Low-Density Parity-Check Codes—A Statistical Physics Perspective

RENATO VICENTE,[1,*] DAVID SAAD[1] AND YOSHIYUKI KABASHIMA[2]

[1]*Neural Computing Research Group, University of Aston, Birmingham B4 7ET, United Kingdom*
[2]*Department of Computational Intelligence and Systems Science, Tokyo Institute of Technology, Yokohama 2268502, Japan*

*Current affiliation: Departamento de Física Geral, Instituto de Física, Universidade de São Paulo, 05315-970, São Paulo–SP, Brazil; to whom correspondence should be addressed (rvicente@if.usp.br).

ISSN 1076-5670/02 $35.00

I. Introduction

A. Error Correction

The way we communicate has been deeply transformed during the twentieth century. Telegraph, telephone, radio, and television technologies have brought to reality instantaneous long distance communication. Satellite and digital technologies have made global high-fidelity communication possible.

Two obvious common features of modern digital communication systems are that typically the message to be transmitted (e.g., images, text, computer programs) is redundant and the medium used for transmission (e.g., deep-space, atmosphere, optical fibers, etc.) is noisy. The key issues in modern communication are, therefore, saving storage space and computing time by

eliminating redundancies (*source coding* or *compression*) and making transmissions reliable by employing error-correction techniques (*channel coding*). Shannon was one of the first to point out these key issues. In his influential 1948 papers, Shannon proved general results on the natural limits of compression and error-correction by setting up the framework for what is now known as information theory.

Shannon's channel coding theorem states that error-free communication is possible if some redundancy is added to the original message in the encoding process. A message encoded at rates R (message information content/codeword length) up to the channel capacity C_{channel} can be decoded with a probability of error that decays exponentially with the message length. Shannon's proof was nonconstructive and assumed encoding with unstructured random codes and impractical (nonpolynomial time) (Cover and Thomas, 1991) decoding schemes. Finding practical codes capable of reaching the natural coding limits is one of the central issues in coding theory.

To illustrate the difficulties that may arise when trying to construct high performance codes from first principles, we can use a simple geometric illustration. On the top left of Figure 1 we represent the space of words (a message is a sequence of words), and each circle represents one sequence of binary bits. The word to be sent is represented by a black circle in the left side figure. Corruption by noise in the channel is represented in the top right figure as

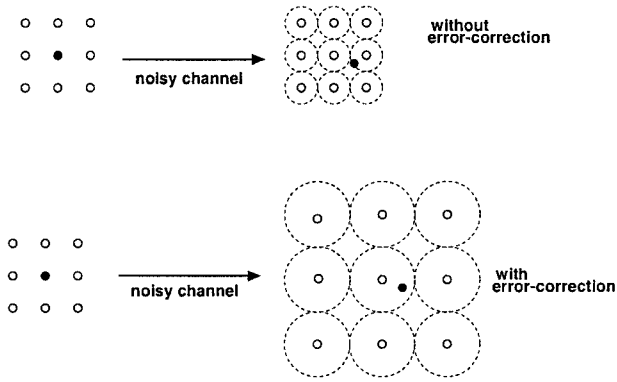

FIGURE 1. In the top figure we illustrate what happens when a word is transmitted without error correction. White circles represent possible word vectors, the black circle represents the word to be sent. The channel noise causes corruption of the original word that is represented by a drift in the top right picture. The dashed circles indicate decision boundaries in the receiver; in the case depicted, noise corruption leads to a transmission error. In the bottom figure we show qualitatively the error-correction mechanism. The redundant information changes the space geometry, increasing the distance between words. The same drift as in the top figure does not result in a transmission error.

a drift in the original word location. The circle around each word represent spheres that provide a decision boundary for each particular word; any signal inside a certain decision region is recognized as representing the word at the center of the sphere. In the case depicted in Figure 1 the drift caused by noise places the received word within the decision boundary of another word vector, causing a transmission error. Error-correction codes are based on mapping the original space of words onto a higher dimensional space in such a way that the typical distance between encoded words (codewords) increases. If the original space is transformed, the same drift shown in the top of Figure 1 is insufficient to push the received signal outside the decision boundary of the transmitted codeword (bottom figure).

Based on this simple picture we can formulate general designing criteria for good error-correcting codes: codewords must be short sequences of binary digits (for fast transmission), the code must allow for a large number of codewords (for a large set of words), and decision spheres must be as large as possible (for large error-correction capability). The general coding problem consists of optimizing one of these conflicting requirements given the other two. So, for example, if the dimension of the lattice and diameter of decision spheres are fixed, the problem is finding the lattice geometry that allows the densest possible sphere packing. This *sphere packing problem* is included in the famous list of problems introduced by Hilbert (it is actually part of the 18th problem). This problem can be solved for a very limited number of dimensions (Conway and Sloane, 1998), but is very difficult in general. As a consequence, constructive procedures are known only for a limited number of small codes.

For a long time, the best practical codes known were Reed–Solomon codes (RS) operating in conjunction with convolutional codes (*concatenated codes*). The current technological standard are RS codes, proposed in 1960, found almost everywhere from compact disks to mobile phones and digital television. Concatenated codes are the current standard in deep-space missions (e.g., Galileo mission) (MacWilliams and Sloane, 1977; Viterbi and Omura, 1979). Recently, *Turbo codes* (Berrou *et al.,* 1993) have been proven to outperform concatenated codes and are becoming increasingly more common. These codes are composed of two convolutional codes working in parallel and show practical performance close to Shannon's bound when decoded with iterative methods known as probability propagation, first studied in the context of coding by Wiberg (1996).

Despite the success of concatenated and Turbo codes, the current performance record is owned by Gallager's low-density parity-check codes (e.g., Chung, 2000; Davey, 1998, 1999). Gallager codes were first proposed in 1962 (Gallager, 1962, 1963) and then were all but forgotten soon after due to computational limitations of the time and due to the success of convolutional codes.

To give an idea of how parity-check codes operate, we exemplify with the simplest code of this type known as *Hamming code* (Hamming, 1950). A (7, 4) Hamming code, where (7, 4) stands for the number of bits in the codeword and input message, respectively, operates by adding 3 extra bits for each 4 message bits; this is done by a linear transformation $\boldsymbol{G}$, called the generator matrix, represented by:

$$\boldsymbol{G} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \end{pmatrix}. \tag{1}$$

When the generator matrix $\boldsymbol{G}$ is applied to a digital message $\boldsymbol{s} = (s_1, s_2, s_3, s_4)$, we get an encoded message defined by $\boldsymbol{t} = \boldsymbol{Gs}$ composed of 4 message bits plus redundant information (*parity-check*) as 3 extra bits $t_5 = s_2 \oplus s_3 \oplus s_4$, $t_6 = s_1 \oplus s_3 \oplus s_4$ and $t_7 = s_1 \oplus s_2 \oplus s_4$ ($\oplus$ indicates binary sums). One interesting point to note is that the transmitted message is such that $t_5 \oplus s_2 \oplus s_3 \oplus s_4 = 0$ and similarly for t_6 and t_7, what allows direct check of single corrupted bits. The decoding procedure relies in a second operator, known as parity-check matrix, with the property $\boldsymbol{HG} = 0$. For the generator (1) the parity-check matrix has the following form:

$$\boldsymbol{H} = \begin{pmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{pmatrix}. \tag{2}$$

The decoding procedure follows from the observation that the received message is corrupted by noise as $\boldsymbol{r} = \boldsymbol{Gs} \oplus \boldsymbol{n}$. By applying the parity-check matrix we get the *syndrome* $\boldsymbol{Hr} = \boldsymbol{Hn} = \boldsymbol{z}$. In the (7, 4) Hamming code the syndrome vector gives the binary representation for the position of the bit where an error has occurred (e.g., if $\boldsymbol{n} = (0, 0, 1, 0, 0, 0, 0)$, $z = (0, 1, 1)$). Due to this nice property, decoding is trivial and this code is known as a perfect single-error-correcting code (Hill, 1986).

Codes in the low-density parity-check family work along the same principles as the simple Hamming code above, the main differences being that they are much longer, the parity-check matrix is very sparse, and multiple errors can be corrected. However, low-density parity-check codes are not perfect and the decoding problem is, in general, significantly more difficult. Luckily, the sparseness of the matrix allows for the decoding process to be carried out by probability propagation methods similar to those employed in Turbo codes. Throughout

this chapter we concentrate on low-density parity-check codes (LDPC) that are state of the art concerning performance and operate along simple principles. We study four variations of LDPCs known as *Sourlas codes, Gallager codes, MacKay–Neal codes,* and *Cascading codes*.

B. Statistical Physics of Coding

The history of statistical physics application to error-correcting codes started in 1989 with a paper by Sourlas relating error-correcting codes to spin glass models (Sourlas, 1989). He showed that the Random Energy Model (Derrida, 1981b; Saakian, 1998; Dorlas and Wedagedera, 1999) can be thought of as an ideal code capable of saturating Shannon's bound at vanishing code rates. He also showed that the SK model (Kirkpatrick and Sherrington, 1978) could operate as a practical code.

In 1995, convolutional codes were analyzed by employing the transfer-matrix formalism and power series expansions (Amic and Luck, 1995).

In 1998, Sourlas work was extended for the case of finite code rates (Kabashima and Saad, 1999a) by employing the replica method. Recently, Turbo codes were also analyzed using the replica method (Montanari and Sourlas, 2000; Montanari, 2000).

In this chapter we present the extension of Sourlas work together with the analysis of other members in the family of LDPCs. We rely mainly on replica calculations (Kabashima *et al.,* 2000; Murayama *et al.,* 2000; Vicente *et al.,* 2000b) and mean-field methods (Kabashima and Saad, 1998; Vicente *et al.,* 2000a). The main idea is to develop the application of statistical physics tools for analyzing error-correcting codes. A number of results obtained are rederivations of well known results in information theory, while others put known results into a new perspective.

The main differences between the statistical physics analysis and traditional results in coding theory are the emphasis on very large systems from the start (thermodynamic limit) and the calculation of ensemble typical performances instead of worst-case bounds. In this sense statistical physics techniques are complementary to traditional methods. As a byproduct of our analysis we connect the iterative decoding methods of probability propagation with well-known mean-field techniques, presenting a framework that might allow a systematic improvement of decoding techniques.

C. Outline

In the next section we provide an overview of results and ideas from information theory that are relevant for understanding of the forthcoming sections. We also discuss more deeply linear encoding and parity-check decoding. We present

the probability propagation algorithm for computing approximate marginal probabilities efficiently and finish by introducing the statistical physics point of view of the decoding problem.

In Section III, we investigate the performance of error-correcting codes based on sparse generator matrices proposed by Sourlas. We employ replica methods to calculate the phase diagram for the system at finite code rates. We then discuss the decoding dynamics of the probability propagation algorithm. Sourlas codes are regarded as a first step toward developing techniques to analyze other more practical codes.

Section IV provides a statistical physics analysis for Gallager codes. These codes use a dense generator and a sparse parity-check matrix. The code is mapped onto a K-body interaction spin system and typical performance is obtained using the replica method. A mean-field solution is also provided by mapping the problem onto a Bethe-like lattice (Husimi cactus), recovering, in the thermodynamic limit, the replica symmetric results and providing a very good approximation for finite systems of moderate size. We show that the probability propagation decoding algorithm emerges naturally from the analysis, and its performance can be predicted by studying the free-energy landscape. A simple technique is introduced to provide upper bounds for the practical performance.

In Section V we investigate MacKay–Neal codes that are a variation of Gallager codes. In these codes, decoding involves two very sparse parity-check matrices, one for the signal with K nonzero elements in each row and a second for the noise with L nonzero elements. We map MN codes onto a spin system with $K + L$ interacting spins. The typical performance is again obtained by using a replica symmetric theory.

A statistical description for the typical PP decoding process for cascading codes is provided in Section VI. We use this description to optimize the construction parameters of a simple code of this type.

We close, in Section VII, with concluding remarks. Appendices with technical details are also provided.

II. Coding and Statistical Physics

A. Mathematical Model for a Communication System

In his papers from 1948, Shannon introduced a mathematical model (schematically represented in Figure 2) incorporating the most basic components of communication systems, and identified key problems and proved some general results. In the following we will introduce the main components of Shannon's communication model, the mathematical objects involved, as well as related general theorems.

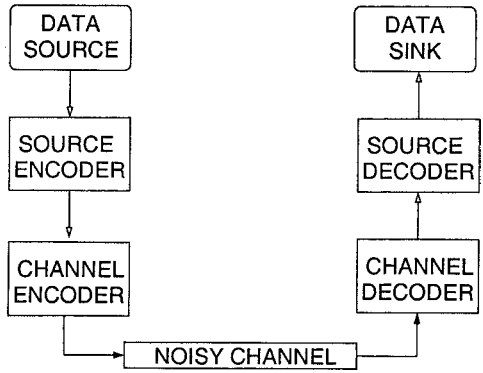

FIGURE 2. Mathematical model for a communication system. Each component is discussed in the text.

1. *Data Source and Sink*

A data source can be discrete or continuous. A discrete source is defined by the pair $(\mathcal{S}, \pi)$, where $\mathcal{S}$ is a set of m symbols (*alphabet*) and π is a probability measure over the space of sequences of symbols with any length (*messages*). In general, any discrete alphabet can be mapped onto sequences of [log m] Boolean digits $\{0, 1\}$. Continuous sources can always be made discrete at the expense of introducing some distortion to the signal (Cover and Thomas, 1991). A source is *memoryless* if each symbol in the sequence is independent of the preceding and succeeding symbols. A data sink is simply the receiver of decoded messages.

2. *Source Encoder and Decoder*

Data sources usually generate redundant messages that can be compressed to vectors of shorter average length. Source encoding, also known as data compression, is the process of mapping sequences of symbols from an alphabet $\mathcal{S}$ onto a shorter representation $\mathcal{A}$.

Shannon employed the statistical physics idea of *entropy* to measure the essential information content of a message. As enunciated by Khinchin (1957), the entropy of Shannon is defined as follows:

Definition II.1 **(Entropy)** Let

$$\begin{pmatrix} a_1 & a_2 & \cdots & a_m \\ p_1 & p_2 & \cdots & p_m \end{pmatrix}$$

be a finite scheme, where a_j are mutually exclusive events and p_j are associated probabilities with $\sum_{j=1}^{m} p_j = 1$. The entropy of the scheme in bits (or shannons) is defined as

$$H_2(A) = -\sum_{j=1}^{m} p_j \log_2 p_j. \tag{3}$$

The entropy is usually interpreted as the amount of information gained by removing the uncertainty and determining which event actually occurs.

Shannon (1948) posed and proved a theorem that establishes the maximal shortening of a message by compression as a function of its entropy. The *compression coefficient* can be defined as $\mu \equiv \overline{\lim}_{N\to\infty} \langle L_N \rangle / N$, where N is the original message length and $\langle L_N \rangle$ is the average length of compressed messages. As presented by Khinchin (1957), the theorem states:

Theorem II.1 ***(Source compression)*** *Given a discrete source with m symbols and entropy of H bits, for any possible compression code, the compression coefficient is such that*

$$\frac{H}{\log_2 m} \leq \mu$$

and there exists a code such that

$$\mu < \frac{H + \epsilon}{\log_2 m},$$

for arbitrarily small ϵ.

A compression scheme that yields a coefficient μ within the bounds above, given that the statistical structure π of the source is known, was proposed in 1952 by Huffman. Several practical algorithms are currently known and the design of more efficient and robust schemes is still a very active research area (Nelson and Gailly, 1995).

3. *Noisy Channels*

Message corruption during transmission can be described by a probabilistic model defined by the conditional probability $P(\boldsymbol{r} \mid \boldsymbol{t})$ where $\boldsymbol{t}$ and $\boldsymbol{r}$ represent transmitted and received messages, respectively. We can assume that in any of the channels used, only one component $t_j, j = 1, \ldots, M$ of the original message is being sent. If there is no interference effects between components, the channel is *memoryless* and the conditional probability factorizes as $P(\boldsymbol{r} \mid \boldsymbol{t}) = \prod_{j=1}^{M} P(r_j \mid t_j)$.

A memoryless channel model is specified by $(\mathcal{T}, P(r \mid t), \mathcal{R})$, where $\mathcal{T}$ and $\mathcal{R}$ are input and output alphabets and $P(r \mid t)$ transition probabilities. The information needed to specify t given the received signal r is the conditional entropy:

$$H_2(T \mid R) = -\sum_{r \in \mathcal{R}} P(r) \left[\sum_{t \in \mathcal{T}} P(t \mid r) \log_2 \left(P(t \mid r) \right) \right]. \tag{4}$$

The information on the original signal t conveyed by the received signal r is given by the mutual information $I(T; R) = H_2(T) - H_2(T \mid R)$, where $H_2(T)$ is defined in (3). The maximal information per bit that the channel can transport defines the *channel capacity* (Cover and Thomas, 1991).

Definition II.2 **(Channel capacity)** Given the channel model, the channel capacity is

$$C_{channel} = \max_{P(t)} I(T; R),$$

where *I(T; R)* is understood as a functional of the transmitted bits distribution *P(t)*. Thus, for example, if $C_{channel} = 1/2$, in the best case, 2 bits must be transmitted for each bit sent.

The following channel model (see MacKay, 1999, 2000a) is of particular interest in this chapter:

Definition II.3 **(Binary symmetric channel)** The memoryless binary symmetric channel (BSC) is defined by binary input and output alphabets $\mathcal{T} = \mathcal{R} = \{0, 1\}$ and by the conditional probability

$$P(r \neq t \mid t) = p \qquad P(r = t \mid t) = 1 - p. \tag{5}$$

The channel capacity of a BSC is given by

$$C_{BSC} = 1 - H_2(p) = 1 + p \, log(p) + (1 - p) \, log(1 - p)$$

In this chapter, we concentrate on the binary symmetric channel due to its simplicity and straightforward mapping onto an Ising spin system. However, there are several other channel types that have been examined in the literature and that play an important role in practical applications (Viterbi and Omura, 1979; Cover and Thomas, 1991). The most important of these is arguably the Gaussian channel; most of the analysis presented in this paper can be carried out in the case of the Gaussian channel as demonstrated in Kabashima and Saad (1999a) and Vicente *et al.* (1999).

message bits	index	codeword	message bits	index	codeword
0000	0	0000000	1000	8	1000011
0001	1	0001111	1001	9	1001100
0010	2	0010110	1010	10	1010101
0011	3	0011001	1011	11	1011010
0100	4	0100101	1100	12	1100110
0101	5	0101010	1101	13	1101101
0110	6	0110011	1110	14	1110010
0111	7	0111100	1111	15	1111111

FIGURE 3. Codebook for the (7, 4) Hamming code defined by (1).

4. *Channel Encoder and Decoder*

Highly reliable communication is possible even through noisy channels. It can be achieved by protecting a message with redundant information using a channel encoder defined as:

Definition II.4 **($(2^N, M)$ Code)** A code of rate $R = N/M$ is an indexed list (codebook) of 2^N codewords $\boldsymbol{t}(i) \in \mathcal{T}$ each of length M. Each index i in the codebook corresponds to a possible sequence of message bits.

In a digital system, a code can be regarded as a map of representations of 2^N symbols as Boolean sequences of N bits onto Boolean sequences of M bits. In Figure 3, we show the codebook for the Hamming code defined by (1) that is a $(2^4, 7)$ code. Each sequence of $N = 4$ message bits is indexed and converted in a codeword with $M = 7$ bits.

A decoding function $\boldsymbol{g}$ is a map of a channel output $\boldsymbol{r} \in \mathcal{R}$ back into a codeword. The probability that a symbol i is decoded incorrectly is given by the *probability of block error:*

$$p_{\text{Block}} = P\{\boldsymbol{g}(\boldsymbol{r}) \neq i \mid \boldsymbol{t} = \boldsymbol{t}(i)\}. \tag{6}$$

The average probability that a decoded bit $\hat{s}_j = g_j(\boldsymbol{r})$ fails to reproduce the original message bits is the *probability of bit error:*

$$p_b = \frac{1}{N} \sum_{j=1}^{N} P\{\hat{s}_j \neq s_j\}. \tag{7}$$

Shannon's coding theorem is as follows (Cover and Thomas, 1991; MacKay, 2000a).

Theorem II.2 ***(Channel coding)*** *The affirmative part of the theorem states: For every rate $R < C_{channel}$, there exists a sequence of $(2^{MR}, M)$ codes with maximum probability of block error $p_{Block}^{(M)} \to 0$. Conversely, any sequence of $(2^{MR}, M)$ codes with $p_{Block}^{(M)} \to 0$ must have $R \leq C_{channel}$.*

The negative part of the theorem is a corollary of the affirmative part and states:

Error-free communication above the capacity $C_{channel}$ is impossible. It is not possible to achieve a rate R with probability of bit error smaller than

$$p_b(R) = H_2^{-1}\left(1 - \frac{C_{channel}}{R}\right). \tag{8}$$

This nonconstructive theorem is obtained by assuming ensembles of random codes and impractical decoding schemes. No practical coding scheme (i.e., that can be encoded and decoded in polynomial time) that saturates the channel capacity is known to date. As Shannon's proof does not deal with complexity issues, there is no guarantee that such practical scheme exists at all.

B. Linear Error-Correcting Codes and the Decoding Problem

Linear error-correction codes add redundancy to the original message $\boldsymbol{s} \in \{0, 1\}^N$ through a linear map like:

$$\boldsymbol{t} = \boldsymbol{G}\boldsymbol{s} \pmod 2, \tag{9}$$

where $\boldsymbol{G}$ is an $M \times N$ Boolean matrix. The received message $\boldsymbol{r} = \boldsymbol{t} + \boldsymbol{n}$ is a corrupted version of the transmitted message. In the simplest form, optimal decoding consists of finding an optimal estimate $\hat{\boldsymbol{s}}(\boldsymbol{r})$ assuming a model for the noisy channel $P(\boldsymbol{r} \mid \boldsymbol{t})$ and a prior distribution for the message source $P(\boldsymbol{s})$.

The definition of the optimal estimator depends on the particular task and loss function assumed. An optimal estimator is defined as follows (see Iba, 1999, and references therein):

Definition II.5 **(Optimal estimator)** An optimal estimator $\hat{\boldsymbol{s}}(\boldsymbol{r})$ for a loss function $L(\boldsymbol{s}, \hat{\boldsymbol{s}}(\boldsymbol{r}))$ minimizes the average of L in relation to the posterior distribution $P(\boldsymbol{s} \mid \boldsymbol{r})$.

A posterior probability of messages given the corrupted message received can be easily found by applying Bayes theorem:

$$P(\boldsymbol{s} \mid \boldsymbol{r}) = \frac{P(\boldsymbol{r} \mid \boldsymbol{t})\, \delta(\boldsymbol{t}; \boldsymbol{G}\boldsymbol{s})\, P(\boldsymbol{s})}{\sum_s P(\boldsymbol{r} \mid \boldsymbol{t})\, \delta(\boldsymbol{t}; \boldsymbol{G}\boldsymbol{s})\, P(\boldsymbol{s})}, \tag{10}$$

where $\delta(x; y) = 1$ if $x = y$ and $\delta(x; y) = 0$, otherwise.

If we define our task to be the decoding of perfectly correct messages (i.e., we are interested in minimizing the probability of block error p_{Block}), we have to employ a two-valued loss function that identifies single mismatches:

$$L(\boldsymbol{s}, \hat{\boldsymbol{s}}(\boldsymbol{r})) = 1 - \prod_{j=1}^{M} \delta(s_j; \hat{s}_j). \tag{11}$$

An optimal estimator for this loss function must minimize the following:

$$\begin{aligned}\langle L(\boldsymbol{s}, \hat{\boldsymbol{s}}(\boldsymbol{r}))\rangle_{P(s|r)} &= \sum_{\boldsymbol{s}} P(\boldsymbol{s} \mid \boldsymbol{r}) L(\boldsymbol{s}, \hat{\boldsymbol{s}}(\boldsymbol{r})) \\ &= 1 - \sum_{\boldsymbol{s}} P(\boldsymbol{s} \mid \boldsymbol{r}) \prod_{j=1}^{M} \delta(s_j; \hat{s}_j) \\ &= 1 - P(\hat{\boldsymbol{s}} \mid \boldsymbol{r}). \end{aligned} \tag{12}$$

Clearly, the optimal estimator in this case is $\hat{\boldsymbol{s}} = \text{argmax}_S P(\boldsymbol{s} \mid \boldsymbol{r})$. This estimator is often called the *Maximum a Posteriori estimator* or simply *MAP*.

If we tolerate a certain degree of error in the decoded message (i.e., we are instead interested in minimizing the probability of bit error p_b), the loss function has to be an error counter like:

$$L(\boldsymbol{s}, \hat{\boldsymbol{s}}(\boldsymbol{r})) = -\sum_{j=1}^{M} s_j \hat{s}_j, \tag{13}$$

where we assume for simplicity the binary alphabet $\boldsymbol{s} \in \{\pm 1\}^N$. The optimal estimator must minimize the following:

$$\langle L(\boldsymbol{s}, \hat{\boldsymbol{s}}(\boldsymbol{r}))\rangle_{P(s|r)} = -\sum_{j=1}^{M} \langle s_j \rangle_{P(s|r)} \hat{s}_j. \tag{14}$$

An obvious choice for the estimator is

$$\begin{aligned}\hat{s}_j &= \frac{\langle s_j \rangle_{P(s|r)}}{|\ \langle s_j \rangle_{P(s|r)}\ |} \\ &= \text{sgn}\big(\langle s_j \rangle_{P(s|r)}\big) \\ &= \text{argmax}_{s_j} P(s_j \mid \boldsymbol{r}), \end{aligned} \tag{15}$$

where $P(s_j \mid \boldsymbol{r}) = \sum_{\{s_k : k \neq j\}} P(\boldsymbol{s} \mid \boldsymbol{r})$ is the marginal posterior distribution. As suggested by Eq. (15), this estimator is often called the *Marginal Posterior Maximizer* or *MPM* for short.

Decoding, namely, the computation of estimators, becomes a hard task, in general, as the message size increases. The MAP estimator requires finding a global maximum of the posterior over a space with 2^N points and the MPM estimator requires to compute long summations of 2^{N-1} terms for finding the two valued marginal posterior. The exponential scaling makes a naïve brute force evaluation quickly impractical. An alternative is to use approximate methods to evaluate posteriors. Popular methods are Monte Carlo sampling and the computationally more efficient probability propagation. In the sequence we will discuss the latter.

C. Probability Propagation Algorithm

The probabilistic dependencies existing in a code can be represented as a bipartite graph (Lauritzen, 1996) where nodes in one layer correspond to the M received bits r_μ and nodes in the other layer to the N message bits s_j. The connections between the two layers are specified by the generator matrix $\boldsymbol{G}$. Decoding requires evaluation of posterior probabilities when the received bits $\boldsymbol{r}$ are known (*evidence*).

The evaluation of the MPM estimator requires the computation of the following marginal joint distribution:

$$\begin{aligned} P(s_j, \boldsymbol{r}) &= \sum_{\{s_i : i \neq j\}} P(\boldsymbol{s} \mid \boldsymbol{r}) P(\boldsymbol{r}) \\ &= \sum_{\{s_i : i \neq j\}} P(\boldsymbol{r} \mid \boldsymbol{s}) P(\boldsymbol{s}) \\ &= \sum_{\{s_i : i \neq j\}} \prod_{\mu=1}^{M} P(r_\mu \mid s_{i_1} \cdots s_{i_K}) \prod_{j=1}^{N} P(s_j), \end{aligned} \tag{16}$$

where $s_{i_1} \cdots s_{i_K}$ are message bits composing the transmitted bit $t_\mu = (Gs)_\mu = s_{i_1} \oplus \cdots \oplus s_{i_K}$ and $\boldsymbol{r}$ is the message received. Equation (16) shows a complex partial factorization that depends on the structure of the generator matrix $\boldsymbol{G}$. We can encode this complex partial factorization on a directed graph known as a *Bayesian network* (Pearl, 1988; Castillo *et al.,* 1997; Jensen, 1996; Kschischang and Frey, 1998; Aji and McEliece, 2000; Frey, 1998, Kschischang *et al.,* 2001). As an example, we show in Figure 4 a simple directed bipartite graph encoding the following joint distribution:

$$\begin{aligned} P(s_1, \ldots, s_4, r_1, \ldots, r_6) = {} & P(r_1 \mid s_1, s_2, s_3) P(r_2 \mid s_3) P(r_3 \mid s_1, s_2) \\ & \times P(r_4 \mid s_3, s_4) P(r_5 \mid s_3) P(r_6 \mid s_3) \\ & \times P(s_1) P(s_2) P(s_3) P(s_4) \end{aligned} \tag{17}$$

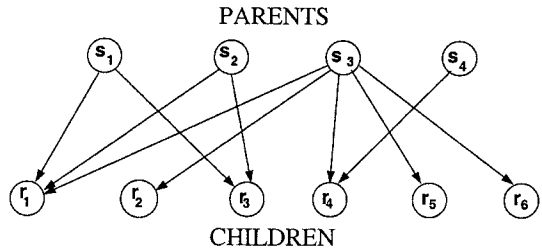

FIGURE 4. Bayesian network representing a linear code of rate 2/3. If there is an arrow from a vertex s_j to a vertex r_μ, s_j is said to be a *parent* and r_μ is said to be a *child*.

The generator matrix for the code in Figure 4 is:

$$\boldsymbol{G} = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}. \tag{18}$$

Given $\boldsymbol{r}$, an exact evaluation of the marginal joint distribution (16) in a space of binary variables $\boldsymbol{s} \in \{\pm 1\}^N$ would require $(N + M)(2^{N-1} - 1) + 1$ operations. In 1988, Pearl proposed an iterative algorithm that requires $\mathcal{O}(N)$ computational steps to calculate approximate marginal probabilities using Bayesian networks. This algorithm is known as *belief propagation* (Pearl, 1988), *probability propagation* (Kschischang and Frey, 1998), *generalized distributive law* (Aji and McEliece, 2000) or *sum-product algorithm* (Frey, 1998; Kschischang *et al.,* 2001; see also Opper and Saad, 2001).

The probability propagation algorithm is exact when the Bayesian network associated to the particular problem is free of loops. To introduce the probability propagation algorithm we start with the simple chain in Figure 5, which represents the following joint distribution:

$$p(s_1, s_2, s_3, s_4, s_5) = p(s_1)p(s_2 \mid s_1)p(s_3 \mid s_2)p(s_4 \mid s_3)p(s_5 \mid s_4). \tag{19}$$

Suppose now that we would like to compute $p(s_3)$, we would then have to compute:

$$p(s_3) = \sum_{s_1, s_2, s_4, s_5} p(s_1)p(s_2 \mid s_1)p(s_3 \mid s_2)p(s_4 \mid s_3)p(s_5 \mid s_4). \tag{20}$$

A brute force evaluation of (20) would take $5 \times (2^4 - 1) + 1 = 61$ operations in a binary field. The probability propagation algorithm reduces significantly the number of operations needed by rationalizing the order in which they are performed. For Figure 5 we can start by marginalizing vertex s_5 and writing:

$$R_{54}(s_4) = \sum_{s_5} p(s_5 \mid s_4). \tag{21}$$

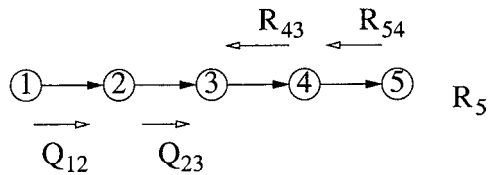

FIGURE 5. Marginal probabilities can be calculated exactly in a Bayesian chain. *R-messages* flow from a child to a parent and *Q-messages* flow from a parent to a child.

The function $R_{54}(s_4)$ can be regarded as a vector (a *message*) carrying information about vertex s_5. In a similar way we can write:

$$R_{43}(s_3) = \sum_{s_4} p(s_4 \mid s_3) R_{54}(s_4). \tag{22}$$

Again $R_{43}(s_3)$ can be seen as a message carrying information about vertices s_4 and s_5. Note that we can write (21) in the same form as (22) by assuming that $R_5(s_5) = 1$ if s_5 is not given or $R_5(s_5) = \delta(s_5; s^*)$ if $s_5 = s^*$, where $\delta(x; y) = 1$ if $x = y$ and $\delta(x; y) = 0$, otherwise.

We can also gather information from vertices to the left of s_3. Firstly, we marginalize s_1 by introducing:

$$Q_{12}(s_1) = p(s_1). \tag{23}$$

We then propagate the message $Q_{12}(s_1)$ to s_2 producing a new message:

$$Q_{23}(s_2) = \sum_{s_1} Q_{12}(s_1) p(s_2 \mid s_1). \tag{24}$$

The marginal probability $p(s_3)$ can be finally computed by:

$$\begin{aligned} p(s_3) &= \sum_{s_2} Q_{23}(s_2) R_{43}(s_3) p(s_3 \mid s_2) \\ &= \sum_{s_2} \sum_{s_1} Q_{12}(s_1) p(s_2 \mid s_1) \sum_{s_4} p(s_4 \mid s_3) R_{54}(s_4) p(s_3 \mid s_2) \\ &= \sum_{s_2} \sum_{s_1} p(s_1) p(s_2 \mid s_1) \sum_{s_4} p(s_4 \mid s_3) \sum_{s_5} p(s_5 \mid s_4) \\ &= \sum_{s_1, s_2, s_4, s_5} p(s_1) p(s_2 \mid s_1) p(s_3 \mid s_2) p(s_4 \mid s_3) p(s_5 \mid s_4). \end{aligned} \tag{25}$$

The evaluation of $p(s_3)$ using probability propagation is exact and requires only 16 operations, much less than the 61 operations required for the brute force calculation.

A slightly more complex situation is shown in Figure 6 representing the following joint distribution:

$$\begin{aligned} p(s_1, \ldots, s_{12}) = {} & p(s_6) p(s_8) p(s_9) p(s_{10}) p(s_{11}) p(s_{12}) p(s_1 \mid s_{10}) p(s_2 \mid s_{11}, s_{12}) \\ & \times p(s_3 \mid s_1, s_2, s_9) p(s_4 \mid s_3, s_8) p(s_5 \mid s_3, s_6) p(s_7 \mid s_4). \end{aligned} \tag{26}$$

Suppose that the variables are binary, s_7 and s_5 are given evidence vertices and we would like to compute the marginal $p(s_3)$. A brute force evaluation would require $11 \times (2^9 - 1) + 1 = 5622$ operations.

In general, we can just initialize the messages with random values, or make use of prior knowledge that may be available, and update the vertices in a

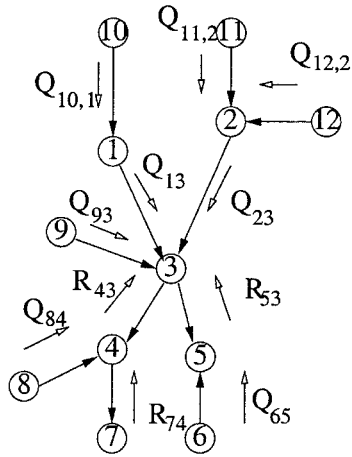

FIGURE 6. Marginal probabilities also can be calculated exactly in a Bayesian tree.

random order, but this may require several iterations for convergence to the correct values. In the particular case of trees there is an obvious optimal scheduling that takes only one iteration per vertex to converge: start at the leaves (vertices with a single edge connected to them) and proceed to the next internal level until the intended vertex. For the tree in Figure 6, the optimal schedule would be as follows:

- $Q_{11,2}$, $Q_{12,2}$, $Q_{10,1}$, Q_{65}, Q_{93}, Q_{84} and Q_{74}
- Q_{13}, Q_{23} and R_{43}, R_{53}

The Q-messages are just the prior probabilities:

$$Q_{j\mu}(s_j) = p(s_j), \tag{27}$$

where $j = 6, 8, 9, 10, 11, 12$.

The R-message between s_7 and s_4 is:

$$R_{74}(s_4) = \sum_{s_7} R_7(s_7) p(s_7 \mid s_4), \tag{28}$$

where $R_7(s_7) = \delta(s_7, s_7^*)$ and s_7^* is the value fixed by the evidence.

Following the schedule, we have the following Q-messages:

$$Q_{13}(s_1) = \sum_{s_{10}} p(s_1 \mid s_{10}) Q_{10,1}(s_{10}) \tag{29}$$

$$Q_{23}(s_2) = \sum_{s_{11}, s_{12}} p(s_2 \mid s_{11}, s_{12}) Q_{11,2}(s_{11}) Q_{12,2}(s_{12}). \tag{30}$$

The remaining R-messages are:

$$R_{43}(s_3) = \sum_{s_4, s_8} p(s_4 \mid s_3, s_8) Q_{84}(s_8) R_{74}(s_4) \tag{31}$$

$$R_{53}(s_3) = \sum_{s_6, s_5} p(s_5 \mid s_3, s_6) Q_{65}(s_6) R_5(s_5), \tag{32}$$

where $R_5(s_5) = \delta(s_5, s_5^*)$ and s_5^* is the value fixed by the evidence.

Finally we can fuse all the messages in the vertex s_3 as follows:

$$p(s_3) = \sum_{s_1, s_2, s_9} p(s_3 \mid s_1, s_2, s_9) Q_{13}(s_1) Q_{23}(s_2) R_{43}(s_3) R_{53}(s_3) Q_{93}(s_9). \tag{33}$$

By substituting the expressions for the messages in (33), it is relatively straightforward to verify that this expression gives the exact value for the marginal of (26). In this case, the probability propagation algorithm requires only 432 operations against 5622 operations required by the brute force evaluation.

We can now summarize the rules for calculating the message that flows through a particular edge:

- Multiply all incoming messages by the local probability table (for example: $p(s_3 \mid s_1, s_2, s_9)$ for vertex s_3) and sum over all vertices not attached to the edge that carries the outgoing message.
- Both Q and R messages must be only functions of the parent in the edge through which the message is flowing.

Probability propagation is only exact if the Bayesian network associated has no cycles. However, we can blindly apply the same algorithm in a general graph hoping that convergence to a good approximation is attained. In this kind of application there is no obvious optimal schedule and nodes can be updated serially, in parallel, or randomly.

Before writing the probability propagation equations for a general graph, let us first provide some definitions. Two vertices s_j and r_μ are adjacent if there is an edge connecting them. If there is an arrow from s_j to r_μ, s_j is said to be a *parent* and r_μ a *child*. The children of s_j are denoted by $\mathcal{M}(j)$ and the parents of r_μ are $\mathcal{L}(\mu)$. Linear codes are specified by bipartite graphs (as in Fig. 4) where all parents are in one layer and all children in the other layer. A *message* is a probability vector $Q = (Q^0, Q^1)$ with $Q^0 + Q^1 = 1$. The probability propagation algorithm in a bipartite graph operates by passing messages between the two layers through the connection edges, first forward from the top layer (parents) to the bottom layer (children), then backward, and so on iteratively. Child-to-parent messages (backward messages in Fig. 4) are R-messages denoted $R_{\mu j}$, while parent-to-child messages (forward messages) are Q-messages denoted by $Q_{j\mu}$.

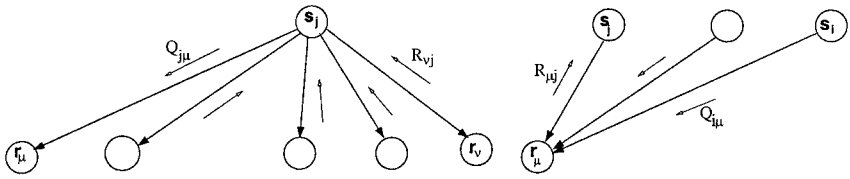

FIGURE 7. Left side: forward (Q) message from parent to child. Right side: backward (R) message from child to parent.

With the help of Figure 7 using the algorithm above, the forward (Q) messages between a parent s_j and child r_μ are just (see also Davey, 1999):

$$Q^a_{j\mu} = P(S_j = a \mid \{J_\nu : \nu \in \mathcal{M}(j)\mu\}) \tag{34}$$

$$= \alpha_{\mu j}\, p(s_j = a) \prod_{\nu \in \mathcal{M}(j)\setminus\mu} R^a_{\nu j}, \tag{35}$$

where $\alpha_{\mu j}$ is a required normalization, $\mathcal{M}(j) \setminus \mu$ stands for all elements in the set $\mathcal{M}(j)$ except μ.

Similarly, we can get the expression for the backward (R) messages between child r_μ and parent s_j:

$$R^a_{\mu j} = \sum_{\{s_i : i \in \mathcal{L}(\mu)\setminus j\}} P(r_\mu \mid s_j = a, \{s_i : i \in \mathcal{L}(\mu) \setminus j\}) \prod_{i \in \mathcal{L}(\mu)\setminus j} Q^{s_i}_{i\mu}. \tag{36}$$

An approximation for the marginal posterior can be obtained by iterating Eqs. (34) and (36) until convergence or some stopping criteria is attained, and fusing all incoming information to a parent node by calculating:

$$Q^a_j = \alpha_j\, p(s_j = a) \prod_{\nu \in \mathcal{M}(j)} R^a_{\mu j}, \tag{37}$$

where α_j is a normalization Q^a_j is an approximation for the marginal posterior $P(s_j \mid \boldsymbol{r})$. Initial conditions can be set to the prior probabilities $Q^s_{j\mu} = p(s)$.

It is clear (see also Pearl, 1988) that the probability propagation (PP) algorithm is exact if the associated graph is a tree and that the convergence for the exact marginal posterior occurs within a number of iterations proportional to the diameter of the tree. However, graphs defining error-correcting codes always have cycles and it has been observed empirically that decoding with the PP algorithm also yields good results (Frey and MacKay, 1998; Cheng, 1997) in spite of that.

There are a number of studies of probability propagation in loopy graphs with a single cycle (Weiss, 1997) and describing Gaussian joint distributions (Freeman, 1999), but no definite explanation for its good performance in this case is known to date.

D. Low-Density Parity-Check Codes

Marginal posteriors can be calculated in $\mathcal{O}(NK)$ steps, where K is the average connectivity of a child node, by using probability propagation. Therefore, the use of very sparse generator matrices ($\sum_{\mu j} G_{\mu j} = \mathcal{O}(N)$) seems favorable. Moreover, it is possible to prove that the probability of a cycle-free path of length l in a random graph decays with $\mathcal{O}(K^l/N)$, that indicates that small cycles are harder to find if the generator matrix is very sparse and that PP decoding is expected to provide better approximations for the marginal posterior (no proof is known for this statement). Encoding is also faster if very sparse matrices are used, requiring $\mathcal{O}(N)$ operations. Despite the advantages, the use of very sparse matrices for encoding has the serious drawback of producing codewords that differ in only $\mathcal{O}(K)$ bits from each other, which leads to a high probability of undetectable errors. Codes with sparse generator matrices are known as *Sourlas codes* and will be our object of study in the next section.

A solution for the bad distance properties of sparse generator codes is to use a dense matrix for encoding (providing a minimum distance between codewords of $\mathcal{O}(N)$), while decoding is carried out in a very sparse graph, allowing efficient use of PP decoding. The method known as parity-check decoding (Hill, 1986; Viterbi and Omura, 1979) is suitable in this situation, as encoding is performed by a generator matrix $\boldsymbol{G}$, while decoding is done by transforming the corrupted received vector $\boldsymbol{r} = \boldsymbol{G}\boldsymbol{s} + \boldsymbol{n}$ (mod 2) with a suitable parity-check matrix $\boldsymbol{H}$ having the property $\boldsymbol{H}\boldsymbol{G}$ (mod 2) $= 0$, yielding the *syndrome vector* $z = \boldsymbol{H}\boldsymbol{n}$ (mod 2).

Decoding reduces to finding the most probable vector $\boldsymbol{n}$ when the syndrome vector z is known, namely, performing MPM estimates that involve the calculation of the marginal posterior $P(n_j \mid z)$. In 1999, MacKay proved that this decoding method can attain vanishing block error probabilities up to the channel capacity if optimally decoded (not necessarily practically decoded).

This type of decoding is the basis for the three families of codes (*Gallager, MacKay–Neal,* and *cascading*) that we study in this chapter.

E. Decoding and Statistical Physics

The connection between spin systems in statistical physics and digital error correcting codes, first noted by Sourlas (1989), is based on the existence of a simple isomorphism between the additive Boolean group ($\{0, 1\}, \oplus$) and the multiplicative binary group ($\{+1, -1\}, \cdot$) defined by:

$$S \cdot X = (-1)^{s \oplus x}, \tag{38}$$

where $S, X \in \{+1, -1\}$ and $s, x \in \{0, 1\}$. Through this isomorphism, every addition on the Boolean group corresponds to a unique product on the binary group and vice-versa. A parity-check bit in a linear code is usually formed by a Boolean sum of K bits of the form $\bigoplus_{j=1}^{K} s_j$ that can be mapped onto a K-spin coupling $\prod_{j=1}^{K} S_j$. The same type of mapping can be applied to other error-correcting codes as convolutional codes (Sourlas, 1994b; Amic and Luck, 1995) and Turbo codes (Montanari and Sourlas, 2000; Montanari, 2000).

The decoding problem depends on posteriors like $P(\boldsymbol{S} \mid \boldsymbol{J})$, where $\boldsymbol{J}$ is the evidence (received message or syndrome vector). By applying Bayes' theorem this posterior can, in general, be written in the form:

$$P_{\alpha\gamma}(\boldsymbol{S} \mid \boldsymbol{J}) = \frac{1}{Z(\boldsymbol{J})} \exp\left[\ln P_{\alpha}(\boldsymbol{J} \mid \boldsymbol{S}) + \ln P_{\gamma}(\boldsymbol{S})\right], \tag{39}$$

where α and γ are hyperparameters assumed to describe features like the encoding scheme, source distribution, and noise level. This form suggests the following family of Gibbs measures:

$$P_{\alpha\beta\gamma}(\boldsymbol{S} \mid \boldsymbol{J}) = \frac{1}{Z} \exp\left[-\beta \mathcal{H}_{\alpha\gamma}(\boldsymbol{S}; \boldsymbol{J})\right] \tag{40}$$

$$\mathcal{H}_{\alpha\gamma}(\boldsymbol{S}; \boldsymbol{J}) = -\ln P_{\alpha}(\boldsymbol{J} \mid \boldsymbol{S}) - \ln P_{\gamma}(\boldsymbol{S}), \tag{41}$$

where $\boldsymbol{J}$ can be regarded as quenched disorder in the system. It is not difficult to see that the MAP estimator is represented by the ground state of the Hamiltonian (40), i.e., by the sign of thermal averages $\hat{S}_j^{\mathrm{MAP}} = \mathrm{sgn}(\langle S_j \rangle_{\beta\to\infty})$ at zero temperature. On the other hand, the MPM estimator is provided by the sign of thermal averages $\hat{S}_j^{\mathrm{MPM}} = \mathrm{sgn}(\langle S_j \rangle_{\beta=1})$ at temperature one. We have seen that if we are concerned with the probability of bit error p_e the optimal choice for an estimator is MPM, this is equivalent to decoding at finite temperature $\beta = 1$, known as the Nishimori temperature (Nishimori, 1980, 1993, 2001; Ruján, 1993).

The evaluation of typical quantities involves the calculation of averages over the quenched disorder (evidence) $\boldsymbol{J}$, namely, averages over:

$$P_{\alpha^*\gamma^*}(\boldsymbol{J}) = \sum_{S} P_{\alpha^*}(\boldsymbol{J} \mid \boldsymbol{S}) P_{\gamma^*}(\boldsymbol{S}), \tag{42}$$

where α^* and γ^* represent the "real" hyperparameters, in other words, the hyperparameters actually used for generating the evidence $\boldsymbol{J}$. Those "real" hyperparameters are, in general, not known to the receiver, but can be estimated from the data. To calculate these estimates we can start by writing free-energy like negative log-likelihoods for the hyperparameters:

$$\langle F(\alpha, \gamma) \rangle_{P_{\alpha^*\gamma^*}} = -\langle \ln P_{\alpha\gamma}(\boldsymbol{J}) \rangle_{P_{\alpha^*\gamma^*}}. \tag{43}$$

This log-likelihood can be regarded as measuring the typical plausibility of α and γ, given the data $\boldsymbol{J}$ (Berger, 1993). This function can be minimized to find the most plausible hyperparameters (known as *type II maximum likelihood hyperparameters* or just *ML-II hyperparameters*) (Berger, 1993).

The ML-II hyperparameters correspond in this case to $\alpha = \alpha^*$ and $\gamma = \gamma^*$, i.e., the "real" hyperparameters must be used in the posterior for decoding. This fact is a consequence of the following inequality:

$$\langle F(\alpha^*, \gamma^*)\rangle_{P_{\alpha^*\gamma^*}} \leq \langle F(\alpha, \gamma)\rangle_{P_{\alpha^*\gamma^*}}. \tag{44}$$

The proof of (44) follows directly from the information inequality (Iba, 1999; Cover and Thomas, 1991), i.e., the nonnegativity of the KL-divergence:

$$\begin{aligned} & D(P_{\alpha^*\gamma^*} \| P_{\alpha\gamma}) \geq 0 \\ & \left\langle \ln\left(\frac{P_{\alpha^*\gamma^*}(\boldsymbol{J})}{P_{\alpha\gamma}(\boldsymbol{J})}\right)\right\rangle_{P_{\alpha^*\gamma^*}} \geq 0 \\ & -\langle \ln P_{\alpha^*\gamma^*}(\boldsymbol{J})\rangle_{P_{\alpha^*\gamma^*}} \leq -\langle \ln P_{\alpha\gamma}(\boldsymbol{J})\rangle_{P_{\alpha^*\gamma^*}}. \end{aligned} \tag{45}$$

When the true and assumed hyperparameters agree, we say that we are at the *Nishimori condition* (Iba, 1999; Nishimori, 2001). At the Nishimori condition many calculations simplify and can be done exactly (for an example, see Appendix B.3). Throughout this chapter we assume, unless otherwise stated, the Nishimori condition.

For background reading about statistical physics methods in general, Nishimori's condition, and its relevance to the current calculation we refer the reader to Nishimori (2001).

III. Sourlas Codes

The code of Sourlas is based on the idea of using a linear operator $\boldsymbol{G}$ (*generator matrix*) to transform a message vector $\boldsymbol{s} \in \{0, 1\}^N$ onto a higher dimensional vector $\boldsymbol{t} \in \{0, 1\}^M$. The encoded vector is $\boldsymbol{t} = \boldsymbol{G}\boldsymbol{s} \pmod 2$, each bit t_k being the Boolean sum of K message bits (*parity-check*). This vector is transmitted through a noisy channel and a corrupted M dimensional vector $\boldsymbol{r}$ is received.

Decoding consists of producing an estimate $\hat{\boldsymbol{s}}$ of the original message. This estimate can be generated by considering a probabilistic model for the communication system. Reduced (order N) time/space requirements for the encoding process and the existence of fast (polynomial time) decoding algorithms are guaranteed by choosing sparse generator matrices, namely, a matrix $\boldsymbol{G}$ with exactly K nonzero elements per row and C nonzero elements per column, where K and C are of order 1. The rate of such a code, in the case of unbiased

messages, is evidently $R = N/M$, as the total number of nonzero elements in $\boldsymbol{G}$ is $MK = NC$ the rate is also $R = K/C$.

In the statistical physics language a binary message vector $\boldsymbol{\xi} \in \{\pm 1\}^N$ is encoded to a higher dimensional vector $\boldsymbol{J}^0 \in \{\pm 1\}^M$ defined as $J^0_{\langle i_1, i_2 \ldots i_K\rangle} = \xi_{i_1}\xi_{i_2}\cdots\xi_{i_K}$, where M sets of K indices are randomly chosen. A corrupted version $\boldsymbol{J}$ of the encoded message $\boldsymbol{J}^0$ has to be decoded for retrieving the original message. The decoding process is the process of calculating an estimate $\hat{\boldsymbol{\xi}}$ to the original message by minimizing a given expected loss $\langle\langle \mathcal{L}(\xi, \hat{\xi})\rangle_{P(J|\xi)}\rangle_{P(\xi)}$ averaged over the indicated probability distributions (Iba, 1999). The definition of the loss depends on the particular task; the overlap $\mathcal{L}(\xi, \hat{\xi}) = \sum_j \xi_j \hat{\xi}_j$ can be used for decoding binary messages. As discussed in Section II.B, an optimal estimator for this particular loss function is $\hat{\xi}_j = \operatorname{sign}\langle S_j\rangle_{P(S_j|J)}$ (Iba, 1999), where $\boldsymbol{S}$ is an N-dimensional binary vector representing the dynamic variables of the decoding process and $P(S_j \mid \boldsymbol{J}) = \sum_{S_k, k\neq j} P(\boldsymbol{S} \mid \boldsymbol{J})$ is the marginal posterior probability. Using Bayes theorem, the posterior probability can be written as:

$$\ln P(\boldsymbol{S} \mid \boldsymbol{J}) = \ln P(\boldsymbol{J} \mid \boldsymbol{S}) + \ln P(\boldsymbol{S}) + \text{const.} \tag{46}$$

The likelihood $P(\boldsymbol{J} \mid \boldsymbol{S})$ has the form:

$$P(\boldsymbol{J} \mid \boldsymbol{S}) = \prod_{\text{chosensets}} \sum_{J^0_{\langle i_1\cdots i_K\rangle}} P\left(J_{\langle i_1\cdots i_K\rangle} \middle| J^0_{\langle i_1\cdots i_K\rangle}\right) P\left(J^0_{\langle i_1\cdots i_K\rangle} \middle| \boldsymbol{S}\right). \tag{47}$$

The term $P(J^0_{\langle i_1\cdots i_K\rangle} \mid \boldsymbol{S})$ models the deterministic encoding process being:

$$P\left(J^0_{\langle i_1\cdots i_K\rangle} \middle| \boldsymbol{S}\right) = \delta\left(J^0_{\langle i_1\cdots i_K\rangle}; S_{i_1}\cdots S_{i_K}\right). \tag{48}$$

The noisy channel is modeled by the term $P(J_{\langle i_1\cdots i_K\rangle} \mid J^0_{\langle i_1\cdots i_K\rangle})$. For the simple case of a memoryless binary symmetric channel (BSC), $\boldsymbol{J}$ is a corrupted version of the transmitted message $\boldsymbol{J}^0$ where each bit is independently flipped with probability p during transmission, in this case (Sourlas, 1994a):

$$\begin{aligned}\ln P\left(J_{\langle i_1\cdots i_K\rangle} \middle| J^0_{\langle i_1\cdots i_K\rangle}\right) &= \frac{1}{2}\left(1 + J^0_{\langle i_1\cdots i_K\rangle}\right) \ln P\left(J_{\langle i_1\cdots i_K\rangle} \middle| + 1\right) \\ &\quad + \frac{1}{2}\left(1 - J^0_{\langle i_1\cdots i_K\rangle}\right) \ln P\left(J_{\langle i_1\cdots i_K\rangle} \middle| - 1\right) \\ &= \text{const} + \frac{1}{2}\ln\left(\frac{1-p}{p}\right) J_{\langle i_1\cdots i_K\rangle} J^0_{\langle i_1\cdots i_K\rangle}. \end{aligned} \tag{49}$$

Putting equations together, we obtain the following Hamiltonian:

$$\ln P(\boldsymbol{S} \mid \boldsymbol{J}) = -\beta_N \mathcal{H}(\boldsymbol{S}) + \text{const} \tag{50}$$

$$= \beta_N \sum_{\mu} \mathcal{A}_\mu J_\mu \prod_{i\in\mathcal{L}(\mu)} S_i + \beta'_N \sum_{j=1}^{N} S_j + \text{const}, \tag{51}$$

where a set of indices is denoted $\mathcal{L}(\mu) = \langle i_1, \ldots i_K \rangle$ and $\mathcal{A}$ is a tensor with the properties $\mathcal{A}_\mu \in \{0, 1\}$ and $\sum_{\{\mu; i \in \mathcal{L}(\mu)\}} \mathcal{A}_\mu = C \forall i$, which determines the M components of the codeword $\boldsymbol{J}^0$. The interaction term is at Nishimori's temperature $\beta_N = \frac{1}{2} \ln(\frac{1-p}{p})$ (Nishimori, 1980, 1993; Iba, 1999; Ruján, 1993), and $\beta'_N = \frac{1}{2} \ln(\frac{1-p_\xi}{p_\xi})$ is the *message prior temperature,* namely, the prior distribution of message bits is assumed to be $P(S_j = +1) = 1 - p_\xi$ and $P(S_j = -1) = p_\xi$.

The decoding procedure translates to finding the thermodynamic spin averages for the system defined by the Hamiltonian (50) at a certain temperature (Nishimori temperature for optimal decoding); as the original message is binary, the retrieved message bits are given by the signs of the corresponding averages.

The performance of the error-correcting process can be measured by the overlap between actual message bits and their estimates for a given scenario characterized by code rate, corruption process, and information content of the message. To assess the typical properties, we average this overlap over all possible codes $\mathcal{A}$ and noise realizations (possible corrupted vectors $\boldsymbol{J}$) given the message $\boldsymbol{\xi}$ and then over all possible messages:

$$\rho = \frac{1}{N} \left\langle \sum_{i=1}^{N} \xi_i \, \langle \operatorname{sign}\langle S_i \rangle \rangle_{\mathcal{A}, J|\xi} \right\rangle_\xi \tag{52}$$

Here $\operatorname{sign}\langle S_i \rangle$ is the sign of the spins thermal average corresponding to the Bayesian optimal decoding. The average error per bit is, therefore, given by $p_b = (1 - \rho)/2$.

The number of checks per bit is analogous to the spin system connectivity and the number of bits in each check is analogous to the number of spins per interaction. The code of Sourlas has been studied in the case of extensive connectivity, where the number of bonds $C \sim \binom{N-1}{K-1}$ scales with the system size. In this case it can be mapped onto known problems in statistical physics such as the SK (Kirkpatrick and Sherrington, 1978) ($K = 2$) and random energy (REM) (Derrida, 1981a) ($K \to \infty$) models. It has been shown that the REM saturates Shannon's bound (Sourlas, 1989). However, it has a rather limited practical relevance as the choice of extensive connectivity corresponds to a vanishingly small code rate.

A. Lower Bound for the Probability of Bit Error

It has been observed in Montanari and Sourlas (2000) that a sparse generator code can only attain vanishing probability of bit error if $K \to \infty$. This fact alone does not rule out the practical use of such codes as they can still be

used if a controlled probability of error is allowed or as part of a concatenated code.

Before engaging in a relatively complex analysis, it is of theoretical interest to establish a detailed picture of how the minimum bit error attainable decays with K. This can be done in quite a simple manner suggested in Montanari and Sourlas (2000). Let us suppose that messages are unbiased and random and that the channel is a BSC of noise level p. Assume, without loss of generality, that the message $\xi_j = 1$ for all j is sent. The bit error probability can be expressed as the sum $p_b = \sum_{l=1}^{N} p_b(l)$, where $p_b(l)$ represents the probability of decoding incorrectly any l bits. Clearly $p_b \geq p_b(1)$.

The probability of decoding incorrectly a single bit can be easily evaluated. A bit j engages in exactly C interactions with different groups of K bits in a way that their contribution to the Hamiltonian is:

$$\mathcal{H}_j = -S_j \sum_{\mu \in \mathcal{M}(j)} J_\mu \prod_{i \in \mathcal{L}(\mu) \setminus j} S_i, \tag{53}$$

where $\mathcal{M}(j)$ is the set of all index sets that contain j. If all bits but j are set to $S_i = 1$, an error in j only can be detected if its contribution to the Hamiltonian is positive; if $\sum_{\mu \in \mathcal{M}(j)} \mathcal{A}_\mu J_\mu \leq 0$ the error is undetectable. The probability of error in a single bit is therefore

$$p_b(1) = P\left\{ \sum_{\mu \in \mathcal{M}(j)} J_\mu \leq 0 \right\}, \tag{54}$$

where $\mathcal{A}_\mu = 1$ for exactly C terms and J_μ can be simply regarded as a random variable taking values $+1$ and -1 with probabilities $1 - p$ and p, respectively; therefore:

$$p_b \geq \sum_{l \in \mathbb{N}, C-2l \leq 0}^{l \leq C} \frac{C!}{(C-l)!\, l!} (1-p)^{C-l} p^l. \tag{55}$$

A lower bound for for p_b in the large C regime can be obtained by using the DeMoivre–Laplace limit theorem (Feller, 1950), writing:

$$p_b \geq \frac{1}{2} \mathrm{erfc}\left(\frac{(1-p)C}{8p} \right) \approx \frac{4p}{\sqrt{\pi}(1-p)C} \exp\left(-\frac{(1-p)^2 C^2}{64 p^2} \right), \tag{56}$$

where $\mathrm{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty du \exp(-u^2)$ and the asymptotic behavior is given in Gradshteyn and Ryzhik (1994, page 940). This bound implies that $K \to \infty$ is a necessary condition for a vanishing bit error probability in sparse generator codes at finite rates $R = K/C$.

B. Replica Theory for the Typical Performance of Sourlas Codes

To calculate the typical performance of Sourlas codes we employ the statistical physics technique known as *replica theory*.

To simplify analysis we use the gauge transformation (Fradkin *et al.,* 1978) $S_i \mapsto S_i \xi_i$ and $J_{\langle i_1 \cdots i_K \rangle} \mapsto J_{\langle i_1 \cdots i_K \rangle} \xi_{i_1} \cdots \xi_{i_K}$ that maps any general message to the configuration defined as $\xi_i^* = 1 \forall i$ (ferromagnetic configuration). By introducing the *external field* $F \equiv \beta'_N / \beta$ we rewrite the Hamiltonian in the form:

$$\mathcal{H}(\boldsymbol{S}) = - \sum_{\langle i_1 \cdots i_K \rangle} \mathcal{A}_{\langle i_1 \cdots i_K \rangle} \, J_{\langle i_1 \cdots i_K \rangle} S_{i_1} \cdots S_{i_K} - F \sum_{j=1}^{N} \xi_j S_j, \tag{57}$$

With the gauge transformation, the bits of the uncorrupted encoded message become $J^0_{\langle i_1 \cdots i_K \rangle} = 1$ and, for the BSC, the corrupted bits can be described as random variables with probability:

$$P(J) = (1 - p)\, \delta(J - 1) + p\, \delta(J + 1), \tag{58}$$

where p is the channel flip rate. For deriving the typical properties we calculate the free-energy following the replica theory prescription:

$$f = -\frac{1}{\beta} \lim_{N \to \infty} \frac{1}{N} \left. \frac{\partial}{\partial n} \right|_{n=0} \langle \mathcal{Z}^n \rangle_{\mathcal{A}, \xi, J}, \tag{59}$$

where $\langle \mathcal{Z}^n \rangle_{\mathcal{A}, \xi, J}$ represents an analytical continuation in the interval $n \in [0, 1]$ of the replicated partition function:

$$\langle \mathcal{Z}^n \rangle_{\mathcal{A}, \xi, J} = \mathrm{Tr}_{\{S_j^\alpha\}} \left\langle e^{\beta F \sum_{\alpha, k} \xi_k S_k^\alpha + \beta \sum_{\alpha, \mu} \mathcal{A}_\mu \, J_\mu \, S_{i_1}^\alpha \cdots S_{i_K}^\alpha} \right\rangle_{\mathcal{A}, J, \xi}. \tag{60}$$

The overlap ρ can be rewritten using gauged variables as:

$$\rho = \frac{1}{N} \sum_{i=1}^{N} \left\langle \langle \mathrm{sign} \langle S_i \rangle \rangle_{\mathcal{A}, J | \xi^*} \right\rangle_\xi, \tag{61}$$

where ξ^* denotes the transformation of a message ξ into the ferromagnetic configuration.

To compute the replicated partition function we closely follow Wong and Sherrington (1987a). We average uniformly over all codes $\mathcal{A}$ such that

$\sum_{\langle i_1=i,i_2\cdots i_K\rangle} \mathcal{A}_{\langle i_1\cdots i_K\rangle} = C\forall i$ to find:

$$\langle \mathcal{Z}^n \rangle_{\mathcal{A},\xi,J} = \exp\left\{ N \mathrm{Extr}_{q,\hat{q}} \left[C - \frac{C}{K} + \frac{C}{K} \left(\sum_{l=0}^{n} \mathcal{T}_l \sum_{\langle \alpha_1 \cdots \alpha_l \rangle} q^K_{\alpha_1 \cdots \alpha_l} \right) \right.\right.$$
$$- C \left(\sum_{l=0}^{n} \sum_{\langle \alpha_1 \cdots \alpha_l \rangle} q_{\alpha_1 \cdots \alpha_l} \hat{q}_{\alpha_1 \cdots \alpha_l} \right)$$
$$\left.\left. + \ln \mathrm{Tr}_{\{S^\alpha\}} \left\langle e^{\beta F \xi \sum_\alpha S^\alpha} \right\rangle_\xi \left(\sum_{l=0}^{n} \sum_{\langle \alpha_1 \cdots \alpha_l \rangle} \hat{q}_{\alpha_1 \cdots \alpha_l} S^{\alpha_1} \ldots S^{\alpha_l} \right)^C \right] \right\}, \tag{62}$$

where $\mathcal{T}_l = \langle \tanh^l(\beta J) \rangle_J$, as in Viana and Bray (1985), $q_0 = 1$ and $\mathrm{Extr}_{q,\hat{q}}$ $f(q,\hat{q})$ denotes the extremum of f (details in Appendix A.1). At the extremum of (62) the order parameters acquire a form similar to those of Wong and Sherrington (1987a):

$$\hat{q}_{\alpha_1,\ldots,\alpha_l} = \mathcal{T}_l \, q^{K-1}_{\alpha_1,\ldots,\alpha_l}$$
$$q_{\alpha_1,\ldots,\alpha_l} = \left\langle \left(\prod_{i=1}^{l} S^{\alpha_i} \right) \left(\sum_{l=0}^{n} \sum_{\langle \alpha_1 \cdots \alpha_l \rangle} \hat{q}_{\alpha_1 \cdots \alpha_l} S^{\alpha_1} \cdots S^{\alpha_l} \right)^{-1} \right\rangle_{\mathcal{X}}. \tag{63}$$

where

$$\mathcal{X} = \left\langle e^{\beta F \xi \sum_\alpha S^\alpha} \right\rangle_\xi \left(\sum_{l=0}^{n} \sum_{\langle \alpha_1 \cdots \alpha_l \rangle} \hat{q}_{\alpha_1 \cdots \alpha_l} S^{\alpha_1} \cdots S^{\alpha_l} \right)^C, \tag{64}$$

and $\langle \cdots \rangle_{\mathcal{X}} = \mathrm{Tr}_{\{S^\alpha\}} \left[(\cdots) \mathcal{X} \right] / \mathrm{Tr}_{\{S^\alpha\}} \left[(\cdots) \right]$.

To compute the partition function it is necessary to assume a replica symmetric (RS) ansätz. It can be done by introducing auxiliary fields $\pi(x)$ and $\hat{\pi}(y)$ (see also Wong and Sherrington, 1987a):

$$\hat{q}_{\alpha_1 \cdots \alpha_l} = \int dy \hat{\pi}(y) \tanh^l(\beta y),$$
$$q_{\alpha_1 \cdots \alpha_l} = \int dx \, \pi(x) \tanh^l(\beta x) \tag{65}$$

for $l = 1, 2, \ldots$.

Plugging (65) into the replicated partition function (62), taking the limit $n \to 0$ and using Eq. (59) (see Appendix A.2 for details):

$$
\begin{aligned}
f = -\frac{1}{\beta} Extr_{\pi,\hat{\pi}} \{ & \alpha \ln \cosh \beta \\
& + \alpha \int \left[\prod_{l=1}^{K} dx_l \, \pi(x_l) \right] \left\langle \ln \left[1 + \tanh \beta J \prod_{j=1}^{K} \tanh \beta x_j \right] \right\rangle_J \\
& - C \int dx \, dy \, \pi(x) \, \hat{\pi}(y) \, \ln \left[1 + \tanh \beta x \, \tanh \beta y \right] \\
& - C \int dy \, \hat{\pi}(y) \, \ln \cosh \beta y \\
& + \int \left[\prod_{l=1}^{C} dy_l \, \hat{\pi}(y_l) \right] \left\langle \ln \left[2 \cosh \beta \left(\sum_{j=1}^{C} y_j + F\xi \right) \right] \right\rangle_\xi \Bigg\}, \qquad (66)
\end{aligned}
$$

where $\alpha = C/K$. The saddle-point equations obtained by calculating functional variatons of Eq. (66) provide a closed set of relations between $\pi(x)$ and $\hat{\pi}(y)$

$$
\begin{aligned}
\pi(x) &= \int \left[\prod_{l=1}^{C-1} dy_l \, \hat{\pi}(y_l) \right] \left\langle \delta \left[x - \sum_{j=1}^{C-1} y_j - F\xi \right] \right\rangle_\xi \\
\hat{\pi}(y) &= \int \left[\prod_{l=1}^{K-1} dx_l \, \pi(x_l) \right] \left\langle \delta \left[y - \frac{\operatorname{atanh}\left(\tanh \beta J \prod_{j=1}^{K-1} \tanh \beta x_j \right)}{\beta} \right] \right\rangle_J .
\end{aligned} \qquad (67)
$$

Later we will show that this self-consistent pair of equations can be seen as a mean-field description of probability propagation decoding.

Using the RS ansätz one can find that the local field distribution is (see Appendix A.3):

$$
P(h) = \int \left[\prod_{l=1}^{C} dy_l \, \hat{\pi}(y_l) \right] \left\langle \delta \left[h - \sum_{j=1}^{C} y_j - F\xi \right] \right\rangle_\xi , \qquad (68)
$$

where $\hat{\pi}(y)$ is given by the saddle-point equations (67).

The overlap (52) can be calculated using:

$$\rho = \int dh \, \text{sign}(h) \, P(h). \tag{69}$$

The code performance is assessed by assuming a prior distribution for the message, solving the saddle-point equations (67) numerically and then computing the overlap.

For Eq. (66) to be valid, the fixed point given by (67) must be stable and the related entropy must be nonnegative. Instabilities within the RS space can be probed by calculating second functional derivatives at the extremum defining the free-energy (66). The solution is expected to be unstable within the space of symmetric replicas for sufficiently low temperatures (large β). For high temperatures we can expand the above expression around small β values to find the stability condition:

$$\langle J \rangle_J \langle x \rangle_\pi^{K-2} \geq 0 \tag{70}$$

The average $\langle x \rangle_\pi = \int dx \, \pi(x) x$ vanishes in the paramagnetic phase and is positive (nonzero when K is even) in the ferromagnetic phase, satisfying the stability condition. We now restrict our study to the unbiased case ($F = 0$), which is of practical relevance, since it is always possible to compress a biased message to an unbiased one.

For the case $K \to \infty$, $C = \alpha K$ we can obtain solutions to the saddle-point equations at arbitrary temperatures. The first saddle-point Eq. (67) can be approximated by:

$$x = \sum_{l=1}^{C-1} y_l \approx (C-1) \langle y \rangle_{\hat{\pi}} = (C-1) \int dy \, y \, \hat{\pi}(y). \tag{71}$$

If $\langle y \rangle_{\hat{\pi}} = 0$ (paramagnetic phase) then $\pi(x)$ must be concentrated at $x = 0$ implying that $\pi(x) = \delta(x)$ and $\hat{\pi}(y) = \delta(y)$ are the only possible solutions. Equation (71) also implies that $x \approx \mathcal{O}(K)$ in the ferromagnetic phase.

Using Eq. (71) and the second saddle-point Eq. (67) we find a self-consistent equation for the mean field $\langle y \rangle_{\hat{\pi}}$:

$$\langle y \rangle_{\hat{\pi}} = \left\langle \frac{1}{\beta} \tanh[\tanh(\beta J) \, [\tanh(\beta (C-1) \langle y \rangle_{\hat{\pi}})]^{K-1}] \right\rangle_J. \tag{72}$$

For the BSC we average over the distribution (58). Computing the average, using $C = \alpha K$ and rescaling the temperature $\beta = \tilde{\beta}(\ln K)/K$, we obtain in the limit $K \to \infty$:

$$\langle y \rangle_{\hat{\pi}} \approx (1-2p) \left[\tanh(\tilde{\beta} \alpha \langle y \rangle_{\hat{\pi}} \ln(K)) \right]^K, \tag{73}$$

where p is the channel flip probability. The mean field $\langle y \rangle_{\hat{\pi}} = 0$ is always a solution to this equation (paramagnetic solution); at $\beta_c = \ln(K)/(2\alpha K(1-2p))$ an extra nontrivial ferromagnetic solution emerges with $\langle y \rangle_{\hat{\pi}} = 1 - 2p$. The connection with the overlap ρ is given by Eqs. (68) and Eq. (69) implying that $\rho = 1$ for the ferromagnetic solution. It is remarkable that the temperature where the ferromagnetic solution emerges is $\beta_c \sim \mathcal{O}(\ln(K)/K)$. Paramagnetic–ferromagnetic barriers emerge at reasonably high temperatures, in a simulated annealing process, implying metastability and, consequently, a very slow convergence. It seems to advocate the use of small K values in practical applications. For $\beta > \beta_c$ both paramagnetic and ferromagnetic solutions exist.

The ferromagnetic free-energy can be obtained from Eq. (66) using Eq. (71), resulting in $f_{\text{FERRO}} = -\alpha(1-2p)$. The corresponding entropy is $s_{\text{FERRO}} = 0$. The paramagnetic free-energy is obtained by plugging $\pi(x) = \delta(x)$ and $\hat{\pi}(y) = \delta(y)$ into Eq. (66):

$$f_{\text{PARA}} = -\frac{1}{\beta}(\alpha \ln(\cosh \beta) + \ln 2), \tag{74}$$

$$s_{\text{PARA}} = \alpha(\ln(\cosh \beta) - \beta \tanh \beta) + \ln 2. \tag{75}$$

Paramagnetic solutions are unphysical for $\alpha > (\ln 2)/[\beta \tanh \beta - \ln(\cosh \beta)]$, since the corresponding entropy is negative. To complete the picture of the phase diagram we have to introduce a replica symmetry breaking scenario that yields sensible physics.

In general, to construct a symmetry breaking solution in finite connectivity systems (see Monasson, 1998b; Franz *et al.,* 2001) is a difficult task. We choose as a first approach a one-step replica symmetry breaking scheme, known as the *frozen spins solution,* that yields exact results for the REM (Gross and Mezard, 1984; Parisi, 1980).

We assume that ergodicity breaks in such a way that the space of configurations is divided in n/m islands. Inside each of these islands there are m identical configurations, implying that the system can freeze in any of n/m microstates. Therefore, in the space of replicas we have the following situation:

$$\frac{1}{N}\sum_{j=1}^{N} S_j^{\alpha} S_j^{\beta} = 1, \text{ if } \alpha \text{ and } \beta \text{ are in the same island}$$

$$\frac{1}{N}\sum_{j=1}^{N} S_j^{\alpha} S_j^{\beta} = q, \text{ otherwise.} \tag{76}$$

By assuming the above structure the replicated partition function has the form:

$$\begin{aligned}\left\langle \mathcal{Z}^n_{\text{RSB}} \right\rangle_{\mathcal{A},\xi,J} &= \left\langle \text{Tr}_{\{S_j^\alpha\}} \exp\left(-\beta \sum_{\alpha=1}^{n} \mathcal{H}(\boldsymbol{S}^\alpha)\right)\right\rangle_{\mathcal{A},J,\xi} \\ &= \left\langle \text{Tr}_{\left\{S_j^1,\ldots,S_j^{n/m}\right\}} \exp\left(-\beta m \sum_{\alpha=1}^{n/m} \mathcal{H}(\boldsymbol{S}^\alpha)\right)\right\rangle_{\mathcal{A},J,\xi} \\ &= \left\langle \prod_{\alpha}^{n/m} \text{Tr}_{\{S_j^\alpha\}} \exp\left(-\beta m\, \mathcal{H}(\boldsymbol{S}^\alpha)\right)\right\rangle_{\mathcal{A},J,\xi} \\ &= \left\langle \mathcal{Z}^{n/m}_{\text{RS}} \right\rangle_{\mathcal{A},\xi,J}, \end{aligned} \tag{77}$$

where in the first line we have used the ansätz with n/m islands with m identical configurations in each and in the last step we have used that the overlap between any two different islands is q. From (77) we have:

$$\begin{aligned}\langle \ln \mathcal{Z}_{\text{RSB}}(\beta)\rangle_{\mathcal{A},\xi,J} &= \left.\frac{\partial}{\partial n}\right|_{n=0} \left\langle \mathcal{Z}^n_{\text{RSB}}(\beta)\right\rangle_{\mathcal{A},\xi,J} \\ &= \frac{1}{m}\langle \ln \mathcal{Z}_{\text{RS}}(\beta m)\rangle_{\mathcal{A},\xi,J}. \end{aligned} \tag{78}$$

The number of configurations per island m must extremize the free-energy, therefore, we have:

$$\frac{\partial}{\partial m}\langle \ln \mathcal{Z}_{\text{RSB}}(\beta)\rangle_{\mathcal{A},\xi,J} = 0, \tag{79}$$

what is equivalent to

$$\begin{aligned} s_{\text{RS}}(\beta_g) &= -\tilde{\beta}^2 \left.\frac{\partial}{\partial \tilde{\beta}}\right|_{\tilde{\beta}=\beta_g} \left[\frac{1}{\tilde{\beta}}\langle \ln \mathcal{Z}_{\text{RS}}(\tilde{\beta})\rangle_{\mathcal{A},\xi,J}\right] \\ &= 0, \end{aligned} \tag{80}$$

where we introduced $\tilde{\beta} = \beta m$. In this way $m = \beta_g/\beta$, with β_g being a root of the replica symmetric paramagnetic entropy (74), satisfying:

$$\alpha(\ln(\cosh \beta_g) - \beta_g \tanh \beta_g) + \ln 2 = 0 \tag{81}$$

The RSB-spin glass free-energy is given by f_{PARA} (74) at temperature β_g:

$$f_{\text{RSB-SG}} = -\frac{1}{\beta_g}(\alpha \ln(\cosh\ \beta_g) + \ln\ 2), \tag{82}$$

consequently the entropy is $s_{\text{RSB-SG}} = 0$. In Figure 8 we show the phase diagram for a given code rate R in the plane of temperature T and noise level p.

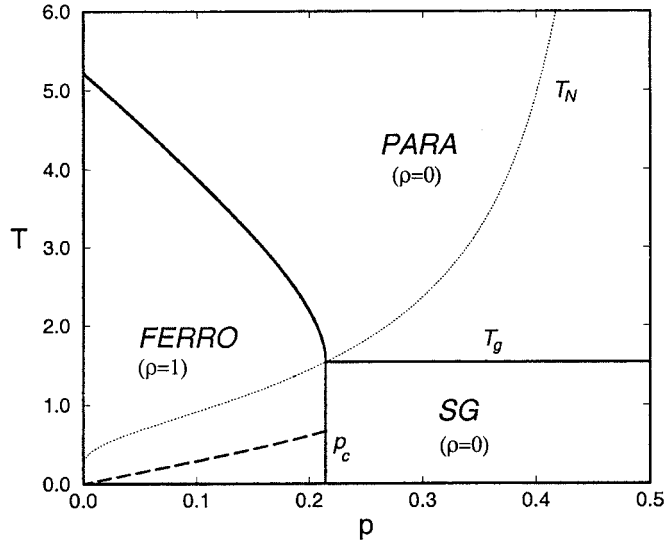

FIGURE 8. Phase diagram in the plane of temperature T versus noise level p for $K \to \infty$ and $C = \alpha K$, with $\alpha = 4$. The dotted line indicates the Nishimori temperature T_N. Full lines represent phase coexistence. The critical noise level is p_c. The necessary condition for stability of the ferromagnetic phase within the replica symmetric space is satisfied above the dashed line.

C. Shannon's Bound

The channel-coding theorem asserts that up to a critical code rate R_c, which equals the channel capacity (*Shannon's bound*), it is possible to recover information with arbitrarily small probability of error. For the BSC:

$$R_c = \frac{1}{\alpha_c} = 1 + p \log_2 p + (1 - p) \log_2 (1 - p). \tag{83}$$

The code of Sourlas, in the case where $K \to \infty$ and $C \sim \mathcal{O}(N^K)$, can be mapped onto the REM and has been shown to saturates the channel capacity in the limit $R \to 0$ (Sourlas, 1989). Shannon's bound can also be attained by Sourlas code at zero temperature for $K \to \infty$ but with connectivity $C = \alpha K$. In this limit the model is analogous to the diluted REM analyzed by Saakian (1998). The errorless phase is manifested in a ferromagnetic phase with total alignment ($\rho = 1$), only attainable for infinite K. Up to a certain critical noise level, a noise level increase produces ergodicity breaking leading to a spin glass phase where the misalignment is maximal ($\rho = 0$). The ferromagnetic–spin glass transition corresponds to the transition from errorless decoding to decoding with errors described by the channel coding theorem. A paramagnetic

phase is also present when the transmitted information is insufficient to recover the original message ($R > 1$).

At zero temperature, saddle-point Eq. (67) can be rewritten as:

$$\pi(x) = \int \left[\prod_{l=1}^{C-1} dy_l \, \hat{\pi}(y_l) \right] \delta \left[x - \sum_{j=1}^{C-1} y_j \right] \tag{84}$$

$$\hat{\pi}(y) = \int \left[\prod_{l=1}^{K-1} dx_l \, \pi(x_l) \right] \times \left\langle \delta \left[y - \text{sign} \left(J \prod_{l=1}^{K-1} x_l \right) \min(|J|, \ldots, |x_{K-1}|) \right] \right\rangle_J , \tag{85}$$

The solutions for these saddle-point equations may result in very structured probability distributions. As an approximation we choose the simplest self-consistent family of solutions which are, since $J = \pm 1$, given by:

$$\hat{\pi}(y) = p_+ \delta(y-1) + p_0 \delta(y) + p_- \delta(y+1) \tag{86}$$

$$\pi(x) = \sum_{l=1-C}^{C-1} T_{[p_\pm, p_0; C-1]}(l) \, \delta(x-l),$$

with

$$T_{[p_+, p_0, p_-; C-1]}(l) = {\sum_{\{k,h,m\}}}' \frac{(C-1)!}{k!\, h!\, m!} \, p_+^k p_0^h p_-^m, \tag{87}$$

where the prime indicates that k,h,m are such that $k - h = l$; $k + h + m = C - 1$. Evidence for this simple ansätz comes from Monte Carlo integration of Eq. (67) at very low temperatures, that shows solutions comprising three dominant peaks and a relatively weak regular part. Plugging this ansätz (86) in the saddle-point equations, we write a closed set of equations in $p_\pm$ and p_0 that can be solved numerically.

Solutions are of three types: ferromagnetic ($p_+ > p_-$), paramagnetic ($p_0 = 1$), and replica symmetric spin glass ($p_- = p_+$). Computing free-energies and entropies enables one to construct the phase diagram. At zero temperature, the paramagnetic free-energy is $f_{\text{PARA}} = -\alpha$ and the entropy is $s_{\text{PARA}} = (1-\alpha)\ln 2$; this phase is physical only for $\alpha < 1$, as is expected since it corresponds exactly to the regime where the transmitted information is insufficient to recover the actual message ($R > 1$).

The ferromagnetic free-energy does not depend on the temperature, having the form $f_{\text{FERRO}} = -\alpha(1-2p)$ with entropy $s_{\text{FERRO}} = 0$. We can find the ferromagnetic–spin glass coexistence line that corresponds to the maximum performance of a Sourlas code by equating Eq. (82) and f_{FERRO}. Observing

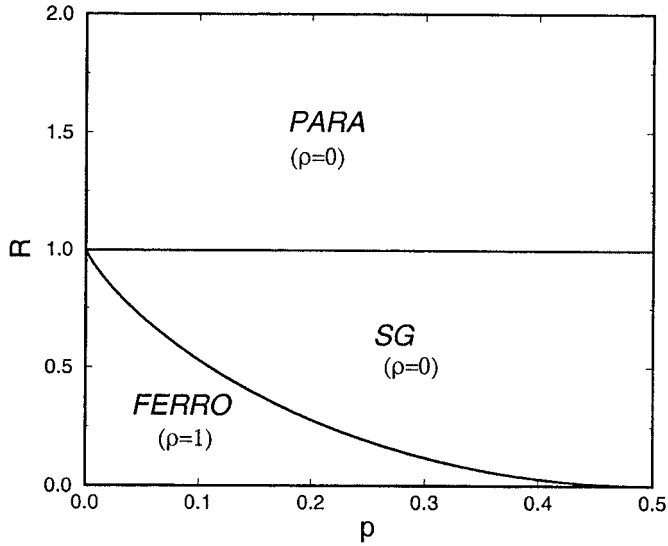

FIGURE 9. Phase diagram in the plane code rate R versus noise level p for $K \to \infty$ and $C = \alpha K$ at zero temperature. The ferromagnetic–spin glass coexistence line corresponds to Shannon's bound.

that $\beta_g = \beta_N(p_c)$ (as seen in Fig. 8) we find that this transition coincides with the channel capacity (83). It is interesting to note that in the large K regime both RS–ferromagnetic and RSB–spin glass free-energies (for $T < T_g$) do not depend on the temperature, it means that Shannon's bound is saturated also for finite temperatures up to T_g. In Figure 9 we represent the complete zero temperature phase diagram.

The bound obtained depends on the stability of the ferromagnetic and paramagnetic solutions within the space of symmetric replicas at zero temperature. Instabilities are found in the ferromagnetic phase for $p > 0$. These instabilities within the replica symmetric space puts in question our result of saturating Shannon's bound, since a correction to the ferromagnetic solution could change the ferromagnetic–spin glass transition line. However, the instability vanishes for high temperatures, which supports the ferromagnetic–spin glass transition line obtained and possible saturation of the bound in some region.

Shannon's bound can only be attained in the limit $K \to \infty$; however, there are some possible drawbacks in using high K values due to large barriers which are expected to occur between the paramagnetic and ferromagnetic phases. We now consider the finite K case, for which we can solve the RS saddle-point Eqs. (67) for arbitrary temperatures using Monte Carlo integration. We can also obtain solutions for the zero temperature case using Eqs. (86) iteratively.

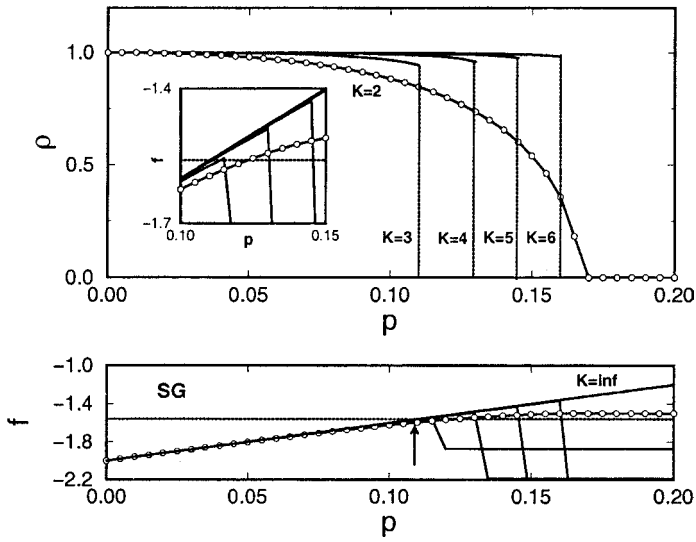

FIGURE 10. Top: zero temperature overlap ρ as a function of the noise level p for various K values at code rate $R = 1/2$, as obtained by the iterative method. Bottom: RS-ferromagnetic free-energies (white circles for $K = 2$ and from the left: $K = 3, 4, 5$, and 6) and RSB-spin glass free-energy (dotted line) as functions of the noise level p. The arrow indicates the region where the RSB–spin glass phase starts to dominate. Inset: a detailed view of the RS–RSB transition region.

It has been shown that $K > 2$ extensively connected models (Gross and Mezard, 1984) exhibit Parisi-type order functions with similar discontinuous structure as found in the $K \to \infty$ case; it was also shown that the one-step RSB frozen spins solution, employed to describe the spin glass phase, is locally stable within the complete replica space and zero field (unbiased messages case) at all temperatures. We, therefore, assume that the ferromagnetic–spin glass transition for $K > 2$ is described by the frozen spins RSB solution.

At the top of Figure 10 we show the zero temperature overlap ρ as a function of the noise level p at code rate $R = 1/2$ obtained by using the three-peaks ansätz. Note that the RSB–spin glass phase dominates for $p > p_c$ (see bottom of Fig. 10). In the bottom figure we plot RS free-energies and RSB frozen spins free-energy, from which we determine the noise level p_c for coexistence of ferromagnetic and spin-glass phases (pointed by an arrow). Above the transition, the system enters in a paramagnetic or RS spin glass phase with free-energies for $K = 3, 4, 5$, and 6 that are lower than the RSB spin glass free-energy; nevertheless, the entropy is negative and these free-energies are therefore unphysical. It is remarkable that the coexistence value does not

change significantly for finite K in comparison to infinite K. Remind that Shannon's bound cannot be attained for finite K, since $\rho \to 1$ ($p_b \to 0$) only if $K \to \infty$.

It is known that the $K = 2$ model with extensive connectivity (SK model) requires a full Parisi solution to recover the concavity of the free-energy (Mezard *et al.,* 1987). No stable solution is known for the intensively connected model (Viana–Bray model). Probability propagation only solves the decoding problem approximately, the approximated solutions are similar to those obtained by supposing replica symmetry. Thus, the theoretical relevance of the RS results for $K = 2$ are to be evaluated by comparison with simulations of probability propagation decoding.

D. Decoding with Probability Propagation

The decoding task consists of evaluating estimates of the form $\hat{\xi}_j = \text{sign}\langle S_j \rangle_{P(S_j|J)}$. The marginal posterior $P(S_j \mid \boldsymbol{J}) = \sum_{S_l, l \neq j} P(\boldsymbol{S} \mid \boldsymbol{J})$ can be, in principle, calculated simply by using Bayes theorem and a proper model for the encoding and corruption processes (namely, coding by a sparse generator matrix with K bit long parity-checks and a memoryless BSC channel) to write:

$$P(S_j \mid \boldsymbol{J}) = \frac{1}{P(\boldsymbol{J})} \sum_{S_l, l \neq j} \prod_{\mu} P(J_\mu \mid S_{i_1} \cdots S_{i_K}) \prod_{i=1}^{N} P(S_i), \tag{88}$$

where $P(\boldsymbol{J})$ is a normalization dependent on $\boldsymbol{J}$ only. A brute force evaluation of the above marginal on a space of binary vectors $\boldsymbol{S} \in \{\pm 1\}^N$ with M checks would take $(M + N + 1)2^N$ operations, what becomes infeasible very quickly. To illustrate how dramatically the computational requirements increase, assume a code of rate $R = 1/2$, if $N = 10$ the number of operations required is 31,744, if one increases the message size to $N = 1000$, 3×10^{304} operations are required! Monte Carlo sampling is an alternative to brute force evaluation; it consists of generating a number (much less than 2^N) of typical vectors $\boldsymbol{S}$. By using this to estimate the marginal posterior, however, the sample size required can prove to be equally prohibitive.

As a solution to these resource problems, we can explore the structure of (88) to devise an algorithm that produces an approximation to $P(S_j \mid \boldsymbol{J})$ in $\mathcal{O}(N)$ operations. We start by concentrating on one particular site S_j; this site interacts directly with a number of other sites through C couplings denoted by $J_{\langle i_1 \cdots i_K \rangle}$ and $\{J_\mu\} = J_{\mu(1)}, \ldots, J_{\mu(C-1)}$. Suppose now that we isolate only the interaction via coupling $J_{\langle i_1 \cdots i_K \rangle}$, if the bipartite Bayesian network representing

the dependencies in the problem is a tree, it is possible to write:

$$P\left(S_j \mid J_{\langle i_1 \cdots i_K \rangle}\right) = \frac{P(S_j)}{P\left(J_{\langle i_1 \cdots i_K \rangle}\right)} \sum_{\{S_{i_1} \cdots S_{i_{K-1}}\}} P\left(J_{\langle i_1 \cdots i_K \rangle} \middle| S_j, S_{i_1} \cdots S_{i_{K-1}}\right) \times \prod_{l=1}^{K-1} P(S_{i_l} \mid \{J_\mu : \mu \in \mathcal{M}(i_l)\}). \tag{89}$$

Terms like $P(S_{i_l} \mid \{J_\mu\})$ can be interpreted simply as updated priors for S_{i_l}. In a tree, these terms factorize like $P(S_{i_l} \mid \{J_\mu\}) = \prod_{j=1}^{C-1} P(S_{i_l} \mid J_{\mu(j)})$ and a recursive relation can be obtained, introducing:

$$Q_{\nu j}^x = P(S_j = x \mid \{J_\mu : \mu \in \mathcal{M}(j) \setminus \nu\}) \tag{90}$$

and

$$R_{\nu j}^x = \sum_{\{S_i : i \in \mathcal{L}(\nu) \setminus j\}} P(J_\nu \mid S_j, \{S_i : i \in \mathcal{L}(\nu) \setminus j\}) \prod_{i \in \mathcal{L}(\nu) \setminus j} Q_{\nu i}^{S_i}, \tag{91}$$

where $\mathcal{M}(j)$ is the set of couplings linked to site j and $\mathcal{L}(\nu)$ is the set of sites linked to coupling ν.

Equation (89) can be rewritten as:

$$Q_{\mu j}^x = a_{\mu j} P(S_j = x) \prod_{\nu \in \mathcal{M}(j) \setminus \mu} R_{\nu j}^x. \tag{92}$$

Equations (91) and (92) can be solved iteratively, requiring $(2^K KC + 2C^2)NT$ operations with T being the (order 1) number of steps needed for convergence. These computational requirements may be further reduced by using Markov chain Monte Carlo methods (MacKay, 1999).

An approximation to the marginal posterior (88) is obtained by counting the influence of all C interactions over each site j and using the assumed factorization property to write:

$$Q_j^x = a_j P(S_j = x) \prod_{\nu \in \mathcal{M}(j)} R_{\nu j}^x. \tag{93}$$

This is an approximation in the sense that the recursion obtained from (89) is only guaranteed to converge to the correct posterior if the system has a tree structure, i.e., every coupling appears only once as one goes backwards in the recursive chain.

By taking advantage of the normalization conditions for the distributions $Q_{\mu j}^{+1} + Q_{\mu j}^{-1} = 1$ and $R_{\mu j}^{+1} + R_{\mu j}^{-1} = 1$, one can change variables and reduce the number of equations by a factor of two $m_{\mu j} = Q_{\mu j}^{+1} - Q_{\mu j}^{-1}$ and $\hat{m}_{\mu j} = R_{\mu j}^{+1} - R_{\mu j}^{-1}$.

The analogy with statistical physics can be exposed by first observing that:

$$P(J_\mu \mid S_j, \{S_i : i \in \mathcal{L}(\mu) \setminus j\}) \sim \exp\left(- \beta J_\mu \prod_{i \in \mathcal{L}(\mu)} S_i \right). \tag{94}$$

That can be also written in the more convenient form:

$$P(J_\mu \mid S_j, \{S_i : i \in \mathcal{L}(\mu) \setminus j\}) \sim \frac{1}{2} \cosh(\beta J_\mu) \left(1 + \tanh(\beta J_\mu) \prod_{j \in \mathcal{L}(\mu)} S_j \right). \tag{95}$$

Plugging Eq. (95) for the likelihood in Eqs. (92), using the fact that the prior probability is given by $P(S_j) = \frac{1}{2}(1 + \tanh(\beta'_N S_j))$ and computing $m_{\mu j}$ and $\hat{m}_{\mu j}$ (see Appendix A.6) one obtains:

$$\begin{aligned} \hat{m}_{\mu j} &= \tanh(\beta J_\mu) \prod_{l \in \mathcal{L}(\mu) \setminus j} m_{\mu l} \\ m_{\mu j} &= \tanh\left(\sum_{\nu \in \mathcal{M}(l) \setminus \mu} \operatorname{atanh}(\hat{m}_{\nu j}) + \beta'_N \right). \end{aligned} \tag{96}$$

The pseudo-posterior can then be calculated:

$$m_j = \tanh\left(\sum_{\nu \in \mathcal{M}(l)} \operatorname{atanh}(\hat{m}_{\nu j}) + \beta'_N \right), \tag{97}$$

providing Bayes optimal decoding $\hat{\xi}_j = \operatorname{sign}(m_j)$.

Equations (96) depend on the received message $\boldsymbol{J}$. In order to make the analysis message independent, we can use a gauge transformation $\hat{m}_{\mu j} \mapsto \xi_j \hat{m}_{\mu j}$ and $m_{\mu j} \mapsto \xi_j m_{\mu j}$ to write:

$$\begin{aligned} \hat{m}_{\mu j} &= \tanh(\beta J) \prod_{l \in \mathcal{L}(\mu) \setminus j} m_{\mu l} \\ m_{\mu j} &= \tanh\left(\sum_{\nu \in \mathcal{M}(l) \setminus \mu} \tanh^{-1}(\hat{m}_{\nu j}) + \beta'_N \xi_j \right). \end{aligned} \tag{98}$$

In the new variables, a decoding success corresponds to $\hat{m}_{\mu j} > 0$ and $m_{\mu j} = 1$ for all μ and j. By transforming these variables as $\hat{m} = \tanh(\beta y)$ and $m = \tanh(\beta x)$ and considering the actual message and noise as quenched disorder,

Eqs. (98) can be rewritten as:

$$y = \frac{1}{\beta}\left\langle \tanh^{-1}\left(\tanh(\beta J)\prod_{j=1}^{K-1}\tanh(\beta x_j)\right)\right\rangle_J$$
$$x = \left\langle \sum_{j=1}^{C-1} y_j + \xi F \right\rangle_\xi . \tag{99}$$

For a large number of iterations, one can expect the ensemble of probability networks to converge to an equilibrium distribution where $\hat{m}$ and m are random variables sampled from distributions $\hat{\phi}(y)$ and $\phi(x)$, respectively. The above relations lead to a dynamics of the distributions $\hat{\phi}(y)$ and $\phi\ (x)$, that is exactly as the one obtained when solving iteratively RS saddle-point Eqs. (67). The probability distributions $\hat{\phi}(y)$ and $\phi\ (x)$ can be, therefore, identified with $\hat{\pi}(y)$ and $\pi\ (x)$, respectively, and the RS solutions correspond to decoding a generic message using probability propagation averaged over an ensemble of different codes, noise, and signals.

Equations (96) are now used to show the agreement between the simulated decoding and analytical calculations. For each run, a fixed code is used to generate 20000-bit codewords from 10000-bit messages; corrupted versions of the codewords are then decoded using (96). Numerical solutions for 10 individual runs are presented in Figures 11 and 12, initial conditions are chosen as $\hat{m}_{\mu l} = 0$ and $m_{\mu l} = \tanh(\beta'_N)$ reflecting the prior beliefs. In Figure 11 we show results for $K = 2$ and $C = 4$ in the unbiased case, at code rate $R = 1/2$ (prior probability $P(S_j = +1) = p_\xi = 0.5$) and low temperature $T = 0.26$ (we avoided $T = 0$ due to numerical difficulties). Solving the saddle-point Eqs. (67) numerically using Monte Carlo integration methods we obtain solutions with good agreement to simulated decoding. In the same figure we show the performance for the case of biased messages ($P(S_j = +1) = p_\xi = 0.1$), at code rate $R = 1/4$. Also here the agreement with Monte Carlo integrations is satisfactory. The third curve in Figure 11 shows the performance for biased messages at the Nishimori temperature T_N, as expected, it is far superior compared to low temperature performance and the agreement with Monte Carlo results is even better.

In Figure 12 we show the results obtained for $K = 5$ and $C = 10$. For unbiased messages the system is extremely sensitive to the choice of initial conditions and does not perform well on average even at the Nishimori temperature. For biased messages ($p_\xi = 0.1$, $R = 1/4$) results are far better and in agreement with Monte Carlo integration of the RS saddle-point equations.

The experiments show that probability propagation methods may be used successfully for decoding Sourlas-type codes in practice, and provide solutions that are consistent with the RS analytical solutions.

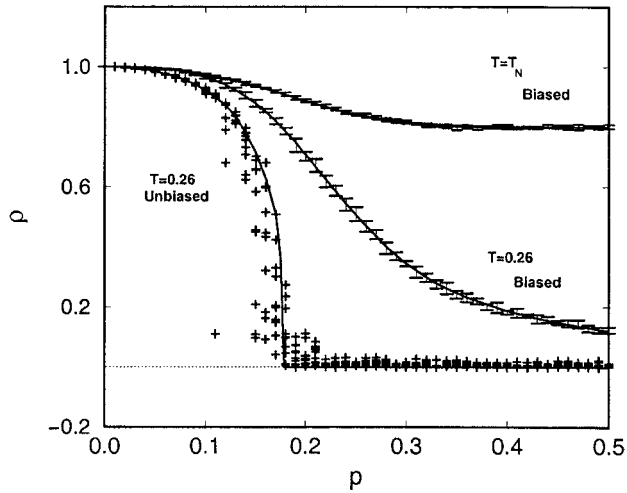

FIGURE 11. Overlap as a function of the flip probability p for decoding using TAP equations for $K = 2$. From the bottom: Monte Carlo solution of the RS saddle-point equations for unbiased message ($p_\xi = 0.5$ at $T = 0.26$ (line) and 10 independent runs of TAP decoding for each flip probability (plus signs), $T = 0.26$ and biased messages ($p_\xi = 0.5$) at the Nishimori temperature T_N.

IV. GALLAGER CODES

In 1962, Gallager proposed a coding scheme which involves sparse linear transformations of binary messages in the decoding stage, while encoding uses a dense matrix. His proposal was overshadowed by convolutional codes due to computational limitations. The best computer available to Gallager in 1962 was an IBM 7090 costing $3 million and with disk capacity of 1 megabyte, while convolutional codes, in comparison, only demanded a simple system of shift registers to process one byte at a time.

Gallager codes have been rediscovered recently by MacKay and Neal (1995) who proposed a closely related code, to be discussed in Section V. This almost coincided with the breakthrough discovery of high performance Turbo codes (Berrou *et al.,* 1993). Variations of Gallager codes have displayed performance comparable (sometimes superior) to Turbo codes (Davey, 1998, 1999), qualifying them as state-of-the-art codes.

A Gallager code is defined by a binary matrix $\boldsymbol{A} = [\boldsymbol{C_1} \mid \boldsymbol{C_2}]$, concatenating two very sparse matrices known to both sender and receiver, with $\boldsymbol{C_2}$ (of dimensionality $(M - N) \times (M - N)$) being invertible and $\boldsymbol{C_1}$ of dimensionality $(M - N) \times N$. A non-systematic Gallager code is defined by a random

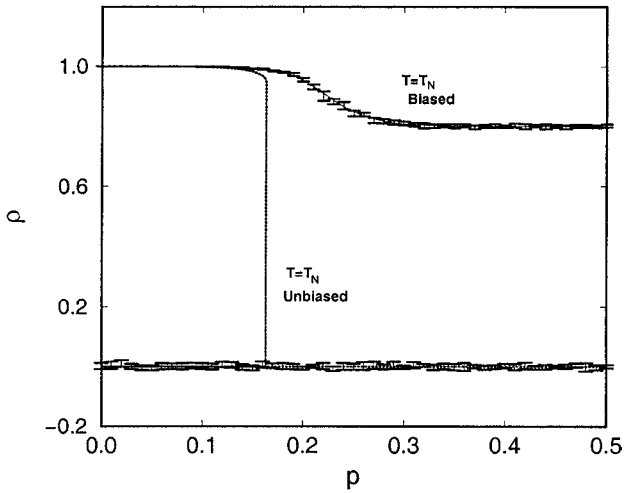

FIGURE 12. Overlap as a function of the flip probability p for decoding using TAP equations for $K = 5$. The dotted line is the replica symmetric saddle-point equations Monte Carlo integration for unbiased messages ($p_\xi = 0.5$) at the Nishimori temperature T_N. The bottom error bars correspond to 10 simulations using the TAP decoding. The decoding performs badly on average in this scenario. The upper curves are for biased messages ($p_\xi = 0.1$) at the Nishimori temperature T_N. The simulations agree with results obtained using the replica symmetric ansätz and Monte Carlo integration.

matrix $\boldsymbol{A}$ of dimensionality $(M - N) \times M$. This matrix can, in general, be organized in a systematic form by eliminating a number $\epsilon \sim \mathcal{O}(1)$ of rows and columns.

Encoding refers to the generation of an M dimensional binary vector $\boldsymbol{t} \in \{0, 1\}^M$ $(M > N)$ from the original message $\boldsymbol{\xi} \in \{0, 1\}^N$ by

$$\boldsymbol{t} = \boldsymbol{G}^T \boldsymbol{\xi} \ (\text{mod } 2), \tag{100}$$

where all operations are performed in the field $\{0, 1\}$ and are indicated by (mod 2). The generator matrix is

$$\boldsymbol{G} = [\boldsymbol{I} \mid \boldsymbol{C}_2^{-1} \boldsymbol{C}_1] \ (\text{mod } 2), \tag{101}$$

where $\boldsymbol{I}$ is the $N \times N$ identity matrix, implying that $\boldsymbol{A}\boldsymbol{G}^T \ (\text{mod } 2) = 0$ and that the first N bits of $\boldsymbol{t}$ are set to the message $\boldsymbol{\xi}$. Note that the generator matrix is dense and each transmitted parity-check carries information about an $\mathcal{O}(N)$ number of message bits. In *regular* Gallager codes the number of nonzero elements in each row of $\boldsymbol{A}$ is chosen to be exactly K . The

number of elements per column is then $C = (1 - R)K$, where the code rate is $R = N/M$ (for unbiased messages). The encoded vector $\boldsymbol{t}$ is then corrupted by noise represented by the vector $\boldsymbol{\zeta} \in \{0, 1\}^M$ with components independently drawn from $P(\zeta) = (1 - p)\delta(\zeta) + p\delta(\zeta - 1)$. The received vector takes the form

$$\boldsymbol{r} = \boldsymbol{G}^T \boldsymbol{\xi} + \boldsymbol{\zeta} \text{ (mod 2)}. \tag{102}$$

Decoding is carried out by multiplying the received message by the matrix $\boldsymbol{A}$ to produce the *syndrome* vector

$$\boldsymbol{z} = \boldsymbol{A}\boldsymbol{r} = \boldsymbol{A}\boldsymbol{\zeta} \text{ (mod 2)}, \tag{103}$$

from which an estimate $\hat{\boldsymbol{\tau}}$ for the noise vector can be produced. An estimate for the original message is then obtained as the first N bits of $\boldsymbol{r} + \hat{\boldsymbol{\tau}}$ (mod 2). The Bayes optimal estimator (also known as *marginal posterior maximizer,* MPM) for the noise is defined as $\hat{\tau}_j = \text{argmax}_{\tau_j} P(\tau_j \mid \boldsymbol{z})$. The performance of this estimator can be measured by the bit error probability $p_b = 1 - 1/M \sum_{j=1}^{M} \delta[\hat{\tau}_j; \zeta_j]$, where $\delta[;]$ is the Kronecker delta. Knowing the matrices $\boldsymbol{C_2}$ and $\boldsymbol{C_1}$, the syndrome vector $\boldsymbol{z}$ and the noise level p, it is possible to apply Bayes theorem and compute the posterior probability

$$P(\boldsymbol{\tau} \mid \boldsymbol{z}) = \frac{1}{Z} \chi\left[\boldsymbol{z} = \boldsymbol{A}\boldsymbol{\tau} \text{ (mod 2)}\right] P(\boldsymbol{\tau}), \tag{104}$$

where $\chi[X]$ is an indicator function providing 1 if X is true and 0 otherwise. To compute the MPM one has to compute the marginal posterior $P(\tau_j \mid \boldsymbol{z}) = \sum_{i \neq j} P(\boldsymbol{\tau} \mid \boldsymbol{z})$, which in general requires $\mathcal{O}(2^M)$ operations, thus becoming impractical for long messages. To solve this problem we can take advantage of the sparseness of $\boldsymbol{A}$ and use probability propagation for decoding, requiring $\mathcal{O}(M)$ operations to perform the same task.

A. Upper Bound on Achievable Rates

It was pointed by MacKay in 1999 that an upper bound for rates achievable for Gallager codes can be found from information theoretic arguments. This upper bound is based on the fact that each bit of the syndrome vector $\boldsymbol{z} = \boldsymbol{A}\boldsymbol{\zeta}$(mod 2) is a sum of K noise bits independently drawn from a bimodal delta distribution $P(\zeta)$ with $P(\zeta = 0) = 1 - p$. The probability of $z_j = 1$ is $p_z^1(K) = \frac{1}{2} - \frac{1}{2}(1 - 2p)^K$ (see Appendix C.1 for details). Therefore, the maximal information content in the syndrome vector is $(M - N)H_2(p_z^1(K))$ (in *bits* or *shannons*), where $H_2(x)$ is the binary entropy. In the decoding process one has to extract information from the syndrome vector in order to reconstruct

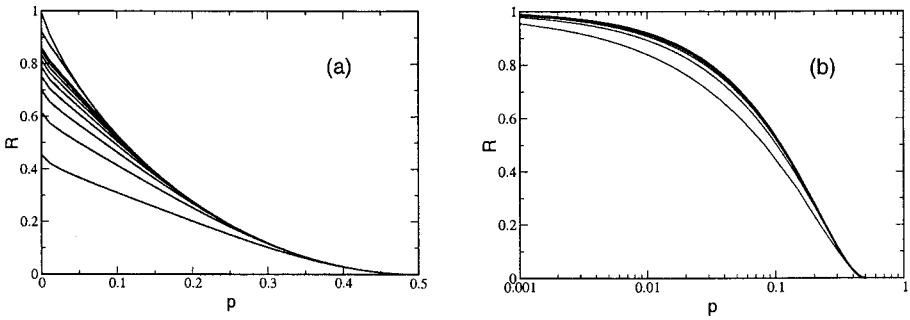

FIGURE 13. (a) Bounds for the rate R as a function of the noise level p for several values of K. From bottom to top: $K = 2$ to 10,20 and Shannon limit. (b) Bounds for several values of C. From bottom to top $C = 2, 3, 4, 5$ and Shannon limit.

a noise vector $\boldsymbol{\zeta}$ which has an information content of $MH_2(p)$. It clearly means that a necessary condition for successful decoding is:

$$(M - N)H_2\left(p_z^1(K)\right) \geq MH_2(p)$$

$$(1 - R)H_2\left(p_z^1(K)\right) \geq H_2(p)$$

$$R \leq 1 - \frac{H_2(p)}{H_2\left(p_z^1(K)\right)}. \tag{105}$$

In Figure 13a we plot this bound by fixing K and finding the minimum value for C such that $R = 1 - C/K$ verifies (105). Observe that as $K \to \infty$, $p_z^1(K) \to 1/2$ and $R \to 1 - H_2(p)$ that corresponds to Shannon's bound.

In Figure 13b we plot the bound by fixing C and finding the maximum K such that $R = 1 - C/K$ satisfies (105), recovering the curves presented in MacKay (1999). Note that $K \to \infty$ implies $C \to \infty$ and vice versa. Gallager codes only can attain Shannon's bound asymptotically in the limit of large K or, equivalently, large C.

B. Statistical Physics Formulation

The connection to statistical physics is made by replacing the field $\{0, 1\}$ by Ising spins $\{\pm 1\}$ and mod 2 sums by products (Sourlas, 1989). The syndrome vector acquires the form of a multispin coupling $\mathcal{J}_\mu = \prod_{j \in \mathcal{L}(\mu)} \zeta_j$ where $j = 1, \ldots, M$ and $\mu = 1, \ldots, (M - N)$. The K indices of nonzero elements in the row μ of $\boldsymbol{A}$ are given by $\mathcal{L}(\mu) = \{j_1, \ldots, j_K\}$, and in a column l are given by $\mathcal{M}(l) = \{\mu_1, \ldots, \mu_C\}$.

The following family of posterior probabilities can be introduced:

$$P_\gamma(\boldsymbol{\tau} \mid \mathcal{J}) = \frac{1}{Z} \exp[-\beta \mathcal{H}_\gamma(\boldsymbol{\tau}; \mathcal{J})] \tag{106}$$

$$\mathcal{H}_\gamma(\boldsymbol{\tau}; \mathcal{J}) = -\gamma \sum_{\mu=1}^{M-N} \left(\mathcal{J}_\mu \prod_{j \in \mathcal{L}(\mu)} \tau_j - 1 \right) - F \sum_{j=1}^{M} \tau_j .$$

The Hamiltonian depends on hyperparameters γ and F. For optimal decoding, γ and F have to be set to specific values that best represent how the encoding process and corruption were performed (*Nishimori condition* (Iba, 1999)). Therefore, γ must be taken to infinity to reflect the hard constraints in Eq. (104) and $F = \operatorname{atanh}(1 - 2p)$, reflecting the channel noise level p. The temperature β must simultaneously be chosen to be the Nishimori temperature $\beta_N = 1$, that will keep the hyperparameters in the correct value.

The disorder in (106) is trivial and can be gauged to $\mathcal{J}_\mu \mapsto 1$ by using $\tau_j \mapsto \tau_j \zeta_j$. The resulting Hamiltonian is a multispin ferromagnet with finite connectivity in a random field $\zeta_j F$:

$$\mathcal{H}_\gamma^{\text{gauge}}(\boldsymbol{\tau}; \boldsymbol{\zeta}) = -\gamma \sum_{\mu=1}^{M-N} \left(\prod_{j \in \mathcal{L}(\mu)} \tau_j - 1 \right) - F \sum_{j=1}^{M} \zeta_j \tau_j . \tag{107}$$

At the Nishimori condition $\gamma \to \infty$ and the model is even simpler, corresponding to a paramagnet with restricted configuration space on a nonuniform external field:

$$\mathcal{H}^{\text{gauge}}(\boldsymbol{\tau} \in \Omega; \boldsymbol{\zeta}) = -F \sum_{j=1}^{M} \zeta_j \tau_j , \tag{108}$$

where

$$\Omega = \left\{ \boldsymbol{\tau} : \prod_{j \in \mathcal{L}(\mu)} \tau_j = 1, \mu = 1, \ldots, M - N \right\}. \tag{109}$$

The optimal decoding process simply corresponds to finding local magnetizations at the Nishimori temperature $m_j = \langle \tau_j \rangle_{\beta_N}$ and calculating Bayesian estimates as $\hat{\tau}_j = \operatorname{sgn}(m_j)$.

In the $\{\pm 1\}$ representation the probability of bit error, acquires the form

$$p_b = \frac{1}{2} - \frac{1}{2M} \sum_{j=1}^{M} \zeta_j \operatorname{sgn}(m_j), \tag{110}$$

connecting the code performance with the computation of local magnetizations.

C. Replica Theory

In this section we use the replica theory for analyzing the typical performance of Gallager codes along the same lines discussed for Sourlas codes. We start by rewriting the gauged Hamiltonian (107) in a form more suitable for computing averages over different codes:

$$\mathcal{H}_{\gamma}^{\text{gauge}}(\boldsymbol{\tau};\boldsymbol{\zeta}) = -\gamma \sum_{\langle i_1 \cdots i_K \rangle} \mathcal{A}_{\langle i_1 \cdots i_K \rangle}(\tau_{i_1} \cdots \tau_{i_K} - 1) - F \sum_{j=1}^{M} \zeta_j \tau_j, \tag{111}$$

where $\mathcal{A}_{\langle i_1 \cdots i_K \rangle} \in \{0, 1\}$ is a random symmetric tensor with the properties:

$$\sum_{\langle i_1 \cdots i_K \rangle} \mathcal{A}_{\langle i_1 \cdots i_K \rangle} = M - N \qquad \sum_{\langle i_1, \ldots, i_j = l, \ldots, i_K \rangle} \mathcal{A}_{\langle i_1, \ldots, i_K \rangle} = C \forall l, \tag{112}$$

that selects $M - N$ sets of indices (*construction*). The construction $\{\mathcal{A}_{\langle i_1 \cdots i_K \rangle}\}$ and the noise vector $\boldsymbol{\zeta}$ are to be regarded as quenched disorder. As usual, the aim is to compute the free-energy:

$$f = -\frac{1}{\beta} \lim_{M \to \infty} \frac{1}{M} \langle \ln \mathcal{Z} \rangle_{\mathcal{A},\zeta}, \tag{113}$$

from which the typical macroscopic (thermodynamic) behavior can be obtained. The partition function $\mathcal{Z}$ is:

$$\mathcal{Z} = \text{Tr}_{\tau} \exp\left(-\beta \mathcal{H}_{\gamma}^{\text{gauge}}(\boldsymbol{\tau};\boldsymbol{\zeta})\right). \tag{114}$$

The free-energy can be evaluated calculating following expression

$$f = -\frac{1}{\beta} \lim_{M \to \infty} \frac{1}{M} \left. \frac{\partial}{\partial n} \right|_{n=0} \langle \mathcal{Z}^n \rangle_{\mathcal{A},\zeta}, \tag{115}$$

where

$$\langle \mathcal{Z}^n \rangle_{\mathcal{A},\zeta} = \sum_{\boldsymbol{\tau}^1, \ldots, \boldsymbol{\tau}^n} \prod_{j=1}^{M} \left\langle \exp\left(F \zeta \beta \sum_{\alpha=1}^{n} \tau_j^{\alpha} \right) \right\rangle_{\zeta} \times \left\langle \prod_{\langle i_1 \cdots i_K \rangle} \prod_{\alpha=1}^{n} \exp\left[\beta \gamma \mathcal{A}_{\langle i_1 \cdots i_K \rangle} \left(\tau_{i_1}^{\alpha} \cdots \tau_{i_K}^{\alpha} - 1 \right) \right] \right\rangle_{\mathcal{A}}. \tag{116}$$

The average over constructions $\langle(\cdots)\rangle_{\mathcal{A}}$ takes the form:

$$\begin{aligned}\langle(\cdots)\rangle_{\mathcal{A}} &= \frac{1}{\mathcal{N}}\sum_{\{\mathcal{A}\}}\prod_{j=1}^{M}\delta\left(\sum_{\langle i_1=j,i_2,\ldots,i_K\rangle}\mathcal{A}_{\langle i_1=j,\ldots,i_K\rangle}-C\right)(\cdots)\\ &= \frac{1}{\mathcal{N}}\sum_{\{\mathcal{A}\}}\prod_{j=1}^{M}\left[\oint\frac{dZ_j}{2\pi i}\frac{1}{Z_j^{C+1}}Z_j^{\sum_{\langle i_1=j,i_2,\ldots,i_K\rangle}\mathcal{A}_{\langle i_1=j,\ldots,i_K\rangle}}\right](\cdots),\end{aligned} \tag{117}$$

and the average $\langle(\cdots)\rangle_{\zeta}$ over the noise is:

$$\langle(\cdots)\rangle_{\zeta} = \sum_{\zeta=-1,+1}(1-p)\delta(\zeta-1)+p\delta(\zeta+1)(\cdots). \tag{118}$$

By computing the averages above and introducing auxiliary variables through the identity

$$\int dq_{\alpha_1\cdots\alpha_m}\delta\left(q_{\alpha_1\cdots\alpha_m}-\frac{1}{M}\sum_{i}^{M}Z_i\tau_i^{\alpha_1}\cdots\tau_i^{\alpha_m}\right)=1 \tag{119}$$

one finds, after using standard techniques (see Appendix B.1 for details), the following expression for the replicated partition function:

$$\begin{aligned}\langle\mathcal{Z}^n\rangle_{\mathcal{A},\zeta} = \frac{1}{\mathcal{N}}\int&\left(\frac{dq_0\,d\hat{q}_0}{2\pi i}\right)\left(\prod_{\alpha=1}^{n}\frac{dq_\alpha\,d\hat{q}_\alpha}{2\pi i}\right)\cdots\\ &\times\exp\left[\frac{M^K}{K!}\sum_{m=0}^{n}\sum_{\langle\alpha_1\cdots\alpha_m\rangle}\mathcal{T}_m q^K_{\alpha_1\cdots\alpha_m}\right.\\ &\left.-M\sum_{m=0}^{n}\sum_{\langle\alpha_1\cdots\alpha_m\rangle}q_{\alpha_1\cdots\alpha_m}\hat{q}_{\alpha_1\cdots\alpha_m}\right]\\ &\times\prod_{j=1}^{M}\mathrm{Tr}_{\{\tau^\alpha\}}\left[\left\langle\exp\left[F\beta\zeta\sum_{\alpha=1}^{n}\tau^\alpha\right]\right\rangle_{\zeta}\right.\\ &\left.\times\oint\frac{dZ}{2\pi i}\frac{\exp\left[Z\sum_{m=0}^{n}\sum_{\langle\alpha_1\cdots\alpha_m\rangle}\hat{q}_{\alpha_1\cdots\alpha_m}\tau^{\alpha_1}\cdots\tau^{\alpha_m}\right]}{Z^{C+1}}\right],\end{aligned} \tag{120}$$

where $\mathcal{T}_m = \mathrm{e}^{-n\beta\gamma}\cosh^n(\beta\gamma)\tanh^m(\beta\gamma)$. Comparing this expression with that obtained for the code of Sourlas in Eq. (A.7), one can see that the differences are

the dimensionality M for Gallager codes instead of N for Sourlas (reflecting the fact that in the former the noise vector of dimension M is the dynamic variable) and the absence of disorder in the couplings, yielding a slightly modified definition for the constants $\mathcal{T}_m$.

D. Replica Symmetric Solution

The replica symmetric ansätz consists of assuming the following form for the order parameters:

$$q_{\alpha_1\cdots\alpha_m} = \int dx\ \pi(x)\ x^m \qquad \hat{q}_{\alpha_1\cdots\alpha_m} = \int d\hat{x}\ \hat{\pi}(\hat{x})\,\hat{x}^{\,m}. \tag{121}$$

By performing the limit $\gamma \to \infty$, plugging (121) into (120), computing the normalization constant $\mathcal{N}$, integrating in the complex variable Z and computing the trace (see Appendix B.2) we find:

$$\begin{aligned}\langle \mathcal{Z}^n \rangle_{\mathcal{A},\zeta} = \mathrm{Extr}_{\pi,\hat{\pi}} \Bigg\{ \exp\Bigg[-MC \left(\int dx\, d\hat{x}\ \pi(x)\,\hat{\pi}(\hat{x})\,(1+x\hat{x})^n - 1 \right) \\ + \left(\frac{MC}{K} \int \prod_{j=1}^{K} dx_j\ \pi(x_j) \left(1 + \prod_{j=1}^{K} x_j \right)^n - 1 \right) \Bigg] \\ \times \left(\int \prod_{j=1}^{C} d\hat{x}_j\ \hat{\pi}(\hat{x}_j) \left\langle \left[\sum_{\sigma=\pm 1} e^{\sigma\beta F\zeta} \prod_{j=1}^{C} (1+\sigma\hat{x}_j) \right]^n \right\rangle_{\zeta} \right)^M \Bigg\}.\end{aligned} \tag{122}$$

Using (115):

$$\begin{aligned} f = \frac{1}{\beta} \mathrm{Extr}_{\pi,\hat{\pi}} \Bigg\{ \frac{C}{K} \ln 2 + C \int dx\, d\hat{x}\ \pi(x)\,\hat{\pi}(\hat{x})\ \ln(1+x\hat{x}) \\ - \frac{C}{K} \int \prod_{j=1}^{K} dx_j\ \pi(x_j) \ln\left(1 + \prod_{j=1}^{K} x_j \right) \\ - \int \prod_{j=1}^{C} d\hat{x}_j\ \hat{\pi}(\hat{x}_j) \left\langle \ln\left[\sum_{\sigma=\pm 1} \mathrm{e}^{\sigma\beta F\zeta} \prod_{j=1}^{C} (1+\sigma\hat{x}_j) \right] \right\rangle_{\zeta} \Bigg\}. \end{aligned} \tag{123}$$

The extremization above yields a pair of saddle-point equations:

$$\hat{\pi}(\hat{x}) = \int \prod_{j=1}^{K-1} dx_j \, \pi(x_j) \, \delta\left[\hat{x} - \prod_{j=1}^{K-1} x_j\right] \tag{124}$$

$$\pi(x) = \int \prod_{l=1}^{C-1} d\hat{x}_l \, \hat{\pi}(\hat{x}_l) \left\langle \delta\left[x - \tanh\left(\beta F \zeta + \sum_{l=1}^{C-1} \operatorname{atanh} \hat{x}_l\right)\right]\right\rangle_{\zeta},$$

where $\beta = 1$ (Nishimori temperature) and $F = \frac{1}{2} \ln\left(\frac{1-p}{p}\right)$ for optimal decoding.

Following the derivation of Appendix A.3 very closely, the typical overlap $\rho = \langle \frac{1}{M} \sum_{j=1}^{M} \zeta_j \hat{\tau}_j \rangle_{\mathcal{A},\zeta}$ between the estimate $\hat{\tau}_j = \operatorname{sgn}(\langle \tau_j \rangle_\beta)$ and the actual noise ζ_j is given by:

$$\rho = \int dh \, P(h) \operatorname{sgn}(h) \tag{125}$$

$$P(h) = \int \prod_{l=1}^{C} d\hat{x}_l \, \hat{\pi}(\hat{x}_l) \left\langle \delta\left[h - \tanh\left(\beta F \zeta + \sum_{l=1}^{C} \operatorname{atanh} \hat{x}_l\right)\right]\right\rangle_{\zeta}.$$

E. Thermodynamic Quantities and Typical Performance

The typical performance of a code as predicted by the replica symmetric theory can be assessed by solving (124) numerically and computing the overlap ρ using (125). The numerical calculation can be done by representing distributions π and $\hat{\pi}$ by histograms (we have used representations with 20000 bins), and performing Monte Carlo integrations in an iterative fashion until a solution is found. Overlaps can be obtained by plugging the distribution $\hat{\pi}$ that is a solution for (124) into (125).

Numerical calculations show the emergence of two solution types; the first corresponds to a totally aligned (ferromagnetic) state with $\rho = 1$ described by:

$$\pi_{\text{FERRO}}(x) = \delta[x - 1] \qquad \hat{\pi}_{\text{FERRO}}(\hat{x}) = \delta[\hat{x} - 1]. \tag{126}$$

The ferromagnetic solution is the only stable solution up to a specific noise level p_s. Above p_s another stable solution with $\rho < 1$ (suboptimal ferromagnetic) can be obtained numerically. This solution is depicted in Figure 14 for $K = 4$, $C = 3$ and $p = 0.20$. The ferromagnetic state is *always* a stable solution for (124) and is present for all choices of noise level or construction parameters C and K . The stability can be verified by introducing small perturbations to the solution and observing that the solution is recovered after a number of iterations of (124).

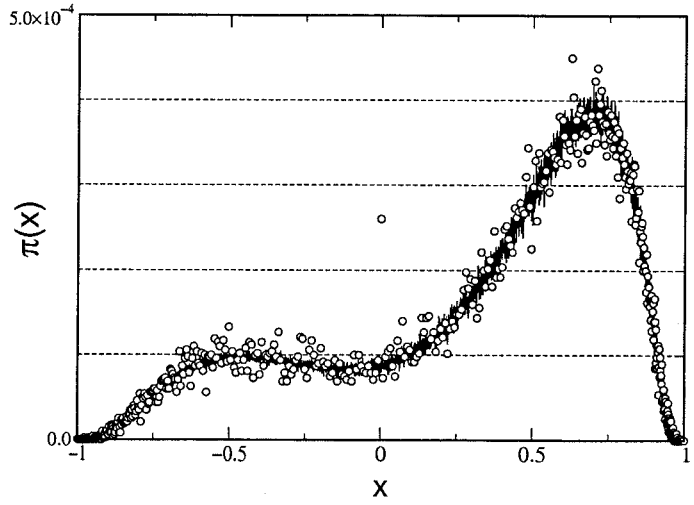

FIGURE 14. Suboptimal ferromagnetic solution $\pi_{\text{NFERRO}}(x)$ for the saddle-point Eqs. (124) obtained numerically. Parameters are $K = 4$, $C = 3$ and $p = 0.20$. Circles correspond to an experimental histogram obtained by decoding with probability propagation in 100 runs for 10 different random constructions.

The free-energy for the ferromagnetic state at Nishimori's temperature is simply $f_{\text{FERRO}} = -F(1 - 2p)$. In Figure 15 we show free-energies for $K = 4$ and $R = 1/4$, p_c indicates the noise level where coexistence between the ferromagnetic and suboptimal ferromagnetic phases occurs. This coexistence noise level coincides, within the numerical precision, with the information theoretic upper bound of Section IV.A. In Figure 16 we show pictorially how the replica symmetric free-energy landscape changes with the noise level p.

In Figure 17 we show the overlap as a function of the noise level, as obtained for $K = 4$ and $R = 1/4$ (therefore $C = 3$). Full lines indicate values corresponding to states of minimum free-energy that are predicted thermodynamically. The general idea is that the macroscopic behavior of the system is dominated by the global minimum of the free-energy (thermodynamic equilibrium state). After a sufficiently long time the system eventually visits configurations consistent with the minimum free-energy state staying there almost all of the time. The whole dynamics is ignored and only the stable equilibrium, in a thermodynamic sense, is taken into account. Also in Figure 17 we show results obtained by simulating probability propagation decoding (black circles). The practical decoding stays in a metastable (in the thermodynamic sense) state between p_s and p_c, and the practical maximum noise level corrected is actually given by p_s. Returning to the pictorial representation in Figure 16, the noise level p_s that provides the practical threshold is signalled by the

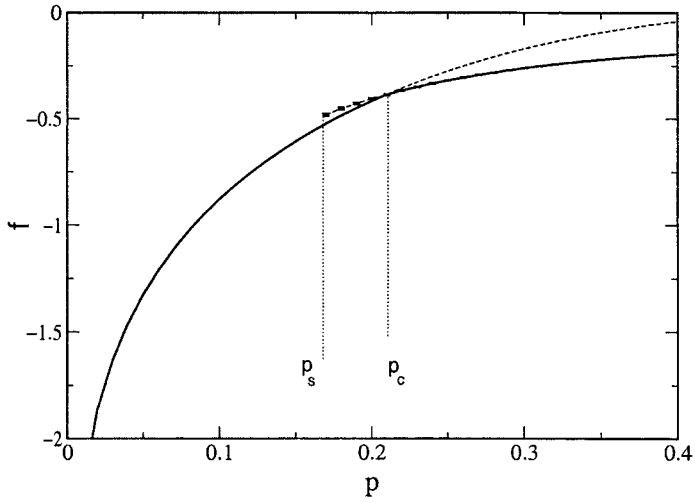

FIGURE 15. Free-energies for $K = 4$, $C = 3$ and $R = 1/4$. The full line corresponds to the free-energy of thermodynamic states. Up to p_s only the ferromagnetic state is present. The ferromagnetic state then dominates the thermodynamics up to p_c, where thermodynamic coexistence with suboptimal ferromagnetic states takes place. Dashed lines correspond to replica symmetric free-energies of nondominant metastable states.

appearance of spinodal points in the replica symmetric free-energy, defined as points separating (meta)stable and unstable regions in the space of thermodynamic configurations (ρ). The noise level p_s may, therefore, be called *spinodal noise level*.

The solutions obtained must produce nonnegative entropies to be physically meaningful. The entropy can be computed from the free-energy (123) as $s = \beta^2 \frac{\partial f}{\partial \beta}$ yielding:

$$s = \beta(u(\beta) - f) \tag{127}$$

$$u(\beta) = -\int \prod_{j=1}^{C} d\hat{x}_j \, \hat{\pi}^*(\hat{x}_j) \left\langle F\zeta \frac{\sum_{\tau=\pm 1} \tau e^{\tau\beta F\zeta} \prod_{j=1}^{C}(1+\tau\hat{x}_j)}{\sum_{\tau=\pm 1} e^{\tau\beta F\zeta} \prod_{j=1}^{C}(1+\tau\hat{x}_j)} \right\rangle_{\zeta},$$

where $\hat{\pi}^*$ is a solution for the saddle-point Eqs. (124) and $u(\beta)$ corresponds to the internal energy density at temperature β. For the ferromagnetic state $s_{\text{FERRO}} = 0$, what indicates that the replica symmetric ferromagnetic solution is physical and that the number of microstates consistent with the ferromagnetic state is at most of polynomial order in N. The entropy of the suboptimal ferromagnetic state can be obtained numerically. Up to the spinodal noise level p_s the entropy vanishes as only the ferromagnetic state is stable. Above p_s, the entropy of the replica symmetric suboptimal ferromagnetic state is negative

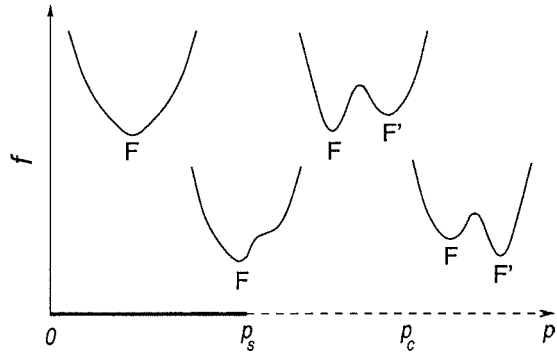

FIGURE 16. Pictorial representation of the replica symmetric free-energy landscape changing with the noise level p. Up to p_s there is only one stable state F corresponding to the ferromagnetic state with $\rho = 1$. At p_s, a second stable suboptimal ferromagnetic state F' emerges with $\rho < 1$, as the noise level increases, coexistence is attained at p_c. Above p_c, F' becomes the global minimum dominating the system thermodynamics.

and, therefore, unphysical. At p_c, the entropy of the suboptimal ferromagnetic state becomes positive again. The internal energy density obtained numerically is depicted in Figure 18 with $u = -F(1-2p)$ for both ferromagnetic and suboptimal ferromagnetic states, justified by assuming Nishimori's condition $\gamma \to \infty$, $\beta = 1$ and $F = \text{atanh}(1-2p)$ (Iba, 1999); (see Appendix B.3).

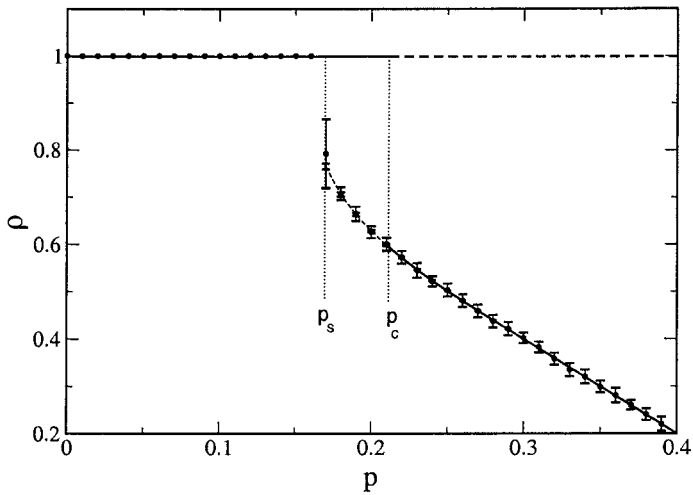

FIGURE 17. Overlaps for $K = 4$, $C = 3$, and $R = 1/4$. The full line corresponds to overlaps predicted by thermodynamic considerations. Up to p_s only the ferromagnetic $\rho = 1$ state is present, it then dominates the thermodynamics up to p_c, where coexistence with suboptimal ferromagnetic states takes place. Dashed lines correspond to overlaps of nondominant states.

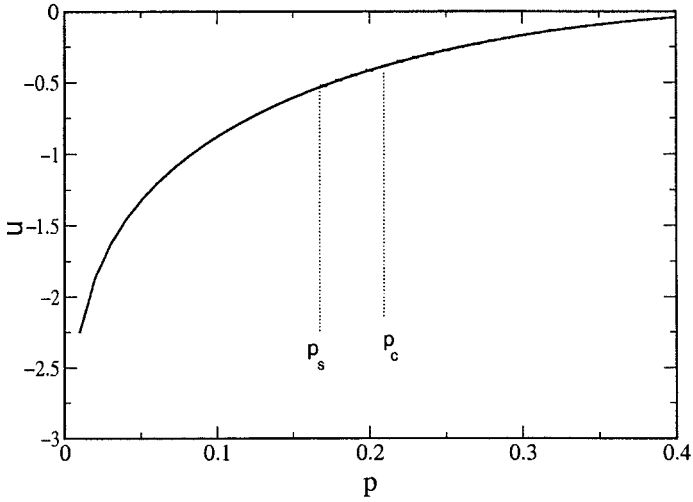

FIGURE 18. Internal energy density for $K = 4$, $C = 3$ and $R = 1/4$ for both ferromagnetic and suboptimal ferromagnetic states. The equality is a consequence of using the Nishimori condition (see Appendix B.3).

The unphysical behavior of the suboptimal ferromagnetic solution between p_s and p_c indicates that the replica symmetric ansätz does not provide the correct physical description of the system. The construction of a complete one-step replica symmetry breaking theory turns out to be a difficult task in the family of models we focus on here (Wong and Sherrington, 1988; Monasson, 1998a,b), although it may be possible in principle using a new method, recently introduced by Mezard and Parisi (2001). An alternative is to consider a frozen spins solution. In this case the entropy in the interval $p_s < p < p_c$ is corrected to $s_{\mathrm{RSB}} = 0$ and the free-energy and internal energy are frozen to the values at p_c.

Any candidate for a physical description for the system would have to be compared with simulations to be validated. Nevertheless, our aim here is to predict the behavior of a particular decoding algorithm, namely, probability propagation. In the next section, we will show that, to this end, the replica symmetric theory will be sufficient.

F. Codes on a Cactus

In this section we present a statistical physics treatment of Gallager codes by employing a mean-field approximation based on the use of a generalized tree structure (Bethe lattice (Wong and Sherrington, 1987b)) known as Husimi

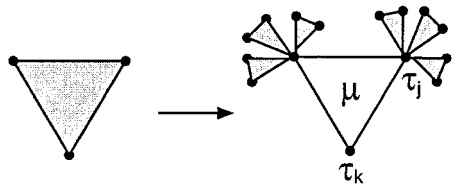

FIGURE 19. First step in the construction of Husimi cactus with $K = 3$ and connectivity $C = 4$.

cactus that is exactly solvable (Gujrati, 1995; Bowman and Levin, 1982; Rieger and Kirkpatrick, 1992; Goldschmidt, 1991).

There are many different ways of building mean-field theories. One can make a perturbative expansion around a tractable model (Plefka, 1982; Tanaka, 2000) or assume a tractable structure and variationally determine the model parameters (Saul and Jordan, 1998). In the approximation we employ, the tractable structure is tree-like and the couplings $\mathcal{J}_\mu$ are just assumed to be those of a model with cycles. In this framework the probability propagation decoding algorithm (PP) emerges naturally, providing an alternative view to the relationship between PP decoding and mean-field approximations already observed in (Kabashima and Saad (1998)). Moreover, this approach has the advantage of being slightly more controlled and easier to understand than replica calculations.

A Husimi cactus with connectivity C is generated starting with a polygon of K vertices with one Ising spin in each vertex (generation 0). All spins in a polygon interact through a single coupling $\mathcal{J}_\mu$ and one of them is called the base spin. In Figure 19 we show the first step in the construction of a Husimi cactus; in a generic step, the base spins of the $(C-1)(K-1)$ polygons in generation $n-1$ are attached to $K-1$ vertices of a polygon in the next generation n. This process is iterated until a maximum generation $n_{\max}$ is reached; the graph is then completed by attaching C uncorrelated branches of $n_{\max}$ generations at their base spins. In this way each spin inside the graph is connected to C polygons exactly. The local magnetization at the center m_j can be obtained by fixing boundary (initial) conditions in the zeroth generation and iterating the related recursion equations until generation $n_{\max}$ is reached. Carrying out the calculation in the thermodynamic limit corresponds to having $n_{\max} \sim \ln M$ generations and $M \to \infty$.

The Hamiltonian of the model has the form (106) where $\mathcal{L}(\mu)$ denotes the polygon μ of the lattice. Due to the tree-like structure, local quantities far from the boundary can be calculated recursively by specifying boundary conditions. The typical decoding performance can therefore be computed exactly without resorting to replica calculations (Gujrati, 1995).

We adopt the approach presented in Rieger and Kirkpatrick (1992) for obtaining recursion relations. The probability distribution $P_{\mu k}(\tau_k)$ for the base spin of the polygon μ is connected to $(C-1)$ $(K-1)$ distributions $P_{\nu j}(\tau_j)$, with $\nu \in \mathcal{M}(j) \setminus \mu$ (all polygons linked to j but μ) of polygons in the previous generation:

$$P_{\mu k}(\tau_k) = \frac{1}{\mathcal{N}} \mathrm{Tr}_{\{\tau_j\}} \exp\left[\beta\gamma\left(\mathcal{J}_\mu \tau_k \prod_{j\in\mathcal{L}(\mu)\setminus k} \tau_j - 1\right) + \beta F \tau_k\right] \times \prod_{\nu\in\mathcal{M}(j)\setminus\mu} \prod_{j\in\mathcal{L}(\mu)\setminus k} P_{\nu j}(\tau_j), \tag{128}$$

where the trace is over the spins τ_j such that $j \in \mathcal{L}(\mu) \setminus k$.

The effective field $\hat{x}_{\nu j}$ on a base spin j due to neighbors in polygon ν can be written as:

$$\mathrm{e}^{-2\hat{x}_{\nu j}} = \mathrm{e}^{2\beta F} \frac{P_{\nu j}(-)}{P_{\nu j}(+)}, \tag{129}$$

Combining (128) and (129) we find the recursion relation (see Appendix B.4 for details):

$$\mathrm{e}^{-2\hat{x}_{\mu k}} = \frac{\mathrm{Tr}_{\{\tau_j\}} e^{-\beta\gamma\mathcal{J}_\mu \prod_{j\in\mathcal{L}(\mu)\setminus k}\tau_j + \sum_{j\in\mathcal{L}(\mu)\setminus k}(\beta F + \sum_{\nu\in\mathcal{M}(j)\setminus\mu}\hat{x}_{\nu j})\tau_j}}{\mathrm{Tr}_{\{\tau_j\}} e^{+\beta\gamma\mathcal{J}_\mu \prod_{j\in\mathcal{L}(\mu)\setminus k}\tau_j + \sum_{j\in\mathcal{L}(\mu)\setminus k}(\beta F + \sum_{\nu\in\mathcal{M}(j)\setminus\mu}\hat{x}_{\nu j})\tau_j}}. \tag{130}$$

By computing the traces and taking $\gamma \to \infty$ and $\beta = 1$ one obtains:

$$\hat{x}_{\mu k} = \mathrm{atanh}\left[\mathcal{J}_\mu \prod_{j\in\mathcal{L}(\mu)\setminus k} \tanh\left(F + \sum_{\nu\in\mathcal{M}(j)\setminus\mu} \hat{x}_{\nu j}\right)\right] \tag{131}$$

The effective local magnetization due to interactions with the nearest neighbors in one branch is given by $\hat{m}_{\mu j} = \tanh(\hat{x}_{\mu j})$. The effective local field on a base spin j of a polygon μ due to $C-1$ branches in the previous generation and due to the external field is $x_{\mu j} = F + \sum_{\nu\in\mathcal{M}(j)\setminus\mu} \hat{x}_{\nu j}$; the effective local magnetization is therefore $m_{\mu j} = \tanh(x_{\mu j})$. Equation (131) can then be rewritten in terms of $\hat{m}_{\mu j}$ and $m_{\mu j}$ and the PP equations (MacKay, 1999; Kabashima and Saad 1998; Kschischang and Frey, 1998) can be recovered:

$$\begin{aligned} m_{\mu k} &= \tanh\left(F + \sum_{\nu\in\mathcal{M}(k)\setminus\mu} \mathrm{atanh}(\hat{m}_{\nu k})\right) \\ \hat{m}_{\mu k} &= \mathcal{J}_\mu \prod_{j\in\mathcal{L}(\mu)\setminus k} m_{\mu j} \end{aligned} \tag{132}$$

Once the magnetization on the boundary (zeroth generation) are assigned, the local magnetization m_j in the central site is determined by iterating (132)

and computing:

$$m_j = \tanh\left(F + \sum_{\nu\in\mathcal{M}(j)} \operatorname{atanh}(\hat{m}_{\nu j})\right) \tag{133}$$

A free-energy can be obtained by integration of (132) (Murayama *et al.,* 2000; Vicente *et al.,* 2000b; Bowman and Levin, 1982). The Eqs. (132) describing PP decoding represent extrema of the following free-energy:

$$\mathcal{F}(\{m_{\mu k}, \hat{m}_{\mu k}\}) = \sum_{\mu=1}^{M-N} \sum_{i\in\mathcal{L}(\mu)} \ln(1 + m_{\mu i}\hat{m}_{\mu i}) - \sum_{\mu=1}^{M-N} \ln\left(1 + \mathcal{J}_\mu \prod_{i\in\mathcal{L}(\mu)} m_{\mu i}\right) - \sum_{j=1}^{M} \ln\left[\mathrm{e}^{F} \prod_{\mu\in\mathcal{M}(j)} (1 + \hat{m}_{\mu j}) + \mathrm{e}^{-F} \prod_{\mu\in\mathcal{M}(j)} (1 - \hat{m}_{\mu j})\right] \tag{134}$$

The iteration of the maps (132) is actually one out of many different methods of finding stable extrema of this free-energy.

The decoding process can be performed by iterating the multidimensional map (132) using some defined scheduling. Assume that the iterations are performed in parallel using the following procedure:

(i) Effective local magnetizations are initialized as $m_{\mu k} = 1 - 2p$, reflecting prior probabilities.
(ii) Conjugate magnetizations $\hat{m}_{\mu k}$ are updated.
(iii) Magnetizations $m_{\mu k}$ are computed.
(iv) If convergence or a maximal number of iterations is attained, stop. Otherwise go to step (ii).

Equations (132) have fixed points that are inconveniently dependent on the particular noise vector $\boldsymbol{\zeta}$. By applying the gauge transformation $\mathcal{J}_\mu \mapsto 1$ and $\tau_j \mapsto \tau_j \zeta_j$ we get a map with noise-independent fixed points that has the following form:

$$m_{\mu k} = \tanh\left(\zeta_k F + \sum_{\nu\in\mathcal{M}(k)\setminus\mu} \operatorname{atanh}(\hat{m}_{\nu k})\right) \tag{135}$$

$$\hat{m}_{\mu k} = \prod_{j\in\mathcal{L}(\mu)\setminus k} m_{\mu j}. \tag{136}$$

In terms of effective fields $x_{\mu k}$ and $\hat{x}_{\mu k}$ we have:

$$x_{\mu k} = \zeta_k F + \sum_{\nu\in\mathcal{M}(k)\setminus\mu} \hat{x}_{\nu k} \qquad \hat{x}_{\mu k} = \operatorname{atanh}\left(\prod_{j\in\mathcal{L}(\mu)\setminus k} \tanh(x_{\mu j})\right). \tag{137}$$

The above equations provide a microscopic description for the dynamics of a probability propagation decoder; a macroscopic description can be constructed by retaining only statistical information about the system, namely, by describing the evolution of histograms of variables $x_{\mu k}$ and $\hat{x}_{\mu k}$.

Assume that the effective fields $x_{\mu k}$ and $\hat{x}_{\mu k}$ are random variables *independently* sampled from the distributions $P(x)$ and $\hat{P}(\hat{x})$, respectively; in the same way assume that ζ_j is sampled from $P(\zeta) = (1-p)\,\delta(\zeta - 1) + \delta(\zeta + 1)$. A recursion relation in the space of probability distributions (Bowman and Levin, 1982) can be found from Eq. (137):

$$P_n(x) = \int d\zeta\, P(\zeta) \int \prod_{l=1}^{C-1} d\hat{x}_l\, \hat{P}_{n-1}(\hat{x}_l)\, \delta\left[x - F\zeta - \sum_{l=1}^{C-1} \hat{x}_l\right]$$

$$\hat{P}_{n-1}(\hat{x}) = \int \prod_{j=1}^{K-1} dx_j\, P_{n-1}(x_j)\, \delta\left[\hat{x} - \operatorname{atanh}\left(\prod_{j=1}^{K-1} \tanh(x_j)\right)\right], \quad (138)$$

where $P_n(x)$ is the distribution of effective fields at the nth generation due to the previous generations and external fields, in the thermodynamic limit the distribution far from the boundary will be $P_\infty(x)$ (generation $n \to \infty$). The local field distribution at the central site is computed by replacing $C - 1$ by C in the first Eq. (138), taking into account C polygons in the generation just before the central site, and inserting the distribution $P_\infty(x)$:

$$P(h) = \int d\zeta\, P(\zeta) \int \prod_{l=1}^{C} d\hat{x}_l\, \hat{P}_\infty(\hat{x}_l)\, \delta\left[x - F\zeta - \sum_{l=1}^{C} \hat{x}_l\right]. \quad (139)$$

It is easy to see that $P_\infty(x)$ and $\hat{P}_\infty(\hat{x})$ satisfy Eqs. (124) obtained by the replica symmetric theory (Kabashima *et al.,* 2000; Murayama *et al.,* 2000; Vicente *et al.,* 2000b), if the variables describing fields are transformed to those of local magnetizations through $x \mapsto \tanh(\beta x)$.

In Figure 14 we show empirical histograms obtained by performing 100 runs of PP decoding for 10 different codes of size $M = 5000$ and compare with a distribution obtained by solving equations like (138). The practical PP decoding is performed by setting initial conditions as $m_{\mu j} = 1 - 2p$ to correspond to the prior probabilities and iterating (132) until stationarity or a maximum number of iterations is attained (MacKay, 1999). The estimate for the noise vector is then produced by computing $\hat{\tau}_j = \operatorname{sign}(m_j)$. At each decoding step the system can be described by histograms of variables (132), this is equivalent to iterating (138) (a similar idea was presented in MacKay (1999) and Davey (1998)).

In Figure 20 we summarize the transitions obtained for $K = 6$ and $K = 10$. A dashed line indicates Shannon's limit, the full line represents the information theoretic upper bound of Section IV.A, white circles stand for the coexistence line obtained numerically. Diamonds represent spinodal noise levels obtained

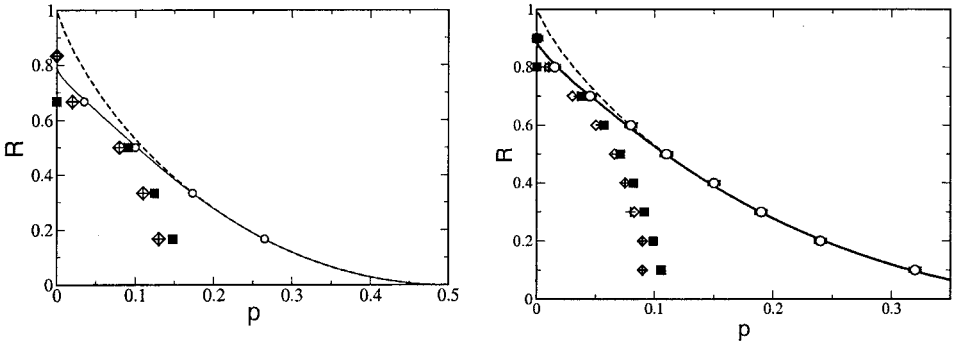

FIGURE 20. Transitions for Gallager codes with $K = 6$ (left) and $K = 10$ (right). Shannon's bound (dashed line), information theory upper bound (full line), and thermodynamic transition obtained numerically (○). Transitions obtained by Monte Carlo integration of Eq. (138) (◇) and by simulations of PP decoding (+, $M = 5000$ averaged over 20 runs) are also shown. Black squares are estimates for practical thresholds based on Sec. IV.H. In both figures, symbols are chosen larger than the error bars.

by solving (138) numerically and (+) are results obtained by performing 20 runs using PP decoding. It is interesting to observe that the practical performance tends to get worse as K grows large, which agrees with the general belief that decoding gets harder as Shannon's limit is approached.

G. Tree-Like Approximation and the Thermodynamic Limit

The geometric structure of a Gallager code defined by the matrix $\boldsymbol{A}$ can be represented by a bipartite graph as in Figure 21 (*Tanner graph*) (Kschischang

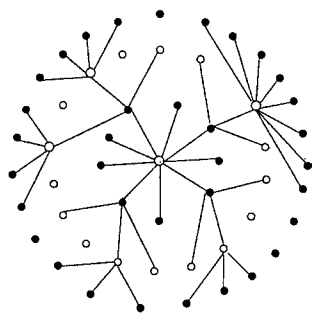

FIGURE 21. Tanner graph representing the neighborhood of a bit node in an irregular Gallager code. Black circles represent checks and white circles represent bits.

and Frey, 1998) with bit and check nodes (in this case, we show an *irregular* constraction where the values of K and C are not fixed). Each column j of $\boldsymbol{A}$ represents a bit node and each row μ represents a check node; $A_{\mu j} = 1$ means that there is an edge linking bit j to check μ. It is possible to show (Richardson and Urbanke, 2001) that for a random ensemble of regular codes, the probability of completing a cycle after walking l edges starting from an arbitrary node is upper bounded by $\mathcal{P}[l; K, C, M] \leq l^2 K^l / M$. It implies that for very large M only cycles of at least order $\ln M$ survive. In the thermodynamic limit $M \to \infty$ and the probability $\mathcal{P}[l; K, C, M] \to 0$ for any finite l and the bulk of the system is effectively tree-like. By mapping each check node to a polygon with K bit nodes as vertices, one can map a Tanner graph into a Husimi lattice that is effectively a tree for any number of generations of order less than $\ln M$. In Figure 22 we show that the number of iterations of (132) required for convergence far from the threshold does not scale with the system size, therefore, it is expected that the interior of a tree-like lattice approximates a Gallager code with increasing accuracy as the system size increases. Figure 23 shows that the approximation is fairly good even for sizes as small as $M = 100$ when compared to theoretical results and simulations for size $M = 5000$. Nevertheless, the difference increases as the spinodal noise level approaches, what seems to indicate the breakdown of the approximation. A possible explanation is that convergence times larger than $\mathcal{O}(\ln M)$ may be required in this region. An interesting analysis of the convergence properties

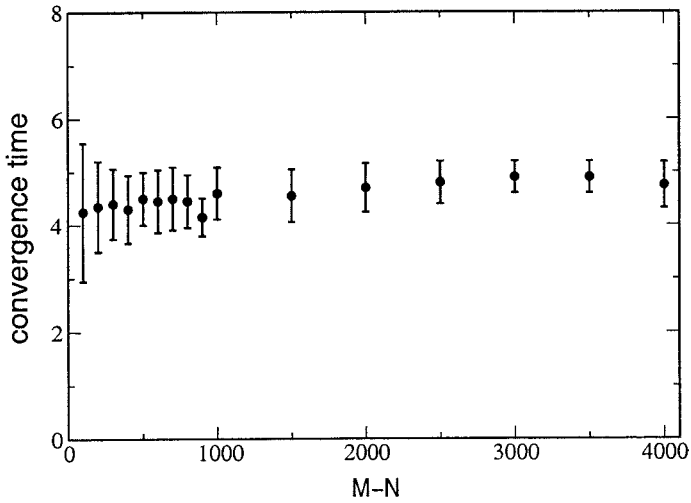

FIGURE 22. PP decoding convergence time as a function of the code size $(M - N)$ for $K = 4C = 3$ and $p = 0.05$, therefore, well below the threshold. The convergence time clearly does not scale with the system size.

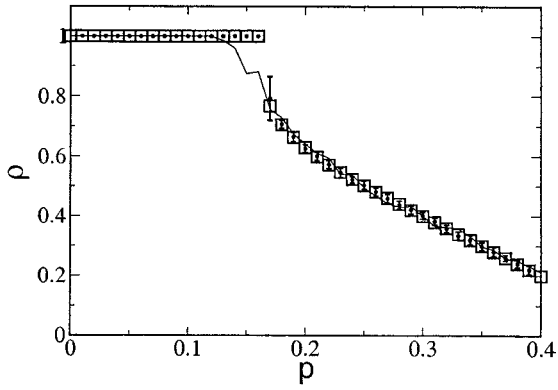

FIGURE 23. Mean normalized overlap ρ between the actual noise vector ζ and decoded noise $\hat{\tau}$ for a Gallager code with $K = 4$ and $C = 3$ (therefore $R = 1/4$). Theoretical values (□) obtained by Monte Carlo integration of Eqs.(138) and averages of 20 simulations of PP decoding for code word lengths $M = 5000$ (●) and $M = 100$ (full line). Symbols are chosen larger than the error bars.

of probability propagation algorithms for some specific graphical models can be found in Weiss (1997).

H. Estimating Spinodal Noise Levels

We now estimate the threshold noise level p_s by introducing a measure for the number of parity checks violated by a bit τ_l:

$$E_l = -\sum_{\mu \in \mathcal{M}(l)} \left(\mathcal{J}_\mu \tau_l \prod_{j \in \mathcal{L}(\mu) \setminus l} \tau_j - 1 \right). \tag{140}$$

By using gauged variables:

$$E_l = -\sum_{\mu \in \mathcal{M}(l)} \left(\tau_l \prod_{j \in \mathcal{L}(\mu) \setminus l} \tau_j - 1 \right). \tag{141}$$

Suppose that random guesses are generated by sampling the prior distribution, their typical overlap will be $\rho = 1 - 2p$. Assume now that the vectors sampled are corrected by flipping τ_l if $E_l = C$. If the landscape has a single dominant minimum, we expect that this procedure will tend to increase the overlap ρ between τ and the actual noise vector ζ in the first step up to the noise level p_s, where suboptimal microscopic configurations are expected to emerge. Above p_s, there is a large number of suboptimal ferromagnetic microstates with an overlap around $\rho = 1 - 2p$ (see Fig. 23), and we expect that if a single bit of a randomly guessed vector is corrected, the overlap will then

either increase or decrease, staying unchanged on average. A vanishing variation in the mean overlap would, therefore, signal the emergence of suboptimal microstates at p_s.

The probability that a bit $\tau_l = +1$ is corrected is:

$$P(E_l = C \mid \tau_l = +1) = \prod_{\mu \in \mathcal{M}(l)} P\left\{ \prod_{j \in \mathcal{L}(\mu) \setminus l} \tau_j = -1 \right\}. \tag{142}$$

For a a bit $\tau_l = -1$:

$$P(E_l = C \mid \tau_l = -1) = \prod_{\mu \in \mathcal{M}(l)} \left[1 - P\left\{ \prod_{j \in \mathcal{L}(\mu) \setminus l} \tau_j = -1 \right\} \right]. \tag{143}$$

Considering vectors sampled from a prior $P(\tau) = (1-p)\,\delta(\tau - 1) + p\,\delta(\tau + 1)$ we have:

$$P\left\{ \prod_{j \in \mathcal{L}(\mu) \setminus l} \tau_j = -1 \right\} = \frac{1}{2} - \frac{1}{2}(1-2p)^{K-1}. \tag{144}$$

The gauged overlap is defined as $\rho = \sum_{j=1}^{M} S_j$ and the variation on the overlap after flipping a bit l is $\Delta\rho = \rho_1 - \rho_0 = S_l^1 - S_l^0$. The mean variation in the overlap due to a flip in a bit τ_l with $E_l = C$ is therefore:

$$\begin{aligned} \frac{1}{2}\langle \Delta\rho \rangle &= P(\tau_l = +1 \mid E_l = C) - P(\tau_l = -1 \mid E_l = C) \\ &= \frac{\sum_{\tau_l = \pm 1} \tau_l P(E_l = C \mid \tau_l) P(\tau_l)}{\sum_{\tau_l = \pm 1} P(E_l = C \mid \tau_l)) P(\tau_l)}, \end{aligned} \tag{145}$$

where we applied the Bayes theorem to obtain the last line.

By plugging the prior probability (142) and (144) into the above expression we get:

$$\frac{1}{2}\langle \Delta\rho \rangle = \frac{[1-(1-2p)^{K-1}]^C (1-p) - [1+(1-2p)^{K-1}]^C p}{[1-(1-2p)^{K-1}]^C (1-p) + [1+(1-2p)^{K-1}]^C p}. \tag{146}$$

At p_s we have $\langle \Delta\rho \rangle = 0$ and:

$$\frac{p_s}{1-p_s} = \left[\frac{1-(1-2p_s)^{K-1}}{1+(1-2p_s)^{K-1}} \right]. \tag{147}$$

The above equation can be solved numerically yielding reasonably accurate estimates for practical thresholds p_s as can be seen in Figure 20.

MacKay (1999) and Gallager (1962, 1963) introduced probabilistic decoding algorithms whose performance analysis is essentially the same those as

presented here. However, the results obtained in Section IV.C put the analysis into a broader perspective: algorithms that generate decoding solutions in polynomial time, as is the case of probabilistic decoding or probability propagation, seem to be bounded by the practical threshold p_s due to the presence of suboptimal solutions. On the other hand, decoding in exponential time is always possible up to the thermodynamic transition at p_c (with p_c attaining channel capacity if $K \to \infty$), by performing an exhaustive search for the global minimum of the free-energy (134).

V. MacKay–Neal Codes

MacKay–Neal (MN) codes were introduced in 1995 as a variation on Gallager codes. As in the case of Gallager codes (see Section IV), MN codes are defined by two very sparse matrices, but with the difference that information on both noise and signal is incorporated to the syndrome vector. MN codes are also decoded using sparse matrices, while encoding uses a dense matrix, which yields good distance properties and a decoding problem solvable in linear time by using the methods of probability propagation.

Cascading codes, a class of constructions inside the MN family recently proposed by Kanter and Saad (1999, 2000a,b) have been shown to outperform some of the cutting-edge Gallager and turbo code constructions. We will discuss cascading codes in the next secion, but this fact alone justifies a thorough study of MN codes.

Theorems showing the asymptotic *goodness* of the MN family have been proved in (MacKay, 1999). By assuming equal message and noise biases (for a BSC), it was proved that the probability of error vanishes as the message length increases and that it is possible to get as close as desired to channel capacity by increasing the number of nonzero elements in a column of the very sparse matrices defining the code.

It can also be shown by a simple upper bound that MN codes, unlike Gallager codes, might, as well, attain Shannon's bound for a finite number of nonzero elements in the columns of the very sparse matrices, given that unbiased messages are used. This upper bound does not guarantee that channel capacity can be attained in polynomial time or even that it can be attained at all. Results obtained using statistical physics techniques (Kabashima *et al.,* 2000; Murayama *et al.,* 2000; Vicente *et al.,* 2000a,b) seem to indicate that Shannon's bound can actually be approached with exponential time decoding. This feature is considered to be new and somewhat surprising (D. MacKay, personal communication, 2000).

Statistical physics has been applied to analyze MN codes and its variants (Kabashima *et al.,* 2000; Murayama *et al.,* 2000; Vicente *et al.,* 2000b). In this

analysis we use the replica symmetric theory to obtain all relevant thermodynamic quantities and to calculate the phase diagram. The theory also yields a noise level where suboptimal solutions emerge that are in connection with the practical thresholds observed when probability propagation decoding is used.

Assuming that a message is represented by a binary vector $\boldsymbol{\xi} \in \{0, 1\}^N$ sampled independently from the distribution $P(\xi) = (1 - p_\xi)\,\delta(\xi) + p_\xi\,\delta(\xi - 1)$, the MN encoding process consists of producing a binary vector $\boldsymbol{t} \in \{0, 1\}^M$ defined by

$$\boldsymbol{t} = \boldsymbol{G}\boldsymbol{\xi} \,(\text{mod } 2), \tag{148}$$

where all operations are performed in the field $\{0, 1\}$ and are indicated by (mod 2). The code rate is, therefore, $R = N/M$.

The generator matrix $\boldsymbol{G}$ is an $M \times N$ dense matrix defined by

$$\boldsymbol{G} = \boldsymbol{C}_n^{-1}\boldsymbol{C}_s \,(\text{mod } 2), \tag{149}$$

with $\boldsymbol{C}_n$ being an $M \times M$ binary invertible sparse matrix and $\boldsymbol{C}_s$ an $M \times N$ binary sparse matrix.

The transmitted vector $\boldsymbol{t}$ is then corrupted by noise. We here assume a memoryless binary symmetric channel (BSC), namely, noise is represented by a binary vector $\boldsymbol{\zeta} \in \{0, 1\}^M$ with components independently drawn from the distribution $P(\zeta) = (1 - p)\,\delta(\zeta) + p\,\delta(\zeta - 1)$.

The received vector takes the form

$$\boldsymbol{r} = \boldsymbol{G}\boldsymbol{\xi} + \boldsymbol{\zeta} \,(\text{mod } 2). \tag{150}$$

Decoding is performed by preprocessing the received message with the matrix $\boldsymbol{C}_n$ and producing the syndrome vector

$$z = \boldsymbol{C}_n\boldsymbol{r} = \boldsymbol{C}_s\boldsymbol{\xi} + \boldsymbol{C}_n\boldsymbol{\zeta} \,(\text{mod } 2), \tag{151}$$

from which an estimate $\hat{\boldsymbol{\xi}}$ for the message can be directly obtained.

An MN code is called *regular* if the number of elements set to one in each row of $\boldsymbol{C}_s$ is chosen to be K and the number of elements in each column is set to be C. For the square matrix $\boldsymbol{C}_n$ the number of elements in each row (or column) is set to L. In this case the total number of ones in the matrix $\boldsymbol{C}_s$ is $MK = NC$, yielding that the rate can alternatively be expressed as $R = K/C$.

In contrast, an MN code is called *irregular* if each row m in $\boldsymbol{C}_s$ and $\boldsymbol{C}_n$ contains K_m and L_m nonzero elements, respectively. In the same way, each column j of $\boldsymbol{C}_s$ contains C_j nonzero elements and each column l of $\boldsymbol{C}_n$ contains D_l nonzero elements.

Counting the number of nonzero elements in the matrices leads to the following relations:

$$\sum_{j=1}^{N} C_j = \sum_{\mu=1}^{M} K_\mu \qquad \sum_{l=1}^{M} D_l = \sum_{\mu=1}^{M} L_\mu, \tag{152}$$

The code rate is, therefore, $R = \overline{K}/\overline{C}$, where:

$$\overline{K} = \frac{1}{M}\sum_{\mu=1}^{M} K_\mu \qquad \overline{C} = \frac{1}{N}\sum_{j=1}^{N} C_j. \tag{153}$$

The Bayes optimal estimator $\hat{\boldsymbol{\xi}}$ for the message $\boldsymbol{\xi}$ is $\hat{\xi}_j = \operatorname{argmax}_{S_j} P(S_j \mid \boldsymbol{z})$. The performance of this estimator is measured by the probability of bit error $p_b = 1 - 1/N \sum_{j=1}^{N} \delta[\hat{\xi}_j; \xi_j]$, where $\delta[;]$ is the Kronecker delta. Knowing the matrices $\boldsymbol{C_s}$ and $\boldsymbol{C_n}$, the syndrome vector $\boldsymbol{z}$, the noise level p, and the message bias p_ξ, the posterior probability is computed by applying Bayes theorem:

$$P(\boldsymbol{S}, \boldsymbol{\tau} \mid \boldsymbol{z}) = \frac{1}{Z}\chi\left[\boldsymbol{z} = \boldsymbol{C_s S} + \boldsymbol{C_n \tau} \pmod 2\right] P(\boldsymbol{S}) P(\boldsymbol{\tau}), \tag{154}$$

where $\chi[X]$ is an indicator function providing 1 if X is true and 0 otherwise.

To obtain the estimate one has to compute the marginal posterior

$$P(S_j \mid \boldsymbol{z}) = \sum_{\{S_i : i \neq j\}} \sum_{\boldsymbol{\tau}} P(\boldsymbol{S}, \boldsymbol{\tau} \mid \boldsymbol{z}), \tag{155}$$

which requires $\mathcal{O}(2^N)$ operations and is impractical for long messages. Again we can use the sparseness of $[\boldsymbol{C_s} \mid \boldsymbol{C_n}]$ and the methods of probability propagation for decoding, which requires only $\mathcal{O}(N)$ operations.

When $p = p_\xi$, MN and Gallager codes are equivalent under a proper transformation of parameters, as the code rate is $R = N/M$ for MN codes and $R = 1 - N/M$ for Gallager codes. The main difference between the codes is in the syndrome vector $\boldsymbol{z}$. For MN codes, the syndrome vector incorporates information on both message and noise while for Gallager codes, only information on the noise is present (see Eq. (103)). This feature opens the possibility of adjusting the code behavior by controlling the message bias p_ξ.

An MN code can be thought of as a nonlinear code (MacKay, 2000b). Redundancy in the original message could be removed (introduced) by using a source (de)compressor defined by some nonlinear function $\boldsymbol{\xi} = g(\boldsymbol{\xi}_0; p_\xi)$, and encoding would then be $\boldsymbol{t} = \boldsymbol{G}\, g(\boldsymbol{\xi}_0; p_\xi) \pmod 2$. In the following we show that other new features emerge due to the introduction of the parameter p_ξ.

A. Upper Bound on Achievable Rates

In a regular MN code, the syndrome vector $\boldsymbol{z} = \boldsymbol{C_s S} + \boldsymbol{C_n \tau}$ (mod 2) is a sum of K message bits drawn from the distribution $P(\xi) = (1 - p_\xi)\,\delta(\xi) + p_\xi\,\delta(\xi - 1)$ and L noise bits drawn from $P(\zeta) = (1 - p)\,\delta(\zeta) + p\,\delta(\zeta - 1)$.

The probability of $z_j = 1$ is (see Appendix C.1)

$$p_z^1(K, L) = \frac{1}{2} - \frac{1}{2}(1 - 2p_\xi)^K (1 - 2p)^L. \tag{156}$$

The maximum information content in the syndrome vector is $M H_2(p_z^1(K, L))$ (in *bits* or *shannons*), where H_2(x) is the binary entropy. The amount of information needed to reconstruct both the message vector $\boldsymbol{\xi}$ and the noise vector $\boldsymbol{\zeta}$ is $N H_2(p_\xi) + M H_2(p)$ (in *bits* or *shannons*). Thus, it is a necessary condition for successful decoding that:

$$\begin{aligned} M H_2\big(p_z^1(K, L)\big) &\geq N H_2(p_\xi) + M H_2(p) \\ H_2\big(p_z^1(K, L)\big) - H_2(p) &\geq R H_2(p_\xi) \\ R &\leq \frac{H_2\big(p_z^1(K, L)\big) - H_2(p)}{H_2(p_\xi)}. \end{aligned} \tag{157}$$

For the case $p_\xi = p$ and $L = C$, we can recover bounds (105) for Gallager codes with dimensions and parameters redefined as $M' = M + N$, $N' = N$ and $K' = K + L$. In MacKay (1999), a theorem stating that channel capacity can be attained when $K \to \infty$ was proved for this particular case.

If unbiased ($p_\xi = 1/2$) messages are used, $H_2(p_\xi) = 1$, $H_2(p_z^1(K, L)) = 1$ and the bound (157) becomes

$$R \leq 1 - H_2(p), \tag{158}$$

i.e., MN codes may be capable of attaining channel capacity even for finite K and L, given that unbiased messages are used.

B. Statistical Physics Formulation

The statistical physics formulation for MN codes is a straightforward extension of the formulation presented for Gallager codes. The field ($\{0, 1\}$, +(mod 2)) is replaced by ($\{\pm 1\}$, $\times$) (Sourlas, 1989) and the syndrome vector acquires the form:

$$\mathcal{J}_\mu = \prod_{j \in \mathcal{L}_s(\mu)} \xi_j \prod_{l \in \mathcal{L}_n(\mu)} \zeta_l \tag{159}$$

where $j = 1, \ldots, N$, $l = 1, \ldots, M$ and $\mu = 1, \ldots, M$.

The K_μ indices of nonzero elements in the row μ of the *signal matrix* $\boldsymbol{C_s}$ are given by $\mathcal{L}_s(\mu) = \{j_1, \ldots, j_{K_\mu}\}$, and in a column j are given by $\mathcal{M}_s(j) = \{\mu_1, \ldots, \mu_{C_j}\}$. In the same way, for the *noise matrix* $\boldsymbol{C_n}$, the L_μ indices of nonzero elements in the row μ are given by $\mathcal{L}_n(\mu) = \{j_1, \ldots, j_{L_\mu}\}$, and in a column l are given by $\mathcal{M}_n(l) = \{\mu_1, \ldots, \mu_{D_l}\}$.

Under the assumption that priors $P(\boldsymbol{S})$ and $P(\boldsymbol{\tau})$ are completely factorizable, the posterior (154) corresponds to the limit $\gamma \to \infty$ and $\beta = 1$ (Nishimori temperature) of:

$$P_\gamma(\boldsymbol{S}, \boldsymbol{\tau} \mid \mathcal{J}) = \frac{1}{Z} \exp[-\beta \mathcal{H}_\gamma(\boldsymbol{S}, \boldsymbol{\tau}; \mathcal{J})] \tag{160}$$

$$\mathcal{H}_\gamma(\boldsymbol{S}, \boldsymbol{\tau}; \mathcal{J}) = -\gamma \sum_{\mu=1}^{M} \left(\mathcal{J}_\mu \prod_{j \in \mathcal{L}_s(\mu)} S_j \prod_{l \in \mathcal{L}_n(\mu)} \tau_l - 1 \right) - F_s \sum_{j=1}^{N} S_j - F_n \sum_{l=1}^{M} \tau_l,$$

with $F_s = \frac{1}{2}\text{atanh}(\frac{1-p_\xi}{p_\xi})$ and $F_n = \frac{1}{2}\text{atanh}(\frac{1-p}{p})$ (*Nishimori condition* (Iba, 1999)).

By applying the gauge transformation $S_j \mapsto S_j \xi_j$ and $\tau_l \mapsto \tau_l \zeta_l$ the couplings can be gauged out $\mathcal{J}_\mu \mapsto 1$, eliminating the disorder. The model is free of frustration (as in Toulouse, 1977, the model is *flat*). Similar to Gallager codes, the resulting Hamiltonian consists of two sublattices interacting via multispin ferromagnetic iterations with finite connectivity in random fields $\xi_j F_s$ and $\zeta_l F_n$:

$$\mathcal{H}_\gamma^{\text{gauge}}(\boldsymbol{S}, \boldsymbol{\tau}; \boldsymbol{\xi}, \boldsymbol{\zeta}) = -\gamma \sum_{\mu=1}^{M} \left(\prod_{j \in \mathcal{L}_s(\mu)} S_j \prod_{l \in \mathcal{L}_n(\mu)} \tau_l 1 \right) - F_s \sum_{j=1}^{N} \xi_j S_j - F_n \sum_{l=1}^{M} \zeta_l \tau_l. \tag{161}$$

At the Nishimori condition $\gamma \to \infty$, the model can also be regarded as a paramagnet with restricted configuration space on a nonuniform external field:

$$\mathcal{H}^{\text{gauge}}((\boldsymbol{S}, \boldsymbol{\tau}) \in \Omega; \boldsymbol{\xi}, \boldsymbol{\zeta}) = -F_s \sum_{j=1}^{N} \xi_j S_j - F_n \sum_{l=1}^{M} \zeta_l \tau_l, \tag{162}$$

where

$$\Omega = \left\{ (\boldsymbol{S}, \boldsymbol{\tau}) : \prod_{j \in \mathcal{L}_s(\mu)} S_j \prod_{l \in \mathcal{L}_n(\mu)} \tau_l = 1, \mu = 1, \ldots, M \right\}. \tag{163}$$

Optimal decoding consists of finding local magnetizations at the Nishimori temperature in the *signal sublattice* $m_j = \langle S_j \rangle_{\beta_N}$ and calculating Bayesian estimates $\hat{\xi}_j = \text{sgn}(m_j)$.

The probability of bit error is

$$p_b = \frac{1}{2} - \frac{1}{2N} \sum_{j=1}^{N} \xi_j \, \mathrm{sgn}(m_j), \tag{164}$$

connecting the code performance with the computation of local magnetizations.

C. Replica Theory

The replica theory for MN codes is the theory constructed for Gallager codes, with the introduction of extra dynamic variables $\boldsymbol{S}$. The gauged Hamiltonian (161) is written as:

$$\begin{aligned} \mathcal{H}_{\gamma}^{\text{gauge}}(\boldsymbol{S}, \boldsymbol{\tau}; \boldsymbol{\xi}, \boldsymbol{\zeta}) = -\gamma \sum_{\langle \boldsymbol{jl} \rangle} \mathcal{A}_{\langle \boldsymbol{jl} \rangle} (S_{j_1} \cdots S_{j_K} \tau_{l_1} \cdots \tau_{l_L} - 1) \\ - F_s \sum_{j=1}^{N} \xi_j S_j - F_n \sum_{l=1}^{M} \zeta_l \tau_l, \end{aligned} \tag{165}$$

where $\langle \boldsymbol{jl} \rangle$ is a shorthand for $\langle j_1 \cdots j_K l_1 \cdots l_L \rangle$.

Code constructions are described by the tensor $\mathcal{A}_{\langle \boldsymbol{il} \rangle} \in \{0, 1\}$ that specifies a set of indices $\langle j_1 \cdots j_K l_1 \cdots l_L \rangle$ corresponding to nonzero elements in a particular row of the matrix $[\boldsymbol{C_s} \mid \boldsymbol{C_n}]$. To cope with noninvertible $\boldsymbol{C_n}$ matrices, we can start by considering an ensemble with uniformly generated $M \times M$ matrices. The noninvertible matrices can be made invertible by eliminating a $\epsilon \sim \mathcal{O}(1)$ number of rows and columns, resulting in an ensemble of $(M - \epsilon) \times (M - \epsilon)$ invertible $\boldsymbol{C_n}$ matrices and $(M - \epsilon) \times (N - \epsilon) \boldsymbol{C_s}$ matrices. As we are interested in the thermodynamic limit, we can neglect $\mathcal{O}(1)$ differences and compute the averages in the original space of $M \times M$ matrices. The averages are then performed over an ensemble of codes generated as follows:

(i) Sets of numbers $\{C_j\}_{j=1}^{N}$ and $\{D_l\}_{l=1}^{M}$ are sampled independently from distributions $\mathcal{P}_C$ and $\mathcal{P}_D$, respectively;
(ii) Tensors $\mathcal{A}_{\langle \boldsymbol{jl} \rangle}$ are generated such that

$$\sum_{\langle \boldsymbol{jl} \rangle} \mathcal{A}_{\langle \boldsymbol{jl} \rangle} = M,$$

$$\sum_{\langle j_1 = j \cdots j_K l_1 \cdots l_L \rangle} \mathcal{A}_{\langle \boldsymbol{jl} \rangle} = C_j \qquad \sum_{\langle j_1 \cdots j_K l_1 = l \cdots l_L \rangle} \mathcal{A}_{\langle \boldsymbol{jl} \rangle} = D_l.$$

The free-energy is computed by the replica method as:

$$f = -\frac{1}{\beta} \lim_{N\to\infty} \frac{1}{N} \left.\frac{\partial}{\partial n}\right|_{n=0} \langle \mathcal{Z}^n \rangle_{\mathcal{A},\xi,\zeta} \tag{166}$$

The replicated partition function is:

$$\begin{aligned}\langle \mathcal{Z}^n \rangle_{\mathcal{A},\xi,\zeta} = & \sum_{\boldsymbol{S}^1,\ldots,\boldsymbol{S}^n} \sum_{\boldsymbol{\tau}^1,\ldots,\boldsymbol{\tau}^n} \prod_{j=1}^{N} \left\langle \exp\left(F_s \xi \beta \sum_{\alpha=1}^{n} S_j^\alpha \right) \right\rangle_\xi \\ & \times \prod_{l=1}^{M} \left\langle \exp\left(F_n \zeta \beta \sum_{\alpha=1}^{n} \tau_l^\alpha \right) \right\rangle_\zeta \\ & \times \left\langle \prod_{\langle \boldsymbol{jl}\rangle} \prod_{\alpha=1}^{n} \exp\left[\beta\gamma \mathcal{A}_{\langle \boldsymbol{jl}\rangle} \left(S_{j_1}^\alpha \cdots S_{j_K}^\alpha \tau_{l_1}^\alpha \cdots \tau_{l_L}^\alpha - 1 \right) \right] \right\rangle_{\mathcal{A}} .\end{aligned} \tag{167}$$

The average over constructions $\langle(\cdots)\rangle_{\mathcal{A}}$ is:

$$\begin{aligned}\langle(\cdots)\rangle_{\mathcal{A}} = & \sum_{\{C_j,D_l\}} \prod_{j=1}^{N} \mathcal{P}_C(C_j) \prod_{l=1}^{M} \mathcal{P}_D(D_l) \frac{1}{\mathcal{N}} \delta\left(\sum_{\langle j_1=j,i_2,\ldots,j_K l\rangle} \mathcal{A}_{\langle \boldsymbol{jl}\rangle} - C_j \right) \\ & \times \delta\left(\sum_{\langle \boldsymbol{j} l_1=l,l_2,\ldots,l_K\rangle} \mathcal{A}_{\langle \boldsymbol{jl}\rangle} - D_l \right) (\cdots) \\ = & \sum_{\{C_j,D_l\}} \prod_{j=1}^{N} \mathcal{P}_C(C_j) \prod_{l=1}^{M} \mathcal{P}_D(D_l) \\ & \times \frac{1}{\mathcal{N}} \sum_{\{\mathcal{A}\}} \prod_{j=1}^{N} \left[\oint \frac{dZ_j}{2\pi i} \frac{1}{Z_j^{C_j+1}} Z_j^{\sum_{\langle i_1=j,i_2,\ldots,i_K l\rangle} \mathcal{A}_{\langle j_1=j,\ldots,j_K l\rangle}} \right] \\ & \times \prod_{l=1}^{M} \left[\oint \frac{dY_l}{2\pi i} \frac{1}{Y_l^{D_l+1}} Y_l^{\sum_{\langle j l_1=l,l_2,\ldots,l_L\rangle} \mathcal{A}_{\langle j l_1=l,\ldots,l_L\rangle}} \right] (\cdots),\end{aligned} \tag{168}$$

where the first sum is over profiles $\{C_j, D_l\}$ composed by N numbers drawn independently from $\mathcal{P}_C(C)$ and M numbers drawn from $\mathcal{P}_D(D)$. The second sum is over constructions $\mathcal{A}$ consistent with the profile $\{C_j, D_l\}$.

The signal average $\langle(\cdots)\rangle_\xi$ has the form:

$$\langle(\cdots)\rangle_\xi = \sum_{\xi=-1,+1} (1-p_\xi)\,\delta(\xi-1) + p_\xi\,\delta(\xi+1)\,(\cdots). \tag{169}$$

Similarly, the noise average $\langle(\cdots)\rangle_\zeta$ is:

$$\langle(\cdots)\rangle_\zeta = \sum_{\zeta=-1,+1} (1-p)\,\delta(\zeta-1) + p\,\delta(\zeta+1)\,(\cdots). \tag{170}$$

Along the same steps described for Gallager codes, we compute averages above and introduce auxiliary variables via

$$\int dq_{\alpha_1\cdots\alpha_m}\delta\left(q_{\alpha_1\cdots\alpha_m} - \frac{1}{N}\sum_i^N Z_i S_i^{\alpha_1}\cdots S_i^{\alpha_m}\right) = 1 \tag{171}$$

$$\int dr_{\alpha_1\cdots\alpha_m}\delta\left(r_{\alpha_1\cdots\alpha_m} - \frac{1}{M}\sum_i^M Y_i \tau_i^{\alpha_1}\cdots\tau_i^{\alpha_m}\right) = 1 \tag{172}$$

Using the same types of techniques employed in the case of Gallager codes (see Appendix C.2 for details), we obtain the following expression for the replicated partition function:

$$\begin{aligned}
\langle \mathcal{Z}^n\rangle_{\mathcal{A},\xi,\zeta} = &\prod_{j=1}^N \sum_{C_j} \mathcal{P}_C(C_j) \prod_{l=1}^M \sum_{D_l} \mathcal{P}_D(D_l) \\
&\times \left(\frac{dq_0 d\hat{q}_0}{2\pi i}\right)\left(\prod_{\alpha=1}^n \frac{dq_\alpha d\hat{q}_\alpha}{2\pi i}\right)\cdots\left(\frac{dr_0 d\hat{r}_0}{2\pi i}\right)\left(\prod_{\alpha=1}^n \frac{dr_\alpha d\hat{r}_\alpha}{2\pi i}\right)\cdots \\
&\times \exp\left[\frac{M^L N^K}{K!L!}\sum_{m=0}^n \sum_{\langle\alpha_1\cdots\alpha_m\rangle} \mathcal{T}_m q_{\alpha_1\cdots\alpha_m}^K r_{\alpha_1\cdots\alpha_m}^L\right. \\
&\left. - N\sum_{m=0}^n \sum_{\langle\alpha_1\cdots\alpha_m\rangle} q_{\alpha_1\cdots\alpha_m}\hat{q}_{\alpha_1\cdots\alpha_m} - M\sum_{m=0}^n \sum_{\langle\alpha_1\cdots\alpha_m\rangle} r_{\alpha_1\cdots\alpha_m}\hat{r}_{\alpha_1\cdots\alpha_m}\right] \\
&\times \frac{1}{\mathcal{N}}\prod_{j=1}^N \mathrm{Tr}_{\{S_j^\alpha\}}\left[\left\langle \exp\left[F_s\beta\xi\sum_{\alpha=1}^n S_j^\alpha\right]\right\rangle_\xi\right. \\
&\left.\times \oint \frac{dZ_j}{2\pi i}\frac{\exp\left[Z_j\sum_{m=0}^n\sum_{\langle\alpha_1\cdots\alpha_m\rangle}\hat{q}_{\alpha_1\cdots\alpha_m}S_j^{\alpha_1}\cdots S_j^{\alpha_m}\right]}{Z_j^{C_j+1}}\right] \\
&\times \prod_{l=1}^M \mathrm{Tr}_{\{\tau_l^\alpha\}}\left[\left\langle \exp\left[F_n\beta\zeta\sum_{\alpha=1}^n \tau_l^\alpha\right]\right\rangle_\zeta\right. \\
&\left.\times \oint \frac{dY_l}{2\pi i}\frac{\exp\left[Y_l\sum_{m=0}^n\sum_{\langle\alpha_1\cdots\alpha_m\rangle}\hat{r}_{\alpha_1\cdots\alpha_m}\tau_l^{\alpha_1}\cdots\tau_l^{\alpha_m}\right]}{Y_l^{D_l+1}}\right],
\end{aligned} \tag{173}$$

where $\mathcal{T}_m = \mathrm{e}^{-n\beta\gamma}\cosh^n(\beta\gamma)\tanh^m(\beta\gamma)$. Note that the above expression is an extension of Eq. (120).

The replica symmetry assumption is enforced by using the ansätze:

$$q_{\alpha_1\cdots\alpha_m} = \int dx\, \pi(x) x^m \qquad \hat{q}_{\alpha_1\cdots\alpha_m} = \int d\hat{x}\, \hat{\pi}(\hat{x})\, \hat{x}^m \tag{174}$$

and

$$r_{\alpha_1\cdots\alpha_m} = \int dy\, \phi(y)\, y^m \qquad \hat{r}_{\alpha_1\cdots\alpha_m} = \int d\hat{y}\, \hat{\phi}(\hat{y})\, \hat{y}^m. \tag{175}$$

By plugging the above ansätze, using the limit $\gamma \to \infty$ and standard techniques (see Appendix C.3 for details) the following expression for the free-energy:

$$\begin{aligned} f = \frac{1}{\beta}\mathrm{Extr}_{\{\hat{\pi},\pi,\hat{\phi},\phi\}} \Bigg\{ & \alpha \ln 2 + \overline{C} \int dx\, \pi(x)\, d\hat{x}\, \hat{\pi}(\hat{x}) \ln(1 + x\hat{x}) \\ & + \alpha \overline{D} \int dy\, \phi(y)\, d\hat{y}\, \hat{\phi}(\hat{y}) \ln(1 + y\hat{y}) \\ & - \alpha \int \left[\prod_{j=1}^{K} dx_j\, \pi(x_j)\right] \left[\prod_{l=1}^{L} dy_l\, \phi(y_l)\right] \ln\left(1 + \prod_{j=1}^{K} x_j \prod_{l=1}^{L} y_l\right) \\ & - \sum_C \mathcal{P}_C \int \left[\prod_{j=1}^{C} d\hat{x}_j\, \hat{\pi}(\hat{x}_j)\right] \left\langle \ln\left[\sum_{\sigma=\pm 1} e^{\sigma\xi\beta F_s} \prod_{j=1}^{C}(1 + \sigma\hat{x}_j)\right]\right\rangle_\xi \\ & - \alpha \sum_D \mathcal{P}_D \int \left[\prod_{l=1}^{D} d\hat{y}_l \hat{\phi}(\hat{y}_l)\right] \left\langle \ln\left[\sum_{\sigma=\pm 1} e^{\sigma\zeta\beta F_n} \prod_{l=1}^{D}(1 + \sigma\hat{y}_l)\right]\right\rangle_\zeta \Bigg\}, \end{aligned} \tag{176}$$

where $\overline{C} = \sum_C C\mathcal{P}_C(C)$, $\overline{D} = \sum_D D\mathcal{P}_D(D)$ and $\alpha = M/N = \overline{C}/\overline{K}$.

By performing the extremization above, restricted to the space of normalized functions, we find the following saddle-point equations:

$$\hat{\pi}(\hat{x}) = \int \prod_{j=1}^{K-1} dx_j\, \pi(x_j) \prod_{l=1}^{L} dy_l\, \phi(y_l) \delta\left[\hat{x} - \prod_{j=1}^{K-1} x_j \prod_{l=1}^{L} y_l\right]$$

$$\begin{aligned} \pi(x) = & \frac{1}{\overline{C}} \sum_C C\mathcal{P}_C \int \prod_{l=1}^{C-1} d\hat{x}_l\, \hat{\pi}(\hat{x}_l) \\ & \times \left\langle \delta\left[x - \tanh\left(\beta F_s \xi + \sum_{l=1}^{C-1} \operatorname{atanh} \hat{x}_l\right)\right]\right\rangle_\xi, \end{aligned}$$

$$\hat{\phi}(\hat{y}) = \int \prod_{l=1}^{L-1} dy_l\, \phi(y_l) \prod_{j=1}^{K} dx_j\, \pi(x_j) \delta\left[\hat{y} - \prod_{l=1}^{L-1} y_l \prod_{j=1}^{K} x_j\right]$$

$$\phi(y) = \frac{1}{D} \sum_{D} D\mathcal{P}_D \int \prod_{l=1}^{D-1} d\hat{y}_l \hat{\phi}(\hat{y}_l) \times \left\langle \delta\left[y - \tanh\left(\beta F_n \zeta + \sum_{l=1}^{D-1} \operatorname{atanh} \hat{y}_l\right)\right]\right\rangle_{\zeta}. \tag{177}$$

The typical overlap $\rho = \langle \frac{1}{N} \sum_{j=1}^{N} \xi_j \hat{\xi}_j \rangle_{\mathcal{A},\zeta,\xi}$ between the estimate $\hat{\xi}_j = \operatorname{sgn}(\langle S_j \rangle_{\beta_N})$ and the actual signal ξ_j is given by (see Appendix A.3):

$$\rho = \int dh\, P(h) \operatorname{sgn}(h) \tag{178}$$

$$P(h) = \sum_{C} \mathcal{P}_C(C) \int \prod_{l=1}^{C} d\hat{x}_l\, \hat{\pi}(\hat{x}_l) \times \left\langle \delta\left[h - \tanh\left(\beta F_s \xi + \sum_{l=1}^{C} \operatorname{atanh} \hat{x}_l\right)\right]\right\rangle_{\xi}.$$

The intensive entropy is simply $s = \beta^2 \frac{\partial f}{\partial \beta}$ yielding:

$$s = \beta(u(\beta) - f) \tag{179}$$

$$u = -\sum_{C} \mathcal{P}_C \int \prod_{j=1}^{C} d\hat{x}_j\, \hat{\pi}^*(\hat{x}_j) \left\langle \frac{F_s \xi \sum_{\sigma=\pm 1} \sigma \mathrm{e}^{\sigma \beta F_s \xi} \prod_j (1 + \sigma \hat{x}_j)}{\sum_{\sigma=\pm 1} \mathrm{e}^{\sigma \beta F_s \xi} \prod_j (1 + \sigma \hat{x}_j)} \right\rangle_{\xi} - \alpha \sum_{D} \mathcal{P}_D \int \prod_{j=1}^{D} d\hat{y}_j \hat{\phi}^*(\hat{y}_j) \left\langle \frac{F_n \zeta \sum_{\sigma=\pm 1} \sigma \mathrm{e}^{\sigma \beta F_n \zeta} \prod_j (1 + \sigma \hat{y}_j)}{\sum_{\sigma=\pm 1} \mathrm{e}^{\sigma \beta F_n \zeta} \prod_j (1 + \sigma \hat{y}_j)} \right\rangle_{\zeta}$$

where starred distributions are solutions for (177) and $u(\beta)$ is the internal energy density.

For optimal decoding the temperature must be chosen to be $\beta = 1$ (Nishimori temperature) and the fields are

$$F_s = \frac{1}{2} \ln\left(\frac{1 - p_\xi}{p_\xi}\right) \qquad F_n = \frac{1}{2} \ln\left(\frac{1 - p}{p}\right).$$

D. Probability Propagation Decoding

In Sections III and IV we derived probability propagation equations first by assuming a set of factorization properties and writing a closed set of equations that allowed the iterative computation of the (approximate) marginal posterior, and second by computing local magnetizations on the interior of a Husimi cactus (Bethe approximation). The two methods are equivalent as the factorization properties assumed in the former are encoded in the geometry of the lattice assumed in the latter.

Here we use insights provided in the last sections to build a decoding algorithm for MN codes directly. From the replica symmetric free-energy (176) we can write the following Bethe free-energy:

$$\begin{aligned}\mathcal{F}(\boldsymbol{m},\hat{\boldsymbol{m}}) = \frac{M}{N}\ln 2 &+ \frac{1}{N}\sum_{\mu=1}^{M}\sum_{i\in\mathcal{L}_s(\mu)}\ln\left(1+m^s_{\mu i}\hat{m}^s_{\mu i}\right)\\ &+\frac{1}{N}\sum_{\mu=1}^{M}\sum_{j\in\mathcal{L}_n(\mu)}\ln\left(1+m^n_{\mu j}\hat{m}^n_{\mu j}\right)\\ &-\frac{1}{N}\sum_{\mu=1}^{M}\ln\left(1+\mathcal{J}_\mu\prod_{i\in\mathcal{L}_s(\mu)}m^s_{\mu i}\prod_{j\in\mathcal{L}_n(\mu)}m^n_{\mu j}\right)\\ &-\frac{1}{N}\sum_{i=1}^{N}\ln\left[\sum_{\sigma=\pm}e^{\sigma F_s}\prod_{\mu\in\mathcal{M}_s(i)}\left(1+\sigma\hat{m}^s_{\mu i}\right)\right]\\ &-\frac{1}{N}\sum_{j=1}^{M}\ln\left[\sum_{\sigma=\pm}e^{\sigma F_n}\prod_{\mu\in\mathcal{M}_n(j)}\left(1+\sigma\hat{m}^n_{\mu j}\right)\right].\end{aligned}\tag{180}$$

The variables $m^s_{\mu j}(m^n_{\mu j})$ are cavity effective magnetizations of signal (noise) bits interacting through the coupling μ, obtained by removing one of the C couplings in $\mathcal{M}_s(j)$ $(\mathcal{M}_n(j))$ from the system. The variables $\hat{m}^s_{\mu j}(\hat{m}^n_{\mu j})$ correspond to effective magnetizations of signal (noise) bits due to the coupling μ only.

The decoding solutions are fixed points of the free-energy (181) given by:

$$\frac{\partial\mathcal{F}(\boldsymbol{m},\hat{\boldsymbol{m}})}{\partial m^s_{\mu j}}=0\qquad\frac{\partial\mathcal{F}(\boldsymbol{m},\hat{\boldsymbol{m}})}{\partial\hat{m}^s_{\mu j}}=0\tag{181}$$

$$\frac{\partial\mathcal{F}(\boldsymbol{m},\hat{\boldsymbol{m}})}{\partial m^n_{\mu j}}=0\qquad\frac{\partial\mathcal{F}(\boldsymbol{m},\hat{\boldsymbol{m}})}{\partial\hat{m}^n_{\mu j}}=0\tag{182}$$

The solutions for the above equations are the equations being solved by the probability propagation decoding algorithm:

$$m^s_{\mu l} = \tanh\left[\sum_{\nu\in\mathcal{M}_s(l)\setminus\mu} \operatorname{atanh}\left(\hat{m}^s_{\nu l}\right) + F_s\right] \tag{183}$$

$$\hat{m}^s_{\mu j} = \mathcal{J}_\mu \prod_{i\in\mathcal{L}_s(\mu)\setminus j} m^s_{\mu i} \prod_{l\in\mathcal{L}_n(\mu)} m^n_{\mu l}, \tag{184}$$

$$m^n_{\mu l} = \tanh\left[\sum_{\nu\in\mathcal{M}_n(l)\setminus\mu} \operatorname{atanh}\left(\hat{m}^n_{\nu l}\right) + F_n\right] \tag{185}$$

$$\hat{m}^n_{\mu j} = \mathcal{J}_\mu \prod_{i\in\mathcal{L}_s(\mu)} m^s_{\mu i} \prod_{l\in\mathcal{L}_n(\mu)\setminus j} m^n_{\mu l}. \tag{186}$$

The estimate for the message is $\hat{\xi}_j = \operatorname{sgn}(m^s_j)$, where $m_j{}^s$ is the local magnetization due to all couplings linked to the site j, can be computed as:

$$m^s_j = \tanh\left[\sum_{\nu\in\mathcal{M}_s(j)} \operatorname{atanh}\left(\hat{m}^s_{\nu j}\right) + F_s\right] \tag{187}$$

One possibility for the decoding dynamics is to update Eqs. (183) and (185) until a certain halting criteria is reached, and then computing the estimate for the message using Eq. (187). The initial conditions are set to reflect the prior knowledge about the message $m^s_{\mu j}(0) = 1 - 2p_\xi$ and noise $m^n_{\mu l}(0) = 1 - 2p$.

As the prior information is limited, a polynomial time decoding algorithm (like PP) will work only if the solution is unique or the initial conditions are inside the correct basin of attraction. In this case, the $2(NK + MC)$ Eqs. (181) only need to be iterated an $\mathcal{O}(1)$ number of times to get a successful decoding. On the other hand, when there are many solutions, it is possible to obtain improved decoding in exponential time by choosing random initial conditions and comparing free-energies of the solutions obtained, selecting a global minimum.

Observe that the free-energy described here is not equivalent to the variational mean-field free-energy introduced in MacKay (1995, 1999). Here no essential correlations are disregarded, except those related to the presence of loops.

In the next section, we will analyze the landscape of the replica symmetric free-energy for three families of construction parameters and will be able to predict the practical performance of a PP decoding algorithm.

E. Equilibrium Results and Decoding Performance

The saddle-point Eqs. (177) can be solved by using Monte Carlo integration iteratively. In this section, we show that MN codes can be divided, as far as performance is concerned, into three parameter groups: $K \geq 3$, $K = 2$, and $K = 1, L > 1$.

We, therefore, treat each of these cases separately in the following.

1. *Analytical Solution: The Case of* $K \geq 3$

Replica symmetric results for the cases of $K \geq 3$ can be obtained analytically; therefore, we focus first on this simple case. For unbiased messages ($F_s = 0$), we can easily verify that the ferromagnetic state, characterized by $\rho = 1$, and the probability distributions

$$\begin{aligned} \pi(x) &= \delta(x-1) \\ \hat{\pi}(\hat{x}) &= \delta(\hat{x}-1) \\ \phi(y) &= \delta(y-1) \\ \hat{\phi}(\hat{y}) &= \delta(\hat{y}-1) \end{aligned} \tag{188}$$

and the paramagnetic state of $\rho = 0$ with the probability distributions

$$\begin{aligned} \pi(x) &= \delta(x) \\ \hat{\pi}(\hat{x}) &= \delta(\hat{x}) \\ \hat{\phi}(\hat{y}) &= \delta(\hat{y}) \\ \phi(y) &= \frac{1+\tanh(F_n)}{2}\,\delta(y-\tanh(F_n)) \\ &\quad + \frac{1-\tanh(F_n)}{2}\,\delta(y+\tanh(F_n)), \end{aligned} \tag{189}$$

satisfy replica symmetric saddle-point Eqs. (177). Other solutions could be obtained numerically. To check for that, we represented the distributions with histograms of 20,000 bins and iterated Eqs. (177) 100–500 times with 2×10^5 Monte Carlo sampling steps for each iteration. No solutions other than ferromagnetic and paramagnetic have been observed.

The thermodynamically dominant state is found by evaluating the free-energy of the two solutions using Eq. (176), which yields

$$f_{\text{FERRO}} = -\frac{C}{K} F_n \tanh(F_n), \tag{190}$$

for the ferromagnetic solution and

$$f_{\mathrm{PARA}} = \frac{C}{K} \ln 2 - \ln 2 - \frac{C}{K} \ln \left(2 \cosh(F_n)\right), \tag{191}$$

for the paramagnetic solution.

Figure 24(a) describes schematically the nature of the solutions for this case, in terms of the replica symmetric free-energy and overlap obtained, for various noise levels p and unbiased messages $p_\xi = 1/2$. The coexistence line in the code rate versus noise level plane is given by

$$f_{\mathrm{FERRO}} - f_{\mathrm{PARA}} = \frac{\ln 2}{R_c} \left[R_c - 1 + H_2(p)\right] = 0. \tag{192}$$

This can be rewritten as

$$R_c = 1 - H_2(p) = 1 + p \log_2(p) + (1-p) \log_2(1-p), \tag{193}$$

which coincides with channel capacity and is represented in Figure 25(a) together with the overlap ρ as a function of the noise level p.

Equation (193) seems to indicate that all constructions with $K \geq 3$ may attain error-free data transmission for $R < R_c$ in the limit where both message and codeword lengths N and M become infinite, thus saturating Shannon's bound. However, as described in Fig. 24(a), the paramagnetic state is also stable for any noise level, which has dynamic implications if a replica symmetric free-energy is to be used for decoding (as is the case in probability propagation decoding).

To validate the solutions obtained we have to make sure that the entropy is positive. Entropies can be computed by simply plugging distributions (189) and (190) into Eq. (179). The energy densities for the unbiased case are $u = u_{\mathrm{PARA}} = u_{\mathrm{FERRO}} = -\alpha F_n(1-2p)$, since the Nishimori condition is employed (see Appendix B.3). Ferromagnetic entropies are $s_{\mathrm{FERRO}} = u - f_{\mathrm{FERRO}} = 0$ and

$$\begin{aligned} s_{\mathrm{PARA}} &= u - f_{\mathrm{PARA}} \\ &= -\alpha F_n(1-2p) - \frac{C}{K} \ln 2 + \ln 2 + \frac{C}{K} \ln \left(2 \cosh(F_n)\right). \end{aligned} \tag{194}$$

It can be seen by using a simple argument that s_{PARA} is negative below p_c. For $p < p_c$, $f_{\mathrm{PARA}} > f_{\mathrm{FERRO}}$ and $u - s_{\mathrm{PARA}} > u - s_{\mathrm{FERRO}}$.

This indicates that the distribution (190) is nonphysical below p_c, despite being a solution of replica symmetric saddle-point equations. This result seems to indicate that the replica symmetric free-energy does not provide the right description below p_c. A simple alternative is to use the frozen spins solution as the formulation of a theory with replica symmetry breaking for highly diluted

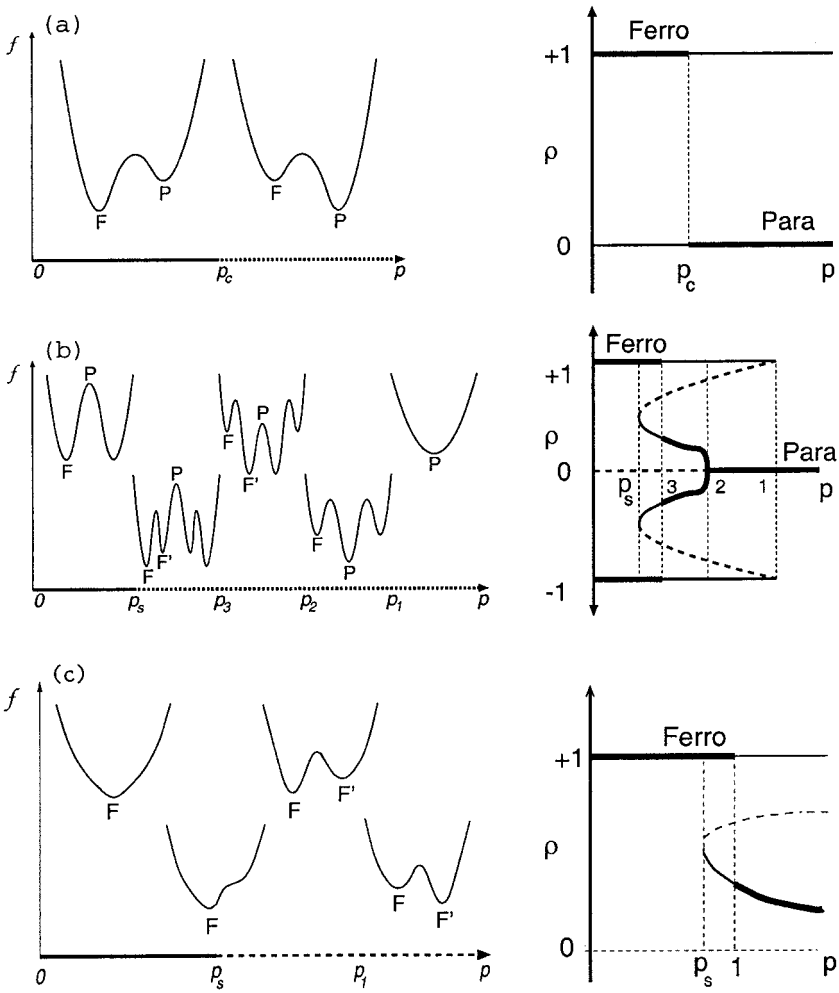

FIGURE 24. Figures on the left side show schematic representations free-energy landscapes while figures on the right show overlaps ρ a function of the noise level p; thick and thin lines denote stable solutions of lower and higher free energies, respectively, dashed lines correspond to unstable solutions. (a) $K \geq 3$—The solid line in the horizontal axis represents the phase where the ferromagnetic solution (F, $\rho = 1$) is thermodynamically dominant. The paramagnetic solution (P, $\rho = 0$) becomes dominant at p_c, which coincides with the channel capacity. (b) $K = 2$—The ferromagnetic solution and its mirror image are the only minima of the free-energy up to p_s (solid line). Above p_s suboptimal ferromagnetic solutions (F′, $\rho < 1$) emerge. The thermodynamic transition occurs at p_3 is below the maximum noise level given by the channel capacity, which implies that these codes do not saturate Shannon's bound even if optimally decoded. (c) $K = 1$—The solid line in the horizontal axis represents the range of noise levels where the ferromagnetic state (F) is the only minimum of the free-energy. The suboptimal ferromagnetic state (F′) appears in the region represented by the dashed line. The dynamic transition is denoted by p_s, where F′ first appears. For higher noise levels, the system becomes bistable and an additional unstable solution for the saddle point equations necessarily appears. The thermodynamic transition occurs at the noise level p_1 where F′ becomes dominant.

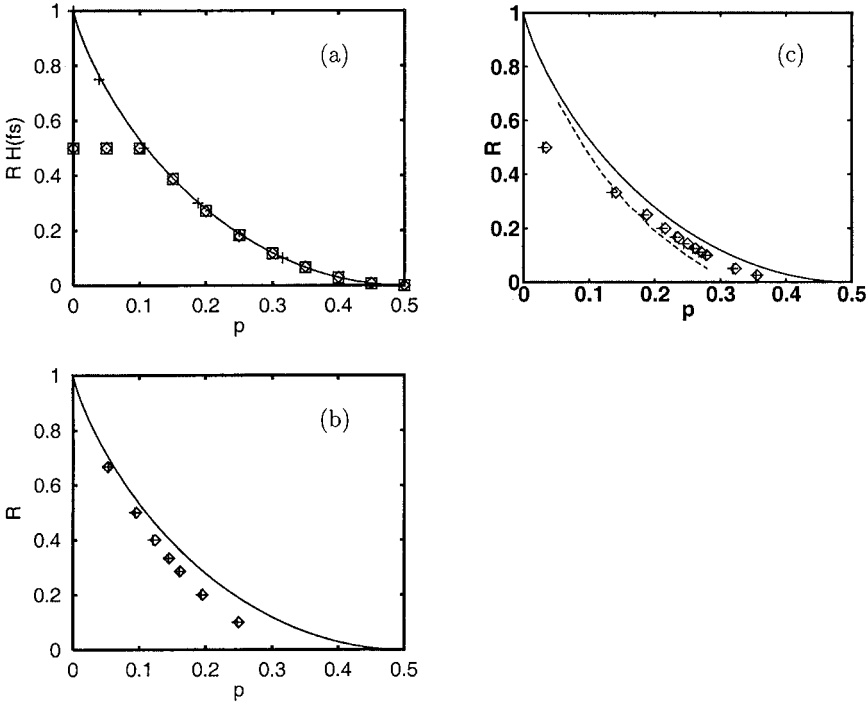

FIGURE 25. Transition lines in the plane rate R versus the flip rate p, obtained from numerical solutions and the TAP approach ($N = 10^4$), and averaged over 10 different initial conditions with error bars much smaller than the symbols size. (a) Numerical solutions for $K = L = 3$, $C = 6$ and varying input bias f_s (□) and TAP solutions for both unbiased (+) and biased (◇) messages; initial conditions were chosen close to the analytical ones. The critical rate is multiplied by the source information content to obtain the maximal information transmission rate, which clearly does not go beyond $R = 3/6$ in the case of biased messages; for unbiased patterns, $H_2(f_s) = 1$. (b) For the unbiased case of $K = L = 2$, initial conditions for the TAP (+) and the numerical solutions (◇) were chosen to be of almost zero magnetization. (c) For the case of $K = 1$, $L = 2$ and unbiased messages. We show numerical solutions of the analytical equations (◇) and those obtained by the TAP approach (+). The dashed line indicates the performance of $K = L = 2$ codes for comparison. Codes with $K = 1$, $L = 2$ outperform $K = L = 2$ for code rates $R < 1/3$.

systems, which is a difficult task (see, e.g., Wong and Sherrington, 1988; Monasson, 1998b).

Nevertheless, the practical performance of the probability propagation decoding is described by the replica symmetric theory, the presence of paramagnetic stable states implies the failure of PP decoding at any noise level. Even without knowing the correct physics below p_c, it is possible to use an

exhaustive search for the global minimum of the free-energy in Section V.D to attain Shannon's bound in exponential time.

2. *The Case of* $K = 2$

All codes with $K \geq 3$ potentially saturate Shannon's bound and are characterized by a first-order phase transition between the ferromagnetic and paramagnetic solutions. Solutions for the case with $K = 2$ can be obtained numerically, yielding significantly different physical behavior as shown in Figure 24(b).

At very large noise levels, the paramagnetic solution (190) gives the unique extremum of the free-energy until the noise level reaches p_1, at which the ferromagnetic solution (189) of higher free-energy becomes locally stable. As the noise level decreases to p_2, the paramagnetic solution becomes unstable and a suboptimal ferromagnetic solution and its mirror image emerge. Those solutions have lower free-energy than the ferromagnetic solution until the noise level reaches p_3. Below p_3, the ferromagnetic solution becomes the global minimum of the free-energy, while the suboptimal ferromagnetic solutions remain locally stable. However, the suboptimal solutions disappear at the spinodal noise level p_s and the ferromagnetic solution (and its mirror image) becomes the unique stable solution of the saddle-point Eqs. (177).

The analysis implies that p_3, the critical noise level below which the ferromagnetic solution becomes thermodynamically dominant, is lower than $p_c = H_2^{-1}(1 - R)$ which corresponds to Shannon's bound. Namely, $K = 2$ does not saturate Shannon's bound in contrast to $K \geq 3$ codes even if decoded in exponential time. Nevertheless, it turns out that the free-energy landscape, with a unique minimum for noise levels $0 < p < p_s$, offers significant advantages in the decoding dynamics compared to that of codes with $K \geq 3$, allowing for the successful use of polynomial time probability propagation decoding.

3. *The Case of* $K = 1$ *and General* $L > 1$

The choice of $K = 1$, independent of the value chosen for $L > 1$, exhibits a different behavior presented schematically in Figure 24(c); also in this case there are no simple analytical solutions and all solutions in this scenario but the ferromagnetic one have been obtained numerically. The first important difference to be noted is that the paramagnetic state (190) is no longer a solution of the saddle-point Eqs. (177) and is being replaced by a suboptimal ferromagnetic state, very much like Gallager codes. Convergence to $\rho = 1$ solution can only be guaranteed for noise levels $p < p_s$, where only the ferromagnetic solution is present.

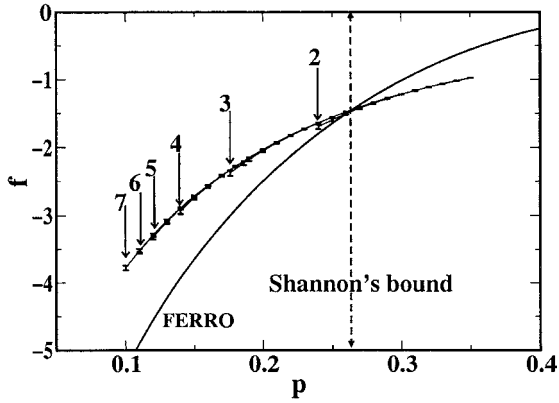

FIGURE 26. Free-energies obtained by solving the analytical equations using Monte Carlo integrations for $K = 1$, $R = 1/6$ and several values of L. Full lines represent the ferromagnetic free-energy (FERRO, higher on the right) and the suboptimal ferromagnetic free-energy (higher on the left) for values of $L = 2, \ldots, 7$. The dashed line indicates Shannon's bound and the arrows represent the spinodal point values p_s for $L = 2, \ldots, 7$. The thermodynamic transition coincides with Shannon's bound.

The $K = 1$ codes do not saturate Shannon's bound in practice; however, we have found that at rates $R < 1/3$ they outperform the $K = L = 2$ code (see Fig. 25) while offering improved decoding times when probability propagation is used. Studying the replica symmetric free-energy in this case shows that as the corruption rate increases, suboptimal ferromagnetic solutions (stable and unstable) emerge at the spinodal point p_s. When the noise increases further, this suboptimal state becomes the global minimum at p_1, dominating the system's thermodynamics. The transition at p_1 must occur at noise levels lower or equal to the value predicted by Shannon's bound.

In Figure 26 we show free-energy values computed for a given code rate and several values of L, denoting Shannon's bound by a dashed line; the thermodynamic transition observed numerically (i.e., the point where the ferromagnetic free-energy equals the suboptimal ferromagnetic free-energy) is closely below Shannon's bound within the numerical precision used. Spinodal noise levels are indicated by arrows. In Figure 27 we show spinodal noise levels as a function of L as predicted by the replica symmetric theory (circles) and obtained by running PP decoding of codes with size 10^4. The optimal parameter choice is $L = 2$.

Due to the simplicity of the saddle-point Eqs. (177) we can deduce the asymptotic behavior of $K = 1$ and $L = 2$ codes for small rates (large C) by computing the two first cumulants of the distributions $\pi, \hat{\pi}, \phi$, and $\hat{\phi}$ (Gaussian approximation). A decoding failure corresponds to $\langle h \rangle \sim \mathcal{O}(1)$ and

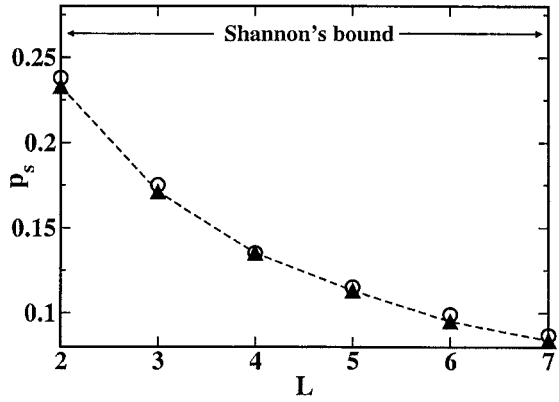

FIGURE 27. Spinodal point noise level p_s for $K = 1$, $R = 1/6$ and several choices of L. Numerical solutions are denoted by circles and PP decoding solutions (10 runs with size $N = 10^4$) by black triangles. Symbols are larger than the error bars.

$\sigma_h^2 \sim \mathcal{O}(1)$. It implies that $\langle \hat{x} \rangle \sim \mathcal{O}(1/C)$ and $\sigma_{\hat{x}} \sim \mathcal{O}(1/C)$. For that, y must be small and we can use atanh(tanh(y_1)tanh(y_2)) $\approx y_1 y_2$ and write:

$$\langle x \rangle \sim \mathcal{O}(1) \qquad \sigma_x^2 \sim \mathcal{O}(1) \tag{195}$$

$$\langle \hat{x} \rangle \approx \langle y \rangle^2 \tag{196}$$

$$\sigma_{\hat{x}}^2 \approx \langle y^2 \rangle^2 - \langle y \rangle^4 \tag{197}$$

$$\langle y \rangle = \langle \hat{y} \rangle + (1-2p)F_n \qquad \sigma_y^2 = \sigma_{\hat{y}}^2 + 4f(1-p)F_n^2 \tag{198}$$

$$\langle \hat{y} \rangle \approx \langle \tanh(x) \rangle \langle y \rangle \tag{199}$$

$$\sigma_{\hat{y}}^2 \approx \langle \tanh^2(x) \rangle \langle y^2 \rangle - \langle \tanh(x) \rangle^2 \langle y \rangle^2 \tag{200}$$

To simplify further we can assume that $p \to 0.5$. Therefore, $F_n \approx (1-2p)$. The critical observation is that in order to have $\langle h \rangle \sim \mathcal{O}(1)$, we need that $\hat{x} \sim \mathcal{O}(1/C)$ and consequently $\langle y \rangle \sim \mathcal{O}(1/\sqrt{C})$. Manipulating the set of equations above:

$$\langle y \rangle \approx \langle \tanh x \rangle \langle y \rangle + (1-2f)^2$$

By imposing the condition over $\langle y \rangle$: $C^{-1/2} \sim (1-2p)^2(1 - \langle \tanh x \rangle)^{-1}$

In terms of the code rate $R = 1/C$:

$$R \sim \frac{(1-2p)^4}{(1 - \langle \tanh x \rangle)^2} \tag{201}$$

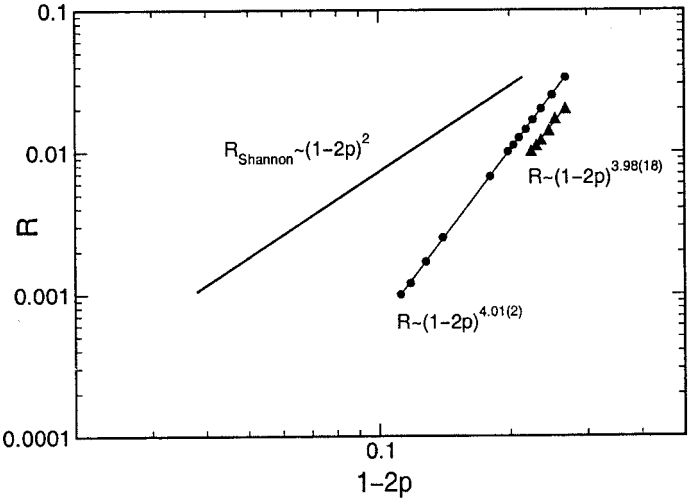

FIGURE 28. Asymptotic behavior of the transition for small rates. The full line represents Shannon's bound, circles represent transitions obtained by using only the first cumulants, and squares correspond to the Gaussian approximation.

The asymptotic behavior of Shannon's bound is given by:

$$R \sim \frac{(1-2p)^2}{\ln 2} \tag{202}$$

Thus, the $K = 1$ and $L = 2$ codes are not optimal asymptotically (large C values). In Figure 28 we verify the relation (201) by iterating first cumulant equations in the delta approximation and first and second cumulant equations in the Gaussian approximation.

F. Error Correction: Regular vs. Irregular Codes

Matrix construction irregularity can improve the practical performance of MN codes. This fact was first reported in the information theory literature (see, e.g., Davey, 1998, 1999; Luby *et al.,* 1998). Here we analyze this problem by using the language and tools of statistical physics. We now use the simplest irregular constructions as an illustration; here, the connectivities of the signal matrix $\boldsymbol{C}_s$ are described by a simple bimodal probability distribution:

$$\mathcal{P}_C(C) = (1-\theta)\,\delta(C - C_o) + \theta\delta(C - C_e). \tag{203}$$

transition. The thermodynamic transition coincides with the upper bound (u.b.) in Section V.A and is very close to, but below, Shannon's limit which is shown for comparison. Similar behavior was observed in regular MN codes with $K = 1$.

G. The Spinodal Noise Level

The PP algorithm can be regarded as an iterative solution of fixed-point equations for the free-energy (181) which is sensitive to the presence of local minima in the system. One can expect convergence to the global minimum of the free-energy from all initial conditions when there is a single minimum or when the landscape is dominated by the basin of attraction of this minimum when random initial conditions are used.

To analyze this point, we run decoding experiments starting from initial conditions $m^s_{\mu j}(0)$ and $m^n_{\mu l}(0)$ that are random perturbations of the ferromagnetic solution drawn from the following distributions:

$$P\left(m^s_{\mu j}(0)\right) = (1-\lambda_s)\,\delta\left(m^s_{\mu j}(0) - \xi_j\right) + \lambda_s\,\delta\left(m^s_{\mu j}(0) + \xi_j\right) \tag{204}$$

and

$$P\left(m^n_{\mu l}(0)\right) = (1-\lambda_n)\,\delta\left(m^n_{\mu l}(0) - \tau_l\right) + \lambda_n\,\delta\left(m^n_{\mu l}(0) + \tau_l\right), \tag{205}$$

where for convenience we choose $0 \le \lambda_s = \lambda_n = \lambda \le 0.5$.

We performed PP decoding several times for different values of λ and noise level p. For $\lambda \le 0.026$, we observed that the system converges to the ferromagnetic state for *all* constructions, message biases p_ξ, and noise levels p examined. It implies that this state is always stable. The convergence occurs for any λ for noise levels below the transition observed in practice.

These observations suggest that the ferromagnetic basin of attraction dominates the landscape up to some noise level p_s. The fact that no other solution is ever observed in this region suggests that p_s is the noise level where suboptimal solutions actually appear, namely, it is the noise level that corresponds to the appearance of spinodal points in the free-energy. The same was observed for regular MN codes with $K = 1$ or $K = 2$.

We have shown that MN codes can be divided into three categories with different equilibrium properties: (i) $K \ge 3$, (ii) $K = 2$, and (iii) general $L > 1$, $K = 1$. In the next two subsections we will discuss these cases separately.

1. *Biased Messages:* $K \ge 3$

To show how irregularity affects codes with this choice of parameters, we choose K, $L = 3$, $C_o = 4$, $C_e = 30$ and biased messages with $p_\xi = 0.3$. These

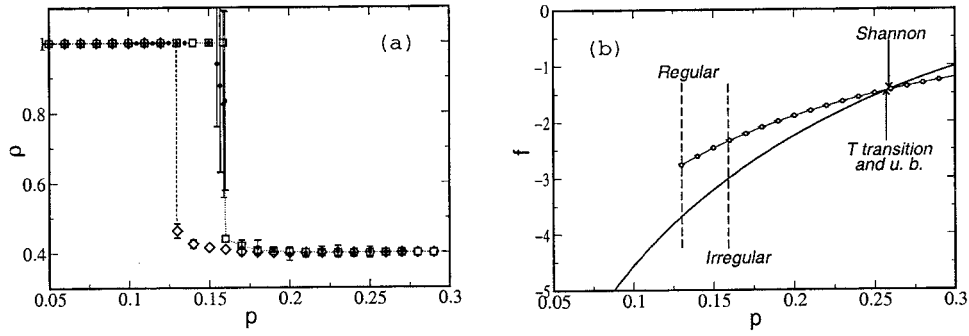

FIGURE 29. (a) Overlap as a function of the noise level p for codes with $K = L = 3$ and $\overline{C} = 15$ with message bias $p_\xi = 0.3$. Analytical RS solutions for the regular code are denoted as $\diamond$ and for the irregular code; with $C_o = 4$ and $C_e = 30$ denoted as $\square$. Results are averages over 10 runs of the PP algorithm in an irregular code of size $N = 6000$ starting from fixed initial conditions (see the text); they are plotted as $\bullet$ in the rightmost curve for comparison. PP results for the regular case agree with the theoretical solutions and have been omitted to avoid overloading the figure. (b) Free-energies for the ferromagnetic state (full line) and for the failure state (line with $\circ$). The transitions observed in (a) are indicated by the dashed lines. Arrows indicate the thermodynamic (T) transition, the upper bound (u.b.) of Section V.A, and Shannon's bound.

The mean connectivity is $\overline{C} = (1-\theta)C_o + \theta C_e$ and $C_o < \overline{C} < C_e$; bits in a group with connectivity C_o will be referred as *ordinary* bits and bits in a group with connectivity C_e as *elite* bits. The noise matrix $\boldsymbol{C_n}$ is chosen to be regular.

To gain some insight into the effect of irregularity on solving the PP Eqs. (183) and (185), we performed several runs starting from the fixed initial conditions $m^s_{\mu j}(0) = 1 - 2p_\xi$ and $m^n_{\mu l}(0) = 1 - 2p$ as prescribed in the last section. For comparison, we also iterated the saddle-point Eqs. (177) obtained by the replica symmetric (RS) analysis, setting the initial conditions to be $\pi_0(x) = (1-p_\xi)\,\delta(x - m^s_{\mu j}(0)) + p_\xi\,\delta(x + m^s_{\mu j}(0))$ and $\rho_0(y) = (1-p)\,\delta(y - m^n_{\mu l}(0)) + p\,\delta(y + m^n_{\mu l}(0))$, as suggested from the interpretation of the fields $\pi(x)$ and $\rho(y)$ in the last section.

In Figure 29(a) we show a typical curve for the overlap ρ as a function of the noise level p. The RS theory agrees very well with PP decoding results. The addition of irregularity improves the performance considerably. In Figure 29(b) we show the free-energies of the two emerging states. The free-energy for the ferromagnetic state with overlap $\rho = 1$ is shown as a full line; the failure suboptimal ferromagnetic state (in Fig. 29(a) with overlap $\rho = 0.4$) is shown as a line marked with $\circ$. The transitions seen in Fig. 29(a) are denoted by dashed lines. It is clear that they are far below the thermodynamic (T) transition, indicating that the system becomes trapped in suboptimal ferromagnetic states for noise levels p between the observed transitions and the thermodynamic

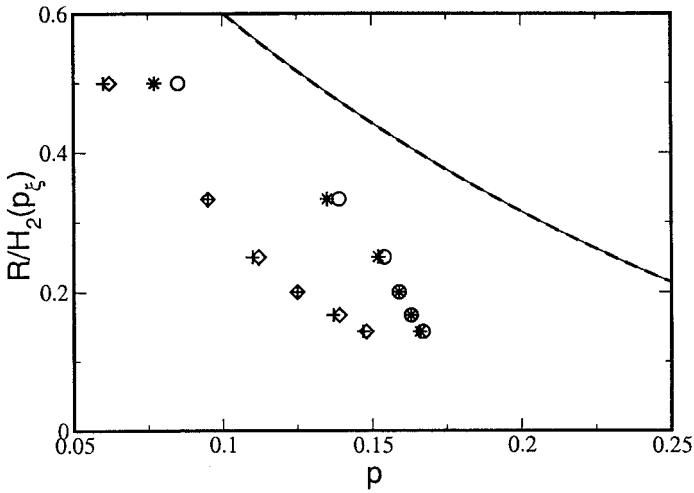

FIGURE 30. *Spinodal* noise level p_s for regular and irregular codes. In both constructions, parameters are set as $K = L = 3$. Irregular codes with $C_o = 4$ and $C_e = 30$ are used. PP decoding is carried out with $N = 5000$ and a maximum of 500 iterations; they are denoted by $+$ (regular) and $*$ (irregular). Numerical solutions for the RS saddle-point equations are denoted by $\diamond$ (regular) and $\circ$ (irregular). Shannon's limit is represented by a full line and the upper bound of Section V.A. is represented by a dashed line. The symbols are chosen to be larger than the actual error bars.

choices are arbitrary but illustrate what happens with the practical decoding performance. In Figure 30 we show the transition from the decoding phase to a failure phase as a function of the noise level p for several rates R in both regular and irregular codes. Practical decoding ($\diamond$ and $\circ$) results are obtained for systems of size $N = 5000$ with a maximum number of iterations set to 500. Random initial conditions are chosen and the whole process repeated 20 times. The practical transition point is found when the number of failures equals the number of successes.

These experiments were compared with the theoretical values for p_s obtained by solving the RS saddle-point Eqs. (177) (represented as $+$ and $*$ in Fig. 30) and finding the noise level for which a second solution appears. For comparison the coding limit is represented in the same figure by a full line.

As the constructions used are chosen arbitrarily, one can expect that these transitions can be further improved, even though the improvement shown in Figuer 30 is already fairly significant.

The analytical solution obtained for $K \geq 3$ and unbiased messages $p_\xi = 1/2$ implies that the system is bistable for arbitrary code constructions when these parameters are chosen. The spinodal noise level is then $p_s = 0$ in this case and cannot be improved by adding irregularity to the construction. Up to the noise

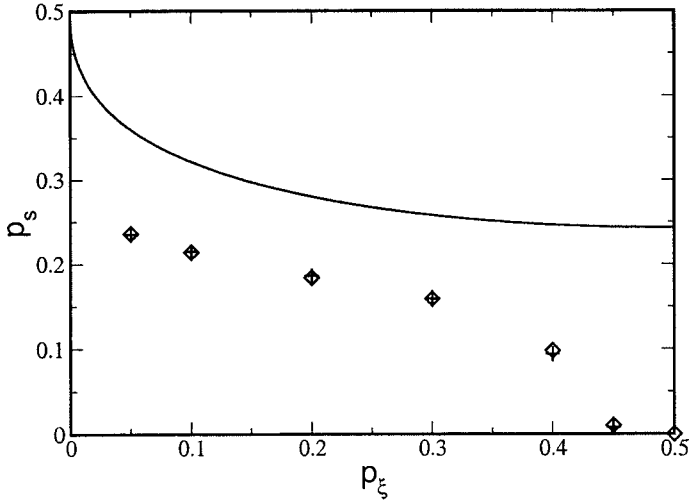

FIGURE 31. Spinodal noise level p_s for irregular codes as a function of the message bias p_ξ. The construction is parameterized by $K = L = 3$, $C_o = 4$, and $C_e = 30$ with $\overline{C} = 15$. PP decoding is carried out with $N = 5000$ and a maximum of 500 iterations, and is represented by $+$, while theoretical RS solutions are represented by $\diamond$. The full line indicates Shannon's limit. Symbols are larger than the actual error bars

level p_c, the ferromagnetic solution is the global minimum of the free-energy, and therefore Shannon's limit is achievable in exponential time; however, the bistability makes these constructions unsuitable for practical decoding with a PP algorithm when unbiased messages are considered.

The situation improves when biased messages are used. Fixing the matrices $\boldsymbol{C_n}$ and $\boldsymbol{C_s}$, one can determine how the spinodal noise level p_s depends on the bias p_ξ. In Figure 31 we compare simulation results with the theoretical predictions of p_s as a function of p_ξ. The spinodal noise level p_s collapses to zero as p_ξ increases toward the unbiased case. It obviously suggests using biased messages for practical MN codes with parameters $K \geq 3$ and PP decoding.

The qualitative pictures of the energy landscape for coding with biased and unbiased messages with $K \geq 3$ differ significatively. In Figure 32 this landscape is sketched as a function of the noise level p for a given bias. Up to the spinodal noise level p_s, the landscape is totally dominated by the ferromagnetic state F. At the spinodal noise level, another suboptimal state F' emerges, dominating the decoding dynamics. At p_c, the suboptimal state F' becomes the global minimum. The bold horizontal line represents the region where the ferromagnetic solution with $\rho = 1$ dominates the decoding dynamics. In the

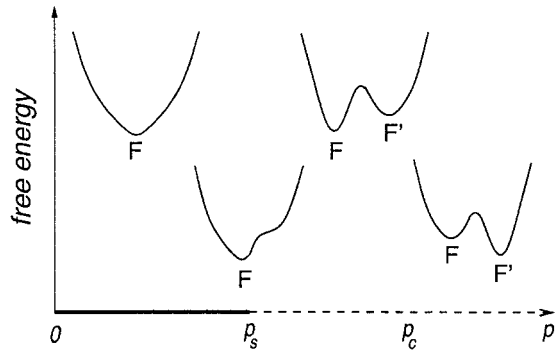

FIGURE 32. Pictorial representation of the free-energy landscape for codes with $K \geq 3$ and biased messages $p_\xi < 0.5$ as a function of the noise level p. Up to the spinodal noise level p_s, there is only the ferromagnetic state F. At p_s, another state F' appears, dominating the decoding dynamics. The critical noise level p_c indicates the point where the state F' becomes the global minimum (thermodynamic transition).

region represented by the dashed line, decoding dynamics is dominated by suboptimal ferromagnetic $\rho < 1$ solutions.

2. *Unbiased Messages*

For the remaining parameter choices, namely general $L > 1$, $K = 1$, and $K = 2$, it was shown that unbiased coding is generally possible, yielding close to Shannon's limit performance.

In the $K \geq 3$ case, the practical performance is defined by the spinodal noise level p_s and the addition of irregularity modifies p_s.

In the general L, $K = 1$ family we illustrate the effect of irregularity by the choice of $L = 2$, $C_o = 4$, and $C_e = 10$. In Figure 33 we show the transitions observed by performing 20 decoding experiments with messages of length $N = 5000$ and a maximal number of iterations set to 500 ($+$ for regular and $*$ for irregular). We compare the experimental results with theoretical predictions based on the RS saddle-point Eqs. (177) ($\diamond$ for regular and $\circ$ for irregular). Shannon's limit is represented by a full line. The improvement is modest, as expected, since regular codes already present close-to-optimal performance. Discrepancies between the theoretical and numerical results are due to finite size effects.

We also performed a set of experiments using $K = L = 2$ with $C_o = 3$ and $C_e = 8$, the same system size $N = 5000$ and maximal number of decoding iterations 500. The transitions obtained experimentally and predicted by theory are shown in Figure 34.

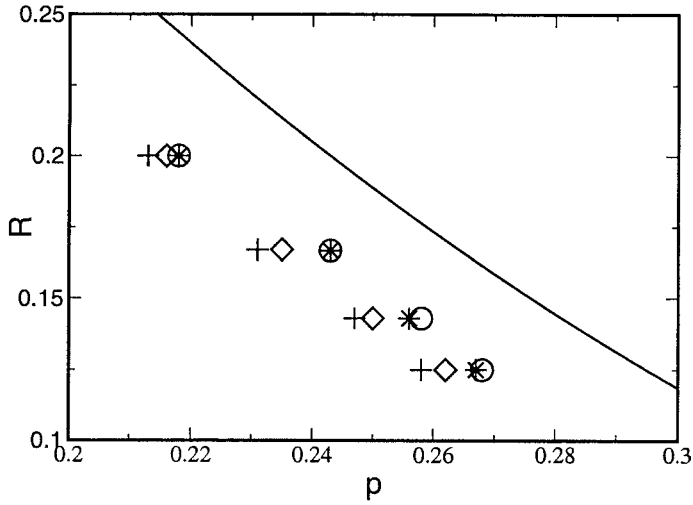

FIGURE 33. Spinodal noise level p_s for regular and irregular codes. The constructions are of $K = 1$ and $L = 2$, irregular codes are parameterized by $C_o = 4$ and $C_e = 10$. PP decoding is carried out with $N = 5000$ and a maximum of 500 iterations; they are denoted by $+$ (regular) and $*$ (irregular). Numerical solutions for RS equations are denoted by $\diamond$ (regular) and $\bigcirc$ (irregular). The coding limit is represented by a line. Symbols are larger than the actual error bars.

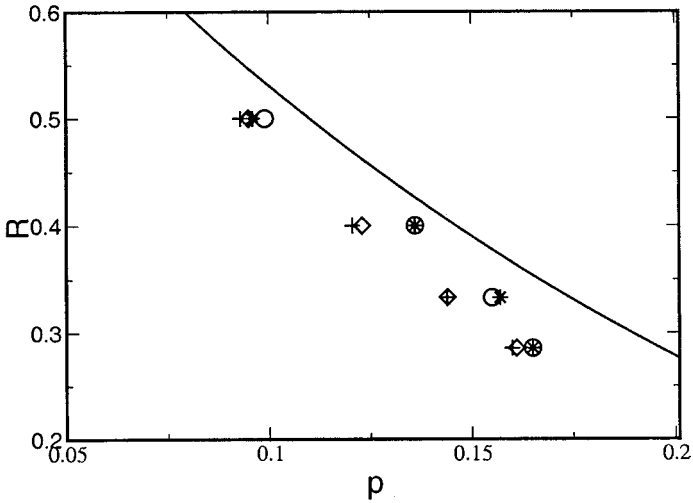

FIGURE 34. Spinodal noise level values p_s for regular and irregular codes. Constructions are of $K = 2$ and $L = 2$, irregular codes are parameterized by $C_o = 3$ and $C_e = 8$. PP decoding is carried out with $N = 5000$ and a maximum of 500 iterations; they are denoted by $+$ (regular) and $*$ (irregular). Theoretical predictions are denoted by $\diamond$ (regular) and $\bigcirc$ (irregular). The coding limit is represented by a line. Symbols are larger than the actual error bars.

$\{\alpha_j\}$ completely specify the construction. A further constraint to the parameters set $\{\alpha_j\}$ is provided by the choice of a code rate, as the inverse code rate is $\alpha = M/N = \sum_{j=1}^{m} \alpha_j$.

Encoding and decoding using cascading codes are performed in exactly the same fashion as described in Section V for MN codes. A binary vector $\boldsymbol{t} \in \{0, 1\}^M$ defined by

$$\boldsymbol{t} = \boldsymbol{G}\boldsymbol{\xi} \text{ (mod 2)}, \tag{206}$$

is produced, where all operations are performed in the field $\{0, 1\}$ and are indicated by (mod 2). The code rate is $R = N/M$. The generator matrix $\boldsymbol{G}$ is a $M \times N$ dense matrix defined by

$$\boldsymbol{G} = \boldsymbol{C}_n^{-1}\boldsymbol{C}_s \text{ (mod 2)}. \tag{207}$$

The transmitted vector $\boldsymbol{\tau}$ is then corrupted by noise. Assuming a memoryless BSC, noise is represented by a binary vector $\boldsymbol{\zeta} \in \{0, 1\}^M$ with components independently drawn from the distribution $P(\zeta) = (1 - p)\,\delta(\zeta) + p\,\delta(\zeta - 1)$.

The received vector is

$$\boldsymbol{r} = \boldsymbol{G}\boldsymbol{\xi} + \boldsymbol{\zeta} \text{ (mod 2)}. \tag{208}$$

Decoding is performed by computing the syndrome vector

$$z = \boldsymbol{C}_n\boldsymbol{r} = \boldsymbol{C}_s\boldsymbol{\xi} + \boldsymbol{C}_n\boldsymbol{\zeta} \text{ (mod 2)}, \tag{209}$$

from which an estimate $\hat{\boldsymbol{\xi}}$ for the message can be obtained.

A. Typical PP Decoding and Saddle-Point-Like Equations

In this section, we show how a statistical description for the typical PP decoding can be constructed without using replica calculations. To keep the analysis as simple as possible, we exemplify the procedure with a KS code with two signal matrices denoted **1s** and **2s** and two noise submatrices denoted **1n** and **2n**. The channel is chosen to be a memoryless BSC. The number of nonzero elements per row is K_1 and K_2, respectively, and the inverse rate is $\alpha = \alpha_1 + \alpha_2$. Therefore, for a fixed code rate, the code construction is specified by a single parameter α_1. We present one code in this family in Figure 37.

The PP decoding dynamics for these codes is described by Eqs. (185). However, due to the irregular character of the construction, sites inside each one of the submatrices are connected differently. Remembering the statistical physics formulation of MN codes presented in Section V.B, nonzero row elements in the matrices depicted in Figure 37 correspond to sites taking part in one multispin interaction. Therefore, signal sites in the submatrix **1s** interact with other

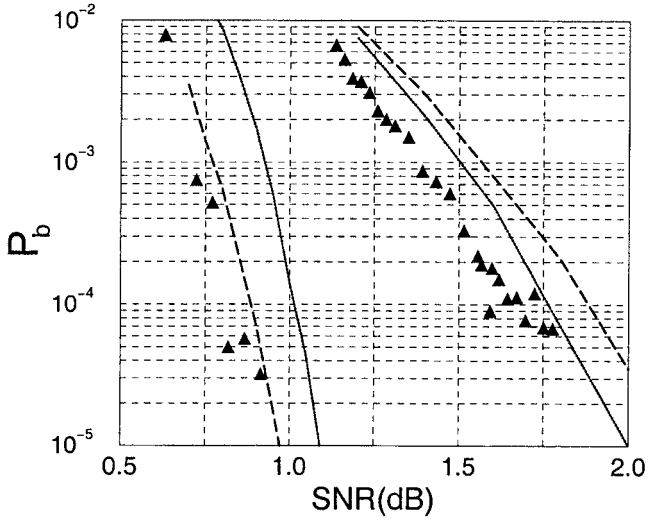

FIGURE 35. Bit error probability p_b as a function of the signal to noise ratio for codes of rate $R = 1/2$, sizes $N = 1000$ (right) and $N = 10000$ (left) in a memoryless Gaussian channel. Black triangles represent cascading codes, dashed lines represent Turbo codes and dotted lines represent optimized irregular Gallager codes of similar sizes (Kanter and Saad, 2000b).

VI. CASCADING CODES

Kanter and Saad (KS) recently proposed a variation of MN codes that has been shown to be capable of attaining close-to-channel capacity performance and outperforming Turbo codes (Kanter and Saad, 1999, 2000a,b). The central idea is to explore the superior dynamic properties (i.e., large basin of attraction) of MN codes with $K = 1, 2$ and the potential for attaining channel capacity of MN codes with $K > 2$ by introducing constructions with intermediate properties. This is done by employing irregular constructions like the one depicted in Figure 35, with the number of nonzero elements per row set to several different values $K_1, \ldots, K_m$.

In Figure 35 we show a performance comparison (presented in (Kanter and Saad, 2000b) of Turbo, KS, and Gallager codes with optimized irregular constructions (Richardson *et al.,* 2001) for a memoryless Gaussian channel. The bit error probability p_b is plotted against the signal to noise ratio in decibels ($10 \log_{10} (S/N)$) for codes of sizes $N = 1000$ and $N = 10000$.

The introduction of multispin interactions of several different orders and of more structured matrices makes the statistical physics of the problem much harder to solve. We, therefore, adopt a different approach: first we write the probability propagation equations and find an appropriate macroscopic

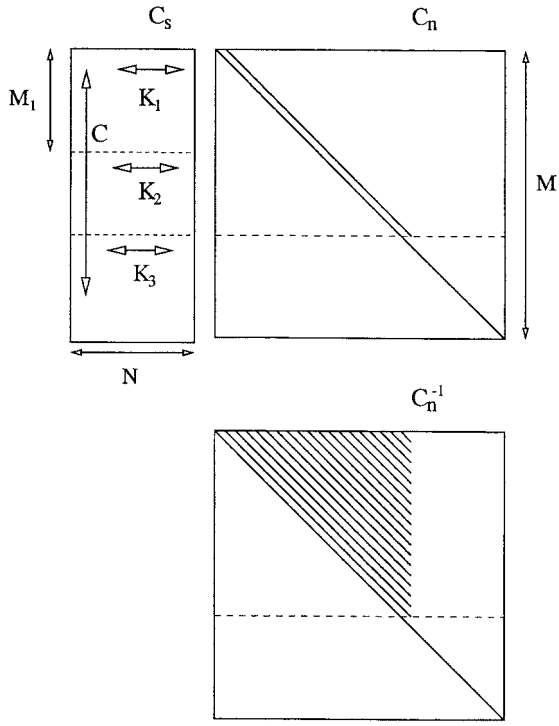

FIGURE 36. Cascading construction with three signal submatrices with K_1, K_2 and K_3 nonzero elements per row, respectively. The number of nonzero elements per column is kept fixed to C. The noise matrix $\boldsymbol{C_n}$ is composed by two submatrices, the nonzero elements are denoted by lines. The inverse $\boldsymbol{C_n^{-1}}$ is also represented.

description in terms of field distributions, we then solve saddle-point-like equations for the field distributions to find the typical performance.

Cascading codes are specific constructions of MN codes. The signal matrix $\boldsymbol{C_s}$ is defined by m random submatrices with $K_1, K_2, \ldots, K_m$ nonzero elements per row, respectively. The matrix $\boldsymbol{C_n}$ is composed of two submatrices: $\boldsymbol{C}^{(1)}_{\boldsymbol{n}_{ij}} = \delta_{i,j} + \delta_{i,j+\Delta}$ and $\boldsymbol{C}^{(2)}_{\boldsymbol{n}_{ij}} = \delta_{i,j}$. The inverse $\boldsymbol{C_n^{-1}}$ used in the encoding process is easily obtainable. In Figure 36 we represent a KS code with three signal submatrices, the nonzero elements in the noise matrix $\boldsymbol{C_n}$ are denoted by lines, we also represent the inverse of the noise matrix $\boldsymbol{C_n^{-1}}$.

The signal matrix $\boldsymbol{C_s}$ is subdivided into $M_j \times N$ submatrices, with $j = 1, \ldots, m$. The total number of nonzero elements is given by $NC = \sum_{j=1}^{m} M_j K_j$ what yields $C = \sum_{j=1}^{m} \alpha_j K_j$, where $\alpha_j = M_j/N$. The code construction is, therefore, parameterized by the set $\{(\alpha_j, K_j)\}$. If we fix $\{K_j\}$, the parameters

For the submatrix **1s** we have:

$$m_{\mu j}^{(1s)} = \tanh\left[\sum_{\nu\in\mathcal{M}_{1s}(j)\backslash\mu} \operatorname{atanh}\left(\hat{m}_{\nu j}^{(1s)}\right) + \sum_{\nu\in\mathcal{M}_{2s}(j)} \operatorname{atanh}\left(\hat{m}_{\nu j}^{(2s)}\right) + F_s\right]$$

$$\hat{m}_{\mu j}^{(1s)} = \mathcal{J}_\mu m_{\mu\mu}^{(1n)} m_{\mu\mu+\Delta}^{(1n)} \prod_{l\in\mathcal{L}_{1s}(\mu)\backslash j} m_{\mu l}^{(1s)}, \tag{211}$$

where the second equation represents interactions with two noise sites and and $K_1 - 1$ signal sites. The first equation represents the $\alpha_1 K_1 + \alpha_2 K_2 - 1$ multispin interactions the site j participates in.

Similarly, for the submatrix **2s** we have:

$$m_{\mu j}^{(2s)} = \tanh\left[\sum_{\nu\in\mathcal{M}_{1s}(j)} \operatorname{atanh}\left(\hat{m}_{\nu j}^{(1s)}\right) + \sum_{\nu\in\mathcal{M}_{2s}(j)\backslash\nu} \operatorname{atanh}\left(\hat{m}_{\nu j}^{(2s)}\right) + F_s\right]$$

$$\hat{m}_{\mu j}^{(2s)} = \mathcal{J}_\mu m_{\mu}^{(2n)} \prod_{l\in\mathcal{L}_{2s}(\mu)\backslash j} m_{\mu l}^{(2s)} \tag{212}$$

For the submatrix **1n** we have:

$$m_{\mu j}^{(1n)} = \tanh\left[\operatorname{atanh}\left(\hat{m}_{\nu j}^{(1n)}\right) + F_n\right] \tag{213}$$

$$\hat{m}_{\mu j}^{(1n)} = \mathcal{J}_\mu m_{\mu i}^{(1n)} \prod_{l\in\mathcal{L}_{1s}(\mu)} m_{\mu l}^{(1s)}, \tag{214}$$

where either $j = \mu, i = \mu + \Delta$ or $j = \mu + \Delta, i = \mu$.

Finally, for submatrix **2n** we have:

$$m_{\mu}^{(2n)} = \tanh\left[F_n\right] \tag{215}$$

$$\hat{m}_{\mu}^{(2n)} = \mathcal{J}_\mu \prod_{l\in\mathcal{L}_{2s}(\mu)} m_{\mu l}^{(2s)} \tag{216}$$

The pseudo-posterior and decoded message are given by:

$$m_j = \tanh\left[\sum_{\nu\in\mathcal{M}_{1s}(j)} \operatorname{atanh}\left(\hat{m}_{\nu j}^{(1s)}\right) + \sum_{\nu\in\mathcal{M}_{2s}(j)} \operatorname{atanh}\left(\hat{m}_{\nu j}^{(2s)}\right)\right] \tag{217}$$

$$\hat{\xi}_j = \operatorname{sgn}(m_j). \tag{218}$$

The above equations provide a microscopic description for the PP decoding process. We can produce a macroscopic description for the typical decoding process by writing equations for probability distributions related to the dynamic variables. It is important to stress that the equations describing the PP decoding are entirely deterministic when couplings $\mathcal{J}_\mu$ and initial conditions are given. The randomness comes into the problem when quenched averages over messages, noise, and constructions are introduced.

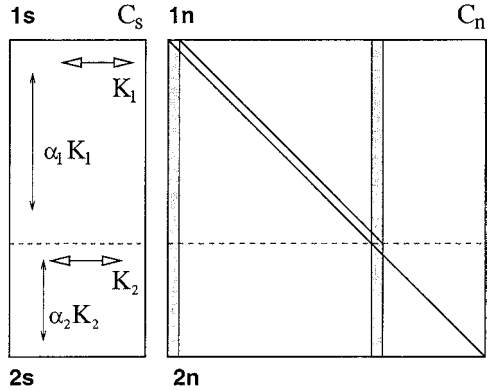

FIGURE 37. Cascading code with two signal matrices with parameters K_1 and K_2. Note that noise sites inside the shaded regions take part in a different number of interactions than the ordinary sites.

$K_1 - 1$ signal sites in **1s** and exactly two noise sites in **1n**. Moreover, the same site takes part in other $\alpha_1 K_1 + \alpha_2 K_2 - 1$ multispin couplings in both **1s** and **2s**. Sites in submatrix **2s** interact with one noise site in **2n** and $K_2 - 1$ signal sites in **2s**, taking part in other $\alpha_1 K_1 + \alpha_2 K_2 - 1$ multispin interaction. Noise sites in the submatrix **1n** interact with another noise site and with K_1 signal sites in **1s**. Finally, noise sites in **2n** interact with K_2 sites in **2s**. Thus, the Hamiltonian for a KS code takes the following form:

$$\mathcal{H} = -\gamma \sum_{\mu=1}^{M_1} (\mathcal{J}_\mu S_{i_1} \cdots S_{i_{K_1}} \tau_\mu \tau_{\mu+\Delta} - 1)$$
$$-\gamma \sum_{\mu=M_1+1}^{M} (\mathcal{J}_\mu S_{i_1} \cdots S_{i_{K_2}} \tau_\mu - 1) - F_n \sum_{l=1}^{M} \tau_l - F_s \sum_{j=1}^{N} S_j, \quad (210)$$

where $\mathcal{J}_\mu = \xi_{i_1} \cdots \xi_{i_{K_1}} \zeta_\mu \zeta_{\mu+\Delta}$, for $\mu = 1, \ldots, M_1$ and $\mathcal{J}_\mu = \xi_{i_1} \cdots \xi_{i_{K_2}} \zeta_\mu$ for $\mu = M_1 + 1, \ldots, M$. Additionally, Nishimori's condition requires that $\gamma \to \infty$, $F_s = \operatorname{atanh}(1 - 2p_\xi)$ and $F_n = \operatorname{atanh}(1 - 2p)$, where the prior probabilities are defined as in the previous chapters.

We can write PP decoding equations for each one of the submatrices **1s**, **2s**, **1n** and **2n**. The shaded regions in Figure 37 have to be described by different equations, but can be disregard if the width Δ is of $\mathcal{O}(1)$, implying $\Delta/N \to 0$ for $N \to \infty$.

By performing the gauge transformation

$$m_{\mu j}^{(as)} \to \xi_j m_{\mu j}^{(as)} \qquad \hat{m}_{vj}^{(as)} \to \xi_j \hat{m}_{\mu j}^{(as)} \tag{219}$$

$$m_{\mu j}^{(an)} \to \zeta_j m_{\mu j}^{(an)} \qquad \hat{m}_{\mu j}^{(an)} \to \zeta_j \hat{m}_{\mu j}^{(an)} \tag{220}$$

$$\mathcal{J}_\mu \to 1 \qquad (a = 1, 2), \tag{221}$$

introducing effective fields $x_{\mu j} = \text{atanh}(m_{\mu j})$, $\hat{x}_{\mu j} = \text{atanh}(\hat{m}_{\mu j})$ and assuming that $x_{\mu j}^{(as)}$, $\hat{x}_{\mu j}^{(as)}$, $y_{\mu j}^{(an)}$, $\hat{y}_{\mu j}^{(an)}$ are independently drawn from distributions $P_a(x)$, $\hat{P}_a(\hat{x})$, $R_a(y)$, $\hat{R}_a(\hat{y})$, respectively, we get the following saddle-point-like equations (for simplicity, we restrict the treatment to the case of unbiased messages $F_s = 0$).

For the submatrix **1s**:

$$P_1(x) = \int \prod_{j=1}^{\alpha_1 K_1 - 1} d\hat{x}_j \hat{P}_1(\hat{x}_j) \prod_{l=1}^{\alpha_1 K_2} d\hat{w}_l \hat{P}_2(\hat{w}_l) \times \delta \left[x - \sum_{j=1}^{\alpha_1 K_1 - 1} x_j - \sum_{l=1}^{\alpha_2 K_2} w_l \right] \tag{222}$$

$$\hat{P}_1(\hat{x}) = \int \prod_{j=1}^{K_1 - 1} dx_j P_1(x_j)\, dy_1\, R_1(y_1)\, dy_2 R_1(y_2) \times \delta \left[\hat{x} - \text{atanh} \left(\tanh(y_1)\tanh(y_2) \prod_{j=1}^{K_1 - 1} \tanh(x_j) \right) \right] \tag{223}$$

For **2s**:

$$P_2(x) = \int \prod_{j=1}^{\alpha_1 K_1} d\hat{x}_j \hat{P}_1(\hat{x}_j) \prod_{l=1}^{\alpha_1 K_2 - 1} d\hat{w}_l \hat{P}_2(\hat{w}_l) \times \delta \left[x - \sum_{j=1}^{\alpha_1 K_1} x_j - \sum_{l=1}^{\alpha_2 K_2 - 1} w_l \right] \tag{224}$$

$$\hat{P}_2(\hat{x}) = \int \prod_{j=1}^{K_2 - 1} dx_j P_2(x_j)\, dy R_2(y) \times \delta \left[\hat{x} - \text{atanh} \left(\tanh(y) \prod_{j=1}^{K_2 - 1} \tanh(x_j) \right) \right] \tag{225}$$

For **1n** we have:

$$R_1(y) = \int d\hat{y}\, \hat{R}_1(\hat{y})\, \langle \delta\,[y - \hat{y} - \zeta F_n] \rangle_\zeta$$

$$\hat{R}_1(\hat{y}) = \int \prod_{j=1}^{K_1} dx_j P_1(x_j)\, dy R_1(y) \times \delta \left[\hat{x} - \operatorname{atanh} \left(\tanh(y) \prod_{j=1}^{K_1} \tanh(x_j) \right) \right] \tag{226}$$

Finally, for submatrix **2n**:

$$R_2(y) = \langle \delta\,[y - \zeta F_n] \rangle_\zeta \tag{227}$$

$$\hat{R}_2(\hat{y}) = \int \prod_{j=1}^{K_2} dx_j P_2(x_j) \delta \left[\hat{x} - \operatorname{atanh} \left(\prod_{j=1}^{K_2} \tanh(x_j) \right) \right]$$

The typical overlap can then be obtained as in the case of MN codes by computing:

$$\rho = \int dh\, P(h)\, \operatorname{sgn}(h) \tag{228}$$

$$P(h) = \int \prod_{j=1}^{\alpha_1 K_1} d\hat{x}_j\, \hat{P}_1(\hat{x}_j) \prod_{l=1}^{\alpha_1 K_2} d\hat{w}_l\, \hat{P}_2(\hat{w}_l)\, \delta \left[h - \sum_{j=1}^{\alpha_1 K_1} x_j - \sum_{l=1}^{\alpha_2 K_2} w_l \right] \tag{229}$$

The numerical solution of these equations provides the typical overlap for cascading codes with two signal matrices parameterized by $\alpha_1 (\alpha_2 = \alpha - \alpha 1)$. In Figure 38 we compare results obtained by solving the above equations numerically (Monte Carlo integration with 4000 bins) and PP decoding simulations (10 runs, $N = 5000$) with $R = 1/5$ and $\alpha_1 = 3$. The agreement between theory and experiments supports the assumptions employed to obtain the saddle-point-like equations.

B. Optimizing Construction Parameters

Equations (222) to (229) can be used to optimize code constructions within a given family. For the family introduced in Figure 37 with fixed parameters K_1 and K_2, the optimization requires finding the value of α_1 that produces the highest threshold p_s. In Figure 39 we show the threshold (spinodal noise level) p_s for a KS code with $K_1 = 1$, $K_2 = 3$ and rate $R = 1/5$ ($\alpha = 5$). The optimal performance is obtained by selecting $\alpha_1 = 3$ and is very close to the channel capacity.

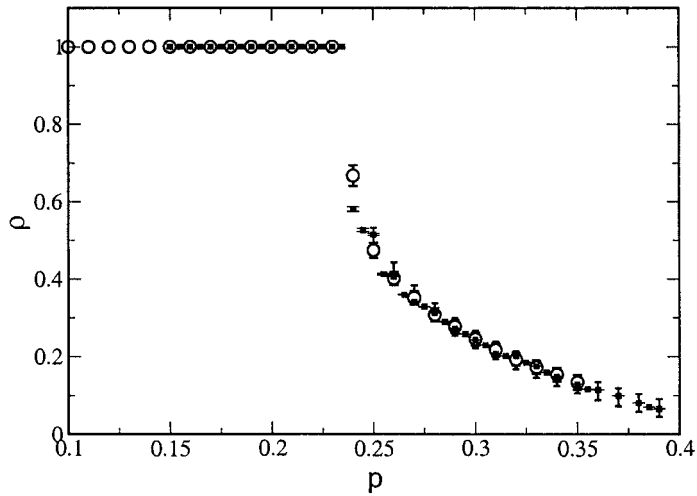

FIGURE 38. Monte Carlo integration of field distributions and simulations for a KS code with two signal matrices ($K_1 = 1$ and $K_2 = 3$), $\alpha = 5$ ($R = 1/5$) and $\alpha_1 = 3$. Circles: full statistics (4000 bins). Squares: simulations $N = 5000$.

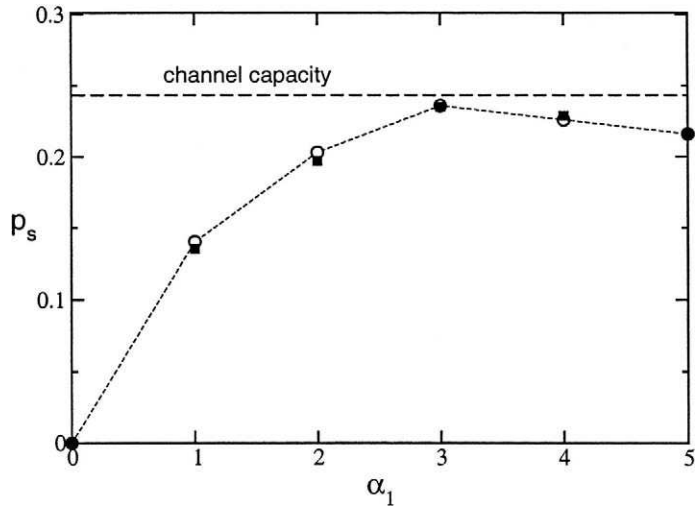

FIGURE 39. Spinodal noise level p_s as a function of α_1 for a KS code with $K_1 = 1$, $K_2 = 3$ and $R = 1/5$ ($\alpha = 5$). Circles: Monte Carlo integrations of saddle-point equations (4000 bins). Squares: PP decoding simulations (10 runs with size $N = 5000$). The best performance is reached for $\alpha_1 = 3$ and is close to the channel capacity for a BSC (indicated by a dashed line).

VII. Conclusions and Perspectives

In this chapter we have analyzed error-correcting codes based on very sparse matrices by mapping them onto spin systems of the statistical physics. The equivalence between coding concepts and statistical physics is summarized in a table.

Coding theory	Statistical physics
Message bits $\boldsymbol{s}$	Spins $\boldsymbol{S}$
Received bits $\boldsymbol{r}$	Multispin disordered couplings $\boldsymbol{J}$ (Sourlas)
Syndrome bits $\boldsymbol{z}$	Multispin couplings $\mathcal{J}$ (Gallager, MN, KS)
Bit error probability p_e	Gauged magnetization ρ (overlap)
Posterior probability	Boltzmann weight
MAP estimator	Ground state
MPM estimator	Thermal average at Nishimori's temperature

In the statistical physics framework, random parity-check matrices (or generator matrices as in the case of Sourlas codes), random messages, and noise are treated as quenched disorder and the replica method is employed to compute the free-energy. Under the assumption of replica symmetry, we found in most of the cases that two phases emerge: a successful decoding ($\rho = 1$) and failure ($\rho < 1$) phases. For MN codes with $K = 2$ or $K = 1$, three phases emerge, representing successful decoding, failure, and catastrophic failure.

The general picture that emerges shows a phase transition between successful and failure states that coincides with the information theory upper bounds in most cases, the exception being MN codes with $K = 2$ (and to some extent $K = 1$) where the transition is bellow the upper bound.

A careful analysis of replica symmetric quantities reveals unphysical behavior for low noise levels with the appearance of negative entropies. This question is resolved in the case of Sourlas codes with $K \to \infty$ by the introduction of a simple frozen spins first-step replica symmetry breaking ansatz. Despite the difficulties in the replica symmetric analysis, threshold noise values observed in simulations using probability propagation (PP) decoding agree with the noise level where metastable states (or spinodal points) appear in the replica symmetric free-energy.

A mean-field (Bethe) theory based on the use of a tree-like lattice (Husimi cactus) exposes the relationship between PP decoding and statistical physics and supports the agreement between theory and simulations as PP decoding can be reinterpreted as a method for finding local minima of a Bethe free-energy. Those minima can be described by distributions of cavity local fields that are solutions of the replica symmetric saddle-point equations.

The performance of the decoding process with probability propagation can be obtained by looking at the Bethe free-energy landscape (or the replica symmetric landscape); in this way we can show that information theoretic upper bounds can be attained by looking for global minima of the Bethe free-energy, which may require computing time that grows exponentially with the system size. In practical time scales, simple decoding procedures that simply find minima become trapped in metastable states. That is the reason practical thresholds are linked to the appearance of spinodal points in the Bethe free-energy.

For cascading codes, we adopted a different approach for the analysis. Using the insights obtained in the analysis of the other codes, we started by writing down the PP decoding equations and writing the Bethe free-energy and the saddle-point-like equations for distributions of cavity fields. The transitions predicted by these saddle-point-like equations were shown to agree with experiments. We then employed this procedure to optimize parameters of one simple family of cascading codes.

By studying the replica symmetric landscape we classified the various codes by their construction parameters, we also showed that modifications in code construction, such as the use of irregular matrices, can improve the performance by changing the way the free-energy landscape evolves with the noise level. We summarize in a table the results obtained.

	Channel capacity	Practical decoding of unbiased messages
Sourlas	$K \to \infty$	$K = 2$
Gallager	$K \to \infty$	Any K
MacKay–Neal	$K > 2$	$K = 1$, any $L > 1$ or $K = 2$
Cascading	Still unclear	$K_j = 1,\ 2$ for some j

These results shed light on the properties that limit the theoretical and practical performance of parity-check codes, explain the differences between Gallager and MN constructions, and explore the role of irregularity in LDPC error-correcting codes.

Some new directions are now being pursued and are worth mentioning. The statistical physics of Gallager codes with nonbinary alphabets is investigated in Nakamura *et al.* (2001). In Kabashima *et al.* (2001), the performance of error-correcting codes in the case of finite message lengths has been addressed, yielding tighter general reliability bounds. New analytical methods to investigate practical noise thresholds using statistical physics have been proposed in van Mourik *et al.* (2001) and in Kabashima *et al.* (2001), while the nature of Gallager codes phase diagram was studied in detail in Montanari (2001).

We believe that methods developed over the years in the statistical physics community can make a significant contribution in other areas of information theory. Research in some of these areas, such as CDMA and image restoration, is already underway.

Appendix A. Sourlas Codes: Technical Details

1. *Free-Energy*

To compute free-energies we need to calculate the replicated partition function (62). We can start from Eq. (60):

$$\langle \mathcal{Z}^n \rangle_{\mathcal{A},\xi,J} = \mathrm{Tr}_{\{S_j^\alpha\}}\Big[\big\langle \exp\big(-\beta \mathcal{H}^{(n)}(\{\boldsymbol{S}^\alpha\})\big)\big\rangle_{\mathcal{A},J,\xi}\Big], \tag{A.1}$$

where $\mathcal{H}^{(n)}(\{\boldsymbol{S}^\alpha\})$ represents the replicated Hamiltonian and α the replica indices. First, we average over the parity-check tensors $\mathcal{A}$; for that, an appropriate distribution has to be introduced, denoting $\mu \equiv \langle i_1, \ldots, i_K \rangle$ for a specific set of indices:

$$\langle \mathcal{Z}^n \rangle = \left\langle \frac{1}{\mathcal{N}} \sum_{\{\mathcal{A}\}} \prod_i \delta\left(\sum_{\mu \backslash i} \mathcal{A}_\mu - C \right) \mathrm{Tr}_{\{S_j^\alpha\}} \mathrm{e}^{-\beta \mathcal{H}^{(n)}(\{\boldsymbol{S}^\alpha\})} \right\rangle_{J,\xi}, \tag{A.2}$$

where the δ distribution imposes a restriction on the connectivity per spin, $\mathcal{N}$ is a normalization coefficient, and the notation $\mu \backslash i$ means the set μ except the element i. Using integral representations for the delta functions and rearranging:

$$\begin{aligned} \langle \mathcal{Z}^n \rangle = \mathrm{Tr}_{\{S_j^\alpha\}} \Bigg\langle & \frac{1}{\mathcal{N}} \left(\prod_i \oint \frac{dZ_i}{2\pi i} \frac{1}{Z_i^{C+1}} \right) \\ & \times \sum_{\{\mathcal{A}\}} \left(\prod_\mu \left(\prod_{i \in \mu} Z_i \right)^{\mathcal{A}_\mu} \right) \exp\big(-\beta \mathcal{H}^{(n)}(\{\boldsymbol{S}^\alpha\})\big) \Bigg\rangle_{J,\xi}. \end{aligned} \tag{A.3}$$

Remembering that $\mathcal{A} \in \{0, 1\}$, and using the expression (50) for the Hamiltonian, we can change the order of the summation and the product above and sum over $\mathcal{A}$:

$$\begin{aligned} \langle \mathcal{Z}^n \rangle = \mathrm{Tr}_{\{S_j^\alpha\}} \Bigg\langle & \frac{1}{\mathcal{N}} \left(\prod_i \oint \frac{dZ_i}{2\pi i} \frac{1}{Z_i^{C+1}} \right) e^{\beta F \sum_{\alpha,i} \xi_i S_i^\alpha} \\ & \times \prod_\mu \left[1 + \left(\prod_{i \in \mu} Z_i \right) \exp\left(\beta J_\mu \sum_\alpha \prod_{i \in \mu} S_i^\alpha \right) \right] \Bigg\rangle_{J,\xi}. \end{aligned} \tag{A.4}$$

Using the identity $\exp(\beta J_\mu \prod_{i\in\mu} S_i^\alpha) = \cosh(\beta)\,[1 + (\prod_{i\in\mu} S_i^\alpha)\tanh(\beta J_\mu)]$, we can perform the product over α to write:

$$\begin{aligned}\langle \mathcal{Z}^n \rangle = \mathrm{Tr}_{\{S_j^\alpha\}} \frac{1}{\mathcal{N}} \left(\prod_i \oint \frac{dZ_i}{2\pi i} \frac{1}{Z_i^{C+1}} \right) \left\langle e^{\beta F \sum_{\alpha,i} \xi_i S_i^\alpha} \right\rangle_\xi \\ \times \prod_\mu \left[1 + \left(\prod_{i\in\mu} Z_i \right) \cosh^n(\beta) \left(1 + \langle \tanh(\beta J) \rangle_J \sum_\alpha \prod_{i\in\mu} S_i^\alpha \right.\right. \\ \left.\left. + \langle \tanh^2(\beta J) \rangle_J \sum_{\langle \alpha_1 \alpha_2 \rangle} \prod_{i\in\mu} S_i^{\alpha_1} \prod_{j\in\mu} S_j^{\alpha_2} + \cdots \right) \right]. \end{aligned} \tag{A.5}$$

Defining $\langle \mu_1, \mu_2, \ldots, \mu_l \rangle$ as an ordered set of sets, and observing that for large N, $\sum_{\langle \mu_1 \cdots \mu_l \rangle} (\cdots) = \frac{1}{l!} (\sum_\mu (\cdots))^l$, we can perform the product over the sets μ and replace the energy series by an exponential:

$$\begin{aligned}\langle \mathcal{Z}^n \rangle = \mathrm{Tr}_{\{S_j^\alpha\}} \frac{1}{\mathcal{N}} \left(\prod_i \oint \frac{dZ_i}{2\pi i} \frac{1}{Z_i^{C+1}} \right) \left\langle e^{\beta F \sum \alpha, i \xi_i S_i^\alpha} \right\rangle_\xi \\ \times \exp \left[\cosh^n(\beta) \left(\sum_\mu \left(\prod_{i\in\mu} Z_i \right) + \langle \tanh(\beta J) \rangle_J \sum_\alpha \sum_\mu \prod_{i\in\mu} z_i S_i^\alpha \right.\right. \\ \left.\left. + \langle \tanh^2(\beta J) \rangle_J \sum_{\langle \alpha_1 \alpha_2 \rangle} \sum_\mu \prod_{i\in\mu} Z_i S_i^{\alpha_1} S_i^{\alpha_2} + \cdots \right) \right]. \end{aligned} \tag{A.6}$$

Observing that $\sum_\mu = 1/K! \sum_{i_1,\ldots i_K}$, defining $\mathcal{T}_l = \langle \cosh^n(\beta J) \tanh^l(\beta J) \rangle_J$ and introducing auxiliary variables $q_{\alpha_1 \cdots \alpha_m} = \frac{1}{N} \sum_i Z_i S_i^{\alpha_1} \cdots S_i^{\alpha_m}$ we find:

$$\begin{aligned}\langle \mathcal{Z}^n \rangle_{\mathcal{A},\xi,J} = \frac{1}{\mathcal{N}} \left(\prod_i \oint \frac{dZ_i}{2\pi i} \frac{1}{Z_i^{C+1}} \right) \left(\int \frac{dq_0\, d\hat{q}_0}{2\pi i} \right) \times \left(\prod_\alpha \int \frac{dq_\alpha d\hat{q}_\alpha}{2\pi i} \right) \cdots \\ \times \exp \left[\frac{N^K}{K!} \left(\mathcal{T}_0 q_0^K + \mathcal{T}_1 \sum_\alpha q_\alpha^K + \mathcal{T}_2 \sum_{\langle \alpha_1 \alpha_2 \rangle} q_{\alpha_1 \alpha_2}^K + \cdots \right) \right] \\ \times \exp \left[-N \left(q_0 \hat{q}_0 + \sum_\alpha q_\alpha \hat{q}_\alpha + \sum_{\langle \alpha_1 \alpha_2 \rangle} q_{\alpha_1 \alpha_2} \hat{q}_{\alpha_1 \alpha_2} + \cdots \right) \right] \\ \times \mathrm{Tr}_{\{S_j^\alpha\}} \left[\left\langle e^{\beta F \sum_{\alpha,i} \xi_i S_i^\alpha} \right\rangle_\xi \exp \sum_i \left(\hat{q}_0 Z_i + \sum_\alpha \hat{q}_\alpha Z_i S_i^\alpha + \cdots \right) \right]. \end{aligned} \tag{A.7}$$

The normalization constant is given by:

$$\mathcal{N} = \sum_{\{\mathcal{A}\}} \prod_i \delta\left(\sum_{\mu\backslash i} \mathcal{A}_\mu - C\right), \tag{A.8}$$

and can be computed using exactly the same methods as above, resulting in:

$$\mathcal{N} = \left(\prod_i \oint \frac{dZ_i}{2\pi i} \frac{1}{Z_i^{C+1}}\right)\left(\int \frac{dq_0\, d\hat{q}_0}{2\pi i}\right) \times \exp\left[\frac{N^K}{K!} q_0^K - N q_0\, \hat{q}_0 + \hat{q}_0 \sum_i Z_i\right]. \tag{A.9}$$

Computing the integrals over Z_is and using Laplace method to compute the integrals over q_0 and $\hat{q}_0$ we obtain:

$$\mathcal{N} = \exp\left\{\mathrm{Extr}_{q_0,\hat{q}_0}\left[\frac{N^K}{K!} q_0^K - N q_0 \hat{q}_0 + N \ln\left(\frac{\hat{q}_0^C}{C!}\right)\right]\right\}. \tag{A.10}$$

The extremum point is given by

$$q_0 = N^{(1-K)/K}[(K-1)!C]^{1/K}$$

and

$$\hat{q}_0 = (C\,N)^{(K-1/K)}\,[(K-1)!]^{-1/K}\,.$$

Replacing the auxiliary variables in Eq. (A.7) using $q_{\alpha_1\cdots\alpha_m}/q_0 \to q_{\alpha_1\cdots\alpha_m}$ and $\hat{q}_{\alpha_1\cdots\alpha_m}/q_0 \to \hat{q}_{\alpha_1\cdots\alpha_m}$, computing the integrals over Z_i and using Laplace method to evaluate the integrals, we finally find Eq. (62).

2. *Replica Symmetric Solution*

The replica symmetric free-energy (66) can be obtained by plugging the ansätz (65) into Eq. (A.7). Using Laplace method we obtain:

$$\langle \mathcal{Z}^n \rangle_{\mathcal{A},\xi,J} = \frac{1}{\mathcal{N}} \exp\left\{N\,\mathrm{Extr}_{\pi,\hat{\pi}}\left[\frac{C}{K}\mathcal{G}_1 - C\,\mathcal{G}_2 + \mathcal{G}_3\right]\right\}, \tag{A.11}$$

where:

$$\mathcal{G}_1 = \mathcal{T}_0 + \mathcal{T}_1 \sum_\alpha \int \prod_j^K (dx_j\, \pi(x_j) \tanh(\beta x_j)) + \mathcal{T}_2 \sum_{\langle\alpha_1\alpha_2\rangle} \int \prod_j^K (dx_j\, \pi(x_j) \tanh^2(\beta x_j)) + \cdots, \tag{A.12}$$

Putting everything together, using Eq. (59) and some simple manipulation we find Eq. (66).

3. *Local Field Distribution*

Here we derive explicitly Eq. (68). The gauge transformed overlap can be written as

$$\rho = \frac{1}{N}\sum_{i=1}^{N} \langle \mathrm{sign}(m_i)\rangle_{\mathcal{A},J,\xi}\,, \tag{A.19}$$

introducing the notation $m_i = \langle S_i\rangle$, where $\langle\cdots\rangle$ is a gauged average.

For an arbitrary natural number p, one can compute pth moment of m_i

$$\left\langle m_i^p\right\rangle_{\mathcal{A},J,\xi} = \lim_{n\to 0}\left\langle \sum_{S^1,\ldots,S^n} S_i^1\cdot S_i^2\cdot\cdots\cdot S_i^p e^{-\beta\sum_{\alpha=1}^n \mathcal{H}^{(\alpha)}}\right\rangle_{\mathcal{A},J,\xi}, \tag{A.20}$$

where $\mathcal{H}^{(\alpha)}$ denotes the gauged Hamiltonian of the αth replica. By performing the same steps described in the Appendices A.1 and A.2, introducing the auxiliary functions $\pi(x)$ and $\hat{\pi}(y)$ defined in Eqs. (65), one obtains

$$\left\langle m_i^p\right\rangle_{\mathcal{A},J,\xi} = \int\prod_{j=1}^{C} dy_j\hat{\pi}(y_j)\left\langle \tanh^p\left(\beta F\xi + \beta\sum_{j=1}^{C} y_j\right)\right\rangle_{\xi}. \tag{A.21}$$

Employing the identity

$$\mathrm{sign}(x)+1 = 2\lim_{n\to\infty}\sum_{m=0}^{n}\frac{2n!}{(2n-m)!m!}\left(\frac{1+x}{2}\right)^{2n-m}\left(\frac{1-x}{2}\right)^{m} \tag{A.22}$$

which holds for any arbitrary real number $x\in[-1,1]$ and Eqs. (A.21) and (A.22), one obtains

$$\begin{aligned}\langle \mathrm{sign}(m_i)\rangle_{\mathcal{A},J,\xi} + 1 &= 2\int dh\,P(h)\\ &\quad\times\lim_{n\to\infty}\sum_{m=0}^{n} C_{2n,m}\left(\frac{1+h}{2}\right)^{2n-m}\left(\frac{1-h}{2}\right)^{m}\\ &= \int dh\,P(h)\,\mathrm{sign}(h),\end{aligned} \tag{A.23}$$

where we introduced the local fields distribution

$$P(h) = \int\prod_{j=1}^{C} dy_j\,\hat{\pi}(y_j)\left\langle\delta\left(h - F\xi - \sum_{j=1}^{C} y_j\right)\right\rangle_{\xi}, \tag{A.24}$$

thus reproducing Eq. (68).

$$\mathcal{G}_2 = 1 + \sum_{\alpha} \int dx\, dy\, \pi(x)\, \hat{\pi}(y) \tanh(\beta x) \tanh(\beta y) + \sum_{\langle \alpha_1 \alpha_2 \rangle} \int dx\, dy\, \pi(x)\, \hat{\pi}(y) \tanh^2(\beta x) \tanh^2(\beta y) + \cdots \quad \text{(A.13)}$$

and

$$\mathcal{G}_3 = \frac{1}{N} \ln \left\{ \left(\prod_i \oint \frac{dZ_i}{2\pi i} \frac{1}{Z_i^{C+1}} \right) \mathrm{Tr}_{\{S_j^\alpha\}} \left[\left\langle \exp \beta F \sum_{\alpha,i} \xi_i S_i^\alpha \right\rangle_\xi \times \exp \hat{q}_0 \left(\sum_i Z_i + \sum_\alpha \sum_i Z_i S_i^\alpha \int dy\, \hat{\pi}(y) \tanh(\beta y) + \sum_{\langle \alpha_1 \alpha_2 \rangle} \sum_i Z_i S_i^{\alpha_1} S_i^{\alpha_2} \int dy\, \hat{\pi}(y) \tanh^2(\beta y) + \cdots \right) \right] \right\}. \quad \text{(A.14)}$$

The equation for $\mathcal{G}_1$ can be worked out by using the definition of $\mathcal{T}_m$ and the fact that $(\sum_{\langle \alpha_1 \ldots \alpha_l \rangle} 1) = \binom{n}{l}$ to write:

$$\mathcal{G}_1 = \left\langle \cosh^n(\beta J) \int \left(\prod_{j=1}^K dx_j\, \pi(x_j) \right) \left(1 + \tanh(\beta J) \prod_{j=1}^K \tanh(\beta x_j) \right)^n \right\rangle_J . \quad \text{(A.15)}$$

Following exactly the same steps we obtain:

$$\mathcal{G}_2 = \int dx\, dy\, \pi(x)\, \hat{\pi}(y)\, (1 + \tanh(\beta x) \tanh(\beta y))^n , \quad \text{(A.16)}$$

and

$$\mathcal{G}_3 = \ln \left\{ \mathrm{Tr}_{\{S^\alpha\}} \left[\left\langle \exp \left(\beta F \xi \sum_\alpha S^\alpha \right) \right\rangle_\xi \times \oint \frac{dZ}{2\pi i} \frac{1}{Z^{C+1}} \exp \left(\hat{q}_0\, Z \int dy\, \hat{\pi}(y) \prod_{\alpha=1}^n (1 + S^\alpha \tanh(\beta y)) \right) \right] \right\}. \quad \text{(A.17)}$$

Computing the integral over Z_i and the trace, we finally find:

$$\mathcal{G}_3 = \ln \left\{ \frac{\hat{q}_0^C}{C!} \int \prod_{l=1}^C dy_l\, \hat{\pi}(y_l) \left[\sum_{\sigma=\pm 1} \langle e^{\sigma \beta F \xi} \rangle_\xi \prod_{l=1}^C (1 + \sigma \tanh(\beta y_l)) \right]^n \right\}. \quad \text{(A.18)}$$

4. *Zero Temperature Self-Consistent Equations*

In this section we describe how one can write a set of self-consistent equations to solve the zero temperature saddle-point Eqs. (84). Supposing a three peaks ansätz given by:

$$\hat{\pi}(y) = p_+\delta(y-1) + p_0\delta(y) + p_-\delta(y+1) \tag{A.25}$$

$$\pi(x) = \sum_{l=1-C}^{C-1} T_{[p_\pm, p_0; C-1]}(l)\,\delta(x-l), \tag{A.26}$$

with

$$T_{[p_+,p_0,p_-;C]}(l) = \sum_{\{k,h,m;\,k-h=l;\;k+h+m=C-1\}} \frac{(C-1)!}{k!h!m!}\, p_+^k p_0^h p_-^m. \tag{A.27}$$

We can consider the problem as a random walk, where $\hat{\pi}(y)$ describes the probability of one step of length y ($y > 0$ means one step to the right) and $\pi(x)$ describes the probability of being at distance x from the origin after $C-1$ steps. With this idea in mind, it is relatively easy to understand $T_{[p_+,p_0,p_-;C-1]}(l)$ as the probability of walking the distance l after $C-1$ steps with the probabilities p_+, p_- and p_0 of, respectively, moving right, left, and staying at the same position. We define the probabilities of walking right/left as $\psi_\pm = \sum_l^{C-1} T_{[p_+,p_0,p_-;C-1]}(\pm l)$. Using second saddle-point Eqs. (84):

$$p_+ = \int \left[\prod_{l=1}^{K-1} dx_l\, \pi(x_l)\right] \left\langle \delta\left[1 - \operatorname{sign}\left(J\prod_{l=1}^{K-1} x_l\right) \min(|J|, |x_1|, \ldots |\right]\right\rangle_J . \tag{A.28}$$

The right side of the above equality can be read as the probability of making $K-1$ independent walks, such that after $C-1$ steps: none is at origin and an even (for $J=+1$) or odd (for $J=-1$) number of walks is at the left side.

Using this reasoning for p_- and p_0, we can finally write:

$$\begin{aligned} p_+ = (1-p) &\sum_{j=0}^{\lfloor \frac{K-1}{2} \rfloor} \frac{(K-1)!}{2j!(K-1-2j)!} \psi_-^{2j} \psi_+^{K-2j-1} \\ &+ p \sum_{j=0}^{\lfloor \frac{K-1}{2} \rfloor - 1} \frac{(K-1)!}{(2j+1)!(K-2-2j)!} \psi_-^{2j+1} \psi_+^{K-2j-2} \\ &+ p\,\psi_-^{K-1}\,\mathrm{odd}(K-1) \end{aligned} \tag{A.29}$$

$$p_- = (1-p)\sum_{j=0}^{\lfloor\frac{K-1}{2}\rfloor-1}\frac{(K-1)!}{(K-2j-2)!(2j+1)!}\psi_-^{2j+1}\psi_+^{K-2j-2} + p\sum_{j=0}^{\lfloor\frac{K-1}{2}\rfloor-1}\frac{(K-1)!}{(K-2j-1)!2j!}\psi_-^{2j}\psi_+^{K-2j-1} + (1-p)\psi_-^{K-1}\mathrm{odd}(K-1), \tag{A.30}$$

where $\mathrm{odd}(x) = 1(0)$ if x is odd (even). Using that $p_+ + p_- + p_0 = 1$, one can obtain p_0. A similar set of equations can be obtained for a five-peaks ansätz leading to the same set of solutions for the ferromagnetic and paramagnetic phases. The paramagnetic solution $p_0 = 1$ is always a solution, for $C > K$ a ferromagnetic solution with $p_+ > p_- > 0$ emerges.

5. *Symmetric Channels Averages at Nishimori's Temperature*

Here we establish the identity $\langle J\rangle_J = \langle J\tanh(\beta_N J)\rangle_J$ for symmetric channels. It was shown in Sourlas (1994a) that:

$$\beta_N J = \frac{1}{2}\ln\left(\frac{p(J\mid 1)}{p(J\mid -1)}\right), \tag{A.31}$$

where β_N is the Nishimori temperature and $p(J\mid J^0)$ are the probabilities that a transmitted bit J^0 is received as J. From this we can easily find:

$$\tanh(\beta_N J) = \frac{p(J\mid 1) - p(J\mid -1)}{p(J\mid 1) + p(J\mid -1)}. \tag{A.32}$$

In a symmetric channel ($p(J\mid -J^0) = p(-J\mid J^0)$), it is also represented as

$$\tanh(\beta_N J) = \frac{p(J\mid 1) - p(-J\mid 1)}{p(J\mid 1) + p(-J\mid 1)}. \tag{A.33}$$

Therefore,

$$\begin{aligned}\langle J\tanh(\beta_N J)\rangle_J &= \mathrm{Tr}_J\, p(J\mid 1)\frac{Jp(J\mid 1)}{p(J\mid 1)+p(-J\mid 1)} \\ &\quad + \mathrm{Tr}_J\, p(J\mid 1)\frac{(-J)p(-J\mid 1)}{p(J\mid 1)+p(-J\mid 1)} \\ &= \mathrm{Tr}_J\, p(J\mid 1)\frac{Jp(J\mid 1)}{p(J\mid 1)+p(-J\mid 1)} \\ &\quad + \mathrm{Tr}_J\, p(-J\mid 1)\frac{Jp(J\mid 1)}{p(-J\mid 1)+p(J\mid 1)} \\ &= \mathrm{Tr}_J\, Jp(J\mid 1) \\ &= \langle J\rangle_J. \end{aligned} \tag{A.34}$$

6. *Probability Propagation Equations*

In this section we derive the probability propagation Eqs. (36) and (34) in the form (96). We start by introducing the following representation for the variables $Q_{\mu k}^{S_k}$ and $R_{\mu k}^{S_k}$:

$$Q_{\mu k}^{S_k} = \frac{1}{2}(1 + m_{\mu k} S_k) \qquad R_{\mu k}^{S_k} = \frac{1}{2}(1 + \hat{m}_{\mu k} S_k). \tag{A.35}$$

We can now put (91), (95), and (A.35) together to write:

$$\begin{aligned} R_{\mu j}^{S_k} &= \frac{1}{a_\mu} \sum_{\{S_k : k \in \mathcal{L}(\mu)\setminus j\}} \frac{1}{2}\cosh(\beta J_\mu) \left(1 + \tanh(\beta J_\mu) \prod_{j \in \mathcal{L}\mu} S_j\right) \\ &\quad \times \prod_{k \in \mathcal{L}(\mu)\setminus j} \frac{1}{2}\left(1 + m_{\mu k} S_k\right) \\ &= \frac{1}{2^K}\frac{1}{a_\mu} \sum_{\{S_k : k \in \mathcal{L}(\mu)\setminus j\}} \cosh(\beta J_\mu) \left(1 + \tanh(\beta J_\mu) \prod_{j \in \mathcal{L}(\mu)} S_j\right) \\ &\quad \times \left(1 + \sum_{k \in \mathcal{L}(\mu)\setminus j} m_{\mu k} S_k + \sum_{k \neq l \in \mathcal{L}(\mu)\setminus j} m_{\mu k} m_{\mu l} S_k S_l + \cdots\right) \\ &= \frac{1}{2^K}\frac{1}{a_\mu}\cosh(\beta J_\mu) \left(1 + \tanh(\beta J_\mu) S_j \prod_{k \in \mathcal{L}(\mu)\setminus j} m_{\mu k}\right) \\ &= \frac{1}{2}\left(1 + \tanh(\beta J_\mu) S_j \prod_{k \in \mathcal{L}(\mu)\setminus j} m_{\mu k}\right). \end{aligned} \tag{A.36}$$

To obtain the last line, we used that the normalization constant is $a_\mu = \frac{1}{2^{K-1}}\cosh(\beta J_\mu)$. Writing the above equation in terms of the new variable $\hat{m}_{\mu k}$ we obtain the first Eq. (96):

$$\begin{aligned} \hat{m}_{\mu k} &= R_{\mu k}^{(+)} - R_{\mu k}^{(-)} \\ &= \frac{1}{2}\left(1 + \tanh(\beta J_\mu) \prod_{k \in \mathcal{L}(\mu)\setminus j} m_{\mu k}\right) - \frac{1}{2}\left(1 - \tanh(\beta J_\mu) \prod_{k \in \mathcal{L}(\mu)\setminus j} m_{\mu k}\right) \\ &= \tanh(\beta J_\mu) \prod_{k \in \mathcal{L}(\mu)\setminus j} m_{\mu k}. \end{aligned} \tag{A.37}$$

To obtain the second Eq. (96), we write:

$$Q_{\mu k}^{S_k} = a_{\mu k} \frac{1}{2}(1 + \tanh(\beta'_N S_k)) \prod_{\nu \in \mathcal{M}(k)\backslash\mu} \frac{1}{2}(1 + \hat{m}_{\nu k} S_k). \tag{A.38}$$

In the new variables $m_{\mu k}$:

$$\begin{aligned} m_{\mu k} = a_{\mu k} \frac{1}{2^K} \Bigg\{ & (1 + \tanh(\beta'_N)) \prod_{\nu \in \mathcal{M}(k)\backslash\mu} (1 + \hat{m}_{\nu k}) \\ & - (1 - \tanh(\beta'_N)) \prod_{\nu \in \mathcal{M}(k)\backslash\mu} (1 - \hat{m}_{\nu k}) \Bigg\} \end{aligned} \tag{A.39}$$

By using the identity $e^{\sigma x} = \cosh(x)(1 + \sigma \tanh(x))$ we can write:

$$\begin{aligned} m_{\mu k} = & \frac{\exp\left[\sum_{\nu \in \mathcal{M}(k)\backslash\mu} \operatorname{atanh}(m_{\nu k}) + \beta'_N\right]}{a_{\mu k}^{-1} 2^K \cosh(\beta'_N) \prod_{\nu \in \mathcal{M}(k)\backslash\mu} \cosh(\operatorname{atanh}(m_{\nu k}))} \\ & - \frac{\exp\left[-\sum_{\nu \in \mathcal{M}(k)\backslash\mu} \operatorname{atanh}(m_{\nu k}) - \beta'_N\right]}{a_{\mu k}^{-1} 2^K \cosh(\beta'_N) \prod_{\nu \in \mathcal{M}(k)\backslash\mu} \cosh(\operatorname{atanh}(m_{\nu k}))} \end{aligned} \tag{A.40}$$

Computing the normalization $a_{\mu j}$ along the same lines gives:

$$\begin{aligned} a_{\mu k}^{-1} = & \frac{\exp\left[\sum_{\nu \in \mathcal{M}(k)\backslash\mu} \operatorname{atanh}(m_{\nu k}) + \beta'_N\right]}{2^K \cosh(\beta'_N) \prod_{\nu \in \mathcal{M}(k)\backslash\mu} \cosh(\operatorname{atanh}(m_{\nu k}))} \\ & + \frac{\exp\left[-\sum_{\nu \in \mathcal{M}(k)\backslash\mu} \operatorname{atanh}(m_{\nu k}) - \beta'_N\right]}{2^K \cosh(\beta'_N) \prod_{\nu \in \mathcal{M}(k)\backslash\mu} \cosh(\operatorname{atanh}(m_{\nu k}))} \end{aligned} \tag{A.41}$$

Inserting (A.41) into (A.40) gives:

$$m_{\mu k} = \tanh\left[\sum_{\nu \in \mathcal{M}(k)\backslash\mu} \operatorname{atanh}(m_{\nu k}) + \beta'_N\right]. \tag{A.42}$$

to write:

$$\langle \mathcal{Z}^n \rangle_{\mathcal{A},\zeta} = \frac{1}{\mathcal{N}} \sum_{\tau^1,\ldots,\tau^n} \prod_{j=1}^{M} \left\langle \exp\left(F\zeta\beta \sum_{\alpha=1}^{n} \tau_j^\alpha \right) \right\rangle_\zeta$$
$$\times \sum_{\{\mathcal{A}\}} \prod_{j=1}^{M} \left[\oint \frac{dZ_j}{2\pi i} \frac{1}{Z_j^{C+1}} \right]$$
$$\times \prod_{\langle i_1 \cdots i_K \rangle} \left\{ 1 + \frac{\cosh^n(\beta\gamma)}{\mathrm{e}^{n\beta\gamma}} (Z_{i_1} \cdots Z_{i_K}) \right.$$
$$\left. \times \prod_{\alpha=1}^{n} \left[1 + \tau_{i_1}^\alpha \cdots \tau_{i_K}^\alpha \tanh(\beta\gamma) \right] \right\}. \tag{B.4}$$

By following Appendix A.1 from Eq. (A.5), we can finally find Eq. (120).

2. *Replica Symmetric Solution*

As in the code of Sourlas (Appendix A.2), the replicated partition function can be put into the form:

$$\langle \mathcal{Z}^n \rangle_{\mathcal{A},\zeta} = \frac{1}{\mathcal{N}} \exp\left\{ M \operatorname{Extr}_{\pi,\hat{\pi}} \left[\frac{C}{K} \mathcal{G}_1 - C\,\mathcal{G}_2 + \mathcal{G}_3 \right] \right\}. \tag{B.5}$$

Introducing the replica symmetric ansätz (121) into the functions $\mathcal{G}_1, \mathcal{G}_2$, and $\mathcal{G}_3$, we obtain:

$$\mathcal{G}_1(n) = \mathcal{T}_0 + \mathcal{T}_1 \sum_\alpha q_\alpha^K + \mathcal{T}_2 \sum_{\langle \alpha_1 \alpha 2 \rangle} q_{\alpha_1\alpha_2}^K + \cdots$$
$$= \frac{\cosh^n(\beta\gamma)}{\mathrm{e}^{n\gamma\beta}} \int \prod_{j=1}^{K} dx_j\, \pi(x_j) \left[1 + \frac{n!}{(n-1)!} \tanh(\beta\gamma) \prod_{j=1}^{K} x_j \right.$$
$$\left. + \frac{n!}{(n-2)!2!} \tanh^2(\beta\gamma) \prod_{j=1}^{K} x_j^2 + \cdots \right]$$
$$= \frac{\cosh^n(\beta\gamma)}{\mathrm{e}^{n\gamma\beta}} \int \prod_{j=1}^{K} dx_j \pi(x_j) \left[1 + \tanh(\beta\gamma) \prod_{j=1}^{K} x_j \right]^n$$
$$\xrightarrow{\gamma\to\infty} \frac{1}{2^n} \int \prod_{j=1}^{K} dx_j\, \pi(x_j) \left[1 + \prod_{j=1}^{K} x_j \right]^n, \tag{B.6}$$

Appendix B. Gallager Codes: Technical Details

1. *Replica Theory*

The replica theory for Gallager codes is very similar to the theory obtained for Sourlas codes (see Appendix A). We start with Eq. (116):

$$\langle \mathcal{Z}^n \rangle_{\mathcal{A},\zeta} = \sum_{\tau^1,\ldots,\tau^n} \prod_{j=1}^{M} \left\langle \exp\left(F\zeta\beta \sum_{\alpha=1}^{n} \tau_j^\alpha \right) \right\rangle_\zeta \times \left\langle \prod_{\langle i_1 \cdots i_K \rangle} \prod_{\alpha=1}^{n} \exp\left[\beta\gamma \mathcal{A}_{\langle i_1 \cdots i_K \rangle} \left(\tau_{i_1}^\alpha \cdots \tau_{i_K}^\alpha - 1 \right) \right] \right\rangle_{\mathcal{A}}. \quad \text{(B.1)}$$

The average over constructions $\mathcal{A}$ is then introduced using Eq. (117):

$$\langle \mathcal{Z}^n \rangle_{\mathcal{A},\zeta} = \frac{1}{\mathcal{N}} \sum_{\tau^1,\ldots,\tau^n} \prod_{j=1}^{M} \left\langle \exp\left(F\zeta\beta \sum_{\alpha=1}^{n} \tau_j^\alpha \right) \right\rangle_\zeta \times \sum_{\{\mathcal{A}\}} \prod_{j=1}^{M} \left[\oint \frac{dZ_j}{2\pi i} \frac{1}{Z_j^{C+1}} Z_j^{\sum_{\langle i_1 = j, i_2, \ldots, i_K \rangle} \mathcal{A}_{\langle i_1 = j, \ldots, i_K \rangle}} \right] \times \prod_{\langle i_1 \cdots i_K \rangle} \exp\left[\beta\gamma \mathcal{A}_{\langle i_1 \cdots i_K \rangle} \sum_{\alpha=1}^{n} \left(\tau_{i_1}^\alpha \cdots \tau_{i_K}^\alpha - 1 \right) \right]. \quad \text{(B.2)}$$

After observing that

$$\prod_{j=1}^{M} Z_j^{\sum_{\langle i_1 = j, i_2, \ldots, i_K \rangle} \mathcal{A}_{\langle i_1 = j, \ldots, i_K \rangle}} = \prod_{\langle i_1 \cdots i_K \rangle} (Z_{i_1} \cdots Z_{i_K})^{\mathcal{A}_{\langle i_1 \cdots i_K \rangle}},$$

we can compute the sum over $\mathcal{A}_{\langle i_1 \cdots i_K \rangle} \in \{0, 1\}$:

$$\langle \mathcal{Z}^n \rangle_{\mathcal{A},\zeta} = \frac{1}{\mathcal{N}} \sum_{\tau^1,\ldots,\tau^n} \prod_{j=1}^{M} \left\langle \exp\left(F\zeta\beta \sum_{\alpha=1}^{n} \tau_j^\alpha \right) \right\rangle_\zeta \times \prod_{j=1}^{M} \left[\oint \frac{dZ_j}{2\pi i} \frac{1}{Z_j^{C+1}} \right] \times \prod_{\langle i_1 \cdots i_K \rangle} \left\{ 1 + \frac{Z_{i_1} \cdots Z_{i_K}}{e^{n\beta\gamma}} \prod_{\alpha=1}^{n} \exp\left[\beta\gamma \left(\tau_{i_1}^\alpha \cdots \tau_{i_K}^\alpha \right) \right] \right\}. \quad \text{(B.3)}$$

We can now use the identity $e^{x\sigma} = \cosh(x)(1 + \sigma \tanh(x))$, where $\sigma = \pm 1$,

where the hyperparameters γ^*, F^* are used in the Hamiltonian $\mathcal{H}$ and β^* is the temperature, while γ, F and β are the actual parameters of the encoding and corruption processes.

The Nishimori condition is defined by setting the temperature and all hyperparameters of the Hamiltonian to the values in the encoding and corruption processes. If this is done, the expression for the energy can be rewritten:

$$U = \frac{\sum_{\mathcal{J},\tau} \mathcal{H}(\gamma, F) P_{\gamma\beta}(\{\mathcal{J}_\mu\} \mid \boldsymbol{\tau}) P_{F\beta}(\boldsymbol{\tau})}{\sum_{\mathcal{J},\tau} P_{\gamma\beta}(\{\mathcal{J}_\mu\} \mid \boldsymbol{\tau}) P_{F\beta}(\boldsymbol{\zeta})}. \tag{B.12}$$

By plugging (106) for the likelihood $P_{\gamma\beta}(\{\mathcal{J}_\mu\} \mid \boldsymbol{\tau})$ and for the prior $P_{F\beta}(\boldsymbol{\zeta})$; setting the hyperparameters to $\gamma \to \infty$, $\beta = 1$ and $F = \operatorname{atanh}(1-2p)$ and performing the summation over $\mathcal{J}$ first, we easily get:

$$u = \lim_{M\to\infty} \frac{U}{M} = -F(1-2p). \tag{B.13}$$

Note that this expression is independent of the macroscopic state of the system.

4. *Recursion Relations*

We start by introducing the effective field $\hat{x}_{\nu j}$:

$$\tanh(\beta \hat{x}_{\nu j}) = \frac{P_{\nu j}(+)\mathrm{e}^{-\beta F} - P_{\nu j}(-)\mathrm{e}^{+\beta F}}{P_{\nu j}(+)\mathrm{e}^{-\beta F} + P_{\nu j}(-)\mathrm{e}^{+\beta F}}. \tag{B.14}$$

Equation (129) can be easily obtained from the equation above. Equation (130) is then obtained by introducing Eq. (128) into Eq. (129), and performing a straightforward manipulation, we obtain Eq. (131):

$$\exp(-2\beta \hat{x}_{\mu k}) = \frac{\mathrm{Tr}_{\{\tau_j\}} \mathrm{e}^{\beta\gamma(-\mathcal{J}_\mu \prod_j'' \tau_j - 1)} \prod_\nu' \prod_j'' \mathrm{e}^{\beta F \tau_j + \beta \hat{x}_{\nu j}(\tau_j - 1)}}{\mathrm{Tr}_{\{\tau_j\}} \mathrm{e}^{\beta\gamma(+\mathcal{J}_\mu \tau_k \prod_j'' \tau_j - 1)} \prod_\nu' \prod_j'' \mathrm{e}^{\beta F \tau_j + \beta \hat{x}_{\nu j}(\tau_j - 1)}}, \tag{B.15}$$

where

$$\exp\left(\beta \hat{x}_{\nu j}(\tau_j - 1)\right) = \frac{P_{\nu j}(\tau_j)\mathrm{e}^{-\beta F \tau_j}}{P_{\nu j}(+)\mathrm{e}^{-\beta F}}$$

and the products $\prod_\nu'$ and $\prod_j''$ are over $\nu \in \mathcal{M}(j) \setminus \mu$ and $j \in \mathcal{L}(\mu) \setminus k$, respectively.

The above equation can be rewritten as:

$$\mathrm{e}^{-2\beta \hat{x}_{\mu k}} = \frac{\mathrm{Tr}_{\{\tau_j\}} \prod_j'' \mathrm{e}^{(\beta F + \sum_\nu' \hat{x}_{\nu j})\tau_j} \left[\left(1 - \mathcal{J}_\mu \prod_j'' \tau_j \tanh(\beta\gamma)\right)\right]}{\mathrm{Tr}_{\{\tau_j\}} \prod_j'' \mathrm{e}^{(\beta F + \sum_\nu' \hat{x}_{\nu j})\tau_j} \left[\left(1 + \mathcal{J}_\mu \prod_j'' \tau_j \tanh(\beta\gamma)\right)\right]}. \tag{B.16}$$

where we use the Nishimori condition $\gamma \to \infty$, $\beta = 1$ to obtain the last line.

$$\begin{aligned}\mathcal{G}_2(n) &= 1 + \sum_{\alpha} q_\alpha \hat{q}_\alpha + \sum_{\langle \alpha_1 \alpha_2 \rangle} q_{\alpha_1 \alpha_2} \hat{q}_{\alpha_1 \alpha_2} + \cdots \\ &= \int dx d\hat{x}\, \pi(x) \hat{\pi}(\hat{x}) \left[1 + x\hat{x}\right]^n . \end{aligned} \tag{B.7}$$

and

$$\begin{aligned}\mathcal{G}_3(n) &= \frac{1}{M} \ln \mathrm{Tr}_{\{\tau^\alpha\}} \left[\left\langle \exp\left[F\beta\zeta \sum_{\alpha=1}^{n} \tau^\alpha \right] \right\rangle_\zeta \right. \\ &\quad \left. \times \oint \frac{dZ}{2\pi i} \frac{\exp\left[Z \sum_{m=0}^{n} \sum_{\langle \alpha_1 \cdots \alpha_m \rangle} \hat{q}_{\alpha_1 \cdots \alpha_m} \tau^{\alpha_1} \cdots \tau^{\alpha_m} \right]}{Z^{C+1}} \right] \\ &= \frac{1}{M} \ln \mathrm{Tr}_{\{\tau^\alpha\}} \left[\left\langle \exp\left[F\beta\zeta \sum_{\alpha=1}^{n} \tau^\alpha \right] \right\rangle_\zeta \right. \\ &\quad \left. \times \oint \frac{dZ}{2\pi i} \frac{\exp\left[Z \int d\hat{x}\, \hat{\pi}(\hat{x}) \prod_{\alpha=1}^{n} (1 + \tau^\alpha \hat{x}) \right]}{Z^{C+1}} \right] \\ &= \frac{1}{M} \ln \frac{\hat{q}_0^C}{C!} \int \prod_{l=1}^{C} d\hat{x}_l\, \hat{\pi}(\hat{x}_l) \left[\sum_{\tau=\pm 1} \left\langle e^{F\beta\zeta\tau} \right\rangle_\zeta \prod_{l=1}^{C} (1 + \tau \hat{x}_l) \right]^n \end{aligned} \tag{B.8}$$

By using Eq. (115) we can write

$$f = -\frac{1}{\beta} \mathrm{Extr}_{\pi, \hat{\pi}} \frac{\partial}{\partial n} \Big|_{n=0} \left[\frac{C}{K} \mathcal{G}_1(n) - C\mathcal{G}_2(n) + \mathcal{G}_3(n) \right], \tag{B.9}$$

what yields the free-energy (123).

3. *Energy Density at the Nishimori Condition*

In general, the average internal energy is evaluated as:

$$U = \langle\langle \mathcal{H}(\gamma^*, F^*) \rangle_{\beta^*} \rangle_{\mathcal{J}, \zeta} \tag{B.10}$$

$$\begin{aligned} &= \sum_{\mathcal{J}} \frac{\sum_\zeta P_{\gamma\beta}(\{\mathcal{J}_\mu\} \mid \boldsymbol{\zeta}) P_{F\beta}(\boldsymbol{\zeta})}{\sum_{\tilde{\mathcal{J}}, \tilde{\zeta}} P_{\gamma\beta}(\{\tilde{\mathcal{J}}_\mu\} \mid \tilde{\boldsymbol{\zeta}}) P_{F\beta}(\tilde{\boldsymbol{\zeta}})} \\ &\quad \times \frac{\sum_\tau \mathcal{H}(\gamma^*, F^*) P_{\gamma^*\beta^*}(\{\mathcal{J}_\mu\} \mid \boldsymbol{\tau}) P_{F^*\beta^*}(\boldsymbol{\tau})}{\sum_{\tilde{\tau}} P_{\gamma^*\beta^*}(\{\mathcal{J}_\mu\} \mid \tilde{\boldsymbol{\tau}}) P_{F^*\beta^*}(\tilde{\boldsymbol{\tau}})}, \end{aligned} \tag{B.11}$$

From Eqs. (C.2) and (C.3) above we can write:

$$1 - 2p_z^1(K) = \sum_{l\,\text{odd}}^{K} (-1)^l \frac{K!}{(K-l)!l!} p^l (1-p)^{K-l}$$
$$= (1 - p - p)^K = (1 - 2p)^K. \tag{C.4}$$

From which we find:

$$p_z^1(K) = \frac{1}{2} - \frac{1}{2}(1 - 2p)^K. \tag{C.5}$$

For MN codes, syndrome bits have the form:

$$z_\mu = \xi_{j_1} \oplus \cdots \oplus \xi_{j_K} \oplus \zeta_{l_1} \oplus \cdots \oplus \zeta_{l_L}, \tag{C.6}$$

where signal bits ξ_j are randomly drawn with probability $P(\xi = 1) = p_\xi$ and noise bits ζ_l are drawn with probability $P(\zeta = 1) = p$.

The probability $p_z^0(K, L)$ of $z_\mu = 0$ is therefore:

$$p_z^0(K, L) = p_z^0(K)p_z^0(L) + p_z^1(K)p_z^1(L)$$
$$= 1 - p_z^1(K) - p_z^1(L) + 2p_z^1(K)p_z^1(L). \tag{C.7}$$

where $p_z^x(K)$ and $p_z^0(L)$ stand for probabilities involving the K signal bits and L noise bits, respectively.

By plugging Eq. (C.5) into Eq. (C.7), we get:

$$p_z^1(K, L) = 1 - p_z^0(K, L)$$
$$= \frac{1}{2} - \frac{1}{2}(1 - 2p_\xi)^K (1 - 2p)^L. \tag{C.8}$$

2. *Replica Theory*

For MN codes the replicated partition function has the following form:

$$\langle \mathcal{Z}^n \rangle_{\mathcal{A},\xi,\zeta} = \sum_{\boldsymbol{S}^1,\ldots,\boldsymbol{S}^n} \sum_{\boldsymbol{\tau}^1,\ldots,\boldsymbol{\tau}^n} \prod_{j=1}^{N} \left\langle \exp\left(F_s \xi \beta \sum_{\alpha=1}^{n} S_j^\alpha \right) \right\rangle_\xi$$
$$\times \prod_{l=1}^{M} \left\langle \exp\left(F_n \zeta \beta \sum_{\alpha=1}^{n} \tau_l^\alpha \right) \right\rangle_\zeta$$
$$\times \left\langle \prod_{\langle jl \rangle} \prod_{\alpha=1}^{n} \exp\left[\beta\gamma \mathcal{A}_{\langle jl \rangle} \left(S_{j_1}^\alpha \cdots S_{j_K}^\alpha \tau_{l_1}^\alpha \cdots \tau_{l_L}^\alpha - 1 \right) \right] \right\rangle_{\mathcal{A}}. \tag{C.9}$$

By introducing the Nishimori condition $\beta = 1$ and $\gamma \to \infty$ and computing traces:

$$\exp(-2\beta\hat{x}_{\mu k}) = \frac{\prod_{j\in\mathcal{L}(\mu)\backslash k}\sum_{\tau=\pm1} e^{x_{\mu j}\tau} - \mathcal{J}_\mu \prod_{j\in\mathcal{L}(\mu)\backslash k}\sum_{\tau=\pm1}\tau e^{x_{\mu j}\tau}}{\prod_{j\in\mathcal{L}(\mu)\backslash k}\sum_{\tau=\pm1} e^{x_{\mu j}\tau} + \mathcal{J}_\mu \prod_{j\in\mathcal{L}(\mu)\backslash k}\sum_{\tau=\pm1}\tau e^{x_{\mu j}\tau}}$$
$$= \frac{1 - \mathcal{J}_\mu \prod_{j\in\mathcal{L}(\mu)\backslash k}\tanh(x_{\mu j})}{1 + \mathcal{J}_\mu \prod_{j\in\mathcal{L}(\mu)\backslash k}\tanh(x_{\mu j})}, \tag{B.17}$$

where we have introduced

$$x_{\mu j} = F + \sum_{\nu\in\mathcal{M}(j)\backslash\mu} \hat{x}_{\nu j}.$$

A brief manipulation of the equation above yields Eq. (131).

Appendix C. MN Codes: Technical Details

1. *Distribution of Syndrome Bits*

In this section we evaluate probabilities p_z^x associated with syndrome bits in MN and Gallager codes.

In the case of Gallager codes, a syndrome bit μ has the form

$$z_\mu = \zeta_{l_1} \oplus \cdots \oplus \zeta_{l_K}, \tag{C.1}$$

where $\zeta \in \{0, 1\}$ and $\oplus$ denotes mod 2 sums. Each bit ζ_1 is randomly drawn with probabilities $P(\zeta = 1) = p$ and $P(\zeta = 0) = 1 - p$. The probability $p_z^0(K)$ of $z_\mu = 0$ equates with the probability of having an even number of $\zeta_l = 1$ in the summation, therefore:

$$p_z^0(K) = \sum_{l\,\text{even}}^{K} \frac{K!}{(K-l)!l!} p^l (1-p)^{K-l}$$
$$= \sum_{l\,\text{even}}^{K} (-1)^l \frac{K!}{(K-l)!l!} p^l (1-p)^{K-l}. \tag{C.2}$$

Consequently

$$p_z^1(K) = \sum_{l\,\text{odd}}^{K} \frac{K!}{(K-l)!l!} p^l (1-p)^{K-l}$$
$$= -\sum_{l\,\text{odd}}^{K} (-1)^l \frac{K!}{(K-l)!l!} p^l (1-p)^{K-l}. \tag{C.3}$$

We use the identity $e^{x\sigma} = \cosh(x)(1 + \sigma \tanh(x))$, where $\sigma = \pm 1$, to write:

$$\begin{aligned}
\langle \mathcal{Z}^n \rangle_{\mathcal{A},\xi,\zeta} = & \sum_{\boldsymbol{S}^1,\ldots,\boldsymbol{S}^n} \sum_{\boldsymbol{\tau}^1,\ldots,\boldsymbol{\tau}^n} \prod_{j=1}^{N} \left\langle \exp\left(F_s \xi \beta \sum_{\alpha=1}^{n} S_j^\alpha \right) \right\rangle_\xi \\
& \times \prod_{j=1}^{M} \left\langle \exp\left(F_n \zeta \beta \sum_{\alpha=1}^{n} \tau_j^\alpha \right) \right\rangle_\zeta \\
& \times \sum_{\{C_j, D_l\}} \prod_{j=1}^{N} \mathcal{P}_C(C_j) \prod_{l=1}^{M} \mathcal{P}_D(D_l) \\
& \times \frac{1}{\mathcal{N}} \oint \frac{dZ_j}{2\pi i} \frac{1}{Z_j^{C_j+1}} \oint \frac{dY_l}{2\pi i} \frac{1}{Y_l^{D_l+1}} \\
& \times \prod_{\langle \boldsymbol{il} \rangle} \left\{ 1 + \frac{\cosh^n(\beta\gamma)}{e^{n\beta\gamma}} (Z_{i_1} \cdots Z_{i_K} Y_{l_1} \cdots Y_{l_L}) \right. \\
& \left. \times \prod_{\alpha=1}^{n} \left[1 + S_{i_1}^\alpha \cdots S_{i_K}^\alpha \tau_{l_1}^\alpha \cdots \tau_{l_L}^\alpha \tanh(\beta\gamma) \right] \right\}. \qquad \text{(C.12)}
\end{aligned}$$

The product in the replica index α yields:

$$\begin{aligned}
\prod_{\alpha=1}^{n} \left[1 + S_{i_1}^\alpha \cdots S_{i_K}^\alpha \tau_{l_1}^\alpha \cdots \tau_{l_L}^\alpha \tanh(\beta\gamma) \right] = & \sum_{m=0}^{n} \left[\tanh^m(\beta\gamma) \right. \\
& \left. \times \sum_{\langle \alpha_1,\ldots,\alpha_m \rangle} S_{i_1}^{\alpha_1} \cdots S_{i_1}^{\alpha_m} \cdots S_{i_K}^{\alpha_1} \cdots S_{i_K}^{\alpha_m} \tau_{l_1}^{\alpha_1} \cdots \tau_{l_1}^{\alpha_m} \tau_{l_L}^{\alpha_1} \cdots \tau_{l_L}^{\alpha_m} \right], \qquad \text{(C.13)}
\end{aligned}$$

where $\langle \alpha_1, \ldots, \alpha_m \rangle = \{\alpha_1, \ldots, \alpha_m : \alpha_1 < \cdots < \alpha_m\}$.

The product in the multi-indices $\langle \boldsymbol{il} \rangle$ can be computed by observing that the following relation holds in the thermodynamic limit:

$$\prod_{\langle \boldsymbol{il} \rangle} \left(1 + \psi_{\langle il \rangle} \right) = \sum_{m=0}^{mmax} \sum_{\langle \langle il \rangle_1, \ldots, \langle il \rangle_m \rangle} \psi_{\langle il \rangle_1} \cdots \psi_{\langle il \rangle_m} \xrightarrow{N \to \infty} \exp\left[\sum_{\langle \boldsymbol{il} \rangle} \psi_{\langle il \rangle} \right], \qquad \text{(C.14)}$$

with $mmax \sim (N^K M^L)/K!L!$.

By introducing averages over constructions (117) as described in Appendix B.1 we find:

$$\begin{aligned}\langle \mathcal{Z}^n \rangle_{\mathcal{A},\xi,\zeta} = & \sum_{\boldsymbol{S}^1,\ldots,\boldsymbol{S}^n} \sum_{\boldsymbol{\tau}^1,\ldots,\boldsymbol{\tau}^n} \prod_{j=1}^{N} \left\langle \exp\left(F_s \xi \beta \sum_{\alpha=1}^{n} S_j^\alpha \right) \right\rangle_\xi \\ & \times \prod_{j=1}^{M} \left\langle \exp\left(F_n \zeta \beta \sum_{\alpha=1}^{n} \tau_j^\alpha \right) \right\rangle_\zeta \\ & \times \sum_{\{C_j, D_l\}} \prod_{j=1}^{N} \mathcal{P}_C(C_j) \prod_{l=1}^{M} \mathcal{P}_D(D_l) \\ & \times \frac{1}{\mathcal{N}} \sum_{\{\mathcal{A}\}} \prod_{j=1}^{N} \left[\oint \frac{dZ_j}{2\pi i} \frac{1}{Z_j^{C_j+1}} Z_j^{\sum_{\langle j_1=j, j_2,\ldots,j_K, l\rangle} \mathcal{A}_{\langle j_1=j,\ldots,j_K,l\rangle}} \right] \\ & \times \prod_{l=1}^{M} \left[\oint \frac{dY_l}{2\pi i} \frac{1}{Y_l^{D_l+1}} Y_l^{\sum_{\langle j, l_1=l, l_2,\ldots,l_L\rangle} \mathcal{A}_{\langle j, l_1=l,\ldots,l_L\rangle}} \right] \\ & \times \prod_{\langle \boldsymbol{jl}\rangle} \exp\left[\beta\gamma \mathcal{A}_{\langle \boldsymbol{jl}\rangle} \sum_{\alpha=1}^{n} \left(S_{j_1}^\alpha \cdots S_{j_K}^\alpha \tau_{l_1}^\alpha \cdots \tau_{l_L}^\alpha - 1 \right) \right]. \end{aligned} \tag{C.10}$$

Computing the sum over $\mathcal{A}$ we get:

$$\begin{aligned}\langle \mathcal{Z}^n \rangle_{\mathcal{A},\xi,\zeta} = & \sum_{\boldsymbol{S}^1,\ldots,\boldsymbol{S}^n} \sum_{\boldsymbol{\tau}^1,\ldots,\boldsymbol{\tau}^n} \prod_{j=1}^{N} \left\langle \exp\left(F_s \xi \beta \sum_{\alpha=1}^{n} S_j^\alpha \right) \right\rangle_\xi \\ & \times \prod_{j=1}^{M} \left\langle \exp\left(F_n \zeta \beta \sum_{\alpha=1}^{n} \tau_j^\alpha \right) \right\rangle_\zeta \\ & \times \sum_{\{C_j, D_l\}} \prod_{j=1}^{N} \mathcal{P}_C(C_j) \prod_{l=1}^{M} \mathcal{P}_D(D_l) \frac{1}{\mathcal{N}} \\ & \times \oint \frac{dZ_j}{2\pi i} \frac{1}{Z_j^{C_j+1}} \oint \frac{dY_l}{2\pi i} \frac{1}{Y_l^{D_l+1}} \\ & \times \prod_{\langle \boldsymbol{il}\rangle} \left\{ 1 + \frac{Z_{i_1} \cdots Z_{i_K} Y_{l_1} \cdots Y_{l_L}}{\mathrm{e}^{n\beta\gamma}} \right. \\ & \left. \times \prod_{\alpha=1}^{n} \exp\left[\beta\gamma \left(S_{i_1}^\alpha \cdots S_{i_K}^\alpha \tau_{l_1}^\alpha \cdots \tau_{l_L}^\alpha \right) \right] \right\}. \end{aligned} \tag{C.11}$$

The variables can be normalized as:

$$\frac{q_{\alpha_1\cdots\alpha_m}}{q_0} \mapsto q_{\alpha_1\cdots\alpha_m} \qquad \frac{r_{\alpha_1\cdots\alpha_m}}{r_0} \mapsto r_{\alpha_1\cdots\alpha_m}. \tag{C.21}$$

By plugging Eqs. (C.17), (C.18), the above transformation into (173) and by using Laplace's method, we obtain:

$$\begin{aligned}
\langle \mathcal{Z}^n \rangle_{\mathcal{A},\xi,\zeta} = \mathrm{Extr}_{q,r,\hat{q},\hat{r}} \Bigg\{ & \exp\Bigg[N\frac{\overline{C}}{K} \sum_{m=1}^{n} \sum_{\langle \alpha_1\cdots\alpha_m \rangle} \mathcal{T}_m q^K_{\alpha_1\cdots\alpha_m} r^L_{\alpha_1\cdots\alpha_m} \\
& - N\overline{C} \sum_{m=1}^{n} \sum_{\langle \alpha_1\cdots\alpha_m \rangle} q_{\alpha_1\cdots\alpha_m} \hat{q}_{\alpha_1\cdots\alpha_m} \\
& - M\overline{L} \sum_{m=1}^{n} \sum_{\langle \alpha_1\cdots\alpha_m \rangle} r_{\alpha_1\cdots\alpha_m} \hat{r}_{\alpha_1\cdots\alpha_m} \Bigg] \\
& \times \prod_{j=1}^{N} \sum_{C_j} \mathcal{P}_C(C_j) \prod_{l=1}^{M} \sum_{D_l} \mathcal{P}_D(D_l) \\
& \times \prod_{j=1}^{N} \left(\frac{C_j!}{\hat{q}_0^{C_j}} \right) \mathrm{Tr}_{\{S_j^\alpha\}} \Bigg[\left\langle \exp\left[F_s \beta \xi \sum_{\alpha=1}^{n} S^\alpha \right] \right\rangle_\xi \\
& \times \oint \frac{dZ_j}{2\pi i} \frac{\exp\left[Z_j \sum_{m=0}^{n} \sum_{\langle \alpha_1\cdots\alpha_m \rangle} \hat{q}_{\alpha_1\cdots\alpha_m} S^{\alpha_1} \cdots S^{\alpha_m} \right]}{Z_j^{C_j+1}} \Bigg] \\
& \times \prod_{l=1}^{M} \left(\frac{D_l!}{\hat{r}_0^{D_l}} \right) \mathrm{Tr}_{\{\tau_l^\alpha\}} \Bigg[\left\langle \exp\left[F_n \beta \zeta \sum_{\alpha=1}^{n} \tau_l^\alpha \right] \right\rangle_\zeta \\
& \times \oint \frac{dY_l}{2\pi i} \frac{\exp\left[Y_l \sum_{m=0}^{n} \sum_{\langle \alpha_1\cdots\alpha_m \rangle} \hat{r}_{\alpha_1\cdots\alpha_m} \tau^{\alpha_1} \cdots \tau^{\alpha_m} \right]}{Y_l^{D_l+1}} \Bigg] \Bigg\}
\end{aligned} \tag{C.22}$$

where $\mathcal{T}_m = \mathrm{e}^{-n\beta\gamma} \cosh^n(\beta\gamma) \tanh^m(\beta\gamma)$.

We can rewrite the replicated partition function as:

$$\langle \mathcal{Z}^n \rangle_{\mathcal{A},\xi,\zeta} = \exp\left\{ N \mathrm{Extr}_{q,r,\hat{q},\hat{r}} \left[\frac{\overline{C}}{K} \mathcal{G}_1 - \overline{C}\mathcal{G}_2 - \overline{L}\mathcal{G}_3 + \mathcal{G}_4 + \mathcal{G}_5 \right] \right\} \tag{C.23}$$

We find Eq. (173) by putting Eqs. (C.14) and (C.13) into (C.12) and using the following identities to introduce auxiliary variables:

$$\int dq_{\alpha_1\cdots\alpha_m}\delta\left[q_{\alpha_1\cdots\alpha_m}-\frac{1}{N}\sum_{j=1}^{N}Z_j S_j^{\alpha_1}\cdots S_j^{\alpha_m}\right]=1$$

$$\int dr_{\alpha_1\cdots\alpha_m}\delta\left[r_{\alpha_1\cdots\alpha_m}-\frac{1}{M}\sum_{l=1}^{M}Y_l \tau_l^{\alpha_1}\cdots\tau_l^{\alpha_m}\right]=1 \tag{C.15}$$

3. *Replica Symmetric Free-Energy*

We first compute the normalization $\mathcal{N}$ for a given:

$$\begin{aligned}\mathcal{N}=&\int\left(\frac{dq_0 d\hat{q}_0}{2\pi i}\right)\int\left(\frac{dr_0 d\hat{r}_0}{2\pi i}\right)\\&\times\exp\left[\frac{M^L N^K}{K!L!}\mathcal{T}_0 q_0^K r_0^L-Nq_0\hat{q}_0-Mr_0\hat{r}_0\right]\\&\times\prod_{j=1}^{N}\oint\frac{dZ_j}{2\pi i}\frac{\exp\left[Z_j\hat{q}_0\right]}{Z_j^{C_j+1}}\prod_{l=1}^{M}\oint\frac{dY_l}{2\pi i}\frac{\exp\left[Y_l\hat{r}_0\right]}{Y_l^{D_l+1}}\end{aligned} \tag{C.16}$$

By using Cauchy's integrals to integrate in Z_j and Y_l and Laplace's method, we get:

$$\begin{aligned}\mathcal{N}=\exp\Bigg\{\mathrm{Extr}_{q_0,\hat{q}_0,r_0,\hat{r}_0}\Bigg[&\frac{M^L N^K}{K!L!}\mathcal{T}_0 q_0^K r_0^L-Nq_0\hat{q}_0-Mr_0\hat{r}_0\\&+\sum_{j=1}^{N}\ln\left(\frac{\hat{q}_0^{C_j}}{C_j!}\right)+\sum_{l=1}^{M}\ln\left(\frac{\hat{r}_0^{L_l}}{L_l!}\right)\Bigg]\Bigg\}.\end{aligned} \tag{C.17}$$

The extremization above yields the following equations:

$$q_0\hat{q}_0=\frac{1}{N}\sum_{j=1}^{N}C_j=\overline{C} \tag{C.18}$$

$$r_0\hat{r}_0=\frac{1}{M}\sum_{l=1}^{M}L_l=\overline{L} \tag{C.19}$$

$$q_0^K r_0^L=\overline{C}\frac{(K-1)!L!}{N^{K-1}M^L}. \tag{C.20}$$

$$\mathcal{G}_4(n) = \frac{1}{N}\sum_{j=1}^{N} \ln \sum_{C_j} \mathcal{P}_C(C_j) \left(\frac{C_j!}{\hat{q}_0^{C_j}}\right) \mathrm{Tr}_{\{S^\alpha\}} \left[\left\langle \exp\left[F_s \beta \xi \sum_{\alpha=1}^{n} S_j^\alpha \right] \right\rangle_\xi \right.$$
$$\left. \times \oint \frac{dZ_j}{2\pi i} \frac{\exp\left[Z_j \sum_{m=0}^{n} \sum_{\langle \alpha_1 \cdots \alpha_m \rangle} \hat{q}_{\alpha_1 \cdots \alpha_m} S_j^{\alpha_1} \cdots S_j^{\alpha_m} \right]}{Z_j^{C_j+1}} \right]$$
$$= \frac{1}{N}\sum_{j=1}^{N} \ln \sum_{C_j} \mathcal{P}_C(C_j) \left(\frac{C_j!}{\hat{q}_0^{C_j}}\right) \mathrm{Tr}_{\{S_j^\alpha\}} \left[\left\langle \exp\left[F_s \beta \xi \sum_{\alpha=1}^{n} S_j^\alpha \right] \right\rangle_\xi \right.$$
$$\left. \times \oint \frac{dZ_j}{2\pi i} \frac{\exp\left[Z_j \int d\hat{x}\, \hat{\pi}(\hat{x}) \prod_{\alpha=1}^{n} (1 + S_j^\alpha \hat{x}) \right]}{Z_j^{C_j+1}} \right]$$
$$= \ln \sum_{C_j} \mathcal{P}_C(C_j) \int \prod_{l=1}^{C_j} d\hat{x}_l\, \hat{\pi}(\hat{x}_l) \left[\sum_{S=\pm 1} \left\langle \mathrm{e}^{F_s \beta \xi S} \right\rangle_\xi \prod_{i=1}^{C_j} (1 + S\hat{x}_i) \right]^n \tag{C.29}$$

In the same way:

$$\mathcal{G}_5(n) = \frac{1}{M}\sum_{l=1}^{M} \ln \sum_{D_l} \mathcal{P}_D(D_l) \left(\frac{D_l!}{\hat{r}_0^{D_l}}\right) \mathrm{Tr}_{\{\tau^\alpha\}} \left[\left\langle \exp\left[F_n \beta \zeta \sum_{\alpha=1}^{n} \tau_l^\alpha \right] \right\rangle_\zeta \right.$$
$$\left. \times \oint \frac{dY_l}{2\pi i} \frac{\exp\left[Y_l \sum_{m=0}^{n} \sum_{\langle \alpha_1 \cdots \alpha_m \rangle} \hat{r}_{\alpha_1 \cdots \alpha_m} \tau_l^{\alpha_1} \cdots \tau_l^{\alpha_m} \right]}{Y_l^{D_l+1}} \right]$$
$$= \frac{1}{M}\sum_{l=1}^{M} \ln \sum_{D_l} \mathcal{P}_D(D_l) \left(\frac{D_l!}{\hat{r}_0^{D_l}}\right) \mathrm{Tr}_{\{D_l^\alpha\}} \left[\left\langle \exp\left[F_n \beta \zeta \sum_{\alpha=1}^{n} \tau_l^\alpha \right] \right\rangle_\zeta \right.$$
$$\left. \times \oint \frac{dY_l}{2\pi i} \frac{\exp\left[Y_l \int d\hat{y}\, \hat{\phi}(\hat{y}) \prod_{\alpha=1}^{n} (1 + \tau_l^\alpha \hat{y}) \right]}{Y_l^{D_l+1}} \right]$$
$$= \ln \sum_{D_l} \mathcal{P}_C(D_l) \int \prod_{l=1}^{D_l} d\hat{y}_l\, \hat{\phi}(\hat{y}_l) \left[\sum_{\tau=\pm 1} \left\langle \mathrm{e}^{F_n \beta \zeta \tau} \right\rangle_\zeta \prod_{i=1}^{D_l} (1 + \tau \hat{y}_i) \right]^n \tag{C.30}$$

By using Eq. (166) we can write

$$f = -\frac{1}{\beta} \mathrm{Extr}_{\pi, \hat{\pi}, \phi, \hat{\phi}} \left. \frac{\partial}{\partial n} \right|_{n=0} \left[\frac{\overline{C}}{K} \mathcal{G}_1(n) - \overline{C} \mathcal{G}_2(n) - \overline{L} \mathcal{G}_3(n) + \mathcal{G}_4(n) + \mathcal{G}_5(n) \right], \tag{C.31}$$

what yields free-energy (176).

Introducing the replica symmetric ansätze:

$$q_{\alpha_1\cdots\alpha_m} = \int dx\, \pi(x)\, x^m \qquad \hat{q}_{\alpha_1\cdots\alpha_m} = \int d\hat{x}\, \hat{\pi}(\hat{x})\, \hat{x}^m \tag{C.24}$$

and

$$r_{\alpha_1\cdots\alpha_m} = \int dy\, \phi(y)\, y^m \qquad \hat{r}_{\alpha_1\cdots\alpha_m} = \int d\hat{y}\, \hat{\phi}(\hat{y})\, \hat{y}^m. \tag{C.25}$$

By introducing Nishimori's condition $\gamma \to \infty$ and $\beta = 1$, we can work each term on (C.23) out and find:

$$\begin{aligned}
\mathcal{G}_1(n) &= \mathcal{T}_0 + \mathcal{T}_1 \sum_{\alpha} q_{\alpha}^K r_{\alpha}^L + \mathcal{T}_2 \sum_{\langle\alpha_1\alpha 2\rangle} q_{\alpha_1\alpha_2}^K r_{\alpha_1\alpha_2}^L + \cdots \\
&= \frac{\cosh^n(\beta\gamma)}{\mathrm{e}^{n\gamma\beta}} \int \prod_{j=1}^{K} dx_j \prod_{l=1}^{L} dy_l\, \phi(y_l) \\
&\quad \times \left[1 + \frac{n!}{(n-1)!} \tanh(\beta\gamma) \prod_{j=1}^{K} x_j \prod_{l=1}^{L} y_l \right. \\
&\quad \left. + \frac{n!}{(n-2)!2!} \tanh^2(\beta\gamma) \prod_{j=1}^{K} x_j^2 \prod_{l=1}^{L} y_l^2 + \cdots \right] \\
&= \frac{\cosh^n(\beta\gamma)}{\mathrm{e}^{n\gamma\beta}} \int \prod_{j=1}^{K} dx_j\, \pi(x_j) \prod_{l=1}^{L} dy_l\, \phi(y_l) \left[1 + \tanh(\beta\gamma) \prod_{j=1}^{K} x_j \prod_{l=1}^{L} y_l \right]^n \\
&\xrightarrow{\gamma\to\infty} \frac{1}{2^n} \int \prod_{j=1}^{K} dx_j\, \pi(x_j) \prod_{l=1}^{L} dy_l\, \phi(y_l) \left[1 + \prod_{j=1}^{K} x_j \prod_{l=1}^{L} y_l \right]^n, \qquad \text{(C.26)}
\end{aligned}$$

$$\begin{aligned}
\mathcal{G}_2(n) &= 1 + \sum_{\alpha} q_{\alpha}\hat{q}_{\alpha} + \sum_{\langle\alpha_1\alpha_2\rangle} q_{\alpha_1\alpha_2}\hat{q}_{\alpha_1\alpha_2} + \cdots \\
&= \int dx d\hat{x}\, \pi(x)\hat{\pi}(\hat{x}) \left[1 + x\hat{x}\right]^n. \qquad \text{(C.27)}
\end{aligned}$$

Similarly,

$$\begin{aligned}
\mathcal{G}_3(n) &= 1 + \sum_{\alpha} r_{\alpha}\hat{r}_{\alpha} + \sum_{\langle\alpha_1\alpha_2\rangle} r_{\alpha_1\alpha_2}\hat{r}_{\alpha_1\alpha_2} + \cdots \\
&= \int dy d\hat{y}\, \phi(y)\, \hat{\phi}(\hat{y}) \left[1 + y\hat{y}\right]^n. \qquad \text{(C.28)}
\end{aligned}$$

Since the variance of a Poisson distribution is given by the square root of the mean in the thermodynamic limit:

$$P\left\{\sum_{\langle jl\rangle}\mathcal{A}_{\langle jl\rangle}=x\right\}\stackrel{M\to\infty}{\longrightarrow}\delta\left(x-M\right). \tag{C.40}$$

The Poisson distribution for the construction variables C and L will imply that a fraction $Ne^{-\overline{C}}$ of the signal bits and $Me^{-\overline{L}}$ of the noise bits will be decoupled from the system. These unchecked bits have to be estimate by randomly sampling the prior probability $P(S_j)$, implying that the overlap ρ is upper bounded by:

$$\begin{aligned}\rho &\leq \frac{1}{N}\left[N-Ne^{-\overline{C}}+Ne^{-\overline{C}}(1-2p_\xi)\right]\\ &\leq 1-e^{-\overline{C}}+e^{-\overline{C}}(1-2p_\xi)\\ &\leq 1-2p_\xi e^{-\overline{C}}.\end{aligned} \tag{C.41}$$

Therefore, a VB-like code has necessarily an error-floor that decays exponentially with the $\overline{C}$ chosen.

Acknowledgments

Support by Grants-in-Aid, MEXT (13680400) and JSPS (YK), The Royal Society and EPSRC-GR/N00562 (DS) is acknowledged. We acknowledge the contribution of Tatsuto Murayama to this research effort.

References

Aji, S., and McEliece, R. (2000). The generalized distributive law. *IEEE Trans. Inf. Theory* **46,** 325–343.

Amic, C. D. E., and Luck, J. (1995). Zero-temperature error-correcting code for a binary symmetric channel. *J. Phys. A* **28,** 135–147.

Berger, J. (1993). *Statistical Decision Theory and Bayesian Analysis.* New York: Springer-Verlag.

Berrou, C., Glavieux, A., and Thitimajshima, P. (1993). Near Shannon limit error-correcting coding and decoding: Turbo codex, in *Proc. IEEE Int. Conf. Commun. (ICC),* Geneva, Switzerland. pp. 1064–1070.

Bowman, D., and Levin, K. (1982). Spin-glass in the Bethe approximation: Insights and problems. *Phys. Rev. B* **25,** 3438–3441.

Castillo, E., Gutiérrez, J., and Hadi, A. (1997). *Expert Systems and Probabilistic Network Models.* New York: Springer-Verlag.

Cheng, J.-F. (1997). Iterative decoding. Ph.D. thesis, California Institute of Technology, Pasadena, CA.

4. *Viana–Bray Model: Poisson Constructions*

The Viana-Bray (VB) model is a multispin system with random couplings and strong dilution (Viana and Bray, 1985). We can introduce a VB version of our statistical mechanical formulation for MN codes. The Hamiltonian for a VB-like code is identical to Eq. (160):

$$\mathcal{H}_{\gamma}^{\text{gauge}}(\boldsymbol{S},\boldsymbol{\tau};\boldsymbol{\xi},\boldsymbol{\zeta}) = -\gamma \sum_{\langle \boldsymbol{jl}\rangle} \mathcal{A}_{\langle \boldsymbol{jl}\rangle} \left(S_{j_1} \cdots S_{j_K} \tau_{l_1} \cdots \tau_{l_L} - 1 \right) - F_s \sum_{j=1}^{N} \xi_j S_j - F_n \sum_{l=1}^{M} \zeta_l \tau_l. \tag{C.32}$$

The variables $\mathcal{A}_{\langle \boldsymbol{jl}\rangle}$ are independently drawn from the distribution:

$$P(\mathcal{A}) = \left(1 - \frac{L!K!}{M^{L-1}N^K}\right)\delta(\mathcal{A}) + \frac{L!K!}{M^{L-1}N^K}\delta(\mathcal{A}-1). \tag{C.33}$$

The above distribution will yield the following averages:

$$\left\langle \sum_{\langle \boldsymbol{jl}\rangle} \mathcal{A}_{\langle \boldsymbol{jl}\rangle} \right\rangle_{\mathcal{A}} = M \tag{C.34}$$

$$\left\langle \sum_{\langle j_1 = j \cdots j_K l_1 \cdots l_L\rangle} \mathcal{A}_{\langle \boldsymbol{jl}\rangle} \right\rangle_{\mathcal{A}} = C \tag{C.35}$$

$$\left\langle \sum_{\langle j_1 \cdots j_K l_1 = l \cdots l_L\rangle} \mathcal{A}_{\langle \boldsymbol{jl}\rangle} \right\rangle_{\mathcal{A}} = L. \tag{C.36}$$

In the thermodynamic limit the above summations are random variabels with a Poisson distributions:

$$P\left\{ \sum_{\langle \boldsymbol{jl}\rangle} \mathcal{A}_{\langle \boldsymbol{jl}\rangle} = x \right\} = e^{-M} \frac{M^x}{x!} \tag{C.37}$$

$$P\left\{ \sum_{\langle j_1 = j \cdots j_K l_1 \cdots l_L\rangle} \mathcal{A}_{\langle \boldsymbol{jl}\rangle} = x \right\} = e^{-\overline{C}} \frac{\overline{C}^x}{x!} \tag{C.38}$$

$$P\left\{ \sum_{\langle j_1 \cdots j_K l_1 = l \cdots l_L\rangle} \mathcal{A}_{\langle \boldsymbol{jl}\rangle} = x \right\} = e^{-\overline{L}} \frac{\overline{L}^x}{x!}. \tag{C.39}$$

Khinchin, A. (1957). *Mathematical Foundations of Information Theory.* Dover Publications, Inc., New York, NY.

Kabashima, Y., Murayama, T., and Saad, D. (2000). Typical performance of Gallager-type error-correcting codes. *Phys. Rev. Lett.* **84,** 1355–1358.

Kabashima, Y., Nakamura, K., and van Mourik, J. (June 2001). Statistical mechanics of typical set decoding. cond-mat/0106323.

Kirkpatrick, S., and Sherrington, D. (1978). Infinite-ranged models of spin-glasses. *Phys. Rev. B* **17,** 4384–4403.

Kabashima, Y., and Saad, D. (1998). Belief propagation vs. TAP for decoding corrupted messages. *Europhys. Lett.* **44,** 668–674.

Kabashima, Y., and Saad, D. (1999). Statistical physics of error-correcting codes. *Europhys. Lett.* **45,** 97–103.

Kanter, I., and Saad, D. (1999). Error-correcting codes that nearly saturate Shannon's bound. *Phys. Rev. Lett.* **83,** 2660–2663.

Kanter, I., and Saad, D. (2000a). Cascading parity-check error-correcting codes. *Phys. Rev. E* **61,** 2137–2140.

Kanter, I., and Saad, D. (2000b). Finite-size effects and error-free communication in Gaussian channels. *J. Phys. A* **33,** 1675–1681.

Kabashima, Y., Sazuka, N., Nakamura, K., and Saad, D. (2001). Tighter decoding reliability bound for gallager's error-correcting codes. *Phys. Rev. E* **64,** art. no. 046113.

Lauritzen, S. (1996). *Graphical Models.* Oxford University Press, New York, NY.

Luby, M., Mitzenmacher, M., Shokrollahi, A., and Spielman, D. (1998). Improved low-density parity-check codes using irregular graphs and belief propagation. Digital Systems Research Center, http://www.research.digital.com/SRC/, Tech. Report SRC 1998-009.

MacKay, D. (1995). Free energy minimization algorithm for decoding and cryptanalysis. *Electronics Letters* **31,** 446–447.

MacKay, D. (1999). Good error-correcting codes based on very sparse matrices. *IEEE Trans. on Info. Theory* **45,** 399–431.

MacKay, D. (2000a). Information theory, inference and learning algorithms. Available at http://wol.ra.phy.cam.ac.uk/mackay/.

MacKay, D. (2000b). Relationships between sparse graph codes. *Proceedings of the 2000 Workshop on Information-Based Induction Sciences (IBIS2000).* Izu, Japan, pp. 257–270.

Murayama, T., Kabashima, Y., Saad, D., and Vicente, R. (2000). Statistical physics of regular low-density parity-check error-correcting codes. *Phys. Rev. E* **62,** 1577–1591.

MacKay, D., and Neal, R. (1995). Good codes based on very sparse matrices. *Lecture Notes in Computer Science,* Vol. 1025, Spriger, Berlin, pp. 100–111.

Monasson, R. (1998a). Optimization problems and replica symmetry breaking in finite connectivity spin glasses. *J. Phys. A* **31,** 513–529.

Monasson, R. (1998b). Some remarks on hierarchical replica symmetry breaking in finite-connectivity systems. *Philos. Mag. B* **77,** 1515–1521.

Montanari, A. (2000). Turbo codes: the phase transition. *Eur. Phys. J. B* **18,** 121–136.

Montanari, A. (2001). The glassy phase of gallager codes. *Eur. Phys. J. B* **23,** 121–136.

Mezard, M., and Parisi, G. (2001). The bethe lattice spin glass revisited. *Eur. Phys. J. B* **20,** 217–233.

Mezard, M., Parisi, G., and Virasoro, M. (1987). *Spin Glass Theory and Beyond.* World Scientific Publishing Co., Singapore.

MacWilliams, F., and Sloane, N. (1977). *The Theory of Error-Correcting Codes.* North-Holland, Amsterdam.

Montanari, A., and Sourlas, N. (2000). The statistical mechanics of turbo codes. *Eur. Phys. J. B* **18,** 107–119.

Nelson, M., and Gailly, J. (1995). *The Data Compression Book.* M & T Books, New York, NY.

Chung, S.-Y. (2000). On the construction of some capacity-approaching coding schemes. Ph.D. thesis, Massachusetts Institute of Technology.

Conway, J., and Sloane, N. (1998). *Sphere Packings, Lattices and Groups.* New York: Springer-Verlag.

Cover, T., and Thomas, J. (1991). *Elements of Information Theory.* New York: John Wiley & Sons.

Davey, M. (1998). Record-breaking correction using low-density parity-check codes. Hamilton prize essay, Gonville and Caius College, Cambridge.

Davey, M. (December 1999). Error-correction using low-density parity-check codes. Ph.D. thesis, University of Cambridge.

Derrida, B. (1981). Random-energy model: An exactly solvable model of disordered systems. *Phys. Rev. B* **24,** 2613–2626.

Derrida, B. (1981). Random Energy Model: An exactly solvable model of disordered systems. *Phys. Rev. B* **24,** 238–251.

Dorlas, T., and Wedagedera, J. (1999). Phase diagram of the random energy model with higher order ferromagnetic term and error correcting codes due to Sourlas. *Phys. Rev. Lett.* **83,** 4441–4444.

Feller, W. (1950). *An Introduction to Probability Theory and Its Applications.* Vol. 1, John Wiley & Sons, New York, NY.

Fradkin, E., Huberman, B., and Shenker, S. (1978). Gauge symmetries in random magnetic systems. *Phys. Rev. B* **18,** 4879–4814.

Franz, S., Leone, M., Ricci-Tersenghi, F., and Zecchina, R. (2001). Exact solutions for diluted spin glasses and optimization problems. *Phys. Rev. Lett.* **87,** No. 127209.

Frey, B., and MacKay, D. (1998). A revolution: Belief propagation in graphs with cycles. in *Advances in Neural Information Processing Systems 10,* edited by Jordan, M., Kearns, M., and Solla, S., Cambridge, MA: The MIT Press, pp. 479–485.

Frey, B. (1998). *Graphical Models for Machine Learning and Digital Communication.* Cambridge, MA: MIT Press.

Freeman, Y. W. W. (1999). Correctness of belief propagation in Gaussian graphical models of arbitrary topology. *Tech. Report TR UCB–CSD-99-1046,* UC Berkeley CS Department TR UCB–CSD-99-1046.

Gallager, R. (1962). Low density parity check codes. *IRE Trans. Inf. Theory.* **IT-8,** 21–28.

Gallager, R. (1963). Low-density parity-check codes. *Research Monograph series. No. 21.* Cambridge, MA: MIT Press.

Gross, D., and Mezard, M. (1984). The simplest spin glass. *Nucl. Phys. B* **240,** 431–452.

Goldschmidt, Y. (1991). Spin glass on the finite-connectivity lattice: The replica solution without replicas. *Phys. Rev. B* **43,** 8148–8152.

Gradshteyn, I., and Ryzhik, I. (1994). *Table of Integrals, Series and Products.* London: Academic Press, *London*.

Gujrati, P. (1995). Bethe or Bethe-like lattice calculations are more reliable than conventional mean-field calculations. *Phys. Rev. Lett.* **74,** 809–812.

Hamming, R. (1950). Error detecting and error correcting codes. *Bell Sys. Tech. J.* **26,** 147–160.

Hill, R. (1986). *A First Course in Coding Theory.* Oxford, Claredon Press.

Huffman, D. (1952). A method for construction of minimum redundancy codes. *Proc. IRE* **40,** 1098–1101.

Iba, Y. (1999). The Nishimori line and Bayesian statistics. *J. Phys. A* **32,** 3875–3888.

Jensen, F. (1996). *An Introduction to Bayesian Networks,* London, UCL Press.

Kschischang, F., and Frey, B. (1998). Iterative decoding of compound codes by probability propagation in graphical models. *IEEE J. Selected Areas Commun.* **2,** 153–159.

Kschischang, F., Frey, B., and Loeliger, H.-A. (2001). Factor graphs and the sum-product algorithm. *IEEE Trans. Inf. Theory* **47,** 498–519.

Nishimori, H. (1980). Exact results and critical properties of the Ising model with competing interactions. *J. Phys. C* **13,** 4071–4076.

Nishimori, H. (1993). Optimal decoding for error-correcting codes. *J. Phys. Soc. Japan* **62,** 2973–2975.

Nishimori, H. (2001). *Statistical Physics of Spin Glasses and Information Processing.* Oxford University Press, Oxford, UK.

Nakamura, K., Kabashima, Y., and Saad, D. (2001). Statistical mechanics of low-density parity check error-correcting codes over galois fields. *Europhys. Lett.* **56,** 610–616.

Opper, M., and Saad, D. (2001). *Advanced Mean Field Methods: Theory and Practice.* MIT Press, Cambridge, MA.

Parisi, G. (1980). The order parameter for spin glasses: a function on the interval 0-1, *J. Phys. A* **13,** 1101–1112.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems.* Morgan Kaufmann Publishers, Inc., San Francisco, CA.

Plefka, T. (1982). Convergence condition of the TAP equation for the infinite-ranged ising spin glass model. *J. Phys. A* **15,** 1971–1978.

Rieger, H., and Kirkpatrick, T. (1992). Disordered p-spin interaction models on Husimi trees. *Phys. Rev. B* **45,** 9772–9777.

Richardson, T., Shokrollahi, A., and Urbanke, R. (2001). Design of prvably good low-density parity check codes. *IEEE Trans. Inf. Theory* **47,** 619–637.

Richardson, T., and Urbanke, R. (2001). The capacity of low-density parity check codes under message-passing decoding. *IEEE Trans. Inf. Theory* **47,** 599–618.

Ruján, P. (1993). Finite temperature error-correcting codes. *Phys. Rev. Lett.* **70,** 2968–2971.

Saakian, D. (1998). Diluted generalized random energy model. *JETP Lett.* **67,** 440–444.

Shannon, C. (1948). Mathematical theory of communication. *Bell Sys. Tech. J.* **27,** (pt. I) 379–423 (pt. II) 623–656.

Saul, L., and Jordan, M. (1998). Exploiting tractable substructures in intractable, in *Advances in Neural Information Processing Systems 10* edited by Touretzky, D., Mozer, M., and Hasselmo, M. E., Cambridge, MA: MIT Press, pp. 479–485.

Sourlas, N. (1989). Spin-glass models as error-correcting codes. *Nature* **339,** 693–695.

Sourlas, N. (1994). Spin-glasses, error-correcting codes and finite-temperature decoding. *Europhys. Lett.* **25,** 159–164.

Sourlas, N. (1994). Statistical mechanics and error-correcting codes, in *From Statistical Physics to Statistical Inference and Back* edited by Grassberger, P., and Nadal, J.-P., NATO ASI Series, Vol. 428, Kluwer Academic Publishers. pp. 195–204.

Tanaka, T. (2000). Information geometry of mean field approximation. *Neural Computation* **12,** 1951–1968.

Toulouse, G. (1977). Theory of the frustration effect in spin glasses: I. *Commun. Phys.* **2,** 115–119.

Viana, L., and Bray, A. (1985). Phase diagrams for dilute spin glasses. *J. Phys. C* **18,** 3037–3051.

van Mourik, J., Saad, D., and Kabashima, Y. (2001). Weight vs. magnetization enumerator for gallager codes, in *Cryptography and Coding, 8-th IMA International Conference* (Berlin, Germany) (Honary, B., ed.), Springer, pp. 148–157.

Viterbi, A., and Omura, J. (1979). *Principles of Digital Communication and Coding.* McGraw-Hill Book Co., Singapore.

Vicente, R., Saad, D., and Kabashima, Y. (1999). Finite-connectivity systems as error-correcting codes. *Phys. Rev. E* **60,** 5352–5366.

Vicente, R., Saad, D., and Kabashima, Y. (2000). Error-correcting code on a cactus: a solvable model. *Europhys. Lett.* **51,** 698–704.

Vicente, R., Saad, D., and Kabashima, Y. (2000). Statistical mechanics of irregular low-density parity-check codes. *J. Phys. A* **33,** 6527–6542.

Weiss, Y. (1997). Belief propagation and revision in networks with loops. *Tech. Report A.I. Memo 1616,* MIT.

Wiberg, N. (1996). Codes and decoding on general graphs. Ph.D. thesis, Dep. of Electrical Engineering, Linköping University.

Wong, K., and Sherrington, D. (1987). Graph bipartitioning and spin glasses on a random network of fixed finite valence. *J. Phys. A* **20,** L793–L799.

Wong, K., and Sherrington, D. (1987). Graph bipartitioning and the Bethe spin-glass. *J. Phys. A* **20,** L785–L791.

Wong, K., and Sherrington, D. (1988). Intensively connected spin glasses: towards a replica-symmetry-breaking solution of the groud state. *J. Phys. A* **21,** L459–L466.