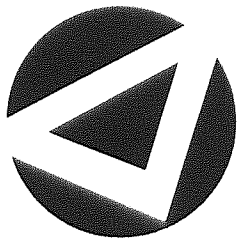
Statistical Physics of Error-Correcting Codes

RENATO VICENTE

Doctor of Philosophy



ASTON UNIVERSITY

October 2000

ASTON UNIVERSITY

Statistical Physics of Error-Correcting Codes

RENATO VICENTE

Doctor of Philosophy, 2000

Thesis Summary

In this thesis we use statistical physics techniques to study the typical performance of four families of error-correcting codes based on very sparse linear transformations: Sourlas codes, Gallager codes, MacKay-Neal codes and Kanter-Saad codes. We map the decoding problem onto an Ising spin system with many-spins interactions. We then employ the replica method to calculate averages over the quenched disorder represented by the code constructions, the arbitrary messages and the random noise vectors. We find, as the noise level increases, a phase transition between successful decoding and failure phases. This phase transition coincides with upper bounds derived in the information theory literature in most of the cases. We connect the practical decoding algorithm known as probability propagation with the task of finding local minima of the related Bethe free-energy. We show that the practical decoding thresholds correspond to noise levels where suboptimal minima of the free-energy emerge. Simulations of practical decoding scenarios using probability propagation agree with theoretical predictions of the replica symmetric theory. The typical performance predicted by the thermodynamic phase transitions is shown to be attainable in computation times that grow exponentially with the system size. We use the insights obtained to design a method to calculate the performance and optimise parameters of the high performance codes proposed by Kanter and Saad.

Keywords: Information theory, Ising systems, coding, approximate inference, statistical physics

Acknowledgements

I would like to thank Professor David Saad and Professor Yoshiyuki Kabashima for the very productive collaboration and helpful supervision.

I would like also to thank David Barber for many discussions and for sharing with me some of his knowledge on probability propagation algorithms and graphical models.

I am grateful to all the nice people in the Neural Computing Research Group for providing a truly stimulating environment. I would like to thank the **NCRG**, specially Professor David Saad, for the financial support.

I would like also to thank my wife Samira for her love and for sharing with me these (difficult) years.

I am grateful to my family for making it all possible.

I wish to thank Nestor Caticha, Osame Kinouchi and Mauro Copelli for all their advice, opportunities and friendship.

For making my life in B'ham easier I would like to thank my new friends from all parts of the known (and unknown) Universe: Mehdi, Ragnar, Lehel, Wei Lee, David Evans, Dörthe, Ernest, Sun Yi, Lars, Silvia, Francesco, Vicky and the french guys in the MSc.

I also would like to thank my Brazilian friends Marcelo, Zé, Luiz, Luciana, Claudio, Silvia, Josué, Cristiano, Roberta and Rafael always very patient and supportive.

Finally, I would like to thank **FAPESP** for the financial support in the early stages of my PhD research.

Contents

List of Figures

Declaration

This thesis describes the work carried out between March 1998 and October 2000 in the Neural Computing Research Group at Aston University under the supervision of Professor David Saad in collaboration with Professor Yoshiyuki Kabashima from the Tokyo Institute of Technology.

Most of the work reported in this thesis has been entirely executed by myself. Part of the work reported in Chapter 5 was done in collaboration with Tatsuto Murayama from the Tokyo Institute of Technology.

This thesis has been composed by myself and has not, nor any similar dissertation, submitted in any previous application for a degree.

Chapter 1

Introduction

1.1 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 usually the messages to be transmitted (*e.g.* images, English text, computer programs) are redundant and the media used for transmission (*e.g.* deep-space, atmosphere, optical fibres, etc...) are noisy. The key issues are to save space and time by eliminating redundancies (*source coding* or *compression*) and to make transmissions reliable by error correction (*channel coding*). Shannon was the first to identify these key issues in his very influential 1948 papers [Sha48]. He did not solve the practical problems but was able to prove general results showing the natural limits of compression and error-correction, he also set up a new framework that gave birth to the information theory.

The surprising fact that error-free communication is possible if the message is encoded to include a minimum amount of redundancy is the content of Shannon's channel coding theorem. He proved that a message encoded at rates R (message information content/code-word length) up to the channel capacity C_{channel} can be decoded with a probability of error that decays exponentially with the message length. His proof was non-constructive and assumed encoding with unstructured random codes and impractical (non-polynomial time) [CT91] decoding schemes. After Shannon's papers the issue of finding practical codes capable of reaching the theoretical limit became a central problem 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 idea. On the top left of Fig. 1.1 we represent the space of words (a message is a sequence of words), 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 a drift in the original word location. The circles

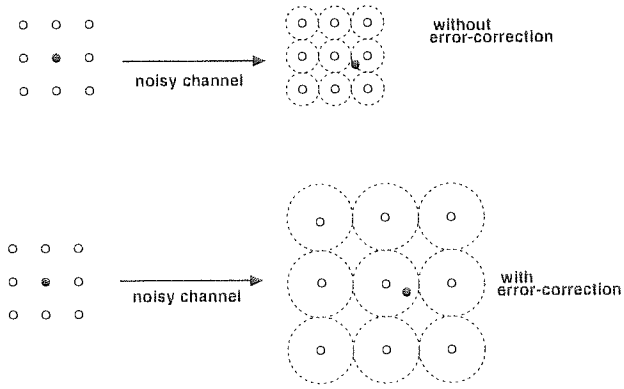


Figure 1.1: Geometry of error-correction. In the top figure we represent what happens with a word transmitted without error-correction. White circles represent possible word vectors, the black circle represents the word to be sent. The channel noise causes corruption of the original word that is represented 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 the action of an error-correcting code. The redundant information changes the space geometry, increasing the distance between words. The same drift of the top figure is not sufficient to cause a transmission error.

around each word represent spheres that provide a decision boundary for each particular word, any signal inside a particular decision region is recognised as representing the word in the center of the sphere. In the case depicted in Fig. 1.1 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 a way that the typical distances between encoded words (codewords) increase. Note that the code rate R measures the size of the codewords space. In the bottom figure we show what happens if the original space is transformed. In this case the same drift by noise shown in the top of Fig. 1.1 is not sufficient to push the received signal outside the decision boundary of the transmitted codeword.

Based on this geometrical 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 a large number of codewords (for a large variety of words) and decision spheres must be as large as possible (for large error-correction capability). Therefore, the general coding problem consists of optimising 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 sphere packing. This *sphere packing problem* is part of the famous list of problems created by Hilbert (it is actually part of the 18th problem). This problem can be solved for a very limited number of dimensions [CS98], but it is very difficult in general. As a consequence, there are constructive procedures only for a limited number of small codes.

For a long time the best practical codes were Reed-Solomon codes (RS) operating in combination with convolutional codes (*concatenated codes*). RS codes, proposed in 1960, are the current techno-

logical standard being found almost everywhere, from compact disks to mobile phones and digital television. Concatenated codes are the present standard in deep-space scientific missions (e.g. Galileo mission) [MS77, OO79]. Recently, *Turbo codes* [BGT93] 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 [Wib96]. A brief description convolutional and Turbo codes is provided in Appendix A.

Despite the success of concatenated and Turbo codes the current practical performance record is owned by the conceptually much simpler Gallager codes [Dav99, Dav98] that ironically are linear block codes, the first family of codes ever proposed. More specifically, they are low-density parity-check codes. Gallager codes were first proposed in 1962 [Gal62, Gal63] 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 a first idea of how parity-check codes operate, we exemplify with the simplest code of this type known as *Hamming code* [Ham50] . A $(7,4)$ Hamming code, where $(7,4)$ stands for the number of bits in the output and input 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:

$$\boldsymbol{G} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \end{pmatrix}. \tag{1.1}$$

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 $\boldsymbol{t} = \boldsymbol{G}\boldsymbol{s}$ composed of 4 message bits plus redundant information (*parity-check*) contained in 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 $\boldsymbol{H}\boldsymbol{G} = 0$. For the generator (1.1) the parity check matrix has the following form:

$$\boldsymbol{H} = \begin{pmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{pmatrix}. \tag{1.2}$$

The decoding procedure follows from the observation that the received message is corrupted by noise as $\boldsymbol{r} = \boldsymbol{G}\boldsymbol{s} \oplus \boldsymbol{n}$. By applying the parity-check matrix we get the *syndrome* $\boldsymbol{H}\boldsymbol{r} = \boldsymbol{H}\boldsymbol{n} = \boldsymbol{z}$. 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 $\boldsymbol{n} = (0,0,1,0,0,0,0)$, $\boldsymbol{z} = (0,1,1)$). Due to this nice property

decoding is trivial and this code is known as a perfect single-error-correcting code [Hil86].

Codes in the low-density parity-check family work along the same principles as the simple Hamming code above, the differences are that they are much larger, the parity-check matrix is very sparse and more than a single error can be corrected. As they are not perfect codes, the decoding problem is much more difficult, but the sparseness of the matrix allows decoding by probability propagation methods similar to those employed in Turbo codes. In this thesis we will concentrate in low-density parity-check codes (LDPC) that are state-of-the art-codes in terms of their performance and, at the same time, operate along simple principles. We will study four variations of LDPCs known as *Sourlas codes*, *Gallager codes*, *MacKay-Neal codes* and *Kanter-Saad codes*.

1.2 Statistical physics of coding

The history of application of statistical physics to error-correcting codes started in 1989 with a paper in Nature by Sourlas relating error-correcting codes to spin glass models [Sou89]. He showed that the Random Energy model [Der81b, Saa98, DW99] 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 [KS78] could operate as a practical code.

In 1995 convolutional codes were analysed by employing the transfer-matrix formalism and power series expansions [AL95].

In 1998 Sourlas work was extended for the case of finite code rates [KS99a, VSK99] by employing the replica method. Recently also Turbo codes were analysed using the replica method [MS99, Mon00].

In this thesis we present the extension of Sourlas work together with analysis of other codes in the low-density parity-check family. We rely on replica calculations [KMS00, MKSV00, VSK00c] as well as mean-field methods [KS98, VSK00a]. The main idea is to develop the application of statistical physics tools to analyse error-correcting codes. Many of the results obtained are rederivations of well known results of information theory, while others put known results into a new perspective.

The main differences between 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 are able to 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.

1.3 Outline

In the next chapter we provide an overview of results and ideas from information theory that are relevant for the understanding of the forthcoming chapters. 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 Chapter 3, we investigate the performance of error-correcting codes based on sparse generator matrices proposed by Sourlas. We employ replica calculations to present the phase diagram for the system at finite code rates. We then discuss the decoding dynamics of the probability propagation algorithm. We regard Sourlas codes as a first step towards developing techniques to analyse other more practical codes.

Chapter 4 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 the 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 Chapter 5 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 non-zero elements in each row and a second for the noise with L non-zero 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 the codes proposed by Kanter and Saad is provided in Chapter 6. We use this description to optimise the construction parameters of a simple code of this type.

Conclusions and perspectives for future work are discussed in Chapter 7.

Five appendices with technical details on the material presented are provided in the end of this thesis.

Chapter 2

Coding and Statistical Physics

In this chapter we provide an overview of results and ideas from information theory that are relevant for the following chapters. We introduce linear encoding and parity-check decoding. We present the probability propagation algorithm for efficiently computing approximate marginal posterior probabilities and discuss the relationship between statistical physics and the decoding problem.

2.1 Mathematical model for a communication system

In his 1948 papers [Sha48] Shannon introduced a mathematical model (pictorially represented in Fig. 2.1) describing the common features present in communication systems, he also identified key problems and proved general results about them. Among Shannon's most celebrated results are the source and channel coding theorems. In the following subsections we will introduce the components of Shannon's communication model as well as the mathematical objects and related general theorems.

2.1.1 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*). For example, a discrete source transmitting texts in Portuguese can be described by an alphabet containing about 76 symbols with a probability measure defined by the Portuguese semantics and grammatical rules. Genomes can be described as discrete sources with alphabet $\{A, G, C, T\}$ and probability measure defined by biochemical functions. In general any discrete alphabet can be mapped onto sequences of $\lceil \log m \rceil$ Boolean digits $\{0, 1\}$. Continuous sources can always be made discrete at the expense of introducing some distortion to the signal [CT91]. A source is *memoryless* if each symbol in the sequence is independent of the preceding and succeeding symbols. The data sink is simply the receiver of decoded messages.

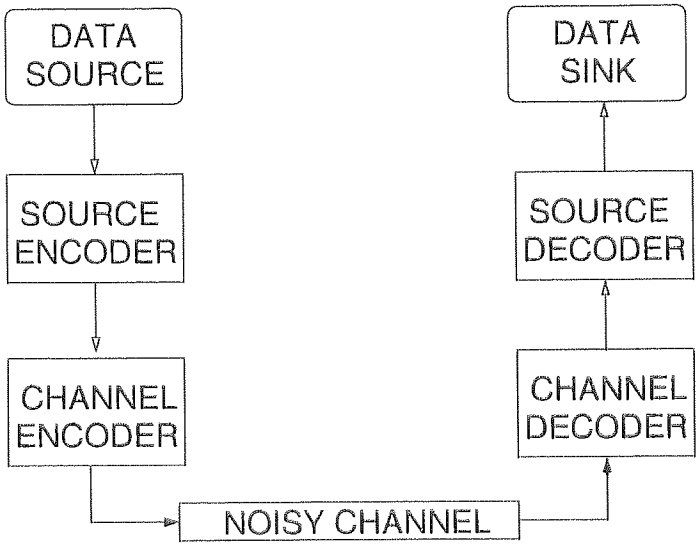


Figure 2.1: Mathematical model for a communication system. The role of each of the components is discussed in the text.

2.1.2 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 more economical representation $\mathcal{A}$.

In his seminal 1948 paper, Claude Shannon borrowed the idea of *entropy* from statistical physics and defined a quantity that can measure the essential information content of a message. As enunciated by Khinchin [Khi57], the entropy of Shannon is defined as follows:

Definition 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

$$H_2(A) = -\sum_{j=1}^{m} p_j \log_2 p_j. \tag{2.1}$$

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

Shannon [Sha48] was able to pose and prove a theorem that establishes what is the maximal shortening of a message by compression as a function of the 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 [Khi57] the theorem states:

Theorem 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 these bounds, given that the statistical structure π of the source is known, was found in 1952 by Huffman [Huf52]. Several practical algorithms are currently known and the design of more efficient and robust schemes still is a very active area [NG95].

2.1.3 Noisy channels

Message corruption during transmission can be described by a probabilistic model defined by the conditional probability $P(r \mid t)$ where t and r represent transmitted and received messages respectively. One can assume that in each channel use only one component t_j, $j = 1, \cdots, M$ of the original message is sent, if there is no interference effects between different components, the channel is called *memoryless* and the conditional probability factorises as $P(r \mid t) = \prod_{j=1}^{M} P(r_j \mid t_j)$.

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:

$$H_2(T \mid R) = \sum_{r \in \mathcal{R}} P(r) \left[\sum_{t \in \mathcal{T}} P(t \mid r) \log_2 \left(\frac{1}{P(t \mid r)} \right) \right]. \tag{2.2}$$

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 (2.1). The maximal information that the channel can retain defines the *channel capacity* [CT91].

Definition 2 (Channel capacity)

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

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.

In this thesis we will be interested in the following channel models (see [Mac99, Mac00a]):

Definition 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*

$$P(r \neq t \mid t) = p \qquad P(r = t \mid t) = 1 - p. \tag{2.3}$$

The channel capacity of a BSC is given by

$$\begin{aligned} C_{BSC} &= 1 - H_2(p) \\ &= 1 + p\ log\ (p) + (1-p)\ log\ (1-p) \end{aligned} \tag{2.4}$$

Definition 4 (Gaussian channel) *Input and output alphabets are the real numbers* $\mathcal{T} = \mathcal{R} = \mathbb{R}$ *and the channel is described by*

$$P(r \mid t) = \frac{1}{\sqrt{2\pi\sigma^2}}\ exp\left[\frac{-(r-t)^2}{2\sigma^2}\right]. \tag{2.5}$$

The channel capacity is

$$C_{Gaussian} = \frac{1}{2}\ log\left(1 + \frac{S}{N}\right), \tag{2.6}$$

where $S/N = \langle t^2 \rangle / \sigma^2$ *is the signal to noise ratio and* $\langle t^2 \rangle = \int dt\ P(t)\ t^2$.

Definition 5 (Binary Gaussian channel) *The input alphabet is binary* $\mathcal{T} = \{-t_0, t_0\}$ *and the output alphabet are the real numbers* $\mathcal{R} = \mathbb{R}$.

$$P(r \mid t) = \frac{1}{\sqrt{2\pi\sigma^2}}\ exp\left[\frac{-(r-t)^2}{2\sigma^2}\right]. \tag{2.7}$$

The channel capacity is

$$C_{Binary} = -\int dr\ P(r)\ log\ P(r) + \int dr\ P(r \mid t = t_0)\ log\ P(r \mid t = t_0), \tag{2.8}$$

where

$$P(r) = \frac{1}{2}\frac{1}{\sqrt{2\pi\sigma^2}}\left[e^{(r-t_0)^2/(2\sigma^2)} + e^{(r+t_0)^2/(2\sigma^2)}\right]. \tag{2.9}$$

2.1.4 Channel encoder and decoder

Shannon was the first to show the surprising fact that highly reliable communication is possible even through noisy channels. He showed that it is possible to protect a message by adding redundant information into the transmission, the operation is carried out by the channel encoder defined as:

Definition 6 ($(2^N, M)$ 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 *that are inputs of the noisy channel. 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 N bits Boolean representations of 2^N symbols onto Boolean sequences of M bits. In Fig. 2.2 we show the codebook for the Hamming code defined by (1.1) 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 g is a map of a channel output $r \in \mathcal{R}$ back into a codeword. The probability that a symbol i is decoded incorrectly is given by the *probability of block error*:

$$p_{\text{Block}} = P\{g(r) \neq i \mid t = t(i)\}. \tag{2.10}$$

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 2.2: Codebook for the $(7,4)$ Hamming code defined by (1.1).

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

$$p_b = \frac{1}{N}\sum_{j=1}^{N} P\{\widehat{s}_j \neq s_j\}. \tag{2.11}$$

Shannon's coding theorem can be enunciated as follows [CT91, Mac00a]:

Theorem 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 actually 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

$$p_b(R) = H_2^{-1}\left(1 - \frac{C_{channel}}{R}\right). \tag{2.12}$$

This theorem is non-constructive, it is obtained by assuming ensembles of random codes and non-practical 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. In fact, as Shannon's proof does not deal with complexity issues, there is no guarantee that such practical scheme exists at all.

2.2 Linear error-correcting codes and the decoding problem

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

$$\boldsymbol{t} = G\boldsymbol{s} \pmod 2, \tag{2.13}$$

where G is an $M \times N$ Boolean matrix. The received message $\boldsymbol{r} = \boldsymbol{t} + \boldsymbol{n}$ is a version of the transmitted message corrupted by the noise vector $\boldsymbol{n}$. In the simplest form, optimal decoding consists of finding

an optimal estimate $\widehat{s}(r)$ assuming a model for the noisy channel $P(r \mid t)$ and a prior distribution for the message source $P(s)$.

The definition of the optimal estimator depends on the particular task and loss function assumed. An optimal estimator is defined in general as follows (see [Iba99] and references therein):

Definition 7 (Optimal estimator) *An optimal estimator $\widehat{s}(r)$ for a loss function $L(s, \widehat{s}(r))$ is an estimator that minimises the average of L in relation to the posterior distribution $P(s \mid r)$.*

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

$$P(s \mid r) = \frac{P(r \mid t)\, \delta\left(t; Gs\right) P(s)}{\sum_s P(r \mid t)\, \delta\left(t; Gs\right) P(s)}, \tag{2.14}$$

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

If in our task we only can accept messages that are totally correct (i.e. we are interested in minimising the probability of block error p_{Block}) we have to assume a loss function that indicates single bit mismatches as an error measure:

$$L(s, \widehat{s}(r)) = 1 - \prod_{j=1}^{M} \delta(s_j; \widehat{s}_j). \tag{2.15}$$

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

$$\begin{aligned} \langle L(s, \widehat{s}(r)) \rangle_{P(s|r)} &= \sum_s P(s \mid r) L(s, \widehat{s}(r)) \\ &= 1 - \sum_s P(s \mid r) \prod_{j=1}^{M} \delta(s_j; \widehat{s}_j) \\ &= 1 - P(\widehat{s} \mid r). \end{aligned} \tag{2.16}$$

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

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

$$L(s, \widehat{s}(r)) = -\sum_{j=1}^{M} s_j \widehat{s}_j, \tag{2.17}$$

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

$$\langle L(s, \widehat{s}(r)) \rangle_{P(s|r)} = -\sum_{j=1}^{M} \langle s_j \rangle_{P(s|r)} \widehat{s}_j. \tag{2.18}$$

An obvious choice for the estimator is

$$\begin{aligned} \widehat{s}_j &= \frac{\langle s_j \rangle_{P(s|r)}}{|\, \langle s_j \rangle_{P(s|r)} \,|} \\ &= \text{sgn}(\langle s_j \rangle_{P(s|r)}) \\ &= \text{argmax}_{s_j} P(s_j \mid r), \end{aligned} \tag{2.19}$$

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 it is suggested by Eq. (2.19), this estimator is often called the *Marginal Posterior Maximiser* or *MPM* for short.

Decoding, namely, the computation of estimators, in general becomes a hard task very quickly as the message size increases. The MAP estimator requires finding a global maximum of the posterior over a space with 2^N possible solutions 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 the task quickly impractical, at least in a naïve brute force evaluation. 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.

2.3 Probability propagation algorithm

The probabilistic dependencies existing in a code can be graphically represented as a bipartite graph [Lau96] 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:

$$\begin{aligned} P(s_j, \boldsymbol{r}) &= \sum_{\{s_i : i \neq j\}} P(\boldsymbol{s} \mid \boldsymbol{r}) P(\boldsymbol{r}) \\ &= \sum_{\{s_i : i \neq j\}} P(\boldsymbol{r} \mid \boldsymbol{s}) P(\boldsymbol{s}) \\ &= \sum_{\{s_i : i \neq j\}} \prod_{\mu=1}^{M} P(r_\mu \mid s_{i_1} \cdots s_{i_K}) \prod_{j=1}^{N} P(s_j), \end{aligned} \tag{2.20}$$

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 distorted message received, or the evidence. Equation (2.20) shows a complex partial factorisation that depends on the structure of the generator matrix $\boldsymbol{G}$. We can encode this complex partial factorisation on a directed graph known as a *Bayesian network* [Pea88, CGH97, Jen96, KF98, AM00, Fre98, KFL98]. As an example, we show in Fig. 2.3 a simple directed bipartite graph encoding the following joint distribution:

$$\begin{aligned} P(s_1, \cdots, s_4, r_1, \cdots, r_6) &= P(r_1 \mid s_1, s_2, s_3) P(r_2 \mid s_3) P(r_3 \mid s_1, s_2) P(r_4 \mid s_3, s_4) \\ &\times P(r_5 \mid s_3) P(r_6 \mid s_3) P(s_1) P(s_2) P(s_3) P(s_4) \end{aligned} \tag{2.21}$$

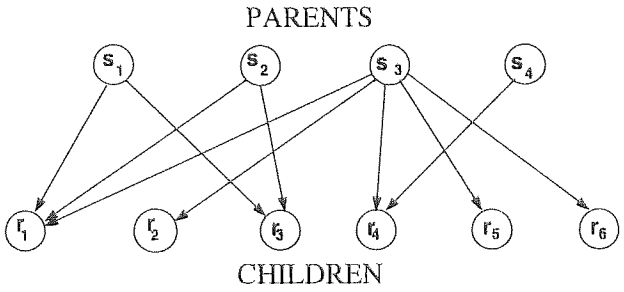


Figure 2.3: Bayesian Network representing a linear code of rate 2/3. If there is an arrow from a vertex s_j to a vertex r_μ, s_j is said to be a *parent* and r_μ is said to be a *child*.

The generator matrix specifying the code in Fig. 2.3 is:

$$G = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}. \tag{2.22}$$

Given the received evidence $\boldsymbol{r}$, an exact evaluation of the marginal joint distribution (2.20) in a space of binary variables $\boldsymbol{s} \in \{\pm 1\}^N$ would require $(N + M)2^{N-1}$ operations. In the eighties Pearl [Pea88] proposed an iterative algorithm that requires $\mathcal{O}(N)$ time to calculate approximate marginal probabilities using Bayesian networks. This algorithm is known as *belief propagation* [Pea88], *probability propagation* [KF98], *generalised distributive law* [AM00] or *sum-product algorithm* [Fre98, KFL98] (see also [SO]).

The probability propagation algorithm is exact when the Bayesian network associated to the particular problem is free of loops (*i.e.* it is a *tree*). To introduce the probability propagation algorithm we start with the simple chain in Fig. 2.4, which represents the following joint distribution:

$$p(s_1, s_2, s_3, s_4, s_5) = p(s_1)p(s_2 \mid s_1)p(s_3 \mid s_2)p(s_4 \mid s_3)p(s_5 \mid s_4). \tag{2.23}$$

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

$$p(s_3) = \sum_{s_1, s_2, s_4, s_5} p(s_1)p(s_2 \mid s_1)p(s_3 \mid s_2)p(s_4 \mid s_3)p(s_5 \mid s_4). \tag{2.24}$$

A brute force evaluation of (2.24) would take $5 \times (2^4) = 80$ operations in a binary field. The probability propagation algorithm reduces significantly the number of operations needed by rationalising the order in which they are performed. For Fig. 2.4 we can start by marginalising out vertex s_5 and writing:

$$R_{54}(s_4) = \sum_{s_5} p(s_5 \mid s_4). \tag{2.25}$$

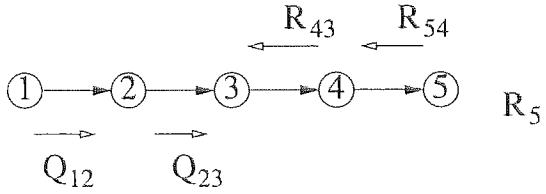


Figure 2.4: Marginal probabilities can be calculated exactly in a Bayesian chain. *R-messages* flow from a child to a parent and *Q-messages* flow from a parent to a child.

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:

$$R_{43}(s_3) = \sum_{s_4} p(s_4 \mid s_3) R_{54}(s_4). \tag{2.26}$$

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 (2.25) in the same form as (2.26) 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 in the left side of s_3. Firstly, we marginalise s_1 by introducing:

$$Q_{12}(s_1) = p(s_1). \tag{2.27}$$

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

$$Q_{23}(s_2) = \sum_{s_1} Q_{12}(s_1) p(s_2 \mid s_1). \tag{2.28}$$

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

$$\begin{aligned} p(s_3) &= \sum_{s_2} Q_{23}(s_2) R_{43}(s_3) p(s_3 \mid s_2) \\ &= \sum_{s_2} \sum_{s_1} Q_{12}(s_1) p(s_2 \mid s_1) \sum_{s_4} p(s_4 \mid s_3) R_{54}(s_4) p(s_3 \mid s_2) \\ &= \sum_{s_2} \sum_{s_1} p(s_1) p(s_2 \mid s_1) \sum_{s_4} p(s_4 \mid s_3) \sum_{s_5} p(s_5 \mid s_4) \\ &= \sum_{s_1, s_2, s_4, s_5} p(s_1) p(s_2 \mid s_1) p(s_3 \mid s_2) p(s_4 \mid s_3) p(s_5 \mid s_4). \end{aligned} \tag{2.29}$$

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 Fig. 2.5 representing the following joint distribution:

$$\begin{aligned} p(s_1, ..., s_{12}) &= p(s_6) p(s_8) p(s_9) p(s_{10}) p(s_{11}) p(s_{12}) p(s_1 \mid s_{10}) p(s_2 \mid s_{11}, s_{12}) \\ &\times \; p(s_3 \mid s_1, s_2, s_9) p(s_4 \mid s_3, s_8) p(s_5 \mid s_3, s_6) p(s_7 \mid s_4). \end{aligned} \tag{2.30}$$

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 $12 \times 2^9 = 6144$ operations.

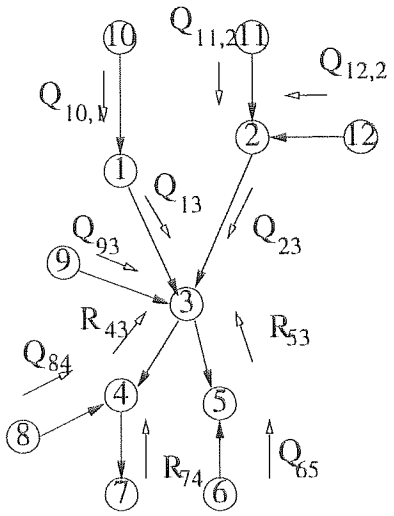


Figure 2.5: Marginal probabilities also can be calculated exactly in a Bayesian tree.

In general we can just initialise the messages with random values or make use of the prior knowledge that may be available and update the vertices in a 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, to say 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 Fig. 2.5 the optimal schedule would be as follows:

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

The Q-messages from parents to children are just the prior probabilities:

$$Q_{j\mu}(s_j) = p(s_j), \tag{2.31}$$

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

The R-message between s_7 and s_4 is:

$$R_{74}(s_4) = \sum_{s_7} R_7(s_7) p(s_7 \mid s_4), \tag{2.32}$$

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:

$$Q_{13}(s_1) = \sum_{s_{10}} p(s_1 \mid s_{10}) Q_{10,1}(s_{10}) \tag{2.33}$$

$$Q_{23}(s_2) = \sum_{s_{11}, s_{12}} p(s_2 \mid s_{11}, s_{12}) Q_{11,2}(s_{11}) Q_{12,2}(s_{12}). \tag{2.34}$$

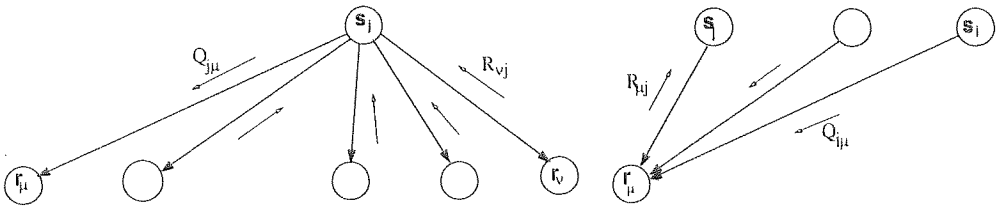


Figure 2.6: Left side: forward (Q) message from parent to child. Right side: backward (R) message from child to parent.

The remaining R-messages are:

$$R_{43}(s_3) = \sum_{s_4,s_8} p(s_4 \mid s_3, s_8) Q_{84}(s_8) R_{74}(s_4) \tag{2.35}$$

$$R_{53}(s_3) = \sum_{s_6,s_5} p(s_5 \mid s_3, s_6) Q_{65}(s_6) R_5(s_5), \tag{2.36}$$

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:

$$p(s_3) = \sum_{s_1,s_2,s_9} p(s_3 \mid s_1, s_2, s_9) Q_{13}(s_1) Q_{23}(s_2) R_{43}(s_3) R_{53}(s_3) Q_{93}(s_9). \tag{2.37}$$

By substituting the expressions for the messages in (2.37) it is relatively straightforward to verify that this expression gives the exact value for the marginal of (2.30). In this case the probability propagation algorithm requires only 432 operations against 6144 operations required by the brute force evaluation.

We can now summarise 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 contained in 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 is a tree (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 fix some definitions. Two vertices s_j and r_μ are adjacent if there is a line 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 (like in Fig. 2.3) 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 forwards from the top layer (parents) to the bottom layer (children), then backwards, and so on iteratively. Child-to-parent messages (backward messages in Fig. 2.3) are R-messages denoted $R_{\mu j}$, while parent-to-child messages (forward messages) are Q-messages denoted by $Q_{j\mu}$.

With the help of Fig. 2.6 using the algorithm above the forward (Q) messages between a parent s_j and child r_μ are just (see also [Dav99]):

$$Q^a_{j\mu} = P(S_j = a \mid \{J_\nu : \nu \in \mathcal{M}(j)\mu\}) \tag{2.38}$$

$$= \alpha_{\mu j}\, p(s_j = a) \prod_{\nu \in \mathcal{M}(j)\setminus\mu} R^a_{\nu j}, \tag{2.39}$$

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:

$$R^a_{\mu j} = \sum_{\{s_i : i \in \mathcal{L}(\mu)\setminus j\}} P(r_\mu \mid s_j = a, \{s_i : i \in \mathcal{L}(\mu) \setminus j\}) \prod_{i \in \mathcal{L}(\mu)\setminus j} Q^{s_i}_{i\mu}. \tag{2.40}$$

An approximation for the marginal posterior can be obtained by iterating Equations (2.38) and (2.40) until convergence or some stopping criteria is attained, and fusing all incoming information to a parent node by calculating:

$$Q^a_j = \alpha_j\, p(s_j = a) \prod_{\nu \in \mathcal{M}(j)} R^a_{\mu j}, \tag{2.41}$$

where α_j is a normalisation 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 [Pea88]) 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 is observed empirically that decoding with the PP algorithm also yields good results [FM98, Che97] in spite of that.

There are a limited number of studies of probability propagation in loopy graphs with a single cycle [Wei97] and describing Gaussian joint distributions [Fre99] but no definite explanation for its good performance in this case is know to date.

2.4 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 favourable. Moreover, it is possible to prove that the probability of a cycle-free path of length l in a random graph decays with $\mathcal{O}(K^l/N)$ (see Appendix F) what indicates that small cycles are harder to find if the generator matrix is very sparse and PP decoding is expected

to provide better approximates 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, what leads to a high probability of undetectable errors. Codes with sparse generator matrices are known as *Sourlas codes* and will be our first object of study in Chapter 3.

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 done in a very sparse graph, allowing efficient use of PP decoding. The method known as parity-check decoding [Hil86, OO79] 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 $\boldsymbol{H}$ with the property $\boldsymbol{H}\boldsymbol{G}$ (mod 2) $= 0$, yielding the *syndrome vector* $\boldsymbol{z} = \boldsymbol{H}\boldsymbol{n}$ (mod 2) .

Decoding reduces to finding $\boldsymbol{n}$ when the syndrome vector $\boldsymbol{z}$ is known, namely, MPM estimates involve the calculation of the marginal posterior $P(n_j \mid \boldsymbol{z})$. In [Mac99] MacKay was able to prove 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 *Kanter-Saad*) we study in Chapters 4, 5 and 6.

2.5 Decoding and statistical physics

The connection between spin systems in statistical physics and digital error correcting codes was first noted by Sourlas [Sou89]. This connection 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:

$$S \cdot X = (-1)^{s \oplus x}, \tag{2.42}$$

where $S, X \in \{+1,-1\}$ and $s, x \in \{0,1\}$. Trough this isomorphism every addition on the Boolean group corresponds to an 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 $\bigoplus_{j=1}^{K} s_j$ what 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 [Sou94b, AL95] and Turbo codes [MS99, Mon00] (see Appendix A).

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:

$$P_{\alpha\gamma}(\boldsymbol{S} \mid \boldsymbol{J}) = \frac{1}{Z(\boldsymbol{J})} \exp\left[\ln P_{\alpha}(\boldsymbol{J} \mid \boldsymbol{S}) + \ln P_{\gamma}(\boldsymbol{S})\right], \tag{2.43}$$

where α and γ are hyper-parameters assumed to describe features like the encoding scheme, source

distribution and noise level. This form suggests the following family of Gibbs measures:

$$P_{\alpha\beta\gamma}(\boldsymbol{S} \mid \boldsymbol{J}) = \frac{1}{Z} \exp\left[-\beta \mathcal{H}_{\alpha\gamma}(\boldsymbol{S}; \boldsymbol{J})\right] \tag{2.44}$$

$$\mathcal{H}_{\alpha\gamma}(\boldsymbol{S}; \boldsymbol{J}) = -\ln P_{\alpha}(\boldsymbol{J} \mid \boldsymbol{S}) - \ln P_{\gamma}(\boldsymbol{S}), \tag{2.45}$$

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 (2.44), i.e. by the sign of thermal averages $\widehat{S}_j^{\mathrm{MAP}} = \mathrm{sgn}(\langle S_j \rangle_{\beta \to \infty})$ at zero temperature. On the other hand the MPM estimator is provided by the sign of thermal averages $\widehat{S}_j^{\mathrm{MPM}} = \mathrm{sgn}(\langle S_j \rangle_{\beta=1})$ at temperature one. We have seen in Section 2.2 that if we are concerned with the probability of bit error p_e the optimal choice for an estimator is MPM, that is equivalent to decoding at finite temperature $\beta = 1$, known as the Nishimori temperature [Nis80, Nis93, Nis00, Ruj93].

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

$$P_{\alpha^*\gamma^*}(\boldsymbol{J}) = \sum_{\boldsymbol{S}} P_{\alpha^*}(\boldsymbol{J} \mid \boldsymbol{S}) P_{\gamma^*}(\boldsymbol{S}), \tag{2.46}$$

where α^* and γ^* represent the "real" hyper-parameters, in other words, the hyper-parameters actually used for generating the evidence $\boldsymbol{J}$. Those "real" hyper-parameters are, in general, not known by 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 hyper-parameters:

$$\langle F(\alpha, \gamma) \rangle_{P_{\alpha^*\gamma^*}} = -\langle \ln P_{\alpha\gamma}(\boldsymbol{J}) \rangle_{P_{\alpha^*\gamma^*}} . \tag{2.47}$$

This log-likelihood can be regarded as measuring the typical plausibility of α and γ, given the data $\boldsymbol{J}$ [Ber93]. This function can be minimised to find the most plausible hyper-parameters (known as *type II maximum likelihood hyper-parameters* or just *ML-II hyper-parameters*) [Ber93].

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

$$\langle F(\alpha^*, \gamma^*) \rangle_{P_{\alpha^*\gamma^*}} \leq \langle F(\alpha, \gamma) \rangle_{P_{\alpha^*\gamma^*}} . \tag{2.48}$$

The proof of (2.48) follows directly from the information inequality [Iba99, CT91], i.e. the non-negativity of the KL-divergence :

$$\begin{aligned} D(P_{\alpha^*\gamma^*} \| P_{\alpha\gamma}) &\geq 0 \\ \left\langle \ln \left(\frac{P_{\alpha^*\gamma^*}(\boldsymbol{J})}{P_{\alpha\gamma}(\boldsymbol{J})} \right) \right\rangle_{P_{\alpha^*\gamma^*}} &\geq 0 \\ -\langle \ln P_{\alpha^*\gamma^*}(\boldsymbol{J}) \rangle_{P_{\alpha^*\gamma^*}} &\leq -\langle \ln P_{\alpha\gamma}(\boldsymbol{J}) \rangle_{P_{\alpha^*\gamma^*}} . \end{aligned} \tag{2.49}$$

When "real" and assumed hyper-parameters agree, we say that we are at the *Nishimori condition* [Iba99, Nis00]. At the Nishimori condition many calculations simplify and can be done exactly (see Appendix C.3 for an example). Through this thesis we will be, most of the time, assuming the Nishimori condition.

Chapter 3

Sourlas Codes

In this chapter we investigate the performance of error-correcting codes based on sparse generator matrices using the mapping onto Ising spin systems proposed by Sourlas. We study codes where each parity-check comprises products of K bits selected from the original digital message with exactly C checks per message bit. We show, using the replica method, that these codes saturate Shannon's coding bound for $K \to \infty$ when the code rate K/C is finite. We then examine the finite temperature case to assess the use of simulated annealing methods for decoding, we study the performance of the finite K case and extend the analysis to accommodate different types of noisy channels. The connection between statistical physics and probability propagation decoders is discussed and the dynamics of the decoding itself is analysed. Further insight into new approaches for improving the code performance is given. The content of this chapter appeared in [VŠK99]

3.1 Introduction

The code of Sourlas is based on the simple idea of using a linear operation $\boldsymbol{G}$ (*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 then $\boldsymbol{t} = \boldsymbol{G}\boldsymbol{s} \pmod 2$, each bit t_k being the Boolean sum of K message bits (*parity check*). This vector is then corrupted when transmitted through a noisy channel and a corrupted M dimensional vector $\boldsymbol{r}$ is received.

Decoding consists of producing an estimate $\widehat{\boldsymbol{s}}$ of the original message. This estimate can be generated by considering a probabilistic model for the encoding/corruption/decoding 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 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 1989 Sourlas [Sou89, Sou94a] proposed that, due to the equivalence between addition over the

field $\{0,1\}$ and multiplication over $\{\pm 1\}$, many error-correcting codes can be mapped onto many-body spin-glasses with appropriately defined couplings. This observation opened the possibility of applying techniques from statistical physics to study coding systems.

In the mapping proposed by Sourlas, a message is represented by a binary vector $\boldsymbol{\xi} \in \{\pm 1\}^N$ encoded to a higher dimensional vector $\boldsymbol{J^0} \in \{\pm 1\}^M$ defined as $J^0_{\langle i_1, i_2 \ldots i_K \rangle} = \xi_{i_1} \xi_{i_2} \ldots \xi_{i_K}$, 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 can be viewed as a statistical process [Iba99] (see Fig.3.1), where an estimate $\widehat{\boldsymbol{\xi}}$ to the original message that minimises a given expected loss $\langle \langle \mathcal{L}(\xi, \widehat{\xi}) \rangle_{P(J|\xi)} \rangle_{P(\xi)}$ averaged over the indicated probability distributions is produced. The definition of the loss depends on the particular task; the overlap $\mathcal{L}(\xi, \widehat{\xi}) = \sum_j \xi_j \widehat{\xi}_j$ can be used for decoding binary messages. As discussed in Section 2.2, an optimal estimator for this particular loss function is $\widehat{\xi}_j = \text{sign}\langle S_j \rangle_{P(S_j|J)}$ [Iba99], where $\boldsymbol{S}$ is an N dimensional binary vector representing the dynamical 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:

$$\ln P(\boldsymbol{S} \mid \boldsymbol{J}) = \ln P(\boldsymbol{J} \mid \boldsymbol{S}) + \ln P(\boldsymbol{S}) + \text{const.} \tag{3.1}$$

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

$$P(\boldsymbol{J} \mid \boldsymbol{S}) = \prod_{\text{chosen sets}} \sum_{J^0_{\langle i_1 \cdots i_K \rangle}} P(J_{\langle i_1 \cdots i_K \rangle} \mid J^0_{\langle i_1 \cdots i_K \rangle}) \, P(J^0_{\langle i_1 \cdots i_K \rangle} \mid \boldsymbol{S}). \tag{3.2}$$

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

$$P(J^0_{\langle i_1 \cdots i_K \rangle} \mid \boldsymbol{S}) = \delta(J^0_{\langle i_1 \cdots i_K \rangle}; S_{i_1} \cdots S_{i_K}). \tag{3.3}$$

The noisy channel is modelled 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 [Sou94a]:

$$\begin{aligned} \ln P(J_{\langle i_1 \cdots i_K \rangle} \mid J^0_{\langle i_1 \cdots i_K \rangle}) &= \frac{1}{2}(1 + J^0_{\langle i_1 \cdots i_K \rangle}) \ln P(J_{\langle i_1 \cdots i_K \rangle} \mid +1) \\ &+ \frac{1}{2}(1 - J^0_{\langle i_1 \cdots i_K \rangle}) \ln P(J_{\langle i_1 \cdots i_K \rangle} \mid -1) \\ &= \text{constant} + \frac{1}{2} \ln \left(\frac{1-p}{p} \right) J_{\langle i_1 \cdots i_K \rangle} \, J^0_{\langle i_1 \cdots i_K \rangle}. \end{aligned} \tag{3.4}$$

Putting everything together one obtains a Hamiltonian for the code of Sourlas as:

$$\ln P(\boldsymbol{S} \mid \boldsymbol{J}) = -\beta_N \, \mathcal{H}(\boldsymbol{S}) \tag{3.5}$$

$$= \beta_N \sum_{\mu} \mathcal{A}_\mu \, J_\mu \prod_{i \in \mathcal{L}(\mu)} S_i + \beta'_N \sum_{j=1}^{N} S_j, \tag{3.6}$$

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 $\sum_{\mathcal{L}(\mu) \setminus i} \mathcal{A}_\mu = C \ \forall i$, which determines the M components of the codeword $\boldsymbol{J^0}$. The temperature of the interaction term used here is $\beta_N = \frac{1}{2} \ln \left(\frac{1-p}{p} \right)$, known as the Nishimori temperature

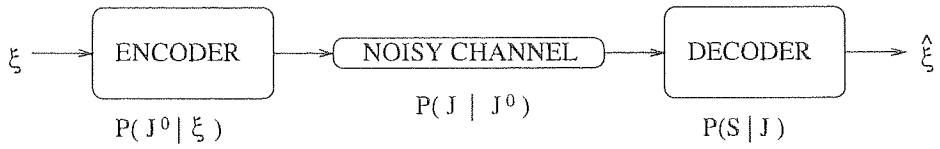


Figure 3.1: The encoding, message corruption in the noisy channel and decoding can be represented as a Markovian process. The aim is to obtain a good estimate $\widehat{\boldsymbol{\xi}}$ for the original message $\boldsymbol{\xi}$.

[Nis80, Iba99, Ruj93, Nis93], and $\beta'_N = \frac{1}{2}\ln\left(\frac{1-p_\xi}{p_\xi}\right)$ 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 (3.5) 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.

In the statistical physics framework the performance of the error-correcting process can be measured by the overlap between actual message and estimate for a given scenario characterised by a 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 realisations (possible corrupted vectors $\boldsymbol{J}$) given the message $\boldsymbol{\xi}$ and then over all possible messages:

$$\rho = \frac{1}{N}\left\langle \sum_{i=1}^{N} \xi_i \ \langle \mathrm{sign}\langle S_i\rangle\rangle_{\mathcal{A},J|\xi} \right\rangle_{\xi} \tag{3.7}$$

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

From the statistical physics point of view, 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 $C \sim \begin{pmatrix} N-1 \\ K-1 \end{pmatrix}$ scales with the system size. In this case it can be mapped onto known problems in statistical physics such as the SK [KS78] (K=2) and Random Energy (REM) [Der81a] ($K \to \infty$) models. It has been shown that the REM saturates Shannon's bound [Sou89]. However, it has a rather limited practical relevance as the choice of extensive connectivity corresponds to a vanishingly small code rate.

Here we present the analysis for the code of Sourlas in the case of finite connectivity where the code rate is finite, extending the analysis in [KS98, KS99a]. We show that Shannon's bound can also be attained at finite code rates. We study the decoding dynamics and discuss the connections between statistical physics and the methods of probability propagation.

This chapter is organised as follows: in Section 3.2 we present a lower bound for the probability of bit error, in Section 3.3 we discuss a naïve mean field approximation that yields unphysical results. Section 3.4 describes the statistical physics treatment of the code of Sourlas showing that Shannon's bound can be attained for finite code rates if $K \to \infty$. The finite K case and the Gaussian noise are

also discussed in Section 3.4. The decoding dynamics is analysed in Section 3.5. Concluding remarks are given in Section 3.6. Appendices with the detailed calculations are also provided.

3.2 Lower bound for the probability of bit error

It was observed in [MS99] 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 [MS99]. 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:

$$\mathcal{H}_j = -S_j \sum_{\mu \in \mathcal{M}(j)} J_\mu \prod_{i \in \mathcal{L}(\mu) \setminus j} S_i, \tag{3.8}$$

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

$$p_b(1) = P\{ \sum_{\mu \in \mathcal{M}(j)} J_\mu \leq 0 \}, \tag{3.9}$$

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. Regarding the sequence $\{J_1, \cdots, J_C\}$ as a random walk we find:

$$p_b \geq \sum_{l \in \mathbb{N}, C-2l \leq 0}^{l \leq C} \frac{C!}{(C-l)!\, l!} (1-p)^{C-l}\, p^l. \tag{3.10}$$

A lower bound for for p_b in the large C regime can be obtained by using the DeMoivre-Laplace limit theorem [Fel50], writing:

$$p_b \geq \frac{1}{2} \mathrm{erfc}\left(\frac{(1-p)C}{8p} \right) \approx \frac{4p}{\sqrt{\pi}(1-p)C} \exp\left(-\frac{(1-p)^2 C^2}{64 p^2} \right), \tag{3.11}$$

where $\mathrm{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty du\, \exp(-u^2)$ and the asymptotic behaviour is given in [GR94] (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 $R = K/C$.

3.3 Naïve mean field

As a very simple start, we consider a message $\xi_j = 1$ for all j (so $m = 1$ corresponds to perfect decoding and we can identify the overlap ρ with the magnetisation m), and perform a naïve mean field theory as a first approximation.

Let us start by writing the Hamiltonian of the system for unbiased messages with $\beta'_N = 0$ as:

$$H(\{S_j\}) = -\sum_{\mu} J_\mu \prod_{i \in \mathcal{L}(\mu)} S_i, \tag{3.12}$$

where $\sum_\mu(...)$ represents a sum over M randomly chosen index sets and $\mathcal{L}(\mu)$ represents K indices in a particular index set μ.

Omitting the dependency on the quenched disorder J_μ we can write:

$$P(\{S_j\}]) = \frac{\mathrm{e}^{-\beta H(\{S_j\})}}{Z}. \tag{3.13}$$

We would like to approximate the above distribution with a more tractable distribution $Q(\{S_j\})$ which is factorisable. The most general totally factorisable distribution in the space of binary vectord $\{\pm 1\}$ is as follows:

$$\begin{aligned} Q(\{S_j\}) &= \prod_j Q_j(S_j) \\ &= \prod_j \frac{(1 + S_j m_j)}{2}. \end{aligned} \tag{3.14}$$

It is clear that the parameters of the approximating distribution are such that $m_j = \langle S_j \rangle_Q$. An even simpler approximation assumes only one parameter to describe the approximating distribution:

$$Q(\{S_j\}) = \prod_j \frac{(1 + S_j m)}{2}. \tag{3.15}$$

The framework for selecting appropriate parameters in the approximation consists in projecting the original distribution P in the space of approximating distributions Q by minimising the KL-divergence between these distributions defined as [SO]:

$$\begin{aligned} D(Q\|P) &= \sum_{\{S_j\}} Q(\{S_j\}) \ln \left[\frac{Q(\{S_j\})}{P(\{S_j\})}\right] \\ &= \ln(Z) + \sum_{\{S_j\}} Q(\{S_j\}) \ln [Q(\{S_j\})] + \beta \sum_{\{S_j\}} Q(\{S_j\}) H(\{S_j\}) \\ &= \ln(Z) - \mathcal{S}[Q] + \beta U[Q], \end{aligned} \tag{3.16}$$

where $\mathcal{S}[Q] = -\sum_{\{S\}} Q(\{S\}) \ln [Q(\{S\})]$ is the entropy of the approximating distribution and $U[Q] = \sum_{\{S\}} Q(\{S\}) H(\{S_j\})$ is the energy averaged over the approximating distribution.

The approximating distribution Q^* is such that $D(Q^*\|P) \leq D(Q\|P)$ for all Q. Firstly we introduce the free-energy functional as $\beta F[Q] = \beta U[Q] - \mathcal{S}[Q]$. Therefore, the KL-divergence can minimised by imposing:

$$\frac{\partial \beta F[Q](\{m_j\})}{\partial \{m_j\}} = 0, \tag{3.17}$$

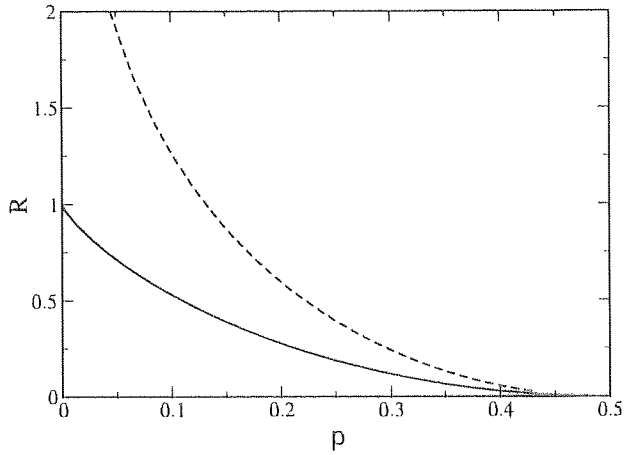


Figure 3.2: Coexistence line in the plane rate *versus* noise level for the Weiss approximation with $K = 10$ (dashed line) and Shannon's bound for the BSC.

as the term $\ln(Z)$ involving the partition function does not depend on the distribution Q.

By plugging the approximating distribution (3.15) into the expression for the free-energy functional $F[Q]$ we get:

$$\begin{aligned} U[Q] &= \sum_{\{S_j\}} Q(\{S_j\}) H(\{S_j\}) \\ &= -\sum_{\{S_j\}} \prod_j \frac{1+S_j m}{2} \sum_\mu J_\mu \prod_{i \in \mathcal{L}(\mu)} S_i \\ &= -\sum_\mu J_\mu m^K \end{aligned} \tag{3.18}$$

and

$$\begin{aligned} \mathcal{S}[Q] &= -\sum_{\{S_j\}} Q(\{S_j\}) \ln\left[Q(\{S_j\})\right] \\ &= -\sum_{\{S_j\}} \prod_j \frac{1+S_j m}{2} \ln\left[\prod_j \frac{1+S_j m}{2}\right] \\ &= -\sum_l \sum_{\{S_j\}} \prod_j \frac{1+S_j m}{2} \ln\left[\frac{1+S_l m}{2}\right] \\ &= -N\left(\frac{1+m}{2}\ln\left[\frac{1+m}{2}\right] + \frac{1-m}{2}\ln\left[\frac{1-m}{2}\right]\right). \end{aligned} \tag{3.19}$$

The free-energy quenched average over the disorder $\{J_\mu\}$ is therefore:

$$\begin{aligned} \langle \beta F(\{m_j\}) \rangle_J &= -\beta M\ (1-2p) m^K \\ &+ N\left(\frac{1+m}{2}\ln\left[\frac{1+m}{2}\right] + \frac{1-m}{2}\ln\left[\frac{1-m}{2}\right]\right) \end{aligned} \tag{3.20}$$

The extremisation (3.17) finally gives:

$$m = \tanh\left[C\beta(1-2p)m^{K-1}\right]. \tag{3.21}$$

The intensive free-energy of the paramagnetic state $f_{\mathrm{PARA}} = \ln 2$ can be equated to the free-energy f_{FERRO} of the state of maximum magnetisation to find the coexistence line. In Fig. 3.2 we show the coexistence line for large K and optimal decoding temperature and compare with the Shannon bound. Evidently this coexistence line in inconsistent with the channel coding theorem as it predicts the possibility of errorless (for $K \to \infty$) communication beyond channel capacity. In the following sections we will discuss a less naïve approach to the problem.

3.4 Equilibrium

3.4.1 Replica theory

In the following subsections we will develop the replica symmetric theory for Sourlas codes and show that, in addition to providing a good description of the equilibrium, it describes the typical decoding dynamics using probability propagation methods.

The previous naïve "all ones" messages assumption can be formally translated to the gauge transformation [FHS78] $S_i \mapsto S_i \xi_i$ 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 ferromagnetic configuration defined as $\xi_i^* = 1 \; \forall i$. By introducing the *external field* $F \equiv \beta_N'/\beta$ one can rewrite the Hamiltonian in the form:

$$\mathcal{H}(\mathbb{S}) = -\sum_{\langle i_1 \cdots i_K \rangle} \mathcal{A}_{\langle i_1 \cdots i_K \rangle} \; J_{\langle i_1 \cdots i_K \rangle} \; S_{i_1} \cdots S_{i_K} - F \sum_{j=1}^{N} \xi_j S_j \;, \tag{3.22}$$

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:

$$P(J) = (1-p)\, \delta(J-1) + p\, \delta(J+1)\,, \tag{3.23}$$

where p is the channel flip rate. For deriving the typical properties we have to obtain an expression for the free-energy by invoking the replica technique where the free-energy is calculated using the identity:

$$f = -\frac{1}{\beta} \lim_{N \to \infty} \frac{1}{N} \frac{\partial}{\partial n}\bigg|_{n=0} \langle \mathcal{Z}^n \rangle_{\mathcal{A},\xi,J}, \tag{3.24}$$

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:

$$\langle \mathcal{Z}^n \rangle_{\mathcal{A},\xi,J} = \mathrm{Tr}_{\{S_j^\alpha\}} \left[\left\langle e^{\beta F \sum_{\alpha,k} \xi_k S_k^\alpha} \right\rangle_\xi \left\langle \exp\left(\beta \sum_{\alpha,\langle i_1 \cdots i_K \rangle} \mathcal{A}_{\langle i_1 \cdots i_K \rangle} \; J_{\langle i_1 \cdots i_K \rangle} \; S_{i_1}^\alpha \cdots S_{i_K}^\alpha \right) \right\rangle_{\mathcal{A},J} \right]. \tag{3.25}$$

The overlap ρ can be rewritten using gauged variables as :

$$\rho = \frac{1}{N} \sum_{i=1}^{N} \left\langle \langle \mathrm{sign}\langle S_i \rangle \rangle_{\mathcal{A},J|\xi^*} \right\rangle_\xi, \tag{3.26}$$

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

To compute the replicated partition function we closely follow [WS87a]. We average uniformly over all codes $\mathcal{A}$ such that $\sum_{\langle i_1=ii_2\cdots i_K\rangle}\mathcal{A}_{\langle i_1\cdots i_K\rangle}=C\ \forall i$ to find:

$$\begin{aligned}\langle\mathcal{Z}^n\rangle_{\mathcal{A},\xi,J} &= \exp\left\{N\ Extr_{q,\widehat{q}}\left[C-\frac{C}{K}+\frac{C}{K}\left(\sum_{l=0}^{n}\mathcal{T}_l\sum_{\langle\alpha_1\ldots\alpha_l\rangle}q^K_{\alpha_1\ldots\alpha_l}\right)\right.\right.\\ &- C\left(\sum_{l=0}^{n}\sum_{\langle\alpha_1\ldots\alpha_l\rangle}q_{\alpha_1\ldots\alpha_l}\widehat{q}_{\alpha_1\ldots\alpha_l}\right)\\ &+ \left.\left.\ln\mathrm{Tr}_{\{S^\alpha\}}\left\langle e^{\beta F\xi\sum_\alpha S^\alpha}\right\rangle_\xi\left(\sum_{l=0}^{n}\sum_{\langle\alpha_1\ldots\alpha_l\rangle}\widehat{q}_{\alpha_1\ldots\alpha_l}S^{\alpha_1}\ldots S^{\alpha_l}\right)^C\right]\right\},\end{aligned}\tag{3.27}$$

where $\mathcal{T}_l=\langle\tanh^l(\beta J)\rangle_J$, as in [VB85], and $q_0=1$. We give details of this calculation in the Appendix B.1 . At the extremum of (3.27) the order parameters acquire a form similar to those of [WS87a]:

$$\begin{aligned}\widehat{q}_{\alpha_1,\ldots,\alpha_l} &= \mathcal{T}_l\,q^{K-1}_{\alpha_1,\ldots,\alpha_l}\\ q_{\alpha_1,\ldots,\alpha_l} &= \left\langle\left(\prod_{i=1}^{l}S^{\alpha_i}\right)\left(\sum_{l=0}^{n}\sum_{\langle\alpha_1\ldots\alpha_l\rangle}\widehat{q}_{\alpha_1\ldots\alpha_l}S^{\alpha_1}\ldots S^{\alpha_l}\right)^{-1}\right\rangle_{\mathcal{X}}.\end{aligned}\tag{3.28}$$

where

$$\mathcal{X}=\left\langle e^{\beta F\xi\sum_\alpha S^\alpha}\right\rangle_\xi\left(\sum_{l=0}^{n}\sum_{\langle\alpha_1\ldots\alpha_l\rangle}\widehat{q}_{\alpha_1\ldots\alpha_l}S^{\alpha_1}\ldots S^{\alpha_l}\right)^C,\tag{3.29}$$

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

3.4.2 Replica symmetric solution

The replica symmetric (RS) ansatz can be introduced via the auxiliary fields $\pi(x)$ and $\widehat{\pi}(y)$ in the following way (see also [WS87a]):

$$\begin{aligned}\widehat{q}_{\alpha_1\ldots\alpha_l} &= \int dy\ \widehat{\pi}(y)\tanh^l(\beta y),\\ q_{\alpha_1\ldots\alpha_l} &= \int dx\ \pi(x)\tanh^l(\beta x)\end{aligned}\tag{3.30}$$

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

Plugging (3.30) into the replicated partition function (3.27), taking the limit $n\to 0$ and using

Eq.(3.24) (see Appendix B.2 for details) one obtains:

$$\begin{aligned} f &= -\frac{1}{\beta}\, Extr_{\pi,\widehat{\pi}} \{\alpha \ln\cosh\beta \\ &+ \alpha \int \left[\prod_{l=1}^{K} dx_l\ \pi(x_l)\right] \left\langle \ln\left[1+\tanh\beta J \prod_{j=1}^{K}\tanh\beta x_j\right]\right\rangle_J \\ &- C\int dx\ dy\ \pi(x)\ \widehat{\pi}(y)\ \ln[1+\tanh\beta x\ \tanh\beta y] \\ &- C\int dy\ \widehat{\pi}(y)\ \ln\cosh\beta y \\ &+ \int \left[\prod_{l=1}^{C} dy_l\ \widehat{\pi}(y_l)\right] \left\langle \ln\left[2\cosh\beta\left(\sum_{j=1}^{C} y_j + F\xi\right)\right]\right\rangle_\xi \Bigg\}, \end{aligned} \tag{3.31}$$

where $\alpha = C/K$. The saddle-point equations, obtained by varying Eq. (3.31) with respect to the probability distributions, provide a closed set of relations between $\pi(x)$ and $\widehat{\pi}(y)$

$$\begin{aligned} \pi(x) &= \int \left[\prod_{l=1}^{C-1} dy_l\ \widehat{\pi}(y_l)\right] \left\langle \delta\left[x - \sum_{j=1}^{C-1} y_j - F\xi\right]\right\rangle_\xi \\ \widehat{\pi}(y) &= \int \left[\prod_{l=1}^{K-1} dx_l\ \pi(x_l)\right] \left\langle \delta\left[y - \frac{1}{\beta}\mathrm{atanh}\left(\tanh\beta J \prod_{j=1}^{K-1}\tanh\beta x_j\right)\right]\right\rangle_J . \end{aligned} \tag{3.32}$$

Later we will show that this self-consistent pair of equations can be seen as a macroscopic version for the probability propagation decoding.

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

$$P(h) = \int \left[\prod_{l=1}^{C} dy_l\ \widehat{\pi}(y_l)\right] \left\langle \delta\left[h - \sum_{j=1}^{C} y_j - F\xi\right]\right\rangle_\xi , \tag{3.33}$$

where $\widehat{\pi}(y)$ is given by the saddle-point equations (3.32).

The overlap (3.7) can then be calculated using:

$$\rho = \int dh\ \mathrm{sign}(h)\, P(h). \tag{3.34}$$

The code performance can be assessed by assuming a particular prior distribution for the message bits, solving the saddle-point equations (3.32) numerically and then computing the overlap.

The replica symmetric solution is expected to be stable at the Nishimori temperature β_N as is shown for $K = 2$ in [NS00].

The stability of the fix points defining the replica symmetric free-energy can be probed looking at second functional derivatives of (3.31). The simplest necessary condition for stability is having non-negative second functional derivatives in relation to $\pi(x)$ (and $\widehat{\pi}(y)$) :

$$\frac{1}{\beta}\int \left[\prod_{l=1}^{K-2} dx_l\ \pi(x_l)\right] \left\langle \ln\left[1+\tanh\beta J\ \tanh^2\beta x \prod_{j=1}^{K-2}\tanh\beta x_j\right]\right\rangle_J \geq 0, \tag{3.35}$$

for all x.

The solutions may be unstable for sufficiently low temperatures (large β). For high temperatures we can expand the above expression around small β values to find the stability condition:

$$\langle J \rangle_J \langle x \rangle_\pi^{K-2} \geq 0 \tag{3.36}$$

The average $\langle x \rangle_\pi = \int dx\, \pi(x)\, x$ vanishes in the paramagnetic phase and is positive (non-zero when K is even) in the ferromagnetic phase, satisfying the stability condition. This result is still generally inconclusive, but provides some evidence that can be examined numerically. In Section 3.4.5 we will test the stability of the solutions using the condition (3.35).

In the next sections we 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.

3.4.3 Case $K \to \infty$, $C = \alpha K$

For this case we can obtain solutions to the saddle-point equations at arbitrary temperatures. The first saddle-point equation (3.32) can be approximated by:

$$x = \sum_{l=1}^{C-1} y_l \approx (C-1)\langle y \rangle_{\widehat{\pi}} = (C-1) \int dy\, y\, \widehat{\pi}(y). \tag{3.37}$$

It means that if $\langle y \rangle_{\widehat{\pi}} = 0$ (paramagnetic phase) then $\pi(x)$ must be concentrated at $x = 0$ implying that $\pi(x) = \delta(x)$ and $\widehat{\pi}(y) = \delta(y)$ are the only possible solutions. Moreover, Eq.(3.37) implies that in the ferromagnetic phase one can expect $x \approx \mathcal{O}(K)$.

Using Eq.(3.37) and the second saddle-point equation (3.32) one can find a self-consistent equation for the mean field $\langle y \rangle_{\widehat{\pi}}$:

$$\langle y \rangle_{\widehat{\pi}} = \left\langle \frac{1}{\beta} \operatorname{atanh} \left[\tanh(\beta J) \left[\tanh(\beta(C-1)\langle y \rangle_{\widehat{\pi}}) \right]^{K-1} \right] \right\rangle_J . \tag{3.38}$$

For the BSC the above average is over the distribution (3.23). Computing the average, using $C = \alpha K$ and rescaling the temperature $\beta = \tilde{\beta}(\ln K)/K$, we obtain in the limit $K \to \infty$:

$$\langle y \rangle_{\widehat{\pi}} \approx (1-2p) \left[\tanh(\tilde{\beta}\alpha \langle y \rangle_{\widehat{\pi}} \ln(K)) \right]^K , \tag{3.39}$$

where p is the channel flip probability. The mean field $\langle y \rangle_{\widehat{\pi}} = 0$ is always a solution to this equation (paramagnetic solution); at $\beta_c = \ln(K)/(2\alpha K(1-2p))$ an extra non-trivial ferromagnetic solution emerges with $\langle y \rangle_{\widehat{\pi}} = 1 - 2p$. As the connection with the overlap ρ is given by Eqs. (3.33) and (3.34); it is not difficult to see that it implies that $\rho = 1$ for the ferromagnetic solution. It is remarkable that the temperature were the ferromagnetic solution emerges is $\beta_c \sim \mathcal{O}(\ln(K)/K)$; it means that in a simulated annealing process paramagnetic-ferromagnetic barriers emerge at high temperatures, implying metastability and, consequently, a very slow convergence. It seems to advocate the use of small K values in practical applications. This case is analysed in Section 3.4.6. For $\beta > \beta_c$ both paramagnetic and ferromagnetic solutions exist.

The ferromagnetic free-energy can be obtained from Eq.(3.31) using Eq.(3.37), 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 $\widehat{\pi}(y) = \delta(y)$ into Equation (3.31):

$$f_{\rm PARA} = -\frac{1}{\beta}(\alpha \ln(\cosh \beta) + \ln 2), \tag{3.40}$$

$$s_{\rm PARA} = \alpha(\ln(\cosh \beta) - \beta \tanh \beta) + \ln 2. \tag{3.41}$$

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.

3.4.4 Replica symmetry breaking: the frozen spins scenario

We have seen in Section 3.4.3 that the replica symmetric paramagnetic solution for $K \to \infty$ is unphysical for $\alpha > (\ln 2)/[\beta \tanh \beta - \ln(\cosh \beta)]$. In order to construct a solution with non-negative entropy one has to break the replica symmetry. In general, it is a difficult task to implement a symmetry breaking scheme in finite connectivity systems (see [Mon98b]). Here we choose as a first candidate a very simple one-step replica symmetry breaking scheme that yields exact results for the REM [GM84, Par80].

This simple one-step replica symmetry breaking solution is known as the *frozen spins* solution. The idea consists in assuming that the 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:

$$\begin{aligned} \frac{1}{N}\sum_{j=1}^{N} S_j^\alpha S_j^\beta &= 1 \text{ , if } \alpha \text{ and } \beta \text{ are in the same island} \\ \frac{1}{N}\sum_{j=1}^{N} S_j^\alpha S_j^\beta &= q \text{ , otherwise.} \end{aligned} \tag{3.42}$$

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

$$\begin{aligned} \langle \mathcal{Z}_{\rm RSB}^n \rangle_{\mathcal{A},\xi,J} &= \left\langle {\rm Tr}_{\{S_j^\alpha\}} \exp\left(-\beta \sum_{\alpha=1}^{n} \mathcal{H}(\boldsymbol{S}^\alpha)\right)\right\rangle_{\mathcal{A},J,\xi} \\ &= \left\langle {\rm Tr}_{\{S_j^1,\cdots,S_j^{n/m}\}} \exp\left(-\beta m \sum_{\alpha=1}^{n/m} \mathcal{H}(\boldsymbol{S}^\alpha)\right)\right\rangle_{\mathcal{A},J,\xi} \\ &= \left\langle \prod_{\alpha}^{n/m} {\rm Tr}_{\{S_j^\alpha\}} \exp\left(-\beta m\, \mathcal{H}(\boldsymbol{S}^\alpha)\right)\right\rangle_{\mathcal{A},J,\xi} \\ &= \langle \mathcal{Z}_{\rm RS}^{n/m} \rangle_{\mathcal{A},\xi,J}, \end{aligned} \tag{3.43}$$

where in the first step we have used the ansatz 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

(3.43) we have:

$$\begin{aligned} \langle \ln \mathcal{Z}_{\mathrm{RSB}}(\beta)\rangle_{\mathcal{A},\xi,J} &= \left.\frac{\partial}{\partial n}\right|_{n=0} \langle \mathcal{Z}^n_{\mathrm{RSB}}(\beta)\rangle_{\mathcal{A},\xi,J} \\ &= \frac{1}{m}\langle \ln \mathcal{Z}_{\mathrm{RS}}(\beta m)\rangle_{\mathcal{A},\xi,J}. \end{aligned} \tag{3.44}$$

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

$$\frac{\partial}{\partial m}\langle \ln \mathcal{Z}_{\mathrm{RSB}}(\beta)\rangle_{\mathcal{A},\xi,J} = 0, \tag{3.45}$$

what is equivalent to

$$\begin{aligned} s_{\mathrm{RS}}(\beta_g) &= -\tilde{\beta}^2 \left.\frac{\partial}{\partial\tilde{\beta}}\right|_{\tilde{\beta}=\beta_g} \left[\frac{1}{\tilde{\beta}}\langle \ln \mathcal{Z}_{\mathrm{RS}}(\tilde{\beta})\rangle_{\mathcal{A},\xi,J}\right] \\ &= 0, \end{aligned} \tag{3.46}$$

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 (3.40), satisfying:

$$\alpha(\ln(\cosh\,\beta_g) - \beta_g\,\tanh\,\beta_g) + \ln\,2 = 0 \tag{3.47}$$

The RSB-spin glass free-energy is given by f_{PARA} (3.40) at temperature β_g:

$$f_{\text{RSB-SG}} = -\frac{1}{\beta_g}\,(\alpha\,\ln(\cosh\,\beta_g) + \ln\,2), \tag{3.48}$$

consequently the entropy is $s_{\text{RSB-SG}} = 0$. In Fig.3.3 we show the phase diagram for a given code rate R in the temperature T versus noise level p plane.

3.4.5 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 :

$$R_c = \frac{1}{\alpha_c} = 1 + p\,\log_2\,p + (1-p)\,\log_2\,(1-p). \tag{3.49}$$

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 be capable of saturating channel capacity in the limit $R \to 0$ [Sou89]. In this section we extend the analysis to show that Shannon's bound can be attained by the Sourlas code at zero temperature also for $K \to \infty$ but with connectivity $C = \alpha K$. In this limit the model is analogous to the diluted REM analysed by Saakian in [Saa98]. The errorless phase is manifested in a ferromagnetic phase with total alignment ($\rho = 1$) (condition that is only possible for infinite K, see Section 3.2) up to a certain critical noise level; a further 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

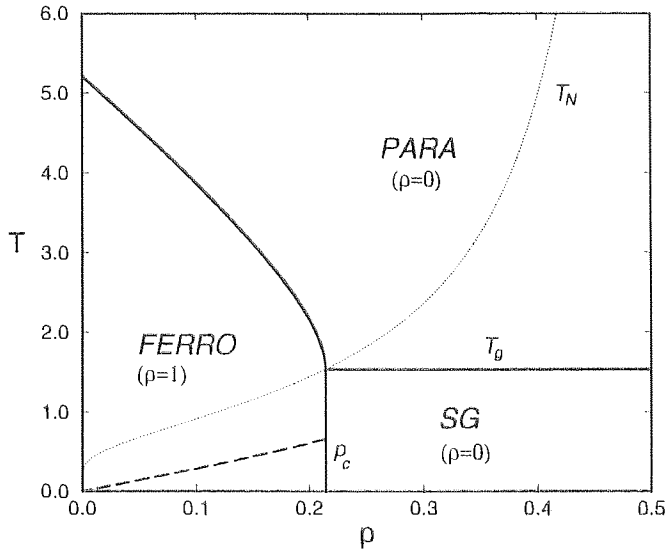


Figure 3.3: Phase diagram in the plane temperature T versus noise level p for $K \to \infty$ and $C = \alpha K$, with $\alpha = 4$. The dotted line indicates the Nishimori temperature T_N . Full lines represent coexistence. The critical noise level is p_c. The necessary condition for stability of the fixed point defining the replica symmetric ferromagnetic state is satisfied above the dashed line.

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 equations (3.32) can be rewritten as:

$$\begin{aligned} \pi(x) &= \int \left[\prod_{l=1}^{C-1} dy_l\ \widehat{\pi}(y_l)\right]\ \delta\left[x - \sum_{j=1}^{C-1} y_j\right] \\ \widehat{\pi}(y) &= \int \left[\prod_{l=1}^{K-1} dx_l\ \pi(x_l)\right]\ \left\langle \delta\left[y - \mathrm{sign}(J \prod_{l=1}^{K-1} x_l)\min(\mid J \mid, ..., \mid x_{K-1} \mid)\right]\right\rangle_J , \end{aligned} \tag{3.50}$$

The solutions for these saddle-point equations may, in general, result in probability distributions with singular and regular parts. As a first approximation we choose the simplest self-consistent family of solutions which are, since $J = \pm 1$, given by:

$$\begin{aligned} \widehat{\pi}(y) &= p_+\delta(y-1) + p_0\delta(y) + p_-\delta(y+1) \\ \pi(x) &= \sum_{l=1-C}^{C-1} T_{[p_\pm, p_0; C-1]}(l)\,\delta(x-l), \end{aligned} \tag{3.51}$$

with

$$T_{[p_+,p_0,p_-;C-1]}(l) = \sum_{\{k,h,m\}}{}' \frac{(C-1)!}{k!\ h!\ m!}\ p_+^k\ p_0^h\ p_-^m, \tag{3.52}$$

where the prime indicates that k, h, m are such that $k - h = l$; $k + h + m = C - 1$. Evidence for this simple ansatz comes from Monte-Carlo integration of Eq. (3.32) at very low temperatures, that shows solutions comprising three dominant peaks and a relatively weak regular part. Having employed a

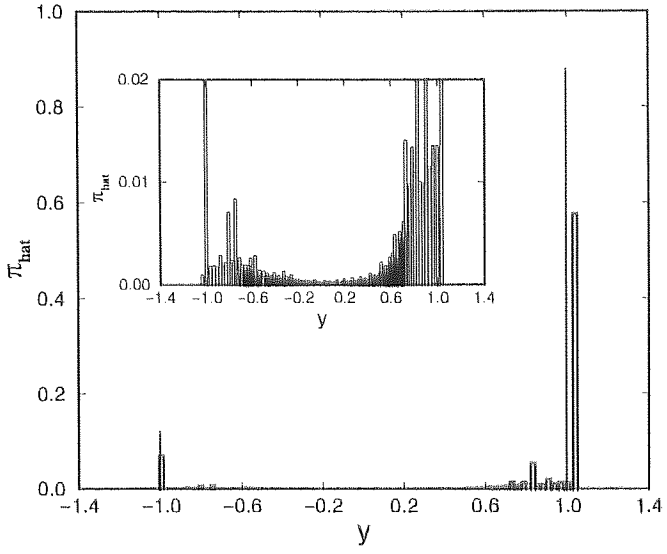


Figure 3.4: Histogram representing the mean field distribution $\widehat{\pi}(y)$ obtained by Monte-Carlo integration at low temperature ($\beta = 10$, $K = 3$, $C = 6$ and $p = 0.1$). Dotted lines represent solutions obtained by iterating self-consistent equations both with five peak and three peak ansätze. Inset: detailed view of the weak regular part arising in the Monte-Carlo integration.

more complex singular solution inside the ferromagnetic and the paramagnetic phases, comprising five peaks $\widehat{\pi}(y) = p_{+2}\delta(y-1) + p_{+}\delta(y-0.5) + p_0\delta(y) + p_{-}\delta(y+0.5) + p_{-2}\delta(y+1)$; we have observed that it collapses back to the simpler three peak solution. In Fig.3.4 we show a typical result of a Monte-Carlo integration for the field $\widehat{\pi}(y)$. The two peaks that emerge by using either the three peak ansatz or the five peak ansatz are shown as dotted lines. In the inset we show the weak regular part of the Monte-Carlo solution.

Plugging the ansatz (3.51) in the saddle-point equations one can write a closed set of equations in $p_{\pm}$ and p_0 that can be solved numerically (see Appendix B.4 for details).

The three peaks solution can be 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 $f_{\mathrm{PARA}} = -\alpha$ and the entropy is $s_{\mathrm{PARA}} = (1-\alpha)\ln 2$, this phase is physical only for $\alpha < 1$, what is expected since it corresponds exactly to the regime where the transmitted information is not sufficient to recover the actual message ($R > 1$).

The ferromagnetic free-energy does not depend on the temperature, having the form $f_{\mathrm{FERRO}} = -\alpha(1-2p)$ with entropy $s_{\mathrm{FERRO}} = 0$. We can find the ferromagnetic-spin glass coexistence line that corresponds to the maximum performance of a Sourlas code by equating Eq. (3.48) and f_{FERRO}. Observing that $\beta_g = \beta_N(p_c)$ (as seen in Fig. 3.3) we find that this transition coincides with the channel capacity (3.49). 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

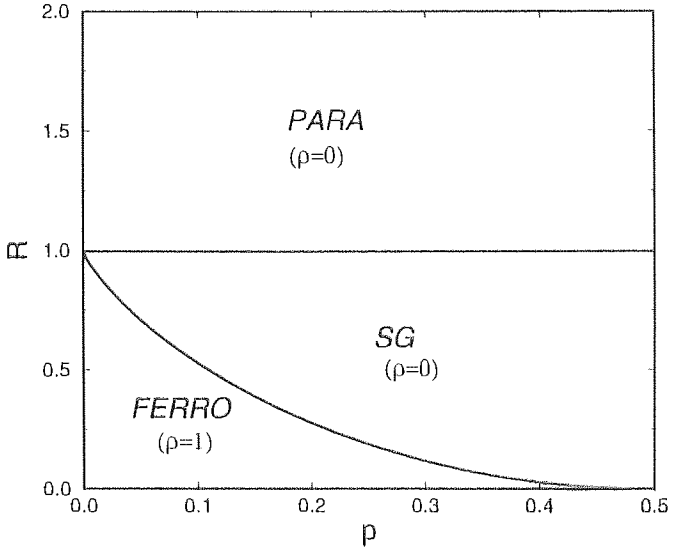


Figure 3.5: Phase diagram in the plane code rate R versus noise level p for $K \to \infty$ and $C = \alpha K$ at zero temperature. The ferromagnetic-spin glass coexistence line corresponds to the Shannon bound.

Shannon's bound is valid also for finite temperatures up to T_g. In Fig. 3.5 we give the complete zero temperature phase diagram.

The bound obtained depends on the stability of the extrema defining the ferromagnetic and paramagnetic solutions at zero temperature. This stability can be checked by using Eq.(3.35) at zero temperature:

$$\lim_{\beta\to\infty} \frac{1}{\beta} \int \left[\prod_{l=1}^{K-2} dx_l \; \pi(x_l)\right] \left\langle \ln \left[1 + \tanh \beta J \; \tanh^2 \beta x \prod_{j=1}^{K-2} \tanh \beta x_j\right] \right\rangle_J \geq 0, \tag{3.53}$$

for all x.

For the paramagnetic solutions the above integral vanishes, trivially satisfying the condition, while for the ferromagnetic solution in the large K regime, $x_l \approx \mathcal{O}(K)$ and the integral becomes

$$-2p\left[(1 - \Theta(x+1)) + |x|\left(\Theta(x+1) - \Theta(x-1)\right) + \Theta(x-1)\right], \tag{3.54}$$

where $\Theta(x) = 1$ for $x \geq 0$ and 0 otherwise, indicating instability for $p > 0$. For the noiseless case $p = 0$ the stability condition is satisfied. The instability of the ferromagnetic state may imply that the ferromagnetic-spin glass transition saturating Shannon's bound is actually unphysical. However, it was shown in Sec. 3.4.2 that this instability vanishes for large temperatures, which supports the ferromagnetic-spin glass transition line obtained and the saturation of the bound in some region.

For finite temperatures the stability condition for the ferromagnetic solution can be rewritten as:

$$\left(1 + \tanh(\beta)\tanh^2(\beta x)\right)^{(1-p)} \left(1 - \tanh(\beta)\tanh^2(\beta x)\right)^{p} \geq 1 \; \forall x. \tag{3.55}$$

For $p = 0$ the condition is clearly satisfied. For finite p, a critical temperature above which the stability condition is fulfilled can be found numerically. The occurrence of instabilities imply that

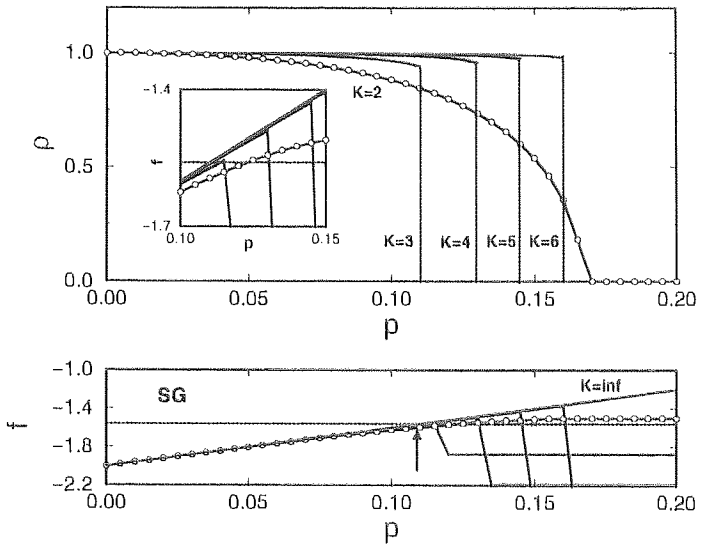


Figure 3.6: Top: zero temperature overlap ρ as a function of the noise level p for various K values at code rate $R = 1/2$, as obtained by the iterative method. Bottom: RS-ferromagnetic free-energies (white circles for $K = 2$ and from the left: $K = 3, 4, 5$ and 6) and RSB-spin glass free-energy (dotted 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.

another stable replica symmetric state may exist or that the replica symmetric ansatz is unphysical, a complete explanation for this is still an open problem.

In Fig. 3.3 we show this temperature in the phase diagram; one can see that there is a considerable region in which our result that Sourlas code can saturate Shannon's bound is supported.

3.4.6 Finite K case

Although Shannon's bound only can be attained in the limit $K \to \infty$, it was shown in Section 3.4.3 that there are some possible drawbacks, mainly in the decoding of messages encoded by large K codes, due to large barriers which are expected to occur between the paramagnetic and ferromagnetic phases. In this section we consider the finite K case, for which we can solve the RS saddle-point equations (3.32) for arbitrary temperatures using Monte-Carlo integration. We can also obtain solutions for the zero temperature case using the simple iterative method described in Section 3.4.5.

It has been shown that $K > 2$ extensively connected models [GM84] 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 Fig.3.6 we show the zero temperature overlap ρ as a function of the noise level p at

code rate $R = 1/2$. These curves were obtained by using the three peak ansatz of the Section 3.4.5. Note that the RSB spin glass phase dominates for $p > p_c$ (see bottom of Fig.3.6). 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. Observe that Shannon's bound cannot be attained for finite K, since $\rho \to 1$ ($p_b \to 0$) only if $K \to \infty$ (Section 3.2).

It is known that the $K = 2$ model with extensive connectivity (SK) requires a full Parisi solution to recover the concavity of the free-energy [MPV87]. No stable solution is known for the intensively connected model (Viana-Bray model). We will see that probability propagation only solves the decoding problem approximately, the approximated solutions are the same obtained by supposing replica symmetry. Thus, we measure the theoretical relevance of the RS results for $K = 2$ by comparison with simulations of probability propagation decoding.

3.4.7 Gaussian noise

Using the replica symmetric free-energy (3.31) and the frozen spins RSB free-energy (3.48) one can easily extend the analysis to other noise types. The general paramagnetic free-energy and entropy can be written:

$$\begin{aligned} f_{\text{PARA}} &= -\frac{1}{\beta}\left(\alpha \left\langle \ln\left(\cosh \beta J\right)\right\rangle_J + \ln\ 2\right) \\ s_{\text{PARA}} &= \alpha\left(\left\langle \ln\left(\cosh \beta J\right)\right\rangle_J - \beta\left\langle J \tanh\left(\beta J\right)\right\rangle_J\right) + \ln 2. \end{aligned} \tag{3.56}$$

The spin glass-RSB free-energy is given by :

$$f_{\text{SG-RSB}} = -\frac{1}{\beta_g}\left(\alpha \left\langle \ln\ \left(\cosh\ \beta_g J\right)\right\rangle_J + \ln\ 2\right), \tag{3.57}$$

with β_g defined as the solution of

$$\alpha\ \left(\left\langle \ln\ \left(\cosh\ \beta_g J\right)\right\rangle_J - \beta_g\left\langle J \tanh\ \left(\beta_g J\right)\right\rangle_J\right) + \ln\ 2 = 0. \tag{3.58}$$

The ferromagnetic free-energy is in general given by (see Appendix B.5)

$$f_{\text{FERRO}} = -\alpha\left\langle J\right\rangle_J = -\alpha\left\langle J \tanh\ \left(\beta_N J\right)\right\rangle_J. \tag{3.59}$$

The maximum performance of the code is defined by the coexistence line :

$$\alpha\left(\left\langle \ln(\cosh\ \beta_g J)\right\rangle_J - \beta_g\left\langle J \tanh(\beta_N J)\right\rangle_J\right) + \ln\ 2 = 0, \tag{3.60}$$

obtained by equating free-energies in the RSB spin glass and ferromagnetic phases. Comparing this expression with entropy (3.58) it can be seen that $\beta_g = \beta_N$ at the coexistence line; the same behaviour

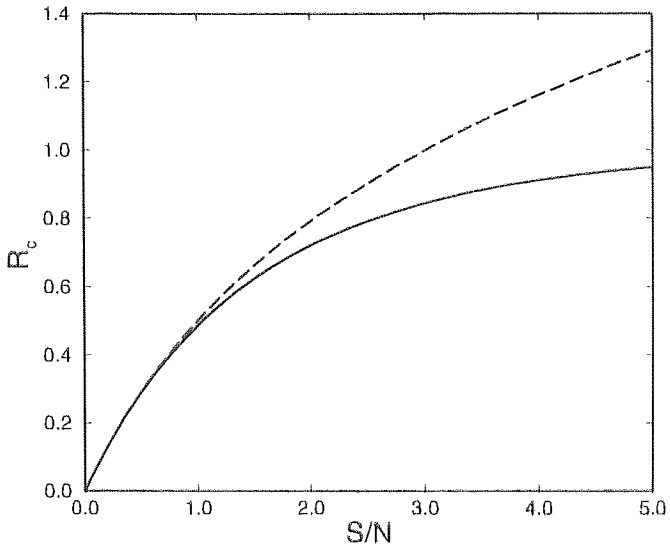


Figure 3.7: Critical code rate R_c and channel capacity for a binary Gaussian channel as a function of the signal to noise rate S/N (solid line). The code of Sourlas saturates Shannon's bound. Channel capacity of the unconstrained Gaussian channel (dashed line).

observed in the BSC case. From Eq. (3.60) one can write:

$$R_c = \beta_N^2 \frac{\partial}{\partial \beta} \left[\frac{1}{\beta} \langle \log_2 \cosh(\beta J) \rangle_J \right]_{\beta=\beta_N}, \tag{3.61}$$

that can be used to compute the performance of the code for arbitrary symmetric noise.

Supposing that the encoded bits can acquire totally unconstrained values Shannon's bound for Gaussian noise is given by $R_c = \frac{1}{2} \log_2(1 + S/N)$, where S/N is the signal to noise ratio, defined as the ratio of source energy per bit (squared amplitude) divided by the spectral density of the noise (variance). However, if one constrains the encoded bits to binary values $\{\pm 1\}$ the capacity of a Gaussian channel becomes:

$$R_c = \int dJ \; P(J \mid 1) \log_2 P(J \mid 1) - \int dJ \; P(J) \log_2 P(J), \tag{3.62}$$

where $P(J \mid J^0) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{(J-J^0)^2}{2\sigma^2})$.

In Fig. 3.7 we show the performance of Sourlas code in a Gaussian channel together with the capacities of the unconstrained and binary Gaussian channels. We show that $K \to \infty$, $C = \alpha K$ Sourlas' code saturates Shannon's bound for the binary Gaussian channel as well. The significantly lower performance with respect to the unconstrained Gaussian channel is due to the binary coding scheme while signal and noise are allowed to acquire real values.

3.5 Decoding with probability propagation

3.5.1 Probability propagation and statistical physics

The probability propagation algorithm was shown in [Mac99] to outperform other methods such as simulated annealing in decoding tasks. In [KS98] it was proposed that this framework can be reinterpreted using statistical physics.

The decoding task consists in the evaluation of estimates of the kind $\widehat{\xi}_j = \text{sign}\langle S_j \rangle_{P(S_j|J)}$. The marginal posterior $P(S_j \mid \boldsymbol{J}) = \sum_{S_l, l \neq j} P(\boldsymbol{S} \mid \boldsymbol{J})$ can in principle be 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 bits long parity-checks and a memoryless BSC channel) to write:

$$P(S_j \mid \boldsymbol{J}) = \frac{1}{P(\boldsymbol{J})} \sum_{S_l, l \neq j} \prod_{\text{checks}} P(J_{\langle i_1 \cdots i_K \rangle} \mid S_{i_1} \cdots S_{i_K}) \prod_{i=1}^{N} P(S_i), \tag{3.63}$$

where $P(\boldsymbol{J})$ is a normalisation 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. As an illustration of how dramatically the computational requirements increase, assume a code with rate $R = 1/2$, if $N = 10$ the number of operations required is 31744, 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 that consists in generating a number (much less than 2^N) of typical vectors $\boldsymbol{S}$ and using this to estimate the marginal posterior, however the sample size can prove to be equally prohibitive.

As a solution to these resource problems, one can explore the structure of (3.63) to devise an algorithm that produces an approximation to $P(S_j \mid \boldsymbol{J})$ in order 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 $\{J_\mu\} = J_{\mu(1)}, \cdots, J_{\mu(C-1)}$. Suppose now that we isolate only the interaction via coupling $J_{\langle i_1 \cdots i_K \rangle}$. When the bipartite graph representing the dependencies in the problem is a tree it is always possible to write:

$$P(S_j \mid J_{\langle i_1 \cdots i_K \rangle}) = \frac{P(S_j)}{P(J_{\langle i_1 \cdots i_K \rangle})} \sum_{\{S_{i_1} \cdots S_{i_{K-1}}\}} P(J_{\langle i_1 \cdots i_K \rangle} \mid S_j, S_{i_1} \cdots S_{i_{K-1}}) \prod_{l=1}^{K-1} P(S_{i_l} \mid \{J_\mu : \mu \in \mathcal{M}(i_l)\}). \tag{3.64}$$

Terms like $P(S_{i_l} \mid \{J_\mu\})$ can be interpreted simply like updated priors for S_{i_l}. In a tree these terms factorise 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 by introducing:

$$Q^x_{\nu j} = P(S_j = x \mid \{J_\mu : \mu \in \mathcal{M}(j) \setminus \nu\}) \tag{3.65}$$

and

$$R^x_{\nu j} = \sum_{\{S_i : i \in \mathcal{L}(\nu) \setminus j\}} P(J_\nu \mid S_j, \{S_i : i \in \mathcal{L}(\nu) \setminus j\}) \prod_{i \in \mathcal{L}(\nu) \setminus j} Q^{S_i}_{\nu i}, \tag{3.66}$$

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 (3.64) can be rewritten as:

$$Q^x_{\mu j} = a_{\mu j} P(S_j = x) \prod_{\nu \in \mathcal{M}(j) \setminus \mu} R^x_{\nu j}. \tag{3.67}$$

Equations (3.66) and (3.67) 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 can be further reduced by using Markov chain Monte-Carlo methods [Mac99].

An approximation to the marginal posterior (3.63) is obtained by counting the influence of all C interactions over each site j and using the assumed factorisation property to write:

$$Q^x_j = a_j P(S_j = x) \prod_{\nu \in \mathcal{M}(j)} R^x_{\nu j}. \tag{3.68}$$

This is an approximation in the sense that the recursion obtained from (3.64) 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 normalisation conditions for the distributions $Q^{+1}_{\mu j} + Q^{-1}_{\mu j} = 1$ and $R^{+1}_{\mu j} + R^{-1}_{\mu j} = 1$ one can change variables and reduce the number of equations by a factor of two $m_{\mu j} = Q^{+1}_{\mu j} - Q^{-1}_{\mu j}$ and $\widehat{m}_{\mu j} = R^{+1}_{\mu j} - R^{-1}_{\mu j}$.

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

$$P(J_\mu \mid S_j, \{S_i : i \in \mathcal{L}(\mu) \setminus j\}) \sim \exp\left(-\beta J_\mu \prod_{i \in \mathcal{L}(\mu)} S_i\right). \tag{3.69}$$

That can be also written in the more convenient form:

$$P(J_\mu \mid S_j, \{S_i : i \in \mathcal{L}(\mu) \setminus j\}) \sim \frac{1}{2}\cosh(\beta J_\mu)\left(1 + \tanh(\beta J_\mu) \prod_{j \in \mathcal{L}(\mu)} S_j\right). \tag{3.70}$$

Plugging Eq. (3.70) for the likelihood in equations (3.67), using the fact that the prior probability is given by $P(S_j) = \frac{1}{2}\left(1 + \tanh(\beta'_N S_j)\right)$ and computing $m_{\mu j}$ and $\widehat{m}_{\mu j}$ (see Appendix B.6):

$$\begin{aligned} \widehat{m}_{\mu j} &= \tanh(\beta J_\mu) \prod_{l \in \mathcal{L}(\mu) \setminus j} m_{\mu l} \\ m_{\mu j} &= \tanh\left(\sum_{\nu \in \mathcal{M}(l) \setminus \mu} \operatorname{atanh}(\widehat{m}_{\nu j}) + \beta'_N\right). \end{aligned} \tag{3.71}$$

The pseudo-posterior can then be calculated:

$$m_j = \tanh\left(\sum_{\nu \in \mathcal{M}(l)} \operatorname{atanh}(\widehat{m}_{\nu j}) + \beta'_N\right), \tag{3.72}$$

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

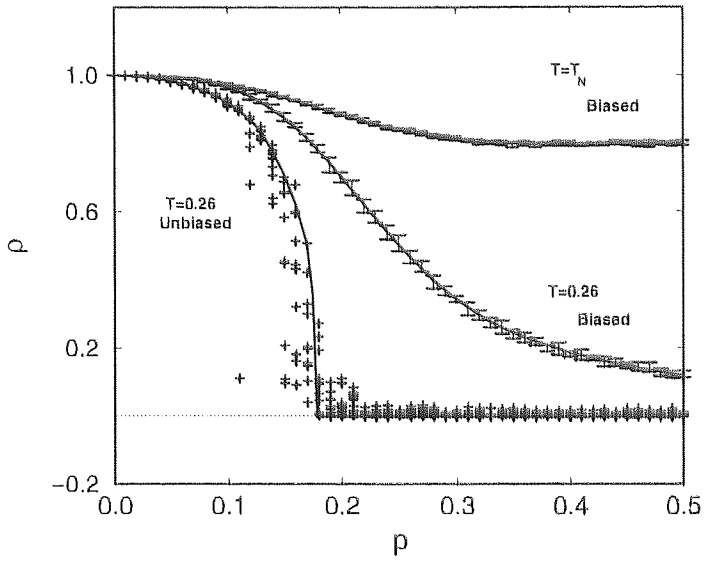


Figure 3.8: Overlap as a function of the flip probability p for decoding using TAP equations for $K = 2$. From the bottom: Monte-Carlo solution of the RS saddle-point equations for unbiased messages ($p_\xi = 0.5$) at $T = 0.26$ (line) and 10 independent runs of TAP decoding for each flip probability (plus signs), $T = 0.26$ and biased messages ($p_\xi = 0.1$) at the Nishimori temperature T_N.

Equations (3.71) depend on the particular received message J. In order to make the analysis message independent, one can use a gauge transformation $\widehat{m}_{\mu j} \mapsto \xi_j \widehat{m}_{\mu j}$ and $m_{\mu j} \mapsto \xi_j m_{\mu j}$ to write:

$$\begin{aligned} \widehat{m}_{\mu j} &= \tanh(\beta J) \prod_{l \in \mathcal{L}(\mu) \setminus j} m_{\mu l} \\ m_{\mu j} &= \tanh\left(\sum_{\nu \in \mathcal{M}(l) \setminus \mu} \tanh^{-1}(\widehat{m}_{\nu j}) + \beta'_N \xi_j \right). \end{aligned} \tag{3.73}$$

In this form, a success in the decoding process correspond to $\widehat{m}_{\mu j} > 0$ and $m_{\mu j} = 1$ for all μ and j. For a large number of iterations, one can expect the ensemble of probability networks to converge to an equilibrium distribution where $\widehat{m}$ and m are random variables sampled from distributions $\widehat{\phi}(y)$ and $\phi(x)$ respectively. By transforming these variables as $\widehat{m} = \tanh(\beta y)$ and $m = \tanh(\beta x)$ and considering the actual message and noise as quenched disorder, Eqs. (3.73) can be rewritten as:

$$\begin{aligned} y &= \frac{1}{\beta} \left\langle \tanh^{-1} \left(\tanh(\beta J) \prod_{j=1}^{K-1} \tanh(\beta x_j) \right) \right\rangle_J \\ x &= \left\langle \sum_{j=1}^{C-1} y_j + \xi F \right\rangle_\xi . \end{aligned} \tag{3.74}$$

The above relations lead to a dynamics on the distributions $\widehat{\phi}(y)$ and $\phi(x)$, that is exactly the same obtained when solving iteratively RS saddle-point equations (3.32). The probability distributions $\widehat{\phi}(y)$ and $\phi(x)$ can be, therefore, identified with $\widehat{\pi}(y)$ and $\pi(x)$ respectively and the RS solutions correspond

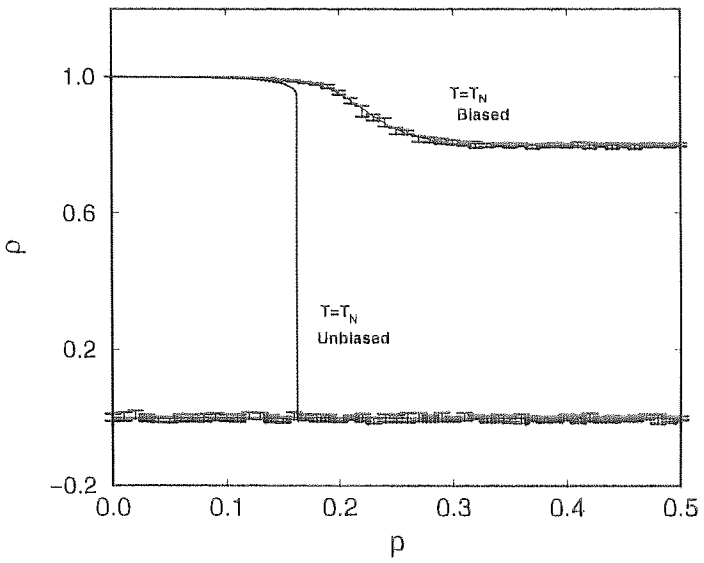


Figure 3.9: Overlap as a function of the flip probability p for decoding using TAP equations for $K = 5$. The dotted line is the replica symmetric saddle-point equations Monte-Carlo integration for unbiased messages ($p_\xi = 0.5$) at the Nishimori temperature T_N. The bottom error bars correspond to 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 ansatz and Monte-Carlo integration.

to decoding a generic message using probability propagation averaged over an ensemble of different codes, noise and signals.

Equations (3.71) 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 (3.71). Numerical solutions for 10 individual runs are presented in Figs. 3.8 and 3.9, initial conditions are chosen as $\widehat{m}_{\mu l} = 0$ and $m_{\mu l} = \tanh(\beta'_N)$ reflecting the prior beliefs. In Fig. 3.8 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 equations (3.32) 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 rather convincing. The third curve in Fig. 3.8 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 Fig.3.9 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 in average even at the Nishimori temperature. For biased messages ($p_\xi = 0.1$, $R = 1/4$) results are far better and in

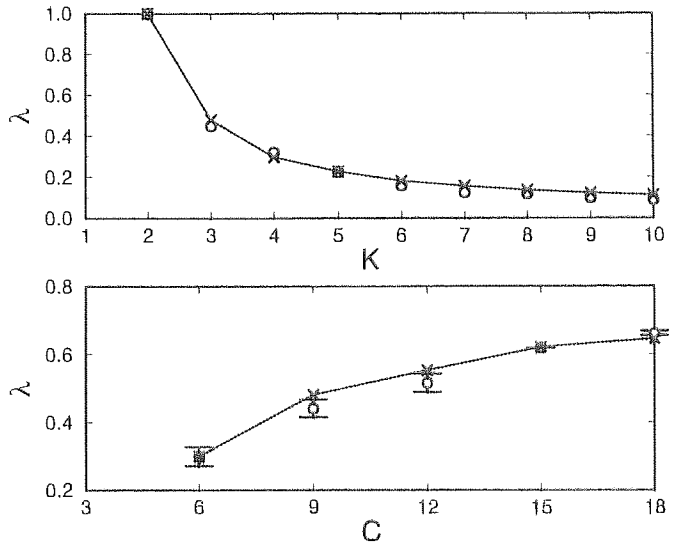


Figure 3.10: Maximum initial overlap λ for decoding. Top: λ as function of the number of interactions K. Circles are averages over 10 different codes with $N = 300$, $R = 1/3$ and noise level $p = 0.1$. Symbol sizes are larger than the error bars. Bottom: λ as function of the connectivity C. Circles are averages over 10 codes with $N = 300$, $K = 3$ and noise level $p = 0.1$. Lines and $\times$'s correspond to the RS dynamics described by the saddle-point equations.

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.

3.5.2 Basin of attraction

To assess the size of the basin of attraction we consider the decoding process as a dynamics with $m_{\mu j}$ as dynamical variables. In gauged transformed equations (3.73) , the perfect decoding of a message correspond to $m^*_{\mu j} = 1$. To analyse the basin of attraction we start with random initial values with a given normalised overlap from the perfect decoding $\lambda = 1 - \frac{1}{NC} \sum_{\mu j} m^0_{\mu j} m^*_{\mu j}$.

In Fig. 3.10 we show the maximal overlap in initial conditions required for successful decoding. The top figure shows an average over 10 different experiments with $N = 300$ (circles) for a fixed code rate $R = 1/3$, fixed noise level $p = 0.1$ and increasing K. The bottom figure shows the maximal overlap in initial conditions for a fixed number of spins per interaction $K = 3$, noise level $p = 0.1$ and increasing C. We confirm the accuracy of the RS description by comparing the experimental results with the basin of attraction predicted by saddle-point equations (3.32). We can interpret these equations as dynamics in the space of distributions $\pi(x)$. Performing the transformation $X = \tanh(\beta x)$, one can move to the space of distributions $\Pi(X)$ with support over $[-1, +1]$. The initial conditions can then be described simply as $\Pi^0(X) = (1 - \frac{\lambda}{2})\delta(X - 1) + \frac{\lambda}{2}\delta(X + 1)$. In Fig. 3.10 we show the basin of

attraction of this dynamics as lines and $\times$'s.

The $K = 2$ case is the only practical code from a dynamical point of view, since it has the largest basin of attraction and no prior knowledge on the message is necessary for decoding. Nevertheless, the code performance degrades faster than the $K > 2$ case as shown in Section 3.4, which points to a compromise between good dynamical properties on the one hand and good performance on the other. One idea could be having a code with changing K, starting with $K = 2$ to guarantee convergence and progressively increasing its values to improve the performance [KS99b].

On the other hand, the basin of attraction increases with C. Again it points to a trade off between good equilibrium properties (small C and large code rates) and good dynamical properties (large C, large basin of attraction). Mixing small and large C values in the same code seems to be a way to take advantage of this trade-off [LMSS98, MWD99, VSK00c].

3.6 Conclusion

In this chapter we studied, using the replica approach, a finite connectivity many-body spin glass that corresponds to the code of Sourlas for finite code rates. We have shown, using a simplified one step RSB solution for spin glass phase, that for $K \to \infty$ and $C = \alpha K$ regime at low temperatures the system exhibits a ferromagnetic-spin glass phase transition that corresponds to Shannon's bound. However, we have also shown that the decoding problem for large K has poor convergence properties when simulated annealing methods are used.

We were able to find replica symmetric solutions for finite K and found good agreement with practical decoding performance using probability propagation. Moreover, we have shown that the RS saddle-point equations describe the typical behaviour of probability propagation algorithms.

We studied the dynamical properties of probability propagation and compared to statistical physics predictions, confirming the validity of the description. The basin of attraction was shown to depend on K and C. Strategies for improving the performance were briefly mentioned.

Chapter 4

Gallager Codes

In this chapter we provide a statistical physics treatment for the encoding/decoding scheme proposed by Gallager. In this scheme the generator of the encoding process is a dense matrix while the decoding matrix is sparse. The code is mapped onto a K-bodies interaction spin system and the typical performance is obtained by using the replica method. A phase transition between the decoding success and failure phases is found to coincide with the information-theoretic upper bound. Channel capacity is shown to be attainable in the $K \to \infty$ limit. A solution is also provided by mapping the problem onto a Husimi cactus, obtaining results that, in the thermodynamic limit, recover the replica symmetric results and provide a very good approximation for finite systems of moderate size. 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. Part of the content in this chapter appeared in [VSK00a, VSK00b].

4.1 Introduction

In 1962 Gallager [Gal62] 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 soon after by the emergency of convolutional codes due to computational limitations of the time. In fact, the best computer available to Gallager in 1962 was an IBM 7090, the first transistorised computer, costing US$ 3 million and with only 1 Megabyte of disc! Convolutional codes only needed a simple system of shift registers to process one byte at a time.

Recently, Gallager codes have been rediscovered by MacKay and Neal that proposed a closely related code [MN95] to be discussed in Chapter 5. Curiously, this almost coincided with the breakthrough discovery of the high performance Turbo codes [BGT93]. Variations of Gallager codes have displayed performance comparable (and sometimes superior) to Turbo codes [Dav98, Dav99], qualifying them as state-of-the-art codes.

In Chapter 3 we have used statistical physics for discussing simple error-correcting codes based on

binary sparse generator matrices first introduced by Sourlas [Sou89]. In the limit of infinite number of nonzero elements in each row of the generator matrix the codes of Sourlas are equivalent to the Random Energy Model [Der81b, Saa98, DW99] and can be thought of as an ideal code capable of saturating Shannon's bound at vanishing code rates. In Chapter 3 we showed that they are ideal codes at finite code rates as well [KS99a, VSK99].

In this chapter we analyse the performance of Gallager error-correcting codes first by using a replica calculation and then by using a generalisation of a Bethe lattice known as Husimi cactus [RK92]. We show that both methods lead to the same results for the typical performance, yielding the threshold noise level that corresponds to the phase transition between perfect decoding and decoding failure phases, this appears to coincide with existing information-theoretic upper bounds.

We also show that the probability propagation (PP) decoding algorithm emerges naturally from the treatment based on a Husimi cactus, allowing for the analysis of the practical performance employing PP decoding. We connect this practical performance to the emergence of metastable states in the replica symmetric free-energy landscape.

We concentrate on analysing a simple communication model whereby messages are represented by binary vectors and are communicated through a Binary Symmetric Channel (BSC) where uncorrelated bit flips appear with probability p. The extension to different channel models is not expected to be difficult as was shown in Section 3.4.7 for a Gaussian channel.

This chapter is organised as follows: Section 4.2 describes the code of Gallager. An upper bound based on information theoretic considerations is discussed in Section 4.3. Section 4.4 presents the statistical physics formulation for Gallager codes. The replica theory is discussed in Section 4.5. Section 4.6 provides an alternative derivation for the probability propagation decoding algorithm based on a tree-like (or cactus-like) approximation for the lattice. In Section 4.7 we introduce a simple method for estimating the practical performance based on the statistical physics picture. Conclusions are given in Section 4.8. Appendices with technical details are also provided.

4.2 The code of Gallager

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 matrix $\boldsymbol{A}$ of dimensionality $(M-N)\times M$. This matrix can, in general, be organised 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

$$\boldsymbol{t} = \boldsymbol{G}^T \boldsymbol{\xi} \pmod 2, \tag{4.1}$$

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

matrix is

$$G = [I \mid C_2^{-1} C_1] \pmod 2, \tag{4.2}$$

where I is the $N \times N$ identity matrix, implying that $AG^T \pmod 2 = 0$ and that the first N bits of t are set to the message ξ. 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 non-zero elements in each row of 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 t is then corrupted by noise represented by the vector $\zeta \in \{0, 1\}^M$ with components independently drawn from $P(\zeta) = (1 - p)\delta(\zeta) + p\delta(\zeta - 1)$. The received vector takes the form

$$r = G^T \xi + \zeta \pmod 2. \tag{4.3}$$

Decoding is carried out by multiplying the received message by the matrix A to produce the *syndrome* vector

$$z = Ar = A\zeta \pmod 2, \tag{4.4}$$

from which an estimate $\hat{\tau}$ for the noise vector can be produced. An estimate for the original message is then obtained as the first N bits of $r + \hat{\tau} \pmod 2$. The Bayes optimal estimator (also known as *marginal posterior maximiser*, MPM) for the noise is defined as $\hat{\tau}_j = \text{argmax}_{\tau_j} P(\tau_j \mid z)$. 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 C_2 and C_1, the syndrome vector z and the noise level p it is possible to apply Bayes theorem and compute the posterior probability

$$P(\tau \mid z) = \frac{1}{Z} \chi[z = A\tau \pmod 2] \, P(\tau), \tag{4.5}$$

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 $P(\tau_j \mid z) = \sum_{i \neq j} P(\tau \mid z)$, 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 A and use probability propagation for decoding, requiring $\mathcal{O}(M)$ operations to perform the same task.

4.3 Upper bound on achievable rates

It was pointed by MacKay in [Mac99] that an upper bound for rates achievable for Gallager codes can be found by an information theoretic argument. This upper bound is based on the fact that each bit of the syndrome vector $z = A\zeta \pmod 2$ 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 D.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

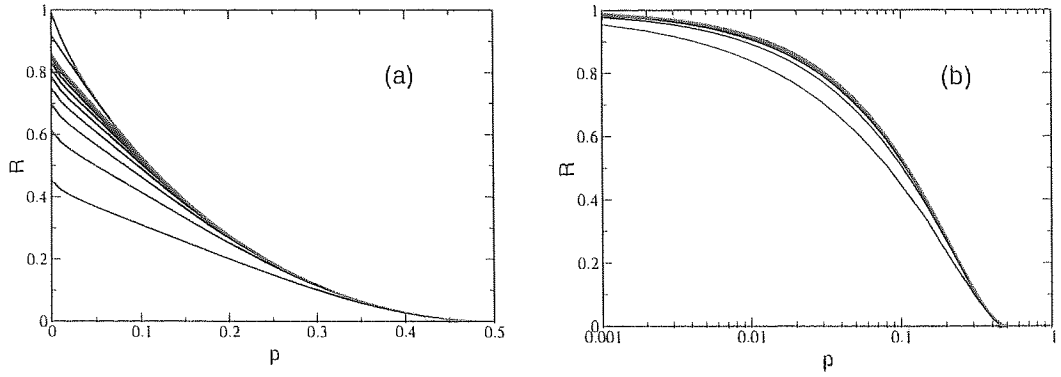


Figure 4.1: (a) Bounds for the rate R as a function of the noise level p for several values of K. From bottom to top: $K = 2$ to $10, 20$ and Shannon limit. (b) Bounds for several values of C. From bottom to top $C = 2, 3, 4, 5$ and Shannon limit.

process one has to extract information from the syndrome vector in order to reconstruct a noise vector ζ which has an information content of $MH_2(p)$. It clearly means that a necessary condition for successful decoding is:

$$\begin{aligned} (M-N)H_2(p_z^1(K)) &\geq MH_2(p) \\ (1-R)H_2(p_z^1(K)) &\geq H_2(p) \\ R &\leq 1 - \frac{H_2(p)}{H_2(p_z^1(K))}. \end{aligned} \tag{4.6}$$

In Fig. 4.1a we plot this bound by fixing K and finding the minimum value for C such that $R = 1 - C/K$ verifies (4.6). Observe that as $K \to \infty$, $p_z^1(K) \to 1/2$ and $R \leq 1 - H_2(p)$ that corresponds to Shannon limit.

In Fig. 4.1b we plot the bound by fixing C and finding the maximum K such that $R = 1 - C/K$ verifies (4.6), recovering the curves presented in [Mac99]. Note that $K \to \infty$ implies $C \to \infty$ and vice-versa. Gallager codes only can attain the Shannon limit asymptotically in the limit of large K or, equivalently, large C.

4.4 Statistical physics formulation

The connection to statistical physics becomes clear when the field $\{0, 1\}$ is replaced by Ising spins $\{\pm 1\}$ and mod 2 sums by products [Sou89]. The syndrome vector acquires the form of a multi-spin coupling $\mathcal{J}_\mu = \prod_{j \in \mathcal{L}(\mu)} \zeta_j$ where $j = 1, \cdots, M$ and $\mu = 1, \cdots, (M - N)$. The K indices of nonzero elements in the row μ of $\boldsymbol{A}$ are given by $\mathcal{L}(\mu) = \{j_1, \cdots, j_K\}$, and in a column l are given by $\mathcal{M}(l) = \{\mu_1, \cdots, \mu_C\}$.

The following family of posterior probabilities can be introduced:

$$\begin{aligned} P_\gamma(\boldsymbol{\tau} \mid \mathcal{J}) &= \frac{1}{Z}\exp\left[-\beta\mathcal{H}_\gamma(\boldsymbol{\tau};\mathcal{J})\right] \\ \mathcal{H}_\gamma(\boldsymbol{\tau};\mathcal{J}) &= -\gamma\sum_{\mu=1}^{M-N}\left(\mathcal{J}_\mu\prod_{j\in\mathcal{L}(\mu)}\tau_j - 1\right) - F\sum_{j=1}^{M}\tau_j\ . \end{aligned} \tag{4.7}$$

The Hamiltonian depends on hyper-parameters 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* [Iba99]). Therefore, γ must be taken to infinity to reflect the hard constraints in Eq. (4.5) and $F = \text{atanh}(1-2p)$, 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 hyper-parameters in the correct scale.

The disorder in (4.7) 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 multi-spin ferromagnet with finite connectivity in a random field $\zeta_j F$:

$$\mathcal{H}_\gamma^{\text{gauge}}(\boldsymbol{\tau};\boldsymbol{\zeta}) = -\gamma\sum_{\mu=1}^{M-N}\left(\prod_{j\in\mathcal{L}(\mu)}\tau_j - 1\right) - F\sum_{j=1}^{M}\zeta_j\tau_j. \tag{4.8}$$

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

$$\mathcal{H}^{\text{gauge}}(\boldsymbol{\tau}\in\Omega;\boldsymbol{\zeta}) = -F\sum_{j=1}^{M}\zeta_j\tau_j, \tag{4.9}$$

where

$$\Omega = \{\boldsymbol{\tau} : \prod_{j\in\mathcal{L}(\mu)}\tau_j = 1\ , \mu = 1,\cdots,M-N\}. \tag{4.10}$$

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

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

$$p_b = \frac{1}{2} - \frac{1}{2M}\sum_{j=1}^{M}\zeta_j\ \text{sgn}(m_j), \tag{4.11}$$

connecting the code performance with the computation of local magnetisations.

4.5 Equilibrium theory

4.5.1 Replica theory

In this section we use the replica theory for analysing the typical performance of Gallager codes along the same lines discussed in Chapter 3 for Sourlas codes. We start by rewriting the gauged Hamiltonian (4.8) in a form more suitable for computing averages over different codes:

$$\mathcal{H}_\gamma^{\text{gauge}}(\boldsymbol{\tau};\boldsymbol{\zeta}) = -\gamma\sum_{(i_1\cdots i_K)}\mathcal{A}_{(i_1\cdots i_K)}\left(\tau_{i_1}\cdots\tau_{i_K} - 1\right) - F\sum_{j=1}^{M}\zeta_j\tau_j, \tag{4.12}$$

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

$$\sum_{\langle i_1 \cdots i_K \rangle} \mathcal{A}_{\langle i_1 \cdots i_K \rangle} = M - N \qquad \sum_{\langle i_1, \cdots, i_j = l, \cdots, i_K \rangle} \mathcal{A}_{\langle i_1, \cdots, i_K \rangle} = C \ \forall l, \tag{4.13}$$

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:

$$f = -\frac{1}{\beta} \lim_{M \to \infty} \frac{1}{M} \langle \ln \mathcal{Z} \rangle_{\mathcal{A}, \boldsymbol{\zeta}}, \tag{4.14}$$

from which the typical macroscopic (thermodynamic) behaviour can be obtained.

The partition function $\mathcal{Z}$ is:

$$\mathcal{Z} = \mathrm{Tr}_{\boldsymbol{\tau}} \exp\left(-\beta \mathcal{H}_{\gamma}^{\text{gauge}}(\boldsymbol{\tau}; \boldsymbol{\zeta})\right). \tag{4.15}$$

The replica method is based on in using

$$f = -\frac{1}{\beta} \lim_{M \to \infty} \frac{1}{M} \left. \frac{\partial}{\partial n} \right|_{n=0} \langle \mathcal{Z}^n \rangle_{\mathcal{A}, \boldsymbol{\zeta}} \tag{4.16}$$

to evaluate the free-energy, what requires computing the averaged replicated partition function:

$$\langle \mathcal{Z}^n \rangle_{\mathcal{A}, \boldsymbol{\zeta}} = \sum_{\boldsymbol{\tau}^1, \cdots, \boldsymbol{\tau}^n} \prod_{j=1}^{M} \left\langle \exp\left(F \zeta \beta \sum_{\alpha=1}^{n} \tau_j^{\alpha}\right) \right\rangle_{\zeta} \left\langle \prod_{\langle i_1 \cdots i_K \rangle} \prod_{\alpha=1}^{n} \exp\left[\beta \gamma \mathcal{A}_{\langle i_1 \cdots i_K \rangle} (\tau_{i_1}^{\alpha} \cdots \tau_{i_K}^{\alpha} - 1)\right] \right\rangle_{\mathcal{A}}. \tag{4.17}$$

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

$$\begin{aligned} \langle (\cdots) \rangle_{\mathcal{A}} &= \frac{1}{\mathcal{N}} \sum_{\{\mathcal{A}\}} \prod_{j=1}^{M} \delta\left(\sum_{\langle i_1 = j, i_2, \cdots, i_K \rangle} \mathcal{A}_{\langle i_1 = j, \cdots, i_K \rangle} - C \right) (\cdots) \\ &= \frac{1}{\mathcal{N}} \sum_{\{\mathcal{A}\}} \prod_{j=1}^{M} \left[\oint \frac{dZ_j}{2\pi i} \frac{1}{Z_j^{C+1}} Z_j^{\sum_{\langle i_1 = j, i_2, \cdots, i_K \rangle} \mathcal{A}_{\langle i_1 = j, \cdots, i_K \rangle}} \right] (\cdots). \end{aligned} \tag{4.18}$$

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

$$\langle (\cdots) \rangle_{\boldsymbol{\zeta}} = \sum_{\zeta = -1, +1} (1-p)\delta(\zeta - 1) \ + \ p\delta(\zeta + 1) \ (\cdots). \tag{4.19}$$

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

$$\int dq_{\alpha_1 \cdots \alpha_m} \delta\left(q_{\alpha_1 \cdots \alpha_m} - \frac{1}{M} \sum_{i}^{M} Z_i \tau_i^{\alpha_1} \cdots \tau_i^{\alpha_m} \right) = 1 \tag{4.20}$$

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

$$\begin{aligned} \langle \mathcal{Z}^n \rangle_{\mathcal{A}, \boldsymbol{\zeta}} &= \frac{1}{\mathcal{N}} \left(\frac{dq_0 d\widehat{q}_0}{2\pi i} \right) \left(\prod_{\alpha=1}^{n} \frac{dq_{\alpha} d\widehat{q}_{\alpha}}{2\pi i} \right) \cdots \\ &\times \exp\left[\frac{M^K}{K!} \sum_{m=0}^{n} \sum_{\langle \alpha_1 \cdots \alpha_m \rangle} \mathcal{T}_m q_{\alpha_1 \cdots \alpha_m}^K - M \sum_{m=0}^{n} \sum_{\langle \alpha_1 \cdots \alpha_m \rangle} q_{\alpha_1 \cdots \alpha_m} \widehat{q}_{\alpha_1 \cdots \alpha_m} \right] \\ &\times \prod_{j=1}^{M} \mathrm{Tr}_{\{\tau^{\alpha}\}} \left[\left\langle \exp\left[F\beta\zeta \sum_{\alpha=1}^{n} \tau^{\alpha} \right] \right\rangle_{\zeta} \right. \\ &\times \left. \oint \frac{dZ}{2\pi i} \frac{\exp\left[Z \sum_{m=0}^{n} \sum_{\langle \alpha_1 \cdots \alpha_m \rangle} \widehat{q}_{\alpha_1 \cdots \alpha_m} \tau^{\alpha_1} \cdots \tau^{\alpha_m} \right]}{Z^{C+1}} \right], \end{aligned} \tag{4.21}$$

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. (B.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 dynamical variable) and the absence of disorder in the couplings, yielding a slightly modified definition for the constants $\mathcal{T}_m$.

4.5.2 Replica symmetric solution

The replica symmetric ansatz consists in assuming:

$$q_{\alpha_1\cdots\alpha_m} = \int dx\ \pi(x)\ x^m \qquad \widehat{q}_{\alpha_1\cdots\alpha_m} = \int d\widehat{x}\ \widehat{\pi}(\widehat{x})\ \widehat{x}^m. \tag{4.22}$$

By performing the limit $\gamma \to \infty$, plugging (4.22) into (4.21), computing the normalisation constant $\mathcal{N}$, integrating in the complex variable Z and computing the trace (see Appendix C.2) one finds:

$$\begin{aligned}\langle \mathcal{Z}^n\rangle_{\mathcal{A},\zeta} &= \mathrm{Extr}_{\pi,\widehat{\pi}}\left\{\exp\left[-MC\left(\int dxd\widehat{x}\ \pi(x)\ \widehat{\pi}(\widehat{x})\ (1+x\widehat{x})^n - 1\right)\right.\right. \\ &+ \left.\left.\left(\frac{MC}{K}\int\prod_{j=1}^{K} dx_j\ \pi(x_j)\ (1+\prod_{j=1}^{K} x_j)^n - 1\right)\right] \right. \\ &\times \left.\left(\int\prod_{j=1}^{C} d\widehat{x}_j\ \widehat{\pi}(\widehat{x}_j)\left\langle\left[\mathrm{e}^{\beta F\zeta}\prod_{j=1}^{C}(1+\widehat{x}_j) + \mathrm{e}^{-\beta F\zeta}\prod_{j=1}^{C}(1-\widehat{x}_j)\right]^n\right\rangle_\zeta\right)^M\right\}.\end{aligned} \tag{4.23}$$

By using (4.16), one finally finds:

$$\begin{aligned} f &= \frac{1}{\beta}\,\mathrm{Extr}_{\pi,\widehat{\pi}}\left\{\frac{C}{K}\ln 2 + C\int dxd\widehat{x}\ \pi(x)\ \widehat{\pi}(\widehat{x})\ \ln(1+x\widehat{x})\right. \\ &- \frac{C}{K}\int\prod_{j=1}^{K} dx_j\ \pi(x_j)\ \ln(1+\prod_{j=1}^{K} x_j) \\ &- \left.\int\prod_{j=1}^{C} d\widehat{x}_j\ \widehat{\pi}(\widehat{x}_j)\left\langle\ln\left[\mathrm{e}^{\beta F\zeta}\prod_{j=1}^{C}(1+\widehat{x}_j) + \mathrm{e}^{-\beta F\zeta}\prod_{j=1}^{C}(1-\widehat{x}_j)\right]\right\rangle_\zeta\right\}. \end{aligned} \tag{4.24}$$

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

$$\begin{aligned}\widehat{\pi}(\widehat{x}) &= \int\prod_{j=1}^{K-1} dx_j\ \pi(x_j)\ \delta\left[\widehat{x} - \prod_{j=1}^{K-1} x_j\right] \\ \pi(x) &= \int\prod_{l=1}^{C-1} d\widehat{x}_l\ \widehat{\pi}(\widehat{x}_l)\left\langle\delta\left[x - \tanh\left(\beta F\zeta + \sum_{l=1}^{C-1}\mathrm{atanh}\ \widehat{x}_l\right)\right]\right\rangle_\zeta,\end{aligned} \tag{4.25}$$

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

$$\begin{aligned}\rho &= \int dh\ P(h)\ \mathrm{sgn}(h) \\ P(h) &= \int\prod_{l=1}^{C} d\widehat{x}_l\ \widehat{\pi}(\widehat{x}_l)\left\langle\delta\left[h - \tanh\left(\beta F\zeta + \sum_{l=1}^{C}\mathrm{atanh}\ \widehat{x}_l\right)\right]\right\rangle_\zeta.\end{aligned} \tag{4.26}$$

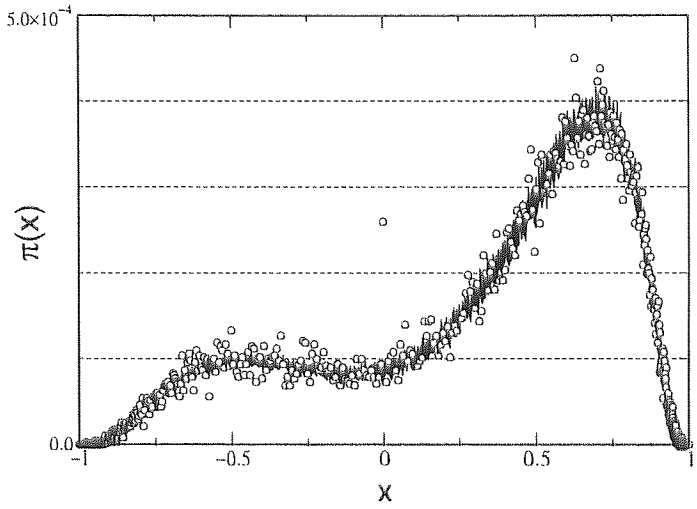


Figure 4.2: Suboptimal ferromagnetic solution $\pi_{\text{NFERRO}}(x)$ for the saddle-point equations (4.25) obtained numerically. Parameters are $K = 4$, $C = 3$ and $p = 0.20$. Circles correspond to an experimental histogram obtained by decoding with probability propagation in 100 runs with 10 different random constructions.

4.5.3 Thermodynamic quantities and typical performance

The typical performance of a code as predicted by the replica symmetric theory can be assessed by solving (4.25) numerically and computing the overlap ρ using (4.26). The numerical calculation can be done by representing distributions π and $\widehat{\pi}$ with 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 $\widehat{\pi}$ that is a solution for (4.25) into (4.26).

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

$$\pi_{\text{FERRO}}(x) = \delta[x - 1] \qquad \widehat{\pi}_{\text{FERRO}}(\widehat{x}) = \delta[\widehat{x} - 1]. \tag{4.27}$$

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 numerically obtained. This solution is depicted in Fig. 4.2 for $K = 4$, $C = 3$ and $p = 0.20$. The ferromagnetic state is *always* a stable solution for (4.25) 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 (4.25).

The free-energy for the ferromagnetic state at Nishimori's temperature is simply $f_{\text{FERRO}} = -F(1 - 2p)$. In Fig. 4.3 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 in Section 4.3. In Fig. 4.4 we show pictorially how the replica symmetric free-energy landscape changes

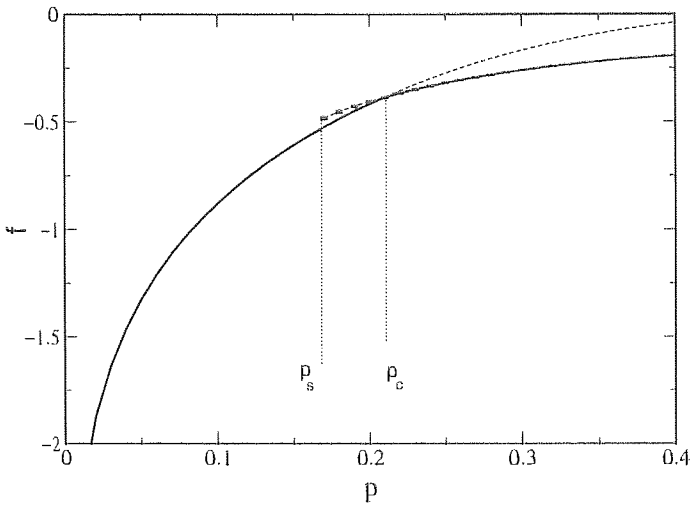


Figure 4.3: Free-energies for $K = 4$, $C = 3$ and $R = 1/4$. The full line corresponds to the free-energy of thermodynamic states. Up to p_s only the ferromagnetic state is present. The ferromagnetic state then dominates the thermodynamics up to p_c, where thermodynamic coexistence with suboptimal ferromagnetic states takes place. Dashed lines correspond to free-energies of non-dominant metastable states.

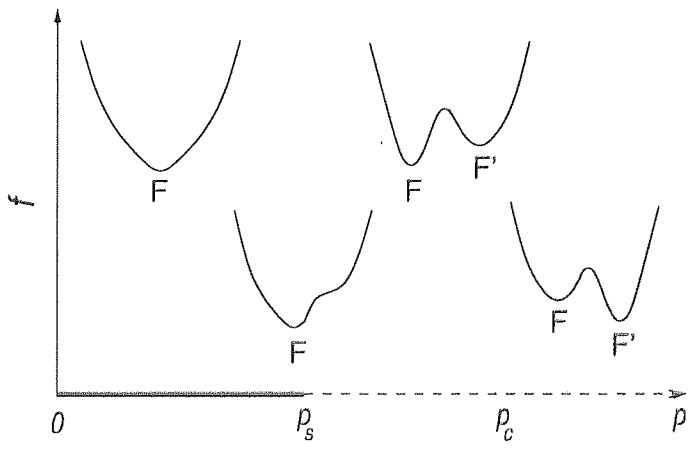


Figure 4.4: Pictorial representation of the replica symmetric free-energy landscape changing with the noise level p. Up to p_s there is only one stable state F corresponding to the ferromagnetic state with $\rho = 1$. At p_s, a second stable suboptimal ferromagnetic state F' emerges with $\rho < 1$, as the noise level increases coexistence is attained at p_c. Above p_c, F' becomes the global minimum dominating the system thermodynamics.

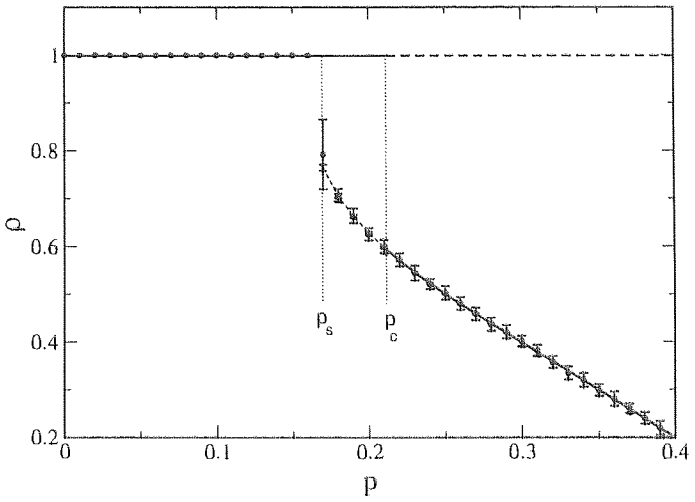


Figure 4.5: Overlaps for $K = 4$, $C = 3$ and $R = 1/4$. The full line corresponds to the overlaps predicted by thermodynamics. Up to p_s only the ferromagnetic $\rho = 1$ state is present, it then dominates the thermodynamics up to p_c, where coexistence with suboptimal ferromagnetic states takes place. Dashed lines correspond to overlaps of non-dominant states.

with the noise level p.

In Fig. 4.5 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 behaviour 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 Fig. 4.5 we show results obtained by simulating probability propagation decoding (black circles). The practical decoding stays in a meta-stable (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 Fig. 4.4, the noise level p_s that provides the practical threshold is signalled by the appearance of spinodal points in the replica symmetric free-energy, defined as points separating (meta)stable and unstable regions in the space of thermodynamical configurations (ρ). The noise level p_s may, therefore, be called *spinodal noise level.*

The solutions obtained must produce non-negative entropies to be physically meaningful. The entropy can be computed from the free-energy (4.24) as $s = \beta^2 \frac{\partial f}{\partial \beta}$ yielding:

$$\begin{aligned} s &= \beta(u(\beta) - f) \\ u(\beta) &= -\int \prod_{j=1}^{C} d\widehat{x}_j \; \widehat{\pi}^*(\widehat{x}_j) \left\langle F\zeta \frac{e^{\beta F\zeta} \prod_{j=1}^{C}(1+\widehat{x}_j) - e^{-\beta F\zeta} \prod_{j=1}^{C}(1-\widehat{x}_j)}{e^{\beta F\zeta} \prod_{j=1}^{C}(1+\widehat{x}_j) + e^{-\beta F\zeta} \prod_{j=1}^{C}(1-\widehat{x}_j)} \right\rangle_{\zeta}, \end{aligned} \tag{4.28}$$

where $\widehat{\pi}^*$ is a solution for the saddle-point equations (4.25) and $u(\beta)$ corresponds to the internal

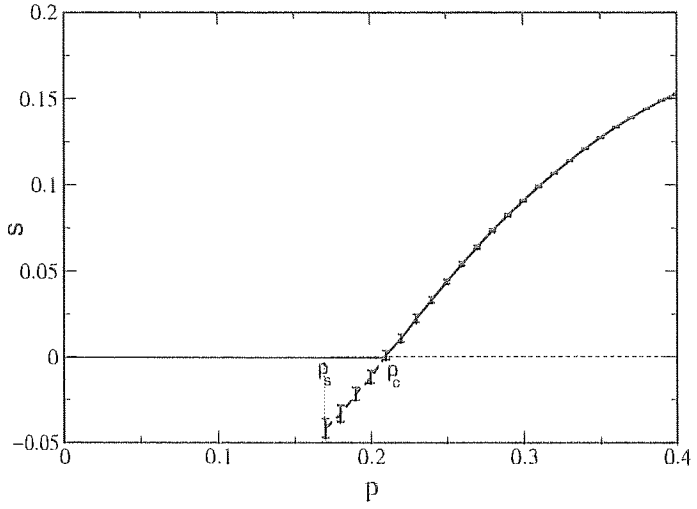


Figure 4.6: Entropy for $K = 4$, $C = 3$ and $R = 1/4$. Negative values imply unphysical behaviour of the metastable replica symmetric state between p_s and p_c.

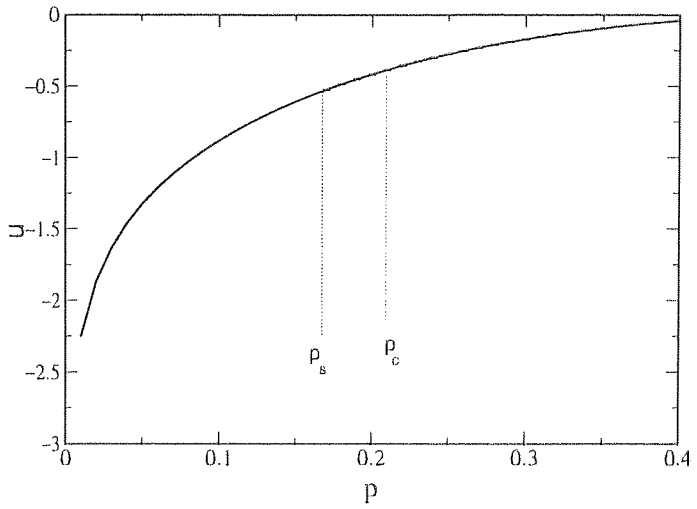


Figure 4.7: Internal energy density for $K = 4$, $C = 3$ and $R = 1/4$ for both ferromagnetic and suboptimal ferromagnetic states. The equality is a consequence of using the Nishimori condition (see Appendix C.3).

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 micro-states consistent with the ferromagnetic state is at most of polynomial order in N. The entropy of the suboptimal ferromagnetic state can be obtained numerically. In Fig. 4.6 we plot the entropy as a function of the noise level for $K = 4$, $C = 3$ and $R = 1/4$. 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 and, therefore, unphysical. At p_c the entropy of the suboptimal ferromagnetic state becomes positive. The internal energy density obtained numerically is depicted in Fig. 4.7 being $u = -F(1-2p)$ for both ferromagnetic and suboptimal ferromagnetic states, what can be justified by assuming the Nishimori condition $\gamma \to \infty$, $\beta = 1$ and $F = \text{atanh}(1-2p)$ [Iba99] (see Appendix C.3).

The unphysical behaviour of the suboptimal ferromagnetic solution between p_s and p_c indicates that the replica symmetric ansatz does not provide the correct physical description for 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 [WS88, Mon98b, Mon98a]. An alternative is to consider a frozen spins solution as discussed in Section 3.4.4 for Sourlas codes. In this case the entropy in the interval $p_s < p < p_c$ is corrected to $s_{\text{RSB}} = 0$ and the free-energy and internal energy are frozen to the values at p_c.

Any candidate to a physical description for the system would have to be compared with simulations to be validated. Nevertheless, our aim here is predicting the behaviour 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.

4.6 Code 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 generalised tree structure (Bethe lattice [WS87b]) known as Husimi cactus that is exactly solvable [Guj95, BL82, RK92, Gol91].

There are many different ways of building mean-field theories. One can make a perturbative expansion around a tractable model [Ple82, Tan00] or assume a tractable structure and variationally determine the model parameters [SJ98]. 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 [KS98]. Moreover, this approach has the advantage of being slightly more controlled and easier to understand than replica calculations.

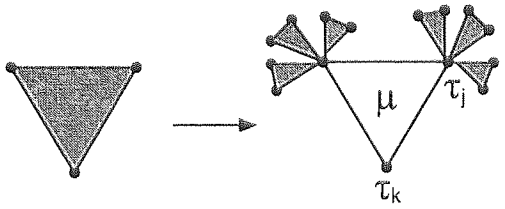


Figure 4.8: First step in the construction of Husimi cactus with $K = 3$ and connectivity $C = 4$.

4.6.1 The Husimi cactus

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 Fig. 4.8 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 magnetisation at the centre m_j can be obtained by fixing boundary (initial) conditions in the 0-th 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 (4.7) 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 [Guj95].

4.6.2 Recursion relations: probability propagation

We adopt the approach presented in [RK92] 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:

$$P_{\mu k}(\tau_k) = \frac{1}{\mathcal{N}} \operatorname{Tr}_{\{\tau_j\}} \exp\left[\beta\gamma\left(\mathcal{J}_\mu \tau_k \prod_{j\in\mathcal{L}(\mu)\setminus k} \tau_j - 1\right) + \beta F \tau_k\right] \prod_{\nu\in\mathcal{M}(j)\setminus\mu} \prod_{j\in\mathcal{L}(\mu)\setminus k} P_{\nu j}(\tau_j), \qquad (4.29)$$

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

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

$$\exp\left(-2\widehat{x}_{\nu j}\right) = \mathrm{e}^{2\beta F} \frac{P_{\nu j}(-)}{P_{\nu j}(+)}, \qquad (4.30)$$

Combining (4.29) and (4.30) one finds the recursion relation (see Appendix C.4 for details):

$$\exp\left(-2\widehat{x}_{\mu k}\right) = \frac{\mathrm{Tr}_{\{\tau_j\}} \exp\left[-\beta\gamma\mathcal{J}_\mu \prod_{j\in\mathcal{L}(\mu)\setminus k}\tau_j + \sum_{j\in\mathcal{L}(\mu)\setminus k}(\beta F + \sum_{\nu\in\mathcal{M}(j)\setminus\mu}\widehat{x}_{\nu j})\tau_j\right]}{\mathrm{Tr}_{\{\tau_j\}} \exp\left[+\beta\gamma\mathcal{J}_\mu \prod_{j\in\mathcal{L}(\mu)\setminus k}\tau_j + \sum_{j\in\mathcal{L}(\mu)\setminus k}(\beta F + \sum_{\nu\in\mathcal{M}(j)\setminus\mu}\widehat{x}_{\nu j})\tau_j\right]}. \tag{4.31}$$

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

$$\widehat{x}_{\mu k} = \mathrm{atanh}\left[\mathcal{J}_\mu \prod_{j\in\mathcal{L}(\mu)\setminus k} \tanh(F + \sum_{\nu\in\mathcal{M}(j)\setminus\mu}\widehat{x}_{\nu j})\right] \tag{4.32}$$

The effective local magnetisation due to interactions with the nearest neighbours in one branch is given by $\widehat{m}_{\mu j} = \tanh(\widehat{x}_{\mu j})$. 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 $x_{\mu j} = F + \sum_{\nu\in\mathcal{M}(j)\setminus\mu}\widehat{x}_{\nu j}$; the effective local magnetisation is, therefore, $m_{\mu j} = \tanh(x_{\mu j})$. Equation (4.32) can then be rewritten in terms of $\widehat{m}_{\mu j}$ and $m_{\mu j}$ and the PP equations [Mac99, KS98, KF98] can be recovered:

$$m_{\mu k} = \tanh\left(F + \sum_{\nu\in\mathcal{M}(k)\setminus\mu} \mathrm{atanh}\,(\widehat{m}_{\nu k})\right) \qquad \widehat{m}_{\mu k} = \mathcal{J}_\mu \prod_{j\in\mathcal{L}(\mu)\setminus k} m_{\mu j} \tag{4.33}$$

Once the magnetisations on the boundary (0-th generation) are assigned, the local magnetisation m_j in the central site is determined by iterating (4.33) and computing :

$$m_j = \tanh\left(F + \sum_{\nu\in\mathcal{M}(j)} \mathrm{atanh}\,(\widehat{m}_{\nu j})\right) \tag{4.34}$$

A free-energy can be obtained by integration of (4.33) [MKSV00, VSK00c, BL82]. The equations (4.33) describing PP decoding represent extrema of the following free-energy:

$$\begin{aligned}\mathcal{F}(\{m_{\mu k}, \widehat{m}_{\mu k}\}) &= \sum_{\mu=1}^{M-N}\sum_{i\in\mathcal{L}(\mu)} \ln(1 + m_{\mu i}\widehat{m}_{\mu i}) - \sum_{\mu=1}^{M-N} \ln(1 + \mathcal{J}_\mu \prod_{i\in\mathcal{L}(\mu)} m_{\mu i}) \\ &- \sum_{j=1}^{M} \ln\left[\mathrm{e}^{F}\prod_{\mu\in\mathcal{M}(j)}(1+\widehat{m}_{\mu j}) + \mathrm{e}^{-F}\prod_{\mu\in\mathcal{M}(j)}(1-\widehat{m}_{\mu j})\right]\end{aligned} \tag{4.35}$$

The iteration of the maps (4.33) is actually one out of many different methods of finding stable extrema of this free-energy.

4.6.3 Macroscopic description and replica symmetric solution

The decoding process can be performed by iterating the multidimensional map (4.33) using some defined scheduling. Assume that the iterations are performed in parallel using the following procedure:

1. Effective local magnetisations are initialised as $m_{\mu k} = 1 - 2p$, reflecting prior probabilities.
2. Conjugate magnetisations $\widehat{m}_{\mu k}$ are updated.
3. Magnetisations $m_{\mu k}$ are computed.
4. If convergence or a maximal number of iterations is attained, stop. Otherwise go to step 2.

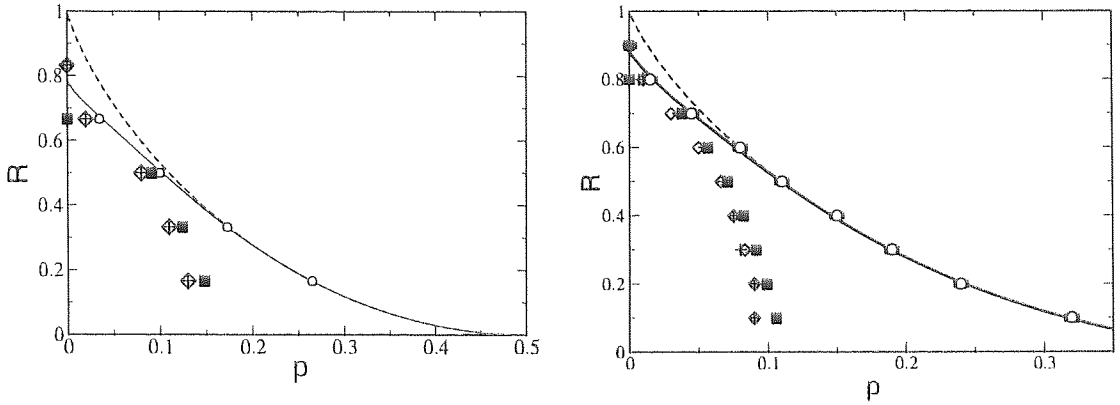


Figure 4.9: Transitions for Gallager codes with $K = 6$ (left) and $K = 10$ (right). Shannon's bound (dashed line), information theory upper bound (full line) and thermodynamic transition obtained numerically ($\circ$). Transitions obtained by Monte-carlo integration of Eq. (4.38) ($\diamond$) and by simulations of PP decoding ($+$, $M = 5000$ averaged over 20 runs) are also shown. Black squares are estimates for practical thresholds based on Sec . 4.7. In both figures, symbols are chosen larger than the error bars.

Equations (4.33) have fixed points that are inconveniently dependent on the particular noise vector ζ. 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:

$$m_{\mu k} = \tanh\left(\zeta_k F + \sum_{\nu \in \mathcal{M}(k)\backslash\mu} \operatorname{atanh}\,(\widehat{m}_{\nu k})\right) \qquad \widehat{m}_{\mu k} = \prod_{j \in \mathcal{L}(\mu)\backslash k} m_{\mu j}. \tag{4.36}$$

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

$$x_{\mu k} = \zeta_k F + \sum_{\nu \in \mathcal{M}(k)\backslash\mu} \widehat{x}_{\nu k} \qquad \widehat{x}_{\mu k} = \operatorname{atanh}\left(\prod_{j \in \mathcal{L}(\mu)\backslash k} \tanh(x_{\mu j})\right). \tag{4.37}$$

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 $\widehat{x}_{\mu k}$.

Assume that the effective fields $x_{\mu k}$ and $\widehat{x}_{\mu k}$ are random variables *independently* sampled from distributions $P(x)$ and $\widehat{P}(\widehat{x})$ respectively, in the same way assume that ζ_j is sampled from $P(\zeta) = (1-p)\,\delta(\zeta-1) + \delta(\zeta+1)$. From Eq. (4.37) a recursion relation in the space of probability distributions [BL82] can be found:

$$\begin{aligned} P_n(x) &= \int d\zeta\; P(\zeta) \int \prod_{l=1}^{C-1} d\widehat{x}_l\; \widehat{P}_{n-1}(\widehat{x}_l)\; \delta\left[x - F\zeta - \sum_{l=1}^{C-1} \widehat{x}_l\right] \\ \widehat{P}_{n-1}(\widehat{x}) &= \int \prod_{j=1}^{K-1} dx_j\; P_{n-1}(x_j)\; \delta\left[\widehat{x} - \operatorname{atanh}\left(\prod_{j=1}^{K-1} \tanh(x_j)\right)\right], \end{aligned} \tag{4.38}$$

where $P_n(x)$ is the distribution of effective fields at the n-th generation due to the previous generations and external fields, in the thermodynamic limit the distribution far from the boundary will be $P_\infty(x)$

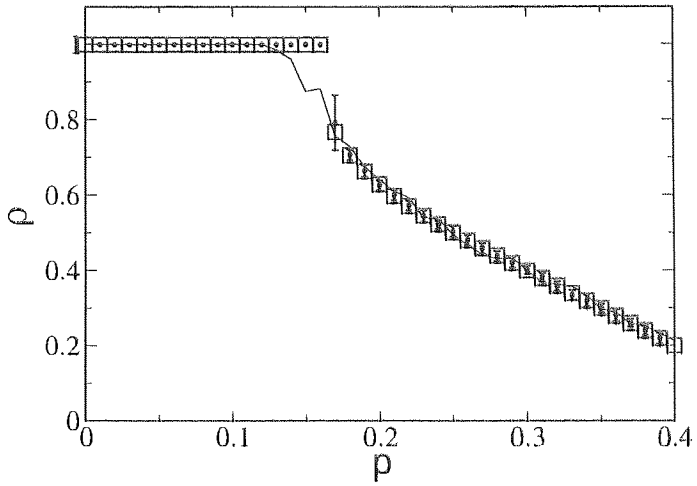


Figure 4.10: Mean normalised overlap ρ between the actual noise vector ζ and decoded noise $\widehat{\tau}$ for a Gallager code with $K = 4$ and $C = 3$ (therefore $R = 1/4$). Theoretical values ($\square$) obtained by Monte-carlo integration of Eq.(4.38) and averages of 20 simulations of PP decoding for code word lengths $M = 5000$ ($\bullet$) and $M = 100$ (full line). Symbols are chosen larger than the error bars.

(generation $n \to \infty$). The local field distribution at the central site is computed by replacing $C - 1$ by C in the first equation (4.38), taking into account C polygons in the generation just before the central site, and inserting the distribution $P_\infty(x)$:

$$P(h) = \int d\zeta \; P(\zeta) \int \prod_{l=1}^{C} d\widehat{x}_l \; \widehat{P}_\infty(\widehat{x}_l) \; \delta \left[x - F\zeta - \sum_{l=1}^{C} \widehat{x}_l \right]. \tag{4.39}$$

Equations (4.38) are identical to equations (4.25) obtained by the replica symmetric theory [KMS00, MKSV00, VSK00c] if the variables describing fields are transformed to those of local magnetisations through $x \mapsto \tanh(\beta x)$.

In Fig. 4.2 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 (4.38). The practical PP decoding is performed by setting initial conditions as $m_{\mu j} = 1 - 2p$ to correspond to the prior probabilities and iterating (4.33) until stationarity or a maximum number of iterations is attained [Mac99]. The estimate for the noise vector is then produced by computing $\widehat{\tau}_j = \text{sign}(m_j)$. At each decoding step the system can be described by histograms of variables (4.33), this is equivalent to iterating (4.38) (a similar idea was presented in [Mac99, Dav98]).

In Fig. 4.9 we summarise the transitions obtained for $K = 6$ and $K = 10$. A dashed line indicates the Shannon limit, the full line represents the information theoretic upper bound of Section 4.3, white circles stand for the coexistence line obtained numerically. Diamonds represent spinodal noise levels obtained by solving (4.38) 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, what agrees with the general belief that decoding gets harder as the Shannon limit is approached.

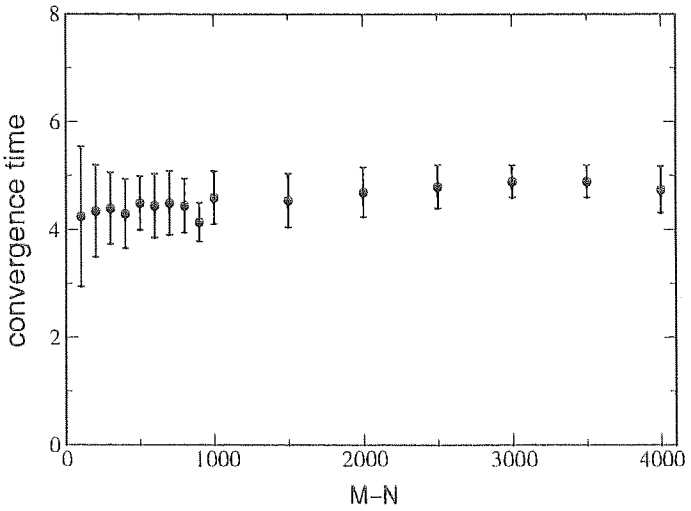


Figure 4.11: PP decoding convergence time as a function of the code size $(M - N)$ for $K = 4$ $C = 3$ and $p = 0.05$, therefore, well below the threshold. The convergence time clearly does not scale with the system size.

4.6.4 Tree-like approximation and the thermodynamic limit

The geometrical structure of a Gallager code defined by the matrix $\boldsymbol{A}$ can be represented by a bipartite graph as represented in Fig. (4.12) (*Tanner graph*) [KF98] with bit and check nodes. 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 [RU98] 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 $\mathcal{P}[l; K, C, M] \leq l^2 K^l / M$ (see Appendix F). It implies that for very large M only cycles of at least order $\ln M$ survive. In the thermodynamic limit $M \to \infty$ 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 Fig. 4.11 we show that the number of iterations of (4.33) 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 4.10 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 of probability propagation algorithms for some specific graphical models can be found in [Wei97].

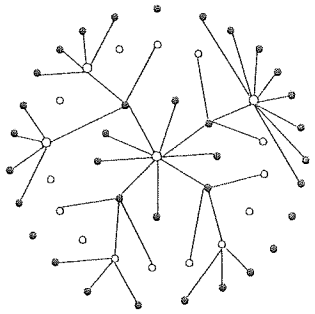


Figure 4.12: Tanner graph representing the neighbourhood of a bit node in an irregular MN code. Black circles represent checks and white circles represent bits.

4.7 Estimating spinodal noise levels

In this section we use insights provided in Section 4.5 to estimate the threshold noise level p_s. For that we introduce a measure for the number of parity-checks violated by a bit τ_l:

$$E_l = -\sum_{\mu \in \mathcal{M}(l)} \left(\mathcal{J}_\mu \tau_l \prod_{j \in \mathcal{L}(\mu) \setminus l} \tau_j \; - \; 1 \right). \tag{4.40}$$

By using gauged variables:

$$E_l = -\sum_{\mu \in \mathcal{M}(l)} \left(\tau_l \prod_{j \in \mathcal{L}(\mu) \setminus l} \tau_j \; - \; 1 \right). \tag{4.41}$$

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 a bit 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 micro-states with overlap around $\rho = 1 - 2p$ (see Fig. 4.10) and we expect that if a single bit, of a randomly guessed vector, is corrected the overlap will either increase or decrease, staying unchanged on average. A vanishing variation in the mean overlap would, therefore, be the signature of the emergence of suboptimal micro-states at p_s.

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

$$P(E_l = C \mid \tau_l = +1) = \prod_{\mu \in \mathcal{M}(l)} P\left\{ \prod_{j \in \mathcal{L}(\mu) \setminus l} \tau_j = -1 \right\}. \tag{4.42}$$

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

$$P(E_l = C \mid \tau_l = -1) = \prod_{\mu \in \mathcal{M}(l)} \left[1 - P\left\{ \prod_{j \in \mathcal{L}(\mu) \setminus l} \tau_j = -1 \right\} \right]. \tag{4.43}$$

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

$$P\left\{ \prod_{j \in \mathcal{L}(\mu) \setminus l} \tau_j = -1 \right\} = \frac{1}{2} \; - \; \frac{1}{2}\,(1-2p)^{K-1}. \tag{4.44}$$

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:

$$\begin{aligned} \frac{1}{2}\langle\Delta\rho\rangle &= P(\tau_l = +1 \mid E_l = C) \; - \; P(\tau_l = -1 \mid E_l = C) \\ &= \frac{P(E_l = C \mid \tau_l = +1)P(\tau_l = +1) \; - \; P(E_l = C \mid \tau_l = -1)P(\tau_l = -1)}{P(E_l = C \mid \tau_l = +1))P(\tau_l = +1) \; + \; P(E_l = C \mid \tau_l = -1)P(\tau_l = -1)}, \end{aligned} \tag{4.45}$$

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

By plugging the prior probability, (4.42) and (4.44) into the above expression we get:

$$\frac{1}{2}\langle\Delta\rho\rangle = \frac{\left[1-(1-2p)^{K-1}\right]^C (1-p) \; - \; \left[1+(1-2p)^{K-1}\right]^C p}{\left[1-(1-2p)^{K-1}\right]^C (1-p) \; + \; \left[1+(1-2p)^{K-1}\right]^C p}. \tag{4.46}$$

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

$$\frac{p_s}{1-p_s} = \left[\frac{1-(1-2p_s)^{K-1}}{1+(1-2p_s)^{K-1}}\right]. \tag{4.47}$$

The above equation can be solved numerically yielding reasonably accurate estimates for practical thresholds p_s as can be seen in Fig. 4.9.

MacKay [Mac99] and Gallager [Gal62, Gal63] introduced probabilistic decoding algorithms whose performance analysis is essentially the same those as presented here. However, the results obtained in Section 4.5 put the analysis into a broader perspective: algorithms that generate decoding solutions in polynomial time, as it 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 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 (4.35).

4.8 Conclusion

We mapped a Gallager code onto a multi-spins interaction Ising system. We then computed the typical performance of a Gallager code by using the replica symmetric theory. A thermodynamic transition between a decodable (ferromagnetic) and an undecodable (suboptimal ferromagnetic) phase at a noise level p_c is found to coincide with information theory based arguments. The Shannon limit is shown to be attainable in the limit $K \to \infty$. The emergence of suboptimal solutions at $p_s < p_c$ separates regions that are decodable in polynomial time and exponential time.

We also introduce a mean-field approximation that transforms the lattice onto a tree-like geometry known as a Husimi cactus. We show that this approximation allows an exact solution without resorting to the replica method. We use this approach to recover the replica symmetric solutions and show that the predictions made are in agreement with numerical experiments carried out in systems of moderate size. The probability propagation algorithm emerges naturally from this mean-field analysis. A simple procedure for computing reasonably accurate estimates for the practical thresholding noise levels p_s is also presented.

Chapter 5

MacKay-Neal Codes

In this chapter we study the statistical physics of the family of codes proposed by MacKay and Neal (MN). The encoding process is performed by a dense matrix while decoding involves two very sparse parity-check matrices, one for the signal with K non-zero elements in each row and the second with L non-zero elements for the noise. We map MN codes onto a spin system with $K + L$-bodies per interaction. The typical performance is then obtained by using a replica symmetric theory. We show that codes with $K, L > 2$ can attain channel capacity in exponential time, while $K = L = 2$ and $K = 1$, $L = 2$ codes can be decoded successfully by a probability propagation decoder, being more practical. The thermodynamic transitions predicted by the replica symmetric theory coincides with information theoretic upper bounds for all cases but $K = L = 2$. We also introduce irregular constructions and show that they improve the spinodal noise level p_s and, consequently, the practical performance. Part of the content of this chapter appeared in [KMS00, KMSV00, MKSV00, VSK00c].

5.1 Introduction

MacKay-Neal (MN) codes were introduced [MN95] as a variation on Gallager codes. As Gallager codes (see Chapter 4), 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, what yields good distance properties and a decoding problem solvable in linear time by using the methods of probability propagation.

A class of constructions inside the MN family was recently proposed by Kanter and Saad [KS99b, KS00b, KS00a] and shown to be capable of outperforming not only Gallager codes but also the cutting-edge Turbo codes. We will discuss Kanter-Saad codes in the next chapter, but this fact alone justifies a thorough study of MN codes.

A number of theorems showing the asymptotic *goodness* of the MN family has been proved in [Mac99]. By assuming that 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 non-zero 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 (see Section 4.3), might as well attain Shannon's bound for a finite number of non-zero 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. However, results obtained using statistical physics techniques [KMS00, MKSV00, VSK00c, VSK00a] seem to indicate that Shannon's bound can actually be approached asymptotically with exponential time decoding. This feature is considered to be new and somewhat surprising [Mac00b].

Statistical physics has been applied to analyse MN codes and its variants in [KMS00, MKSV00, VSK00c]. The analysis follows along the same lines of Chapter 4, namely, we use the replica symmetric theory to obtain all relevant thermodynamic quantities and to calculate coexistence lines. The theory also yields a noise level where suboptimal solutions emerge that is in connection with the practical thresholds observed when probability propagation decoding is used.

This Chapter is organised as follows: Section 5.2 describes MN codes in detail. Section 5.3 derives an information theoretic upper bound. The statistical physics formulation and the replica theory are developed in Sections 5.4 and 5.5 respectively. Section 5.6 briefly discusses the probability (PP) decoding algorithm as applied to MN codes. The PP decoding performance for several families of construction parameters is presented in Section 5.7. Sections 5.8 and 5.9 analyse the improvement in the practical performance attained by the introduction of irregular constructions. Conclusions are given in the final Section 5.10. Appendices with the technical details are also provided.

5.2 MN codes

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

$$\boldsymbol{t} = \boldsymbol{G}\boldsymbol{\xi} \pmod 2, \tag{5.1}$$

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

$$\boldsymbol{G} = \boldsymbol{C}_n^{-1}\boldsymbol{C}_s \pmod 2, \tag{5.2}$$

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 will 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 extension

for other types of noise is expected to be relatively straightforward as was shown for Sourlas codes in Section 3.4.7.

The received vector takes the form

$$\boldsymbol{r} = \boldsymbol{G}\boldsymbol{\xi} + \boldsymbol{\zeta} \pmod 2. \tag{5.3}$$

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

$$\boldsymbol{z} = \boldsymbol{C}_n\boldsymbol{r} = \boldsymbol{C}_s\boldsymbol{\xi} + \boldsymbol{C}_n\boldsymbol{\zeta} \pmod 2, \tag{5.4}$$

from which an estimate $\widehat{\boldsymbol{\xi}}$ 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 a row m in $\boldsymbol{C}_s$ and $\boldsymbol{C}_n$ contains K_m and L_m non-zero elements respectively. In the same way, each column j of $\boldsymbol{C}_s$ contains C_j non-zero elements and each column l of $\boldsymbol{C}_n$ contains D_l non-zero elements.

Counting the number of non-zero elements in the matrices leads to the following relations:

$$\sum_{j=1}^{N} C_j = \sum_{\mu=1}^{M} K_\mu \qquad \sum_{l=1}^{M} D_l = \sum_{\mu=1}^{M} L_\mu, \tag{5.5}$$

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

$$\overline{K} = \frac{1}{M}\sum_{\mu=1}^{M} K_\mu \qquad \overline{C} = \frac{1}{N}\sum_{j=1}^{N} C_j. \tag{5.6}$$

The Bayes optimal estimator $\widehat{\boldsymbol{\xi}}$ for the message $\boldsymbol{\xi}$ is $\widehat{\xi}_j = \mathrm{argmax}_{S_j} P(S_j \mid \boldsymbol{z})$. The performance of this estimator is measured by the probability of bit error $p_b = 1 - 1/N \sum_{j=1}^{N} \delta[\widehat{\xi}_j, \xi_j]$, 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 the Bayes theorem:

$$P(\boldsymbol{S}, \boldsymbol{\tau} \mid \boldsymbol{z}) = \frac{1}{Z}\chi\left[\boldsymbol{z} = \boldsymbol{C}_s\boldsymbol{S} + \boldsymbol{C}_n\boldsymbol{\tau} \pmod 2\right] P(\boldsymbol{S})P(\boldsymbol{\tau}), \tag{5.7}$$

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

$$P(S_j \mid \boldsymbol{z}) = \sum_{\{S_i; i \neq j\}} \sum_{\boldsymbol{\tau}} P(\boldsymbol{S}, \boldsymbol{\tau} \mid \boldsymbol{z}), \tag{5.8}$$

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, what requires only $\mathcal{O}(N)$ operations.

When $p = p_\xi$, MN and Gallager codes are equivalent under a certain 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.(4.4)). This feature opens the possibility of adjusting the code behaviour by controlling the message bias p_ξ.

An MN code can be thought as a non-linear code [Mac00c]. Redundancy in the original message could be removed (introduced) by using a source (de)compressor defined by some non-linear function $\boldsymbol{\xi} = g(\boldsymbol{\xi}_0; p_\xi)$. Encoding would then be $\boldsymbol{t} = \boldsymbol{G}g(\boldsymbol{\xi}_0; p_\xi)$ (mod 2). In the next section we show that some new features emerge due to the introduction of the parameter p_ξ.

5.3 Upper bound on achievable rates

The same kind of information theoretic argument discussed in Section 4.3 can be applied to MN codes. For a regular code the syndrome vector $\boldsymbol{z} = \boldsymbol{C}_s\boldsymbol{S} + \boldsymbol{C}_n\boldsymbol{\tau}$ (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 D.1)

$$p_z^1(K, L) = \frac{1}{2} - \frac{1}{2}(1 - 2p_\xi)^K (1 - 2p)^L. \tag{5.9}$$

The maximum information content in the syndrome vector is $MH_2(p_z^1(K, L))$ (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 $NH_2(p_\xi) + MH_2(p)$ (in *bits* or *shannons*). Thus, it is a necessary condition for successful decoding that:

$$\begin{aligned} M\ H_2(p_z^1(K, L)) &\geq N\ H_2(p_\xi) + M\ H_2(p) \\ H_2(p_z^1(K, L)) - H_2(p) &\geq R\ H_2(p_\xi) \\ R &\leq \frac{H_2(p_z^1(K, L)) - H_2(p)}{H_2(p_\xi)}. \end{aligned} \tag{5.10}$$

For the case $p_\xi = p$ and $L = C$, we can recover bounds (4.6) for Gallager codes with sizes and parameters redefined as $M' = M + N$, $N' = N$ and $K' = K + L$. In [Mac99] 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 (5.10) becomes

$$R \leq 1 - H_2(p), \tag{5.11}$$

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

5.4 Statistical physics formulation

The statistical physics formulation for MN codes is a straightforward extension of the formulation presented in Section 4.4 for Gallager codes. The field $(\{0,1\}, + \pmod 2)$ is replaced by $(\{\pm 1\}, \times)$ [Sou89] and the syndrome vector acquires the form :

$$\mathcal{J}_\mu = \prod_{j \in \mathcal{L}_s(\mu)} \xi_j \prod_{l \in \mathcal{L}_n(\mu)} \zeta_l \tag{5.12}$$

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

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

Under the assumption that priors $P(\boldsymbol{S})$ and $P(\boldsymbol{\tau})$ are completely factorisable, the posterior (5.7) corresponds to the limit $\gamma \to \infty$ and $\beta = 1$ (Nishimori temperature) of:

$$\begin{aligned} P_\gamma(\boldsymbol{S}, \boldsymbol{\tau} \mid \mathcal{J}) &= \frac{1}{Z} \exp\left[-\beta \mathcal{H}_\gamma(\boldsymbol{S}, \boldsymbol{\tau}; \mathcal{J})\right] \\ \mathcal{H}_\gamma(\boldsymbol{S}, \boldsymbol{\tau}; \mathcal{J}) &= -\gamma \sum_{\mu=1}^{M} \left(\mathcal{J}_\mu \prod_{j \in \mathcal{L}_s(\mu)} S_j \prod_{l \in \mathcal{L}_n(\mu)} \tau_l - 1 \right) - F_s \sum_{j=1}^{N} S_j - F_n \sum_{l=1}^{M} \tau_l, \end{aligned} \tag{5.13}$$

with $F_s = \frac{1}{2} \operatorname{atanh}(\frac{1-p_\xi}{p_\xi})$ and $F_n = \frac{1}{2} \operatorname{atanh}(\frac{1-p}{p})$ (*Nishimori condition* [Iba99]).

By applying the gauge transform $S_j \mapsto S_j \xi_j$ and $\tau_l \mapsto \tau_l \zeta_l$ the couplings can be transformed as $\mathcal{J}_\mu \mapsto 1$, eliminating the disorder. The model is free of frustration (as in [Tou77], the model is *flat*) . Similarly to Gallager codes, the resulting Hamiltonian consists of two sub-lattices interacting via multi-spin ferromagnetic iterations with finite connectivity in random fields $\xi_j F_s$ and $\zeta_l F_n$:

$$\mathcal{H}_\gamma^{\text{gauge}}(\boldsymbol{S}, \boldsymbol{\tau}; \boldsymbol{\xi}, \boldsymbol{\zeta}) = -\gamma \sum_{\mu=1}^{M} \left(\prod_{j \in \mathcal{L}_s(\mu)} S_j \prod_{l \in \mathcal{L}_n(\mu)} \tau_l - 1 \right) - F_s \sum_{j=1}^{N} \xi_j S_j - F_n \sum_{l=1}^{M} \zeta_l \tau_l. \tag{5.14}$$

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

$$\mathcal{H}^{\text{gauge}}((\boldsymbol{S}, \boldsymbol{\tau}) \in \Omega; \boldsymbol{\xi}, \boldsymbol{\zeta}) = -F_s \sum_{j=1}^{N} \xi_j S_j - F_n \sum_{l=1}^{M} \zeta_l \tau_l, \tag{5.15}$$

where

$$\Omega = \{(\boldsymbol{S}, \boldsymbol{\tau}) : \prod_{j \in \mathcal{L}_s(\mu)} S_j \prod_{l \in \mathcal{L}_n(\mu)} \tau_l = 1 \, , \mu = 1, \cdots, M\}. \tag{5.16}$$

Optimal decoding consists of finding local magnetisations at the Nishimori temperature in the *signal sub-lattice* $m_j = \langle S_j \rangle_{\beta_N}$ and calculating Bayesian estimates $\widehat{\xi}_j = \operatorname{sgn}(m_j)$.

The probability of bit error is

$$p_b = \frac{1}{2} - \frac{1}{2N} \sum_{j=1}^{N} \xi_j \operatorname{sgn}(m_j), \tag{5.17}$$

connecting the code performance with the computation of local magnetisations.

5.5 Equilibrium theory

5.5.1 Replica theory

The replica theory for MN codes is a straightforward extension of the theory constructed for Gallager codes in Section 4.5.1, with the introduction of extra dynamical variables S. The gauged Hamiltonian (5.14) is written as:

$$\mathcal{H}_\gamma^{\text{gauge}}(\boldsymbol{S},\boldsymbol{\tau};\boldsymbol{\xi},\boldsymbol{\zeta}) = -\gamma \sum_{\langle \boldsymbol{jl}\rangle} \mathcal{A}_{\langle \boldsymbol{jl}\rangle} \left(S_{j_1}\cdots S_{j_K}\tau_{l_1}\cdots\tau_{l_L} - 1\right) - F_s \sum_{j=1}^{N} \xi_j S_j - F_n \sum_{l=1}^{M} \zeta_l \tau_l, \tag{5.18}$$

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

The code construction is 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 non-zero elements in a particular row of the matrix $[\boldsymbol{C}_s \mid \boldsymbol{C}_n]$. To cope with non-invertible $\boldsymbol{C}_n$ matrices one can start by considering an ensemble with uniformly generated $M \times M$ matrices. The non-invertible instances can then 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:

1. 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;

2. 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:

$$f = -\frac{1}{\beta} \lim_{N\to\infty} \frac{1}{N} \frac{\partial}{\partial n}\bigg|_{n=0} \langle \mathcal{Z}^n\rangle_{\mathcal{A},\xi,\zeta} \tag{5.19}$$

The replicated partition function is:

$$\begin{aligned}\langle \mathcal{Z}^n\rangle_{\mathcal{A},\xi,\zeta} &= \sum_{\boldsymbol{S}^1,\cdots,\boldsymbol{S}^n} \sum_{\boldsymbol{\tau}^1,\cdots,\boldsymbol{\tau}^n} \prod_{j=1}^{N} \left\langle \exp\left(F_s\xi_j\beta \sum_{\alpha=1}^{n} S_j^\alpha\right)\right\rangle_\xi \prod_{l=1}^{M} \left\langle \exp\left(F_n\zeta_l\beta \sum_{\alpha=1}^{n} \tau_l^\alpha\right)\right\rangle_\zeta \\ &\times \left\langle \prod_{\langle \boldsymbol{jl}\rangle} \prod_{\alpha=1}^{n} \exp\left[\beta\gamma\mathcal{A}_{\langle \boldsymbol{jl}\rangle}(S_{j_1}^\alpha\cdots S_{j_K}^\alpha \tau_{l_1}^\alpha\cdots\tau_{l_L}^\alpha - 1)\right]\right\rangle_{\mathcal{A}}.\end{aligned} \tag{5.20}$$

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

$$\begin{aligned}\langle(\cdots)\rangle_{\mathcal{A}} &= \sum_{\{C_j,D_l\}}\prod_{j=1}^{N}\mathcal{P}_C(C_j)\prod_{l=1}^{M}\mathcal{P}_D(D_l)\,\frac{1}{\mathcal{N}}\,\delta\left(\sum_{\langle j_1=j,i_2,\cdots,j_K\boldsymbol{l}\rangle}\mathcal{A}_{\langle j\boldsymbol{l}\rangle}-C_j\right)\\ &\times\ \delta\left(\sum_{\langle j\boldsymbol{l}_1=l,l_2,\cdots,l_K\rangle}\mathcal{A}_{\langle j\boldsymbol{l}\rangle}-D_l\right)(\cdots)\\ &= \sum_{\{C_j,D_l\}}\prod_{j=1}^{N}\mathcal{P}_C(C_j)\prod_{l=1}^{M}\mathcal{P}_D(D_l)\,\frac{1}{\mathcal{N}}\sum_{\{\mathcal{A}\}}\prod_{j=1}^{N}\left[\oint\frac{dZ_j}{2\pi i}\frac{1}{Z_j^{C_j+1}}Z_j^{\sum_{\langle i_1=j,i_2,\cdots,i_K\boldsymbol{l}\rangle}\mathcal{A}_{\langle j_1=j,\cdots,j_K\boldsymbol{l}\rangle}}\right]\\ &\times\ \prod_{l=1}^{M}\left[\oint\frac{dY_l}{2\pi i}\frac{1}{Y_l^{D_l+1}}Y_l^{\sum_{\langle j\boldsymbol{i}_1=l,i_2,\cdots,i_L\rangle}\mathcal{A}_{\langle j\boldsymbol{i}_1=l,\cdots,i_L\rangle}}\right](\cdots),\end{aligned}\tag{5.21}$$

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:

$$\langle(\cdots)\rangle_\xi = \sum_{\xi=-1,+1}(1-p_\xi)\,\delta(\xi-1)\ +\ p_\xi\,\delta(\xi+1)\,(\cdots).\tag{5.22}$$

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

$$\langle(\cdots)\rangle_\zeta = \sum_{\zeta=-1,+1}(1-p)\,\delta(\zeta-1)\ +\ p\,\delta(\zeta+1)\,(\cdots).\tag{5.23}$$

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

$$\int dq_{\alpha_1\cdots\alpha_m}\,\delta\left(q_{\alpha_1\cdots\alpha_m}-\frac{1}{N}\sum_{i}^{N}Z_i S_i^{\alpha_1}\cdots S_i^{\alpha_m}\right)=1\tag{5.24}$$

$$\int dr_{\alpha_1\cdots\alpha_m}\,\delta\left(r_{\alpha_1\cdots\alpha_m}-\frac{1}{M}\sum_{i}^{M}Y_i\tau_i^{\alpha_1}\cdots\tau_i^{\alpha_m}\right)=1\tag{5.25}$$

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

$$
\begin{aligned}
\langle \mathcal{Z}^n \rangle_{\mathcal{A},\xi,\zeta} &= \prod_{j=1}^{N} \sum_{C_j} \mathcal{P}_C(C_j) \prod_{l=1}^{M} \sum_{D_l} \mathcal{P}_D(D_l) \\
&\times \left(\frac{dq_0 d\widehat{q}_0}{2\pi i} \right) \left(\prod_{\alpha=1}^{n} \frac{dq_\alpha d\widehat{q}_\alpha}{2\pi i} \right) \cdots \left(\frac{dr_0 d\widehat{r}_0}{2\pi i} \right) \left(\prod_{\alpha=1}^{n} \frac{dr_\alpha d\widehat{r}_\alpha}{2\pi i} \right) \cdots \\
&\times \exp \left[\frac{M^L N^K}{K!L!} \sum_{m=0}^{n} \sum_{\langle \alpha_1 \cdots \alpha_m \rangle} \mathcal{T}_m q^K_{\alpha_1 \cdots \alpha_m} r^L_{\alpha_1 \cdots \alpha_m} \right. \\
&- \left. N \sum_{m=0}^{n} \sum_{\langle \alpha_1 \cdots \alpha_m \rangle} q_{\alpha_1 \cdots \alpha_m} \widehat{q}_{\alpha_1 \cdots \alpha_m} - M \sum_{m=0}^{n} \sum_{\langle \alpha_1 \cdots \alpha_m \rangle} r_{\alpha_1 \cdots \alpha_m} \widehat{r}_{\alpha_1 \cdots \alpha_m} \right] \\
&\times \frac{1}{\mathcal{N}} \prod_{j=1}^{N} \mathrm{Tr}_{\{S_j^\alpha\}} \left[\left\langle \exp \left[F_s \beta \xi \sum_{\alpha=1}^{n} S_j^\alpha \right] \right\rangle_\xi \right. \\
&\times \left. \oint \frac{dZ_j}{2\pi i} \frac{\exp \left[Z_j \sum_{m=0}^{n} \sum_{\langle \alpha_1 \cdots \alpha_m \rangle} \widehat{q}_{\alpha_1 \cdots \alpha_m} S_j^{\alpha_1} \cdots S_j^{\alpha_m} \right]}{Z_j^{C_j+1}} \right] \\
&\times \prod_{l=1}^{M} \mathrm{Tr}_{\{\tau_l^\alpha\}} \left[\left\langle \exp \left[F_n \beta \zeta \sum_{\alpha=1}^{n} \tau_l^\alpha \right] \right\rangle_\zeta \right. \\
&\times \left. \oint \frac{dY_l}{2\pi i} \frac{\exp \left[Y_l \sum_{m=0}^{n} \sum_{\langle \alpha_1 \cdots \alpha_m \rangle} \widehat{r}_{\alpha_1 \cdots \alpha_m} \tau_l^{\alpha_1} \cdots \tau_l^{\alpha_m} \right]}{Y_l^{D_l+1}} \right], \qquad (5.26)
\end{aligned}
$$

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. (4.21).

5.5.2 Replica symmetric solution

The replica symmetry assumption is enforced by using the ansätze:

$$
q_{\alpha_1 \cdots \alpha_m} = \int dx \; \pi(x) \; x^m \qquad \widehat{q}_{\alpha_1 \cdots \alpha_m} = \int d\widehat{x} \; \widehat{\pi}(\widehat{x}) \; \widehat{x}^m \qquad (5.27)
$$

and

$$
r_{\alpha_1 \cdots \alpha_m} = \int dy \; \phi(y) \; y^m \qquad \widehat{r}_{\alpha_1 \cdots \alpha_m} = \int d\widehat{y} \; \widehat{\phi}(\widehat{y}) \; \widehat{y}^m. \qquad (5.28)
$$

By plugging the above ansätze, using the limit $\gamma \to \infty$ and standard techniques (see Appendix

D.3 for details) the following expression for the free-energy:

$$
\begin{aligned}
f &= \frac{1}{\beta}\,\mathrm{Extr}_{\{\widehat{\pi},\pi,\widehat{\phi},\phi\}}\Bigg\{\alpha\ \ln 2 \\
&+ \overline{C}\int dx\ \pi(x)\ d\widehat{x}\ \widehat{\pi}(\widehat{x})\ \ln\left(1+x\widehat{x}\right) + \alpha\,\overline{D}\int dy\ \phi(y)\ d\widehat{y}\ \widehat{\phi}(\widehat{y})\ \ln\left(1+y\widehat{y}\right) \\
&- \alpha\int\left[\prod_{j=1}^{K} dx_j\pi(x_j)\right]\left[\prod_{l=1}^{L} dy_l\phi(y_l)\right]\ \ln\left(1+\prod_{j=1}^{K}x_j\prod_{l=1}^{L}y_l\right) \\
&- \sum_{C}\mathcal{P}_C(C)\int\left[\prod_{j=1}^{C} d\widehat{x}_j\ \widehat{\pi}(\widehat{x}_j)\right]\left\langle\ \ln\left[e^{\xi\beta F_s}\prod_{j=1}^{C}(1+\widehat{x}_j)+e^{-\xi\beta F_s}\prod_{j=1}^{C}(1-\widehat{x}_j)\right]\right\rangle_{\xi} \\
&- \alpha\sum_{D}\mathcal{P}_D(D)\int\left[\prod_{l=1}^{D} d\widehat{y}_l\ \widehat{\phi}(\widehat{y}_l)\right]\left\langle\ \ln\left[e^{\zeta\beta F_n}\prod_{l=1}^{D}(1+\widehat{y}_l)+e^{-\zeta\beta F_n}\prod_{l=1}^{D}(1-\widehat{y}_l)\right]\right\rangle_{\zeta}\Bigg\},
\end{aligned}
\tag{5.29}
$$

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 extremisation above, restricted to the space of normalised functions, we find the following saddle-point equations:

$$
\begin{aligned}
\widehat{\pi}(\widehat{x}) &= \int\prod_{j=1}^{K-1} dx_j\ \pi(x_j)\prod_{l=1}^{L} dy_l\ \phi(y_l)\ \delta\left[\widehat{x}-\prod_{j=1}^{K-1}x_j\prod_{l=1}^{L}y_l\right] \\
\pi(x) &= \frac{1}{\overline{C}}\sum_{C} C\ \mathcal{P}_C(C)\int\prod_{l=1}^{C-1} d\widehat{x}_l\ \widehat{\pi}(\widehat{x}_l)\ \left\langle\delta\left[x-\tanh\left(\beta F_s\xi\ +\ \sum_{l=1}^{C-1}\mathrm{atanh}\ \widehat{x}_l\right)\right]\right\rangle_{\xi}, \\
\widehat{\phi}(\widehat{y}) &= \int\prod_{l=1}^{L-1} dy_l\ \phi(y_l)\prod_{j=1}^{K} dx_j\ \pi(x_j)\ \delta\left[\widehat{y}-\prod_{l=1}^{L-1}y_l\prod_{j=1}^{K}x_j\right] \\
\phi(y) &= \frac{1}{\overline{D}}\sum_{D} D\ \mathcal{P}_D(D)\int\prod_{l=1}^{D-1} d\widehat{y}_l\ \widehat{\phi}(\widehat{y}_l)\ \left\langle\delta\left[y-\tanh\left(\beta F_n\zeta\ +\ \sum_{l=1}^{D-1}\mathrm{atanh}\ \widehat{y}_l\right)\right]\right\rangle_{\zeta}.
\end{aligned}
\tag{5.30}
$$

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

$$
\begin{aligned}
\rho &= \int dh\ P(h)\ \mathrm{sgn}(h) \\
P(h) &= \sum_{C}\mathcal{P}_C(C)\int\prod_{l=1}^{C} d\widehat{x}_l\ \widehat{\pi}(\widehat{x}_l)\ \left\langle\delta\left[h-\tanh\left(\beta F_s\xi\ +\ \sum_{l=1}^{C}\mathrm{atanh}\ \widehat{x}_l\right)\right]\right\rangle_{\xi}.
\end{aligned}
\tag{5.31}
$$

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

$$
\begin{aligned}
s &= \beta(u(\beta)\ -\ f) \\
u(\beta) &= -\sum_{C}\mathcal{P}_C(C)\int\prod_{j=1}^{C} d\widehat{x}_j\ \widehat{\pi}^*(\widehat{x}_j)\left\langle F_s\xi\frac{e^{\beta F_s\xi}\prod_{j=1}^{C}(1+\widehat{x}_j)\ -\ e^{-\beta F_s\xi}\prod_{j=1}^{C}(1-\widehat{x}_j)}{e^{\beta F_s\xi}\prod_{j=1}^{C}(1+\widehat{x}_j)\ +\ e^{-\beta F_s\xi}\prod_{j=1}^{C}(1-\widehat{x}_j)}\right\rangle_{\xi} \\
&- \alpha\sum_{D}\mathcal{P}_D(D)\int\prod_{j=1}^{D} d\widehat{y}_j\ \widehat{\phi}^*(\widehat{y}_j)\left\langle F_n\zeta\frac{e^{\beta F_n\zeta}\prod_{j=1}^{D}(1+\widehat{y}_j)\ -\ e^{-\beta F_n\zeta}\prod_{j=1}^{D}(1-\widehat{y}_j)}{e^{\beta F_n\zeta}\prod_{j=1}^{D}(1+\widehat{y}_j)\ +\ e^{-\beta F_n\zeta}\prod_{j=1}^{D}(1-\widehat{y}_j)}\right\rangle_{\zeta},
\end{aligned}
\tag{5.32}
$$

where starred distributions are solutions for (5.30) 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

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

5.6 Probability propagation decoding

In Chapters 3 and 4 we derived probability propagation equations firstly by assuming a set of factorisation properties and writing a closed set of equations that allowed the iterative computation of the (approximate) marginal posterior and secondly by computing local magnetisations on the interior of a Husimi cactus (Bethe approximation). The two methods are equivalent as the factorisation properties assumed in the former are encoded in the geometry of the lattice assumed in the second derivation.

Here we use the insights provided in the last chapters to directly build a decoding algorithm for MN codes. From the replica symmetric free-energy (5.29) we can write the following Bethe free-energy:

$$\begin{aligned}
\mathcal{F}(\boldsymbol{m}, \widehat{\boldsymbol{m}}) \;=\; & \frac{M}{N} \ln 2 + \frac{1}{N} \sum_{\mu=1}^{M} \sum_{i \in \mathcal{L}_s(\mu)} \ln \left(1 + m^s_{\mu i} \widehat{m}^s_{\mu i} \right) + \frac{1}{N} \sum_{\mu=1}^{M} \sum_{j \in \mathcal{L}_n(\mu)} \ln \left(1 + m^n_{\mu j} \widehat{m}^n_{\mu j} \right) \\
& - \frac{1}{N} \sum_{\mu=1}^{M} \ln \left(1 + \mathcal{J}_\mu \prod_{i \in \mathcal{L}_s(\mu)} m^s_{\mu i} \prod_{j \in \mathcal{L}_n(\mu)} m^n_{\mu j} \right) \\
& - \frac{1}{N} \sum_{i=1}^{N} \ln \left[e^{F_s} \prod_{\mu \in \mathcal{M}_s(i)} \left(1 + \widehat{m}^s_{\mu i} \right) + e^{-F_s} \prod_{\mu \in \mathcal{M}_s(i)} \left(1 - \widehat{m}^s_{\mu i} \right) \right] \\
& - \frac{1}{N} \sum_{j=1}^{M} \ln \left[e^{F_n} \prod_{\mu \in \mathcal{M}_n(j)} \left(1 + \widehat{m}^n_{\mu j} \right) + e^{-F_n} \prod_{\mu \in \mathcal{M}_n(j)} \left(1 - \widehat{m}^n_{\mu j} \right) \right] .
\end{aligned} \tag{5.33}$$

The variables $m^s_{\mu j}$ ($m^n_{\mu j}$) are cavity effective magnetisations 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 $\widehat{m}^s_{\mu j}$ ($\widehat{m}^n_{\mu j}$) correspond to effective magnetisations of signal (noise) bits due to the coupling μ only.

The decoding solutions are fixed points of the free-energy (5.33) given by :

$$\frac{\partial \mathcal{F}(\boldsymbol{m}, \widehat{\boldsymbol{m}})}{\partial m^s_{\mu j}} = 0 \qquad \frac{\partial \mathcal{F}(\boldsymbol{m}, \widehat{\boldsymbol{m}})}{\partial \widehat{m}^s_{\mu j}} = 0 \tag{5.34}$$

$$\frac{\partial \mathcal{F}(\boldsymbol{m}, \widehat{\boldsymbol{m}})}{\partial m^n_{\mu j}} = 0 \qquad \frac{\partial \mathcal{F}(\boldsymbol{m}, \widehat{\boldsymbol{m}})}{\partial \widehat{m}^n_{\mu j}} = 0 \tag{5.35}$$

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

$$m^s_{\mu l} = \tanh \left[\sum_{\nu \in \mathcal{M}_s(l) \backslash \mu} \operatorname{atanh}(\widehat{m}^s_{\nu l}) + F_s \right] \qquad \widehat{m}^s_{\mu j} = \mathcal{J}_\mu \prod_{i \in \mathcal{L}_s(\mu) \backslash j} m^s_{\mu i} \prod_{l \in \mathcal{L}_n(\mu)} m^n_{\mu l}, \tag{5.36}$$

$$m^n_{\mu l} = \tanh \left[\sum_{\nu \in \mathcal{M}_n(l) \backslash \mu} \operatorname{atanh}(\widehat{m}^n_{\nu l}) + F_n \right] \qquad \widehat{m}^n_{\mu j} = \mathcal{J}_\mu \prod_{i \in \mathcal{L}_s(\mu)} m^s_{\mu i} \prod_{l \in \mathcal{L}_n(\mu) \backslash j} m^n_{\mu l}. \tag{5.37}$$

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

$$m_j^s = \tanh\left[\sum_{\nu\in\mathcal{M}_s(j)} \mathrm{atanh}(\widehat{m}_{\nu j}^s) + F_s\right] \tag{5.38}$$

One possibility for a decoding dynamics is to update Eqs. (5.36) and (5.37) until a certain halting criteria is reached, and then computing the estimate for the message using equation (5.38). The initial conditions are set to reflect the prior knowledge about the message $m_{\mu j}^s(0) = 1 - 2p_\xi$ and noise $m_{\mu l}^n(0) = 1 - 2p$.

As the prior information is fairly 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 $2(NK + MC)$ equations (5.34) only need to be iterated a $\mathcal{O}(1)$ number of times to get a successful decoding. In 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 [Mac95, Mac99]. Here no essential correlations except those related to the presence of loops are disregarded.

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

5.7 Equilibrium results and decoding performance

The saddle-point equations (5.30) can be solved by using Monte-Carlo integration iteratively. In this section we show, that in terms of the typical performance, MN codes can be divided into three parameter groups: $K \geq 3$ (or $L \geq 3$) and $K > 1$, $K = 2$ and $L = 2$ and $K = 1$. We, therefore, treat each these cases separately in the following.

5.7.1 Analytical solution: the case of $K \geq 3$ or $L \geq 3$ and $K > 1$

Replica symmetric results for the cases of $K \geq 3$ or $L \geq 3$, $K > 1$ 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, characterised by $\rho = 1$, and the probability distributions

$$\pi(x) = \delta(x-1),\ \widehat{\pi}(\widehat{x}) = \delta(\widehat{x}-1),\ \phi(y) = \delta(y-1),\ \widehat{\phi}(\widehat{y}) = \delta(\widehat{y}-1)\ ; \tag{5.39}$$

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

$$\begin{aligned} \pi(x) &= \delta(x),\ \widehat{\pi}(\widehat{x}) = \delta(\widehat{x}),\ \widehat{\phi}(\widehat{y}) = \delta(\widehat{y}), \\ \phi(y) &= \frac{1+\tanh(F_n)}{2}\,\delta(y-\tanh(F_n)) + \frac{1-\tanh(F_n)}{2}\,\delta(y+\tanh(F_n)), \end{aligned} \tag{5.40}$$

satisfy replica symmetric saddle-point equations (5.30). Other solutions could be obtained numerically. To check for that, we represented the distributions with histograms of 20000 bins and iterated Eqs. (5.30) $100-500$ 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. (5.29), which yields

$$f_{\rm FERRO} = -\frac{C}{K}\, F_n \,\tanh(F_n), \tag{5.41}$$

for the ferromagnetic solution and

$$f_{\rm PARA} = \frac{C}{K}\,\ln 2 - \ln 2 - \frac{C}{K}\,\ln\,\left(2\,\cosh(F_n)\right), \tag{5.42}$$

for the paramagnetic solution.

Figure 5.1(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 is given by

$$f_{\rm FERRO} - f_{\rm PARA} = \frac{\ln 2}{R_c}\left[R_c - 1 + H_2(p)\right] = 0. \tag{5.43}$$

This can be rewritten as

$$R_c = 1 - H_2(p) = 1 + p\log_2(p) + (1-p)\log_2(1-p), \tag{5.44}$$

which coincides with Shannon's channel capacity and is represented in Fig. 5.2(a) together with the overlap ρ as a function of the noise level p.

Equation (5.44) seems to indicate that all constructions with either $K \geq 3$ or $L \geq 3$ (and $K > 1$) might 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 it is described in Fig. 5.1(a), the paramagnetic state is also stable for any noise level what has dynamical 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. The entropies can be computed by simply plugging distributions (5.39) and (5.41) into Eq. (5.32). The energy densities for the unbiased case are $u = u_{\rm PARA} = u_{\rm FERRO} = -\alpha\, F_n\,(1-2p)$, since the Nishimori condition is employed (see Appendix C.3). The ferromagnetic entropies are $s_{\rm FERRO} = u - f_{\rm FERRO} = 0$ and

$$\begin{aligned} s_{\rm PARA} &= u - f_{\rm PARA} \\ &= -\alpha\, F_n\,(1-2p) - \frac{C}{K}\,\ln 2 + \ln 2 + \frac{C}{K}\,\ln\,\left(2\,\cosh(F_n)\right). \end{aligned} \tag{5.45}$$

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

This indicates that the distribution (5.41) is non-physical below p_c, despite being a solution of replica symmetric saddle-point equations. This result seems to indicate that the replica symmetric free-energy do not provide the right description below p_c. A simple alternative is to use a frozen spins solution as the formulation of a theory with replica symmetry breaking for highly diluted systems is a difficult task (see, for example, [WS88, Mon98b]).

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. Moreover, even without knowing the correct physics below p_c, it is possible in this case use an exhaustive search of the global minimum of the free-energy in Section 5.6 to attain Shannon's bound in exponential time.

5.7.2 The case of $K = 2$ and $L = 2$

All codes with either $K = 3$ or $L = 3$, $K > 1$ potentially saturate Shannon's bound and are characterised by a first order phase transition between the ferromagnetic and paramagnetic solutions. Solutions for the case with $K = L = 2$ can be obtained numerically, yielding significantly different physical behaviour as shown in Fig.5.1(b).

At very large noise levels, the paramagnetic solution (5.41) gives the unique extremum of the free-energy until the noise level reaches p_1, at which the ferromagnetic solution (5.39) of higher free-energy becomes locally stable. As the noise level decreases to p_2 the paramagnetic solution becomes unstable and a sub-optimal 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 sub-optimal ferromagnetic solutions remain locally stable. However, the sub-optimal 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.(5.30).

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, $K = L = 2$ 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 an unique minimum for noise levels $0 < p < p_s$, offers significant advantages in the decoding dynamics comparing to that of codes with $K \geq 3$ or $L \geq 3$, $K > 1$, allowing the successful use of the polynomial time probability propagation decoding.

5.7.3 The case of $K = 1$ and general L

The choice of $K = 1$, independently of the value chosen for L, exhibits a different behaviour presented schematically in Fig.5.1(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

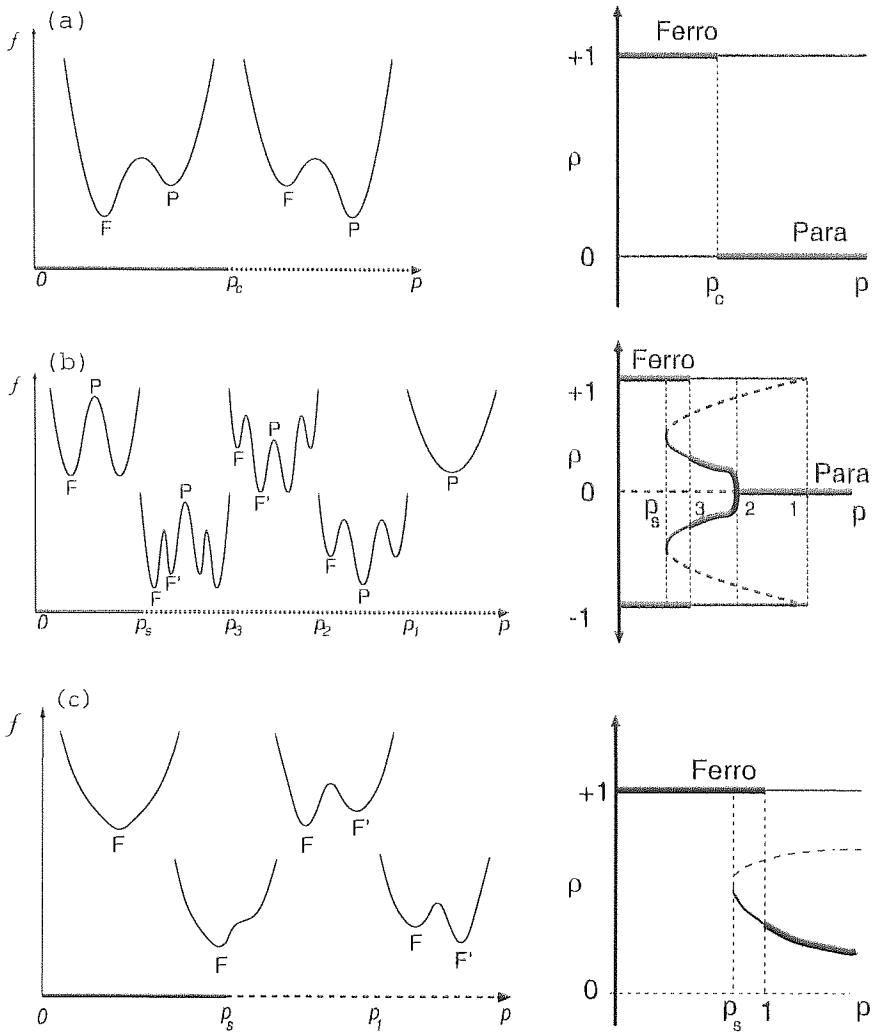


Figure 5.1: Figures in the left side show schematic representations free-energy landscapes while figures on the right show overlaps ρ a function of the noise level p; thick and thin lines denote stable solutions of lower and higher free energies respectively, dashed lines correspond to unstable solutions. (a) $K \geq 3$ or $L \geq 3$, $K > 1$. 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, that coincides with the channel capacity. (b) $K = 2$ and $L = 2$; the ferromagnetic solution and its mirror image are the only minima of the free-energy up to p_s (solid line). Above p_s sub-optimal 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 sub-optimal ferromagnetic state (F') appears in the region represented by the dashed line. The dynamical 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.

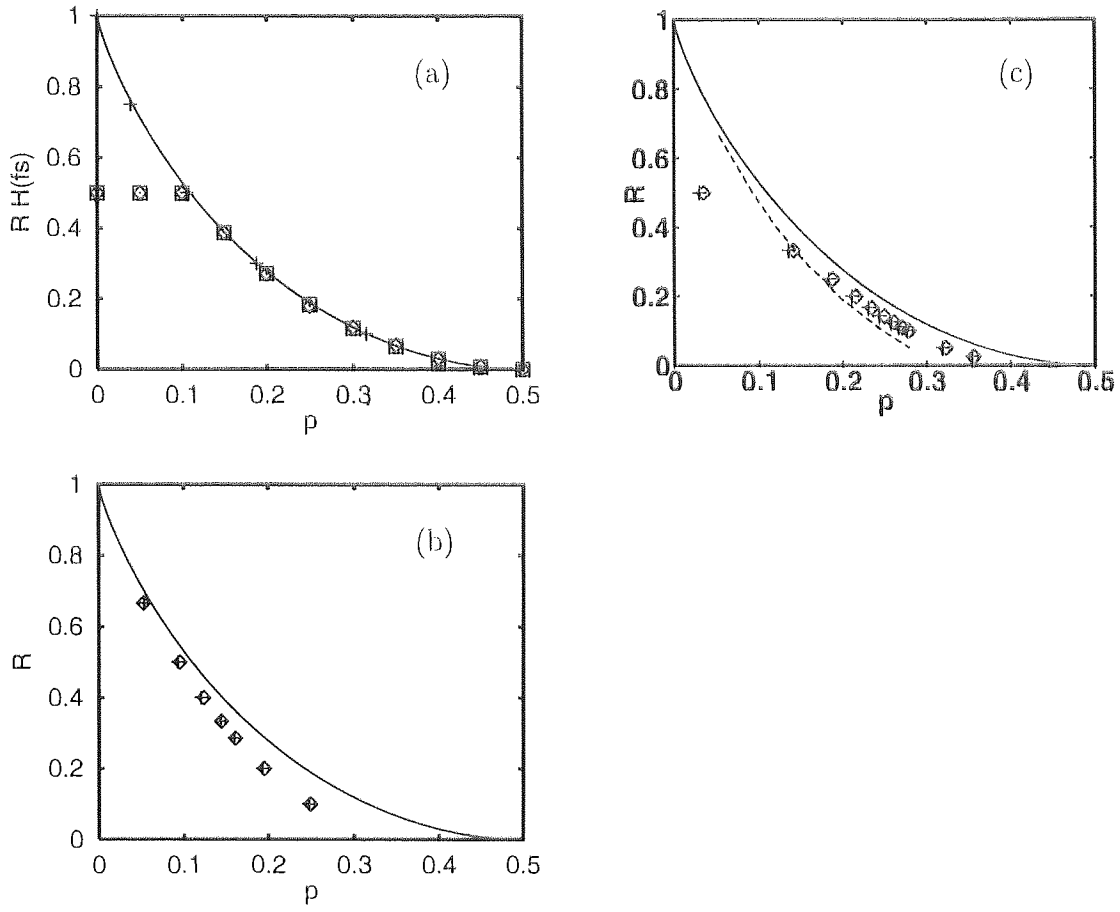


Figure 5.2: Transition lines in the plane rate R *versus* the flip rate p, obtained from numerical solutions and the TAP approach ($N = 10^4$), and averaged over 10 different initial conditions with error bars much smaller than the symbols size. (a) Numerical solutions for $K = L = 3$, $C = 6$ and varying input bias f_s ($\Box$) and TAP solutions for both unbiased (+) and biased ($\diamond$) 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 ($\diamond$) were chosen to be of almost zero magnetisation. (c) For the case of $K = 1$, $L = 2$ and unbiased messages. We show numerical solutions of the analytical equations ($\diamond$) 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 $R < 1/3$.

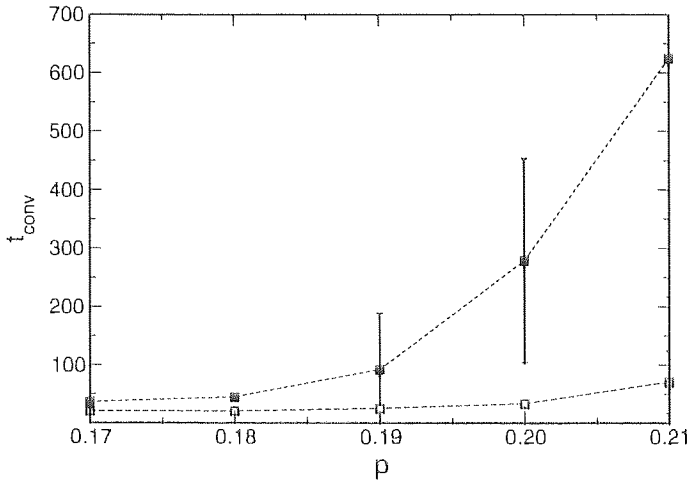


Figure 5.3: Number $t_{\rm conv}$ of iterations needed for convergence of the probability propagation algorithm versus noise level p. The system size is $N = 5000$, rate $R = 1/5$, $K = 2$ codes are represented by black triangles while $K = 1$ codes are represented by white squares. averages are over 2 random codes and 10 runs for each code.

to be noted is that the paramagnetic state (5.41) is no longer a solution of the saddle-point equations (5.30) and is being replaced by a sub-optimal 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.

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. 5.2) while offering improved decoding times when probability propagation is used (Fig. 5.3). Studying the replica symmetric free-energy in this case shows that as the corruption rate increases, sub-optimal ferromagnetic solutions (stable and unstable) emerge at the spinodal point p_s. When the noise increases further this sub-optimal 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 Fig.5.4 we show free-energy values computed for a given code rate and several values of L, pointing Shannon's bound by a dashed line; the thermodynamic transition observed numerically (i.e. the point where the ferromagnetic free-energy equals the sub-optimal ferromagnetic free-energy) coincides with Shannon's bound within the numerical precision used. Spinodal noise levels are indicated by arrows. In Fig. 5.5 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 equations (5.30) we can deduce the asymptotic behaviour of $K = 1$ and $L = 2$ codes for small rates (large C) by computing the two first cummulants of the distributions $\pi, \widehat{\pi}, \phi$ and $\widehat{\phi}$ (Gaussian approximation). A decoding failure corresponds to $\langle h \rangle \sim \mathcal{O}(1)$ and $\sigma_h^2 \sim \mathcal{O}(1)$. It implies that $\langle \widehat{x} \rangle \sim \mathcal{O}(1/C)$ and $\sigma_{\widehat{x}} \sim \mathcal{O}(1/C)$. For that y must be small and we

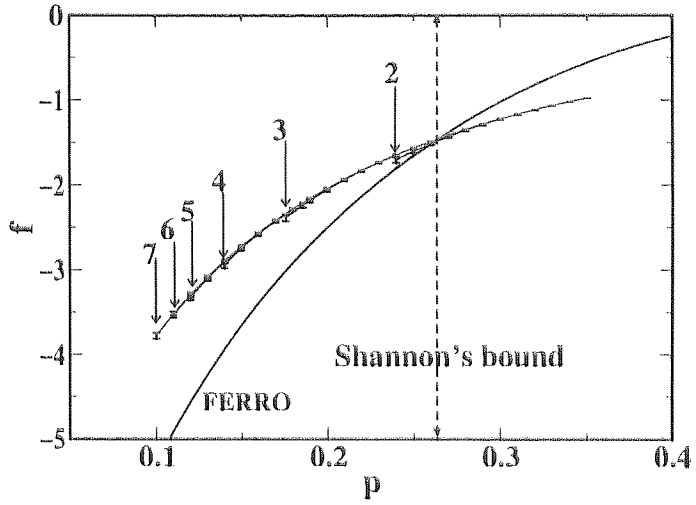


Figure 5.4: Free-energies obtained by solving the analytical equations using Monte-Carlo integrations for $K = 1$, $R = 1/6$ and several values of L. Full lines represent the ferromagnetic free-energy (FERRO, higher on the right) and the suboptimal ferromagnetic free-energy (higher on the left) for values of $L = 1, ..., 7$. The dashed line indicates Shannon's bound and the arrows represent the spinodal point values p_s for $L = 2, ..., 7$. The thermodynamic transition coincides with Shannon's bound.

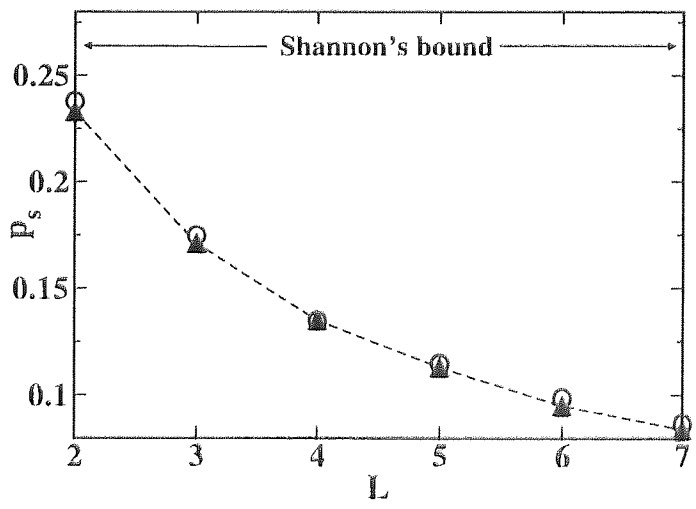


Figure 5.5: Spinodal point noise level p_s for $K = 1$, $R = 1/6$ and several choices of L. Numerical solutions are denoted by circles and PP decoding solutions (10 runs with size $N = 10^4$) by black triangles. Symbols are larger than the error bars.

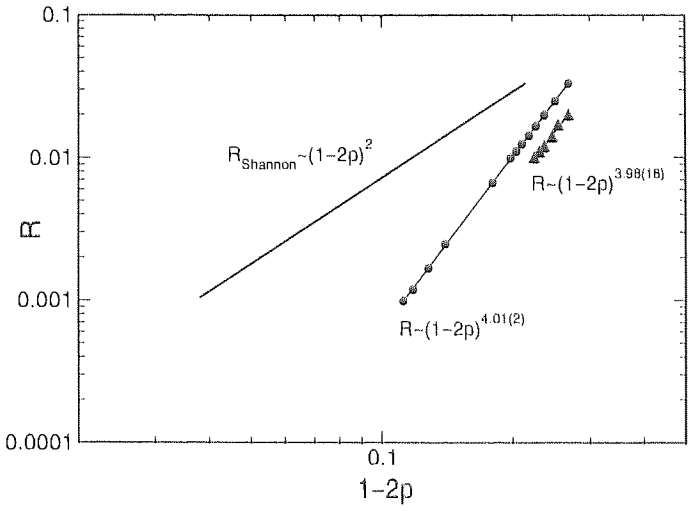


Figure 5.6: Asymptotic behaviour of the transition for small rates. The full line represents Shannon's bound, circles represent transitions obtained by using only the first cummulants and squares correspond to the Gaussian approximation.

can use $\text{atanh}(\tanh(y_1)\tanh(y_2)) \approx y_1 y_2$ and write :

$$\langle x \rangle \sim \mathcal{O}(1) \qquad \sigma_x^2 \sim \mathcal{O}(1) \tag{5.46}$$

$$\langle \widehat{x} \rangle \approx \langle y \rangle^2 \tag{5.47}$$

$$\sigma_{\widehat{x}}^2 \approx \langle y^2 \rangle^2 - \langle y \rangle^4 \tag{5.48}$$

$$\langle y \rangle = \langle \widehat{y} \rangle + (1-2p)F_n \qquad \sigma_y^2 = \sigma_{\widehat{y}}^2 + 4f(1-p)F_n^2 \tag{5.49}$$

$$\langle \widehat{y} \rangle \approx \langle \tanh(x) \rangle \langle y \rangle \tag{5.50}$$

$$\sigma_{\widehat{y}}^2 \approx \langle \tanh^2(x) \rangle \langle y^2 \rangle - \langle \tanh(x) \rangle^2 \langle y \rangle^2 \tag{5.51}$$

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 $\widehat{x} \sim \mathcal{O}(1/C)$ 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 $R = 1/C$:

$$R \sim \frac{(1-2p)^4}{(1-\langle \tanh x \rangle)^2} \tag{5.52}$$

The asymptotic behaviour of the Shannon bound is given by :

$$R \sim \frac{(1-2p)^2}{\ln 2} \tag{5.53}$$

Thus, the $K = 1$ and $L = 2$ codes are not asymptotically optimal. In Fig. 5.6 we verify (5.52) by iterating equations for first cummulants (delta approximation) and by first and second cummulants (Gaussian approximation).

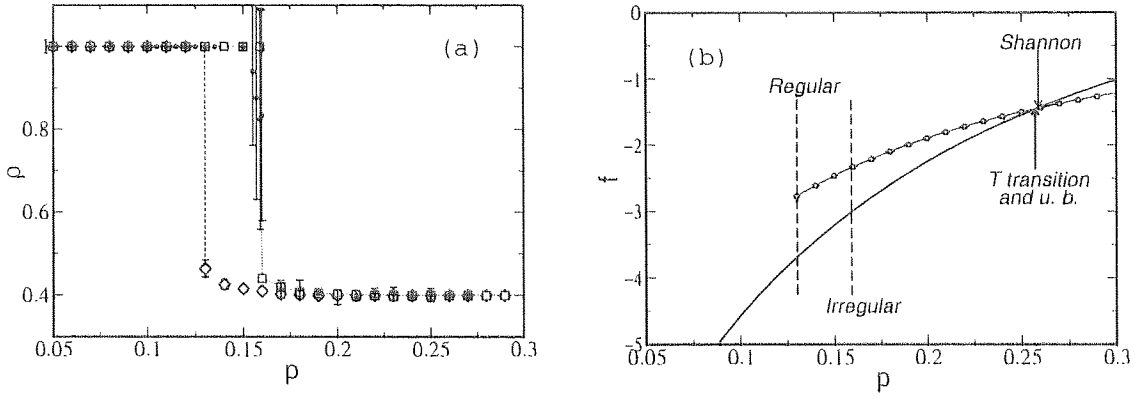


Figure 5.7: (a) Overlap as a function of the noise level p for codes with $K = L = 3$ and $\overline{C} = 15$ with message bias $p_\xi = 0.3$. Analytical RS solutions for the regular code are denoted as $\diamond$ and for the irregular code; with $C_o = 4$ and $C_e = 30$ denoted as $\square$. Results 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 5.3 and Shannon's limit.

5.8 Error-correction: regular vs. irregular codes

Irregularity can improve the practical performance of MN codes. This fact has been already reported in the information theory literature (see for example [Dav99, Dav98, LMSS98]). Here we analyse this problem by using the language and tools of statistical physics. We now use the simplest irregular constructions as an illustration, to say, the connectivities of the signal matrix $\boldsymbol{C}_s$ are described by a simple bimodal probability distribution:

$$\mathcal{P}_C(C) = (1-\theta)\,\delta(C - C_o) \;+\; \theta\,\delta(C - C_e). \tag{5.54}$$

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 on the effect of irregularity on solving the PP equations (5.36) and (5.37) 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 equations (5.30) obtained in the replica symmetric (RS) theory, setting the initial conditions to be $\pi_0(x) = (1 - p_x i)\;\delta(x - m^s_{\mu j}(0)) \;+\; p_x i\;\delta(x + m^s_{\mu j}(0))$ 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 Fig. 5.7 (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 per-

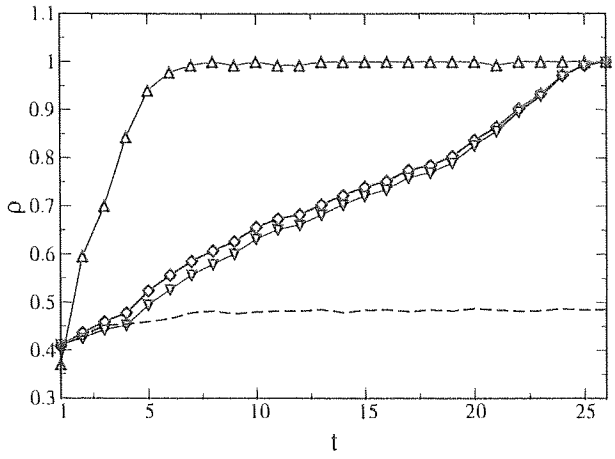


Figure 5.8: Overlap monitored during the PP decoding process as a function of the number of iterations for $N = 4000$. Elite nodes overlap is represented by $\triangle$. Ordinary nodes overlap is represented by $\bigtriangledown$. The overall overlap is represented by $\diamond$. The long dashed line shows the dynamics of the regular code. The constructions employed have parameters $K = L = 3$, $\overline{C} = 6$, $C_e = 20$ and $C_o = 5$. The noise level is $p = 0.065$ and the message bias is $p_\xi = 0.3$.

formance considerably. In Fig. 5.7 (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. 5.7 (a) with overlap $\rho = 0.4$) is shown as a line marked with $\circ$. The transitions seen in Fig. 5.7(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 transition. The thermodynamic transition coincides with the upper bound (u.b.) in Section 5.3 and is very close to, but below, Shannon's limit which is shown for comparison. Similar behaviour was observed in regular MN codes with $K = 1$ in Section 5.7.3.

It is instructive to look how the overlap of elite (m_e) and ordinary (m_o) nodes evolve throughout the iterative decoding process. In Fig. 5.8 we show this dynamics for a regular and an irregular code at a noise level where the irregular code converges to the ferromagnetic state while the regular code fails (long-dashed lines). One can see that the overlap of ordinary nodes follow that of the regular code in the first iterations, elite nodes are then corrected quickly achieving high overlaps. These highly reliable nodes then lead the correction of ordinary nodes, producing successful decoding. From the decoding dynamics point of view the typical performance of irregular MN codes can be qualitatively regarded as a mixture of low and highly connected regular codes where elite nodes can tolerate higher noise levels while ordinary nodes allow for higher code rates.

5.9 The *spinodal* noise level

In the last section we gained some insight into how irregularity affects the practical performance of codes. The dynamical decoding process shown in Figure 5.8 only provides a qualitative explanation and does not seem to allow some simple analysis.

A possible alternative is to relate the observation that the system gets trapped in suboptimal states to global properties of the free-energy. The PP algorithm can be regarded as an iterative solution of fixed point equations for the free-energy (5.33), 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 analyse this point we rerun the 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:

$$P\left(m^s_{\mu j}(0)\right) = (1-\lambda_s)\,\delta(m^s_{\mu j}(0)-\xi_j) \;+\; \lambda_s\,\delta(m^s_{\mu j}(0)+\xi_j) \tag{5.55}$$

and

$$P\left(m^n_{\mu l}(0)\right) = (1-\lambda_n)\,\delta(m^n_{\mu l}(0)-\tau_l) \;+\; \lambda_n\,\delta(m^n_{\mu l}(0)+\tau_l), \tag{5.56}$$

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

We performed PP decoding several times for different values of λ and noise level p. For $\lambda \leq 0.026$ 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. This behaviour have already been observed for regular MN codes with $K = 1$ or $K = L = 2$ in Sections 5.7.3 and 5.7.2.

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

5.9.1 Biased messages: $K \geq 3$ or $L \geq 3$

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 choices are arbitrary but illustrate what happens with the practical decoding performance. In Fig. 5.9 we show the transition from the decoding phase to the failure phase as a function of the noise level p for several rates R in both regular and irregular

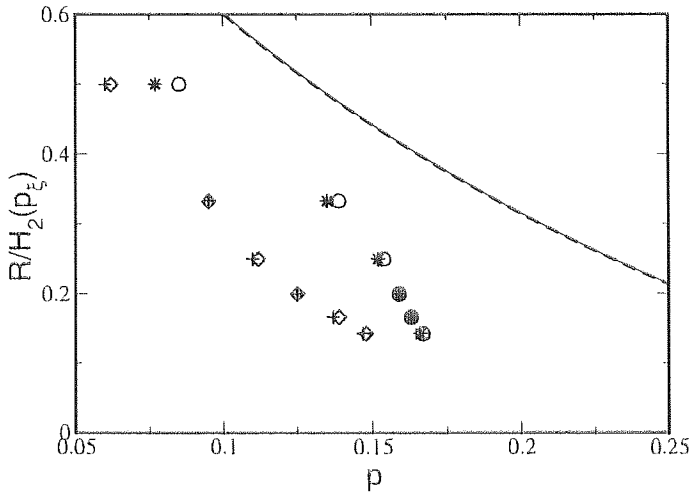


Figure 5.9: *Spinodal* noise level p_s for regular and irregular codes. In both constructions parameters are set as $K = L = 3$. Irregular codes with $C_o = 4$ and $C_e = 30$ are used. PP decoding is carried out with $N = 5000$ and a maximum of 500 iterations; they are denoted by $+$ (regular) and $*$ (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 5.3 is represented by a dashed line. The symbols are chosen to be larger than the actual error bars.

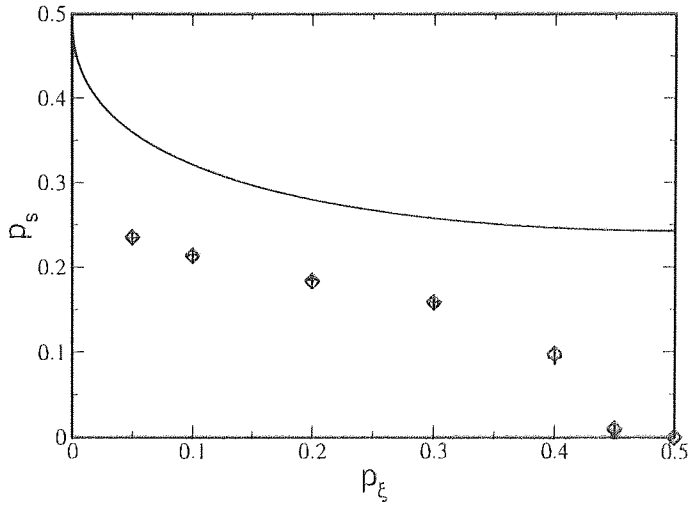


Figure 5.10: Spinodal noise level p_s for irregular codes as a function of the message bias p_ξ. The construction is parameterised by $K = L = 3$, $C_o = 4$ and $C_e = 30$ with $\overline{C} = 15$. PP decoding is carried out with $N = 5000$ and a maximum of 500 iterations, and is represented by $+$, while theoretical RS solutions are represented by $\diamond$. The full line indicates Shannon's limit. Symbols are larger than the actual error bars

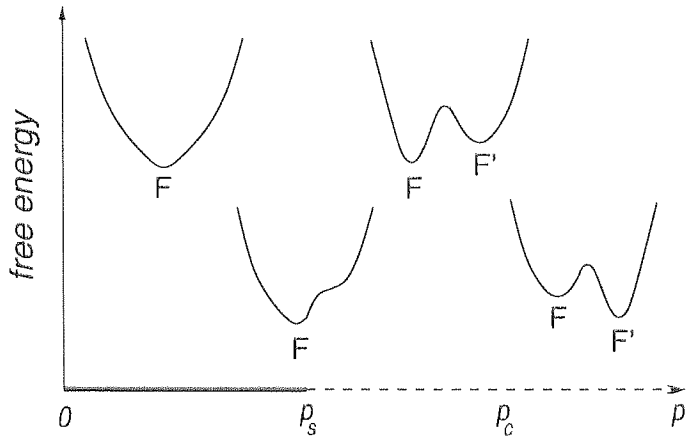


Figure 5.11: Pictorial representation of the free-energy landscape for codes with $K \geq 3$ or $L \geq 3$, $K > 1$ and biased messages $p_\xi < 0.5$ as a function of the noise level p. Up to the spinodal noise level p_s there is only the ferromagnetic state F. At p_s another state F' appears , dominating the decoding dynamics. The critical noise level p_c indicates the point where the state F' becomes the global minimum (thermodynamic transition).

codes. Practical decoding ($\diamond$ an $\circ$) 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 equations (5.30) (represented as + and $*$ in Fig. 5.9) 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 Figure 5.9 is already fairly significant.

The analytical solution obtained in Section 5.7.1, for $K \geq 3$ or $L \geq 3$, $K > 1$ 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 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 C_n and C_s one can determine how the spinodal noise level p_s depends on the bias p_ξ. In Figure 5.10 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 towards the unbiased case. It obviously suggests the use of biased messages for practical use of MN codes with parameters $K \geq 3$ or $L \geq 3$, $K > 1$ under PP decoding.

For biased messages with $K \geq 3$ or $L \geq 3$, $K > 1$ the qualitative picture of the energy landscape differs from the unbiased coding presented in Section 5.7.1. In Fig. 5.11 this landscape is sketched

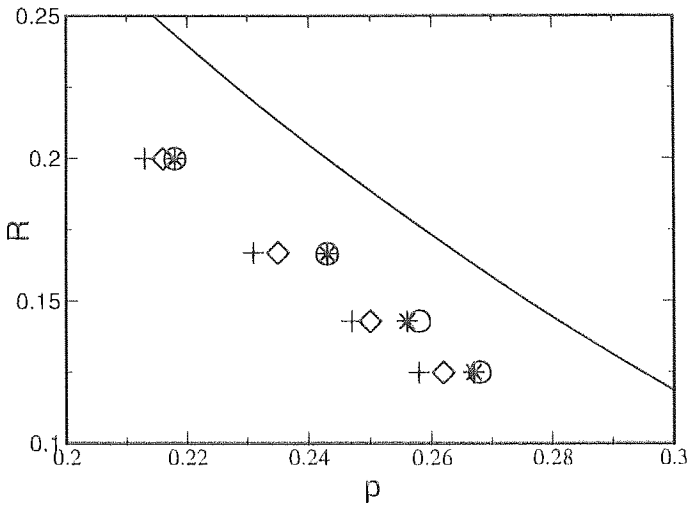


Figure 5.12: Spinodal noise level p_s for regular and irregular codes. The constructions are of $K = 1$ and $L = 2$, irregular codes are parameterised by $C_o = 4$ and $C_e = 10$. PP decoding is carried out with $N = 5000$ and a maximum of 500 iterations ; they are denoted by + (regular) and $*$ (irregular). Numerical solutions for RS equations are denoted by $\diamond$ (regular) and $\circ$ (irregular). The coding limit is represented by a line. Symbols are larger than the actual error bars.

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

5.9.2 Unbiased messages

For the remaining parameter choices, namely general L, $K = 1$ and $K = L = 2$, it was shown in Sections 5.7.2 and 5.7.3 that unbiased coding is generally possible yielding close to Shannon's limit performance.

In the same way as in the $K \geq 3$ case the practical performance is defined by the spinodal noise level p_s. The addition of irregularity also changes p_s in these cases.

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 Fig. 5.12 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 equations (5.30) ($\diamond$ 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.

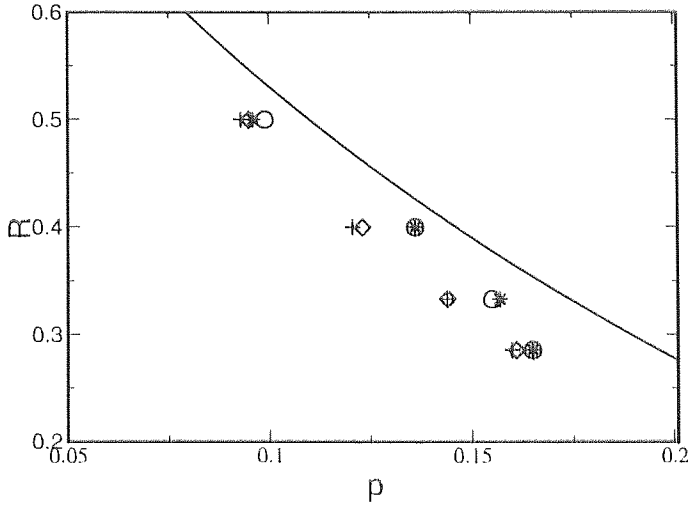


Figure 5.13: Spinodal noise level values p_s for regular and irregular codes. Constructions are of $K = 2$ and $L = 2$, irregular codes are parameterised by $C_o = 3$ and $C_e = 8$. PP decoding is carried out with $N = 5000$ and a maximum of 500 iterations; they are denoted by + (regular) and $*$ (irregular). Theoretical predictions are denoted by $\diamond$ (regular) and $\circ$ (irregular). The coding limit is represented by a line. Symbols are larger than the actual error bars.

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 Fig. 5.13.

5.10 Conclusion

We mapped a general irregular MN codes onto a multi-spins systems and employed the replica theory to compute their typical performance. We divided the codes in terms of performance to three categories: (i) $K > 2$ or $L > 2$, $K > 1$; (ii) $K = L = 2$ and (iii) $K = 1$. For unbiased messages (p_ξ), we obtained an analytical replica symmetric solution for the first case, showing that the thermodynamic transition coincides with the channel capacity and with an information theoretic upper bound. For the second case, we obtained numerical solutions and a transition below the information theoretic upper bound. For $K = 1$ codes we numerically obtained a transition that coincides with the channel capacity within the numerical precision used.

The probability propagation algorithm was linked directly to the replica symmetric free-energy. We discussed the connection between the noise level where spinodal points emerge in the free-energy (p_s) and the performance of PP decoding. We showed that, since unbiased $K > 2$ or $L > 2$, $K > 1$ codes have a bistable free-energy, only $K = L = 2$ and $K = 1$ are suitable for PP decoding, unless biased messages are used. $K = 1$ codes were shown to attain better decoding times.

We also introduced simple irregular constructions, we analysed each one of the three parameter groups, discussing the practical use of biased messages with $K > 2$ or $L > 2$, $K > 1$ and of unbiased

messages with the other two groups. We showed that irregularity changes the free-energy global topology and can easily improve the practical performance threshold p_s.

Chapter 6

Kanter-Saad Codes

In this chapter we obtain a statistical description for the typical PP decoding process of the codes presented by Kanter and Saad, a variation of MN codes. We use this description to optimise the construction of a simple Kanter-Saad code.

6.1 Introduction

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 [KS99b, KS00b, KS00a]. The central idea is to explore the superior dynamical 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 Fig. 6.2, with the number of non-zero elements per row set to several different values $K_1, \cdots, K_m$.

In Fig. 6.1 we show a performance comparison (presented in [KS00b]) of Turbo, KS and Gallager codes with optimised irregular constructions [RSU99] 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 multi-spin interactions of several different orders and of more structured matrices makes the statistical physics of the problem harder to solve. In this chapter we, therefore, adopt a different approach. We first write the probability propagation equations and find an appropriate macroscopic description of them in terms of field distributions. We then solve saddle-point like equations for the field distributions to find the typical performance.

This chapter is organised as follows. Section 6.2 presents the KS constructions in detail. In Section 6.3 a macroscopic description for the decoding process in terms of field distributions is provided. In Section 6.4 we use the macroscopic description to optimise a simple KS construction. Conclusions are provided in Section 6.5. An appendix on the use of cumulant expansions for optimisation is also

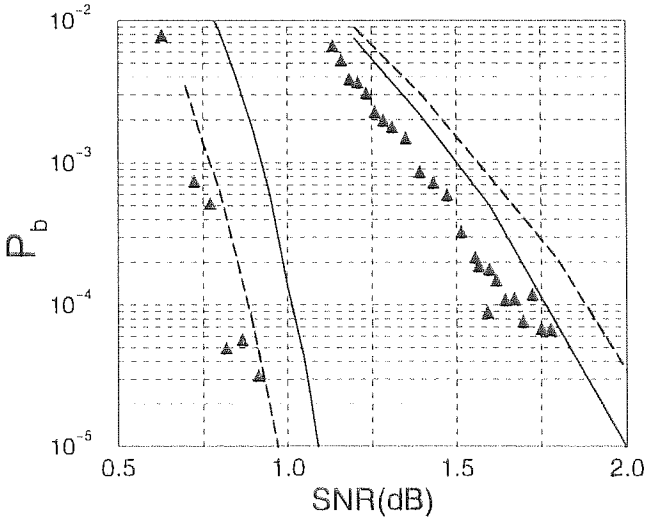


Figure 6.1: Bit error probability p_b as a function of the signal to noise ratio for codes of rate $R = 1/2$, sizes $N = 1000$ (right) and $N = 10000$ (left) in a memoryless Gaussian channel. Black triangles represent KS codes, dashed lines represent Turbo codes and dotted lines represent optimised irregular Gallager codes of similar sizes. This figure was extracted from [KS00b].

provided.

6.2 KS codes

KS codes are specific constructions of MN codes. The signal matrix $\boldsymbol{C}_s$ is defined by m random sub-matrices with $K_1, K_2, \cdots, K_m$ non-zero elements per row respectively. The matrix $\boldsymbol{C}_n$ is composed by two sub-matrices: $\boldsymbol{C}_{n_{ij}}^{(1)} = \delta_{i,j} + \delta_{i,j+\Delta}$ and $\boldsymbol{C}_{n_{ij}}^{(2)} = \delta_{i,j}$. The inverse $\boldsymbol{C}_n^{-1}$ used in the encoding process is easily obtainable. In Fig. 6.2 we represent a KS code with three signal sub-matrices, the non-zero 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 onto $M_j \times N$ sub-matrices, with $j = 1, \cdots, m$. The total number of non-zero elements is given by $N\ C = \sum_{j=1}^{m} M_j\ K_j$ what yields $C = \sum_{j=1}^{m} \alpha_j\ K_j$, where $\alpha_j = M_j/N$. The code construction is, therefore, parametrised by the set $\{(\alpha_j, K_j)\}$. If we fix $\{K_j\}$, the parameters $\{\alpha_j\}$ 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 KS codes are performed in exactly the same fashion as described in Section 5.2 for MN codes. A binary vector $\boldsymbol{t} \in \{0,1\}^M$ defined by

$$\boldsymbol{t} = \boldsymbol{G}\boldsymbol{\xi} \pmod 2, \tag{6.1}$$

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

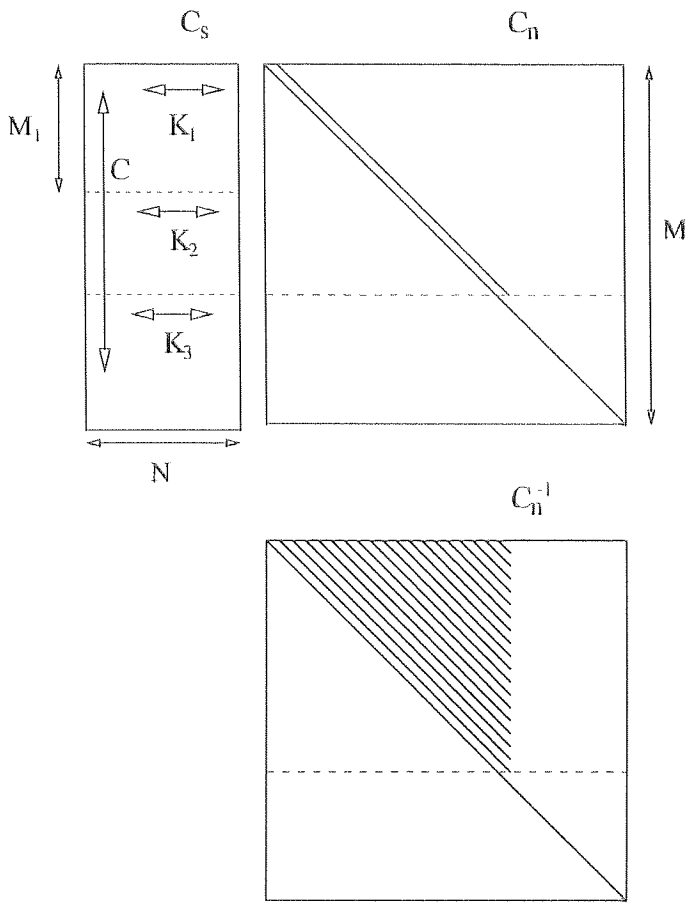


Figure 6.2: KS construction with three signal sub-matrices with K_1, K_2 and K_3 non-zero elements per row, respectively. The number of non-zero elements per column is kept fixed to C. The noise matrix C_n is composed by two sub-matrices, the non-zero elements are denoted by lines. The inverse C_n^{-1} is also represented.

code rate is $R = N/M$. The generator matrix G is a $M \times N$ dense matrix defined by

$$G = C_n^{-1} C_s \pmod 2. \tag{6.2}$$

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

The received vector is

$$r = G\xi + \zeta \pmod 2. \tag{6.3}$$

Decoding is performed by computing the syndrome vector

$$z = C_n r = C_s \xi + C_n \zeta \pmod 2, \tag{6.4}$$

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

6.3 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 sub-matrices denoted **1n** and **2n**. The channel is chosen to be a memoryless binary symmetric channel (BSC). The number of non-zero 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 represent a code in this family in Fig.6.3.

The PP decoding dynamics for these codes is described by Eqs.(5.37). However, due to the irregular character of the construction, sites inside each one of the sub-matrices are connected differently. Reminding the statistical physics formulation of MN codes presented in Section 5.4, non-zero row elements in the matrices depicted in Fig.6.3 correspond to sites taking part in one multi-spin interaction. Therefore, signal sites in the sub-matrix **1s** interact with other $K_1 - 1$ signal sites in **1s** and exactly two noise sites in **1n**. Moreover, the same site takes part in other $\alpha_1 K_1 + \alpha_2 K_2 - 1$ multi-spin couplings in both **1s** and **2s**. Sites in sub-matrix **2s** interact with one noise site in **2n** and $K_2 - 1$ signal sites in **2s**, taking part in other $\alpha_1 K_1 + \alpha_2 K_2 - 1$ multi-spin interaction. Noise sites in the sub-matrix **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:

$$\mathcal{H} = -\gamma \sum_{\mu=1}^{M_1} (\mathcal{J}_\mu\, S_{i_1} \cdots S_{i_{K_1}} \tau_\mu \tau_{\mu+\Delta} - 1) - \gamma \sum_{\mu=M_1+1}^{M} (\mathcal{J}_\mu\, S_{i_1} \cdots S_{i_{K_2}} \tau_\mu - 1) - F_n \sum_{l=1}^{M} \tau_l - F_s \sum_{j=1}^{N} S_j, \tag{6.5}$$

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

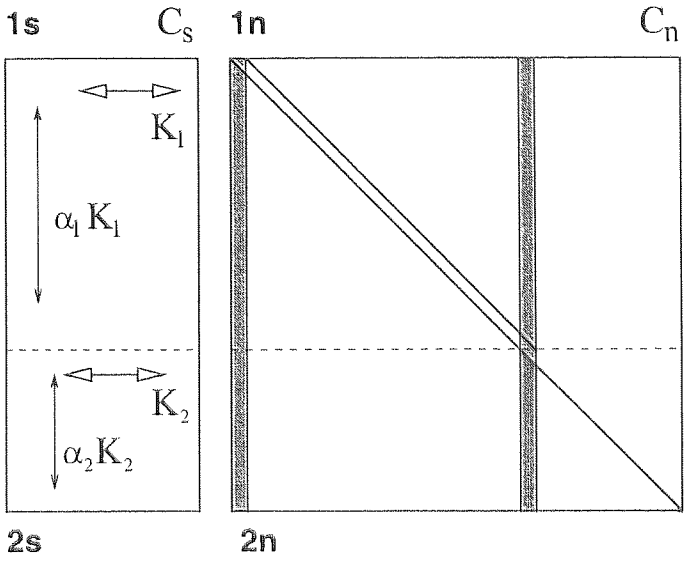


Figure 6.3: KS code with two signal matrices with parameters K_1 and K_2. Note that noise sites inside the shaded regions take part in a different number of interactions than the ordinary sites.

We can write PP decoding equations for each one of the sub-matrices **1s**, **2s**, **1n** and **2n**. The shaded regions in Fig. 6.3 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 $N \to \infty$.

For the sub-matrix **1s** we have:

$$m_{\mu j}^{(1s)} = \tanh\left[\sum_{\nu \in \mathcal{M}_{1s}(j)\setminus\mu} \operatorname{atanh}(\widehat{m}_{\nu j}^{(1s)}) + \sum_{\nu \in \mathcal{M}_{2s}(j)} \operatorname{atanh}(\widehat{m}_{\nu j}^{(2s)}) + F_s\right] \tag{6.6}$$

$$\widehat{m}_{\mu j}^{(1s)} = \mathcal{J}_\mu \, m_{\mu\mu}^{(1n)} m_{\mu\mu+\Delta}^{(1n)} \prod_{l \in \mathcal{L}_{1s}(\mu)\setminus j} m_{\mu l}^{(1s)}, \tag{6.7}$$

where the second equation represents interactions with two noise sites and and $K_1 - 1$ signal sites. The first equation represents the $\alpha_1 K_1 + \alpha_2 K_2 - 1$ multi-spin interactions the site j belongs to.

Similarly, for the sub-matrix **2s** we have:

$$m_{\mu j}^{(2s)} = \tanh\left[\sum_{\nu \in \mathcal{M}_{1s}(j)} \operatorname{atanh}(\widehat{m}_{\nu j}^{(1s)}) + \sum_{\nu \in \mathcal{M}_{2s}(j)\setminus\nu} \operatorname{atanh}(\widehat{m}_{\nu j}^{(2s)}) + F_s\right] \tag{6.8}$$

$$\widehat{m}_{\mu j}^{(2s)} = \mathcal{J}_\mu \, m_{\mu}^{(2n)} \prod_{l \in \mathcal{L}_{2s}(\mu)\setminus j} m_{\mu l}^{(2s)} \tag{6.9}$$

For the sub-matrix **1n** we have:

$$m_{\mu j}^{(1n)} = \tanh\left[\operatorname{atanh}(\widehat{m}_{\nu j}^{(1n)}) + F_n\right] \tag{6.10}$$

$$\widehat{m}_{\mu j}^{(1n)} = \mathcal{J}_\mu \, m_{\mu i}^{(1n)} \prod_{l \in \mathcal{L}_{1s}(\mu)} m_{\mu l}^{(1s)}, \tag{6.11}$$

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

Finally, for sub-matrix **2n** we have:

$$m_{\mu}^{(2n)} = \tanh\left[F_n\right] \tag{6.12}$$

$$\widehat{m}_{\mu}^{(2n)} = \mathcal{J}_{\mu} \prod_{l \in \mathcal{L}_{2s}(\mu)} m_{\mu l}^{(2s)} \tag{6.13}$$

The pseudo-posterior and decoded message are given by :

$$m_j = \tanh\left[\sum_{\nu \in \mathcal{M}_{1s}(j)} \operatorname{atanh}(\widehat{m}_{\nu j}^{(1s)}) + \sum_{\nu \in \mathcal{M}_{2s}(j)} \operatorname{atanh}(\widehat{m}_{\nu j}^{(2s)})\right] \tag{6.14}$$

$$\widehat{\xi}_j = \operatorname{sgn}(m_j). \tag{6.15}$$

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 dynamical 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.

By performing the gauge transformation

$$m_{\mu j}^{(as)} \to \xi_j m_{\mu j}^{(as)} \qquad \widehat{m}_{\nu j}^{(as)} \to \xi_j \widehat{m}_{\mu j}^{(as)} \tag{6.16}$$

$$m_{\mu j}^{(an)} \to \zeta_j m_{\mu j}^{(an)} \qquad \widehat{m}_{\mu j}^{(an)} \to \zeta_j \widehat{m}_{\mu j}^{(an)} \tag{6.17}$$

$$\mathcal{J}_{\mu} \to 1 \qquad (a = 1, 2), \tag{6.18}$$

introducing effective fields $x_{\mu j} = \operatorname{atanh}(m_{\mu j})$, $\widehat{x}_{\mu j} = \operatorname{atanh}(\widehat{m}_{\mu j})$ and assuming that $x_{\mu j}^{(as)}$, $\widehat{x}_{\mu j}^{(as)}$, $y_{\mu j}^{(an)}$, $\widehat{y}_{\mu j}^{(an)}$ are independently drawn from distributions $P_a(x)$, $\widehat{P}_a(\widehat{x})$, $R_a(y)$, $\widehat{R}_a(\widehat{y})$, 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 sub-matrix **1s**:

$$P_1(x) = \int \prod_{j=1}^{\alpha_1 K_1 - 1} d\widehat{x}_j \widehat{P}_1(\widehat{x}_j) \prod_{l=1}^{\alpha_1 K_2} d\widehat{w}_l \widehat{P}_2(\widehat{w}_l)\, \delta\left[x - \sum_{j=1}^{\alpha_1 K_1 - 1} x_j - \sum_{l=1}^{\alpha_2 K_2} w_l\right] \tag{6.19}$$

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

For **2s**:

$$P_2(x) = \int \prod_{j=1}^{\alpha_1 K_1} d\widehat{x}_j \widehat{P}_1(\widehat{x}_j) \prod_{l=1}^{\alpha_1 K_2 - 1} d\widehat{w}_l \widehat{P}_2(\widehat{w}_l)\, \delta\left[x - \sum_{j=1}^{\alpha_1 K_1} x_j - \sum_{l=1}^{\alpha_2 K_2 - 1} w_l\right] \tag{6.20}$$

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

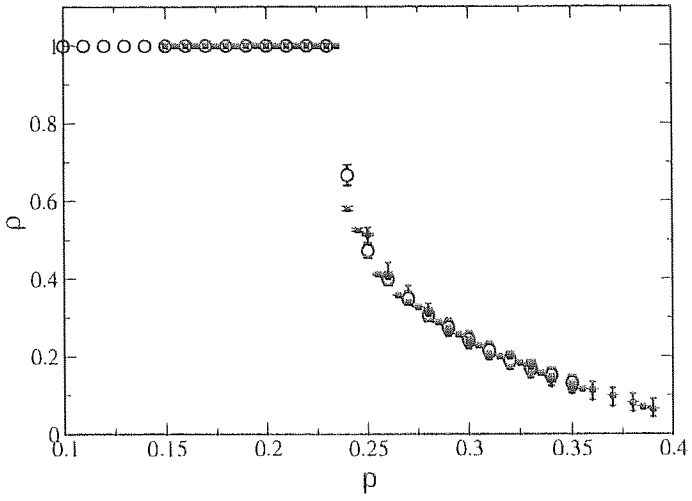


Figure 6.4: Monte-Carlo integration of field distributions and simulations for a KS code with two signal matrices ($K_1 = 1$ and $K_2 = 3$) , $\alpha = 5$ ($R = 1/5$) and $\alpha_1 = 3$. Circles: full statistics (4000 bins). Squares: simulations $N = 5000$

For **1n** we have:

$$R_1(y) = \int d\widehat{y}\,\widehat{R}_1(\widehat{y})\,\langle\delta\left[y-\widehat{y}-\zeta F_n\right]\rangle_\zeta \tag{6.21}$$

$$\widehat{R}_1(\widehat{y}) = \int \prod_{j=1}^{K_1} dx_j P_1(x_j) dy R_1(y)\;\delta\left[\widehat{x}-\text{atanh}(\tanh(y)\prod_{j=1}^{K_1}\tanh(x_j))\right]$$

Finally, for sub-matrix **2n**:

$$R_2(y) = \langle\delta\left[y-\zeta F_n\right]\rangle_\zeta \tag{6.22}$$

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

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

$$\rho = \int dh\; P(h)\;\text{sgn}(h) \tag{6.23}$$

$$P(h) = \int \prod_{j=1}^{\alpha_1 K_1} d\widehat{x}_j \widehat{P}_1(\widehat{x}_j) \prod_{l=1}^{\alpha_1 K_2} d\widehat{w}_l \widehat{P}_2(\widehat{w}_l)\;\delta\left[h-\sum_{j=1}^{\alpha_1 K_1} x_j - \sum_{l=1}^{\alpha_2 K_2} w_l\right] \tag{6.24}$$

The numerical solution of these equations provides the typical overlap for KS codes with two signal matrices parametrised by α_1 ($\alpha_2 = \alpha - \alpha 1$). In Fig 6.4 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.

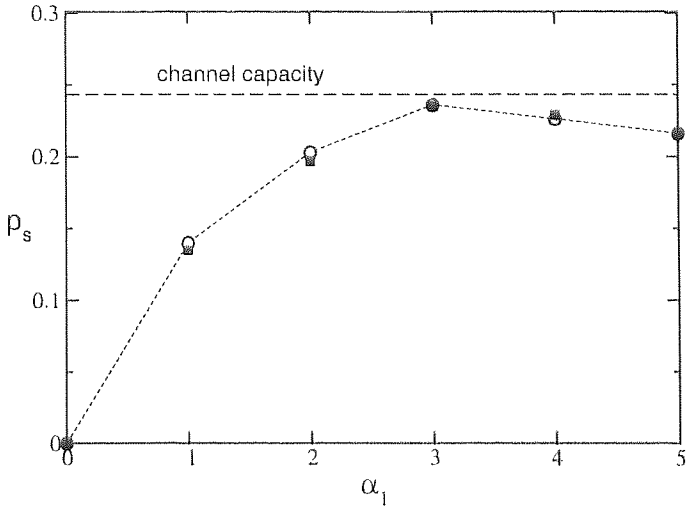


Figure 6.5: Spinodal noise level p_s as a function of α_1 for a KS code with $K_1 = 1$, $K_2 = 3$ and $R = 1/5$ ($\alpha = 5$). Circles: Monte-Carlo integrations of saddle-point equations (4000 bins). Squares: PP decoding simulations (10 runs with size $N = 5000$). The best performance is reached for $\alpha_1 = 3$ and is close to the channel capacity for a BSC (indicated by a dashed line).

6.4 Optimising construction parameters

Equations (6.19) to (6.24) can be used to optimise code constructions within a given family. For the family introduced in Fig. 6.3 with parameters K_1 and K_2 fixed the optimisation reduces to finding the value of α_1 that produces the highest threshold p_s. In Fig. 6.4 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.

6.5 Conclusion

We introduced a high performance construction of MN codes named KS codes. A macroscopic description for the typical PP decoding was obtained by writing saddle-point like equations for effective fields. Numerical simulations were shown to agree with the theory. We then used the macroscopic description in terms of effective fields distributions to optimise a single parameter family of KS codes. The optimised construction was shown to attain close to channel capacity performance in a BSC.

Chapter 7

Conclusions and Perspectives

7.1 Overview

In this thesis we analysed 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 summarised in the following table.

Coding Theory	*Statistical Physics*
message bits s	spins S
received bits r	multi-spin disordered couplings J (Sourlas)
syndrome bits z	multi-spin couplings $\mathcal{J}$ (Gallager, MN, KS)
bit error probability p_e	gauged magnetisation 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 $K = L = 2$ 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 $K = L = 2$ where the transition is bellow the upper bound.

A careful analysis of replica symmetric quantities reveals unphysical behaviour 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 of the replica symmetric theory, threshold noise values observed in simulations of probability propagation (PP) decoding agree with the noise level where meta-stable 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 justifies 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 take time that grows exponentially with the system size. In practical time scales, simple decoding procedures that simply find minima become trapped in meta-stable states. That is the reason why practical thresholds are linked to the appearance of spinodal points in the Bethe free-energy.

For KS 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 optimise parameters of one simple family of KS codes.

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

	Channel capacity	*Practical decoding of unbiased messages*
Sourlas	$K \to \infty$	$K = 2$
Gallager	$K \to \infty$	any K
MacKay-Neal	$K, L > 2$	$K = 1$, any L or $K = L = 2$
Kanter-Saad	still unclear	$K_j = 1, 2$ for some j

7.2 Some future directions

A tentative list of possible future research directions and open problems suggested by the research presented in this thesis is the following:

- **Optimisation of irregular constructions**: Is it possible to use statistical physics to optimise general irregular constructions from first principles ?

- **KS codes and Shannon's bound**: Are KS codes capable of attaining Shannon's bound in some region of the parameter space ?

- **Enhanced decoding algorithms**: Probability propagation decoding is based on a mean-field approximation that assumes a tree-like lattice; is it possible to design better decoding algorithms by using methods of statistical physics (*e.g.* Kikuchi approximations or variational methods) ?

- **Calculating the typical probability of block error**: The framework presented in this thesis used the overlap $\rho = \frac{1}{N}\sum_{j=1}^{N} \xi_j \widehat{\xi}_j$ between message and estimate as an order parameter. This is equivalent to calculate the typical probability of bit error as a performance measure. The same calculation using the typical probability of block error as a performance measure is still an open problem.

- **Replica symmetry breaking**: The role of replica symmetry breaking is still unclear in the low noise region. What are the practical implications of breaking the replica symmetry ?

- **Phase diagram outside Nishimori's condition**: Most of the calculations in this thesis were performed using Nishimori's condition, that is equivalent to consider that real and assumed parameters (temperature, noise level, message bias) match. It would be interesting to determine the complete phase diagram by analysing the situation where real and assumed parameters do not match.

- **Replica symmetry breaking and Nishimori's condition**: It has been recently shown that replica symmetry breaking is not expected to occur over the Nishimori line in the SK model [NS00, Nis00]. However it is still unclear the same arguments can be extended to the somewhat pathological models describing Gallager and MN codes. If the argument can be extended, how to explain the negative entropies emerging in the low noise level region ?

- **Concatenated $K = L = 2$ codes**: We have demosntrated that MN codes with $K = L = 2$ are practical but have performance far below Shannon's bound. Is that possible to produce practical codes with near Shannon's bound performance by combining two $K = L = 2$ codes in a concatenated code as internal and external codes ?

Appendix A

Convolutional and Turbo Codes

In this appendix we briefly present convolutional and Turbo codes.

A.1 Convolutional codes

Convolutional codes operate by storing K message bits at a time in a set of shift registers $D^0, D^1, ..., D^K$ (K is known as the *range*) and a number of bits is produced from these K bits. The shift registers are initialised with $D^j = 0$. At each time step the content of the registers is updated as $D^{j+1}(t+1) = D^j(t)$ for $j = 1, ..., K-1$, the next message bit is then stored in D^0. A simple convolutional code with rate $R = 1/2$ is represented in Figure A.1.

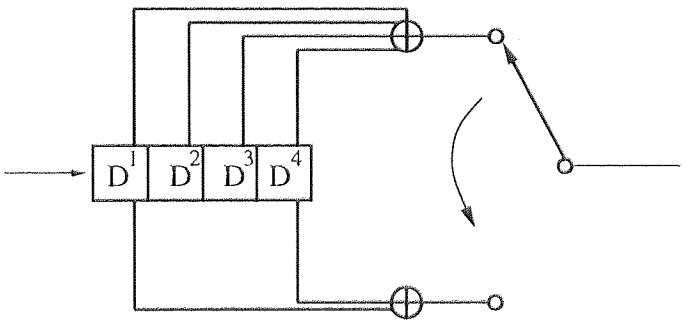


Figure A.1: Schematic representation of a convolutional code with rate $R = 1/2$. At each step four message bits are stored in the shift register and two bits are produced, the registers are then shifted to the right and a new message bit is stored in D^3.

It is interesting to observe that a convolutional code can be represented as linear code. For the code represented in Figure A.1 the generator matrix G has the following form:

$$G = \begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & \dots \\ 0 & 1 & 1 & 1 & 1 & 0 & 0 & \dots \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 & \dots \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & \dots \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \dots \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & \dots \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & \dots \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & \dots \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & \dots \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \dots \end{pmatrix}. \tag{A.1}$$

The decoding can proceed as described in this thesis.

A.2 Turbo codes

Turbo codes [BGT93] are made of two convolutional codes operating in parallel, the first receiving message bits and the second receiving a random permutation of the message produced by the *interleaver*. In Figure A.2 a Turbo code with rate $R = 1/3$ is represented.

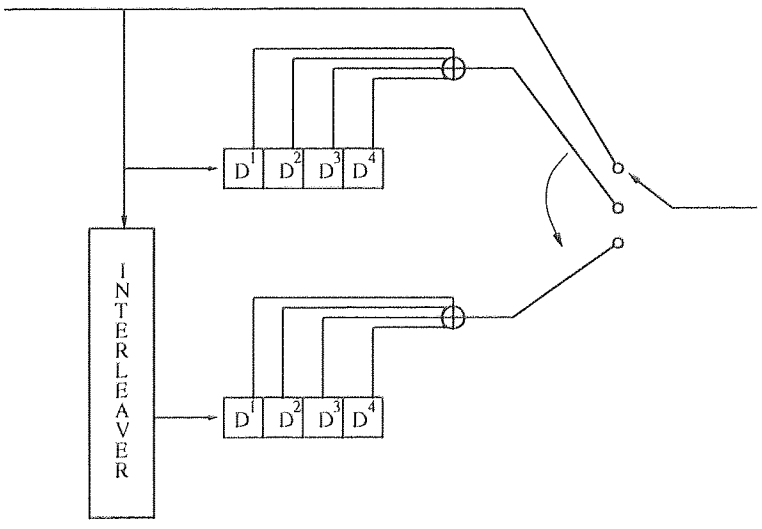


Figure A.2: Schematic representation of a Turbo code with rate $R = 1/3$. The interleaver produces a random permutation of the original block of message bits. the bits are then processed for each convolutional code.

Turbo codes can also be represented in the form of a generator matrix G, the code in Figure A.2

has the following representation:

$$G = \begin{pmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & \dots \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & \dots \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & \dots \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & \dots \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & \dots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \dots \\
1 & 1 & 1 & 1 & 0 & 0 & 0 & \dots \\
0 & 1 & 1 & 1 & 1 & 0 & 0 & \dots \\
0 & 0 & 1 & 1 & 1 & 1 & 0 & \dots \\
0 & 0 & 0 & 1 & 1 & 1 & 1 & \dots \\
0 & 0 & 0 & 0 & 1 & 1 & 1 & \dots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \dots \\
\Pi_{11} & \Pi_{12} & \Pi_{13} & \Pi_{14} & \Pi_{15} & \Pi_{16} & \Pi_{17} & \dots \\
\Pi_{21} & \Pi_{22} & \Pi_{23} & \Pi_{24} & \Pi_{25} & \Pi_{26} & \Pi_{27} & \dots \\
\Pi_{31} & \Pi_{32} & \Pi_{33} & \Pi_{34} & \Pi_{35} & \Pi_{36} & \Pi_{37} & \dots \\
\Pi_{41} & \Pi_{42} & \Pi_{43} & \Pi_{44} & \Pi_{45} & \Pi_{46} & \Pi_{47} & \dots \\
\Pi_{51} & \Pi_{52} & \Pi_{53} & \Pi_{54} & \Pi_{55} & \Pi_{56} & \Pi_{57} & \dots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \dots
\end{pmatrix}, \tag{A.2}$$

where Π is a submatrix obtained by permutation of columns of the submatrix in the middle.

Appendix B

Sourlas Codes: Technical Details

B.1 Free-energy

In order to compute free-energies we need to calculate the replicated partition function (3.27). We can start from Equation (3.25):

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

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 $\mu \equiv \langle i_1, ..., i_K \rangle$ for a specific set of indices:

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

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

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

Remembering that $\mathcal{A} \in \{0, 1\}$, and using the expression (3.5) for the Hamiltonian we can change the order of the summation and the product above and sum over $\mathcal{A}$:

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

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

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

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

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

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 $q_{\alpha_1 \ldots \alpha_m} = \frac{1}{N} \sum_i Z_i S_i^{\alpha_1} \ldots S_i^{\alpha_m}$ we find:

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

The normalisation constant is given by:

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

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

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

Computing the integrals over Z_i's and using Laplace method to compute the integrals over q_0 and $\widehat{q}_0$ we obtain:

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

The extremum point is given by

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

and

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

Replacing the auxiliary variables in Eq.(B.7) using $q_{\alpha_1\ldots\alpha_m}/q_0 \to q_{\alpha_1\ldots\alpha_m}$ and $\widehat{q}_{\alpha_1\ldots\alpha_m}/q_0 \to \widehat{q}_{\alpha_1\ldots\alpha_m}$, computing the integrals over Z_i and using Laplace method to evaluate the integrals we finally find Eq. (3.27).

B.2 Replica symmetric solution

The replica symmetric free-energy (3.31) can be obtained by plugging the ansatz (3.30) into Eq. (B.7).Using Laplace method we obtain:

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

where:

$$\begin{aligned}\mathcal{G}_1 &= \mathcal{T}_0 + \mathcal{T}_1\sum_\alpha\int\prod_j^K\left(dx_j\,\pi(x_j)\tanh(\beta x_j)\right)\\ &+\mathcal{T}_2\sum_{\langle\alpha_1\alpha_2\rangle}\int\prod_j^K\left(dx_j\,\pi(x_j)\tanh^2(\beta x_j)\right)+\ldots,\end{aligned} \tag{B.12}$$

$$\begin{aligned}\mathcal{G}_2 &= 1+\sum_\alpha\int dx\,dy\,\pi(x)\,\widehat{\pi}(y)\tanh(\beta x)\tanh(\beta y)\\ &+\sum_{\langle\alpha_1\alpha_2\rangle}\int dx\,dy\,\pi(x)\,\widehat{\pi}(y)\tanh^2(\beta x)\tanh^2(\beta y)+\ldots\end{aligned} \tag{B.13}$$

and

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

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:

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

Following exactly the same steps we obtain:

$$\mathcal{G}_2 = \int dx \, dy \pi(x) \, \widehat{\pi}(y) \, (1 + \tanh(\beta x) \tanh(\beta y))^n , \tag{B.16}$$

and

$$\mathcal{G}_3 = \ln \left\{ \mathrm{Tr}_{\{S^\alpha\}} \left[\left\langle \exp \left(\beta F \xi \sum_\alpha S^\alpha \right) \right\rangle_\xi \right.\right.$$
$$\left.\left. \times \oint \frac{dZ}{2\pi i} \frac{1}{Z^{C+1}} \exp \left(\widehat{q}_0 \, Z \int dy \, \widehat{\pi}(y) \prod_{\alpha=1}^{n} (1 + S^\alpha \tanh(\beta y)) \right) \right] \right\} . \tag{B.17}$$

Computing the integral over Z_i and the trace we finally find:

$$\mathcal{G}_3 = \ln \left\{ \frac{\widehat{q}_0^C}{C!} \int \prod_{l=1}^{C} dy_l \widehat{\pi}(y_l) \left[\sum_{\sigma=\pm 1} \langle e^{\sigma \beta F \xi} \rangle_\xi \prod_{l=1}^{C} (1 + \sigma \tanh(\beta y_l)) \right]^n \right\} . \tag{B.18}$$

Putting everything together, using Eq. (3.24) and some simple manipulation we find Eq. (3.31).

B.3 Local field distribution

In this appendix we derive explicitly Eq. (3.33). The gauge transformed overlap can be written as

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

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 p-th moment of m_i

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

where $\mathcal{H}^{(\alpha)}$ denotes the gauged Hamiltonian of the α-th replica. By performing the same steps described in the Appendices B.1 and B.2, introducing the auxiliary functions $\pi(x)$ and $\widehat{\pi}(y)$ defined in Eqs. (3.30), one obtains

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

Employing the identity

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

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

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

where we introduced the local fields distribution

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

thus reproducing Eq. (3.33).

B.4 Zero temperature self-consistent equations

In this appendix we describe how one can write a set of self-consistent equations to solve the zero temperature saddle-point equations (3.50). Supposing a three peaks ansatz given by:

$$\hat{\pi}(y) = p_+ \delta(y-1) + p_0 \delta(y) + p_- \delta(y+1) \tag{B.25}$$

$$\pi(x) = \sum_{l=1-C}^{C-1} T_{[p_\pm, p_0; C-1]}(l)\ \delta(x-l), \tag{B.26}$$

with

$$T_{[p_+,p_0,p_-;C]}(l) = \sum_{\{k,h,m\ ;\ k-h=l\ ;\ k+h+m=C-1\}} \frac{(C-1)!}{k!\ h!\ m!}\ p_+^k\ p_0^h\ p_-^m. \tag{B.27}$$

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 equations (3.50):

$$p_+ = \int \left[\prod_{l=1}^{K-1} dx_l\ \pi(x_l)\right] \left\langle \delta \left[1 - \mathrm{sign}(J \prod_{l=1}^{K-1} x_l)\ \min(\mid J \mid, \mid x_1 \mid, \ldots, \mid x_{K-1} \mid)\right] \right\rangle_J \tag{B.28}$$

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 :

$$p_+ = (1-p)\sum_{j=0}^{\lfloor\frac{K-1}{2}\rfloor}\frac{(K-1)!}{2j!(K-1-2j)!}\psi_-^{2j}\psi_+^{K-2j-1} \tag{B.29}$$
$$+ \; p\sum_{j=0}^{\lfloor\frac{K-1}{2}\rfloor-1}\frac{(K-1)!}{(2j+1)!(K-2-2j)!}\psi_-^{2j+1}\psi_+^{K-2j-2} + p\,\psi_-^{K-1}\;\mathrm{odd}(K-1)$$

$$p_- = (1-p)\sum_{j=0}^{\lfloor\frac{K-1}{2}\rfloor-1}\frac{(K-1)!}{(K-2j-2)!(2j+1)!}\psi_-^{2j+1}\psi_+^{K-2j-2} \tag{B.30}$$
$$+ \; p\sum_{j=0}^{\lfloor\frac{K-1}{2}\rfloor-1}\frac{(K-1)!}{(K-2j-1)!2j!}\psi_-^{2j}\psi_+^{K-2j-1} + (1-p)\,\psi_-^{K-1}\;\mathrm{odd}(K-1),$$

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 ansatz 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.

B.5 $\langle J\rangle_J = \langle J\,\tanh(\beta_N J)\rangle_J$

In this appendix we establish the identity $\langle J\rangle_J = \langle J\,\tanh(\beta_N J)\rangle_J$ for symmetric channels. It was shown in [Sou94a] that :

$$\beta_N\, J = \frac{1}{2}\,\ln\left(\frac{p(J\mid 1)}{p(J\mid -1)}\right), \tag{B.31}$$

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:

$$\tanh\,(\beta_N\, J) = \frac{p(J\mid 1) - p(J\mid -1)}{p(J\mid 1) + p(J\mid -1)}. \tag{B.32}$$

In a symmetric channel ($p(J\mid -J^0) = p(-J\mid J^0)$), it is also represented as

$$\tanh\,(\beta_N\, J) = \frac{p(J\mid 1) - p(-J\mid 1)}{p(J\mid 1) + p(-J\mid 1)}. \tag{B.33}$$

Therefore,

$$\begin{aligned}
\langle J\tanh\,(\beta_N\, J)\rangle_J &= \mathrm{Tr}_J\; p(J\mid 1)\;\frac{J\,p(J\mid 1)}{p(J\mid 1)+p(-J\mid 1)}\\
&+ \mathrm{Tr}_J\; p(J\mid 1)\;\frac{(-J)\,p(-J\mid 1)}{p(J\mid 1)+p(-J\mid 1)}\\
&= \mathrm{Tr}_J\; p(J\mid 1)\;\frac{J\,p(J\mid 1)}{p(J\mid 1)+p(-J\mid 1)}\\
&+ \mathrm{Tr}_J\; p(-J\mid 1)\;\frac{J\,p(J\mid 1)}{p(-J\mid 1)+p(J\mid 1)}\\
&= \mathrm{Tr}_J\; J\,p(J\mid 1)\\
&= \langle J\rangle_J.
\end{aligned} \tag{B.34}$$

B.6 Probability propagation equations

In this section we derive the probability propagation equations (2.40) and (2.38) in the form (3.71). We start by introducing the following representation for the variables $Q_{\mu k}^{S_k}$ and $R_{\mu k}^{S_k}$:

$$Q_{\mu k}^{S_k} = \frac{1}{2}\left(1 + m_{\mu k} S_k\right) \qquad R_{\mu k}^{S_k} = \frac{1}{2}\left(1 + \widehat{m}_{\mu k} S_k\right). \tag{B.35}$$

We can now put (3.66), (3.70) and (B.35) together to write:

$$\begin{aligned}
R_{\mu j}^{S_k} &= \frac{1}{a_\mu} \sum_{\{S_k : k \in \mathcal{L}(\mu)\backslash j\}} \frac{1}{2}\cosh(\beta J_\mu) \left(1 + \tanh(\beta J_\mu) \prod_{j \in \mathcal{L}_\mu} S_j\right) \prod_{k \in \mathcal{L}(\mu)\backslash j} \frac{1}{2}\left(1 + m_{\mu k} S_k\right), \\
&= \frac{1}{2^K}\frac{1}{a_\mu} \sum_{\{S_k : k \in \mathcal{L}(\mu)\backslash j\}} \cosh(\beta J_\mu) \left(1 + \tanh(\beta J_\mu) \prod_{j \in \mathcal{L}(\mu)} S_j\right) \\
&\times \left(1 + \sum_{k \in \mathcal{L}(\mu)\backslash j} m_{\mu k} S_k + \sum_{k \neq l \in \mathcal{L}(\mu)\backslash j} m_{\mu k} m_{\mu l} S_k S_l + \cdots\right) \\
&= \frac{1}{2^K}\frac{1}{a_\mu} \cosh(\beta J_\mu) \left(1 + \tanh(\beta J_\mu)\, S_j \prod_{k \in \mathcal{L}(\mu)\backslash j} m_{\mu k}\right) \\
&= \frac{1}{2}\left(1 + \tanh(\beta J_\mu)\, S_j \prod_{k \in \mathcal{L}(\mu)\backslash j} m_{\mu k}\right).
\end{aligned} \tag{B.36}$$

To obtain the last line we used that the normalisation constant is $a_\mu = \frac{1}{2^{K-1}}\cosh(\beta J_\mu)$. Writing the above equation in terms of the new variable $\widehat{m}_{\mu k}$ we obtain the first equation (3.71):

$$\begin{aligned}
\widehat{m}_{\mu k} &= R_{\mu k}^{(+)} - R_{\mu k}^{(-)} \\
&= \frac{1}{2}\left(1 + \tanh(\beta J_\mu) \prod_{k \in \mathcal{L}(\mu)\backslash j} m_{\mu k}\right) - \frac{1}{2}\left(1 - \tanh(\beta J_\mu) \prod_{k \in \mathcal{L}(\mu)\backslash j} m_{\mu k}\right) \\
&= \tanh(\beta J_\mu) \prod_{k \in \mathcal{L}(\mu)\backslash j} m_{\mu k}.
\end{aligned} \tag{B.37}$$

To obtain the second equation (3.71), we write:

$$Q_{\mu k}^{S_k} = a_{\mu k} \frac{1}{2}\left(1 + \tanh(\beta'_N S_k)\right) \prod_{\nu \in \mathcal{M}(k)\backslash \mu} \frac{1}{2}\left(1 + \widehat{m}_{\nu k} S_k\right). \tag{B.38}$$

In terms of the new variables $m_{\mu k}$ we have:

$$m_{\mu k} = a_{\mu k} \frac{1}{2^K}\left\{\left(1 + \tanh(\beta'_N)\right) \prod_{\nu \in \mathcal{M}(k)\backslash \mu} \left(1 + \widehat{m}_{\nu k}\right) - \left(1 - \tanh(\beta'_N)\right) \prod_{\nu \in \mathcal{M}(k)\backslash \mu} \left(1 - \widehat{m}_{\nu k}\right)\right\}$$

By using the identity $e^{\sigma x} = \cosh(x)(1 + \sigma \tanh(x))$ we can write:

$$m_{\mu k} = \frac{\exp\left[\sum_{\nu \in \mathcal{M}(k)\backslash \mu} \operatorname{atanh}(m_{\nu k}) + \beta'_N\right] - \exp\left[-\sum_{\nu \in \mathcal{M}(k)\backslash \mu} \operatorname{atanh}(m_{\nu k}) - \beta'_N\right]}{a_{\mu k}^{-1} 2^K \cosh(\beta'_N) \prod_{\nu \in \mathcal{M}(k)\backslash \mu} \cosh(\operatorname{atanh}(m_{\nu k}))} \tag{B.39}$$

Computing the normalisation $a_{\mu j}$ along the same lines gives:

$$a_{\mu k}^{-1} = \frac{\exp\left[\sum_{\nu \in \mathcal{M}(k)\backslash \mu} \operatorname{atanh}(m_{\nu k}) + \beta'_N\right] + \exp\left[-\sum_{\nu \in \mathcal{M}(k)\backslash \mu} \operatorname{atanh}(m_{\nu k}) - \beta'_N\right]}{2^K \cosh(\beta'_N) \prod_{\nu \in \mathcal{M}(k)\backslash \mu} \cosh(\operatorname{atanh}(m_{\nu k}))} \tag{B.40}$$

Inserting (B.40) into (B.39) gives:

$$m_{\mu k} = \tanh\left[\sum_{\nu\in\mathcal{M}(k)\setminus\mu} \operatorname{atanh}(m_{\nu k}) + \beta'_N\right]. \tag{B.41}$$

Appendix C

Gallager Codes: Technical Details

C.1 Replica theory

The replica theory for Gallager codes is very similar to the theory obtained for Sourlas codes (see Appendix B). We start with Eq. (4.17):

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

The average over constructions $\mathcal{A}$ is then introduced using Eq. (4.18) :

$$\begin{aligned} \langle \mathcal{Z}^n \rangle_{\mathcal{A},\zeta} &= \frac{1}{\mathcal{N}} \sum_{\tau^1,\cdots,\tau^n} \prod_{j=1}^{M} \left\langle \exp\left(F\zeta\beta \sum_{\alpha=1}^{n} \tau_j^\alpha \right) \right\rangle_\zeta \\ &\times \sum_{\{\mathcal{A}\}} \prod_{j=1}^{M} \left[\oint \frac{dZ_j}{2\pi i} \frac{1}{Z_j^{C+1}} Z_j^{\sum_{\langle i_1=j,i_2,\cdots,i_K \rangle} \mathcal{A}_{\langle i_1=j,\cdots,i_K \rangle}} \right] \\ &\times \prod_{\langle i_1 \cdots i_K \rangle} \exp\left[\beta\gamma \mathcal{A}_{\langle i_1 \cdots i_K \rangle} \sum_{\alpha=1}^{n} (\tau_{i_1}^\alpha \cdots \tau_{i_K}^\alpha - 1) \right] . \end{aligned} \tag{C.2}$$

After observing that

$$\prod_{j=1}^{M} Z_j^{\sum_{\langle i_1=j,i_2,\cdots,i_K \rangle} \mathcal{A}_{\langle i_1=j,\cdots,i_K \rangle}} = \prod_{\langle i_1 \cdots i_K \rangle} (Z_{i_1} \cdots Z_{i_K})^{\mathcal{A}_{\langle i_1 \cdots i_K \rangle}},$$

we can compute the summation over $\mathcal{A}_{\langle i_1 \cdots i_K \rangle} \in \{0,1\}$:

$$\begin{aligned} \langle \mathcal{Z}^n \rangle_{\mathcal{A},\zeta} &= \frac{1}{\mathcal{N}} \sum_{\tau^1,\cdots,\tau^n} \prod_{j=1}^{M} \left\langle \exp\left(F\zeta\beta \sum_{\alpha=1}^{n} \tau_j^\alpha \right) \right\rangle_\zeta \prod_{j=1}^{M} \left[\oint \frac{dZ_j}{2\pi i} \frac{1}{Z_j^{C+1}} \right] \\ &\times \prod_{\langle i_1 \cdots i_K \rangle} \left\{ 1 + \frac{Z_{i_1} \cdots Z_{i_K}}{e^{n\beta\gamma}} \prod_{\alpha=1}^{n} \exp\left[\beta\gamma (\tau_{i_1}^\alpha \cdots \tau_{i_K}^\alpha) \right] \right\} . \end{aligned} \tag{C.3}$$

We can now use the identity $e^{x\sigma} = \cosh(x)(1 + \sigma\tanh(x))$, where $\sigma = \pm 1$, to write:

$$\langle \mathcal{Z}^n \rangle_{\mathcal{A},\zeta} = \frac{1}{\mathcal{N}} \sum_{\tau^1,\cdots,\tau^n} \prod_{j=1}^{M} \left\langle \exp\left(F\zeta\beta \sum_{\alpha=1}^{n} \tau_j^\alpha \right) \right\rangle_\zeta \sum_{\{\mathcal{A}\}} \prod_{j=1}^{M} \left[\oint \frac{dZ_j}{2\pi i} \frac{1}{Z_j^{C+1}} \right] \tag{C.4}$$
$$\times \prod_{\langle i_1 \cdots i_K \rangle} \left\{ 1 + \frac{\cosh^n(\beta\gamma)}{e^{n\beta\gamma}} (Z_{i_1} \cdots Z_{i_K}) \prod_{\alpha=1}^{n} \left[1 + \tau_{i_1}^\alpha \cdots \tau_{i_K}^\alpha \tanh(\beta\gamma) \right] \right\}.$$

By following Appendix B.1 from Eq. (B.5) we can finally find Eq. (4.21).

C.2 Replica symmetric solution

As in the code of Sourlas (Appendix B.2) the replicated partition function can be put into the form:

$$\langle \mathcal{Z}^n \rangle_{\mathcal{A},\zeta} = \frac{1}{\mathcal{N}} \exp\left\{ M \, \mathrm{Extr}_{\pi,\widehat{\pi}} \left[\frac{C}{K} \mathcal{G}_1 - C\,\mathcal{G}_2 + \mathcal{G}_3 \right] \right\}. \tag{C.5}$$

Introducing the replica symmetric ansatz (4.22) into the functions $\mathcal{G}_1$, $\mathcal{G}_2$ and $\mathcal{G}_3$ we obtain:

$$\begin{aligned}
\mathcal{G}_1(n) &= \mathcal{T}_0 + \mathcal{T}_1 \sum_\alpha q_\alpha^K + \mathcal{T}_2 \sum_{\langle \alpha_1 \alpha 2 \rangle} q_{\alpha_1\alpha_2}^K + \cdots \qquad \text{(C.6)} \\
&= \frac{\cosh^n(\beta\gamma)}{e^{n\gamma\beta}} \int \prod_{j=1}^{K} dx_j \, \pi(x_j) \left[1 + \frac{n!}{(n-1)!} \tanh(\beta\gamma) \prod_{j=1}^{K} x_j \right. \\
&+ \left. \frac{n!}{(n-2)!2!} \tanh^2(\beta\gamma) \prod_{j=1}^{K} x_j^2 + \cdots \right] \\
&= \frac{\cosh^n(\beta\gamma)}{e^{n\gamma\beta}} \int \prod_{j=1}^{K} dx_j \, \pi(x_j) \left[1 + \tanh(\beta\gamma) \prod_{j=1}^{K} x_j \right]^n \\
&\overset{\gamma\to\infty}{\to} \frac{1}{2^n} \int \prod_{j=1}^{K} dx_j \, \pi(x_j) \left[1 + \prod_{j=1}^{K} x_j \right]^n,
\end{aligned}$$

where we use the Nishimori condition $\gamma \to \infty$, $\beta = 1$ to obtain the last line.

$$\begin{aligned}
\mathcal{G}_2(n) &= 1 + \sum_\alpha q_\alpha \widehat{q}_\alpha + \sum_{\langle \alpha_1 \alpha_2 \rangle} q_{\alpha_1\alpha_2} \widehat{q}_{\alpha_1\alpha_2} + \cdots \\
&= \int dx d\widehat{x} \, \pi(x) \widehat{\pi}(\widehat{x}) \left[1 + x\widehat{x} \right]^n. \qquad \text{(C.7)}
\end{aligned}$$

and

$$\begin{aligned}
\mathcal{G}_3(n) &= \frac{1}{M}\ln \mathrm{Tr}_{\{\tau^\alpha\}}\left[\left\langle \exp\left[F\beta\zeta\sum_{\alpha=1}^{n}\tau^\alpha\right]\right\rangle_\zeta\right. \\
&\times \left.\oint\frac{dZ}{2\pi i}\frac{\exp\left[Z\sum_{m=0}^{n}\sum_{\langle\alpha_1\cdots\alpha_m\rangle}\widehat{q}_{\alpha_1\cdots\alpha_m}\,\tau^{\alpha_1}\cdots\tau^{\alpha_m}\right]}{Z^{C+1}}\right] \\
&= \frac{1}{M}\ln \mathrm{Tr}_{\{\tau^\alpha\}}\left[\left\langle \exp\left[F\beta\zeta\sum_{\alpha=1}^{n}\tau^\alpha\right]\right\rangle_\zeta\right. \\
&\times \left.\oint\frac{dZ}{2\pi i}\frac{\exp\left[Z\int d\widehat{x}\widehat{\pi}(\widehat{x})\,\prod_{\alpha=1}^{n}(1+\tau^\alpha\widehat{x})\right]}{Z^{C+1}}\right] \\
&= \frac{1}{M}\ln\frac{\widehat{q}_0^C}{C!}\int\prod_{l=1}^{C}d\widehat{x}_l\,\widehat{\pi}(\widehat{x}_l)\left[\sum_{\tau=\pm1}\langle\exp\left[F\beta\zeta\tau\right]\rangle_\zeta\prod_{l=1}^{C}(1+\tau\widehat{x}_l)\right]^n
\end{aligned} \tag{C.8}$$

By using Eq. (4.16) we can write

$$f = -\frac{1}{\beta}\mathrm{Extr}_{\pi,\widehat{\pi}}\left[\frac{C}{K}\left.\frac{\partial}{\partial n}\right|_{n=0}\mathcal{G}_1(n) - C\left.\frac{\partial}{\partial n}\right|_{n=0}\mathcal{G}_2(n) + \left.\frac{\partial}{\partial n}\right|_{n=0}\mathcal{G}_3(n)\right], \tag{C.9}$$

what yields the free-energy (4.24).

C.3 Energy density at the Nishimori condition

In general the average internal energy is evaluated as:

$$\begin{aligned}
U &= \langle\langle\mathcal{H}(\gamma^*,F^*)\rangle_{\beta^*}\rangle_{\mathcal{J},\zeta} \\
&= \sum_{\mathcal{J}}\frac{\sum_\zeta P_{\gamma\beta}(\{\mathcal{J}_\mu\}\mid\zeta)\,P_{F\beta}(\zeta)}{\sum_{\tilde{\mathcal{J}},\tilde{\zeta}}P_{\gamma\beta}(\{\tilde{\mathcal{J}}_\mu\}\mid\tilde{\zeta})\,P_{F\beta}(\tilde{\zeta})}\,\frac{\sum_\tau\mathcal{H}(\gamma^*,F^*)\,P_{\gamma^*\beta^*}(\{\mathcal{J}_\mu\}\mid\tau)\,P_{F^*\beta^*}(\tau)}{\sum_{\tilde{\tau}}P_{\gamma^*\beta^*}(\{\mathcal{J}_\mu\}\mid\tilde{\tau})\,P_{F^*\beta^*}(\tilde{\tau})},
\end{aligned} \tag{C.10}$$

where the hyper-parameters 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 hyper-parameters of the Hamiltonian to the values in the encoding and corruption processes. If this is done, the expression for the energy can be rewritten:

$$U = \frac{\sum_{\mathcal{J},\tau}\mathcal{H}(\gamma,F)\,P_{\gamma\beta}(\{\mathcal{J}_\mu\}\mid\tau)\,P_{F\beta}(\tau)}{\sum_{\mathcal{J},\tau}P_{\gamma\beta}(\{\mathcal{J}_\mu\}\mid\tau)\,P_{F\beta}(\zeta)}. \tag{C.11}$$

By plugging (4.7) for the likelihood $P_{\gamma\beta}(\{\mathcal{J}_\mu\}\mid\tau)$ and for the prior $P_{F\beta}(\zeta)$; setting the hyperparameters to $\gamma\to\infty$, $\beta=1$ and $F=\mathrm{atanh}(1-2p)$ and performing the summation over $\mathcal{J}$ first, we easily get:

$$u = \lim_{M\to\infty}\frac{U}{M} = -F\,(1-2p). \tag{C.12}$$

Note that this expression is independent of the macroscopic state of the system.

C.4 Recursion relations

We start by introducing the effective field $\widehat{x}_{\nu j}$:

$$\tanh(\beta\widehat{x}_{\nu j}) = \frac{P_{\nu j}(+)\mathrm{e}^{-\beta F} - P_{\nu j}(-)\mathrm{e}^{+\beta F}}{P_{\nu j}(+)\mathrm{e}^{-\beta F} + P_{\nu j}(-)\mathrm{e}^{+\beta F}}. \tag{C.13}$$

Equation (4.30) can be easily obtained from the equation above. Equation(4.31) is then obtained by introducing Eq. (4.29) into Eq. (4.30) and performing a straightforward manipulation we obtain Eq. (4.32):

$$\exp(-2\beta\widehat{x}_{\mu k}) = \frac{\mathrm{Tr}_{\{\tau_j\}} \exp\left[\beta\gamma\left(-\mathcal{J}_\mu \prod_j^{''} \tau_j - 1\right)\right] \prod_\nu^{'} \prod_j^{''} \mathrm{e}^{\beta F \tau_j} \exp\left(\beta\widehat{x}_{\nu j}(\tau_j - 1)\right)}{\mathrm{Tr}_{\{\tau_j\}} \exp\left[\beta\gamma\left(+\mathcal{J}_\mu \tau_k \prod_j^{''} \tau_j - 1\right)\right] \prod_\nu^{'} \prod_j^{''} \mathrm{e}^{\beta F \tau_j} \exp\left(\beta\widehat{x}_{\nu j}(\tau_j - 1)\right)}, \tag{C.14}$$

where

$$\exp\left(\beta\widehat{x}_{\nu j}(\tau_j - 1)\right) = \frac{P_{\nu j}(\tau_j)\mathrm{e}^{-\beta F \tau_j}}{P_{\nu j}(+)\mathrm{e}^{-\beta F}}$$

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:

$$\exp(-2\beta\widehat{x}_{\mu k}) = \frac{\mathrm{Tr}_{\{\tau_j\}} \prod_j^{''} \exp\left[\left(\beta F + \sum_\nu^{'} \widehat{x}_{\nu j}\right)\tau_j\right] \left[\cosh(\beta\gamma)\left(1 - \mathcal{J}_\mu \prod_j^{''} \tau_j \tanh(\beta\gamma)\right)\right]}{\mathrm{Tr}_{\{\tau_j\}} \prod_j^{''} \exp\left[\left(\beta F + \sum_\nu^{'} \widehat{x}_{\nu j}\right)\tau_j\right] \left[\cosh(\beta\gamma)\left(1 + \mathcal{J}_\mu \prod_j^{''} \tau_j \tanh(\beta\gamma)\right)\right]}. \tag{C.15}$$

By introducing the Nishimori condition $\beta = 1$ and $\gamma \to \infty$ and computing traces:

$$\begin{aligned} \exp(-2\beta\widehat{x}_{\mu k}) &= \frac{\prod_{j\in\mathcal{L}(\mu)\setminus k} \sum_{\tau=\pm1} \mathrm{e}^{x_{\mu j}\tau} - \mathcal{J}_\mu \prod_{j\in\mathcal{L}(\mu)\setminus k} \sum_{\tau=\pm1} \tau \mathrm{e}^{x_{\mu j}\tau}}{\prod_{j\in\mathcal{L}(\mu)\setminus k} \sum_{\tau=\pm1} \mathrm{e}^{x_{\mu j}\tau} + \mathcal{J}_\mu \prod_{j\in\mathcal{L}(\mu)\setminus k} \sum_{\tau=\pm1} \tau \mathrm{e}^{x_{\mu j}\tau}} \\ &= \frac{1 - \mathcal{J}_\mu \prod_{j\in\mathcal{L}(\mu)\setminus k} \tanh(x_{\mu j})}{1 + \mathcal{J}_\mu \prod_{j\in\mathcal{L}(\mu)\setminus k} \tanh(x_{\mu j})}, \end{aligned} \tag{C.16}$$

where we have introduced

$$x_{\mu j} = F + \sum_{\nu\in\mathcal{M}(j)\setminus\mu} \widehat{x}_{\nu j}.$$

A brief manipulation of the equation above yields Eq. (4.32).

Appendix D

MN codes: technical details

D.1 Distribution of syndrome bits

In this section we evaluate probabilities p_z^x associated to syndrome bits in MN and Gallager codes.

In the case of Gallager codes a syndrome bit μ has the form

$$z_\mu = \zeta_{l_1} \oplus \cdots \oplus \zeta_{l_K}, \tag{D.1}$$

where $\zeta \in \{0,1\}$ and $\oplus$ denotes mod 2 sums. Each bit ζ_l 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:

$$\begin{aligned} p_z^0(K) &= \sum_{l \text{ even}}^{K} \frac{K!}{(K-l)!l!}\, p^l (1-p)^{K-l} \\ &= \sum_{l \text{ even}}^{K} (-1)^l \frac{K!}{(K-l)!l!}\, p^l (1-p)^{K-l}. \end{aligned} \tag{D.2}$$

Consequently

$$\begin{aligned} p_z^1(K) &= \sum_{l \text{ odd}}^{K} \frac{K!}{(K-l)!l!}\, p^l (1-p)^{K-l} \\ &= -\sum_{l \text{ odd}}^{K} (-1)^l \frac{K!}{(K-l)!l!}\, p^l (1-p)^{K-l}. \end{aligned} \tag{D.3}$$

From Equations (D.2) and (D.3) above we can write:

$$\begin{aligned} 1 - 2\, p_z^1(K) &= \sum_{l \text{ odd}}^{K} (-1)^l \frac{K!}{(K-l)!l!}\, p^l (1-p)^{K-l} \\ &= (1-p-p)^K = (1-2p)^K. \end{aligned} \tag{D.4}$$

From what we find:

$$p_z^1(K) = \frac{1}{2} - \frac{1}{2}(1-2p)^K. \tag{D.5}$$

For MN codes syndrome bits have the form:

$$z_\mu = \xi_{j_1} \oplus \cdots \oplus \xi_{j_K} \oplus \zeta_{l_1} \oplus \cdots \oplus \zeta_{l_L}, \tag{D.6}$$

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:

$$\begin{aligned} p_z^0(K,L) &= p_z^0(K)p_z^0(L) + p_z^1(K)p_z^1(L) \\ &= 1 - p_z^1(K) - p_z^1(L) + 2\, p_z^1(K)p_z^1(L). \end{aligned} \tag{D.7}$$

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 Equation (D.5) into Equation (D.7) we get:

$$\begin{aligned} p_z^1(K,L) &= 1 - p_z^0(K,L) \\ &= \frac{1}{2} - \frac{1}{2}(1-2p_\xi)^K\,(1-2p)^L. \end{aligned} \tag{D.8}$$

D.2 Replica theory

For MN codes the replicated partition function has the following form:

$$\begin{aligned} \langle \mathcal{Z}^n \rangle_{\mathcal{A},\xi,\zeta} &= \sum_{\boldsymbol{S}^1,\cdots,\boldsymbol{S}^n} \sum_{\boldsymbol{\tau}^1,\cdots,\boldsymbol{\tau}^n} \prod_{j=1}^{N} \left\langle \exp\left(F_s \xi \beta \sum_{\alpha=1}^{n} S_j^\alpha\right)\right\rangle_\xi \prod_{l=1}^{M} \left\langle \exp\left(F_n \zeta \beta \sum_{\alpha=1}^{n} \tau_l^\alpha\right)\right\rangle_\zeta \\ &\times \left\langle \prod_{\langle jl\rangle} \prod_{\alpha=1}^{n} \exp\left[\beta\gamma \mathcal{A}_{\langle jl\rangle}(S_{j_1}^\alpha \cdots S_{j_K}^\alpha \tau_{l_1}^\alpha \cdots \tau_{l_L}^\alpha - 1)\right]\right\rangle_{\mathcal{A}}. \end{aligned} \tag{D.9}$$

By introducing averages over constructions (4.18) as described in Appendix C.1 we find:

$$\begin{aligned} \langle \mathcal{Z}^n \rangle_{\mathcal{A},\xi,\zeta} &= \sum_{\boldsymbol{S}^1,\cdots,\boldsymbol{S}^n} \sum_{\boldsymbol{\tau}^1,\cdots,\boldsymbol{\tau}^n} \prod_{j=1}^{N} \left\langle \exp\left(F_s \xi \beta \sum_{\alpha=1}^{n} S_j^\alpha\right)\right\rangle_\xi \prod_{j=1}^{M} \left\langle \exp\left(F_n \zeta \beta \sum_{\alpha=1}^{n} \tau_j^\alpha\right)\right\rangle_\zeta \\ &\times \sum_{\{C_j,D_l\}} \prod_{j=1}^{N} \mathcal{P}_C(C_j) \prod_{l=1}^{M} \mathcal{P}_D(D_l) \\ &\times \frac{1}{\mathcal{N}} \sum_{\{\mathcal{A}\}} \prod_{j=1}^{N} \left[\oint \frac{dZ_j}{2\pi i} \frac{1}{Z_j^{C_j+1}} Z_j^{\sum_{\langle j_1=j,j_2,\cdots,j_K,l\rangle} \mathcal{A}_{\langle j_1=j,\cdots,j_K,l\rangle}}\right] \\ &\times \prod_{l=1}^{M} \left[\oint \frac{dY_l}{2\pi i} \frac{1}{Y_l^{D_l+1}} Y_l^{\sum_{\langle j,l_1=l,l_2,\cdots,l_L\rangle} \mathcal{A}_{\langle j,l_1=l,\cdots,l_L\rangle}}\right] \\ &\times \prod_{\langle jl\rangle} \exp\left[\beta\gamma \mathcal{A}_{\langle jl\rangle} \sum_{\alpha=1}^{n} (S_{j_1}^\alpha \cdots S_{j_K}^\alpha \tau_{l_1}^\alpha \cdots \tau_{l_L}^\alpha - 1)\right]. \end{aligned} \tag{D.10}$$

Computing the sum over $\mathcal{A}$ we get:

$$\begin{aligned}\langle \mathcal{Z}^n \rangle_{\mathcal{A},\xi,\zeta} &= \sum_{\boldsymbol{S}^1,\cdots,\boldsymbol{S}^n} \sum_{\boldsymbol{\tau}^1,\cdots,\boldsymbol{\tau}^n} \prod_{j=1}^{N} \left\langle \exp\left(F_s \xi \beta \sum_{\alpha=1}^{n} S_j^\alpha\right)\right\rangle_\xi \prod_{j=1}^{M} \left\langle \exp\left(F_n \zeta \beta \sum_{\alpha=1}^{n} \tau_j^\alpha\right)\right\rangle_\zeta \\ &\times \sum_{\{C_j,D_l\}} \prod_{j=1}^{N} \mathcal{P}_C(C_j) \prod_{l=1}^{M} \mathcal{P}_D(D_l) \frac{1}{\mathcal{N}} \oint \frac{dZ_j}{2\pi i} \frac{1}{Z_j^{C_j+1}} \oint \frac{dY_l}{2\pi i} \frac{1}{Y_l^{D_l+1}} \\ &\times \prod_{\langle \boldsymbol{il}\rangle} \left\{ 1 + \frac{Z_{i_1}\cdots Z_{i_K} Y_{l_1}\cdots Y_{l_L}}{\mathrm{e}^{n\beta\gamma}} \prod_{\alpha=1}^{n} \exp\left[\beta\gamma(S_{i_1}^\alpha \cdots S_{i_K}^\alpha \tau_{l_1}^\alpha \cdots \tau_{l_L}^\alpha)\right] \right\}. \end{aligned} \tag{D.11}$$

We use the identity $\mathrm{e}^{x\sigma} = \cosh(x)(1 + \sigma \tanh(x))$, where $\sigma = \pm 1$, to write:

$$\begin{aligned}\langle \mathcal{Z}^n \rangle_{\mathcal{A},\xi,\zeta} &= \sum_{\boldsymbol{S}^1,\cdots,\boldsymbol{S}^n} \sum_{\boldsymbol{\tau}^1,\cdots,\boldsymbol{\tau}^n} \prod_{j=1}^{N} \left\langle \exp\left(F_s \xi \beta \sum_{\alpha=1}^{n} S_j^\alpha\right)\right\rangle_\xi \prod_{j=1}^{M} \left\langle \exp\left(F_n \zeta \beta \sum_{\alpha=1}^{n} \tau_j^\alpha\right)\right\rangle_\zeta \\ &\times \sum_{\{C_j,D_l\}} \prod_{j=1}^{N} \mathcal{P}_C(C_j) \prod_{l=1}^{M} \mathcal{P}_D(D_l) \frac{1}{\mathcal{N}} \oint \frac{dZ_j}{2\pi i} \frac{1}{Z_j^{C_j+1}} \oint \frac{dY_l}{2\pi i} \frac{1}{Y_l^{D_l+1}} \\ &\times \prod_{\langle \boldsymbol{il}\rangle} \left\{ 1 + \frac{\cosh^n(\beta\gamma)}{\mathrm{e}^{n\beta\gamma}} (Z_{i_1}\cdots Z_{i_K} Y_{l_1}\cdots Y_{l_L}) \prod_{\alpha=1}^{n} \left[1 + S_{i_1}^\alpha \cdots S_{i_K}^\alpha \tau_{l_1}^\alpha \cdots \tau_{l_L}^\alpha \tanh(\beta\gamma)\right] \right\}. \end{aligned} \tag{D.12}$$

The product in the replica index α yields:

$$\begin{aligned}\prod_{\alpha=1}^{n} \left[1 + S_{i_1}^\alpha \cdots S_{i_K}^\alpha \tau_{l_1}^\alpha \cdots \tau_{l_L}^\alpha \tanh(\beta\gamma)\right] &= \sum_{m=0}^{n} \Bigg[\tanh^m(\beta\gamma) \\ &\times \sum_{\langle \alpha_1,\cdots,\alpha_m\rangle} S_{i_1}^{\alpha_1}\cdots S_{i_1}^{\alpha_m} \cdots S_{i_K}^{\alpha_1}\cdots S_{i_K}^{\alpha_m} \tau_{l_1}^{\alpha_1}\cdots\tau_{l_1}^{\alpha_m} \tau_{l_L}^{\alpha_1}\cdots\tau_{l_L}^{\alpha_m} \Bigg], \end{aligned} \tag{D.13}$$

where $\langle \alpha_1,\cdots,\alpha_m\rangle = \{\alpha_1,\cdots,\alpha_m : \alpha_1 < \cdots < \alpha_m\}$.

The product in the multi-indices $\langle \boldsymbol{il}\rangle$ can be computed by observing that the following relation holds in the thermodynamic limit:

$$\begin{aligned}\prod_{\langle \boldsymbol{il}\rangle} (1 + \psi_{\langle il\rangle}) &= \sum_{m=0}^{mmax} \sum_{\langle \langle il\rangle_1,\cdots,\langle il\rangle_m\rangle} \psi_{\langle il\rangle_1}\cdots\psi_{\langle il\rangle_m} \\ &\xrightarrow{N\to\infty} \exp\left[\sum_{\langle \boldsymbol{il}\rangle} \psi_{\langle il\rangle}\right], \end{aligned} \tag{D.14}$$

with $mmax \sim (N^K M^L)/K!L!$.

We find Eq. (5.26) by putting Eqs. (D.14) and (D.13) into (D.12) and using the following identities to introduce auxiliary variables:

$$\begin{aligned}\int dq_{\alpha_1\cdots\alpha_m}\, \delta\left[q_{\alpha_1\cdots\alpha_m} - \frac{1}{N}\sum_{j=1}^{N} Z_j S_j^{\alpha_1}\cdots S_j^{\alpha_m}\right] &= 1 \\ \int dr_{\alpha_1\cdots\alpha_m}\, \delta\left[r_{\alpha_1\cdots\alpha_m} - \frac{1}{M}\sum_{l=1}^{M} Y_l \tau_l^{\alpha_1}\cdots\tau_l^{\alpha_m}\right] &= 1 \end{aligned} \tag{D.15}$$

D.3 Replica symmetric free-energy

We first compute the normalisation $\mathcal{N}$ for a given :

$$\begin{aligned}\mathcal{N} &= \int\left(\frac{dq_0 d\widehat{q}_0}{2\pi i}\right)\int\left(\frac{dr_0 d\widehat{r}_0}{2\pi i}\right)\exp\left[\frac{M^L N^K}{K!L!}\mathcal{T}_0 q_0^K\, r_0^L \;-\; N q_0\widehat{q}_0 \;-\; M r_0\widehat{r}_0\right]\\ &\times \prod_{j=1}^{N}\oint\frac{dZ_j}{2\pi i}\frac{\exp\left[Z_j\widehat{q}_0\right]}{Z_j^{C_j+1}}\prod_{l=1}^{M}\oint\frac{dY_l}{2\pi i}\frac{\exp\left[Y_l\widehat{r}_0\right]}{Y_l^{D_l+1}}\end{aligned} \tag{D.16}$$

By using Cauchy's integrals to integrate in Z_j and Y_l and Laplace's method we get:

$$\begin{aligned}\mathcal{N} &= \exp\left\{\mathrm{Extr}_{q_0,\widehat{q}_0,r_0,\widehat{r}_0}\left[\frac{M^L N^K}{K!L!}\mathcal{T}_0 q_0^K\, r_0^L \;-\; N q_0\widehat{q}_0 \;-\; M r_0\widehat{r}_0\right.\right.\\ &+ \left.\left.\sum_{j=1}^{N}\ln\left(\frac{\widehat{q}_0^{C_j}}{C_j!}\right) + \sum_{l=1}^{M}\ln\left(\frac{\widehat{r}_0^{L_l}}{L_l!}\right)\right]\right\}.\end{aligned} \tag{D.17}$$

The extremisation above yields the following equations:

$$q_0\widehat{q}_0 = \frac{1}{N}\sum_{j=1}^{N}C_j = \overline{C} \tag{D.18}$$

$$r_0\widehat{r}_0 = \frac{1}{M}\sum_{l=1}^{M}L_l = \overline{L} \tag{D.19}$$

$$q_0^K r_0^L = \overline{C}\frac{(K-1)!L!}{N^{K-1}M^L}. \tag{D.20}$$

The variables can be normalised as:

$$\frac{q_{\alpha_1\cdots\alpha_m}}{q_0}\mapsto q_{\alpha_1\cdots\alpha_m} \qquad \frac{r_{\alpha_1\cdots\alpha_m}}{r_0}\mapsto r_{\alpha_1\cdots\alpha_m}. \tag{D.21}$$

By plugging Eqs. (D.17), (D.18), the above transformation into (5.26) and by using Laplace's method we obtain:

$$\begin{aligned}\langle\mathcal{Z}^n\rangle_{\mathcal{A},\xi,\zeta} &= \mathrm{Extr}_{q,r,\widehat{q},\widehat{r}}\left\{\exp\left[N\frac{\overline{C}}{K}\sum_{m=1}^{n}\sum_{\langle\alpha_1\cdots\alpha_m\rangle}\mathcal{T}_m q_{\alpha_1\cdots\alpha_m}^K\, r_{\alpha_1\cdots\alpha_m}^L\right.\right.\\ &- N\overline{C}\sum_{m=1}^{n}\sum_{\langle\alpha_1\cdots\alpha_m\rangle}q_{\alpha_1\cdots\alpha_m}\widehat{q}_{\alpha_1\cdots\alpha_m} - M\overline{L}\sum_{m=1}^{n}\sum_{\langle\alpha_1\cdots\alpha_m\rangle}r_{\alpha_1\cdots\alpha_m}\widehat{r}_{\alpha_1\cdots\alpha_m}\Bigg]\\ &\times \prod_{j=1}^{N}\sum_{C_j}\mathcal{P}_C(C_j)\prod_{l=1}^{M}\sum_{D_l}\mathcal{P}_D(D_l)\\ &\times \prod_{j=1}^{N}\left(\frac{C_j!}{\widehat{q}_0^{C_j}}\right)\mathrm{Tr}_{\{S_j^\alpha\}}\left[\left\langle\exp\left[F_s\beta\xi\sum_{\alpha=1}^{n}S^\alpha\right]\right\rangle_\xi\right.\\ &\times \left.\oint\frac{dZ_j}{2\pi i}\frac{\exp\left[Z_j\sum_{m=0}^{n}\sum_{\langle\alpha_1\cdots\alpha_m\rangle}\widehat{q}_{\alpha_1\cdots\alpha_m}\,S^{\alpha_1}\cdots S^{\alpha_m}\right]}{Z_j^{C_j+1}}\right]\\ &\times \prod_{l=1}^{M}\left(\frac{D_l!}{\widehat{r}_0^{D_l}}\right)\mathrm{Tr}_{\{\tau_l^\alpha\}}\left[\left\langle\exp\left[F_n\beta\zeta\sum_{\alpha=1}^{n}\tau_l^\alpha\right]\right\rangle_\zeta\right.\\ &\times \left.\left.\oint\frac{dY_l}{2\pi i}\frac{\exp\left[Y_l\sum_{m=0}^{n}\sum_{\langle\alpha_1\cdots\alpha_m\rangle}\widehat{r}_{\alpha_1\cdots\alpha_m}\,\tau^{\alpha_1}\cdots\tau^{\alpha_m}\right]}{Y_l^{D_l+1}}\right]\right\},\end{aligned} \tag{D.22}$$

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:

$$\langle \mathcal{Z}^n \rangle_{\mathcal{A},\xi,\zeta} = \exp\left\{ N\mathrm{Extr}_{q,r,\widehat{q},\widehat{r}} \left[\frac{\overline{C}}{K}\mathcal{G}_1 - \overline{C}\mathcal{G}_2 - \overline{L}\mathcal{G}_3 + \mathcal{G}_4 + \mathcal{G}_5 \right] \right\} \tag{D.23}$$

Introducing the replica symmetric ansätze:

$$q_{\alpha_1\cdots\alpha_m} = \int dx\ \pi(x)\ x^m \qquad \widehat{q}_{\alpha_1\cdots\alpha_m} = \int d\widehat{x}\ \widehat{\pi}(\widehat{x})\ \widehat{x}^m \tag{D.24}$$

and

$$r_{\alpha_1\cdots\alpha_m} = \int dy\ \phi(y)\ y^m \qquad \widehat{r}_{\alpha_1\cdots\alpha_m} = \int d\widehat{y}\ \widehat{\phi}(\widehat{y})\ \widehat{y}^m. \tag{D.25}$$

By introducing the Nishimori condition $\gamma \to \infty$ and $\beta = 1$, we can work each term on (D.23) out and find:

$$\begin{aligned}
\mathcal{G}_1(n) &= \mathcal{T}_0 + \mathcal{T}_1 \sum_\alpha q_\alpha^K r_\alpha^L + \mathcal{T}_2 \sum_{\langle\alpha_1\alpha 2\rangle} q_{\alpha_1\alpha_2}^K r_{\alpha_1\alpha_2}^L + \cdots \qquad \text{(D.26)}\\
&= \frac{\cosh^n(\beta\gamma)}{\mathrm{e}^{n\gamma\beta}} \int \prod_{j=1}^K dx_j \prod_{l=1}^L dy_l\ \phi(y_l) \left[1 + \frac{n!}{(n-1)!}\tanh(\beta\gamma) \prod_{j=1}^K x_j \prod_{l=1}^L y_l \right.\\
&+ \left. \frac{n!}{(n-2)!2!}\tanh^2(\beta\gamma) \prod_{j=1}^K x_j^2 \prod_{l=1}^L y_l^2 + \cdots \right]\\
&= \frac{\cosh^n(\beta\gamma)}{\mathrm{e}^{n\gamma\beta}} \int \prod_{j=1}^K dx_j\ \pi(x_j) \prod_{l=1}^L dy_l\ \phi(y_l) \left[1 + \tanh(\beta\gamma) \prod_{j=1}^K x_j \prod_{l=1}^L y_l \right]^n\\
&\overset{\gamma\to\infty}{\to} \frac{1}{2^n} \int \prod_{j=1}^K dx_j \pi(x_j) \prod_{l=1}^L dy_l \phi(y_l) \left[1 + \prod_{j=1}^K x_j \prod_{l=1}^L y_l \right]^n,
\end{aligned}$$

$$\begin{aligned}
\mathcal{G}_2(n) &= 1 + \sum_\alpha q_\alpha \widehat{q}_\alpha + \sum_{\langle\alpha_1\alpha_2\rangle} q_{\alpha_1\alpha_2}\widehat{q}_{\alpha_1\alpha_2} + \cdots\\
&= \int dx d\widehat{x}\ \pi(x)\widehat{\pi}(\widehat{x})\left[1 + x\widehat{x}\right]^n. \qquad \text{(D.27)}
\end{aligned}$$

Similarly,

$$\begin{aligned}
\mathcal{G}_3(n) &= 1 + \sum_\alpha r_\alpha \widehat{r}_\alpha + \sum_{\langle\alpha_1\alpha_2\rangle} r_{\alpha_1\alpha_2}\widehat{r}_{\alpha_1\alpha_2} + \cdots\\
&= \int dy d\widehat{y}\ \phi(y)\widehat{\phi}(\widehat{y})\left[1 + y\widehat{y}\right]^n. \qquad \text{(D.28)}
\end{aligned}$$

$$
\begin{aligned}
\mathcal{G}_4(n) &= \frac{1}{N}\sum_{j=1}^{N}\ln\sum_{C_j}\mathcal{P}_C(C_j)\left(\frac{C_j!}{\widehat{q}_0^{C_j}}\right)\mathrm{Tr}_{\{S^\alpha\}}\left[\left\langle\exp\left[F_s\beta\xi\sum_{\alpha=1}^{n}S_j^\alpha\right]\right\rangle_\xi\right.\\
&\times\left.\oint\frac{dZ_j}{2\pi i}\frac{\exp\left[Z_j\sum_{m=0}^{n}\sum_{\langle\alpha_1\cdots\alpha_m\rangle}\widehat{q}_{\alpha_1\cdots\alpha_m}\,S_j^{\alpha_1}\cdots S_j^{\alpha_m}\right]}{Z_j^{C_j+1}}\right]\\
&= \frac{1}{N}\sum_{j=1}^{N}\ln\sum_{C_j}\mathcal{P}_C(C_j)\left(\frac{C_j!}{\widehat{q}_0^{C_j}}\right)\mathrm{Tr}_{\{S_j^\alpha\}}\left[\left\langle\exp\left[F_s\beta\xi\sum_{\alpha=1}^{n}S_j^\alpha\right]\right\rangle_\xi\right.\\
&\times\left.\oint\frac{dZ_j}{2\pi i}\frac{\exp\left[Z_j\int d\widehat{x}\widehat{\pi}(\widehat{x})\,\prod_{\alpha=1}^{n}(1+S_j^\alpha\widehat{x})\right]}{Z_j^{C_j+1}}\right]\\
&= \ln\sum_{C_j}\mathcal{P}_C(C_j)\int\prod_{l=1}^{C_j}d\widehat{x}_l\,\widehat{\pi}(\widehat{x}_l)\left[\sum_{S=\pm1}\langle\exp\left[F_s\beta\xi S\right]\rangle_\xi\prod_{i=1}^{C_j}(1+S\widehat{x}_i)\right]^n
\end{aligned}
\tag{D.29}
$$

In the same way:

$$
\begin{aligned}
\mathcal{G}_5(n) &= \frac{1}{M}\sum_{l=1}^{M}\ln\sum_{D_l}\mathcal{P}_D(D_l)\left(\frac{D_l!}{\widehat{r}_0^{D_l}}\right)\mathrm{Tr}_{\{\tau^\alpha\}}\left[\left\langle\exp\left[F_n\beta\zeta\sum_{\alpha=1}^{n}\tau_l^\alpha\right]\right\rangle_\zeta\right.\\
&\times\left.\oint\frac{dY_l}{2\pi i}\frac{\exp\left[Y_l\sum_{m=0}^{n}\sum_{\langle\alpha_1\cdots\alpha_m\rangle}\widehat{r}_{\alpha_1\cdots\alpha_m}\,\tau_l^{\alpha_1}\cdots\tau_l^{\alpha_m}\right]}{Y_l^{D_l+1}}\right]\\
&= \frac{1}{M}\sum_{l=1}^{M}\ln\sum_{D_l}\mathcal{P}_D(D_l)\left(\frac{D_l!}{\widehat{r}_0^{D_l}}\right)\mathrm{Tr}_{\{D_l^\alpha\}}\left[\left\langle\exp\left[F_n\beta\zeta\sum_{\alpha=1}^{n}\tau_l^\alpha\right]\right\rangle_\zeta\right.\\
&\times\left.\oint\frac{dY_l}{2\pi i}\frac{\exp\left[Y_l\int d\widehat{y}\widehat{\phi}(\widehat{y})\,\prod_{\alpha=1}^{n}(1+\tau_l^\alpha\widehat{y})\right]}{Y_l^{D_l+1}}\right]\\
&= \ln\sum_{D_l}\mathcal{P}_C(D_l)\int\prod_{l=1}^{D_l}d\widehat{y}_l\,\widehat{\phi}(\widehat{y}_l)\left[\sum_{\tau=\pm1}\langle\exp\left[F_n\beta\zeta\tau\right]\rangle_\zeta\prod_{i=1}^{D_l}(1+\tau\widehat{y}_i)\right]^n
\end{aligned}
\tag{D.30}
$$

By using Eq. (5.19) we can write

$$
\begin{aligned}
f &= -\frac{1}{\beta}\mathrm{Extr}_{\pi,\widehat{\pi},\phi,\widehat{\phi}}\left[\frac{\overline{C}}{K}\left.\frac{\partial}{\partial n}\right|_{n=0}\mathcal{G}_1(n)\;-\;\overline{C}\left.\frac{\partial}{\partial n}\right|_{n=0}\mathcal{G}_2(n)\;-\;\overline{L}\left.\frac{\partial}{\partial n}\right|_{n=0}\mathcal{G}_3(n)\right.\\
&+\left.\left.\frac{\partial}{\partial n}\right|_{n=0}\mathcal{G}_4(n)\;+\;\left.\frac{\partial}{\partial n}\right|_{n=0}\mathcal{G}_5(n)\right],
\end{aligned}
\tag{D.31}
$$

what yields free-energy (5.29).

D.4 Viana-Bray model: Poisson constructions

The Viana-Bray (VB) model is a multi-spin system with random couplings and strong dilution [VB85]. 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. (5.13):

$$
\mathcal{H}_\gamma^{\text{gauge}}(\boldsymbol{S},\boldsymbol{\tau};\boldsymbol{\xi},\boldsymbol{\zeta}) = -\gamma\sum_{\langle jl\rangle}\mathcal{A}_{\langle jl\rangle}\left(S_{j_1}\cdots S_{j_K}\tau_{l_1}\cdots\tau_{l_L}-1\right)-F_s\sum_{j=1}^{N}\xi_j S_j-F_n\sum_{l=1}^{M}\zeta_l\tau_l. \tag{D.32}
$$

The variables $\mathcal{A}_{\langle jl \rangle}$ are independently drawn from the distribution:

$$P(\mathcal{A}) = \left(1 - \frac{L!K!}{M^{L-1}N^K}\right) \delta(\mathcal{A}) + \frac{L!K!}{M^{L-1}N^K} \delta(\mathcal{A} - 1). \tag{D.33}$$

The above distribution will yield the following averages:

$$\left\langle \sum_{\langle jl \rangle} \mathcal{A}_{\langle jl \rangle} \right\rangle_{\mathcal{A}} = M \tag{D.34}$$

$$\left\langle \sum_{\langle j_1 = j \cdots j_K l_1 \cdots l_L \rangle} \mathcal{A}_{\langle jl \rangle} \right\rangle_{\mathcal{A}} = C \tag{D.35}$$

$$\left\langle \sum_{\langle j_1 \cdots j_K l_1 = l \cdots l_L \rangle} \mathcal{A}_{\langle jl \rangle} \right\rangle_{\mathcal{A}} = L. \tag{D.36}$$

In the thermodynamic limit the above summations are random variabels with a Poisson distributions:

$$P\left\{ \sum_{\langle jl \rangle} \mathcal{A}_{\langle jl \rangle} = x \right\} = e^{-M} \frac{M^x}{x!} \tag{D.37}$$

$$P\left\{ \sum_{\langle j_1 = j \cdots j_K l_1 \cdots l_L \rangle} \mathcal{A}_{\langle jl \rangle} = x \right\} = e^{-\overline{C}} \frac{\overline{C}^x}{x!} \tag{D.38}$$

$$P\left\{ \sum_{\langle j_1 \cdots j_K l_1 = l \cdots l_L \rangle} \mathcal{A}_{\langle jl \rangle} = x \right\} = e^{-\overline{L}} \frac{\overline{L}^x}{x!}. \tag{D.39}$$

Since the variance of a Poisson distribution is given by the square root of the mean in the thermodynamic limit:

$$P\left\{ \sum_{\langle jl \rangle} \mathcal{A}_{\langle jl \rangle} = x \right\} \stackrel{M \to \infty}{\longrightarrow} \delta(x - M). \tag{D.40}$$

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:

$$\begin{aligned} \rho &\leq \frac{1}{N}\left[N - Ne^{-\overline{C}} + Ne^{-\overline{C}}(1 - 2p_\xi)\right] \\ &\leq 1 - e^{-\overline{C}} + e^{-\overline{C}}(1 - 2p_\xi) \\ &\leq 1 - 2\, p_\xi e^{-\overline{C}}. \end{aligned} \tag{D.41}$$

Therefore, a VB-like code has necessarily an error-floor that decays exponentially with the $\overline{C}$ chosen.

Appendix E

KS codes: cumulant expansion

In this appendix we propose to expand Eqs. (6.19) to (6.24) into cumulants and truncate the series, speeding up the optimisation of constructions and avoiding multi-dimensional Monte-Carlo integrations. To exemplify the procedure, the same family of codes with two signal sub-matrices presented in Section 6.3 is analysed and equations for first (delta approximation) and second order (Gaussian approximation) cumulant expansions are obtained.

A n-order cumulant $[x]_n$ of a probability distribution $P(x)$ is defined as:

$$[x]_n = \frac{\partial^n}{\partial \omega^n} \ln \mathcal{F}P(\omega = 0), \tag{E.1}$$

where $\mathcal{F}P(\omega) = \int dx\, e^{-ix\omega}\, P(x)$ is the Fourier transform of $P(x)$.

Equations (6.19) to (6.24) can be expanded in series of cumulants by observing the following relationships:

$$\frac{\partial^n}{\partial \omega^n} \ln \mathcal{F}P_j(\omega = 0) = (\alpha_j K_j - 1) \frac{\partial^n}{\partial \omega^n} \ln \mathcal{F}\widehat{P}_j(\omega = 0) + (\alpha_i K_i) \frac{\partial^n}{\partial \omega^n} \ln \mathcal{F}\widehat{P}_i(\omega = 0). \tag{E.2}$$

In the same way:

$$\frac{\partial^n}{\partial \omega^n} \ln \mathcal{F}R_1(\omega = 0) = \frac{\partial^n}{\partial \omega} \ln \mathcal{F}\widehat{R}_1(\omega = 0) + \frac{\partial^n}{\partial \omega^n} \ln \left\langle e^{-i\omega\zeta F_n} \right\rangle_\zeta |_{\omega=0} \tag{E.3}$$

$$\frac{\partial^n}{\partial \omega^n} \ln \mathcal{F}R_2(\omega = 0) = \frac{\partial^n}{\partial \omega^n} \ln \left\langle e^{-i\omega\zeta F_n} \right\rangle_\zeta |_{\omega=0} \tag{E.4}$$

For the conjugate distributions we have:

$$\frac{\partial^n}{\partial \omega^n} \ln \mathcal{F}\widehat{P}_1(\omega = 0) = \frac{\partial^n}{\partial \omega^n} \ln \left[\int \prod_{j=1}^{K_1 - 1} dx_j P_1(x_j) dy_1 R_1(y_1) dy_2 R_1(y_2) \times \right. \tag{E.5}$$
$$\left. e^{-i\omega[\text{atanh}(\tanh(y_1)\tanh(y_2)\prod_{j=1}^{K_1-1}\tanh(x_j))]} \right]_{\omega=0},$$

$$\frac{\partial^n}{\partial \omega^n} \ln \mathcal{F}\widehat{P}_2(\omega = 0) = \frac{\partial^n}{\partial \omega^n} \ln \left[\int \prod_{j=1}^{K_2 - 1} dx_j P_2(x_j) dy R_2(y)\, e^{-i\omega[\text{atanh}(\tanh(y)\prod_{j=1}^{K_2-1}\tanh(x_j))]} \right]_{\omega=0} \tag{E.6}$$

$$\frac{\partial^n}{\partial\omega^n}\ln\mathcal{F}\widehat{R}_1(\omega=0)=\frac{\partial^n}{\partial\omega^n}\ln\left[\int\prod_{j=1}^{K_1}dx_jP_1(x_j)dyR_1(y)\;e^{-i\omega[\text{atanh}(\tanh(y)\prod_{j=1}^{K_1}\tanh(x_j))]}\right]_{\omega=0}, \quad \text{(E.7)}$$

$$\frac{\partial^n}{\partial\omega^n}\ln\mathcal{F}\widehat{R}_2(\omega=0)=\frac{\partial^n}{\partial\omega^n}\ln\left[\int\prod_{j=1}^{K_1}dx_jP_2(x_j)\;e^{-i\omega[\text{atanh}(\prod_{j=1}^{K_2}\tanh(x_j))]}\right]_{\omega=0}. \quad \text{(E.8)}$$

The equations for the first order ($n=1$) cumulant expansion are:

$$x_1=(\alpha_1K_1-1)\widehat{x}_1+\alpha_2K_2\widehat{x}_2 \qquad x_2=(\alpha_1K_1)\widehat{x}_1+(\alpha_2K_2-1)\widehat{x}_2 \quad \text{(E.9)}$$

$$\widehat{x}_1=\text{atanh}\left[\tanh^2(y_1)\tanh^{K_1-1}(x_1)\right] \qquad \widehat{x}_2=\text{atanh}\left[\tanh(y_2)\tanh^{K_2-1}(x_2)\right] \quad \text{(E.10)}$$

$$y_1=\widehat{y}_1+(1-2p)F_n \qquad y_2=(1-2p)F_n \quad \text{(E.11)}$$

$$\widehat{y}_1=\text{atanh}\left[\tanh(y_1)\tanh^{K_1}(x_1)\right] \qquad \widehat{y}_2=\text{atanh}\left[\tanh^{K_2}(x_2)\right]. \quad \text{(E.12)}$$

The field generating the decoded message is given by:

$$h=\alpha_1K_1\widehat{x}_1+\alpha_2K_2\widehat{x}_2. \quad \text{(E.13)}$$

The first-order approximation is equivalent to assuming that the probability distributions in Eqs. (6.19) to (6.24) are delta distributions.

Eliminating the equations for the section **2n**, the second order (Gaussian, $n=2$) expansion gives:

$$\langle x_1\rangle = (\alpha_1K_1-1)\langle\widehat{x}_1\rangle+(\alpha_2K_2)\langle\widehat{x}_2\rangle \qquad \sigma^2_{x_1}=(\alpha_1K_1-1)\sigma^2_{\widehat{x}_1}+(\alpha_2K_2)\sigma^2_{\widehat{x}_2} \quad \text{(E.14)}$$

$$\langle x_2\rangle = (\alpha_1K_1)\langle\widehat{x}_1\rangle+(\alpha_2K_2-1)\langle\widehat{x}_2\rangle \qquad \sigma^2_{x_2}=(\alpha_1K_1-1)\sigma^2_{\widehat{x}_1}+(\alpha_2K_2-1)\sigma^2_{\widehat{x}_2} \quad \text{(E.15)}$$

$$\langle y\rangle = \langle\widehat{y}\rangle+(1-2p)F_n \qquad \sigma^2_y=\sigma^2_{\widehat{y}}+4f(1-p)F_n^2 \quad \text{(E.16)}$$

$$\langle\widehat{y}\rangle = \left\langle\text{atanh}\left[\tanh(y_1)\prod_{j=1,\ldots,K_1}\tanh(x_{1,j})\right]\right\rangle_{y_1,x_{1,j}} \quad \text{(E.17)}$$

$$\sigma^2_{\widehat{y}} = \left\langle\text{atanh}^2\left[\tanh(y_1)\prod_{j=1,\ldots,K_1}\tanh(x_{1,j})\right]\right\rangle_{y_1,x_{1,j}}-\langle\widehat{y}\rangle^2 \quad \text{(E.18)}$$

$$\langle\widehat{x}_1\rangle = \left\langle\text{atanh}\left[\tanh(y_{1,1})\tanh(y_{1,2})\prod_{j=1,\ldots,K_1-1}\tanh(x_{1,j})\right]\right\rangle_{y_{1,1},y_{1,2}x_{1,j}} \quad \text{(E.19)}$$

$$\sigma^2_{\widehat{x}_1} = \left\langle\text{atanh}^2\left[\tanh(y_{1,1})\tanh(y_{1,2})\prod_{j=1,\ldots,K_1-1}\tanh(x_{1,j})\right]\right\rangle_{y_{1,1},y_{1,2},x_{1,j}}-\langle\widehat{x}_1\rangle^2 \quad \text{(E.20)}$$

$$\langle\widehat{x}_2\rangle = (1-2p)\left\langle\text{atanh}\left[\tanh(F_n)\prod_{j=1,\ldots,K_2-1}\tanh(x_{2,j})\right]\right\rangle_{x_{2,j}} \quad \text{(E.21)}$$

$$\sigma^2_{\widehat{x}_2} = \left\langle\text{atanh}^2\left[\tanh(F_n)\prod_{j=1,\ldots,K_2-1}\tanh(x_{2,j})\right]\right\rangle_{x_{2,j}}-\langle\widehat{x}_2\rangle^2 \quad \text{(E.22)}$$

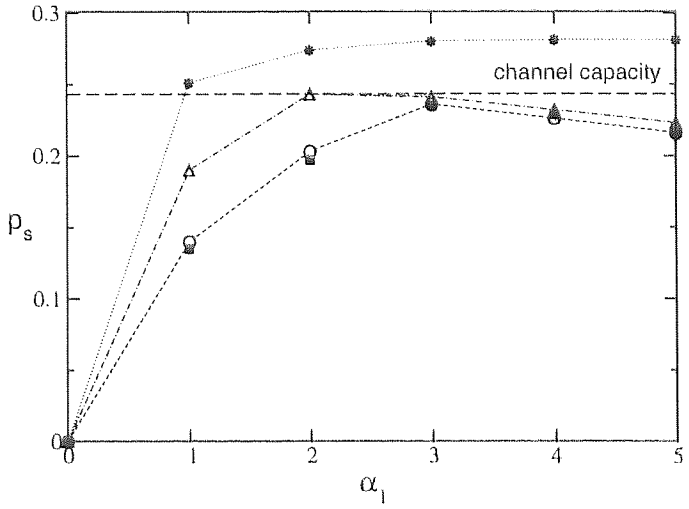


Figure E.1: Threshold p_s as a function of α_1 for a KS code with $K_1 = 1$, $K_2 = 3$ and $R = 1/5$ ($\alpha = 5$). Circles: Monte-Carlo integrations of saddle-point equations (4000 bins). Squares: PP decoding simulations (10 runs with size $N = 5000$). Triangles: Gaussian approximation. Stars: delta approximation.

The decoding field is then :

$$\langle h \rangle = (\alpha_1 K_1)\langle \widehat{x}_1 \rangle + (\alpha_2 K_2)\langle \widehat{x}_2 \rangle \tag{E.23}$$

$$\sigma_h^2 = (\alpha_1 K_1)\sigma_{\widehat{x}_1}^2 + (\alpha_2 K_2)\sigma_{\widehat{x}_2}^2 \tag{E.24}$$

In the above equations $\langle \cdots \rangle$ indicates Gaussian means and σ^2 variances. This approximation is, therefore, equivalent to assuming that distributions in Eqs. (6.19) to (6.24) are Gaussians.

In Fig. E we show thresholds p_s obtained by the delta and Gaussian approximations and compare with simulations and results obtained by Monte-Carlo integration of Eqs. (6.19) to (6.24). The approximations considerably overestimate the performance and generate poor optimisations ($\alpha_1 = 2$ for the Gaussian approximation and $\alpha_1 = 3, 4$ or 5 for the delta approximation).

Appendix F

Path lengths on random graphs

In this appendix we evaluate the probability distribution of lengths of paths starting in a randomly selected node.

A regular Gallager code of rate $R = N/M$ can be represented by a bipartite graph $G(M, M-N)$ with M nodes in the *signal class* and $M-N$ nodes in the *check class*. The degree of each signal node is C and the degree of each check node is K. A path is a subgraph with no cycles.

Lets organise the graph in generations such that in generation 0 (g_0) we have one randomly chosen signal nodes and the C check nodes connected to it. The following generations g_l are then recursively produced from this initial generation.

A subgraph with n generations is denoted by $\mathcal{G}^n$. If this subgraph $\mathcal{G}^n$ has a tree structure, to say $g_l \cap g_m = \emptyset$ for $0 \leq l, m \leq n$ $l \neq m$, one can denote $\mathcal{G}^n \in \mathcal{T}^n$. Suppose that one has $\mathcal{G}^n \in \mathcal{T}^n$. What is then the probability that the addition of the generation g_{n+1} produces a cycle of any size ?

The total number of signal nodes is M, the total number of check nodes is $M-N$. The number of signal nodes in $\mathcal{G}^n$ is

$$N_n = \sum_{j=0}^{n} (C-1)^j \, (K-1)^j. \tag{F.1}$$

The number of check nodes in $\mathcal{G}^n$ is

$$M_n = 1 \; + \; (C-1) \sum_{j=0}^{n-1} (C-1)^j \, (K-1)^j. \tag{F.2}$$

The probability of cycle not emerging in the generation g_{n+1} is then :

$$\mathcal{P}\{\mathcal{G}^{n+1} \in \mathcal{T}^{n+1} \mid \mathcal{G}^n \in \mathcal{T}^n\} = \prod_{j=1}^{M_{n+1}-M_n} \left(1 - \frac{M_n + j - 1}{M-N}\right) \prod_{j=1}^{N_{n+1}-N_n} \left(1 - \frac{N_n + j - 1}{M}\right) \tag{F.3}$$

The probability that a cycle emerges in generation g_n is then:

$$\mathcal{P}(n) = \mathcal{P}\{\mathcal{G}^n \notin \mathcal{T}^n \mid \mathcal{G}^n \in \mathcal{T}^n\} \prod_{j}^{n-1} \mathcal{P}\{\mathcal{G}^j \in \mathcal{T}^j \mid \mathcal{G}^{j-1} \in \mathcal{T}^{j-1}\} \tag{F.4}$$

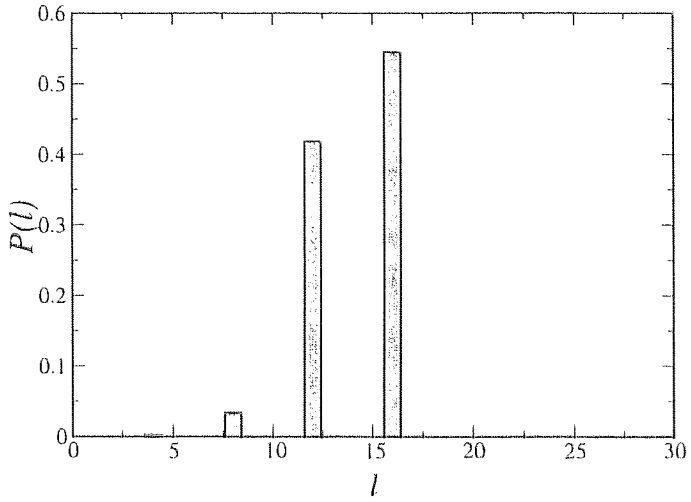


Figure F.1: $P(l)$ for $K = 5$, $M = 5000$, $M - N = 2000$.

A simpler lower bound for Equation (F.3) is:

$$\mathcal{P}\{\mathcal{G}^{n+1} \in \mathcal{T}^{n+1} \mid \mathcal{G}^n \in \mathcal{T}^n\} \geq \prod_{j=1}^{M_{n+1}} \left(1 - \frac{M_n + j - 1}{M - N}\right) \prod_{j=1}^{N_{n+1}} \left(1 - \frac{N_n + j - 1}{M}\right),$$

as $M_{n+1} - M_n \leq M_{n+1}$ and $N_{n+1} - N_n \leq N_{n+1}$.

Simplifying further:

$$\mathcal{P}\{\mathcal{G}^{n+1} \in \mathcal{T}^{n+1} \mid \mathcal{G}^n \in \mathcal{T}^n\} \geq \left(1 - \frac{2\,M_{n+1}}{M - N}\right)^{M_{n+1}} \left(1 - \frac{2\,N_{n+1}}{M}\right)^{N_{n+1}} \tag{F.5}$$

$$\geq 1 - \frac{2\,M_{n+1}^2}{M - N} - \frac{2\,N_{n+1}^2}{M} + \sum_{j+l>1}^{j=M_{n+1}, l=N_{n+1}} C_{M_{n+1},j}\, C_{N_{n+1},l} (-1)^{j+l} \left(\frac{2\,M_{n+1}}{M - N}\right)^j \left(\frac{2\,N_{n+1}}{M}\right)^l,$$

Where $C_{M_{n+1},j}$ is a combinatorial factor. Observing that the series in the left hand side has positive sign (an order M number of terms is needed to cancel the sign of the leading order positive term, but only an order 1 number of terms is available in the whole summation) one can write:

$$\mathcal{P}\{\mathcal{G}^{n+1} \in \mathcal{T}^{n+1} \mid \mathcal{G}^n \in \mathcal{T}^n\} \geq 1 - \frac{2}{\alpha} \frac{M_{n+1}^2 + \alpha N_{n+1}^2}{M}, \tag{F.6}$$

where $\alpha = 1 - M/N = 1 - R = C/K$. The probability that a cycle of any size emerges in the g_{n+1} generation is then:

$$\mathcal{P}\{\mathcal{G}^{n+1} \notin \mathcal{T}^{n+1} \mid \mathcal{G}^n \in \mathcal{T}^n\} \leq \frac{\phi(C, K, n)}{M}, \tag{F.7}$$

where

$$\phi(C, K, n) = 2\,\frac{K}{C}\left(M_{n+1}^2 + \alpha N_{n+1}^2\right). \tag{F.8}$$

Since n, C, K are of $\mathcal{O}(1)$, $\phi(C, K, n)$ is also of $\mathcal{O}(1)$ and the probability that a cycle emerges vanishes in the thermodynamical limit.

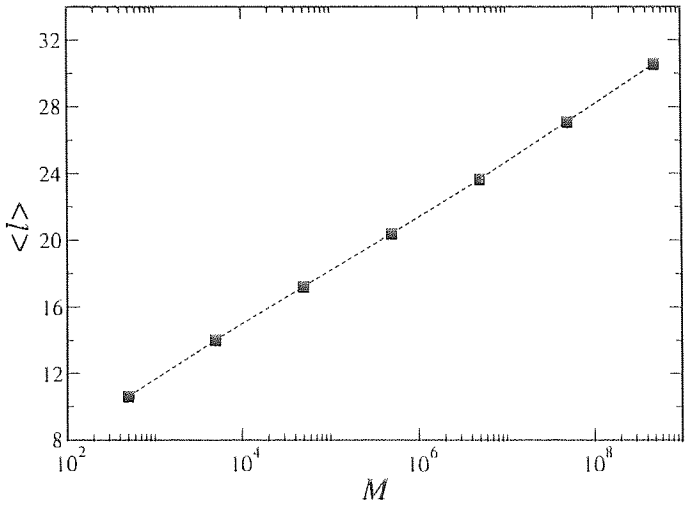


Figure F.2: Mean cycle-free length $\langle l \rangle$ as a function of the size M

Another upper bound can be obtained by using that

$$N_n, M_n \leq (1-R)^n n K^{2n},$$

resulting in

$$\mathcal{P}\{\mathcal{G}^{n+1} \notin \mathcal{T}^{n+1} \mid \mathcal{G}^n \in \mathcal{T}^n\} \leq \frac{2}{\alpha} \frac{(1-R)^{2n} n^2 K^{4n}(1+\alpha)}{M} \tag{F.9}$$

It can be deduced from (F.4) $\mathcal{P}(n) \leq \mathcal{P}\{\mathcal{G}^n \notin \mathcal{T}^n \mid \mathcal{G}^{n-1} \in \mathcal{T}^{n-1}\}$ and one can find the upper bound:

$$\mathcal{P}(n) < \frac{4n^2 K^{4n}}{M} \tag{F.10}$$

For large M one has $\mathcal{P}(n) \approx \mathcal{P}\{\mathcal{G}^n \notin \mathcal{T}^n \mid \mathcal{G}^{n-1} \in \mathcal{T}^{n-1}\}$. It can be seen in (F.10) that for large M only paths with length of at least order $\ln M$ can have finite probability.

The length of a path crossing n generations is $l = 4n$, the upper bound becomes:

$$\mathcal{P}(l) < \frac{l^2 K^l}{4M} \tag{F.11}$$

Equation (F.4) can be computed numerically. In Figure F.1 the distribution of path lengths obtained for a system of size $M = 5000$ with $K = 5$ and $N = 3000$ is shown. In Figure F.2 we show the mean path length $\langle l \rangle$ as a function of the size M.

Appendix G

Publications related to this thesis

The research reported in this thesis resulted in the following publications:

- Vicente, R., Saad, D. and Kabashima, Y., *Finite-connectivity systems as error-correcting codes*, Phys. Rev. E **60** (1999), 532-5366.
- Kabashima, Y., Murayama, T., Saad, D. and Vicente, R., *Regular and irregular Gallager-type error-correcting codes*, in Advances in Neural Information Processing Systems 12, edited by S. A. Solla, T. K. Leen and K. R. Müler (MIT Press, Cambridge, MA) (1999), pp. 271-278.
- Murayama, T., Kabashima, Y., Saad, D. and Vicente, R., *Statistical physics of regular low-density parity-check error-correcting codes*, Phys. Rev E **62**, (2000) 1577-1591.
- Vicente, R., Saad, D. and Kabashima, Y., *Statistical mechanics of irregular low-density parity-check codes*, J. Phys . A **33**, (2000) 6527-6542.
- Vicente, R., Saad, D. and Kabashima, Y., *Error-correcting code on a cactus: a solvable model*, Europhys. Lett. **51**, (2000) 698-704.
- Vicente, R., Saad, D. and Kabashima, Y., *Error-correcting codes on a Bethe-like lattice*, to appear in Advances in Neural Information Processing Systems 13.
- Saad, D., Kabashima, Y. and Vicente, R., *TAP for parity-check error-correcting codes*, to appear in Saad, D. and Opper, M. (eds.) Advanced Mean Field Methods. MIT Press.

Bibliography

[AL95] Amic, C. D. E., and Luck, J., *Zero-temperature error-correcting code for a binary symmetric channel*, J. Phys. A **28** (1995), 135–147.

[AM00] Aji, S., and McEliece, R., *The generalized distributive law*, IEEE Trans. Info. Theory **46** (2000), 325–343.

[Ber93] Berger, J., *Statistical decision theory and Bayesian analysis*, Springer-Verlag, New York, NY, 1993.

[BGT93] Berrou, C., Glavieux, A., and Thitimajshima, P., *Near Shannon limit error-correcting coding and decoding: Turbo codes*, in Proc. IEEE Int. Conf. Commun (ICC) (Geneva, Switzerland), 1993, pp. 1064–1070.

[BL82] Bowman, D., and Levin, K., *Spin-glass in the Bethe approximation: Insights and problems*, Phys. Rev. B **25** (1982), 3438–3441.

[CGH97] Castillo, E., Gutiérrez, J., and Hadi, A., *Expert systems and probabilistic network models*, Springer-Verlag, New York, NY, 1997.

[Che97] Cheng, J.-F., *Iterative decoding*, Ph.D. thesis, California Institute of Technology, March 1997.

[CS98] Conway, J., and Sloane, N., *Sphere packings, lattices and groups*, Springer-Verlag, New York, 1998.

[CT91] Cover, T., and Thomas, J., *Elements of information theory*, John Wiley & Sons, New York, NY, 1991.

[Dav98] Davey, M., *Record-breaking correction using low-density parity-check codes*, Hamilton prize essay, Gonville and Caius College, Cambridge, October 1998.

[Dav99] Davey, M., *Error-correction using low-density parity-check codes*, Ph.D. thesis, University of Cambridge, December 1999.

[Der81a] Derrida, B., *Random-energy model: An exactly solvable model of disordered systems*, Phys. Rev. B **24** (1981), 2613–2626.

[Der81b] Derrida, B., *Random Energy Model: An exactly solvable model of disordered systems*, Phys. Rev. B **24** (1981), 238–251.

[DW99] Dorlas, T., and Wedagedera, J., *Phase diagram of the Random Energy Model with higher order ferromagnetic term and error correcting codes due to Sourlas*, Phys. Rev. Lett. **83** (1999), 4441–4444.

[Fel50] Feller, W., *An introduction to probability theory and its applications*, vol. 1, John Wiley & Sons, New York, NY, 1950.

[FHS78] Fradkin, E., Huberman, B., and Shenker, S., *Gauge symmetries in random magnetic systems*, Phys. Rev. B **18** (1978), 4879–4814.

[FM98] Frey, B., and MacKay, D., *A revolution: Belief propagation in graphs with cycles*, Advances in Neural Information Processing Systems 10 (Jordan, M. I., Kearns, M. J., and Solla, S. A., eds.), The MIT Press, 1998.

[Fre98] Frey, B., *Graphical models for machine learning and digital communication.*, MIT Press, Cambridge, MA, 1998.

[Fre99] Freeman, Y. W. W., *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, 1999.

[Gal62] Gallager, R., *Low density parity check codes*, IRE Trans. Info. Theory **IT-8** (1962), 21–28.

[Gal63] Gallager, R., *Low-density parity-check codes, Research monograph series*, no. 21, MIT Press, Cambridge,MA, 1963.

[GM84] Gross, D., and Mezard, M., *The simplest spin glass*, Nuclear Pnysics B **240** (1984), 431–452.

[Gol91] Goldschmidt, Y., *Spin glass on the finite-connectivity lattice: The replica solution without replicas*, Phys. Rev. B **43** (1991), 8148–8152.

[Gol92] Goldenfeld, N., *Lectures on phase transitions and the renormalization group*, Perseus Books, Reading, MA, 1992.

[GR94] Gradshteyn, I., and Ryzhik, I., *Table of integrals, series and products*, Academic Press, London, 1994.

[Guj95] Gujrati, P., *Bethe or bethe-like lattice calculations are more reliable than conventional mean-field calculations*, Phys. Rev. Lett. **74** (1995), 809–812.

[Ham50] Hamming, R., *Error detecting and error correcting codes*, Bell. Sys. Tech. J. **26** (1950), 147–160.

[Hil86] Hill, R., *A first course in coding theory*, Claredon Press, Oxford, 1986.

[Huf52] Huffman, D., *A method for construction of minimum redundancy codes*, Proc. IRE **40** (1952), 1098–1101.

[Iba99] Iba, Y., *The Nishimori line and Bayesian statistics*, J. Phys. A **32** (1999), 3875–3888.

[Jen96] Jensen, F., *An introduction to Bayesian networks*, UCL Press, London, 1996.

[KF98] Kschischang, F., and Frey, B., *Iterative decoding of compound codes by probability propagation in graphical models*, IEEE J. Selected Areas in Commun. **2** (1998), 153–159.

[KFL98] Kschischang, F., Frey, B., and Loeliger, H.-A., *Factor graphs and the sum-product algorithm*, submitted to IEEE Trans. on Info. Theory, available at http://www.cs.toronto.edu/ frey/papers/publications.html, 1998.

[Khi57] Khinchin, A., *Mathematical foundations of information theory*, Dover Publications, Inc., New York, NY, 1957.

[KMS00] Kabashima, Y., Murayama, T., and Saad, D., *Typical performance of Gallager-type error-correcting codes*, Phys. Rev. Lett. **84** (2000), 1355–1358.

[KMSV00] Kabashima, Y., Murayama, T., Saad, D., and Vicente, R., *Regular and irregular Gallager-type error-correcting codes*, Advances in Neural Information Processing Systems 12 (Cambridge, MA) (Solla, S., Leen, T., and Müler, K., eds.), MIT Press, 2000, pp. 272–278.

[KS78] Kirkpatrick, S., and Sherrington, D., *Infinite-ranged models of spin-glasses*, Phys. Rev. B **17** (1978), 4384–4403.

[KS98] Kabashima, Y., and Saad, D., *Belief propagation vs. TAP for decoding corrupted messages*, Europhys. Lett. **44** (1998), 668–674.

[KS99a] Kabashima, Y., and Saad, D., *Statistical physics of error-correcting codes*, Europhys. Lett. **45** (1999), 97–103.

[KS99b] Kanter, I., and Saad, D., *Error-correcting codes that nearly saturate Shannon's bound*, Phys. Rev. Lett. **83** (1999), 2660–2663.

[KS00a] Kanter, I., and Saad, D., *Cascading parity-check error-correcting codes*, Phys. Rev. E **61** (2000), 2137–2140.

[KS00b] Kanter, I., and Saad, D., *Finite-size effects and error-free communication in Gaussian channels*, J. Phys. A **33** (2000), 1675–1681.

[Lau96] Lauritzen, S., *Graphical models*, Oxford University Press, New York, NY, 1996.

[LMSS98] Luby, M., Mitzenmacher, M., Shokrollahi, A., and Spielman, D., *Improved low-density parity-check codes using irregular graphs and belief propagation*, Tech. Report SRC 1998-009, Digital Systems Research Center, http://www.research.digital.com/SRC/, 1998.

[Mac95] MacKay, D., *Free energy minimization algorithm for decoding and cryptanalysis*, Electronics Letters **31** (1995), 446–447.

[Mac99] MacKay, D., *Good error-correcting codes based on very sparse matrices*, IEEE Trans. on Info. Theory **45** (1999), 399–431.

[Mac00a] MacKay, D., *Information theory, inference and learning algorithms*, 2000, available at http://wol.ra.phy.cam.ac.uk/mackay/.

[Mac00b] MacKay, D., *private communication*, 2000.

[Mac00c] MacKay, D., *Relationships between sparse graph codes*, IBIS, Japan, 2000.

[MKSV00] Murayama, T., Kabashima, Y., Saad, D., and Vicente, R., *Statistical physics of regular low-density parity-check error-correcting codes*, Phys. Rev. E **62** (2000), 1577–1591.

[MN95] MacKay, D., and Neal, R., *Good codes based on very sparse matrices*, Lecture Notes in Computer Science, vol. 1025, Spriger, Berlin, 1995, pp. 100–111.

[Mon98a] Monasson, R., *Optimization problems and replica symmetry breaking in finite connectivity spin glasses*, J. Phys. A **31** (1998), 513–529.

[Mon98b] Monasson, R., *Some remarks on hierarchical replica symmetry breaking in finite-connectivity systems*, Philos. Mag. B **77** (1998), 1515–1521.

[Mon00] Montanari, A., *Turbo codes: the phase transition*, cond-mat/0003218, March 2000.

[MPV87] Mezard, M., Parisi, G., and Virasoro, M., *Spin glass theory and beyond*, World Scientific Publishing Co., Singapore, 1987.

[MS77] MacWilliams, F., and Sloane, N., *The theory of error-correcting codes*, North-Holland, Amsterdam, 1977.

[MS99] Montanari, A., and Sourlas, N., *The statistical mechanics of turbo codes*, cond-mat/9909018, September 1999.

[MWD99] MacKay, D., Wilson, S., and Davey, M., *Comparison of constructions of irregular Gallager codes*, IEEE Trans. on Communications **47** (1999), 1449–1454.

[NG95] Nelson, M., and Gailly, J., *The data compression book*, M & T Books, New York, NY, 1995.

[Nis80] Nishimori, H., *Exact results and critical properties of the Ising model with competing interactions*, J. Phys. C **13** (1980), 4071–4076.

[Nis93] Nishimori, H., *Optimal decoding for error-correcting codes*, J. Phys. Soc. of Japan **62** (1993), 2973–2975.

[Nis00] Nishimori, H., *Theory of spin glasses and statistical mechanics of information*, 2000, available at http://www.stat.phys.titech.ac.jp/ nishi/index-e.html.

[NS00] Nishimori, H., and Sherrington, D., *Absence of replica symmetry breaking in a region of the phase diagram of the Ising spin glass*, cond-mat/0008139, August 2000.

[OO79] Omura, A., and Omura, J., *Principles of digital communication and coding*, McGraw-Hill Book Co., Singapore, 1979.

[Par80] Parisi, G., *The order parameter for spin glasses: a function on the interval 0-1*, J. Phys. A **13** (1980), 1101–1112.

[Pea88] Pearl, J., *Probabilistic reasoning in intelligent systems*, Morgan Kaufmann Publishers, Inc., San Francisco, CA, 1988.

[Ple82] Plefka, T., *Convergence condition of the TAP equation for the infinite-ranged ising spin glass model*, J. Phys. A **15** (1982), 1971–1978.

[RK92] Rieger, H., and Kirkpatrick, T., *Disordered p-spin interaction models on Husimi trees*, Phys. Rev. B **45** (1992), 9772–9777.

[RSU99] Richardson, T., , Shokrollahi, A., and Urbanke, R., *Design of provably good low-density parity check codes*, submitted to IEEE Trans. on Information Theory, available at http://www.lucent-inferno.com/cm/ms/who/amin/pub.html, April 1999.

[RU98] Richardson, T., and Urbanke, R., *The capacity of low-density parity check codes under message-passing decoding*, submitted to IEEE Trans. on Information Theory, available at http://netlib.lucent.com/cm/ms/former/tjr/pub.html, November 1998.

[Ruj93] Ruján, P., *Finite temperature error-correcting codes*, Phys. Rev. Lett. **70** (1993), 2968–2971.

[Saa98] Saakian, D., *Diluted generalized random energy model*, JETP Lett. **67** (1998), 440–444.

[Sha48] Shannon, C., *Mathematical theory of communication*, Bell. Sys. Tech. J. **27** (1948), (pt. I) 379–423 (pt. II) 623–656.

[SJ98] Saul, L., and Jordan, M., *Exploiting tractable substructures in intractable*, Advances in Neural Information Processing Systems 10 (Cambridge, MA) (Touretzky, D., Mozer, M., and Hasselmo, M. E., eds.), MIT Press, 1998, pp. 479–485.

[SO] Saad, D., and Opper, M., *Advanced mean field methods: Theory and practice*, MIT Press, Cambridge, MA, in press.

[Sou89] Sourlas, N., *Spin-glass models as error-correcting codes*, Nature **339** (1989), 693–695.

[Sou94a] Sourlas, N., *Spin-glasses, error-correcting codes and finite-temperature decoding*, Europhys. Lett. **25** (1994), 159–164.

[Sou94b] Sourlas, N., *Statistical mechanics and error-correcting codes*, From Statistical Physics to Statistical Inference and Back (Grassberger, P., and Nadal, J.-P., eds.), NATO ASI Series, vol. 428, Kluwer Academic Publishers, 1994, pp. 195–204.

[Tan00] Tanaka, T., *Information geometry of mean field approximation*, to appear in Neural Computation, 2000.

[Tou77] Toulouse, G., *Theory of the frustration effect in spin glasses: I*, Communications on Physics **2** (1977), 115–119.

[VB85] Viana, L., and Bray, A., *Phase diagrams for dilute spin glasses*, J. Phys. C **18** (1985), 3037–3051.

[VSK99] Vicente, R., Saad, D., and Kabashima, Y., *Finite-connectivity systems as error-correcting codes*, Phys. Rev. E **60** (1999), 5352–5366.

[VSK00a] Vicente, R., Saad, D., and Kabashima, Y., *Error-correcting code on a cactus: a solvable model*, Europhys. Lett. **51** (2000), 698–704.

[VSK00b] Vicente, R., Saad, D., and Kabashima, Y., *Error-correcting codes on a Bethe-like lattice, to appear in* Advances in Neural Information Processing Systems 13, 2000.

[VSK00c] Vicente, R., Saad, D., and Kabashima, Y., *Statistical mechanics of irregular low-density parity-check codes*, J. Phys. A **33** (2000), 6527–6542.

[Wei97] Weiss, Y., *Belief propagation and revision in networks with loops*, Tech. Report A.I. Memo 1616, MIT, 1997.

[Wib96] Wiberg, N., *Codes and decoding on general graphs*, Ph.D. thesis, Dep. of Electrical Engineering, Linköping University, 1996.

[WS87a] Wong, K., and Sherrington, D., *Graph bipartitioning and spin glasses on a random network of fixed finite valence*, J. Phys. A **20** (1987), L793–L799.

[WS87b] Wong, K., and Sherrington, D., *Graph bipartitioning and the Bethe spin-glass*, J. Phys. A **20** (1987), L785–L791.

[WS88] Wong, K., and Sherrington, D., *Intensively connected spin glasses: towards a replica-symmetry-breaking solution of the groud state*, J. Phys. A **21** (1988), L459–L466.