection to $\zeta(\beta, m)$ and then finding the dominant contribution of m (which is the typical magnetization), one can show (cf. [26],[27],[28]) that there exists a phase transition at $\beta = \beta_c(H)$, where $\beta_c(H)$ is the unique solution to the equation

$$\beta^2 J^2 = 4h\left(\frac{1+\tanh(\beta H)}{2}\right).$$

It is not difficult to see that $\beta_c(H)$ is a non–increasing function of $|H|$ and therefore $T_c(H) = 1/\beta_c(H)$ is non–decreasing, with a minimum at $H = 0$, given by $T_c(0) = J/(2\sqrt{\ln 2})$ (see Fig. 1). For high temperatures ($\beta \le \beta_c(H)$), where the effect of the interactions among the spins is relatively insignificant, one observes the ordinary paramagnetic behavior, with the average magnetization is

$$m = m_p(\beta, H) \stackrel{\Delta}{=} \tanh(\beta H),$$

whereas for low temperatures ($\beta \ge \beta_c(H)$), the system is frozen in the spin glass phase where the magnetization no longer depends on the temperature:

$$m = m_g(H) \stackrel{\Delta}{=} \tanh(\beta_c(H) \cdot H).$$

The free energy per spin is given by

$$F(\beta, H) = -\lim_{n\to\infty} \frac{\ln Z(\beta, H)}{n\beta} = \begin{cases} -\left[\frac{\beta J^2}{4} + \frac{h([1+\tanh(\beta H)]/2)}{\beta} + H\tanh(\beta H)\right] & \beta < \beta_c(H) \\ -\left[J\sqrt{h\left(\frac{1+\tanh(\beta_c(H)H)}{2}\right)} + H\tanh(\beta_c(H)H)\right] & \beta \ge \beta_c(H) \end{cases}$$

As can be seen, no sponteneous magnetization takes place under the REM, even at low temperatures ($H \to 0$ implies $m \to 0$). As for other thermodynamic quantities, we have the average internal energy per spin

$$E(\beta, H) = \frac{\partial}{\partial\beta}[\beta F(\beta, H)] = \begin{cases} -H\tanh(\beta H) - \frac{\beta J^2}{2} & \beta < \beta_c(H) \\ -H\tanh(\beta_c(H)\cdot H) - \frac{\beta_c(H)J^2}{2} & \beta \ge \beta_c(H) \end{cases}$$

the entropy per spin

$$S(\beta, H) = \beta[E(\beta, H) - F(\beta, H)] = \begin{cases} h\left(\frac{1+\tanh(\beta H)}{2}\right) - \frac{\beta^2 J^2}{4} & \beta < \beta_c(H) \\ 0 & \beta \ge \beta_c(H) \end{cases}$$

and the magnetic susceptibility

$$\chi = \left[\frac{\partial m}{\partial H}\right]_{H=0} = \begin{cases} \beta & \beta < \beta_c(H) \\ \beta_c(H) & \beta \ge \beta_c(H) \end{cases}$$

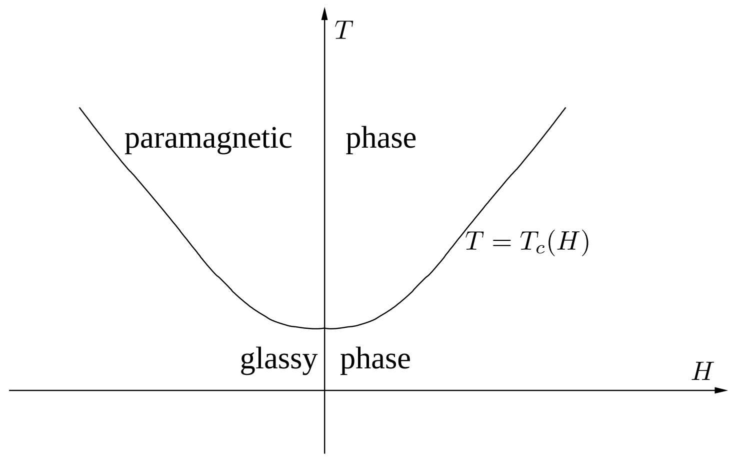

Figure 1: Phase diagram of temperature vs. magnetic field.

2.3 Joint Source–Channel Code Ensembles and the REM in a Magnetic Field

In this subsection, we analyze the behavior of a finite–temperature decoder for a typical randomly selected code using the tools of the analysis of the REM in a magnetic field. Using the viewpoint of the magnetic properties of the REM, it will be seen that the source bits play the role of spins in a magnetic field whose intensity is $H = \frac{1}{2}\ln\frac{q}{1-q}$, where q is the probability that $u_i = 1$ for each i. Accordingly, instead of the binary alphabet $\{0, 1\}$ that we used before, it will prove more convenient to let each u_i assume values in $\{-1, +1\}$. Another slight change in notation, that will take place mostly for the sake of convenience, is that instead of defining channel capacity and coding rates in terms of bits, we will define them in units of *nats*, where 1 nat $= \ln 2$ bits. This means that logarithms will be taken to the natural basis e rather than the base 2. Accordingly, $h(q)$ will be redefined hereafter as $h(q) = -q\ln q - (1-q)\ln(1-q)$ and the capacity of the binary channel considered in Section 2 will be redefined as $C = \ln 2 - h(q) = \ln 2 + q\ln q + (1-q)\ln(1-q)$.

Consider then a binary memoryless source sequence, $u_1, u_2, \ldots, u_i \in \{-1, +1\}$, with a parameter $q = \Pr\{u_i = 1\}$ and a binary symmetric channel with parameter p, which as described in Section 2, is assumed to operate θ times faster than the source, in other words, the channel transmits θ bits during the time that the source generates one bit. The number θ is a positive real which will be assumed fixed throughout the sequel. Consider a joint source–channel code that receives a source vector of length N, $\boldsymbol{u} = (u_1, \ldots, u_N)$, and produces a channel input vector $\boldsymbol{x}(\boldsymbol{u})$ of length $n = N\theta$. The block encoder is generated by random selection: We randomly draw 2^N binary n-vectors,

$\{\boldsymbol{x}(\boldsymbol{u}),\ \boldsymbol{u} \in \{-1,+1\}^N\}$, independently, by fair coin tossing. As described in Section 2, when the input to the encoder is $\boldsymbol{u}$, the encoder transmits the corresponding codeword $\boldsymbol{x} = \boldsymbol{x}(\boldsymbol{u})$, and the decoder, upon receiving the channel output $\boldsymbol{y}$, applies a finite–temperature decoder

$$\hat{u}_i = \text{argmax}_{u\in\{-1,+1\}} \sum_{\boldsymbol{u}:\ u_i=u} [P(\boldsymbol{u})P(\boldsymbol{y}|\boldsymbol{x}(\boldsymbol{u}))]^\beta, \quad i = 1, 2, \ldots, N.$$

We can think of this decoder as a symbol MAP decoder pertaining to a *posterior* distribution given by

$$\begin{aligned} P_\beta(\boldsymbol{u}|\boldsymbol{y}) &= \frac{[P(\boldsymbol{u})P(\boldsymbol{y}|\boldsymbol{x}(\boldsymbol{u}))]^\beta}{\sum_{\boldsymbol{u}'}[P(\boldsymbol{u}')P(\boldsymbol{y}|\boldsymbol{x}(\boldsymbol{u}'))]^\beta} \\ &= \frac{e^{-\beta\ln[1/P(\boldsymbol{u})P(\boldsymbol{y}|\boldsymbol{x}(\boldsymbol{u}))]}}{\sum_{\boldsymbol{u}'} e^{-\beta\ln[1/P(\boldsymbol{u})P(\boldsymbol{y}|\boldsymbol{x}(\boldsymbol{u}))]}} \end{aligned} \tag{7}$$

where in the second line we presented this distribution in the form of the Boltzmann–Gibbs distribution with an Hamiltonian given by $\ln[1/P(\boldsymbol{u})P(\boldsymbol{y}|\boldsymbol{x}(\boldsymbol{u}))]$ (see also e.g., [5, Chap. 6]). The corresponding *partition function* is then

$$\begin{aligned} Z(\beta) &= \sum_{\boldsymbol{u}} [P(\boldsymbol{u})P(\boldsymbol{y}|\boldsymbol{x}(\boldsymbol{u}))]^\beta \\ &= [P(\boldsymbol{u}_0)P(\boldsymbol{y}|\boldsymbol{x}(\boldsymbol{u}_0))]^\beta + \sum_{\boldsymbol{u}\neq\boldsymbol{u}_0} [P(\boldsymbol{u})P(\boldsymbol{y}|\boldsymbol{x}(\boldsymbol{u}))]^\beta \\ &\stackrel{\triangle}{=} Z_c(\beta) + Z_e(\beta), \end{aligned} \tag{8}$$

where we have separated the partition function into two contributions: $Z_c(\beta)$, corresponding to the correct source sequence $\boldsymbol{u}_0$ that was actually generated by the source and fed into the encoder, and $Z_e(\beta)$ corresponding to all other possible messages. Now, since typically, the source produces sequences with about Nq occurrences of +1 and $N(1-q)$ occurrences of -1, and the channel flips about np out of n of the transmitted bits, $Z_c(\beta)$ is typically around $e^{-N\beta h(q)} \cdot e^{-n\beta h(p)} = e^{-N\beta[h(q)+\theta h(p)]}$. On the other hand, as we will show now, $Z_e(\beta)$ behaves like the REM in a magnetic field whose intensity is

$$H = \frac{1}{2}\ln\frac{q}{1-q}.$$

Accordingly, we will henceforth denote $Z_e(\beta)$ also by $Z_e(\beta, H)$, to emphasize the analogy to the REM in a magnetic field.

To see that $Z_e(\beta, H)$ behaves like the REM in a magnetic field, consider the following: first, denote by $N_1(\boldsymbol{u})$ the number of +1's in $\boldsymbol{u}$, so that the magnetization, $m(\boldsymbol{u}) \stackrel{\triangle}{=} \frac{1}{N}[\sum_{i=1}^N 1\{u_i =$

$+1\} - \sum_{i=1}^N 1\{u_i = -1\}]$, pertaining to spin configuration $\boldsymbol{u}$, is given by $m(\boldsymbol{u}) = 2N_1(\boldsymbol{u})/N - 1$. Equivalently, $N_1(\boldsymbol{u}) = N(1 + m(\boldsymbol{u}))/2$, and then

$$\begin{aligned} P(\boldsymbol{u}) &= q^{N_1(\boldsymbol{u})}(1-q)^{N-N_1(\boldsymbol{u})} \\ &= (1-q)^N \left(\frac{q}{1-q}\right)^{N(1+m(\boldsymbol{u}))/2} \\ &= [q(1-q)]^{N/2} \left(\frac{q}{1-q}\right)^{Nm(\boldsymbol{u}))/2} \\ &= [q(1-q)]^{N/2} e^{Nm(\boldsymbol{u})H} \end{aligned} \quad (9)$$

where H is defined as above. By the same token, for the binary symmetric channel we have:

$$P(\boldsymbol{y}|\boldsymbol{x}) = p^{d_H(\boldsymbol{x},\boldsymbol{y})}(1-p)^{n-d_H(\boldsymbol{x},\boldsymbol{y})} = (1-p)^n e^{-Bd_H(\boldsymbol{x},\boldsymbol{y})}$$

where $B = \ln \frac{1-p}{p}$ and $d_H(\boldsymbol{x}, \boldsymbol{y})$ is the Hamming distance between $\boldsymbol{x}$ and $\boldsymbol{y}$, namely, the number of places $\{i\}$ where $y_i \neq x_i$. Thus,

$$\begin{aligned} Z_e(\beta, H) &= [q(1-q)]^{N\beta/2} \sum_m \left[\sum_{\boldsymbol{x}(\boldsymbol{u}):\ m(\boldsymbol{u})=m} e^{-\beta \ln[1/P(\boldsymbol{y}|\boldsymbol{x}(\boldsymbol{u}))]} \right] e^{N\beta mH} \\ &= [q(1-q)]^{\beta N/2}(1-p)^{n\beta} \sum_m \left[\sum_{\boldsymbol{x}(\boldsymbol{u}):\ m(\boldsymbol{u})=m} e^{-\beta B d_H(\boldsymbol{x}(\boldsymbol{u}),\boldsymbol{y})} \right] e^{\beta NmH} \\ &\stackrel{\triangle}{=} [q(1-q)]^{N\beta/2}(1-p)^{n\beta} \sum_m \zeta(\beta, m) e^{\beta NmH} \end{aligned} \quad (10)$$

where the resemblance to eq. (6) is self evident, with $\zeta(\beta, m)$ being redefined as the second bracketed term. In analogy to the above analysis of the REM, $\zeta(\beta, m)$ here behaves like in the REM without a magnetic field, namely, it contains exponentially $e^{Nh((1+m)/2)} = e^{nh((1+m)/2)/\theta}$ terms, with the random energy levels of the REM being replaced now by random Hamming distances $\{d_H(\boldsymbol{x}(\boldsymbol{u}), \boldsymbol{y})\}$ that are induced by the random selection of the code $\{\boldsymbol{x}(\boldsymbol{u})\}$.[3] Using the same considerations as with the REM (see also [5]), $\zeta(\beta, m)$ can be represented as $\sum_\delta N_{\boldsymbol{y},m}(\delta) e^{-\beta Bn\delta}$, where $N_{\boldsymbol{y},m}(\delta)$ is the number of vectors $\{\boldsymbol{u}\}$ with $m(\boldsymbol{u}) = m$ and $d_H(\boldsymbol{x}(\boldsymbol{u}), \boldsymbol{y}) = n\delta$. Since $N_{\boldsymbol{y},m}(\delta)$ is the sum of $e^{nh((1+m)/2)/\theta}$ many i.i.d. binary random variables of the form $1\{d_H(\boldsymbol{x}(\boldsymbol{u}), \boldsymbol{y}) = n\delta\}$ (again, with randomness induced by the random selection of $\boldsymbol{x}(\boldsymbol{u})$), each with expectation given by $\Pr\{d_H(\boldsymbol{x}(\boldsymbol{u}), \boldsymbol{y}) = n\delta\} \doteq e^{n[h(\delta)-\ln 2]}$, then $N_{\boldsymbol{y},m}(\delta)$ is typically zero for all δ such that

[3] Of course, the channel output vector $\boldsymbol{y}$ is also random, but this randomness does not play any essential role here. This discussion applies as well for every given $\boldsymbol{y}$.

$h((1+m)/2)/\theta + h(\delta) - \ln 2 < 0$, and is typically around its expectation, $e^{n[h((1+m)/2)/\theta + h(\delta) - \ln 2]}$, for all δ such that $h((1+m)/2)/\theta + h(\delta) - \ln 2 \geq 0$.

Defining now the Gilbert–Varshamov distance $\delta_{GV}(R)$ [5, Chap. 6] as the solution $\delta \leq 1/2$ to the equation $h(\delta) = \ln 2 - R$, the condition $h((1+m)/2)/\theta + h(\delta) - \ln 2 \geq 0$ is equivalent to the condition $\delta_{GV}(h((1+m)/2)/\theta) \leq \delta \leq 1 - \delta_{GV}(h((1+m)/2)/\theta)$. Thus, for a typical randomly selected code,

$$\begin{aligned}
\phi(\beta, m) &\stackrel{\Delta}{=} \lim_{n\to\infty} \frac{\ln \zeta(\beta, m)}{n} \\
&= \max_{\delta\in[\delta_{GV}(h((1+m)/2)/\theta), 1-\delta_{GV}(h((1+m)/2)/\theta]} \left[\frac{1}{\theta} h\left(\frac{1+m}{2}\right) + h(\delta) - \ln 2 - \beta B \delta\right] \\
&= \left\{ \begin{array}{ll} \frac{1}{\theta} h\left(\frac{1+m}{2}\right) - \ln 2 + h(p_\beta) - \beta B p_\beta & p_\beta \geq \delta_{GV}\left(\frac{1}{\theta} h\left(\frac{1+m}{2}\right)\right) \\ -\beta B \delta_{GV}\left(\frac{1}{\theta} h\left(\frac{1+m}{2}\right)\right) & p_\beta < \delta_{GV}\left(\frac{1}{\theta} h\left(\frac{1+m}{2}\right)\right) \end{array} \right.
\end{aligned} \quad (11)$$

where $p_\beta \stackrel{\Delta}{=} p^\beta/(p^\beta + (1-p)^\beta)$. The condition $p_\beta \geq \delta_{GV}(\frac{1}{\theta} h(\frac{1+m}{2}))$ is equivalent to the condition

$$\beta \leq \beta_0(m) \stackrel{\Delta}{=} \frac{1}{B} \ln \left[\frac{1 - \delta_{GV}(h((1+m)/2)/\theta)}{\delta_{GV}(h((1+m)/2)/\theta)}\right].$$

The exponential order of $\sum_m \zeta(\beta, m) e^{N\beta m H}$, as a function of N is then

$$\psi(\beta, H) \stackrel{\Delta}{=} \lim_{N\to\infty} \frac{1}{N} \ln \left[\sum_m \zeta(\beta, m) e^{N\beta m H}\right] = \max_m [\theta\phi(\beta, m) + \beta m H].$$

For small enough β, the dominant value of m is the one that maximizes $[h((1+m)/2) + \beta m H]$, namely, the well–known paramagnetic magnetization $m = m_p(\beta, H) = \tanh(\beta H)$. This is true as long as $\beta \leq \beta_0(\tanh(\beta H))$. Consider then the equation

$$\beta = \beta_0(\tanh(\beta H))$$

where the unknown is β, or equivalently, the equation

$$\ln 2 - h(p_\beta) = \frac{1}{\theta} h\left(\frac{1 + \tanh(\beta H)}{2}\right).$$

Now $h((1+\tanh(\beta H))/2)$ is decreasing with β, while $[\ln 2 - h(p_\beta)]$ is increasing. At $\beta = 0$, $\ln 2 - h(p_\beta) = 0$ whereas $h((1+\tanh(\beta H))/2)/\theta = \ln 2/\theta$. As $\beta \to \infty$, $\ln 2 - h(p_\beta) \to \ln 2$ whereas $h((1+\tanh(\beta H))/2)/\theta \to 0$, provided that $H \neq 0$. Thus, for $H \neq 0$, there must be a unique solution, which we shall denote by $\beta_{pg}(H)$, where the subscript "pg" stands for the fact that this is the boundary curve between the paramagnetic phase and the glassy phase. Since

$h((1+\tanh(\beta H))/2))/\theta$ is decreasing with $|H|$, $\beta_{pg}(H)$ is decreasing in $|H|$, i.e., the temperature $T_{pg}(H) = 1/\beta_{pg}(H)$ is increasing in $|H|$, as before (see Fig. 2). As for the case $H = 0$, for $\theta > 1$, we have

$$\beta_{pg}(0) = \frac{1}{B}\ln\left[\frac{1 - h^{-1}((1-1/\theta)\ln 2)}{h^{-1}((1-1/\theta)\ln 2)}\right].$$

For $0 < \theta \leq 1$, $\beta_{pg}(0) = \infty$, namely, $T_{pg}(0) = 0$, which means that there is no phase transition as the behavior is paramagnetic at all temperatures. In the same manner, it is easy to see that $\beta_{pg}(\infty) = 0$ for all $\theta > 0$, which is another case where there are no phase transitions, but this time, it is a glassy behavior at all temperatures.

As long as $\beta \leq \beta_{pg}(H)$, we have

$$\psi(\beta,H) = \psi_p(\beta,H) \stackrel{\Delta}{=} h\left(\frac{1+\tanh(\beta H)}{2}\right) - \theta(\ln 2 - h(p_\beta) + \beta B p_\beta) + \beta H\tanh(\beta H).$$

On the other hand, for $\beta > \beta_{pg}(H)$, the system is in the glassy phase. In this case,

$$\psi(\beta,H) = \psi_g(\beta,H) \stackrel{\Delta}{=} \beta\max_m\left[mH - \theta B\delta_{GV}\left(\frac{1}{\theta}h\left(\frac{1+m}{2}\right)\right)\right]$$

thus, the maximizing m depends only on H but not on β. In this case, we have $m = m_g(H) = \tanh(\beta_{pg}(H)\cdot H)$ and so

$$\begin{aligned}\psi_g(\beta,H) &= \beta\left[H\tanh(\beta_{pg}(H)\cdot H) - B\theta\delta_{GV}\left(\frac{1}{\theta}h\left(\frac{1+\tanh(\beta_{pg}(H)\cdot H)}{2}\right)\right)\right]\\ &= \beta\left[H\tanh(\beta_{pg}(H)\cdot H) - B\theta p_{\beta_{pg}(H)}\right]. \end{aligned} \tag{12}$$

The free–energy density associated with erroneous messages is therefore given by

$$F_e(\beta,H) \stackrel{\Delta}{=} -\lim_{N\to\infty}\frac{\ln Z_e(\beta,H)}{N\beta} = -\frac{1}{2}\ln[q(1-q)] - \theta\ln(1-p) - \frac{\psi(\beta,H)}{\beta}$$

i.e.,

$$F_e(\beta,H) = \begin{cases} F_p(\beta,H) & \beta \leq \beta_{pg}(H)\\ F_g(H) & \beta > \beta_{pg}(H)\end{cases}$$

where

$$F_p(\beta,H) = -\frac{1}{2}\ln[q(1-q)] - \theta\ln(1-p) - \frac{1}{\beta}\left[h\left(\frac{1+\tanh(\beta H)}{2}\right) - \theta(\ln 2 - h(p_\beta))\right] + \theta B p_\beta - H\tanh(\beta H)$$

and

$$F_g(H) = -\frac{1}{2}\ln[q(1-q)] - \theta\ln(1-p) - H\tanh(\beta_{pg}(H)\cdot H) + B\theta p_{\beta_{pg}(H)}.$$

The boundary between the ferromagnetic phase (where $Z_c(\beta)$ is the dominant term in $Z(\beta)$) and the glassy phase is the vertical line (see Fig. 2) $H = H_{fg}$, where H_{fg} is the solution to the equation

$$\frac{h(q)}{\theta} + h(p) = -\frac{1}{2\theta}\ln[q(1-q)] - \ln(1-p) - \frac{H\tanh(\beta_{pg}(H)H)}{\theta} + Bp_{\beta_{pg}(H)}$$

which after rearranging terms becomes

$$Bp - \frac{H\tanh(H)}{\theta} = Bp_{\beta_{pg}(H)} - \frac{H\tanh(\beta_{pg}(H)\cdot H)}{\theta},$$

whose solution in turn is achieved when $\beta_{pg}(H) = 1$, i.e.,

$$p = \delta_{GV}\left(\frac{1}{\theta}h\left(\frac{1+\tanh(H)}{2}\right)\right) \equiv \delta_{GV}\left(\frac{h(q)}{\theta}\right),$$

which is nothing but the boundary of reliable communication (1). Thus,

$$H_{fg} = \frac{1}{2}\ln\frac{q^*}{1-q^*} \quad \text{with} \quad q^* = 1 - h^{-1}(\theta(\ln 2 - h(p)),$$

where $h^{-1}(\cdot)$ is the inverse of the function $h(\cdot)$ in the range where the argument is in $[0, \frac{1}{2}]$. The vertical line $H = H_{fg}$ intersects the paramagnetic–glassy boundary curve $T = T_{pg}(H)$ at the triple point $(H,T) = (H_{fg}, 1)$, namely, $T_{pg}(H_{fg}) = 1$. The ferromagnetic region, pertaining to correct decoding (where $m = 2q-1 = \tanh(H)$), is $\{(H,T):\ |H| \geq H_{fg},\ T < T_{pf}(H)\}$, where $T = T_{pf}(H)$ is paramagnetic–ferromagnetic boundary curve (see Fig. 2) given by the solution $\beta = 1/T$ of the equation

$$\beta pB - \frac{\beta H\tanh(H)}{\theta} = \ln 2 + \beta p_\beta B - h(p_\beta) - \frac{1}{\theta}h\left(\frac{1+\tanh(\beta H)}{2}\right) - \frac{\beta}{\theta}H\tanh(\beta H)$$

for every given H which is larger than H_{fg} in absolute value. As can be seen, it also contains the point $(H,T) = (H_{fg}, 1)$.

Discussion: We see that correct decoding occurs in a sufficiently strong magnetic field. This is not surprising as a strong magnetic field corresponds to a low–entropy source which can be transmitted reliably. The above exposition of the magnetization as a function of H and T is instructive for the understanding of typical error patterns in joint source–channel coding. At very low temperatures (like in word MAP decoding, which corresponds to $\beta \to \infty$), the (sub–exponentially few) typical patterns of the erroneneously decoded vectors $\{\boldsymbol{u}\}$ have magnetization dictated by the frozen phase, namely, $m_g(H) = \tanh(\beta_{pg}(H) \cdot H)$, independently of the decoding temperature. For magnetic fields smaller than H_{fg} in absolute value (namely, for sources with high entropy), $\beta_{pg}(H) > 1$, which means that the magnetization of a typical erroneously decoded sequence is *higher* than that of a typical (correct) source sequence which is $m_f = 2q-1 = \tanh(H)$. If the working temperature is lower than $T_{pg}(0)$, this remains true no matter how small $|H|$ is. If, on the other hand, $T_{pg}(0) < T < 1$, then when the magnetic field is reduced, the magnetization of the (exponentially many) erroneously decoded vectors $\{\boldsymbol{u}\}$ is given by $m_p(\beta, H) = \tanh(\beta H)$, which is still higher than that of the typical source vector $\boldsymbol{u}$, but now it is temperature–dependent.

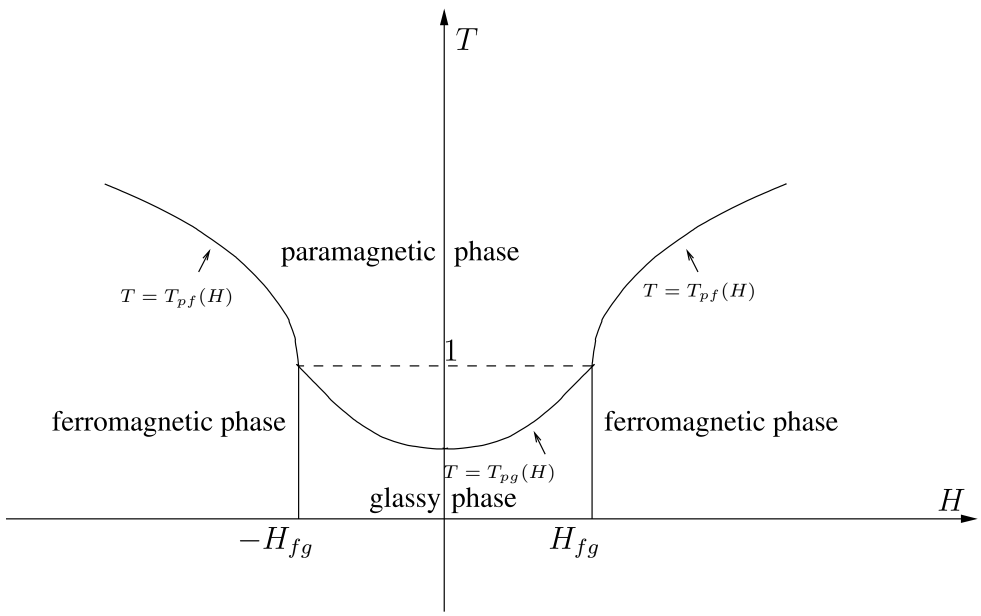

Figure 2: Phase diagram of joint source channel coding: temperature vs. magnetic field.

3 Ensemble Performance of Codes and Free Energies

In this section, we provide bounds on the ensemble performance of joint source channel codes for the binary symmetric source and the binary symmetric channel. In particular, we examine the exponential decay rate of the average probability of correct decoding (the correct decoding exponent, for short) when the condition for reliable communication (1) is violated as well as the exponential decay rate of the average probability of error (error exponent) when this condition holds. As will be seen, the former is intimately related to the free energy in the glassy phase, whereas the latter is strongly related to the free energy in the paramagnetic phase.

The relationship between the previous derivations and both the correct decoding exponent and the error exponent stems from the fact both performance measures are bounded by expressions that are strongly related to the partition function $Z_e(\beta)$.

3.1 The Correct Decoding Exponent

The probability of correct decoding pertaining to the word MAP decoder is well known (and can easily be shown) to be given by

$$\begin{aligned}
P_c &= \sum_{\boldsymbol{y}} \max_{\boldsymbol{u}}[P(\boldsymbol{u})P(\boldsymbol{y}|\boldsymbol{x}(\boldsymbol{u}))] \\
&= \sum_{\boldsymbol{y}} \lim_{\beta\to\infty} \left[\sum_{\boldsymbol{u}} P^{\beta}(\boldsymbol{u})P^{\beta}(\boldsymbol{y}|\boldsymbol{x}(\boldsymbol{u}))\right]^{1/\beta} \\
&= [q(1-q)]^{N/2}(1-p)^n \sum_{\boldsymbol{y}} \lim_{\beta\to\infty} \left[\sum_m \zeta(\beta,m)e^{N\beta mH}\right]^{1/\beta} \\
&\doteq [q(1-q)]^{N/2}(1-p)^n \sum_{\boldsymbol{y}} \lim_{\beta\to\infty} \sum_m \zeta^{1/\beta}(\beta,m)e^{NmH}, \qquad (13)
\end{aligned}$$

where with a slight abuse of notation, here $\zeta(\beta,m)$ is redefined to include *all* messages $\{\boldsymbol{u}\}$, including the correct one. Now, taking the ensemble average:

$$\bar{P}_c \doteq [q(1-q)]^{N/2}(1-p)^n \sum_{\boldsymbol{y}} \lim_{\beta\to\infty} \sum_m \boldsymbol{E}\{\zeta^{1/\beta}(\beta,m)\} \cdot e^{NmH}.$$

Now,

$$\begin{aligned}
\boldsymbol{E}\{\zeta^{1/\beta}(\beta,m)\} &= \boldsymbol{E}\left\{\left[\sum_{\delta} N_{\boldsymbol{y},m}(\delta)e^{-\beta Bn\delta}\right]^{1/\beta}\right\} \\
&\doteq \boldsymbol{E}\left\{\left[\max_{\delta} N_{\boldsymbol{y},m}(\delta)e^{-\beta Bn\delta}\right]^{1/\beta}\right\} \\
&= \boldsymbol{E}\left\{\max_{\delta} N^{1/\beta}_{\boldsymbol{y},m}(\delta)e^{-Bn\delta}\right\} \\
&\doteq \sum_{\delta} \boldsymbol{E}\{N^{1/\beta}_{\boldsymbol{y},m}(\delta)\} \cdot e^{-Bn\delta} \qquad (14)
\end{aligned}$$

where again, $N_{\boldsymbol{y},m}(\delta)$ is the number of codewords $\{\boldsymbol{x}(\boldsymbol{u})\}$, corresponding to source words with $m(\boldsymbol{u}) = m$, which fall at Hamming distance $n\delta$ from $\boldsymbol{y}$. Now, as shown in [29],[36, Appendix],

$$\boldsymbol{E}\{N_{\boldsymbol{y},m}^{1/\beta}(\delta)\} \doteq \begin{cases} \exp\left\{n\left[\frac{1}{\theta}h\left(\frac{1+m}{2}\right)+h(\delta)-\ln 2\right]\right\} & \frac{1}{\theta}h\left(\frac{1+m}{2}\right)+h(\delta)<\ln 2 \\ \exp\left\{n\left[\frac{1}{\theta}h\left(\frac{1+m}{2}\right)+h(\delta)-\ln 2\right]/\beta\right\} & \frac{1}{\theta}h\left(\frac{1+m}{2}\right)+h(\delta)\geq\ln 2 \end{cases} \tag{15}$$

Thus,

$$\lim_{\beta\to\infty}\boldsymbol{E}\{\zeta^{1/\beta}(\beta,m)\} \doteq \exp\left\{n\left[\frac{1}{\theta}h\left(\frac{1+m}{2}\right)-\ln 2+\max_{\delta\leq\delta_{GV}(h((1+m)/2)/\theta)}\{h(\delta)-B\delta\}\right]\right\} \tag{16}$$

and so,

$$\begin{aligned}\bar{P}_c &\doteq [q(1-q)]^{N/2}(1-p)^n\sum_{\boldsymbol{y}}\sum_m\left[\exp\left\{n\left[\frac{1}{\theta}h\left(\frac{1+m}{2}\right)-\ln 2+\right.\right.\right. \\ &\left.\left.\left.\max_{\delta\leq\delta_{GV}(h((1+m)/2)/\theta)}\{h(\delta)-B\delta\}\right]\right\}\right]e^{NmH}. \end{aligned}\tag{17}$$

The dominant m is the one that maximizes

$$\frac{1}{\theta}h\left(\frac{1+m}{2}\right)-\ln 2+\max_{\delta\leq\delta_{GV}(h((1+m)/2)/\theta)}\{h(\delta)-B\delta\}+\frac{mH}{\theta}$$

Now, if $h(p)\leq\ln 2-h((1+m)/2)/\theta$, then the inner maximization is attained at $\delta=p$ and we get

$$\frac{1}{\theta}h\left(\frac{1+\tanh(H)}{2}\right)-\ln 2+h(p)-Bp+\frac{H\tanh(H)}{\theta}.$$

If $h(p)\leq\ln 2-h((1+\tanh(H))/2)/\theta$, namely, the condition for reliable communication holds, this indeed happens. In this case, we get

$$\bar{P}_c \doteq [q(1-q)]^{N/2}(1-p)^n\cdot 2^n\cdot\exp\left\{n\left[\frac{h(q)}{\theta}-\ln 2+h(p)-Bp+\frac{2q-1}{2\theta}\ln\frac{q}{1-q}\right]\right\}=1,$$

as expected. Otherwise, the maximum is attained at the boundary of the allowed range of δ, and we get

$$\max_m\left[\frac{mH}{\theta}-B\delta_{GV}\left(\frac{1}{\theta}h\left(\frac{1+m}{2}\right)\right)\right]=\frac{H}{\theta}\tanh(\beta_{pg}(H)\cdot H)-B\delta_{GV}\left(\frac{1}{\theta}h\left(\frac{1+\tanh(\beta_{pg}(H)\cdot H}{2}\right)\right)$$

and so, the correct decoding exponent is

$$\begin{aligned}E_c &\stackrel{\triangle}{=} -\lim_{N\to\infty}\frac{\ln\bar{P}_c}{N} \\ &= -\frac{1}{2}\ln[q(1-q)]-\theta\ln[2(1-p)]+ \\ &\quad B\theta\delta_{GV}\left(\frac{1}{\theta}h\left(\frac{1+\tanh(\beta_{pg}(H)\cdot H}{2}\right)\right)-H\tanh(\beta_{pg}(H)\cdot H) \\ &= F_g(H)-\theta\ln 2.\end{aligned}\tag{18}$$

Thus, we obtained a very simple relationship between the correct decoding exponent and the glassy free energy. The ferromagnetic–glassy phase transition is exactly the transition from $E_c = 0$ to $E_c > 0$. The dominant magnetization of the correct decoding event is then $m_g(H) = \tanh(\beta_{pg}(H) \cdot H)$, i.e., the dominant (rare) event of correct decoding is when the source vector $\boldsymbol{u}$ has the (non–typical) magentization $m_g(H)$. If the condition of reliable communication does not hold, i.e., $h(q)/\theta > \ln 2 - h(p)$, then the word MAP decoder ($\beta \to \infty$) works in the glassy regime, but the symbol MAP decoder ($\beta = 1$) works in the paramagnetic regime. The computation of $\bar{P}_c$ for the word MAP decoder is carried out also in the glassy regime.

3.2 The Error Exponent

We begin by using Gallager's techniques (see [31, Problem 5.16, pp. 534–535]): The probability of error for a given code and the word MAP decoder is given by

$$P_e = \sum_{\boldsymbol{u}} P(\boldsymbol{u}) \sum_{\boldsymbol{y}} P(\boldsymbol{y}|\boldsymbol{x}(\boldsymbol{u})) \cdot 1\{\exists\ \boldsymbol{u}':\ P(\boldsymbol{u}')P(\boldsymbol{y}|\boldsymbol{x}(\boldsymbol{u}')) \geq P(\boldsymbol{u})P(\boldsymbol{y}|\boldsymbol{x}(\boldsymbol{u}))\}.$$

Now, it is easy to see that whenever an error occurs

$$\sum_{\boldsymbol{u}' \neq \boldsymbol{u}} \left[\frac{P(\boldsymbol{u}')P(\boldsymbol{y}|\boldsymbol{x}(\boldsymbol{u}'))}{P(\boldsymbol{u})P(\boldsymbol{y}|\boldsymbol{x}(\boldsymbol{u}))}\right]^{\beta} \geq 1.$$

for every $\beta \geq 0$. Thus,

$$1\{\exists\ \boldsymbol{u}':\ P(\boldsymbol{u}')P(\boldsymbol{y}|\boldsymbol{x}(\boldsymbol{u}')) \geq P(\boldsymbol{u})P(\boldsymbol{y}|\boldsymbol{x}(\boldsymbol{u}))\} \leq \left(\sum_{\boldsymbol{u}' \neq \boldsymbol{u}} \left[\frac{P(\boldsymbol{u}')P(\boldsymbol{y}|\boldsymbol{x}(\boldsymbol{u}'))}{P(\boldsymbol{u})P(\boldsymbol{y}|\boldsymbol{x}(\boldsymbol{u}))}\right]^{\beta}\right)^{\rho}$$

for every $\rho \geq 0$. Substituting the right–hand side into the expression of P_e, we get the following upper bound:

$$P_e \leq \sum_{\boldsymbol{u}} P(\boldsymbol{u})^{1-\rho\beta} \sum_{\boldsymbol{y}} P(\boldsymbol{y}|\boldsymbol{x}(\boldsymbol{u}))^{1-\rho\beta} \left(\sum_{\boldsymbol{u}' \neq \boldsymbol{u}} [P(\boldsymbol{u}')P(\boldsymbol{y}|\boldsymbol{x}(\boldsymbol{u}'))]^{\beta}\right)^{\rho} \qquad \beta \geq 0,\ \rho \geq 0. \tag{19}$$

Thus, the average error probability over the ensemble of codes is bounded by

$$\bar{P}_e \leq \sum_{\boldsymbol{u}} P(\boldsymbol{u})^{1-\rho\beta} \sum_{\boldsymbol{y}} \boldsymbol{E}\left\{P(\boldsymbol{y}|\boldsymbol{X}(\boldsymbol{u}))^{1-\rho\beta}\right\} \cdot \boldsymbol{E}\left\{\left(\sum_{\boldsymbol{u}' \neq \boldsymbol{u}} [P(\boldsymbol{u}')P(\boldsymbol{y}|\boldsymbol{X}(\boldsymbol{u}'))]^{\beta}\right)^{\rho}\right\}. \tag{20}$$

In the binary symmetric case considered here, the first expectation is given by:

$$\begin{aligned}
\boldsymbol{E}\left\{P(\boldsymbol{y}|\boldsymbol{X}(\boldsymbol{u}))^{1-\rho\beta}\right\} &= 2^{-n}\sum_{\boldsymbol{x}}\prod_{i=1}^{n}P(y_i|x_i)^{1-\rho\beta} \\
&= 2^{-n}[p^{1-\rho\beta}+(1-p)^{1-\rho\beta}]^n \\
&= e^{-n\gamma(1-\rho\beta)} \qquad (21)
\end{aligned}$$

where $\gamma(s) \stackrel{\Delta}{=} \ln 2 - \ln[p^s + (1-p)^s]$. The second expectation is handled as follows. Using the above derived relation:

$$\sum_{\boldsymbol{u}'}[P(\boldsymbol{u}')P(\boldsymbol{y}|\boldsymbol{X}(\boldsymbol{u}'))]^\beta = [q(1-q)]^{N\beta/2}(1-p)^{n\beta}\sum_m \zeta(\beta,m)e^{\beta NmH},$$

we get

$$\begin{aligned}
\left(\sum_{\boldsymbol{u}'\neq\boldsymbol{u}}[P(\boldsymbol{u}')P(\boldsymbol{y}|\boldsymbol{X}(\boldsymbol{u}'))]^\beta\right)^\rho &= \left([q(1-q)]^{N\beta/2}(1-p)^{n\beta}\sum_m \zeta(\beta,m)e^{\beta NmH}\right)^\rho \\
&\doteq [q(1-q)]^{N\beta\rho/2}(1-p)^{n\beta\rho}\sum_m \zeta^\rho(\beta,m)e^{\beta\rho NmH} \qquad (22)
\end{aligned}$$

and so, assuming $\rho \in [0,1]$, and using Jensen's inequality

$$\begin{aligned}
&\boldsymbol{E}\left\{\left(\sum_{\boldsymbol{u}'\neq\boldsymbol{u}}[P(\boldsymbol{u}')P(\boldsymbol{y}|\boldsymbol{X}(\boldsymbol{u}'))]^\beta\right)^\rho\right\} \\
\leq\ & [q(1-q)]^{N\beta\rho/2}(1-p)^{n\beta\rho}\sum_m[\boldsymbol{E}\{\zeta(\beta,m)\}]^\rho e^{\beta\rho mHN} \\
\doteq\ & [q(1-q)]^{N\beta\rho/2}(1-p)^{n\beta\rho}\sum_m\sum_\delta \exp\left\{n\rho\left[\frac{1}{\theta}h\left(\frac{1+m}{2}\right)+h(\delta)-\ln 2-\beta B\delta\right]\right\}\cdot e^{\beta\rho mHN} \\
\doteq\ & [q(1-q)]^{N\beta\rho/2}(1-p)^{n\beta\rho}\times \\
& \exp\left\{n\rho\left[\frac{1}{\theta}h\left(\frac{1+\tanh(\beta H)}{2}\right)+h(p_\beta)-\ln 2-\beta Bp_\beta+\frac{\beta H}{\theta}\tanh(\beta H)\right]\right\} \qquad (23)
\end{aligned}$$

We see that the magnetization that dominates the Gallager bound is the paramagnetic magnetization. By plugging this expression back into the bound on $\bar{P}_e$, we get the error exponent:

$$\begin{aligned}
E &\stackrel{\Delta}{=} -\lim_{N\to\infty}\frac{\ln\bar{P}_e}{N} \\
&\geq -\ln[q^{1-\rho\beta}+(1-q)^{1-\rho\beta}]+\theta[\gamma(1-\rho\beta)-\ln 2]-\frac{\beta\rho}{2}\ln[q(1-q)]-\beta\rho\theta\ln(1-p)- \\
&\quad \rho\left[h\left(\frac{1+\tanh(\beta H)}{2}\right)+\theta[h(p_\beta)-\ln 2-\beta Bp_\beta]+\beta H\tanh(\beta H)\right] \\
&= -\ln[q^{1-\rho\beta}+(1-q)^{1-\rho\beta}]+\theta[\gamma(1-\rho\beta)-\ln 2]+\rho\beta F_p(\beta,H) \\
&= -\ln\{[p^{1-\rho\beta}+(1-p)^{1-\rho\beta}]^\theta\cdot[q^{1-\rho\beta}+(1-q)^{1-\rho\beta}]\}+\rho\beta F_p(\beta,H) \qquad (24)
\end{aligned}$$

Here, unlike in the computation of the correct decoding exponent, there is a mismatch between the phase in the $H - T$ plane at which the decoder operatively works, and the phase at which $\bar{P}_e$ is analyzed: While the former is ferromagnetic, the latter is paramagnetic regardless of the temperature.

References

[1] T. Hosaka and Y. Kabashima, "Statistical mechanical approach to error exponents of lossy data compression," *J. Physical Society of Japan*, vol. 74, no. 1, pp. 488–497, January 2005.

[2] T. Mutayama, "Statistical mechanics of the data compression theorem," *J. Phys. A: Math. Gen.*, vol. 35, pp. L95–L100, 2002.

[3] Y. Kabashima and T. Hosaka, "Statistical mechanics for source coding with a fidelity criterion," *Progress of Theoretical Physics*, Supplement no. 157, pp. 197–204, 2005.

[4] K. Tadaki, "A statistical mechanical interpretation of instantaneous codes," *Proc. ISIT 2007*, pp. 1906–1910, Nice, France, June 2007.

[5] M. Mézard and A. Montanari, *Information, Physics and Computation*, draft, November 9, 2007. Available on–line at: [http://www.stanford.edu/∼montanar/BOOK/book.html].

[6] I. Kanter and D. Saad, "Error–correcting codes that nearly saturate Shannon's bound," *Physical Review Letters*, vol. 83, no. 13, pp. 2660–2663, September 1999.

[7] A. Procacci and B. Scoppola, "Statistical mechanics approach to coding theory," *J. of Statistical Physics*, vol. 96, nos. 3/4, pp. 907–912, 1999.

[8] N. Sourlas, "Spin–glass models as error–correcting codes," *Nature*, pp. 693–695, vol. 339, June 1989.

[9] N. Sourlas, "Spin glasses, error–correcting codes and finite–temperature decoding," *Europhysics Letters*, vol. 25, pp. 159–164, 1994.

[10] Y. Kabashima, N. Sazuka, K. Nakamura, and D. Saad, "Tighter decoding reliability bound for Gallager's error–correcting code," *Physical Review E*, vol. 64, pp. 046113-1–046113-4, 2001.

[11] Y. Kabashima and D. Saad, "Statistical mechanics of error correcting codes," *Europhysics Letters*, vol. 45, no. 1, pp. 97–103, 1999.

[12] O. Shental and I. Kanter, "Shannon capacity of infinite–range spin–glasses," technical report, Bar Ilan University, 2005.

[13] T. Mora and O. Rivoire, "Statistical mechanics of error exponents for error–correcting codes," arXiv:cond-mat/0606696, June 2006.

[14] A. Montanari, "Two lectures on iterative decoding and statistical mechanics," arXiv:cond-mat/0512296, December 14, 2005.

[15] S. Franz, M. Leone, A. Montanari, and F. Ricci–Tersenghi, "The dynamic phase transition for decoding algorithms," arXiv:cond-mat/020505, May 2, 2002.

[16] A. Montanari and R. Urbanke, "Modern coding theory: the statistical mechanics and computer science point of view," preprint 2007.

[17] A. Montanari, "The glassy phase of Gallager codes," arXiv:cond-mat/0104079v1, April 4, 2001.

[18] P. Ruján, "Finite temperature error–correcting codes," *Phys. Rev. Let.*, vol. 70, no. 19, pp. 2968–2971, May 1993.

[19] A. Montanari, "Tight bounds for LDPC and LDGM codes under MAP decoding," *IEEE Trans. Inform. Theory*, vol. 51, no. 9, pp. 3221–3246, September 2005.

[20] T. C. Dorlas and J. R. Wedagedera, "Phase diagram of the random energy model with higher order ferromagnetic term and error correcting codes due to Sourlas," *Phys. Rev. Lett.*, vol. 83, no. 21, pp. 4441–4444, November 1999.

[21] D. B. Saaskyan, "Solution of dilute Derrida model," *Teo. Mat. Fiz.*, vol. 97, pp. 1199–1205, 1993.

[22] D. B. Saaskyan, "Finite–volume effects in the ferromagnetic phase of the Derrida model at absolute zero in connection with coding," *JETP Lett.*, vol. 61, no. 8, pp. 650–654, April 1995.

[23] A. E. Allakhverdyan and D. B. Saaskyan, “Finite–volume corrections to the magnetization in the spin–glass phase of the Derrida model,” *Theoretical and Mathematical Physics*, vol. 109, no. 3, pp. 1574–1577, 1996.

[24] A. E. Allakhverdyan and D. B. Saaskyan, “Finite–size effects for the dilute coupling Derrida model,” *Nuclear Physics B*, vol. 498, pp. 604–618, 1997.

[25] D. B. Saaskian, “Error threshold in optimal coding, numerical criteria, and classes of universality for complexity,” *Phys. Rev.* E 71 016126, 2005.

[26] B. Derrida, “Random–energy model: limit of a family of disordered models,” *Phys. Rev. Lett.*, vol. 45, no. 2, pp. 79–82, July 1980.

[27] B. Derrida, “The random energy model,” *Physics Reports* (Review Section of Physics Letters), vol. 67, no. 1, pp. 29–35, 1980.

[28] B. Derrida, “Random–energy model: an exactly solvable model for disordered systems,” *Phys. Rev. B*, vol. 24, no. 5, pp. 2613–2626, September 1981.

[29] N. Merhav, “Relations between random coding exponents and the statistical physics of random codes,” submitted to *IEEE Trans. Inform. Theory*, August 2007. Available on–line at: [http://www.ee.technion.ac.il/people/merhav/papers/p117.pdf].

[30] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Second Edition, John Wiley & Sons, 2006.

[31] R. G. Gallager, *Information Theory and Reliable Communication*, John Wiley & Sons, 1968.

[32] F. Mandl, *Statistical Physics*, John Wiley & Sons, 1971.

[33] S. F. Edwards and P. W. Anderson, “Theory of spin glasses,” *J. Phys. F*, vol. 5, pp. 965–974, 1975.

[34] D. Sherrington and S. Kirkpatrick, “Solvable model of a spin glass,” *Phys/ Rev. Lett.*, vol. 35, pp. 1792–1796, 1975.

[35] B. Derrida and E. Gardner, “Solution of the generalised random energy model,” *J. Phys. C: Solid State Phys.*, vol. 19, pp. 2253–2274, 1986.

[36] N. Merhav, "Error exponents of erasure/list decoding revisited via moments of distance enumerators," submitted to *IEEE Trans. Inform. Theory*, November 2007. Also, available on–line at: [http://www.ee.technion.ac.il/people/merhav/papers/p119.pdf].